OPTIMISATION VERSUS CERTAINTY; DEVELOPING THE USE OF ECONOMIC EVALUATION FOR DECISION-MAKING

Warren Stevens

Thesis submitted to the Faculty of Science, the University of London for examination for the degree of Doctorate of Philosophy.

Department of Public Health & Policy
London School of Hygiene and Tropical Medicine
Keppel Street
London WC1E 7HT

June 2000

Most of all I would like to thank my mother, Sheila. Without her sacrifice this would not have been possible.

I would also like to thank my supervisors Charles and Sue, both of whom have become close friends despite rather than because they were my supervisors, I am sure. I would particularly like to thank Sue for her prodding and pointing and for Charles for not only showing me where I should be going all those years ago, but also providing the map.

To Jenny, my inspiration.

Abstract

This thesis assesses the methods used in economic evaluation, the relationship of economic evaluation to decision-making and investigates the possible limitations of economic evaluation as it is currently used to support policies aimed at maximising population health gain. It then evaluates alternative methods of analysing data from economic evaluations to better inform policy decisions.

The hypothesis of this thesis is that a greater use of subgroup analysis in policy decisions could potentially improve the efficiency of allocating scarce health care resources. This study aims to investigate the impact on population health gain and service cost-effectiveness of using subgroup analysis within defined parameters to derive and evaluate estimates of effect, and compare it to the more traditional methods of statistical inference.

Data from existing large trials are used to calculate cost-effectiveness ratios for the total study population and for subgroups. Total and subgroup estimates of cost-effectiveness are applied to patient populations through simulation, and outcomes predicted on the assumption that treatment decisions are guided by estimates derived from the trial. The distribution of cost-effectiveness ratios based on different rules for 'allowing' the use of subgroup analysis results is compared with the distribution of cost-effectiveness ratios based on aggregate analyses.

Results show that pre-selected subgroups can provide a stronger likelihood of maximising overall health gain. This thesis argues for optimisation in the use and interpretation of results rather than an over reliance on certainty and the resulting restriction on the use of available data. It concludes that under the scrutiny of a health care system for which the primary goal is health gain maximisation within resource constraints, policy decisions made using the results of subgroup analysis could result in a more efficient allocation of resources.

CONTENTS

Acknowledgements

My supervisors:
Charles Normand
Sue Langham

For the PRAIS UK database:
M Flather
J Collinson
A Bakhai

For the ECMO database:
T Roberts
M Mugford
D Elbourne
CE Normand
D Field
A Grant
C Harris
A Johnson

For their opinions and advice:
Colin Sanderson
Jessica Corner

List of tables

List of figures

1. Introduction

The goal of normative economics is the maximisation of the welfare of the population. Since there are not enough resources to provide all the medical care demanded, choices are inevitable. Within most health care systems the objective is to maximise social welfare in terms of total health benefits or outcomes, but the measurement of these outcomes and the decision of the priority given to the different calls upon these resources is where economic theory must meet practical application.

The aim of this thesis is to look at the methods used in economic evaluation, its relationship to decision-making in health care, to investigate possible limitations of economic evaluation as it is currently used in terms of supporting policies to maximise population health gain and to suggest better methods. The first section gives an overview of the three fields of interest;

- economic evaluation in health care and the relationship between research and decision-making
- the contrasting methods available to analyse economic data and their roots in probability theory
- Methods of data analysis that incorporate or accept the heterogeneity of the studied population

The second section looks in more detail into the theory behind subgroup analysis. It takes forward the probability theories outlined in the first section and explains the intrinsic logic behind subgroup analysis that leads into the conceptual framework of the thesis.

It then assesses the theoretical argument for the use of subgroup analysis and probability and outlines the potential gains that could be found when making resource allocation decisions in health care.

The third section compares subgroup analysis results against the results from the analysis of the data set as a whole. Initially concentrating on the relationship with cost and then with cost-effectiveness ratios. The value to decision-makers of a greater use of subgroup analysis is assessed by comparing the usefulness of evidence gained through subgroup analysis with those gained from traditional analysis of cost-effectiveness ratios. The thesis concludes with a discussion on the relevance of the results of this study to policy makers and what further work should be done to further clarify the role of subgroup analysis in decision-making.

2. Economic evaluation and decision-making in health care

2.1 Introduction

Recently there has been a growth in the number of clinical trials that have included a form of economic evaluation[1]. The aim has been to compare the costs and benefits of new treatments or prevention strategies with other more commonly used methods. The primary objective of trials has been the evaluation of both the safety and efficacy of treatment. However, consideration is now being given to other facets such as resource use. A world-wide increase in health care expenditure has been linked, among other things, to the emergence of new medical procedures. Several countries now require information on the cost-effectiveness of new drugs (e.g. Australia[2] and Canada[3]) and any new health care programme (e.g. in the Netherlands[4]) to ensure greater efficiency in the use of resources.

Closer to home, the recent introduction of the National Institute of Clinical Excellence (NICE) has been set up with the primary aim to ensure that new technologies are assessed for cost-effectiveness as well as clinical effectiveness. The primary purpose of NICE is summarised as follows;

> "NICE will produce clear guidance for clinicians about which treatments work best for which patients. It will assess new drugs, treatments and devices for their clinical and cost-effectiveness." [5]

The main reason for this shift has been the recognition of the fact that diffusion of many medical interventions takes place prior to the associated costs and benefits being

determined[6]. The early inclusion of economic analysis can provide useful information to assist in the rational diffusion of new innovations in health care[7]. Economics can be used in conjunction with clinical trials to help answer the question "should we do this?" rather than confining the conclusion of the trial to the question "does it work?" That is, in comparison with current or alternative treatments, do the marginal benefits of the new treatment outweigh any additional costs? The application of economics to clinical trials does not necessarily mean that less can or should be spent, but rather that the use of resources might be more efficient.

It is argued that clinicians should have an intrinsic responsibility to society as well as to the individual, and that these responsibilities should be incorporated into clinical decisions[8]. Most clinicians would agree that the extra money spent on achieving insignificant improvements in outcome is often not worthwhile. The extra resources used could have been used elsewhere in the health care system achieving greater benefit. These ideas come directly from an understanding of the limitations of finite resources and an aim of maximising health gain through the allocation of these resources. The term "health gain", in this thesis, is defined as a positive change in the status of the health of communities. It occurs if an intervention either initiates or accelerates an improvement in the health of a community, or if it prevents or delays deterioration in the health of a community.

The areas in which economics has the ability to have an impact on improving outcomes are first, to reach definitive conclusions on the most efficient use of health care resources, and second, to provide evidence to enable decision makers to set priorities in the

allocation of total resources between prevention and treatment. However, there is often a sizeable gap between what researchers consider to be good evidence and what decision-makers believe to be good evidence. This concept is not in any way confined to economic evaluation but encompasses clinical trials[9] and operational research[10].

Despite the enormous investment in research, which consumes around $55bn World-wide, every year[11], there is evidence to suggest that little of this research provides evidence which is then used in practice for either policy or clinical decisions[12,13]. Most of the literature[14,15,16] on this gap between research evidence and decision-making focuses on better use of and easier access to the information produced by research, and often overlooks issues raised more specifically by decision-makers, that of interpretation and generalisation of research evidence[17]. This chapter aims to give an overview of the problems of economic evaluation both in terms of its methodology and its relationship with policy and decision making.

2.2 The theory of economic evaluation

Economic evaluation is a two dimensional measurement of the effectiveness of an intervention, where an intervention is assessed by it's ability to produce positive benefits within the constraints of a set level of resources. Resources are usually presented as their equivalent in a monetary currency such as pounds or dollars. The type of economic evaluation will determine the method in which benefits are presented.

Economics looks at how best to allocate scarce, or limited resources to best satisfy unlimited or infinite demand. All resources are scarce, so no matter how much there is of

something it will eventually run out. Economists attempt to maximise aggregate benefit from within the restriction of limited resources. The theory of benefit maximisation is best described in terms of allocative, or Pareto efficiency.

Pareto efficiency was named after the economist who first presented this theory, Vilfredo Pareto. He stated that any redistribution of resources would improve efficiency if it made at least one person better off without making anyone else worse off. Society would, therefore, be most efficient where it is impossible to redistribute resources to make someone better off without simultaneously making someone else worse off. There could be a number of different permutations of resource allocation that resulted in Pareto efficiency, but they would all produce the exact same level of aggregate benefit. This is perfectly acceptable in theory but becomes difficult to measure in practice. For example, how is benefit defined? Do some things benefit some people more than others?

The concept is developed further by the compensation principle introduced by Kaldor[18] and Hicks[19]. It adds the concept of relative or net benefit, making the model more flexible. If a redistribution of resources results in gainers and losers, it is a move towards Pareto-efficiency only if the aggregate benefits of the gainers outweigh the costs of the losers. In principle the gainers can then compensate the losers while still achieving a net benefit. Kaldor and Hicks proposed that the measurement of benefit should be directly comparable with the costs of redistribution and, as such, benefits should be given a monetary value.

These are the roots of the economic evaluation method known as cost-benefit analysis (CBA), the aim of which is to measure the effect of changes in resource allocation in terms of the net benefit to society.

The other form of economic efficiency is technical efficiency. The definition of technical efficiency is that of a production process that utilises the minimum resources. In terms of health care, technical efficiency is achieved where a utilisation of resources produces a level of benefit that cannot be exceeded by utilising those resources in any other way. It is this view of efficiency that forms the basis of cost-effectiveness analysis, which does not attempt to measure benefits against cost, but generally aims to compare interventions with the same goal. Comparisons are made by fixing either the costs or the outcomes, and comparing the remaining variable.

There are limitations with this method in terms of comparing different strategies or intervention that may have different outcome measures. However, a version of cost-effectiveness analysis (CEA), known as cost-utility analysis (CUA) attempts to overcome this by using a common measure of outcome, so the relative cost-effectiveness of all interventions can be compared. In CUA the benefits are expressed in terms of quality-adjusted-life-years (QALYs), disability-adjusted-life years (DALYs) or, the more recently introduced healthy-year-equivalents (HYEs). These methods can be used to compare interventions with different outcomes.

Although CBA is the original method of economic appraisal, CEA is now preferred by decision makers because it doesn't require the translation of health consequences into

monetary units (i.e. quantifying the value of human life). However, there have been signs recently of a re-emergence of CBA[20], with the introduction of a number of new methods that have made the valuation of health consequences less problematic. CUA continues to be the fastest growing form of economic evaluation, although there is an ongoing debate about the use and calculation of the generic measure of benefit, be it QALYs, DALYs or HYEs.

The application of these methods is under continuous review, lacking any consensus of opinion[21]. The choice of method is often not straightforward. Options may be limited due to the type and quality of the information available, as well as possible restraints on time and resources. Although the ideal method can be identified it may not be possible to conduct given such restraints. The goal must therefore be to maximise what can be achieved and to choose the method most appropriate to the data.

A key issue which must also be considered before embarking on an economic evaluation of any kind is the economic viability of economic analysis, i.e. whether it is worthwhile spending resources on the evaluation. The quality or the availability of information for economic analysis may be such that it is not perceived efficient to conduct an economic evaluation due to the unreliability of the results. Ideally economic evaluation, with its associated costs, should only be undertaken if the benefits of improving efficiency in the use of health care resources outweigh these costs. Conversely it may not be worthwhile undertaking the evaluation because the evidence already exists to substantiate the cost-effectiveness.

2.3 Research evidence and decision-making

There has long been an underlying tension between decision-makers and researchers. Much of this tension comes from the belief that those doing research do not understand or appreciate the complexities involved in decision-making. No matter how impressive or definitive the research, it still has to be interpreted at a local level. The problem is due to the fact that researchers, who are trying to reach the widest audience, tend to generalise upward towards a generic or "average" result, whereas those who wish to take advantage of this evidence are trying to generalise downward within a specifically defined population. The result is a dichotomy of interest between both players.

Classical statistical techniques compound this problem by limiting results of trials to a simple right or wrong (or better or worse) outcome, across a general sample. There has been a growing movement towards alternative methods of statistical proof and the summarising of trial results for the benefit of decision-makers. Most of these have tended towards either greater sensitivity analysis or the use of alternative statistical techniques such as Bayesian theory or decision analysis (defined in detail in chapters 3 and 4). The advantages of Bayesian techniques lie in the fact that it is possible to incorporate additional information into the analysis of a particular study (such as specific patient characteristics). This makes it easier for the decision-maker to apply the results of generic research to specific situations and allows the formal inclusion of the heterogeneity of populations into analysis of data, something that cannot be done when using classical statistical techniques.

For example, consider a study that shows that intervention *"A"* leads to a lower rate of post-operative infection. The results using classical statistics will only show the mean difference in infection rate between those who received the intervention and those who didn't, along with a significance test to inform the decision-maker of the likelihood of the results of the trial being true. The decision-maker, however, is also aware that there are a number of other aspects of the treatment that a surgical patient can receive that would also affect the post-operative infection rate, for example, whether they received blood or not. In Bayesian analysis the relationship between the presence of a blood transfusion and post-operative infection rate would have been included in the analysis as prior information, thus the results of the analysis would be more appropriate for decision-makers.

It is not only in the clinical trial arena that there is a need to bridge the gap between research and policy, but also in the field of economic evaluation. There has been much progress in the methods used for economic evaluation in health care in recent years. The majority of the mathematical work has centred on improving the measurement of the statistical inference of cost-effectiveness ratios, rather than the search for alternative methods of presenting economic information and improving its application in decision making[22,23,24,25]. However, there is a growing body of research that has started to challenge the use of classical statistics to present economic data and a move towards methods which are more relevant for decision-making[26,27,28].

Claxton and Posnet[14], along with many others, express the view that a decision-analytic framework should be pursued when attempting to answer economic questions. The reason

for this view is that traditional trial design is not consistent with the concept of efficiency, whereas the decision-analytic approach is as well as with methods used to set priorities in service provision. Claxton and Posnet also state the advantage of incorporating "prior information" as being "handled consistently and open to criticism, alternative formulation and empirical testing," whereas, classical statistical techniques exclude the use of related or prior information in producing results.

Lilford et al[16] also supports the incorporation of prior information via the use of Bayesian statistical methods. They argue that conventional statistical testing is "an improper basis for making decisions because they dichotomise results according to whether they are or are not significant, and do not allow decision makers to take explicit account of additional information." Similar arguments are found in Hornberger et al[29], Van Hout et al[30] and McCloskey et al[31].

2.4 The relationship between economic evaluation and decision making

Evidence suggests that economic evaluations are seldom used as a tool for decision making in policy or management[32,33,34,35]. These studies concluded that the majority of economic evaluations have been done by researchers without prior consultation with policy makers, however, changes are beginning to take place. For example in the UK the Cochrane Collaboration and NICE are initiatives aimed at boosting the acceptance of evidence-based medicine in practice. In Australia a set of economic evaluations[36] have been done in collaboration with the appropriate decision-makers in health departments, which had a definitive impact on policy. This supports the theory that an increased

involvement with decision-makers would improve the impact economic appraisals have on health care policy.

The connection between research and policy has always been a tenuous one, and as economics is still a relatively new facet to the measurement of the effectiveness of health care interventions, the connection with policy is even more tenuous. Australia and Canada are the only countries to have affected an introduction of mandatory economic evaluation for new pharmaceutical products[37], however a number of other countries are now requesting information on economic evaluation in order to aid decisions such as Finland, Switzerland, Sweden and the UK.

A review[38] has cited how economic evaluation has grown in the past three decades. In the five years from 1970-74 there were just 56 published economic evaluations whereas in the period 1985-89 there were 718. The majority of these have been cost analyses and cost-effectiveness analyses. There is evidence, then, that the use of economic evaluation in policy and decision-making is starting to grow. Nevertheless there is still scepticism about the value of economic evaluation in health care, due, in part, to the wide variety in the quality of published economic evaluations. In response to this a number of guidelines on conducting and reporting economic evaluations (e.g. the BMJ guidelines) have been developed to help define appropriate methodological standards[39].

One of the major reasons cited by decision makers[11], for not using information from economic evaluation, is the inability of economists to present the results of their research in a way that is both agreeable and useable for people in the position of making decisions.

If the results of economic evaluations are to be considered in clinical decision-making, a common language must be found between clinicians and economists. Although economic evaluations have increased in both quantity and quality over the years, their acceptance and influence on decision-makers has not grown significantly. The question, therefore, is what forms of clinician/manager education and/or changes in presentation of methodologies and results are required? There have already been a number of developments that have aided the communication. Firstly there is the introduction of health outcomes as a measure of productivity or benefit. There have been a number of innovations in health outcome measures, for example quality adjusted life years, which combined the outcomes of mortality and morbidity into one measure[40]. Second, decision-making has started to move away from the benefit of the individual and towards being based on the benefits to society as a whole. This makes it easier to understand the need to compare benefits with costs.

In addition to the need of presenting economic evaluations in such a way that is both understandable and generalisable to decision-makers, there is also the more practical problem of integration of economic analysis into routine decision-making.

An article by John Hutton[41] recently highlighted the fact that economic evaluation is only done on a one-off basis to solve a specific problem and very rarely used routinely. Hutton drew parallels with the evolution of economic evaluation and Lewis Thomas' classification of medical technologies. Here he states that economic evaluation has moved on from being a non-technology to a half-way technology, but for it to become an

advanced technology it would have to become routine, and an accepted part of clinical decision-making.

2.5 Summary

Economic evaluation, despite the lack of direction of its methodologies and the difficulties of it being accepted in practice, has the potential to be a very valuable tool in helping to maximise health gain within the restraints of limited resources. However evidence suggests that the current methods of presenting results of economic evaluations are not consistent with this goal. Although there has been much talk of bringing economic information and clinical decision making closer together, there has been less evidence of it happening. A first step must be to shift the goals of economic evaluation in terms of the presentation of results closer to those that are relevant and practical to those making the decisions.

The relationship between research and policy is of vital importance, to both groups. They rely heavily on each other for their continued existence, yet often seem to have no common language. There is a growing belief among some that movement away from the strict rules of classical statistics and a wider acceptance of new techniques that take into account the issue of heterogeneity should be considered. These techniques have the advantage of incorporating the specific characteristics of both patient and population, methods such as Bayesian statistics, decision analysis and subgroup analysis could bring the information requirements of the policy maker and the research efforts of the economist closer together.

3. Incorporating heterogeneity into data analysis

3.1 Introduction

The aim of this thesis is to investigate the possibility that for the purpose of answering the question, "which health care interventions should be provided and for whom?" we need to focus more on the relative cost-effectiveness, both costs and outcomes, of health care services within different subgroups.

When policy makers decide what health care interventions should be funded they do so on a basis of knowledge of the demands of their population. Where a population has a high incidence of a certain disease investment in treatment of that disease may take priority over another. Despite the acknowledgement of the heterogeneity of populations there is little acceptance that evidence on the cost-effectiveness of these interventions for different groups within the population itself may be useful in making these decisions.

The purpose of the next two chapters is to look at the theory behind three methods most associated with analysing data sets which take into account the issue of heterogeneity. These are Bayes theory, decision analysis and subgroup analysis. Bayes theory brings the incorporation of prior evidence to the issue of heterogeneity; subgroup analysis concentrates mainly on the issue of heterogeneity alone, and decision analysis takes the complex issue of the non-homogenous effects of interventions into the realm of decision making.

Many of these alternative methods of analysis of clinical trial data have been the subject of widespread debate[42]. In order to assess the relative benefits of all three and to understand how they incorporate the issue of heterogeneity these methods and the theory underlying them will be reviewed in detail.

This chapter reviews the importance of heterogeneity and looks in more detail at the potential benefits and problems with subgroup analysis. Chapter 4 then goes onto review the theoretical concepts underpinning these alternative methods and assesses in more detail the techniques of Bayesian and decision analysis.

3.2 The potential benefits of incorporating heterogeneity into analysis

The benefits of incorporating heterogeneity when undertaking analysis of data is best shown graphically as in figures 3.1 and 3.2. Figure 3.1 shows the average cost-effectiveness ratios for a general population with the minimal data on the characteristics of the population and their impact on cost-effectiveness. Within the block where the resource constraint falls, a proportion of people who have an average cost-effectiveness ratio of "x" will forgo treatment even though some of the people who receive treatment will have a cost-effectiveness ratio lower than "x" and some who don't receive treatment will have a cost-effectiveness ratio greater than "x". Figure 3.2 demonstrates that with greater information on the characteristics of the population and the relationship between those characteristics and cost-effectiveness a clearer picture emerges of the proportion of the population likely to benefit most. This would result in a more efficient allocation of resources. The average cost-effectiveness ratio of those receiving the intervention in

Figure 3.1 would be "x", and the average cost-effectiveness ratio of those receiving the intervention in Figure 3.2 would be between "x" and "y", approximately (x +y) / 2.

This method could also help us to achieve a more efficient balance between different sectors of health care. For example let us assume that there are only two interventions in the prevention and treatment of colorectal cancer, screening and surgery. These interventions would have marginal cost-effectiveness ratio curves (see figure 3.3) and these curves would be made up of subgroups of the population and could be listed in order (see figure 3.4) to give a picture of the extent to which resources are better spent on surgery or screening.

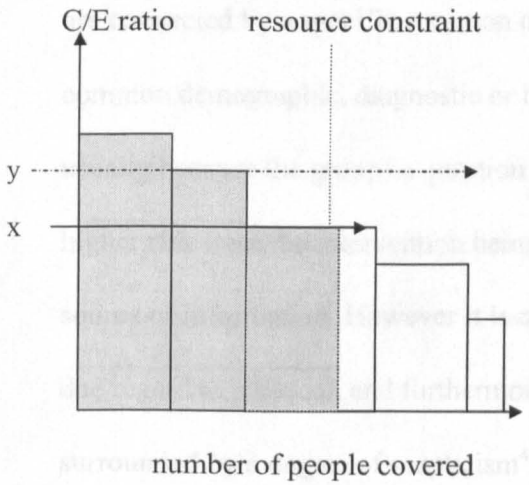


Figure 3.1

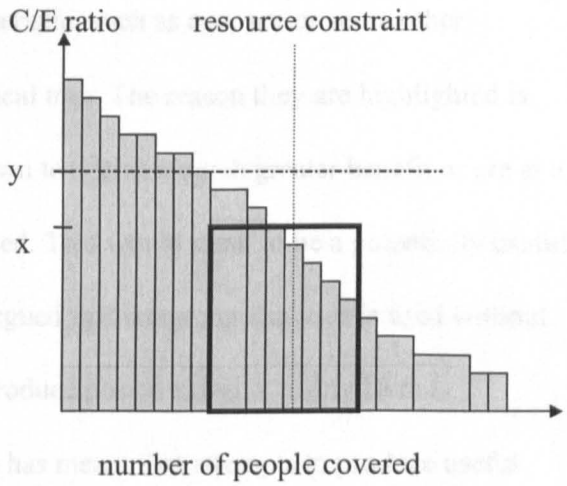


Figure 3.2

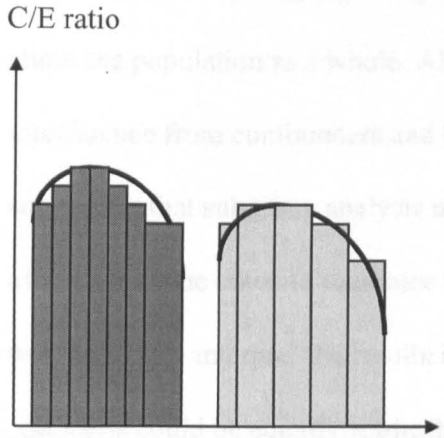


Figure 3.3

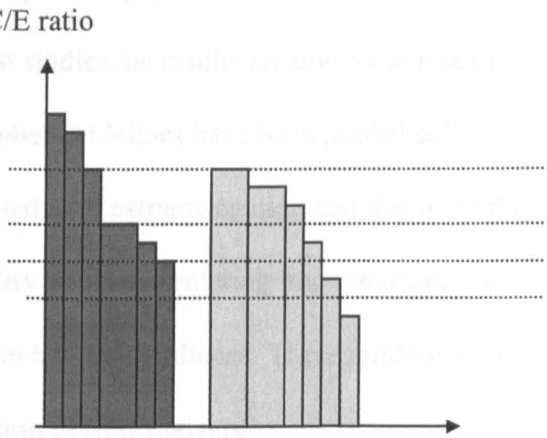


Figure 3.4

In many trials in addition to the main outcomes there are often other results based on one or a number of specific subsets of the population studied. These subsets, or "subgroups", are connected by a specific common characteristic, such as age, sex or some other common demographic, diagnostic or biological trait. The reason they are highlighted is usually because the group in question is shown to derive a much greater benefit or are at a higher risk from the intervention being studied. This would seem to be a potentially useful source of information. However it is often argued that subgroup analyses is used without due regard to protocol, and furthermore to produce positive results in any form is surrounded by a degree of scepticism[43]. This has meant that attempts to produce useful insights into the relative effectiveness of interventions across populations with different characteristics have been looked upon with distrust, regardless of their scientific strength or importance. In reality the identification and use of subgroups in this way may allow a more efficient use of resources.

Like most studies, subgroup analysis uses a sample of a population to draw conclusions about the population as a whole. Also like most studies the results are always at risk of interference from confounders and bias. A number guidelines have been published[44, 21,22] suggesting that subgroup analysis must be treated with extreme caution and that a number of steps must be taken to minimise the possibility of chance entering into the results, as well as how to interpret the results if you are the intended audience. These guidelines and checklists could be equally applicable to any form of trial analysis.

3.3 The potential pitfalls of subgroup analysis

It has been reported that subgroup analysis can be misleading, and that there has been a tendency to over emphasise the importance of subgroup analyses without attention paid to statistical rigour. However, the majority of criticism aimed at published subgroup analyses has been concerned with the potential dangers of misinforming decision-makers while avoiding a more constructive view that points out the poor practical applications of the method. Like any other form of analysis, statistical or otherwise, if undertaken without due regard to the rules by which it is drawn, the results will be unreliable.

For clinical trials good design requires elements of randomisation, masking, completeness of follow-up, and other methods associated with minimising both error and bias. A number of papers have attempted to summarise the problems with conducting and interpreting sub-group analyses[45,46,47]. Consensus moves us to list four main areas of concern. These are;

1) Statistical significance,

2) Clinically important difference,

3) Hypothesis preceding analysis,

4) Dose response and indirect evidence to support the findings.

Statistical significance is an important issue. Subgroup analyses will always include fewer patients than does the overall analysis and as such carries a greater risk of a type II error*. There are a number of techniques for conducting subgroup analyses that are suggested to reduce this such as the Bayesian method[48,49] or those used in decision analysis models. Statistical significance is, however, only a tool for assessing the likelihood of an observed event happening by chance. The possibility of the result being due to chance can be minimised, but never dismissed entirely. This leads onto the second point, the importance of observing a clinically important difference. No trial can control for everything and as such there will always be a measure of difference between different subgroups. It is only when the difference is large enough does the interaction warrants attention. There are no ground rules for size of effect, but generally the larger the difference between a particular subgroup and the overall effect, the more likely that it will be real.

The third factor is the issue of the hypothesis preceding the analysis. It is generally believed that sifting through data for possible interactions is bad science. There are a number of examples of studies finding "apparent" interactions in subgroups where none were found in the overall population. Nevertheless, the issue of whether a hypothesis preceded analysis is not always a cut and dried one. At one end of the spectrum you might have unexpected results in a subgroup detailed after the event that cannot be explained,

* where a relationship shown is more likely to be due to chance.

whereas at the other end, you may have a subgroup result clearly explained in a protocol beforehand that was suggested by previous research. Between these two extremes are an infinite number of combinations. It is normal to discourage conclusions to be drawn from an unexpected correlation, however, if they possess a degree of statistical power they may form a very useful hypothesis for future research.

This brings us to the last point, where results of subgroup analyses may be strengthened by a dose-response or by indirect evidence such as known human biology or behaviour. An example may be where an intervention has minimal benefits across a population, but in older age groups its effects are more pronounced. The effects of the lower age groups may have pulled down the overall average. The evidence could be supported if there was a linear trend in dose-response across the age groups or if there was a physiological or psychological reason why the intervention would be more likely to benefit older people.

3.4 Summary

When studying a multidimensional subject such as a population, there are inevitably issues surrounding the heterogeneity of that sample. It is not under debate that this heterogeneity exists or that there are groups within the population defined by certain characteristics that will mean that a given health care intervention is more or less likely to be cost-effective for them. What is an issue of conjecture is whether this information and these relationships are attainable and whether knowledge of them can be used to improve efficiency in resource allocation.

Subgroup analysis or the measurement of the relative cost-effectiveness of groups within a population, rather than the population as a whole can theoretically improve the efficiency of health care systems. Despite the concern surrounding the methods employed to undertake subgroup analysis, most of these can be overcome with a concentration on good science and the use of accepted methods of measurement.

Having given an overview of the importance of heterogeneity and the value of subgroup analysis, the next chapter will review the theory behind the two more mainstream methods associated with heterogeneity-friendly data analysis, Bayes and decision analysis.

4. Probability, inference and decision analysis

4.1 Introduction

To understand where the heterogeneous methods differ from the more common classical analysis, we must go back to the roots of inference with a description of probability theory and its relation to the various theories of inference and decision analysis. Once we have described the varying theories of probability and their perceived strengths and weaknesses, the mainstream methods of Bayesian inference and decision analysis will be reviewed.

This chapter then goes on to provide a critical analysis of both of these methods in comparison to subgroup analysis. It will look at the strengths and weaknesses of these methods both in terms of the information it provides to decision makers and the acceptance of the tools it uses to achieve these results.

4.2 Theories of Probability.

All theories of statistical inference are rooted in probability theory. Looked at simplistically, it could be said that each different classification of probability theory has led to a theory of statistical inference. As will become apparent, this is not necessarily so. There are no fixed boundaries in terms of belief of one particular theory of probability that automatically disqualifies you from any one theory of statistical inference, but there are some obvious common links that obviate likelihood of one over that of another.

There are basically four theories of probability; the classical, the frequentist, the logical or objective, and the subjective or personalistic theory.

All theories of probability have a number of simple properties. The most obvious of these is that if we have a set of events that are mutually exclusive and which include all the possible events in the given circumstances then the sum of the probabilities of these events happening will be one. The set of these probabilities makes up a probability distribution. Of the four theories of probability, the oldest and simplest is the classical theory of probability.

4.2.1 Classical probability

The first recognisable work on probability was done in the 1600s and centred itself on a framework of "equally likely" outcomes. At this stage there was no attempt to define probability, although some years later Laplaces'[50] classical definition of *probability as the ratio of the number of outcomes favourable to the event to the total number of possible outcomes, each assumed to be equally likely* became the accepted definition. Although forming the roots of modern day thoughts and theories of probability the classical view of probability is generally thought not to stand up to modern methodological standards, mainly due to it's reliance on the concept of "equally likely" outcomes[51].

The definition of "equally likely" is thought to come from the symmetry or homogeneity of the experimental situation. If a two-sided coin is spun, what reasons are there for one side to be favoured over the other? Are the outcomes, therefore, equally likely? This is known as the "principle of cogent reason". The difficulty arises when we study the basis of such reason, for on what basis do the physical properties of such objects imply equal

probability? The basis for the reasoning seems to rely on some prior knowledge about the effect of these physical properties, and at this point the theory becomes somewhat circular. What is really being considered is a theoretical coin that approximates to our chosen coin.

The other attitude to this concept is the "principle of insufficient reason", which argues that if we have no good reason to believe that either side of the coin is more likely to fall face-up, then we must assume that both sides are equally likely to appear. This then becomes a subjective argument based on the level of information we may or may not have about the properties of the coins.

These restrictions aside, the concept of equally likely outcomes can conceivably be used in only a small number of basic problems. Most attempts to use classical probability theory in more widely defined terms have tended to move more towards introduction of the concept of "frequency" probability theory, with the connection of the principle of cogent reason.

4.2.2 Frequency probability theory

A frequency theory of probability started to arise from the attempts to study more diverse topics with the more limited classical theory. Venn[52] first formally defined it as expressing "probability in terms of the limiting values of relative frequencies in infinite sequences of repeatable situations". The mathematical basis for the frequency theory was

not developed until much later by von Mises[53], who, rejecting the limitations of the classical standpoint, saw frequentist probability as

> "the rational concept of probability...(applying) only to problems which either the same event repeats itself again and again, or a great number of uniform elements are involved at the same time. In order to apply (this) theory of probability we must have a practically unlimited sequence of uniform observations."

In essence the restriction of the use of frequency theory seems to be that within the collective or group of identical events being studied, along with the condition of uniformity is a condition that requires the properties of these events are not restricted by their position within the collective. This means that if a selection or subset of those events were selected according to some specific rule, the properties of those events would tend towards the properties of the whole. The basis of this limitation von Mises calls the principle of randomness[54].

It is the inclusion of this concept that differentiates frequency theory from that of subjective probability theory, as in the former an event is not specific but typical. That is, the probability of an event is the probability that the coin will fall heads face-up at any observable occasion. Whereas the subjective viewpoint takes a more personal view of each event, looking to assess the probability of the coin coming up heads on the 10th or 11th toss.

It is may be suggested by subjective probability theorists that like classical probability theory, frequency probability theory has implied limitations to where it can be used. It's use and wide acceptance, as the basis for much of the statistical methods used in population sciences, tends to suggest that these limitations are not very important.

The views range from the admission that situations that do not encompass repetition of events of uniformity are not within the realm of vigorous statistical enquiry. Whereas other go so far as to suggest that this concept itself makes frequentist theory untenable in the study of complex bodies such as populations. It all depends on your viewpoint on the uniformity of the events in question, and how important any amount of lack of uniformity is to the outcomes being studied. No doubt any collective being studied using these concepts could be acceptably defined as uniform or heterogeneous, depending on the viewpoint of the discussant.

4.2.3 Logical probability theory

The logical theory is thought to be somewhat of a bridge between the frequentists and the subjective probability theorists with an objective and typical approach to the collective, but also with an aim to incorporate the degree-of-belief approach. The logical view of probability stands in contrast to the classical or frequentist views of probability, in that it is not designed to be applied to everyday problems of uncertainty. Logical probability is the implication of a certain piece or set of information that makes up evidence, E, to a likely outcome, O. It is the degree to which E implies O, what has been called the "credibility" of O, given E.

The background of logical probability is said to come from an intention to understand the formal structure of probability, rather than to formulate an application to problem solving, while others have used logical probability in conjunction with frequency theory, arguing that the different theories have their roles in distinct circumstances.

With their insistence on the fact that there is no need for numerical values to represent these probabilities, without a great deal of justification, the main criticism of the logical view is precisely that. That is the method of calculating or measuring an appropriate rational degree-of-belief. Subjectivists go further, in criticising the concept of uniqueness of these probabilities. They state that probability will always relate to the individual, as inherent within that individual is a set of information and body of evidence that is unique.

4.2.4 Subjective probability theory

This viewpoint is an important part of subjective probability theory, in that it's most prominent concept and that which holds it apart from the other theories is its concentration on the personal, and it's insistence on the heterogeneity of these persons. It is essentially a personal degree-of-belief model, but can be used to assess probability just as well. The manufacturing of a probability score is in reality simply the representation of someone's degree of belief. In each case subjective probability measures the probability of an event based on conditional information specific to said individual. It contrasts with the logical view only in that it supposes no common rational thread of common circumstance that weighs the subjectivity independent of the individual.

In attempts to quantify a formal theory of subjective probability, the emphasis has been on expressing the probability of an event through betting situations. It is the price an individual would be willing to pay for an amount of money on condition of an event, E, occurring. For example if a coin was thought to be unbiased, one half of that unit would be the maximum that individual would pay in exchange for one unit of that money conditional on the coin landing head-up. Where subjective probability differs is in that the conditions and the probability considered for that toss of the coin are specific only to that toss. If the coin falls head-up what will the individual be willing to 'bet' on a head for the next toss? This is a separate event and as such does not use the same conditions and information as the original toss of the coin.

There has been movement away from the bet analogy in subjective probability theory, as the rational effect of the size of stake became embroiled in criticism of the theory. The concept is irrelevant, but nevertheless was thought to detract from the purpose of the analogy. The concept of comparison of propositions was used first by Ramsey[55]. Here the approach uses an expression of personal probabilities for two events, a head on a toss of a coin and the likelihood of rain before Thursday. If the individual has no reason to believe that the coin is biased, the degree-of-belief of the individual concerning the outcome of imminent precipitation can be measured relative to that of the coin toss. This method does not completely overcome the issue of the need for measures of relative utility connected with different outcomes or events, but subjective theorists argue that this is irrelevant to the proposal of the underlying concept of individual subjectivity in probability. Nevertheless, most work on subjective probability has tended to combine subjective

probability theory with utility to understand the behaviour of individuals in circumstances of uncertainty.

4.3 Probability theory and statistical inference

The evolution of these different theories discussed above can in part be explained by the reason for their original contemplation. Taking the two most commonly considered of the two theories to simplify the argument. The frequentist views came primarily from wanting to understand and solve problems associated with uncertainty. A rational, measurement-based approach to the problem caused by uncertainty and how to overcome them, prior to making the decision. The subjective approach on the other hand comes from the act of decision making itself. A study of why decisions are made, the reasons why individuals make the choices they do and why different individuals make different decisions. Studied after the event of the decision itself[56].

At first glance the former methods seem the more rational and the more easily quantified. These are important factors when the consideration is how best to judge the effectiveness of health care interventions. Is what is required here certainty, a method based on good science and rationality? Almost certainly the answer is yes, but these methods are also marked by a lack of flexibility.

The frequency theory is a straightforward one. If a coin is completely unbiased and is tossed an infinitesimal number of times, half of the time the coin will land head up. There is no need for subjectivity here. There is, nevertheless, a reliance on the coin being

unbiased, a prior conception based on the theoretical nature of the idea, and due in no small part to the construction of theory at a point before the decision takes place. How easy is it to ensure an unbiased coin? How many foibles does the average coin possess? Is it more than the hypothetical, unbiased, coin?

If a coin were found to be more aerodynamic on one side (heads) than the other (tails), then while in the air it would spend fractionally more time head up, with the less aerodynamic side (tails) fighting against air resistance, than tail up. This would lead to an inevitable higher likelihood of landing in that position. Head up. If we accept this to be true, then there is additional information that can be drawn on to come to a more accurate conclusion of the probability of a 'heads' outcome for any one coin.

It is these foibles or imperfections, both in coins and in people, which give us reason to consider the use of subjectivity in the study of the effectiveness of health care interventions. The knowledge that an average person studied will gain four additional quality-adjusted life years at a cost of £20,000 is useful information in the decision on how to treat all the average people in the population. The question is, how useful is this information in deciding how to treat the rest of the population?

There is a similar dichotomy when taking these theories forward into the world of statistical inference. This separation is caused by the need to use information to make decisions under circumstances of uncertainty. Where data is used to inform or to describe a situation is by definition inferential. Whereas where the aim is to go one step further and use these data to inform a decision under uncertainty, this goes beyond simple statistical

inference and into statistical decision making. Although most commentators believe the functions to be inter-linked or at least reliant[57,58], there is still no consensus on where, if at all, subjectivity becomes relevant.

The **classical** approach to inference is designed specifically for the reporting of data alone. It does not attempt to enable decision-making. Neither does it consider prior information or the consequences or costs. This additional or prior information may affect the decisions on what to test using the classical inference approach, but this informality is thought to be imprecise and lacking a scientific basis, the exact criticism classicists have about the more subjective theories. This said, it is widely accepted that were there instances where there was no prior information on the subject at stake, then there is no need for subjectivity. The problem is, how much is there that we can say we know absolutely nothing about.

With the **Bayesian** or **decision-analytic** approaches being rooted in subjective probability, sample data is combined with prior information and utility measures respectively. Thus making results immediately attributable to a probability interpretation. The problem arises in the quantification of this information. Sometimes this information is easily quantified while other times it is less so. Theoretically purist subjectivists would argue that the uniqueness and complexity of individual cases within each sample would make it almost impossible to apply universal affects of any prior information or utility values. The virtues of this approach are weighted on reliance of both the quality of the information available and its relationship to the outcomes studied.

The only pure decision making approach is the decision-analytic approach, or decision theory. This method again uses both sample data and prior information but also incorporates the quantification of losses and gains. Whereas the other two systems aim to show only the likelihood, probability or risk of events, the decision theory model goes one step further and attempts to answer the question of what to do rather than just giving you information on likely outcomes.

In essence the movement between these three methods is an increase in risk and belief in the ability to quantify the previously unquantifiable. The classical approach is ruled by certainty. Information that is subjective, or even directly attributable, to data being studied is ignored either because of a lack of belief in the stated relationship or scientific obstinence against quantifying these relationships. The Bayesian accepts the existence of other information that helps to explain the behaviour of the studied variables and attempts to quantify it. Similarly decision theory tells us that as we can quantify external factors influencing our chosen variables, we can also quantify the importance of the various outcomes.

This movement away from certainty and towards taking a greater risk with the data in an attempt to reach more truly representative outcomes is not just an issue of scientific rigour, but a moral one. A debate between the gains of improving the efficient use of resources and taking risks with outcomes that are thought of as "life-dependent".

Having described and discussed the various theories on probability and how they relate to the movement towards classical or heterogeneous methods of analysis, we now describe the methods of both Bayesian inference and Decision Analysis.

4.4 Bayes Theory

Bayesian statistical inference was the earliest, and is still the accepted, method of calculation of subjective or additional information in inference. The statement of Bayes theorem comes directly from a statement of probability, where A_i is a collection of events, of which one will happen, and Y is another event. Taking the multiplication rule of probability theory, the probability of A_i given Y is the probability of both events divided by the probability of the event of Y, or:

$$P(A_i|Y) = \frac{P(A_i \& Y)}{P(Y)} = \frac{P(Y|A_i)P(A_i)}{P(Y)}$$

the probability of events A_i and Y can be shown as the probability of Y given A_i times the probability of A_i. For a number of events j this is shown as:

$$P(A_i|Y) = \frac{P(Y|A_i)P(A_i)}{\sum_j P(Y|Aj)P(A_j)}$$

None of this, as theory, is seen as particularly contentious by frequentist or subjectivist alike. The area of disagreement is it's use in practice and the ability to scientifically quantify A_i as a statement of belief or as a hypothesis about the relationship between the

subjector and the subjected. Thus Bayesian inference involves making subjective statements about probability and quantifying them in a way classical inference and frequentist probability theorists believe to be unscientific. Their view being that subjectivity is an issue outside inference that must be used in addition to the results of classical inference in making decisions. The opposing view of the Bayesian is that it is in fact unscientific to ignore the potential effect of subjective information at this stage, as any results would be misleading.

Bayesian analysis uses what it calls **priors** as an addition to sample data. Let us go back to our coin, not the hypothetical one, but the real one (the one with the foibles), and we assume that we believe that the curving of one of the surface edges discussed earlier gives the coin an increased probability of landing heads up. We also quantify as the likelihood of this as saying that it will land head-up 75% of the time. I will come back to the issue of quantification later. We can also state that we are 95% certain of this effect. We can then state the probability of a heads outcome at this stage as:

$P(h = 0.75) = 0.95$

$P(h = 0.50) = 0.05$

The next step is to undertake our trial. We toss the coin one hundred times and it lands head-up sixty times of that hundred. Our sample data on it's own suggests to us that P = 0.6. Combining our prior information and our sample data we can produce a third version of our probability of heads, known as the **posterior**. This combines our prior knowledge

of the likelihood of the event with the information we have gathered on this from our sample. Writing the equation in the form,

$$P(A_i|Y) \propto P(Y|A_i)P(A_i)$$

(where the proportionality applies over different choices of i, and the denominator is a constant independent of i)

Our priors were, $P(h = 0.75) = 0.95$

$P(h = 0.50) = 0.05$

Our likelihoods are, $f(Y \mid h = 0.75) = 0.75^6 * 0.25^3 = 0.0028$

$f(Y \mid h = 0.50) = 0.50^6 * 0.50^3 = 0.0019$

Our posteriors become, $P(h = 0.75 \mid Y) \propto 0.95 * 0.0028 = 0.00266$

$P(h = 0.50 \mid Y) \propto 0.05 * 0.0019 = 0.000095$

The denominator in Bayes is the sum of these, 0.002755, so the correct posteriors are,

$P(h = 0.75 \mid Y) = 0.00266 / 0.002755 = 0.97$

$P(h = 0.50 \mid Y) = 0.000095 / 0.002755 = 0.03$

We are now even more certain that the curving of the surface has an effect on the likelihood of the coin landing head up. The measure of our belief has risen from 95% to 97%. Realistically we would not limit our data to point estimates but have distributions

around a mean. For example P_h has a Normal distribution with mean of 0.75 and standard deviation of 0.01, similarly the prior and hence the posterior will possess these attributes.

The weakest part of Bayesian theory lies not in its use of subjective information within the formula of inference but in the quantification of such information. When the basis of the formula relies on the level of trust or knowledge invested within one individual, it loses a degree of certainty. The question for debate is whether this loss of certainty outweighs the benefits in terms of added information brought to bare on the task of answering the question in hand. Classicists say that it must (or that at least it is impossible to say), Bayesians that it does not.

4.4.1 Quantifying prior information

For simplicity, let us assume an economic evaluation has been undertaken to compare two interventions, A and B, which show mean ICERs of μ_A and μ_B, with variance σ^2. An additional question is being asked of the data, regarding the difference of effectiveness within two distinct subgroups. Let us say that a proportion P_{1A} receives treatment A and P_{1B} receive treatment B in subgroup 1. Denote the observed difference in means between treatment A and B in stratum 1 as d_1, where d_1 is the sample estimate of Δ_1, the unknown mean difference in effectiveness for these patients. Denote the same for subgroup B, where $P_{2A} = 1 - P_{1A}$ and $P_{2B} = 1 - P_{1B}$, as shown in table 4.1.

We now use the Bayesian method to incorporate our previous evidence of the effect of our chosen subgroup strata 1 and 2, what we will call ρ and σ_ρ^2, and attempt to estimate

the unknown effect in subgroup 1. This estimate, denoted by M_1, is the posterior mean of Δ_1.

Table 4.1: Samples and mean differences of subgroups

Subgroups	Number of patients		Observed mean difference	Unknown mean difference
	Treatment A	Treatment B		
1	nP_{1A}	nP_{1B}	d_1	Δ_1
2	nP_{2A}	nP_{2B}	d_2	Δ_2
Total	n	n		

Firstly we calculate the quantities which measure the precision of the sample information in each stratum relative to the corresponding amount of prior information. These quantities are given as,

$r_i = \sigma_\rho^2 / Var(d_i), i = 1,2.$ Where $Var(d_i)$ denotes the sampling variance of d_i. Thus large values of r_1 and r_2 imply that there is relatively little prior information available to the investigator as compared to the information supplied by the sample data. From table 1, $Var(d_i)$ can be calculated as,

$$Var(d_i) = \sigma^2\left(\frac{1}{nP_{iA}} + \frac{1}{nP_{iB}}\right), i = 1,2,$$

where σ^2 may be estimated from the sample observations. The posterior mean can now be derived[59] as,

$$M_1 = \frac{r_1[r_2(1-\rho^2)+1]d_1 + r_2\rho d_2}{\{[r_1(1-\rho^2)+1][r_2(1-\rho^2)+1] - \rho^2\}/(1-\rho^2)}$$

Thus M_1 is a linear function of d_1 and d_2, whose coefficients depend on the values of r_1, r_2 and ρ, and can be interpreted as the most likely value of Δ_1 given all the available information. At $\rho = 0$, $M_1 = (d_1 r_1)/(r_1+1)$ and thus, as would be expected with

independent subgroups, the value of d_2 has no effect on d_1 or Δ_1. A small value of r_1, that is strong prior evidence to suggest that a large value of Δ_1 is unlikely, will tend to pull back the estimated effect towards zero. Alternatively as ρ tends towards unity, it can be shown that,

$$M_1 = \frac{r_1 d_1 + r_2 d_2}{r_1 + r_2 + 1}$$

Therefore in the limiting case where Δ_1 and Δ_2 are regarded as indistinguishable, d_1 and d_2 are weighted by the values of r_1 and r_2, respectively, although strong prior evidence of a lack of difference will tend to pull back the estimate, M_1, towards zero. Furthermore by altering the values of ρ and σ_ρ^2 in M_1 to represent differing degrees of belief, the relationship between the strength of prior belief and various levels of confidence intervals can be explored.

To describe the use of this method here, we have used a continuous response variable with only two strata of subgroups. For dichotomous variables, the normal approximation to the binomial distribution may be used if the number of cases in each stratum is relatively high. For the case of more than two strata, Donner[60] has derived general expressions for the stratum specific posterior means $M_1, M_2, \ldots M_k$ of a normally distributed response variable under the assumption of a k-dimensional multivariate prior. Moreover the results here may be generalised to include different prior variances, non-zero means and other distribution shapes[61].

4.5 Decision theory

The use of subjective probability and the incorporation of such information into a decision-making tool rather than simply for inference is only tackled in decision theory. The primary distinction of decision theory is of course that it aims not only to give best evidence or merely to inform, but also to direct towards a best course of action under particular circumstances. As a result the expression of results of decision theory are not expressed in probability terms, as in classical and Bayesian methods. Decision theory prescribes a course, or courses, of action based on average benefit or loss. A disadvantage being that there is no recognition within the final result that the relative advantages of one over another may vary in relation to the probability of it occurring. This is a problem that will be discussed in more detail later. At the same time there is an obvious advantage to a mechanism designed specifically for prescription of action in the face of uncertainty, rather than simply minimising uncertainty on the issue and leaving the decision unmade.

To illustrate the basic formula of decision theory, the example of a cancer-screening programme is used. Apologies are made for any simplification of the issues, but these will be explained and justified as we progress. People are screened for presence of colorectal carcinoma using a faecal occult blood test (FOBT). These people then present with either a negative or positive result. We are interested only in those who present with a positive result.

Looking at the problem under the framework of decision theory, there are two possibilities, or what is known in decision theory as two possible "states of nature"[62]. These are that the individual has a presence of small polyps, that may or may not lead to

cancer, or that the presence of blood in the stool is unrelated to the presence of cancer or polyps. We therefore have two actions open to us. One is to repeat the FOBT test at another time. This assures us that the reason for the positive result is more likely to be cancer-related, or to refer the individual straight away for a colonoscopy, an invasive procedure that can detect the presence of polyps and tumours as well as being able to remove polyps. If we denote actions by a and states of nature by θ then we have:

a_1 = second FOBT,

a_2 = colonoscopy,

θ_1 = no presence of tumour, and

θ_2 = presence of tumour.

Dependent on the different states of nature, the actions will have different implications. If we limit these implications to cost, ignoring the ethics of subjecting a host of healthy individuals to a colonoscopy. We can produce a cost table, or what is known as a "table of losses", that covers each eventuality. Let us assume that an FOBT costs £200 and a colonoscopy costs £500.

Table 4.2: Table of losses

	θ_1	θ_2
a_1	£200	£700
a_2	£500	£500

If we knew θ for every individual, there would be no decision problem, but it is this very 'state of nature' that is not known. Although we do not know θ for each individual, we do have information of the specificity of the single FOBT. Evidence suggests that only 30% of those with an initial positive result on the first FOBT will have a tumour. Therefore we have some prior probability for θ_1 and therefore for θ_2. With this information we can now calculate the average loss from taking action a_1 and a_2;

$a_1 = (200 * 0.7) + (700 * 0.3) = £350$, and

$a_2 = (500 * 0.7) + (500 * 0.3) = £500$.

So in the long run it is worth giving all positives a second FOBT. The result of course relies heavily on the specificity of the test. If this were as high as 70%, then the loss table would point to a system that would mean lead us to give colonoscopies to everyone. If this was the case, and a group of people with irritable bowel syndrome came along, then under this form of decision theory, they will just have to suffer.

This is where the concept of the prior information and subgroups enables us to take this system further, and make it more reactive to a particular set of individuals. Let us consider other information on these individuals that we are presented with. For each of our number who has tested positive we have additional information on symptoms. Either they have;

x_1 = they also complained of pain when passing faeces,

x_2 = they also complained of diarrhoea, and

x_3 = no additional symptoms, other than the positive FOBT.

Such information gives us additional information on the relationship to the state of nature due to the likelihood principle. This could be known from past records, from published literature or from expert opinion. Suppose that from a mixture of these, we create a likelihood table:

Table 4.3: Likelihood table

	x_1	x_2	x_3
θ_1	0.1	0.4	0.5
θ_2	0.7	0.2	0.1

We can now use this additional information on x to help guide our decisions, by taking different decision rules or strategies. A decision rule, $\delta(x)$, determines what action to take given x. In our example there are eight possibilities:

Table 4.4: Decision rule table

	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7	δ_8
x_1	a_1	a_1	a_1	a_1	a_2	a_2	a_2	a_2
x_2	a_1	a_1	a_2	a_2	a_1	a_1	a_2	a_2
x_3	a_1	a_2	a_1	a_2	a_1	a_2	a_1	a_2

In terms of likelihood alone, x_1 favours θ_2, whereas x_2 and x_3 support θ_1. From the loss table, table 1, δ_5 seems to be intuitively correct as it reads a_2 a_1, a_1. Whereas δ_4 seems intuitively to be the worst. However, looked at with recourse to the rule of decision theory that states that we represent each decision as an average loss over all different values of x in each state of nature, we represent the decision rule of x, $\delta(x)$, as a pair of representing θ_1 and θ_2 calculated from the risk function;

$$R(\delta,\theta) = \sum_x L[\delta(x),\theta]p_\theta(x), \qquad \text{where } \theta = (\theta_1, \theta_2)$$

For example, for δ_5 we have $(R_{5,1}, R_{5,2} ....)$, where,

$$\begin{aligned} R(\delta_5, \theta_1) &= L(a_2, \theta_1)\, p_{\theta_1}(x_1) + L(a_1, \theta_1)\, p_{\theta_1}(x_2) + L(a_1, \theta_1)\, p_{\theta_1}(x_3) \\ &= 500 * 0.1 + 200 * 0.4 + 200 * 0.5 \\ &= \underline{230} \end{aligned}$$

This gives us a risk function table for all our decision rules, as follows;

Table 4.5: Risk function table

	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7	δ_8
θ_1	200	350	320	470	230	380	350	500
θ_2	700	680	660	640	560	540	520	500

Once calculated, we can see that no one decision gives us least cost for both states of nature, even though some can be immediately ruled out as having higher costs for both events. These are δ_2 δ_3 δ_4 and δ_6.

A principle that comes from game theory is often considered for making a final choice between the remaining decision strategies, known as the "minimax" principle. This aims to choose the decision where the maximum possible risk is minimised. This would leave us with δ_8, the choice to refer to colonoscopy whatever the circumstances, as at maximum risk this produces the lowest overall costs. This, though, is the pessimistic, slave to certainty attitude that with decision theory we are trying to escape. How can we logically base our actions on the most pessimistic outlook of an event, a point that is all likelihood at the tail-end of a probability distribution.

This brings us back to the use of prior information, the knowledge we have on the likelihood of θ, that is the specificity of our first FOBT. Using this in conjunction with our data on symptoms we can use the prior probability information to avoid both pessimism and over reliance on certainty. Using Bayes equation for posterior probability (page 38) we can create a measure of "posterior expected loss" for the decision rule δ, as follows,

$r(\delta,\pi) = R(\delta, \theta_1)\,\pi(\theta_1) + R(\delta, \theta_2)\,\pi(\theta_2)$, where $(\pi\theta)$ is prior probability of θ_1, θ_2

It seems reasonable, now to choose the so-called Bayes' decision rule (or Bayes' solution), where $r(\delta,\pi)$ is minimised. This is known as the Bayes' risk. Therefore

returning to the specificity of the FOBT at 30%, we have $\pi(\theta_1) = 0.7$ and $\pi(\theta_2) = 0.3$. This gives us values for Bayes' risk as below

Table 4.6: Bayes risk table

	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7	δ_8
r (δ,π)	350	449	422	541	**329**	428	401	500

Therefore the Bayes' solution is δ_5, which dictates that for those presenting with a positive FOBT and also complaining of pain when passing faeces, they should be referred directly for a colonoscopy, whereas those with no other symptoms, or those with diarrhoea should undertake an additional FOBT.

This gives us an average loss (or cost) of £329 compared to our minimax principle ruling of an average loss of £500 and the initial choice of universal action of a_1 of £350. The saving of the Bayes' solution of the initial decision is just £21. Therefore the transactions costs involved in acquiring the additional symptom information must have an average cost of less than £21 to make it the allocatively efficient choice.

So far we have limited our information; losses, probabilities, risks and priors as point estimates. Without great over complication it is possible to incorporate distributions as well as means for these variables[63]. This enables us to incorporate the strength of the data

being used as well as the shape of the distribution of each variable, which will lead to a more representative picture of risk and hence overall average loss.

4.6 Critical appraisal of heterogeneous methods

It is evident that both of these methods of analysis carry the same potential benefits of subgroup analysis, or indeed any other methods that incorporate heterogeneity, in that they tend towards optimisation and away from an over reliance on certainty and risk minimisation. Nevertheless it is only subgroup analysis as explained in the previous chapter that retains the objectivity of interpretation that both Bayes and decision analysis methods do not.

This is not a criticism of the methods themselves but more a criticism of the approach used. The difficulty with which the pro-Bayes and pro-Decision-analysis lobbies have had in incorporating these systems into mainstream economic evaluation is a sign of the mistrust of the scientific establishment of the use of subjective information. It may be too much to expect the acceptance of incorporating the opinions of experts and the independently constructed values of benefit, when the community has been brought up to base their decisions on nothing less than certainty and proof.

The difficulties with the more heterogeneous methods of analysis come from what many consider to be its strengths, the use of subjective or value judgements within the framework. Both Bayesian and decision theories attempt to take a further step towards answering the question that reduces the need for subjective interpretation of the data

policy decision makers are presented with. Bayesian data analysis takes uncertainty and tries to make it seem more certain. This is a laudable cause but it relies still heavily in both belief in the method being used, and a strong subjective element for which there will always be room for disagreement.

Decision theory has the strongest acceptance of the presence of heterogeneity, as it uses probability rather than inference and values rather than outcomes. Once again the movement towards a greater incorporation of the subjective, this time through incorporating the 'values' of different outcomes is a movement towards bridging the gap between research and decision making. Once again it relies on the acceptance by all concerned on the values it lays down in order to construct the answers it produces.

It is the belief of this thesis that the use of subgroup analysis framework for evaluation of cost-effectiveness data has the benefit of incorporating heterogeneity without compromising the objective nature of conventional methods of analysis. Although, in the following chapters consideration will be given to moving the levels of acceptance away from that of the 95% confidence interval, throughout the process of subgroup analysis all evidence used will be empirical and objective.

4.7 Summary

It is deep within the construction of inference methods derived from the various theories of probability that we find justification for examining the value of the less popular, but perhaps more appropriate, methods of inference in reference to economic decision making. Techniques such as Bayesian and decision analysis enables greater utilisation of

information and is more geared towards an action-orientated approach. It is the movement out of the traditional realm of providing information and into the process of decision making that has required the introduction of more subjective measurement. Subgroup analysis itself has many of the advantages of these methods and in addition has the advantage of providing the decision-maker with more objective information.

It is the belief of this thesis that subgroup analysis is the preferred method for incorporating heterogeneity into data analysis. The next chapter looks in more detail at subgroup analysis alone and discusses the theory and conceptual background of the use of subgroup analysis as a tool for decision making. It is then followed by a description of the research question and the methods that will be used to test the hypothesis that subgroup analysis can improve resource allocation and efficiency.

5. Conceptual framework

5.1 Introduction

Very large sums of money are spent every year on clinical trials and economic evaluations. Conventionally these are designed to produce 'average' results for groups of subjects. Conclusions are typically of the form 'treatment X was/was not found to be more cost-effective than the existing standard'. Some trials are designed to provide estimates for specific subgroups of subjects but this adds to the costs through the need for greater sample size. Also researchers and decision-makers are cautious[64] about estimates with low precision and are particularly on their guard against false positive results.

The pursuit of health gain requires an efficient use of health care resources which in turn relies on the presence of *allocative* efficiency, i.e. in order to maximise overall benefit, priority for treatment should be given to patients with the greatest benefit/cost ratio. The finer the differentiation between patients in this respect, the more efficient allocative decisions can be. The crucial difficulty is between-patient variability in response to a given treatment. If patient response is highly uncertain, prioritisation in pursuit of efficiency is made much more difficult. For the pursuit of allocative efficiency, the ability to differentiate between groups, which within them have a greater certainty of patient response, leads to a higher benefit/cost ratio, and resource efficiency becomes more attainable.

The debate over QALYs[65] and the Oregon experiment[66], for example, has drawn attention to the difficulty of seeking to blacklist treatments which can be shown to be beneficial in

some cases, while giving undiscriminating priority to other treatments which are in some cases ineffective and in others potentially harmful.

However, there is a deep-rooted concern among scientists that subgroup analysis may propagate untrue statements. As a result, there is a preference for giving broad-brush but reliable messages about average patients, and leaving clinical decision-makers to bridge the gap between the average and the specific as best they may. This protects the integrity of science, but denies decision-makers evidence that could be more informative because there is a risk that it could be misleading. However, a utilitarian approach to efficiency would suggest that the pursuit of health gain at a population level would imply that getting decisions wrong sometimes is the price to be paid for getting them right as often as possible.

The purpose of this thesis is to explore the trade-off between the needs for more specific information for decision making and the loss of precision this leads to. This question is inherently asking whether or not current attitudes to subgroup analysis, which tend towards scientific caution, are having adverse effects on health gain. The literature on this issue is sparse. An approach to the question is suggested by some recent work on the construction of confidence intervals for cost-effectiveness ratios. It found that "non-parametric boot-strapping"[†] provided the best results when costs and effects were positively correlated[67,68]. However, discussion of subgroup analysis has mainly been from an epidemiological or statistical perspective, rather than a public health or economic

[†] A method of simulating confidence intervals for cost-effectiveness ratios.

one[69,70,71]. The tension involved, or echoes of it, occasionally surfaces in the literature, but mostly in the form of theoretical discussions[72] or disputes in correspondence[73].

There has been no research so far that has looked at the effect a variety of rules for identifying or accepting subgroup analyses, could potentially have in terms of health gain. Being a methodological study, the impact of this thesis could potentially be pervasive. If it transpires that the current cautious approach to subgroup analysis is appropriate, the study will strengthen the hand of advocates of caution and make a small contribution to the avoidance of inappropriate care. If on the other hand it transpires that a less cautious approach can be justified, then this could be a precursor of changes in the approach to the analysis of current and future trials. The extent of the effect on actual health gain and service cost-effectiveness will depend on the heterogeneity of effect among subgroups in different trials, and on clinical judgement, but it should be widespread.

The advancing cause of evidence-based health care has brought into focus the gap between the evidence provided by scientists and the information needed by decision-makers. The response to this has been the introduction of institutions such as NICE, but to make appropriate decisions they will still require the right tools. Pressure on scientists for more subgroup analyses is likely to grow, and the response to this pressure needs to be demonstrably appropriate.

It is on the axes of the relationship between information and decision making that these issues lie. Classical statistical analysis provides us with clear and distinct information about the effects of an intervention throughout a population. Unfortunately this

information, or the form, in which this information is presented, is not always helpful to decision-makers.

5.2 Conceptual approach to subgroup analysis

The problem we have is that an intervention is generally assigned a cost-effectiveness ratio that is taken to apply to all subjects. The cost-effectiveness of an intervention is measured, with some exceptions, as the mean of a sample (with a range within which the mean of the population is expected to lie). In fact the cost-effectiveness of an intervention is a distribution, both in the case of the sample studied and the population it represents. This may or may not approximate to a Normal distribution, but whatever the shape of the distribution, for approximately half of the sample the cost-effectiveness of the intervention will be higher than the stated mean. Additionally, for the other half of the sample the intervention will have a level of cost-effectiveness that is lower than the mean.

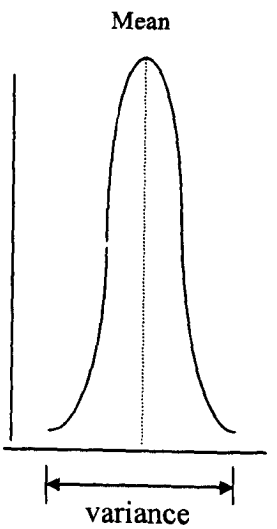


Figure 5.1: little variance about the mean.

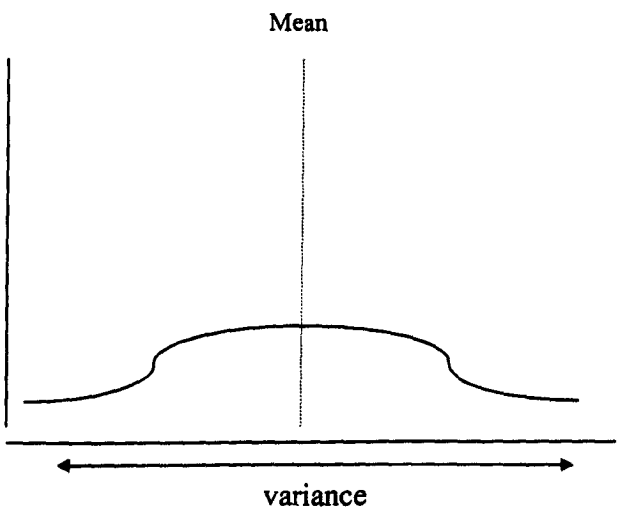


Figure 5.2: Large variance about the mean.

Therefore let us assume a limit is set where provision is too costly for the obtained benefits. If the measure for our chosen intervention lies above that, then it is assumed that this intervention is cost-effective for the population as a whole. If the distribution of relative cost-effectiveness throughout the population is wide, then in reality there is every likelihood that if policy decisions are made using this information, for much of the population this intervention may not be cost-effective under the same constraints. If this is so then resources are being inefficiently allocated.

The importance of the variability of this distribution is the key to the extent of this inefficiency. If variance is small (fig. 5.1), any possible benefits lost or increased costs incurred will be correspondingly small, but if variance is high (fig 5.2), then the inefficiency and inequity resulting could be large.

It is therefore the degree of variation and the kurtosis[‡] in this distribution, which is of most interest to us. Where this variation is large (or the distribution is platykurtic) a large sample is needed to demonstrate a difference in mean effects between intervention and control groups. However, in some cases there is a large average effect, but also a large variance (e.g., in cases where the new treatment can be seen as 'kill or cure'). Even when this is the case it is normal for the recommendation from the trial to adopt the new treatment on grounds that this will increase health gain. Implicitly this is supporting the introduction of treatment that may be systematically inferior for some of the patients.

[‡] the depth and shape of the distribution.

In figure 5.3 the incremental cost effectiveness ratio (ICER) of an intervention is shown, marked with the sparsely broken line that goes through the origin, with its confidence intervals shown as the more regular broken lines either side, tangential to the circle. Our chosen limit for assessing the cost-effectiveness of new interventions is shown as a thick line. Axes are shown as difference in effectiveness (E_T-E_C) and cost (C_T-C_C) between treatment and control groups.

For our intervention to be cost-effective the ICER slope of the intervention must be shallower and hence below this line. In the case of our intervention it clearly is not. Figure 5.4 shows the intervention split into clearly defined subgroups. Here the mean of one of the groups (m_3) is clearly below the line. Unfortunately we cannot say with 95% confidence that the mean of this population will lie under our hypothetical limit. What we can say about these results is that the probability that the slope of the population mean is in the positive section below the thick line is higher than the probability of it being above the line.

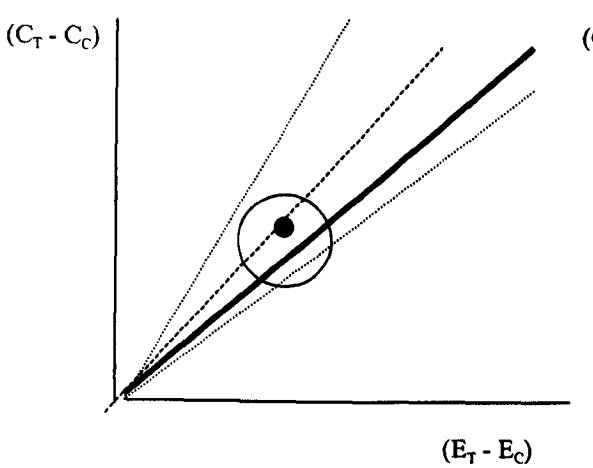


Figure 5.3: Mean and confidence intervals of an intervention, with the thicker line marking the acceptable limit of ICER.

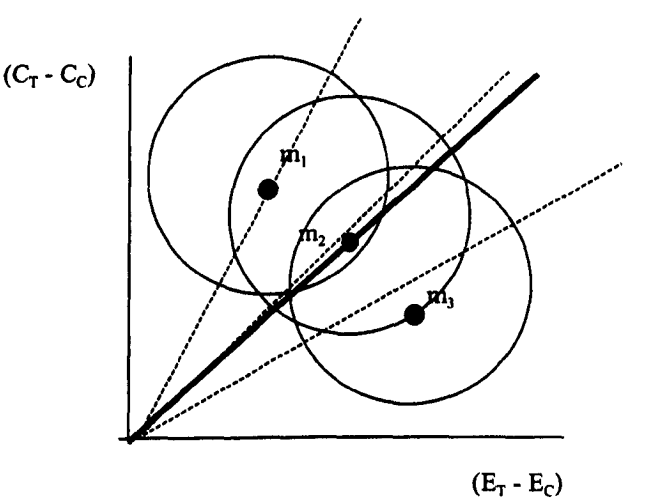


Figure 5.4: Means of sub groups and their confidence intervals.

It is therefore "probable" that a specific subset of the population can be cost-effectively treated with this intervention, but under normal circumstances will either receive no intervention or will receive an alternative intervention that is less cost-effective. Therefore in most circumstances we are likely to have one of three possible alternatives:

1) There are no discernible differences between the cost-effectiveness of an intervention across the whole population (sample) and the cost-effectiveness of that intervention within specific subgroups of that population.
2) Although the intervention is not seen as cost-effective across the whole population, it is seen as cost-effective within one (or more) distinct subset(s) of the population at a statistically significant level.
3) Although the sample size may not allow sub-group analysis to produce differences that are detectable at conventional levels of significance, subgroups can be identified as being distinct at lower than conventional significance levels. What this may say is that one tail of the distribution is in the area that defines the limits between "positive" and "negative" results.

It is this third group which is the area of debate in the probability theory and statistical inference, the concept of the trade-off between certainty and optimisation. Summarised by Fry[74] when he says,

"How much more conclusive must the test data be.... To accept too little implies running greater risks or error than we often desire. To demand too much spells extravagance"

It is the trade-off between optimisation and certainty that is at the heart of the debate around greater use of the data available from trials. The historical link between traditional inference and the need in medical trials to show safety as well as effectiveness as pushed economic evaluation medical interventions towards a greater reliance on certainty than it should. It is the belief of this thesis that greater use of subgroup analysis can go someway to remedying this.

5.3 Summary

The purpose of this chapter was to consider the conceptual and methodological bases of the system proposed to overcome the issues associated with improving evidence for decision-makers. The use of subgroup analysis in achieving solutions has the advantage of both utilising all available information, while being both internally and externally objective with its use of the evidence.

If this conceptual approach to subgroup analysis works in practice it could potentially lead to an improvement in overall health gain for society. This benefit comes with a trade-off against certainty of results (although not likelihood) and it is by no means certain that such a wide distribution of patient specific cost-effectiveness ratios across a diverse sample exists in any or most studies. Even, if such variance does exist it is known whether this level of uncertainty around the mean can be put to use advantageously by the

selection of subgroups with specific characteristics. However, it is the intention of this thesis to find out whether it is possible and whether such methods can be productive.

The next chapter will look at the practical steps of selecting and undertaking a subgroup analysis.

6. The Research Question

6.1 Introduction

The purpose of this chapter is to define the research question and to outline the plan of investigation. The hypothesis is that a cautious attitude to subgroup analysis has had adverse consequences for clinical decision-making and public health, because it potentially recommends inappropriate treatment and a less resource efficient package of services. This study aims to investigate the impact on population health gain and service cost-effectiveness of a variety of 'rules' for deriving and evaluating estimates of cost-effectiveness.

The study involves using data from existing large trials to produce a series of datasets, yielding different estimates of costs and then cost-effectiveness, in aggregate and for subgroups derived by different rules. In the second study aggregate and subgroup estimates of cost-effectiveness will then be used to make hypothetical resource allocation decisions, the outcomes of which will be discussed.

6.2 Methods

Subgroup analysis itself is not a complex procedure. It is simply the analysis of a particular group within a studied data set, which are identified by one or more common characteristics. What makes subgroup analysis complex is the lack of empirical rules associated with its execution. Subgroup analysis, when not used properly, can do more harm than good in terms of providing decision-makers with information on relative effectiveness. This is why there is the need to lay a number of ground rules on exactly

how we will select, or determine our subgroups and how we will calculate ICERs for subgroups. Each of these areas will be discussed in turn.

6.2.1 Subgroup selection

There are three important, as well as interrelated, points often quoted when discussing the issue of subgroup selection[75,76,77]. The time at which the subgroup is identified, the reason for selecting a particular subgroup and the source of the reasoning for it's identification. As has been discussed in chapter three it is generally accepted that the risk of erroneous results emanating from subgroup analysis can be greatly reduced if subgroups are identified prior to analysis, or where possible prior to randomisation and data collection. Additionally it is preferable to have some degree of justification for the choice of this grouping, such as previous evidence of a relationship between the variable concerned and the outcome being measured.

The choice of variable for grouping will be based on a combination of both practical and theoretical grounds. On the practical side it will be limited to variables that are either collectable or already contained within the data set, as well as being limited to variables that, for obvious reasons, can be identified before the intervention in question takes place. The theory on which selection is based will be an accepted or empirically proven reason for likely differences in cost-effectiveness. In short these variables can be classified into three distinct groups;

- demographic,
- diagnostic, and
- characteristic.

Demographic involves the variables of age and sex. Diagnostic includes the specifics of any diagnosis, for example stage of disease, as well as other factors relating to likely prognosis, such as co-morbidity and risk factors. Characteristics contain all other population grouping variables such as social class, education or ethnicity. For some characteristics the variables may overlap into other areas, for example age may well be a risk factor for certain interventions. The classification of variables into one group or another is nevertheless irrelevant, it does not change the analytical method used but is merely there for simplicity of explanation.

The final element of the selection process involves the source of the reasoning behind the selection of the subgroup. There are three types of evidence that can be used for the specification of subgroups; published and non-published research evidence, and expert opinion. Expert opinion is often difficult to quantify and may be unreliable as a source of information, and as such the intention is to limit the evidence for subgroups in this model to published data that show a statistically significant relationship between the variable concerned and the primary outcome.

6.2.2 Measurement of ICERs

The calculation of ICERs will be done taking the incremental change in cost and incremental change in effectiveness. The two methods used to calculate the ICER of a specific subgroup, once selected, will be the bootstrapping approach using a Monte Carlo simulation[§]. Unlike those methods, used by Morris[78] and others, such as by Rubin[79] and

[§] A detailed description of these methods is shown in appendix 3.

that used by Laird and Louis[80], the method we will use does not attempt to counter any effect of regression to the mean.

This problem of regression to the mean lies in the fact that subgroups that lie to the end of a distribution, tend to infer that the reason for any increase (or decrease) in cost-effectiveness could have been for mathematical reasons rather than relating to the chosen variable of study. It is considered not to be an issue of importance in the calculation of ICERs for the following reasons.

Let us assume that the individuals being studied were, for example, in an anti-obesity intervention. At the beginning of the trial, some of these people could have been very overweight. Therefore at the end of an intervention that attempts to help people to reach a target body mass index (BMI), the net difference of these individuals is more likely to be higher than those people just slightly overweight, regardless of the relative effectiveness of the intervention upon those individuals. They just had more to lose. In essence the method used by Morris, similar to the others, shrinks the point estimate of the subgroup towards the overall estimate of the intervention.

The problem of regression to the mean is only relevant, when particular outcome measures are being used. Measures such as height, weight, BMI and blood pressure are obviously effected by this. The question is are the outcomes used in the construction of ICERs liable to be affected by regression to the mean? These being costs and measures such as QALYs and "Life years gained". If an individual within the study is associated with larger than average incremental change in life years gained, or indeed costs, this is

unlikely to be due to the fact that at the start of the trial we expected him or her to live to be a hundred and fifty. It is more likely to be the fact that this individual is more likely to have possessed a number of characteristics that enable the studied intervention to be more effective. If it is the latter, which is the more logical, then it is part of what we are trying to discover an integral part of the outcome. As such there is no reason for this to be of concern when dealing with incremental cost-effectiveness ratios.

6.3 Plan of investigation

The aim is to use data from a randomised controlled trial which has reported the overall cost-effectiveness of an intervention and demonstrate that for many patients in the trial the statement made on cost effectiveness (be it positive or negative) will not be true. These are the cases where use of data on the differences in subgroups is most likely to yield increases in expected health gain.

The basic procedure will be as follows:

1. Choose a trial.
2. Select a number of subgroups for which the variable concerned has an empirical relationship with the primary outcome as dictated by a published study that shows a statistically significant odds-ratio at the 5% level.
3. Derive an estimate of overall incremental cost-effectiveness, and measures of variance, for the main comparison groups using Monte Carlo simulation.
4. For each subgroup derive estimates of the incremental cost-effectiveness ratios, and measures of variance, using Monte Carlo simulation.

5. Compare both the ICER 95% confidence intervals surrounding both subgroups and the overall trial effect

6. Compare the likelihood of outcomes for both subgroups and the overall trial effect

7. Assess the likely health gain from a policy decision based on the overall trial results against the use of the evidence from subgroups.

There will be a number of variants on this basic procedure, with rules for choosing subgroups being derived from the literature. The baseline is 'no subgroup analysis'. Subgroups will be chosen according to *a priori* biological, behavioural or organisational plausibility, ignorant of any subgroup results from the trials.

The 'baseline' estimation procedure in this case will involve standard methods appropriate to the outcomes at issue. The fewer the subjects in any given subgroup, the greater the variability of derived parameter estimates in relation to the actual true values.

6.4 Trial selection

Two trial data sets have been identified. They involve both the measurement of costs and the measurement of the cost-effectiveness of health care interventions. The first is The PRAIS UK study that is a resource-use study on people presenting with acute coronary syndrome. The purpose of the analysis of this first data set is to look at how the characteristics of certain subgroups can have a sizeable effect on cost alone, and as such why subgroup analysis can be so important in the measurement of cost-effectiveness.

The second data set is from the Extracorporeal Membrane Oxygenation (ECMO) study, which is a study comparing the cost-effectiveness of extra corporeal membrane oxygenation with conventional management in neonates born with serious respiratory disease. It is with this data set that a practical attempt to compare the cost-effectiveness of subgroups versus the overall study results will take place.

6.5 Summary

Having outlined the methods to be used to test the value of subgroup analysis these will be put into practice in the next two chapters. Firstly, chapter 7 looks at the value of subgroup analysis in looking at the variation in cost distributions alone, whereas chapter 8 looks at cost-effectiveness ratios under subgroup analysis compared to traditional methods for summarising study data.

7. Variation in cost

7.1 Introduction

This is the first of two chapters that use actual data sets to test the theory that subgroups can provide worthwhile additional information to the decision-maker. The cost-effectiveness ratio is made up of two parts; the cost of an intervention and the effectiveness of that intervention in achieving a desired outcome. In both traditional and subgroup analysis the measurement of cost-effectiveness has traditionally centred on the movement of the level of effectiveness within a population and given little consideration to the patient specific cost of interventions.

More recently methods of economic evaluations have moved more towards incorporating patient specific cost data and as a result more consideration is given to the distributions and variation in costs as compared to effects. This chapter looks at the value of using subgroup analysis in the evaluation of costs and their distributions.

It starts by looking at the shape of a cost distribution and the strengths and weaknesses of classical methods of describing these distributions. It addresses how the debate around cost measurement has led to costs being estimated for individual patients so that the distribution tends to be a specific shape and form.

The traditional methods used for measuring inference in costs and the limitations of this information for decision making is assessed. Alternative descriptive statistics are evaluated leading to a discussion of the value to decision-makers of pre-selected

subgroups of costs and whether they are in fact being undervalued as a source of information to the decision maker. Finally these theories are tested when cost data from an audit of patients presenting with myocardial infarction (MI) are analysed for inference and variance by subgroup. These distributions are then simulated using the Monte Carlo technique, from which conclusions are drawn on the value of the various descriptive statistics.

7.2 The shape of a cost distribution

The distribution of individual patient costs tends to have a very specific pattern, with the majority of values being grouped around a mean at the lower end of the range, with a long flat tail. This is both empirically common[81] and makes intuitive sense. With most cases being closely grouped around a mean there tends to be a minimum or fixed cost element for every value of each process being studied.

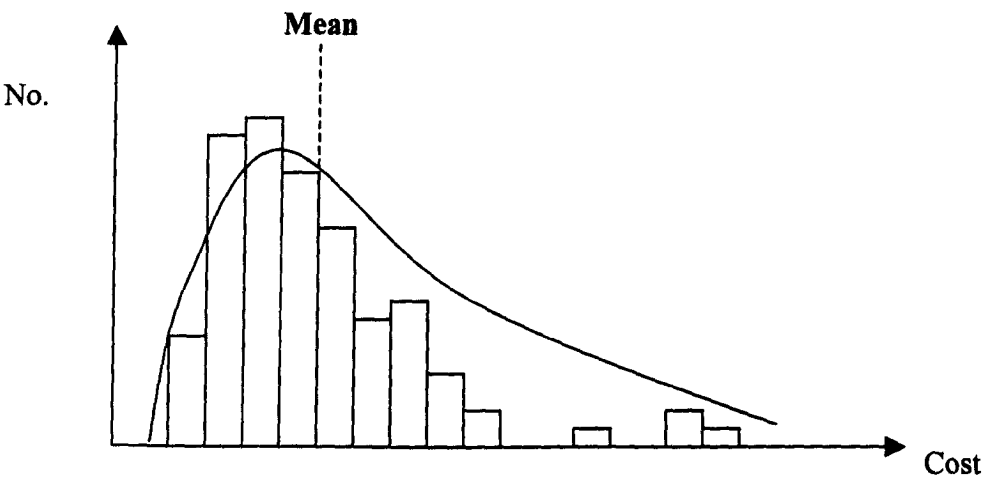


Figure 7.1: Right-skewed curve and histogram most closely associated with a cost distribution

Because of this the possibility of having outliers to the left of the mean is theoretically impossible. Whereas the possibility of outliers to the right of the mean is quite probable as variation in the requirements of each individual case allow high variation in the other element of each value, that of variable cost. We therefore tend to have the shape shown in figure 7.1, one that is skewed to the right rather than symmetrical around the mean[82].

7.3 Measurement of cost

There is much debate surrounding costing of interventions in health economics[83,84,85], most of which can be summarised into a debate surrounding the trade-off between generalisability of costs to wider settings and the detail required to estimate the 'true' costs of the intervention being studied.

The generalisation camp argue that as costs become more and more detailed within the intervention being studied, the more it will incorporate elements of geographical and organisational specificity. For example, it is thought to be unhelpful, or at least not resource-efficient to cost in detail the transport costs of a community-nursing project that was evaluated in a rural area, which will make the results irrelevant to an urban setting. Similarly the costing of a medical intervention being evaluated in a city teaching hospital will not be generalisable to smaller suburban district or community hospitals. As such it puts forward a proposal to use nation-wide or other representative average costs so that firstly decision-makers can interpret them themselves, and so that different costs within economic evaluations can be comparable[86].

The alternative view is that measuring cost is an integral part of economic evaluation and that true cost will, by definition be explicit to the intervention being studied and hence to the outcomes the costs are being associated with[87,88]. For example the cost of employing a specialist nurse in a particular field should be emphasised as a part of the intervention itself, and that assuming a national average cost of a nurse, gives the wrong message to the decision maker as to the true value of the resources used.

It is widely acknowledged that this is a true trade-off and as such there may not be a 'right' answer to the problem[89]. Therefore a best-fit approach is usually suggested, most of these suggestions rely on a greater degree of transparency of the costing process and the splitting up of costing into two parts; resource-use and unit costs[3].

A measure of resource-use is simply a measure of the amount or degree of a specific resource being used. This can be a bed, or a drug or a nurse or doctor. It is this that attempts to deal with the problem of specificity. Once a measure of this resource has been calculated, a unit cost is attributed to it. These are usually based on national averages or other pooled sample data, hence approaching the problem of generalisability.

7.4 Statistical inference and measuring cost

The other advantage of the resource-use measurement method is that each individual being studied has a specific cost for the intervention, rather than being limited to taking an average of all subjects. There has been a tendency to assume a normal distribution around this mean as in the equivalent measure of effectiveness, which makes up the other side of the cost-effectiveness ratio[2]. It is here where we can begin to approach the problem

of using confidence intervals for assessing value of the evidence of cost. Economics concerns itself with finding the optimal use of resources within a population, whereas confidence intervals are used to specify the level of confidence we have in an average.

In statistical inference confidence intervals that surround a mean are a measure of the level of confidence that the mean itself will lie within this range, not a level of confidence that the majority of the distribution of values itself will lie within this range.

As a decision-maker, the information you really need to assess whether an intervention is likely to be an efficient use of resources is the total cost and total benefit of the intervention under local conditions. If the distribution around the mean used to create cost-effectiveness ratios is normal then this information may** be useful, but we know that in the case of cost, this is not true. When measuring the cost in a cost–effectiveness ratio, consideration has to be taken of the shape of the distribution, as well as the mean itself. If the long tail associated with cost distributions is not accounted for, there is a risk that a programme set up on the basis of a cost-effectiveness ratio will create higher costs than expected.

In essence the message is that the length of the tail, or the width of the range of true values within a sample are as important an indicator of cost as the mean cost and the confidence intervals surrounding it. It is imperative then that we incorporate the

** The word 'may' is used because even where the distribution is normal the kurtosis and/or range of the distribution of true values may be significantly different to that of the confidence intervals. So even then the decision-maker does not have a true representation of the cost of the intervention.

measurement of both the skew of the distribution and the kurtosis, or shape of the distribution.

7.5 Measuring the shape and range of a distribution

The shape and size of a distribution can be measured in a number of ways. The three key elements are length or width of the points in a distribution, the shape of the distribution and symmetry of the distribution. The width or length of a distribution, which shows the distance between the lowest point on the distribution and the highest, is known as the range of the distribution. Another measure which is useful for measuring the size of distribution that have long tails is the inter-quartile range, which measures the range between the 25^{th} and 75^{th} percentile of all points within the distribution. Probably the most vigorous method, though is the standard deviation.

The standard deviation is measured as the square of the differences between each value within the distribution and the mean. The differences are squared only because if they were not they would add up to zero, the squaring process is therefore used to get rid of negative values. If a distribution is wide it will tend to have values of difference with the mean that are high, and therefore a high standard deviation. Similarly with a narrow distribution this will give a low value for the standard deviation. The formula for the standard deviation (SD) is thus:

$$SD = \sqrt{\frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2}$$

The shape of a distribution has two areas of comparison; its symmetry and its relative weighting across the range. The symmetry of a distribution is measured by its degree of skew and the weighting by the measure of kurtosis. The first graph in figure 7.2 is symmetrical and has zero skew, whereas the second graph is positively skewed, or skewed to the right. The relative distance between the mean of the distribution and the median measures the degree of skew.

Skew is a measure of the asymmetry of the data around the sample mean. If skew is negative, the data are spread out more to the left of the mean than to the right. If skew is positive, the data are spread out more to the right. The skew of the normal distribution (or any perfectly symmetric distribution) is zero.

The skew of a distribution is defined as:

$$y = \frac{E(x - \mu)^3}{\sigma^3}$$

where E(x) is the expected value of x.

The inter-quartile range of a distribution (IQR) computes the difference between the 75th and the 25th percentiles of the sample in X. The IQR is a robust estimate of the spread of the data, since changes in the upper and lower 25% of the data do not affect it. If there are outliers in the data, then the IQR is more representative than the standard deviation as an estimate of the spread of the body of the data. The IQR is less efficient than the standard deviation as an estimate of the spread, when the data is from a normal distribution.

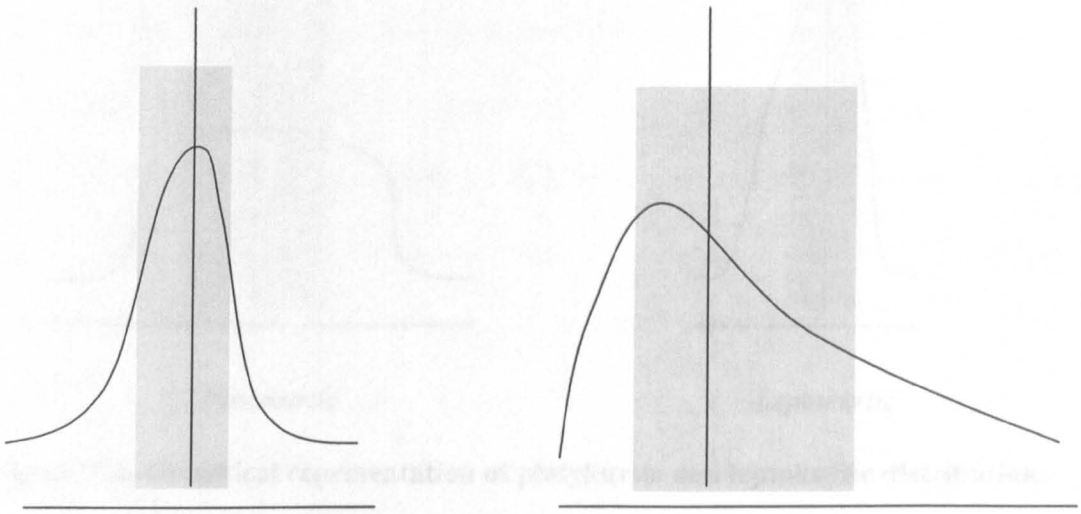
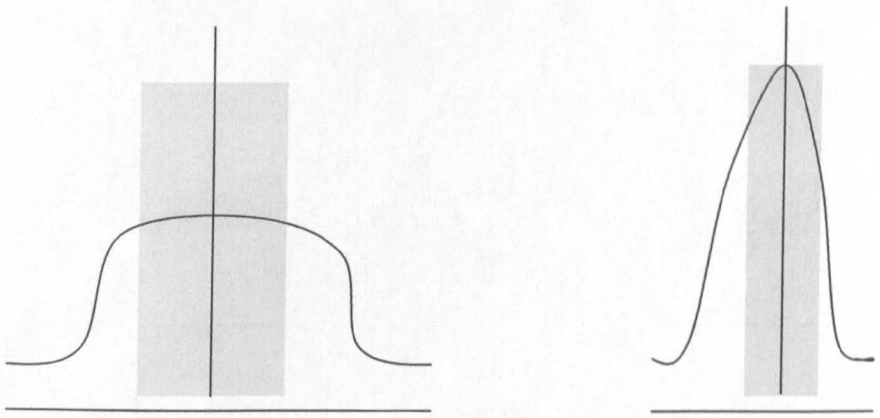


Figure 7.2: Graphical representation of relative mean and inter-quartile range, of a symmetrical and positively skewed distribution.

Kurtosis is a measure of how outlier-prone a distribution is. The kurtosis of the normal distribution is 3. Distributions that are more outlier-prone than the normal distribution have kurtosis greater than 3; distributions that are less outlier-prone have kurtosis less than 3. A flat-topped distribution tends to have a low value of kurtosis and is said to be platykurtic (flat bulging). A sharp-peaked distribution will tend to have a high value of kurtosis and is called leptokurtic (thin bulging). The kurtosis of a distribution is defined as:

$$k = \frac{E(x - \mu)^4}{\sigma^4}$$

where E(x) is the expected value of x.



Platykurtic *Leptokurtic*

Figure 7.3: Graphical representation of platykurtic and leptokurtic distributions

It is the shape and size of cost distributions that makes the mean and confidence intervals an insufficient means of summarising the data. When making a decision based on expected costs of an intervention the degree to which the distribution of costs is skewed is likely to be of some importance. If there is a likelihood that a proportion of those receiving the intervention may cost well in excess of the average cost for the intervention, this provides an added risk to funding the intervention. Similarly if the cost distribution is particularly platykurtic this also implies a degree of unpredictability of the likely cost of any individual.

In theory the decision as to whether to fund an intervention relies heavily on the level of risk associated with the probable cost and benefits of any particular individual. The

average costs and benefits across a wide and varied population, although providing a guide, may not be all that helpful to a decision-maker if this means it is not representative of the vast majority of individuals within the population.

7.6 The measurement of the distribution and confidence in the mean

It is the relative importance of both the shapes of the distribution and confidence in the mean, which is of interest to us in the issue of using evidence from subgroup analysis. The reason for this is that confidence intervals are heavily affected by sample size, which is inevitably smaller in subgroup analyses than in the sample as a whole, hence subgroups will always have relatively wide confidence intervals.

The confidence intervals around a mean are a constant (for example for 95% C.I. it is 1.96) multiplied by the size of the standard error around the mean. The standard error in turn is reliant on two measures, the sample size (n) and the level of variance, measured by the standard deviation of the distribution (σ), in the formula;

$SE = \sigma / \sqrt{n}$.

In terms of the range of values, if subgroups are selected due to a predicted variance in outcome, in this case cost, it is likely that the chosen variable is a key cost driver. Given this it is acceptable to assume that the variance within the subgroups is likely to be narrower than in the sample as a whole. The effect will be twofold, a reduction in the numerator due to lower variation and a reduction in the denominator due to a smaller sample size. In reality the effect on reducing sample size has a disproportionately large effect on the standard error as although the measure of variance is an average across all

variables, the sample size is a sum. Therefore any reduction in sample size is unlikely to be offset by reduced variance within the sample.

In addition under the method used for calculating confidence in the mean the measure of variance is solely based on standard deviation, and takes no account of skew. This is because methods of statistical inference were devised for calculations inferred from a normal distribution, which is rarely seen with cost data. As a result we have a trade-off between restricting the unpredictability of the cost of treating each individual and keeping narrow confidence intervals for the mean. This is shown graphically in figure 7.4. Either we have a reliable estimate of the mean when concentrating on the total sample, at the expense of a leptokurtic, heavily skewed distribution. Or we have the possibility of a less skewed more playtykurtic distribution with a less reliable mean.

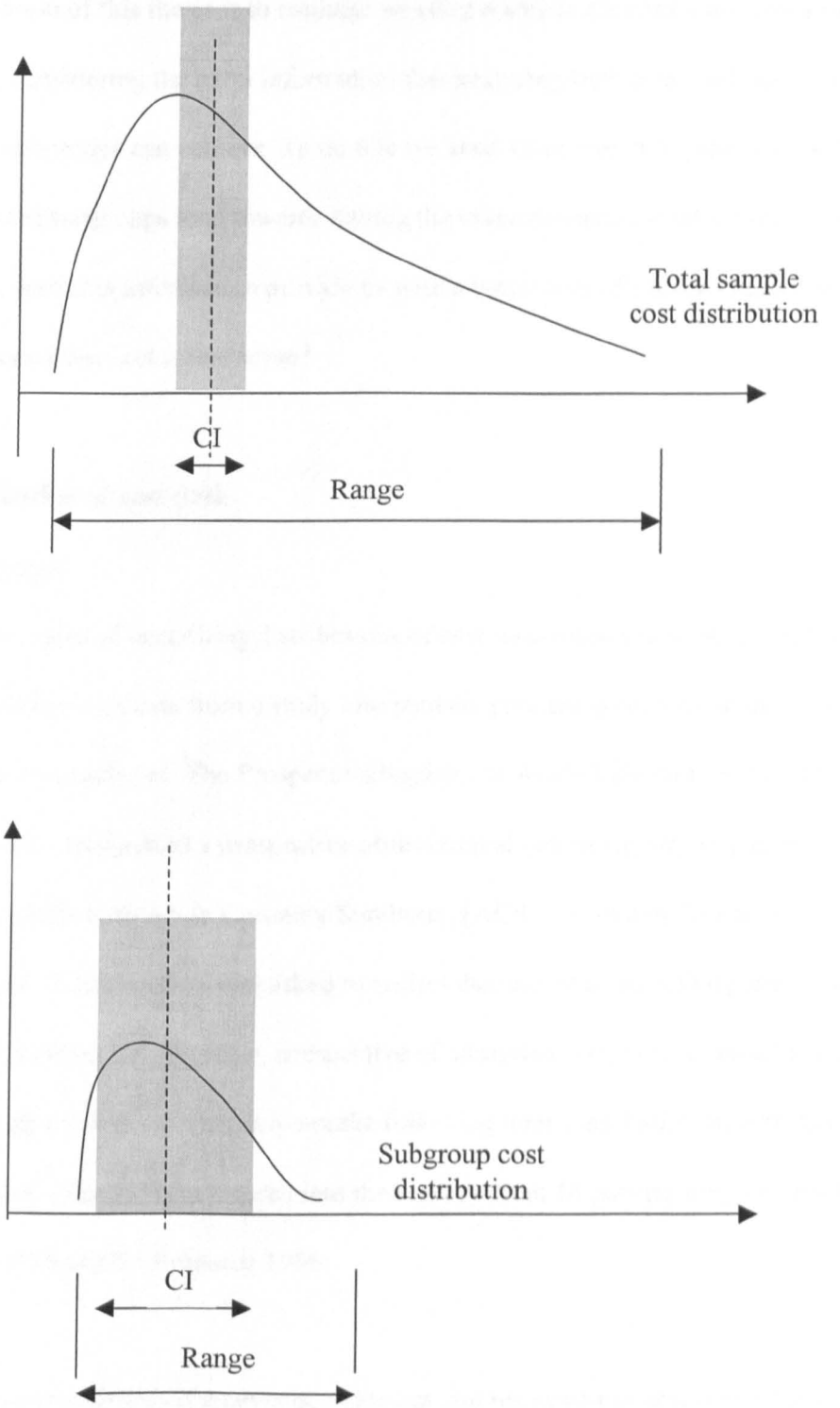


Figure 7.4: A graphical representation of the hypothetical mean, confidence intervals and distribution of costs of a total sample (top) and a pre-selected subgroup (bottom).

The intention of this thesis is to evaluate whether resource allocation decisions can be aided by considering the extra information that analysing both costs and outcomes by pre-selected subgroups can achieve. To do this we need to answer two questions; firstly do pre-selected subgroups tend towards having the characteristics discussed above, while secondly does this information provide us with a better way of measuring, or drawing conclusions from cost information?

7.7 Evaluation of cost data

7.7.1 Methods

To test the value of describing distributions of cost with measures of shape and form as well as inference of data from a study into patients presenting with an acute myocardial infarction was analysed. The Prospective Registry of Acute Ischaemic Syndrome (PRAIS) was designed as a prospective observational cohort registry of patients admitted to UK hospitals with Acute Coronary Syndrome (ACS). A total of 56 UK hospitals participated. Each hospital was asked to collect data on 20 consecutive patients admitted with ACS without ST elevation, irrespective of admission location or consultant team. Patients had a follow-up visit at 6 months following their initial hospital admission. A total of 1046 patients were entered into the registry from 56 participating centres between 23rd May 1998 and 3rd February 1999.

Information was collected on events, treatment and resource-use across the duration of the admission and a follow-up period of the next six months. Table 7.1 lists the measures of resource-use, the unit cost attributed to them and it's source. These unit costs, along with the measure of resource-use from the PRAIS data, are used to create an estimated total

cost for each patient presenting. Figure 7.3 is a histogram of total initial admission cost for all patients in PRAIS.

A literature search was conducted using 'PubMed', 'Medline' and 'HealthPro', looking for evidence of any relationship between a series of patient characteristics at admission and total cost of intervention in the treatment of suspected myocardial infarction (MI). A search was made for published research showing a strong relationship between cost and the following characteristics; sex, age, ECG reading, previous history of event and previous diagnosis of CHD.

The selected subgroups for cost variation were age[90, 91] (below 65, 65 –74, 75+) and ECG category (normal or complications)[92, 93, 94, 95, 96]. These were selected for having a strong empirical relationship with levels of cost or resource-use. The empirical evidence of the relationship between cost and sex, previous history of event and previous diagnosis of CHD, was either too small or inconclusive.

The intention is to compare the relative value of the various statistical descriptive measures of each distribution, both the overall data and the individual subgroups, for relevance to the decision-maker. Also to discover whether the affect that reduced sample size has on measures of inference of subgroups is likely to be similarly negative or conversely positive when measuring distribution shape. The distribution of each subgroup will be plotted and all the descriptive statistics discussed above will be calculated for each. This will then be used to construct probability distribution simulations using the Monte Carlo technique.

The Monte Carlo simulation takes a sample distribution and translates it into a probability distribution where the likelihood of each outcome has a probability and where the probabilities of all possible outcomes sum to one. This helps us gauge the likelihood of one and a group of outcomes. With only one uncertain variable, cost, the shape of the probability distributions will be the same as the sample distributions.

Table 7.1 Resources, unit costs and sources.

Resource	*Unit cost (£) 1998-9 prices*	*Source*
Hospital stay (LOS x 1)		
CCU bed-day	348	CHD costing study [97]
Medical bed-day	119	CHD costing study [17]
In-hospital events		
Angiography	500	Literature [98]
PTCA	1428	CHD costing study [17]
CABG	2310	CHD costing study [17]
Stress test	110.50	Literature [99]
Drugs (daily dose x LOS-except TT)		
aspirin	0.10	British National Formulary
beta blocker	0.40	BNF '99
calcium antagonist	0.42	BNF '99
nitrate(oral)	0.29	BNF '99
potassium channel opener	0.33	BNF '99
nitrate(IV)	16.50	BNF '99
GPIIb/IIIa	1.00	BNF '99
Oral anticoagulant	0.89	BNF '99
Thrombolytic therapy	85.00	BNF '99
heparin	0.62	BNF '99
drugs on discharge (per month)		
antiplatelet (aspirin)	1.55	BNF '99
beta blocker	11.20	BNF '99
calcium antagonist	11.85	BNF '99
nitrate (oral)	8.17	BNF '99
potassium channel opener	9.93	BNF '99
statin	29.69	BNF '99
lipid-lowering (other)	13.38	BNF '99
anticoagulant (oral)	24.92	BNF '99
ACE inhibitor	13.43	BNF '99
Angiotensin 2 inhibitor	15.75	BNF '99

7.7.2 Results

The distributions of the 'cost per admission by individual' for the study as a whole and for the pre-selected subgroups were analysed for both measures of variance and for inference around the mean (table 7.2). As anticipated confidence intervals around the mean were much wider in the subgroups than in the sample as a whole, with sample sizes much smaller but with only marginal differences in standard deviation. In terms of measures of variance, there was a tendency for subgroups to have lower values for skew and kurtosis. This was not universal but where the skew or kurtosis of a subgroup was higher than the total it was predictable. For example, the ageband subgroup of patients over the age of 75, it would be anticipated that a large proportion of the skew of the total was due to outliers in the highest ageband and because of this their length of stay is the least predictable.

An example of the limitation of the confidence interval when dealing with the value of subgroups is illustrated with the comparison of the cost distributions of the total data set and the subgroup of patients who's ECG was normal. As would be expected the subgroup with its smaller sample size (141 versus 868 in the main sample) has far wider confidence intervals (£275 versus £153). But within the smaller subgroup distribution there is far less variance (SD of 825 versus 1145), far less skew (2.14 versus 2.92) and far less kurtosis (4.98 versus 11.67).

This would suggest that the probability of the cost of a random patient being close to the mean is higher in the subgroup than in the main sample. This is the essential difference

between what the confidence interval on a cost distribution tells us, compared to the information as to the probability of random subject having a cost close to the mean.

Table 7.2 Descriptive statistics of the distributions of subgroups

Group	N	Mean	SD	Skew	Kurtosis
Total	868	1092.76	1145.59	2.92	11.67
Age					
< 60	263	1030.63	1079.58	2.29	6.31
60–74	364	1128.83	1210.49	2.73	9.68
75 +	241	1106.09	1116.22	3.88	20.76
ECG					
Normal	141	852.55	825.22	2.14	4.98
Complications	731	1130.03	1192.70	2.89	11.21

Distributions and details of simulations can be found in Appendix 1

Group	Mean	95% confidence intervals		Range of CIs
Total	1092.76	1016.45	1169.08	152.63
Age				
< 60	1030.63	899.55	1161.71	262.16
60–74	1128.83	1004.06	1253.60	249.54
75 +	1106.09	964.45	1247.73	283.28
ECG				
Normal	852.55	715.15	989.94	274.79
Complications	1130.03	1043.57	1216.48	172.91

Distributions and details of simulations can be found in Appendix 1

To take our example further we have simulated both of the distributions using the Monte Carlo method to estimate the likelihood of a random patient falling within the inter-quartile range of the given distribution as compared to the 95% confidence interval range. The distribution of both groups was simulated using the descriptive statistics of the

distributions shown in tables 7.2 and 7.3. Simulations were repeated for a series of iterations between 100 and one million, but little change was seen in the results beyond 10,000[††]. Table 7.3 shows a comparison of the simulations of the total data set against the subgroup for Normal ECG.

Table 7.3 Results of Monte Carlo probability distribution simulation

Group	All patients	ECG-Normal subgroup	Ratio (All / subgroup)
Mean	1065.87	851.09	
Standard deviation	1102.86	815.51	
Skew	2.91	2.16	
Kurtosis	14.60	7.92	
First quartile	395.03	396.08	
Third Quartile	1207.25	965.88	
Inter-quartile range	812.22	569.80	1.42
5^{th} percentile	207	189	
95^{th} percentile	3391	2606	
90 percentile range	3184	2417	1.32
95% C.I.			
lower	1016.45	715.15	
upper	1169.08	989.94	
range	152.63	274.79	0.55

The simulation shows that although the confidence intervals surrounding the mean are narrow in the main sample, the likelihood of a random patient falling into this range is relatively low due to the shape and variance of the distribution, which in turn is due to its unpredictability regarding outliers. The proportion of values within the inter quartile

[††] Distributions and details of the simulations are shown in full in appendix 1.

range that are likely to fall into the range of the confidence interval is just 19% in the main sample as compared to 48% in the subgroup. To give some idea of the importance of predicting the effect of the variance within a distribution, for all the distributions simulated the average proportion of total costs in the upper quartile was 60%.

7.8 Summary

The purpose of this chapter was to show that as well as measures of effectiveness, costs can also be affected by characteristics of the population and as such are just as likely to benefit from subgroup analysis. The example shown in this chapter also shows that, although confidence intervals are a good gauge of the predictability of the mean value, because of the shape of cost distributions they may not be the only measure needed for decision makers to make definitive choices on resource allocation.

The measurement of the mean, and confidence in that mean, are an important descriptive of a distribution. If the distribution is a normal one, it can be argued that it is the only descriptive required to make conclusions of that distribution. Where a distribution is not normal, there are other factors to take into account. When assessing the likely cost of an intervention it is the area under the curve that guides the decision-maker and the likely effect that outliers may have on total cost. The examples analysed in this chapter show that cost distribution can be skewed to the extent where 60% of total costs will be due to 25% of the sample. It is therefore important that the analysis of a cost distribution incorporates a measure of the variance and shape of the distribution.

This chapter has also shown that pre-selected subgroups for which there is good prior evidence of a dependent relationship may in fact be being undervalued by the limits enforced by the measurement of a sample based purely on confidence intervals. With an objective of minimising the risk of unpredictable costs, pre-selected subgroups have been shown to be a potential tool in cost analysis. The next section will take this further by looking at measures of effectiveness and its combination with cost information to inform cost-effectiveness.

8. Measurement of cost-effectiveness

8.1 Introduction

The potential advantages and dangers of using evidence from subgroup analysis to inform policy decisions have been discussed in detail. In the previous chapter we started to look at the characteristics of subgroup analysis in comparison to traditional analysis. We have shown that costs as well as outcome measures can have strong relationships with covariates and that when the relationship is strong, the variance around the mean of a subgroup can be smaller than the study data as a whole. This chapter takes the theory of the potential advantages of subgroup analysis one step further and looks at both the costs and effects of a specific intervention.

Here we look at the data from a trial that evaluated the incremental cost-effectiveness of a new technology in treating severely ill new born babies, the Extracorporeal Membrane Oxygenation (ECMO) trial. First we give details of the background to the ECMO trial, and the results and conclusions that were published. The data is then re-analysed to obtain the measures of variance that had not been calculated in the original study. We then calculate ICERs for selected subgroups and assess the potential value of this information in improving overall health gain.

8.2 Extracorporeal Membrane Oxygenation (ECMO) Trial

The ECMO trial looked to compare the costs and outcomes of conventional treatment and the incorporation of the ECMO technology. Conventional treatment for infants in

respiratory failure is ventilation with high level oxygen. Extracorporeal membrane oxygenation is a technique that oxygenates blood outside the body, obviating the need for gas exchange in the lungs, and, if necessary, providing cardiovascular support. Intention to treat analysis showed that extracorporeal membrane oxygenation was highly clinically effective[100].

A preliminary economic evaluation carried out before the trial suggested that extracorporeal membrane oxygenation was probably more effective and more expensive than conventional management[101]. It also showed that the existing evidence on cost effectiveness was inadequate for setting priorities because the uncertainty surrounding the data was too great[102].

The economic evaluation of the ECMO trial was a cost-effectiveness analysis comparing extracorporeal membrane oxygenation with conventional management based primarily on the principal clinical outcome of the trial (survival without severe disability at age 1 year). The economic evaluation was conducted from the viewpoint of the NHS and so includes only direct costs to the health service[103].

Extracorporeal membrane oxygenation was provided in five centres and babies were recruited from 55 UK neonatal centres. Babies were eligible for the trial if they were mature new-born infants with severe respiratory failure. They were randomised either to be transferred for extracorporeal membrane oxygenation or to receive conventional management. There were 185 babies in the trial, of which 93 (50.3%) were randomised into the ECMO arm with 92 (49.7%) in the conventional treatment arm. A summary of

descriptive variables comparing the two groups at randomisation is shown in Table 8.1, with results at the end of the year one follow-up in Table 8.2.

Health service use was divided into three components: mode of transport used for transfers made after randomisation until discharge; services received in the initial hospital inpatient stay after randomisation, subdivided by level of intensity; and use of health services from discharge up to 1 year of age. The babies' initial hospital treatment was described in terms of five levels: days receiving extracorporeal membrane oxygenation; days receiving maximal intensive care (more than 90% oxygen); days on a ventilator (receiving less than 90% oxygen); days on supplementary oxygen; and days in normal care.

Health service costs for each patient were calculated for initial hospital stay, post discharge health service utilisation and transport costs. The initial stay was estimated based on unit costs of days under extracorporeal membrane oxygenation, days under maximal intensive care (more than 90% oxygen), days on a ventilator (receiving less than 90% oxygen), days on supplementary oxygen and days in normal care. The costs for each group are outlined in Table 8.3.

Table 8.1: Comparison of ECMO and 'conventional treatment' arms

	ECMO (n=93)	*Conventional (n=92)*
Mean Age (hours)	47.2	57.6
Mean Gestation (weeks)	38.8	39.2
Mean Birthweight (grams)	3261	3346
Most recent Oxygenation Index (OI)	69.4	72.1

(Source: Roberts TE, 1998 [4])

Table 8.2: Outcomes at one year of ECMO and 'conventional treatment' arms

	ECMO (n=93)	*Conventional (n=92)*
Death	30	54
Severe disability	1	1
Impairment and some disability	12	4
Impairment and no disability	4	5
No impairment or disability	45	27
Lost to follow-up	1	1

(Source: Roberts TE, 1998 [4])

Table 8.3: Costs at one year of ECMO and 'conventional treatment' arms

	ECMO (n=93)	*Conventional (n=92)*
	£	£
Initial hospital stay	1,603,267	476,409
Additional cost of death	30,352	58,536
Transport	150,146	20,475
Post discharge health service utilisation	153,059	88,739
Total	1,936,824	644,180
Mean cost per case	20,826	7,002

(Source: Roberts TE, 1998 [4])

A cost-effectiveness ratio for ECMO in new-born babies was constructed using the data set described above. The mean cost per case in the conventional management arm was £7,002 compared to £20,826 in the ECMO arm, a difference of £13,824 (95% C.I. £9,660 – £17,984). The primary outcome of the intervention was number of lives saved without resulting in severe disability. For the conventional arm the resulting outcome was achieved in 36 of 92 (39.1%) whereas in the ECMO arm this was achieved in 61 of 93 (65.6%), a difference of 26.5% (95% C.I. 12.4% - 40.5%). The resulting incremental cost per life without severe disability achieved can be estimated as £52,244 within the parameters of the calculations shown in table 8.4.

Table 8.4: Cost-effectiveness estimation at one year of ECMO over conventional treatment.

	ECMO (n=93)	*Conventional (n=92)*	*Conventional # (n equivalent 93)*
Initial hospital stay	£1,603,267	£476,409	£481,587
Additional cost of death	£30,352	£58,536	£59,172
Transport	£150,146	£20,475	£20,698
Post discharge costs	£153,059	£88,739	£89,704
Total	£1,936,824	644,180	£651,161
Survival no disability	61	36	36.4
Average cost per survival with no disability	**£31,751**	**£17,894**	
Incremental cost of ECMO			£1,285,663
Incremental cost per extra survival with no disability			***£52,244***

 # 'Conventional' costs have been weighted so as both sets of costs relate to 93 patients. (Source: Roberts TE, 1998 [4])

The conclusions of the study group were that the evidence so far suggests that Extracorporeal membrane oxygenation can be as cost effective as many other life extending technologies regularly used in developed countries. However, it does include the caveat that until the results of the long term follow up studies become available, this conclusion should be viewed with some caution.

8.3 Stochastic measurement of cost-effectiveness of ECMO

As has been discussed in earlier chapters, cost-effectiveness analysis has moved on recently in an attempt to incorporate measures of uncertainty and this has in general been through the introduction of confidence intervals for cost-effectiveness ratios. The

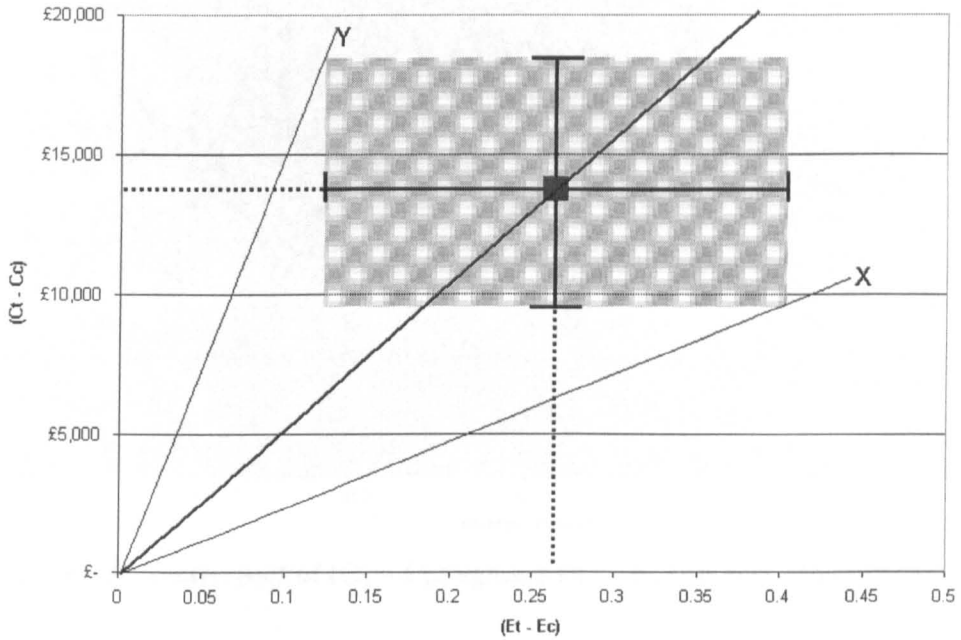
derivation of ICERs where there are individual measures of effectiveness and cost of ***n*** patients allows us to calculate a stochastic ICER. Stochastic measurement is where variables are determined from data sampled from the individuals in the study, as compared to deterministic measurement where variables are analysed as point estimates. If both costs and effects are determined from data sampled from the individual patients in the study, variances are available and so formal statistical tests can be performed on observed differences in costs and effects.

Here effectiveness is measured as a binary outcome where 1 denotes survival without severe disability and 0 denotes death or the presence of severe disability. Our ICER is calculated as

$R = (C_T - C_C) / (E_T - E_C)$, where C_T and C_C are the cost of new treatment and conventional treatment respectively and E_T and E_C are the effects of each.

This gives us a two-dimensional box with the confidence intervals for both our effects and cost sizes, which gives us a quasi-confidence area shown in figure 8.1. In reality the probability of the true ICER of the ECMO intervention lying on either slope X or Y is very low, less than 0.5%, and that both intuitively and mathematically the true confidence interval area is likely to be elliptical in shape[104].

Figure 8.1: Incremental cost-effectiveness ratio (ICER) of ECMO technology on likelihood of survival without severe disability



The next step is to move towards this true form of our confidence area through the estimation of the sampling distribution of the ratio. The most common application for this has been non-parametric bootstrapping[105,106]. The observed data for cost and effects are treated as an empirical probability distribution that is resampled with replacement a given number of times. Each resample is used to give an estimate of R. These estimates are then used to create an empirical distribution of R from which the confidence intervals are constructed. The results of such an exercise with the ECMO data are shown in figures 8.2 and 8.3 and summarised in table 8.5.

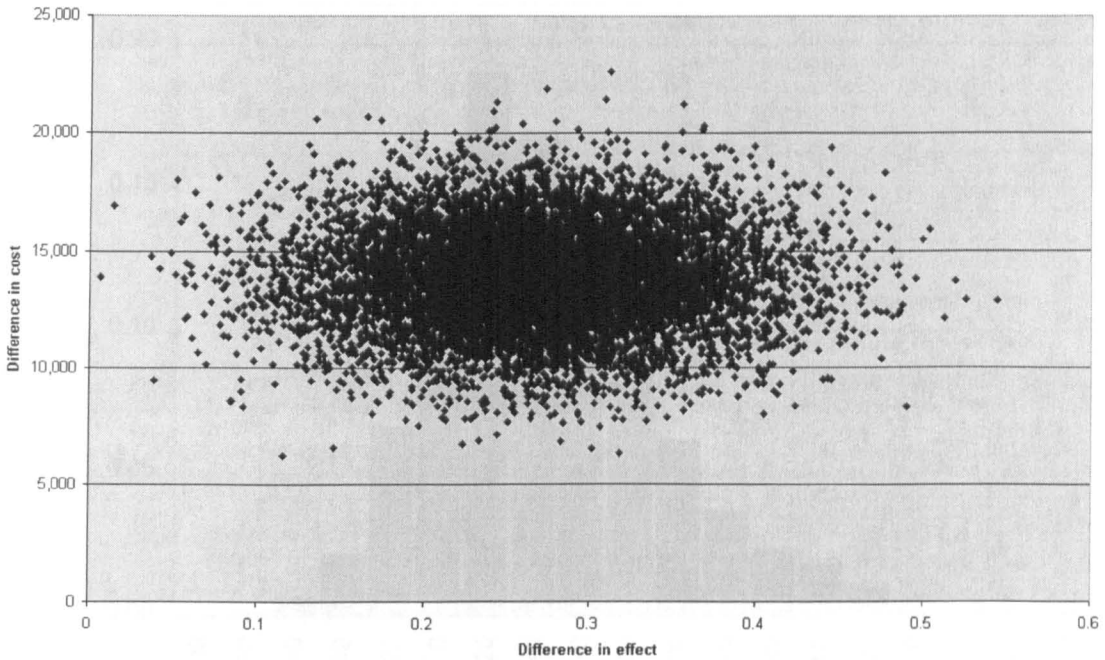


Figure 8.2: Scatterplot of ($C_T - C_C$) against ($E_E - E_C$) on cost-effectiveness plane

Figure 8.3: Histogram of bootstrap replicates of R (ICER) probability distribution.

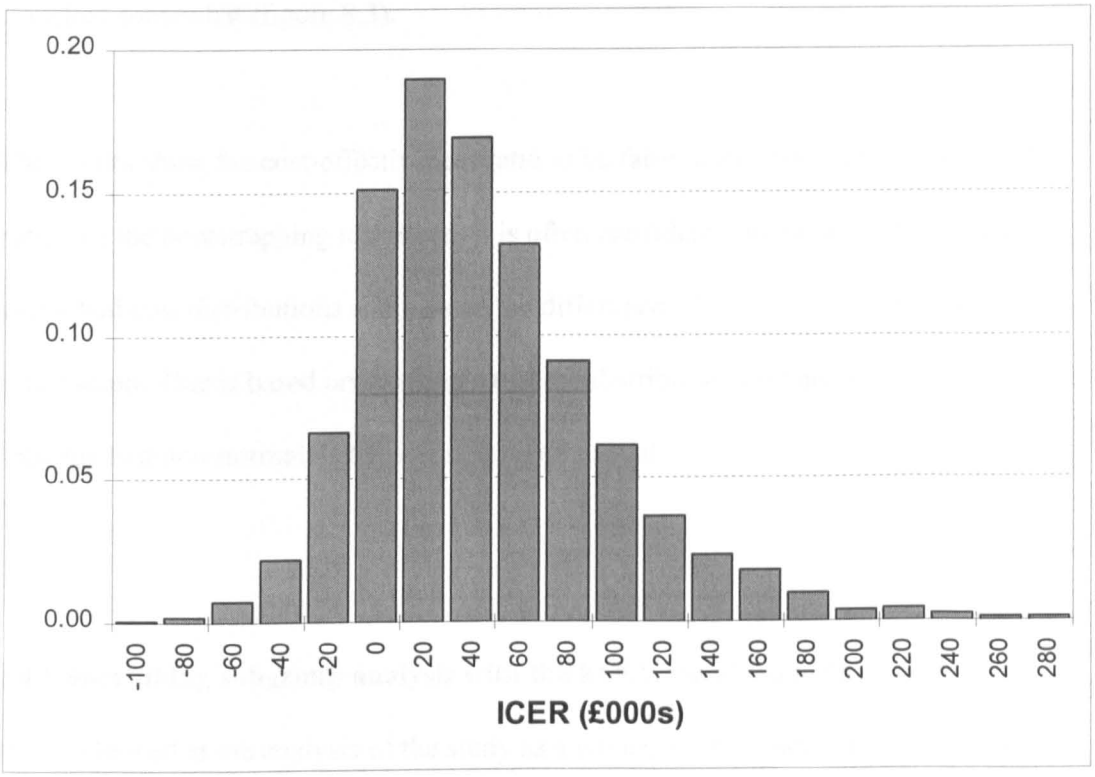


Table 8.5: Resulting 95% percentile ranges from bootstrap replication of ICER.

Variable	*Mean*	*95% confidence interval*
Incremental cost of ECMO	£13,824	(£5,958 – £40,747)
Incremental likelihood of survival	0.27	(0.15 – 0.39)
Incremental cost-effectiveness ratio	£52,244	(32,521 – 95,137)

When these simulations are undertaken, there can be a tendency for replication bias, but this is considered insignificant if measured as a ratio of variance of under 0.1. Bias within the replication was measured as a ratio of variance of just 0.08 and as such was not considered important[107]. The confidence intervals around the incremental costs and effects

alone do not include zero and are positive, whereas the ICER confidence interval is stretched somewhat (figure 8.3).

The results show the cost-effectiveness ratio to be fairly uncertain given the stochastic nature of the bootstrapping technique. It is often considered unnecessary to simulate the individual cost distributions and instead the difference ($C_T - C_C$) is used with a normal distribution. This is based on the theory that the distribution around the mean difference between two non-normal distributions is itself normal.

8.4 Undertaking subgroup analysis with the ECMO trial data set.

Having looked at the analysis of the study as a whole, we now turn our attention to the analysis of specific subgroups. The potential subgroups in the study were that of birth weight, gestation age, diagnosis and oxygenation index (OI). These are listed in table 8.6. There were two subgroups by birth weight, two by gestation, three by diagnosis; congenital diaphragmatic hernia (CDH), idiopathic persistent foetal circulation (IPFC) and persistent pulmonary hypertension (PPH), and two by OI.

A literature search was conducted using 'PubMed', 'Medline' and 'HealthPro', looking for evidence of any relationship between a series of patient characteristics at randomisation and the likely outcome of the usual treatment of new born babies with severe respiratory failure. A search was made for published research showing a strong empirical relationship between the primary outcome and the following characteristics; sex, age, birth weight, gestation age, diagnosis and initial oxygenation index

(OI)[108,109,110,111,112,113,114,115]. An empirical relationship between a characteristic variable and the primary outcome (survival) was assumed to be present where an odds-ratio for survival was significant at the 5% level in at least one paper.

The selected subgroups for outcome variation were birth weight, diagnosis, gestation age, and initial OI. These were selected for having a strong empirical relationship with outcomes. Any relationship between the other variables and the primary outcome were either too small or inconclusive. The intention is to compare the relative value of the various statistical descriptive measures of each distribution, both the overall data and the individual subgroups, for relevance to the decision-maker.

The results in table 8.6 (and in figure 8.4) show that overall the ECMO intervention had relatively wide confidence intervals (95%) spanning from £35,521 to £95,137, which would, under classical analysis, make the acceptance of the intervention unlikely. The subgroups show a wide variety in both means and their measures of variance, shown here in terms of confidence intervals. Although many have confidence intervals (C.I.) that are extremely wide, such as the 'CDH diagnosis', the 'gestation > 275 days' and 'OI > 60', there are also a number which have very narrow C.I., such as the 'diagnosis of IPFC' or 'PPH' and 'OI = 40-59'.

Table 8.6: Deterministic ICERs of subgroups with confidence intervals of stochastic analysis

Subgroup (no.)		*ICER*	*95 percentile confidence intervals*	*Percentile< £50,000/LS*
Birth weight	< 3300 (1)	65,627	(39,639 – 137,317)	20%
	> 3300 (2)	28,738	(9,498 – 75,952)	85%
Gestation	< 275days (3)	63,417	(28,638 – 206,888)	30%
	>275 days (4)	39,281	(22,637 91,789)	70%
Diagnosis	CDH (5)	164,900	(81,305 – 539,475)	0%
	IPFC (6)	16,340	(1,958 – 44,430)	95%
	PPH (7)	33,356	(16,568 – 77,879)	80%
OI category	40-59 (8)	42,359	(22,334 – 86,210)	70%
	>60 (9)	74,218	(29,945 – 318,934)	20%
Total study (10)		£52,244	(32,521 – 95,137)	45%

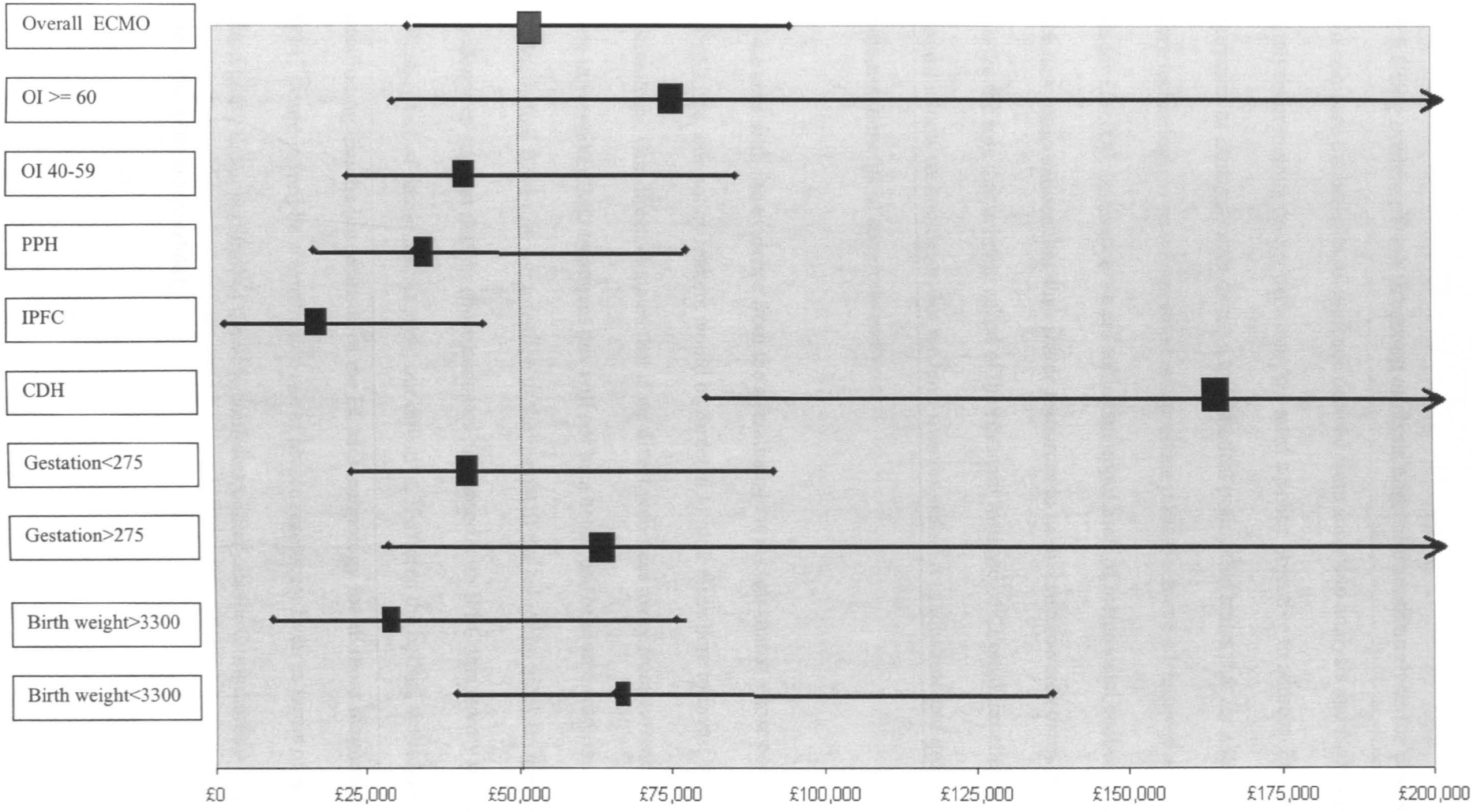
For details of simulations see appendix 2.

PAGE

NUMBERING

AS ORIGINAL

Figure 8.4: Mean ICERs and confidence intervals of main trial and all subgroups.



8.5 Using evidence from subgroup analysis and traditional analysis for policy

To compare the benefits of evidence derived from subgroup analysis and traditional analysis in making policy decisions, we need a policy objective to address. For the purposes of illustration let us use a hypothetical policy that states that for a new health care technology to be incorporated into practice it must achieve a life saved at a cost of under £50,000. In essence we are setting an upper limit of incremental cost-effectiveness for new interventions that wish scarce resources to be diverted towards them. This is not unlike the role that is being asked of the National Institute of Clinical Excellence with regard to new technologies, but we have oversimplified it to illustrate and quantify the potential benefits of subgroup analysis.

If we start with the evidence from the overall study it is clear that at a cost per life saved of £52,244, our policy makers would be directed to look elsewhere to invest in a new technology. Our directive states that if we divert resources away from current activities into universal ECMO treatment this will not be a better use of scarce resources. Nevertheless if we turn our attention to the results of the subgroup analyses there is evidence to suggest that for those neonates diagnosed with IPFC this is very much a cost-effective use of resources. As such, our directive, if offered the option, would be clear in concluding that the introduction of the ECMO technology for all those diagnosed with IPFC would indeed be a worthwhile use of scarce resources. Even in terms of measures of uncertainty the upper limit of the 95% confidence intervals for this subgroup does not cross the £50,000 threshold.

This brings us onto the issue of uncertainty. Although the mean cost per life saved for the PPH subgroup is just £33,356, the upper limit of the confidence interval extends beyond the £50,000 barrier (at £77,879). This would be thought of by many as a good reason to exclude the intervention as not being of statistical significance, whereas in reality 80% of the distribution around the mean is below the £50,000 barrier. There are additional examples of this with regards to birth weight over 3300 grams (85%), gestation over 275 days (70%) and OI category 40-59 (70%).

With all of these subgroups it is more likely than not that the ECMO technology would, for the patient in question, represent a cost-effective treatment option. The decision to proceed here, becomes a subjective one. Probability dictates that cost-effective treatment is likely, but not certain. The cost-effectiveness ratio of treating this baby with ECMO is likely to be very favourable.

Analysis of the incremental cost-effectiveness of ECMO shows us that with a hypothetical limitation of new health care technologies of £50,000 per life saved, ECMO would not be considered a cost-effective use of resources. In fact traditional statistical norms would probably not allow the introduction of this new technology if the limit were set at £60,000 because the 95% confidence intervals do not both lie inside this range.

Nevertheless the analysis of the subgroups show that within the study overall there are those for which the cost-effectiveness of ECMO has its mean value lying under the instituted limit of £50,000. In addition it also has the full range of it's confidence intervals (IPFC subgroup – 95% C.I. 1,958 – 44,430) under £50,000. If the allocation of resources

that leads to a state of optimisation of health benefit relies on performing interventions only if they provide that benefit at less than £50,000 then here is a group of patients and an intervention that fits the bill.

Under the same argument of optimisation, we are aware from evidence in table 8.6 that other groups of patients would more likely than not benefit from the intervention at a cost of less than £50,000. In the case of groups 2,4,6,7 and 8 patients are between 40% and 70% more likely to benefit at a cost of less than £50,000 per life saved than not. In total five of the nine subgroups can justifiably be thought to be more likely than not to receive a treatment that will turn out to be cost effective if undertaking ECMO.

So if our mean ratio as a whole is over the £50,000 barrier, at £52,244, how can it be the case that five from nine subgroups have means lying under the same barrier? The answer is that the mean of the whole sample is dragged up by those subgroups that are over the barrier to a greater extent than the groups that are under (i.e. there is a skewed distribution). This is likely to be common wherever you have a distribution that is heavily skewed to the right. As can be seen in figure 8.4 (and from table 8.6) those subgroup means that lie over the £50,000 have upper distribution intervals that extend away well over the £150,000 and three of them over the £200,000 line. In comparison, of the five groups that have means lying under the £50,000 line, none approach the zero line. From this a logical step may be to calculate the mean ICER for ECMO after truncating the distribution through the exclusion of outliers. This would inevitably bring down the overall mean for the intervention as a whole, but still would not add the value that the

evidence from the subgroup analysis does in pinpointing the areas of best investment for the ECMO technology.

The likelihood of this occurring can be anticipated by looking at figure 8.3, which shows that despite the mean lying over £50,000, the bulk of the histogram of the probability distribution lies under £50,000. This suggests that a randomly chosen case would be more likely to have a cost-effectiveness ratio of under £50,000 than one of over £50,000. Having acknowledged this it is the use of subgroup analysis that allows us to better identify those patients that lie under the £50,000 line and those that lie over it. This will not always be possible as in many circumstances there will be no defining covariates (as was the case in ECMO) and the distribution will be truly random.

The use of a hypothetical maximum tariff put on health benefit is just an example of one method that could be used to direct resource allocation towards the goal of welfare maximisation. It is one of the simpler ones, but whatever system were used to achieve this end, the concept behind how overall health gain can be increased within a limited budget from using subgroup evidence for directing resource allocation is still valid.

To take our example further let us assume that there was only one need for health care, and that was to maximise health benefits for new born infants with severe respiratory failure. There are two methods of treatment; conventional management and conventional management with the addition of ECMO. We are left with a policy decision with which to make the best use of our limited resources which involves choosing from one of three options.

1. Incorporate ECMO into standard care
2. Do not incorporate ECMO into standard care
3. Selectively incorporate ECMO into standard care

If the choice were between 1 and 2, then given that the average cost-effectiveness (ACER) of conventional treatment without ECMO is estimated at £18,000 per survivor without severe disability, and with ECMO it is estimated at £32,000 it is unlikely that ECMO would be funded. Nevertheless if the policy were selective within subgroups, then the marginal cost of ECMO would be considered an efficient use of resources within a limited budget for neonates diagnosed with IPFC. In fact every pound spent on ECMO in these groups would result in more health gain than a pound spent on conventional management.

Furthermore if we look at multi-characteristic subgroups such as neonates with a birth weight over 3300 grams and a diagnosis of IPFC, we find an ICER of just £9,513. Here, a pound spent on ECMO for this group of patients is likely to gain twice the health gain of a pound spent on conventional management. The probability is not 95% but the simulation shows it to be 70% likely that the ICER of ECMO for this group will fall under the ACER of conventional management. To continue with a strategy of not incorporating ECMO into treatment for such a group of patients is to disregard a likely improvement in overall health gain.

8.6 Comparing evidence from traditional and subgroup analysis

To measure the potential value of subgroup analysis there is a need to compare the results of a policy decision based on this method with a policy decision based on the more traditional methods. To do this we will take the evidence from both types of analysis and assume the appropriate decision based on our hypothetical directive. This is to ensure that new technologies must show that they save lives at a cost of less than £50,000. We also assume that this figure is chosen due to the fact that the health care that is already funded is done so at a level not above the cost of £50,000 for every life year gained.

The decision given traditional methods of analysis is a simple one. The ICER is £52,244 and as such the ECMO technology would not be considered cost-effective and as such there will be no change to resource allocation and no resulting increase or decrease in overall health gain.

In the case of subgroup analysis, the results can be used in two ways; to incorporate those subgroups for which the 95% confidence intervals fall below £50,000 (option one). Option two would be to incorporate those subgroups for which more than 50% of the likely distribution has a cost-effectiveness that falls below £50,000.

For option one this is simply the subgroup who are diagnosed with IPFC with a mean ICER of £16,340 (95% C.I. £1,958 – £44,430). If the ECMO technology was offered just to this group the resulting mean health gain per patient would be £33,660 (£50,000 - £16,340) per patient treated, or the equivalent to

$$= 1 \times \text{life saved} \times \frac{£33,660}{£50,000}$$

= 0.67 lives saved per patient treated.

We would also be 95% certain that the mean health gain would be at least £5,570 per patient treated or

$$= 1 \times \text{life saved} \times \frac{£5,570}{£50,000}$$

= 0.11 lives saved per patient treated.

Table 8.7: Mean ICERs for subgroups, proportionate confidence in mean and proportion of total patients.

Subgroup (no.)		*ICER*	*Percentile< £50,000/LS*	*Proportion of n (total)*	*Allocate resources to*
Birthweight	< 3300 (1)	65,627	20%	51%	No
	> 3300 (2)	28,738	85%	49%	Yes
Gestation	< 275days (3)	63,417	30%	46%	No
	>275 days (4)	39,281	70%	54%	Yes
Diagnosis	CDH (5)	164,900	0%	20%	No
	IPFC (6)	16,340	95%	17%	Yes
	PPH (7)	33,356	80%	63%	Yes
OI category	40-59 (8)	42,359	70%	60%	Yes
	>60 (9)	74,218	20%	40%	No
Total study (10)		£52,244	45%	100%	-

With option two the conclusion would be to incorporate the ECMO technology far more than with option one, as it would include subgroups 2,4,6,7 and 8. This would leave us with the policy summarised in table 8.7. We can show that the mean health gain per patient treated can be estimated if we assume the same proportions of each subgroup as was witnessed in the study. As with option one we can show the anticipated mean health

gain on the basis of resources gained or an equivalent lives saved per patient treated assuming that all other interventions currently being provided have a marginal cost-effectiveness of £50,000 per life saved.

Table 8.8: Mean ICERs for subgroups, proportionate confidence in mean, proportion of total patients and marginal lives saved per patient treated.

Subgroup selected (no.)		*ICER*	*Percentile< £50,000/LS*	*Proportion of n (total)*	*Marginal lives saved per patient treated*
Birthweight	> 3300 (2)	28,738	85%	49%	0.43
Gestation	>275 days (4)	39,281	70%	54%	0.21
Diagnosis	IPFC (6)	16,340	95%	17%	0.67
	PPH (7)	33,356	80%	63%	0.33
OI category	40-59 (8)	42,359	70%	60%	0.15
Total		-	80%	100%	0.27

The results in table 8.8 show the mean lives saved per subgroup per patient treated, and also the total mean health gain per person treated given the weighting of the proportion of total patients from each subgroup. However this figure of a mean 0.27 lives saved per person treated is not significant at the 5% level. We do however have 80% confidence that there will be a positive mean improvement in overall health gain given this strategy. In other words it is 80% probable that for every 100 babies treated with ECMO 27 lives (without severe disability) will be saved at a cost of less than £50,000 per life year saved. The obvious conclusion from this is that without the use of subgroup analysis in making a

decision on funding ECMO technology for certain groups of new born babies the result could be a not considerable number of lives lost.

8.7 Summary

We have shown in the analysis of this particular data set that when the goal of health care is to maximise health gain within the financial constraints set upon it, whatever they may be, the decision to treat is a complex one. The choice as to whether to introduce treatments in health care has historically tended to be a simple yes or no. This has merit in having a transparency and measure of robustness that ensures that policy makers are never seen to be making the wrong decision. The question this chapter and indeed this thesis puts forward is while being seen to be making the right decision, are we in fact denying ourselves the opportunity to get closer to achieving our primary goal, that of maximising health gain.

9. Discussion

9.1 Introduction

Policy makers with the goal of maximising health gain for a population require evidence on the health benefits of potential uses of scarce health care resources and the cost of achieving them. This evidence has historically been based on the average effectiveness of an intervention on a representative sample, with the assumption that questions about the value to a population of a given intervention should be answered with a single value. The problem with this is that it ignores the fact that a population is heterogeneous. Just as it is important for a doctor to relate the results of a trial specifically to the characteristics of the patient, it is imperative that policy makers take the evidence from economic evaluations and relate them to any given population or elements within that population.

The mean cost-effectiveness ratio of an intervention is the average of a group of patient specific total costs and patient specific measures of effectiveness for a heterogeneous sample. As such the population from which the study was sampled will often be made up of some patients for whom an intervention is highly cost-effective and some for which that intervention will not be cost-effective. The question is whether it is possible for us to predict those people for whom the cost-effectiveness ratios will be high and those people for which a specific intervention will have a low cost-effectiveness ratio. If this is possible, will the ability to do this lead to a better use of finite resources to improve overall health gain?

This thesis started with the hypothesis that policy decisions regarding resource allocation in health care could be improved by a greater concentration on the use of the results of subgroup analysis. It is argued that information on the heterogeneity of populations is valuable in maximising health gain. The methods used in economic evaluation have progressed enormously over the last twenty years, and one of the biggest areas of progress has been in the movement away from deterministic measurement of cost-effectiveness ratios and towards stochastic methods. It is these methods that have allowed us to measure uncertainty and show the distributions around the means of these ratios. The most common measure of this variance has been the confidence interval.

The purpose of the confidence interval, and of statistical inference in general, is to measure the degree of certainty we have in our estimated mean. The reason such a technique is required is that most of the things we wish to measure we cannot do with great accuracy, and as such attempt to cancel out this 'white noise' by making our measurement repeatedly and taking an average. If our method of measurement were 100% reliable this technique would not be required.

When collecting the patient specific costs and outcomes that are required for stochastic analysis the variation in our data is not just made up of the white noise of measurement distortion, but also includes genuine variation due to real heterogeneity of our sample. This in turn is due to the fact that true heterogeneity exists within a population. The use of statistical inference where we know that any variation will be due to the inadequacies of the measurement tool is sensible and scientific. Where we aware that the cause of variation is at least in part due to the heterogeneity of the variable being measured,

consideration must be given to whether this method is less reliable. In truth, we know that if another sample, however large, was used to measure this variable we could not guarantee a similar result without a similarly heterogeneous sample.

These issues raise a number of issues that relate both to study design for cost-effectiveness analyses and the interpretation of the results of these studies when translating them into policy. We look at both of these areas in turn.

9.2 Implications for cost-effectiveness analysis

As we seek to describe in more detail the relationship between specific health care technologies and their relative value in different situations we start to question the reliance on classical statistical inference methods, which rely so heavily on certainty. Where we have a closer grasp of what, and by how much, other determinant variables effect our chosen measure of outcome this will inevitably help to determine the extent to which that outcome was achieved by the intervention being studied.

The purpose of inference is to try to show that a relationship between an intervention and an outcome exists beyond any reasonable doubt. This will be the goal of a study where, prior to the start of the trial, our knowledge of any relationship is zero. However, this is rarely the case as we often do have knowledge of relationships between factors and likely outcomes, both in terms of cost (as shown in chapter 7) and in effects (as shown in chapter 8). As such what we are searching for is not so much proof as a greater belief in what we already anticipated.

In the previous chapter we compared the relative merits of traditional analysis against subgroup analysis. The conclusions were that subgroup analysis could achieve an improvement in overall health gain based on the hypothetical objectives described. These conclusions were based on a higher likelihood of a positive increase in health gain compared to traditional analysis. The difference was not measured by inference, but by probability. This brings us onto assessing the relative merits of policy decisions based on probability versus those based on the need to prove an outcome beyond any reasonable doubt.

9.3 Implications for policy decisions

This thesis argues for optimisation in the use of data from economic evaluations rather than the current aspiration for certainty. It is the evidence that an event is more likely than not to happen that most decisions are made on. More certainty becomes necessary when the costs of getting the decision wrong outweigh the benefits of getting it right. This is often the case in medical research, where society needs to be assured that a new drug is both effective and safe. The question remains is – is this a lesser concern when dealing with less dichotomous outcomes such as cost-effectiveness?

The costs of approving a new drug or technology that turn out to have significant negative side effects are likely to be high. However, the approving of a new drug or technology that has been shown to be effective based on less certain proof of its cost-effectiveness is unlikely to be considered by society as such a high cost if proved later to be incorrect.

It can be argued that it is just as important to measure cost-effectiveness correctly as it is to measure the effects of an intervention correctly as the results of getting it wrong are the misallocation of resources. This could lead to someone somewhere missing out on valuable treatment. This is not in dispute, but it is suggested that a measurement of a continuous variable such as a cost-effectiveness ratio is more transparent represented as a measure of probability. Whereas a dichotomous outcome, for example to answer whether something is to be considered safe or not, may benefit more from being measured using a method that has the intention of showing some degree of confidence or certainty.

In theory for every potential intervention and for every potential receiver of that intervention there is a specific relationship that will produce a unique cost-effectiveness ratio. The maximisation of health gain comes from finding the cost-effectiveness of every intervention for every individual within our population. Only then can decisions be made on a shift in resource allocation that will improve overall health gain. The problem can therefore be seen as one of availability of information more than of economic theory.

This thesis has shown that there is a wide variance in a singular measure of cost-effectiveness and therefore there exists the possibility for an improvement in health gain if decision-makers were given access to more of the information available in a form that would aid them in making decisions. It is suggested that a greater use of subgroup evidence should be made available to decision-makers. This use of more specific cost-effectiveness ratios for more specific groups within the population could be a movement towards welfare maximisation, which could result in greater efficiency in the delivery of health care.

9.4 In defence of subgroup analysis

A number of concerns could be raised in the increased use of subgroup analysis to inform policy decisions. Two of these are that of sample size and that of 'fishing' for results that may not be there. It is commonly believed that because subgroups have smaller sample sizes, the rigour of the results may suffer. Statistical power, however, relies on two things, effect size and sample size, and as such subgroups can often have narrower confidence intervals and by definition be more robust than the overall effect size of a study due to the greater effect size outweighing the smaller sample size. In addition the greater homogeneity of subgroups, chosen on the basis of reducing the heterogeneity present in the overall study population, will reduce the measure of variance (standard deviation) and correspondingly reduce the width of the confidence intervals.

In terms of the issue of 'fishing' it is imperative that subgroups are selected based on sound evidence as was done in chapters 7 & 8. On a more instinctive level it seems that those against the use of subgroup analysis are actually saying we should put our faith in statistical techniques for one group of numbers (the overall trial results), but that the techniques may be unreliable used in another set of numbers (the subgroups). This is both illogical and lacks objectivity.

9.5 Summary

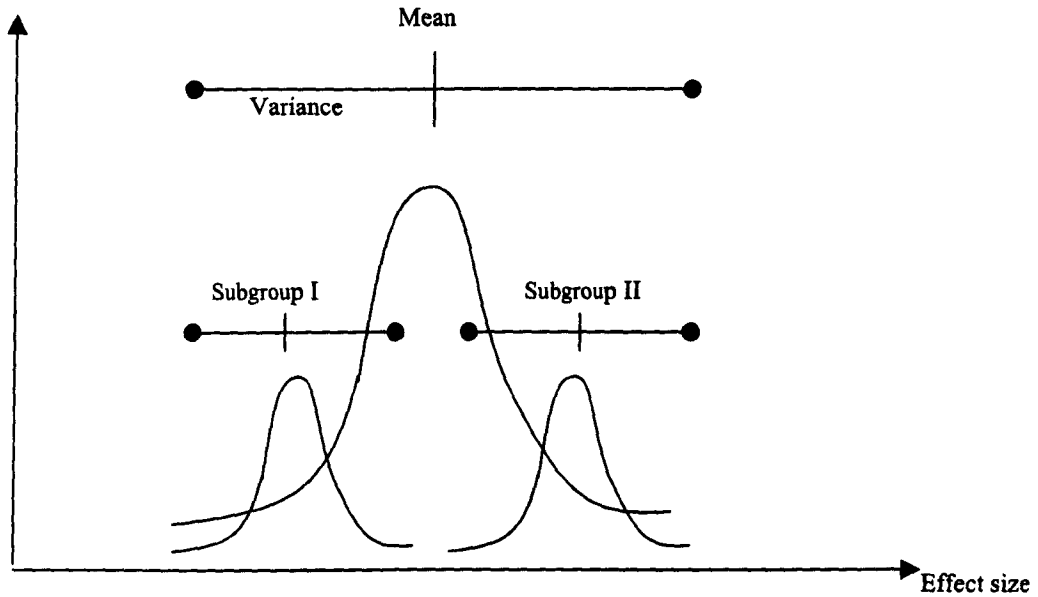
Inevitably the response to a proposal for a greater use of subgroup analysis as a means of increasing overall health gain is an argument for optimisation over certainty, for

probability over statistical significance. It is argued that it is imperative that we are able to assess whether an intervention can improve the health of a patient, but the question we really want an answer to is to what extent can it improve the health of the patient? The certainty associated with statistical significance can only be used to answer a normative question. If the p value is less than 0.05 or the confidence intervals do not include zero we are 95% certain we have an effective intervention. We are not 95% certain that the level of effectiveness is equivalent to the mean.

Such a normative answer would not allow us to assess the value of the intervention compared to others that could be funded in its place. In most cases we are using an estimate of the effectiveness of an intervention for decision making that has no greater specific validity than any other figure that might lie within the confidence range. The confidence is in the range not in the mean. As such the mean provides a measure of likelihood, not of certainty.

The use of such an estimate is not something that this thesis is intending to degrade. It is the belief of this thesis that most effect-size information at our disposal is not much more than a measure of likelihood and those measures of likelihood gained from subgroup analysis are just as robust and valuable as those gained from the totals of trials. It is also suggested that in many cases the measure of likelihood in a subgroup might be more robust than those based on a total, due to the variation of the subgroup being limited to the white noise of measurement distortion and not the effects of heterogeneity.

Figure 9.1: Mean, variance and distribution for total study and for subgroup I & II



For example, we have a variable that we believe to be a factor in the effect size of an intervention and we therefore define two subgroups (I & II) of those with and those without the presence of this variable. If the variable is a factor that does influence the effect size it is likely that the distribution around the mean effect within each of those subgroups will be narrower than that around those receiving the intervention as a whole as the variation will be limited solely to measurement distortion. As such the variance around that mean would be smaller (see figure 9.1). When an intervention is introduced on the basis of an anticipated effectiveness defined by the estimated mean, the benefit will be maximised where the true effectiveness lies closest to that estimated mean, or where the true variance from that mean is small.

The value of evidence from subgroup analysis will always be a matter of debate as far as statisticians are concerned because it is their objective to search for a definitive proof.

This they have subjectively categorised as 95% probable and for which in general we require sizeable sample sizes to achieve. In reality we must question whether the value of this definition of certainty has been pitched too high, and whether the trade-off of being able to use data to answer more questions with less certainty may indeed outweigh the benefits from answering fewer questions with greater certainty.

10. Conclusions

The goal of health economics and of most health systems is primarily one of welfare improvement with an underlying trend towards the maximisation of expected welfare. This thesis questions whether the current system of collecting and using evidence for policy making in health care is the best available method, given this goal, and concludes that it may well not be.

Given the heterogeneous nature of a population, it is likely that for most health care interventions the cost of care and the outcome from that intervention will vary widely as a result of this heterogeneity. The mean ICERs that we tend to use for policy decisions will therefore be surrounded by a distribution of patient specific ICERs that may differ greatly from the mean. This thesis believes the variation surrounding these ICERs can be traced specifically to the heterogeneity of the population, and in some instances be used to greater predict the likely ICER of an intervention for that population.

In addition it is believed that using the relationship between different characteristics of a population and the outcomes of different health care interventions, will lead to a more efficient use of health care resources. The under-utilisation of this information could mean that we continue with an inefficient use of health care resources.

This thesis suggests that there remains an over reliance in cost-effectiveness analysis on classical statistical techniques. This dependency on certainty stems from the fact that the methods involved were originally devised to measure dichotomous outcomes not continuous outcomes such as cost-effectiveness ratios.

It is argued that more consideration be given to the heterogeneous nature of populations when both analysing and using data from economic evaluations in making policy decisions. In addition it is suggested that optimisation in resource allocation can be improved with a greater concentration on likelihood of relative cost-effectiveness rather than the current over reliance on arbitrary measures of certainty.

It is the belief of this thesis that the implications of this given the evidence from chapter 8 are that the health benefits being produced from the scarce health care resources are not being maximised. It is also the belief of this thesis that this is due to the fact that the heterogeneity of populations are not considered enough in decision-making and that subgroup analysis is underused.

These points can only be confirmed with greater research into available evidence from large trials where subgroup analysis is likely to be advantageous, or where there are empirical predictors of outcomes. In addition there needs to be an audit of interventions or technologies which have been accepted into clinical practice due to its overall cost-effectiveness, but which may have subgroups of the population receiving treatment which may not be cost-effective for them and therefore not an optimal use of scarce resources. Such research would serve to validate the results of this thesis and potentially confirm the value of greater use of subgroup analysis in policy decision-making.

References

1 Morris J. and Goddard M. Economic evaluation and quality of life assessments in cancer clinical trials: the CHART trials, *European Journal Cancer*; **1993**; 29A(5):776-770.

2 Commonwealth of Australia. Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee: including submissions involving economic analysis. Canberra: Australian Government Printing Office, **1992**.

3 Schubert, F. History of the development of pharmacoeconomic guidelines, in: Scubert F. (ed.) Proceedings of the Canadian Collaborative Workshop on Pharmacoeconomics, *Experta Medica*. **1993**: 1-2.

4 Rutten FFH, Bonsel GJ. High cost technology in health care: a benefit or a burden? *Social Science in Medicine* **1992**; 4: 567-8.

5 A First Class Service: Quality in the new NHS, Department of Health, London HMSO **1999**.

6 Berenson R, Holohan J. Sources of the growth in Medicare Expenditures. *Journal of the American Medical Association*, **1992**; 267: 687-691.

7 Drummond MF, Davies L. Economic analysis alongside clinical trials: revisiting the methodological issues. *International Journal for Technological Assessment in Health Care*, **1991**;7 :561-573.

8 Eisenberg JM. Clinical Economics, A Guide to the Economic Analysis of Clinical Practices. *Journal of the American Medical Association*, **1989** ;262: 2879-2886.

9 Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. *British Medical Journal*, **1995**; 311: 1356-9.

10 Dawes R, Faust D, Meehl PE. Clinical versus actuarial judgement. *Science* **1989**; 243: 1668-74.

11 Michaud C, Murray CJL, Resources for health research and development in 1992: a global overview. Investing in health research and development: report of the ad hoc committee on health research relating to future intervention options. Geneva: World Health Organisation. **1996.**

12 Lomas J. Words without action? The production, dissemination and impact of consensus recommendations. *Annual Review of Public Health* **1991**; 12 : 41-65

13 Eddy DM, Clinical policies and the quality of clinical practice. *New England Journal of Medicine*. **1982**; 307 :343-347

14 Haynes B, Haines A, Barriers and bridges to evidence based clinical practice. *British Medical Journal.* **1998**; 317: 273-276.

15 Bero LA, Grilli R, Grimshaw JM, Harvey E, Oxman AD, Thomson MA. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. *British Medical Journal*, **1998**; 317: 465-468.

16 Sassi F, McKee M, Roberts J, Economic evaluation of diagnostic technology. Methodological challenges and viable solutions. *International Journal of Technology Assessment in Health Care* **1997**; 13(4): 613-30.

17 Hill SR, Mitchell AS, Henry DA. Problems with the interpretation of pharmacoeconomic analyses: a review of submissions to the Australian Pharmaceutical Benefits Scheme. *Journal of the American Medical Association* **2000** ;283(16):2116-21

18 Kaldor N. Welfare propositions of economics and interpersonal comparisons of utility. *Economic Journal* **1939**; 49:549-551.

19 Hicks JR The valuation of the social income *Economica* **1940**;7:105-124

20 Diener A, O'Brien B, Gafni A. Health care contingent valuation studies; a review and classification of the literature. *Health Economics* **1998**; 7(4): 313-26.

21 Drummond, M.F. et al. Standardising economic evaluation methodologies in health care: practice problems and potential. *International Journal of Technology Assessment in Health Care*, **1993**; 9:26-36.

22 O'Brien, B.J. Drummond, M.F. Labelle, R.J. Willan, A. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care* **1994**; 32(2) : 150-164.

23 Wakker, P. Klassen, M. Confidence intervals for cost-effectiveness ratios. *Health Economics* **1995**; 4(5): 373-382.

24 Chaudhary, M.A. Stearns, S.C. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomised controlled trial. *Statistics in Medicine* **1996**; 15: 1447-1458.

25 Briggs, A.H. Wonderling, D.E. Mooney, C.Z. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Economics* **1997**; 6(4): 327-340.

26 Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Economics* **1996**; 5(6): 513-524.

27 Briggs, A. Fenn, P. Confidence Intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Economics* **1998**; 7: 723 740.

28 Lilford RJ, Braunholtsz D. The statistical basis of public policy: a paradigm shift is overdue. *British Medical Journal* **1996**; 313(7057): 603-607.

29 Hornberger JC, Brown BW Jr, Halpern J. Designing a cost-effective clinical trial. *Statistics in Medicine* **1995**; 14(20):2249-2259.

30 Van Hout VA, Al MJ, Gordon GS, Rutten FF. Cost, effects and C/E ratios alongside a clinical trial. *Health Economics* **1994**; 3(5): 309-319.

31 McCloskey D. The loss function has been mislaid: the rhetoric of significance tests. *American Economics Review* **1985**; 75(2):201-205.

32 Ludbrook A. Mooney G. Economic Appraisal in the NHS: Problems and Challenges, Northern Health Economics, Health Economics Research Unit, University of Aberdeen, Aberdeen, **1984**.

33 Drummond M. Economic evaluation and the rational diffusion and use of health technology, *Health Policy*, **1987**; 7:309-324.

34 Drummond M. Assessing efficiency in the new national health service, Discussion paper 75, Centre for Health Economics, University of York, York. **1990**

35 Gerard K. Cost utility in practice: a policy maker's guide to the state of the art. *Health Policy*, **1992**; 21:249-280.

36 Ross J. The use of economic evaluation in health care: Australian decision makers' perceptions. *Health Policy*, **1995**;31:103-110

37 Commonwealth Department of Health, Housing and Community Services, Guidelines for the Pharmaceutical Benefits Advisory Committee, Including Submissions Involving Economic Analyses, Australian Government Publishing Service, Canberra, **1992**.

38 Backhouse ME, Backhouse RJ, Edey SA. Economic Evaluation Bibliography. *Health Economics*, **1992**; 1(Supplement): 1-236.

39 Stewart, A. Schmier, JK Luce, BR. A survey of standards and guidelines for cost-effectiveness analysis in health care. *American Heart Journal*, **1999**;137:S53-S61

40 Goldhirsch A, Gelber RD, Simes RJ, Glasziou P, Coates AS. Costs and benefits of adjuvant therapy in breast cancer: a quality-adjusted survival analysis. *Journal of Clinical Oncology*, **1989**;17: 1716-1717.

41 Hutton J. Economic evaluation of health care: a halfway technology. *Health Economics*, **1994**; 3(1):1-4.

42 Yusuf S, Collins R, Peto R. Why do we need some large, simple randomised controlled trials? *Statistics in Medicine.* **1984**;3:409-20

43 Counsell CE, Clarke MJ, Slattery J, Sandercock PAG. The miracle of DICE therapy for acute stroke: fact or fictional product of subgroup analysis, *British Medical Journal* **1994**;309:1677-81.

44 Buyse ME. Analysis of clinical trial outcomes: some comments on subgroup analyses. *Controlled clinical trials.***1989**;10:187S-194S

45 Bulpitt, C.J. Subgroup analysis. *Lancet* **1988**;2:31-34.

46 Buyse ME. Analysis of clinical trial outcomes: some comments on subgroup analyses. *Controlled Clinical Trials.* **1989**;10:187S-194S

47 Oxman, AD., Guyatt, GH. A consumers guide to subgroup analyses. *Annals of Internal Medicine.* **1992**;116:78-84.

48 Donner A., A Bayesian approach to the interpretation of subgroup results in clinical trials. *Journal of Chronic Diseases.* **1982**;34:429-35.

49 Davis CE, Leffingwell DP. Empirical Bayes estimates of subgroup effects in clinical trials. *Controlled Clinical Trials.* **1990**;11:37-42.

50 Barnett, V. Comparative Statistical Inference, John Wiley & Sons, London, **1973**

51 Barnett, V. Comparative Statistical Inference, John Wiley & Sons, London, **1973**

52 Venn, J. The Logic of Chance. New York: Chelsea, **1962** (Reprint of Macmillan, London, 1888)

53 von Mises, R. Probability, Statistics and Truth. 2nd Edition, George Allen & Unwin, London, **1957**.

54 von Mises, R. Probability, Statistics and Truth. 2nd Edition, George Allen & Unwin, London, **1957**.

55 Ramsey, F.P. ' Truth and Probability' in Kyburg, H.E. and Smokler, H.E. (eds) Studies in Subjective Probability, Wiley, New York, **1964**.

56 Barnett, V. Comparative Statistical Inference, John Wiley & Sons, London, **1973**

57 Thatcher, A.R. Relationships between Bayesian and confidence limits for predictions, *Journal of the Royal Statistical Society*, **1964**, 26; 176-210.

58 Cox, D.R. Some problems connected with statistical inference, *Annals of Mathematics and Statistics*, **1958**, 29 ; 357-372.

59 Donner, A. A Bayesian approach to the interpretation of subgroup results in clinical trials. *Journal of Chronic Disease*, **1982** ; 35 ; 429-435.

60 Donner, A. The design of a clinical trial with several patient categories. *Canadian Journal of Statistics*, **1979**, 7 ;169.

61 DeGroot, M. Optimal Statistical Decisions. McGraw-Hill, New York, **1970.**

62 Petitti, D. B. Meta-Analysis, Decision Analysis and Cost-effectiveness Analysis, Methods for Quantitative Synthesis in Medicine. Oxford University Press, Oxford, **1994.**

63 Raiffa H. Decision Analysis, Introductory Lectures on Choices under Uncertainty. Addison-Wesley, Reading, Massachusetts, **1970.**

64 Counsell CE, Clarke M, Slattery J, Sandercock PAG. The miracle of DICE therapy for acute stroke: fact or fictional product of subgroup analysis? *British Medical Journal* **1994**; 309: 1677-1681.

65 Loomes G, McKenzie L. The use of QALYs in health care decision-making. *Social Science & Medicine* **1989**; 28: 299-308.

66 Larson EB. Dealing with limited resources. The Oregon decision to curtail funding for organ transplantation. *New England Journal of Medicine* **1988**; 319: 171-173.

67 Polsky D, Glick HA, Willke R, Schulman K. Confidence intervals for cost-effectiveness ratios: a comparison of four methods. *Health Economics* **1997**; 6: 243-252.

68 Chaudary MA, Stearns SC. Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial. *Statistics in Medicine* **1995**; 15: 1447-1458.

69 Buyse ME. Analysis of clinical trial outcomes: some comments on subgroup analysis. *Controlled Clinical Trials* **1989**; 10: S187-19.

70 Oxman AD, Guyatt GH. A consumer's guide to subgroup analysis. *Annals of Internal Medicine* **1992**; 116: 78-84.

71 Bulpitt C. Subgroup analysis. *Lancet* **1988**; 31-34.

72 Michels KB, Rosner BA. Data trawling: to fish or not to fish. *Lancet* **1996**; 348: 1152-3.

73 Peto R. Misleading subgroup analyses in GISSI; and Mauri F, Gasparini M, Barnoaglia L et al, Reply. *American Journal of Cardiology* **1990**; 771-2.

74 Fry, TC. A mathematical theory of rational inference. Notes from a lecture given to the Econometric Society. Chicago, **1933**.

75 Bulpitt C.J. Subgroup analysis. *Lancet* **1988**; 2:31-34.

76 Counsell CE, Clarke MJ, Slattery J, Sandercock PAG. The miracle of DICE therapy for acute stroke: fact or fictional product of subgroup analysis, *British Medical Journal*, **1994**; 309:1677-81.

77 Lee, KL, McNeer, JF, Starmer, CF, Harris, PJ, Rosati, RA. Clinical judgement and statistics - lessons from a simulated randomized controlled trial in coronary heart disease. *Circulation*, **1980**, 61(3); 508-514.

78 Morris CN. Parametric empirical Bayes inference: theory and application. *Journal of the American Statistical Association*, **1983**, 78 ; 47-55

79 Rubin DB, Estimation in parallel randomised experiments. *Journal of Educational Statistics*. **1981**, 6 ;377-401.

80 Laird NM, Louis TA. Empirical Bayes confidence intervals based on bootstrap samples. *Journal of American Statistical Association*. **1987**, 82; 739-750.

81 Coyle D. Statistical analysis in pharmacoeconomic studies. *Pharmacoeconomics* **1996**: 9; 506-516.

82 Briggs A, Gray, A. The distribution of health care costs and their statistical analysis for economic evaluation. *Journal of Health Services Research & Policy*. **1998**: 3(4); 233-245.

83 Smith TD, Taub DR. Activity-based costing; a more accurate alternative. *Strategic Healthcare Excellence* **1999**; 12(2): 8-12.

84 Moore R, Marriott N. Cost and price in the NHS: the importance of monetary value in the decision-making framework--the case of purchasing renal replacement therapy. *Health Services Management Research* **1999**; 12(1): 1-14.

85 Phelan P.D, Tate R, Webster F, Marshall R.P. DRG cost weights-getting it right. *Medical Journal of Australia* **1998**;169 Suppl:S36-8.

86 Drummond MF. Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. *British Medical Journal* **1996**: 313; 275-283.

87 Sassi F, McKee M, Roberts JA. Economic evaluation of diagnostic technology, methodological challenges and viable solutions. *International Journal of Technology Assessment in Health Care* **1997**; 13(4): 613-30.

88 Beck EJ, Beecham J, Mandalia S, Griffith R, Walters MD, Boulton M, Miller DL. What is the cost of getting the price wrong? *Journal of Public Health Medicine* **1999** ;21(3):311-7.

89 Altman DG, Bland JM. Generalisation and extrapolation. *British Medical Journal* **1998**;317(7155):409-10.

90 Paul SD, Eagle KA, Guidry U, DiSalvo TG, Villarreal-Levy G, Smith AJ, O'Donnell CJ, Mahjoub ZA, Muluk V, Newell JB, et al Do gender-based differences in presentation and management influence predictors of hospitalisation costs and length of stay after an acute myocardial infarction? *American Journal of Cardiology* **1995**;76(16):1122-5

91 Haigh R, Castleden M, Woods K, Fletcher S, Bowns I, Gibson M, Soper J Management of myocardial infarction in the elderly: admission and outcome on a coronary care unit. *Health Trends* **1992**;23(4):154-7.

92 Heeschen C, Hamm CW, Goldmann BU, Moeller RH, Meinertz T [Cost-effectiveness of a rapid test for troponin in emergency admissions]. *Deutsch Medica Wochenschr* **1998** ;123(42):1229-34

93 Mattera JA, Arain SA, Sinusas AJ, Finta L, Wackers FJ. Exercise testing with myocardial perfusion imaging in patients with normal baseline electrocardiograms: cost savings with a stepwise diagnostic strategy. *Journal of Nuclear Cardiology* **1998**;5(5):498-506.

94 McCullough PA, Ayad O, O'Neill WW, Goldstein JA Costs and outcomes of patients admitted with chest pain and essentially normal electrocardiograms. *Clinical Cardiology* **1998** ;21(1):22-6

95 Stark ME, Vacek JL. The initial electrocardiogram during admission for myocardial infarction. Use as a predictor of clinical course and facility utilisation. *Archives of Internal Medicine* **1987**;147(5):843-6.

96 Mulley AG, Thibault GE, Hughes RA, Barnett GO, Reder VA, Sherman EL The course of patients with suspected myocardial infarction. The identification of low-risk patients for early transfer from intensive care. *New England Journal of Medicine* **1980**;302(17):943-8

97 Stevens W, Langham S, Normand CE. Calculating the cost of CHD in the North Thames region. Report to the NHSME Regional Office **1998.**

98 Brunelli C, Spallarossa P, Pasdera A, Bezante GP, Zorzet F, Rossettin P [Treatment aspects of unstable angina. Costs and payments for DRG]. *Cardiologia* **1998**;43(1):67-75.

99 Chartered Institute of Public Finance Accountancy. Health Database **1998**, Health Services Financial Database, London.

100 UK Collaborative ECMO (Extracorporeal Membrane Oxygenation) Trial Group. UK collaborative randomised trial of neonatal extracorporeal membrane oxygenation. *Lancet* **1996**; 348: 75-82.

101 Howard S, Mugford M, Normand C, Elbourne D, Grant A, Field D, et al. A cost-effectiveness analysis of neonatal ECMO using existing evidence. *International Journal of Technology & Assessment in Health Care* **1996**; 12: 80-92.

102 Howard S, Normand C, Mugford M, Elbourne D, Field D, Johnson A, et al. Costing neonatal care alongside the collaborative ECMO trial: how much primary research is required? *Health Economics* **1995**; 4: 265-271.

103 Roberts, T.E. Economic evaluation and randomised controlled trial of extracorporeal membrane oxygenation: UK collaborative trial. The Extracorporeal Membrane Oxygenation Economics Working Group. *British Medical Journal.* **1998**;317(7163):911-5.

104 O'Brien, B.J. Drummond, M.F. Labelle, R.J. Willan, A. In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care* **1994**: 32(2);150-163.

105 Efron B, Gong G. A leisurely look at the bootstrap, jack-knife and cross validation. *The American Statistician* **1988**;37:36-38.

106 Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Economics* **1997**;6(4):327-40.

107 Efron, B. Tibshirani, R. An introduction to the Bootstrap. New York. Chapman and Hall **1993**.

108 Peek GJ, Sosnowski AW. Extracorporeal membrane oxygenation for paediatric respiratory failure. *British Medical Bulletin* **1997**;53(4):745-56.

109 The Congenital Diaphragmatic Hernia Study Group. Does extracorporeal membrane oxygenation improve survival in neonates with congenital diaphragmatic hernia? *Journal of Paediatric Surgery* **1999**;34(5):720-4.

110 Pearson GA, Firmin RK, Sosonowski A, Field D. Neonatal extracorporeal membrane oxygenation. *British Journal of Hospital Medicine* **1992** ; 47(9):646-53.

111 Ahmed A, Gangitano E, Odell RM, Doran R, Durand M. Survival, intracranial lesions and neurodevelopmental outcome in infants with congenital diaphragmatic hernia treated with extracorporeal membrane oxygenation. *Journal of Perinatology* **1999**;19(6 pt1):436-40.

112 Nield TA, Langenbacher D, Poulson MK, Platzker AC. Neurodevelopmental outcome at 3.5 years of age in children treated with extracorporeal life support: relationship to primary diagnosis. *Journal of Paediatrics* **2000**; 136(3): 338-44.

113 Zahraa JN, Moler FW, Annich GM, Maxvold NJ, Bartlett RH, Custer JR. Venovenous versus venoarterial extacorporeal life support for paediatric respiratory failure: are there differences in survival and acute complications? *Critical Care medicine* **2000**; 28(2): 521-5.

114 Kumar P, Shankaran S, Bedard MP, Delaney-Black V. Identifying at risk infants following neonatal extracorporeal membrane oxygenation. *Journal of Perinatology* **1999**; 19(5):367-72.

115 Trittenwein G, Pansi H, Graf B, Golej J, Burda G, Hermon M, Marx M, Wollenek G, Trittenwein H, Pollak A. Proposed entry criteria for postoperative cardiac extracorporeal membrane oxygenation after paediatric open heart surgery. *Artificial Organs* **1999**; 23 (11):1010-4.

Appendix 1

Results of simulations of PRAIS UK data set

Figures 1 – 6 Histograms of simulation results

Data tables of PRAIS UK

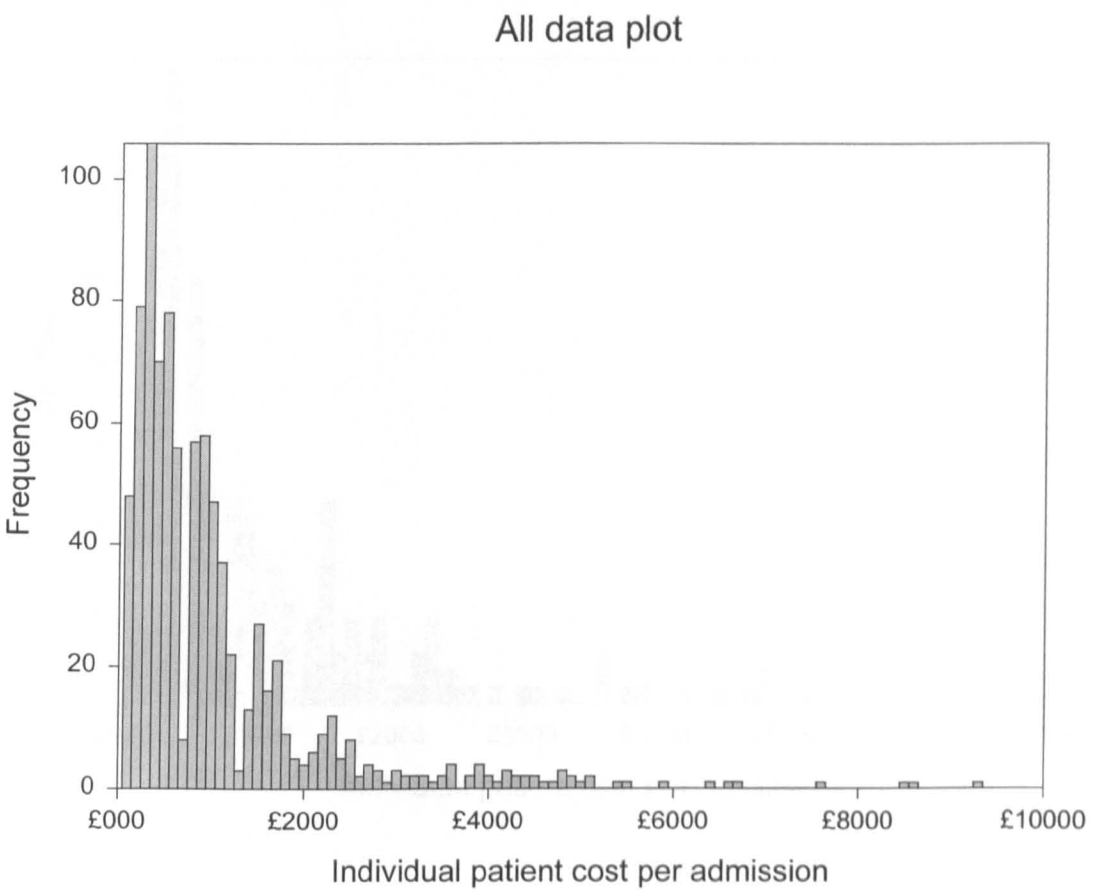


Figure 1: Histogram of total sample – PRAIS UK

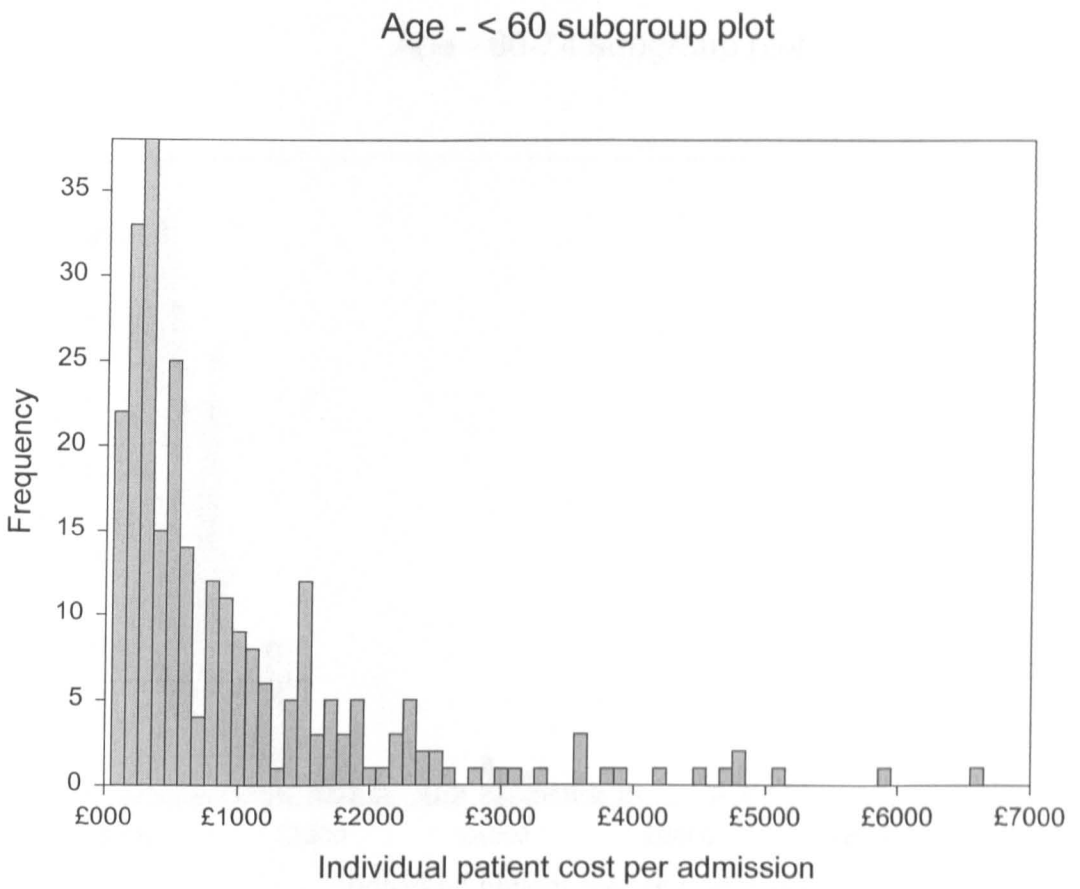


Figure 2: Histogram of subgroup by age: < 60 – PRAIS UK

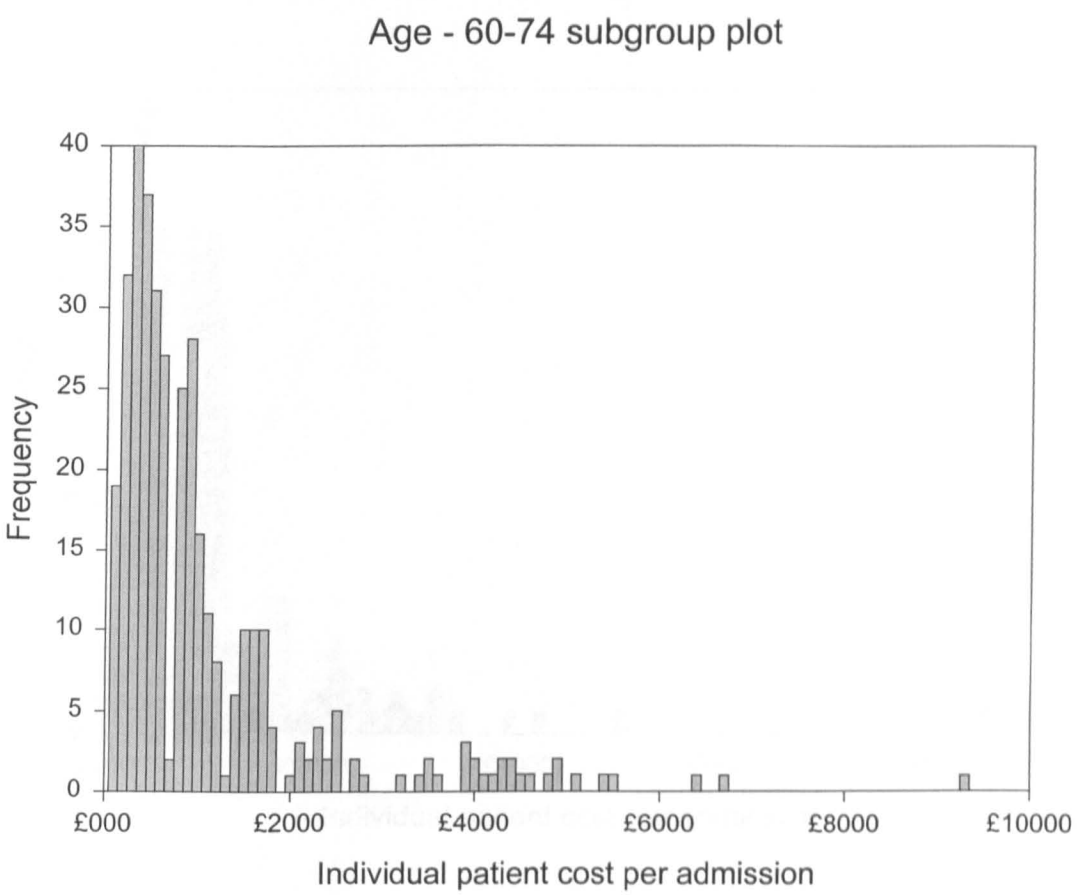


Figure 3: Histogram of subgroup by age: 60-74 – PRAIS UK

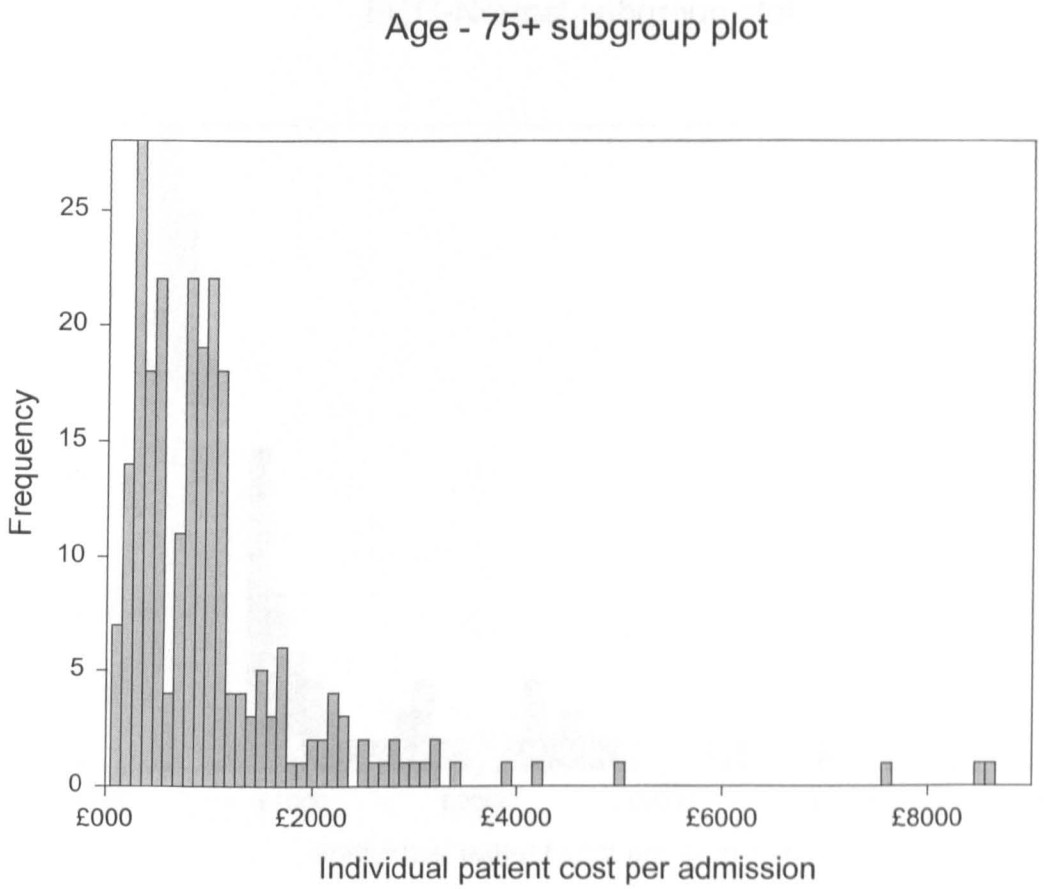


Figure 4: Histogram of subgroup by age: 75+ – PRAIS UK

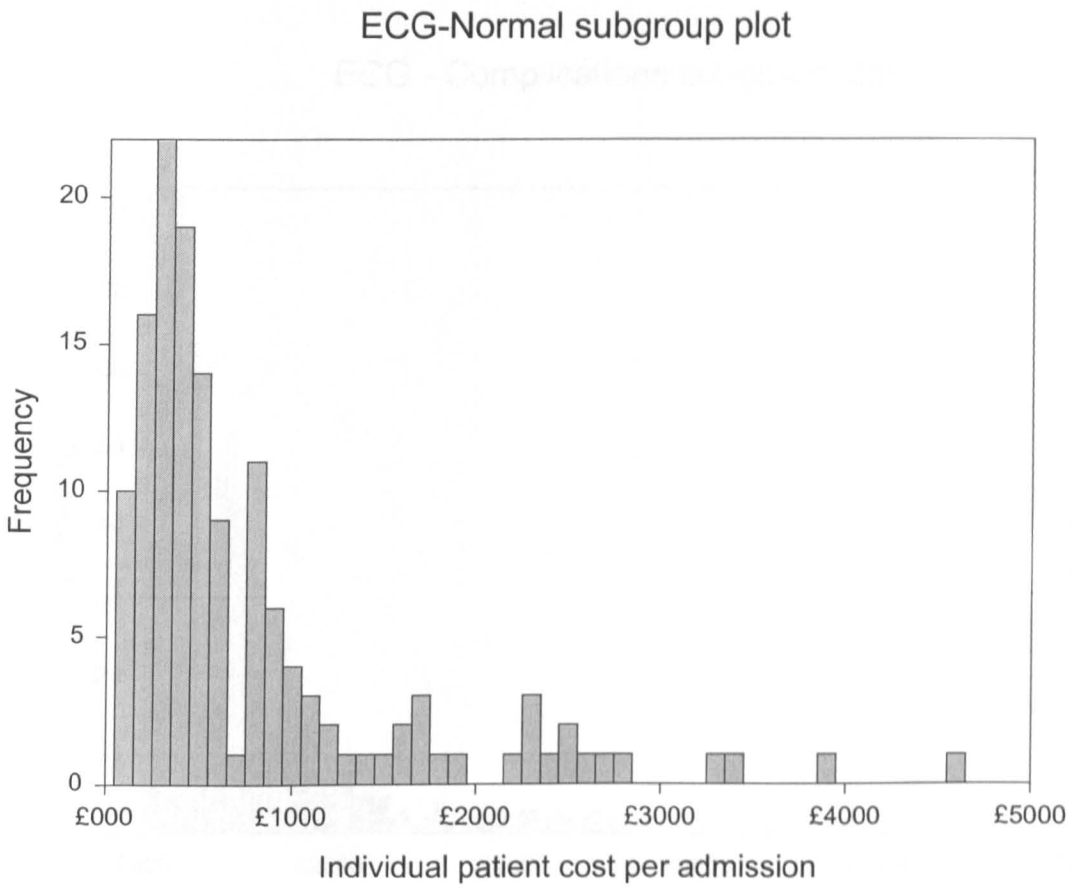


Figure 5: Histogram of subgroup by ECG: Normal – PRAIS UK

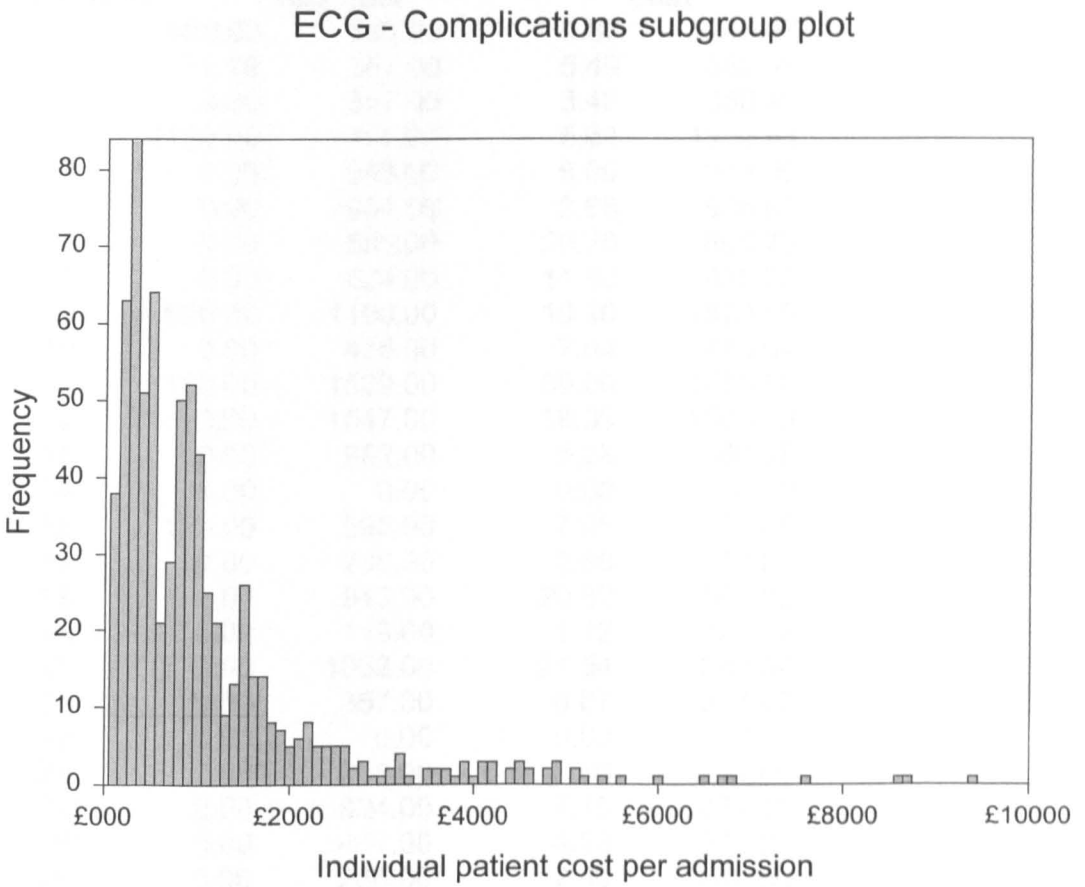


Figure 6: Histogram of subgroup by ECG: complications – PRAIS UK

PRAIS UK study data (n=1046)

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
1	500.00	833.00	9.87	1342.87
2	0.00	357.00	5.49	362.49
3	0.00	357.00	3.42	360.42
4	1428.00	476.00	5.64	1909.64
5	0.00	943.00	6.06	949.06
6	0.00	934.00	2.88	936.88
7	0.00	586.00	20.79	606.79
8	0.00	824.00	11.60	835.60
9	620.00	1190.00	10.10	1820.10
10	0.00	476.00	7.04	483.04
11	120.00	1529.00	89.50	1738.50
12	0.00	1547.00	18.33	1565.33
13	0.00	952.00	8.08	960.08
14	0.00	0.00	0.00	0.00
16	120.00	595.00	7.05	722.05
17	0.00	238.00	2.86	240.86
18	0.00	943.00	20.82	963.82
19	0.00	119.00	1.12	120.12
20	0.00	1062.00	21.54	1083.54
21	0.00	357.00	0.87	357.87
22	0.00	0.00	0.00	0.00
23	0.00	238.00	3.66	241.66
24	0.00	824.00	7.15	831.15
25	0.00	467.00	4.06	471.06
26	0.00	238.00	2.90	240.90
27	1428.00	238.00	2.24	1668.24
28	0.00	833.00	12.32	845.32
29	0.00	952.00	17.28	969.28
30	120.00	476.00	4.48	600.48
31	120.00	943.00	26.94	1089.94
32	500.00	952.00	11.28	1463.28
33	0.00	595.00	3.60	598.60
34	120.00	824.00	25.30	969.30
35	0.00	238.00	2.94	240.94
36	0.00	595.00	5.05	600.05
37	0.00	595.00	5.70	600.70
38	120.00	1172.00	8.58	1300.58
39	0.00	595.00	5.90	600.90
40	1548.00	2133.00	53.30	3734.30
41	1548.00	2948.00	70.78	4566.78
42	0.00	2234.00	66.53	2300.53
43	0.00	1776.00	34.83	1810.83
44	0.00	1520.00	7.07	1527.07
45	1428.00	1172.00	41.58	2641.58
46	0.00	2600.00	60.72	2660.72
47	0.00	714.00	10.98	724.98

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
48	1428.00	357.00	2.37	1787.37
49	0.00	1904.00	22.88	1926.88
50	0.00	357.00	1.56	358.56
51	1428.00	467.00	20.82	1915.82
52	1428.00	2023.00	13.77	3464.77
53	0.00	714.00	0.60	714.60
54	0.00	357.00	5.10	362.10
55	0.00	467.00	19.32	486.32
56	0.00	833.00	7.98	840.98
57	120.00	595.00	3.95	718.95
58	0.00	952.00	6.48	958.48
59	0.00	2838.00	61.60	2899.60
60	120.00	119.00	0.52	239.52
61	0.00	119.00	0.79	119.79
62	0.00	467.00	2.28	469.28
63	0.00	238.00	0.78	238.78
64	0.00	119.00	0.50	119.50
65	0.00	348.00	0.52	348.52
66	0.00	943.00	4.74	947.74
67	0.00	1547.00	10.53	1557.53
68	0.00	1520.00	2.73	1522.73
69	0.00	1529.00	3.51	1532.51
70	0.00	476.00	1.56	477.56
71	0.00	824.00	7.70	831.70
72	0.00	357.00	4.80	361.80
73	0.00	824.00	1.45	825.45
74	0.00	467.00	4.28	471.28
75	0.00	1172.00	2.34	1174.34
76	120.00	1300.00	27.39	1447.39
77	0.00	824.00	3.95	827.95
78	0.00	943.00	3.00	946.00
79	120.00	357.00	2.43	479.43
80	0.00	0.00	14.52	14.52
81	0.00	952.00	6.48	958.48
82	0.00	467.00	0.78	467.78
83	0.00	357.00	3.36	360.36
84	0.00	2243.00	64.50	2307.50
85	0.00	476.00	3.16	479.16
86	0.00	3561.00	33.88	3594.88
87	0.00	3406.00	97.89	3503.89
88	0.00	595.00	6.05	601.05
89	2810.00	1904.00	20.00	4734.00
91	500.00	476.00	0.40	976.40
92	120.00	476.00	0.40	596.40
93	500.00	0.00	1.58	501.58
94	0.00	476.00	3.24	479.24
95	0.00	1291.00	38.53	1329.53
96	120.00	952.00	9.68	1081.68

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
97	2048.00	1181.00	22.98	3251.98
98	0.00	595.00	4.05	599.05
99	2810.00	1181.00	17.30	4008.30
100	0.00	476.00	4.84	480.84
101	0.00	0.00	0.00	0.00
102	0.00	119.00	0.81	119.81
103	0.00	238.00	0.78	238.78
104	0.00	1071.00	3.51	1074.51
105	0.00	0.00	0.00	0.00
106	0.00	1071.00	14.40	1085.40
107	0.00	348.00	0.39	348.39
108	0.00	348.00	0.29	348.29
109	0.00	119.00	0.79	119.79
110	0.00	0.00	0.00	0.00
111	0.00	833.00	10.78	843.78
112	0.00	238.00	1.62	239.62
113	0.00	1547.00	10.53	1557.53
114	0.00	595.00	7.70	602.70
115	0.00	586.00	2.43	588.43
116	0.00	348.00	1.21	349.21
117	0.00	238.00	1.58	239.58
118	0.00	0.00	0.00	0.00
119	0.00	119.00	0.39	119.39
120	0.00	119.00	1.21	120.21
121	120.00	238.00	3.08	361.08
122	0.00	586.00	21.99	607.99
123	0.00	1062.00	5.67	1067.67
124	0.00	943.00	25.08	968.08
125	0.00	357.00	5.49	362.49
126	120.00	586.00	20.73	726.73
127	0.00	586.00	21.78	607.78
128	0.00	595.00	7.15	602.15
129	0.00	476.00	6.16	482.16
130	0.00	476.00	4.56	480.56
131	0.00	824.00	7.05	831.05
132	0.00	824.00	9.15	833.15
133	1428.00	357.00	4.35	1789.35
135	1428.00	833.00	9.45	2270.45
136	1428.00	1053.00	40.15	2521.15
137	0.00	714.00	8.58	722.58
138	0.00	705.00	1.56	706.56
139	1428.00	595.00	10.80	2033.80
140	0.00	357.00	5.49	362.49
141	0.00	934.00	1.56	935.56
142	0.00	357.00	4.74	361.74
143	1428.00	833.00	8.47	2269.47
144	0.00	119.00	0.39	119.39
145	0.00	357.00	2.43	359.43

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
146	0.00	0.00	0.00	0.00
147	0.00	476.00	6.16	482.16
148	0.00	586.00	3.42	589.42
149	120.00	714.00	4.74	838.74
150	0.00	348.00	0.39	348.39
151	0.00	714.00	12.18	726.18
152	0.00	705.00	3.24	708.24
153	120.00	238.00	1.58	359.58
154	0.00	595.00	4.05	599.05
155	0.00	467.00	1.62	468.62
156	0.00	357.00	2.43	359.43
157	0.00	714.00	2.34	716.34
158	0.00	1190.00	19.30	1209.30
159	120.00	595.00	3.95	718.95
160	120.00	1859.00	4.74	1983.74
161	120.00	238.00	0.58	358.58
162	0.00	714.00	9.24	723.24
163	0.00	238.00	2.28	240.28
164	0.00	714.00	4.74	718.74
165	0.00	238.00	1.62	239.62
166	0.00	714.00	7.26	721.26
167	0.00	238.00	0.58	238.58
168	0.00	705.00	6.32	711.32
169	500.00	0.00	0.85	500.85
170	0.00	595.00	6.05	601.05
171	0.00	119.00	1.21	120.21
172	0.00	357.00	4.62	361.62
173	120.00	0.00	0.00	120.00
174	120.00	943.00	25.20	1088.20
175	500.00	0.00	0.00	500.00
176	620.00	1071.00	12.69	1703.69
177	0.00	714.00	8.64	722.64
178	0.00	1053.00	5.20	1058.20
179	0.00	238.00	1.58	239.58
180	0.00	119.00	0.79	119.79
181	0.00	119.00	1.43	120.43
182	0.00	586.00	4.29	590.29
183	0.00	238.00	2.02	240.02
184	0.00	357.00	3.03	360.03
185	0.00	357.00	2.43	359.43
186	0.00	0.00	0.00	0.00
187	0.00	238.00	1.62	239.62
188	0.00	586.00	5.49	591.49
189	0.00	1291.00	40.07	1331.07
190	0.00	705.00	8.64	713.64
191	0.00	952.00	11.28	963.28
192	0.00	705.00	8.64	713.64
193	0.00	238.00	2.86	240.86

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
194	0.00	705.00	6.96	711.96
195	0.00	238.00	1.58	239.58
196	0.00	1309.00	22.66	1331.66
197	0.00	943.00	8.46	951.46
198	0.00	705.00	3.16	708.16
199	2810.00	3689.00	56.73	6555.73
200	120.00	586.00	4.23	710.23
201	120.00	238.00	1.58	359.58
202	0.00	4156.00	29.37	4185.37
203	0.00	1062.00	7.77	1069.77
204	120.00	238.00	1.42	359.42
205	2310.00	2820.00	43.52	5173.52
206	0.00	476.00	1.56	477.56
207	0.00	1776.00	32.23	1808.23
208	0.00	476.00	4.84	480.84
209	1928.00	238.00	2.28	2168.28
210	500.00	1895.00	15.68	2410.68
211	0.00	934.00	3.24	937.24
212	0.00	467.00	1.04	468.04
213	500.00	5485.00	247.68	6232.68
214	500.00	714.00	10.98	1224.98
215	0.00	1282.00	10.05	1292.05
216	500.00	1282.00	58.65	1840.65
217	0.00	815.00	2.07	817.07
218	2048.00	815.00	1.50	2864.50
219	500.00	1053.00	48.25	1601.25
220	0.00	476.00	4.44	480.44
221	500.00	238.00	3.74	741.74
222	1928.00	476.00	9.64	2413.64
223	0.00	1062.00	26.51	1088.51
224	120.00	1520.00	12.81	1652.81
225	0.00	1071.00	3.51	1074.51
226	120.00	238.00	2.82	360.82
227	0.00	357.00	5.49	362.49
228	120.00	238.00	1.04	359.04
229	500.00	1071.00	10.08	1581.08
230	500.00	467.00	1.44	968.44
231	500.00	1071.00	12.69	1583.69
232	500.00	705.00	20.98	1225.98
233	500.00	357.00	5.49	862.49
234	0.00	357.00	4.23	361.23
235	120.00	238.00	4.86	362.86
236	500.00	476.00	11.32	987.32
237	0.00	238.00	2.82	240.82
238	500.00	952.00	14.64	1466.64
239	2810.00	2023.00	23.97	4856.97
240	0.00	476.00	5.72	481.72
241	0.00	714.00	7.68	721.68

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
242	0.00	476.00	4.04	480.04
243	500.00	705.00	22.22	1227.22
244	2810.00	1776.00	39.38	4625.38
245	0.00	119.00	1.12	120.12
246	620.00	595.00	9.15	1224.15
247	0.00	119.00	0.10	119.10
248	2048.00	1309.00	27.94	3384.94
249	0.00	824.00	7.35	831.35
250	0.00	1648.00	43.10	1691.10
251	120.00	357.00	3.93	480.93
252	0.00	357.00	2.43	359.43
253	0.00	1419.00	10.10	1429.10
254	0.00	119.00	0.81	119.81
255	0.00	348.00	0.72	348.72
256	0.00	0.00	0.00	0.00
257	0.00	2737.00	32.89	2769.89
258	0.00	1657.00	33.66	1690.66
259	0.00	238.00	0.78	238.78
260	0.00	1071.00	11.52	1082.52
261	0.00	1630.00	71.46	1701.46
262	2310.00	2856.00	67.92	5233.92
263	0.00	595.00	10.30	605.30
264	0.00	348.00	0.91	348.91
265	0.00	586.00	3.03	589.03
266	0.00	476.00	3.24	479.24
267	0.00	3927.00	47.19	3974.19
268	500.00	2582.00	91.62	3173.62
269	0.00	1053.00	40.15	1093.15
270	120.00	1428.00	17.16	1565.16
271	0.00	1291.00	45.32	1336.32
272	0.00	705.00	21.82	726.82
273	0.00	586.00	21.99	607.99
274	0.00	824.00	8.00	832.00
275	2310.00	1309.00	20.13	3639.13
276	2310.00	7086.00	167.64	9563.64
277	0.00	833.00	10.01	843.01
278	500.00	1895.00	38.06	2433.06
279	2810.00	4798.00	190.59	7798.59
280	0.00	119.00	1.21	120.21
281	0.00	1309.00	11.44	1320.44
282	0.00	476.00	7.72	483.72
283	0.00	833.00	11.20	844.20
284	0.00	1172.00	4.86	1176.86
285	0.00	1776.00	37.30	1813.30
286	0.00	2023.00	22.27	2045.27
287	120.00	1428.00	19.20	1567.20
288	0.00	2380.00	16.20	2396.20
289	0.00	4376.00	81.74	4457.74

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
290	0.00	3561.00	33.04	3594.04
291	500.00	238.00	2.42	740.42
292	0.00	1172.00	4.86	1176.86
293	0.00	1062.00	22.17	1084.17
294	0.00	1181.00	6.32	1187.32
295	0.00	2362.00	57.16	2419.16
296	1428.00	357.00	2.76	1787.76
297	120.00	357.00	1.17	478.17
298	120.00	476.00	2.00	598.00
299	0.00	1657.00	4.68	1661.68
300	0.00	714.00	3.00	717.00
301	0.00	238.00	1.62	239.62
302	0.00	1547.00	18.72	1565.72
303	120.00	238.00	3.66	361.66
304	0.00	476.00	1.56	477.56
305	0.00	238.00	3.08	241.08
306	1928.00	1895.00	46.74	3869.74
307	0.00	119.00	1.21	120.21
308	0.00	586.00	4.29	590.29
309	0.00	357.00	3.63	360.63
310	0.00	586.00	5.49	591.49
311	500.00	2124.00	57.64	2681.64
312	0.00	714.00	10.56	724.56
313	0.00	238.00	2.86	240.86
314	0.00	357.00	5.28	362.28
315	0.00	119.00	0.50	119.50
316	0.00	952.00	20.40	972.40
317	1928.00	2142.00	38.88	4108.88
318	0.00	238.00	3.52	241.52
319	0.00	586.00	6.00	592.00
320	0.00	833.00	12.81	845.81
321	0.00	467.00	1.00	468.00
322	0.00	595.00	1.95	596.95
323	0.00	943.00	7.26	950.26
324	0.00	357.00	2.43	359.43
325	0.00	824.00	2.15	826.15
326	0.00	476.00	8.04	484.04
327	0.00	467.00	3.86	470.86
328	0.00	595.00	6.05	601.05
329	0.00	119.00	0.79	119.79
330	0.00	943.00	4.74	947.74
331	0.00	238.00	1.62	239.62
332	0.00	2975.00	12.50	2987.50
333	0.00	714.00	10.44	724.44
334	0.00	0.00	0.00	0.00
335	0.00	119.00	0.40	119.40
336	0.00	467.00	1.00	468.00
337	0.00	238.00	1.00	239.00

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
338	0.00	0.00	0.00	0.00
339	0.00	714.00	4.86	718.86
340	0.00	476.00	2.00	478.00
341	120.00	595.00	7.15	722.15
342	0.00	833.00	17.85	850.85
343	0.00	1062.00	28.82	1090.82
344	0.00	952.00	4.16	956.16
345	0.00	1785.00	18.15	1803.15
346	0.00	238.00	3.52	241.52
347	0.00	595.00	4.05	599.05
348	500.00	1785.00	26.40	2311.40
349	0.00	1062.00	31.62	1093.62
350	120.00	0.00	0.00	120.00
351	0.00	357.00	4.23	361.23
352	120.00	1190.00	7.20	1317.20
353	0.00	238.00	2.86	240.86
354	0.00	2142.00	24.12	2166.12
355	0.00	833.00	7.84	840.84
356	0.00	595.00	9.15	604.15
357	0.00	705.00	4.04	709.04
358	0.00	934.00	4.04	938.04
359	0.00	0.00	0.00	0.00
360	120.00	238.00	3.08	361.08
361	1428.00	0.00	17.10	1445.10
362	120.00	0.00	4.56	124.56
363	1428.00	0.00	8.69	1436.69
364	0.00	0.00	9.68	9.68
365	0.00	0.00	7.98	7.98
366	0.00	0.00	1.17	1.17
367	0.00	0.00	0.00	0.00
368	0.00	0.00	13.77	13.77
369	0.00	0.00	7.32	7.32
370	1428.00	0.00	2.37	1430.37
371	0.00	0.00	7.26	7.26
372	0.00	0.00	2.37	2.37
373	1428.00	0.00	12.10	1440.10
374	0.00	0.00	8.47	8.47
375	0.00	0.00	11.76	11.76
376	0.00	0.00	0.00	0.00
377	0.00	0.00	5.04	5.04
378	0.00	0.00	1.95	1.95
379	0.00	0.00	3.24	3.24
380	120.00	0.00	8.47	128.47
381	0.00	1062.00	5.53	1067.53
382	0.00	238.00	2.36	240.36
383	120.00	357.00	2.37	479.37
384	0.00	595.00	5.70	600.70
385	0.00	357.00	4.32	361.32

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
386	0.00	357.00	3.54	360.54
387	0.00	833.00	8.47	841.47
388	120.00	595.00	6.05	721.05
389	0.00	238.00	1.62	239.62
390	0.00	119.00	0.79	119.79
391	0.00	238.00	0.78	238.78
392	1428.00	952.00	9.12	2389.12
393	0.00	357.00	2.43	359.43
394	0.00	824.00	3.95	827.95
395	0.00	0.00	0.00	0.00
396	0.00	595.00	3.55	598.55
397	0.00	238.00	2.22	240.22
398	0.00	1172.00	40.26	1212.26
399	0.00	238.00	1.44	239.44
400	0.00	238.00	1.58	239.58
401	0.00	119.00	0.50	119.50
402	0.00	238.00	1.58	239.58
403	0.00	705.00	19.66	724.66
404	0.00	1181.00	3.12	1184.12
405	120.00	357.00	1.50	478.50
406	620.00	1776.00	32.23	2428.23
407	2810.00	5712.00	73.92	8595.92
408	0.00	238.00	0.78	238.78
409	0.00	119.00	0.81	119.81
410	0.00	1785.00	1.50	1786.50
411	0.00	714.00	2.34	716.34
412	500.00	476.00	2.00	978.00
413	120.00	476.00	4.48	600.48
414	0.00	0.00	0.00	0.00
415	120.00	833.00	12.67	965.67
416	0.00	586.00	21.30	607.30
417	120.00	1181.00	4.00	1305.00
418	120.00	943.00	0.60	1063.60
419	0.00	357.00	5.10	362.10
420	0.00	357.00	1.17	358.17
421	0.00	1181.00	11.28	1192.28
422	0.00	833.00	9.87	842.87
423	0.00	1172.00	8.04	1180.04
424	0.00	357.00	4.23	361.23
425	0.00	357.00	4.23	361.23
426	0.00	1300.00	29.19	1329.19
427	500.00	357.00	6.90	863.90
428	0.00	1776.00	41.59	1817.59
429	0.00	4084.00	23.97	4107.97
430	0.00	1538.00	23.76	1561.76
431	0.00	0.00	0.00	0.00
432	0.00	943.00	4.86	947.86
433	0.00	595.00	7.15	602.15

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
434	0.00	1520.00	12.81	1532.81
435	0.00	714.00	4.32	718.32
436	0.00	824.00	7.05	831.05
437	0.00	595.00	7.05	602.05
438	2810.00	2023.00	19.04	4852.04
439	0.00	696.00	3.22	699.22
440	0.00	943.00	12.96	955.96
441	120.00	1071.00	17.10	1208.10
442	0.00	238.00	87.24	325.24
443	0.00	4495.00	133.34	4628.34
444	0.00	476.00	4.56	480.56
445	120.00	476.00	3.16	599.16
446	0.00	476.00	7.04	483.04
447	0.00	238.00	3.08	241.08
448	0.00	476.00	7.32	483.32
449	0.00	238.00	1.62	239.62
450	0.00	824.00	5.60	829.60
451	0.00	952.00	8.88	960.88
452	0.00	238.00	3.08	241.08
453	0.00	815.00	38.19	853.19
454	0.00	833.00	10.57	843.57
455	120.00	119.00	1.83	240.83
456	120.00	833.00	12.81	965.81
457	120.00	467.00	1.64	588.64
458	0.00	586.00	21.72	607.72
459	120.00	238.00	0.20	358.20
460	0.00	705.00	23.82	728.82
461	0.00	595.00	0.50	595.50
462	0.00	952.00	2.32	954.32
463	0.00	595.00	2.50	597.50
464	0.00	1053.00	7.70	1060.70
465	0.00	357.00	2.13	359.13
466	0.00	714.00	3.00	717.00
467	0.00	348.00	0.39	348.39
468	500.00	0.00	0.00	500.00
469	0.00	1181.00	3.12	1184.12
470	0.00	476.00	1.56	477.56
471	0.00	348.00	0.00	348.00
472	0.00	1538.00	4.29	1542.29
473	0.00	476.00	4.88	480.88
474	0.00	2335.00	7.90	2342.90
475	0.00	705.00	4.48	709.48
476	0.00	595.00	5.70	600.70
477	0.00	238.00	0.78	238.78
478	0.00	1547.00	20.02	1567.02
479	500.00	3451.00	73.66	4024.66
480	0.00	357.00	2.37	359.37
481	0.00	705.00	7.72	712.72

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
482	120.00	705.00	3.24	828.24
483	500.00	952.00	6.48	1458.48
484	0.00	119.00	0.00	119.00
485	120.00	1053.00	36.95	1209.95
486	120.00	1300.00	10.89	1430.89
487	0.00	119.00	0.39	119.39
488	500.00	1300.00	23.79	1823.79
489	0.00	348.00	0.81	348.81
490	0.00	476.00	4.84	480.84
491	0.00	934.00	37.84	971.84
492	120.00	595.00	3.95	718.95
493	0.00	714.00	4.74	718.74
494	0.00	238.00	1.58	239.58
495	0.00	696.00	2.62	698.62
497	500.00	2014.00	46.95	2560.95
498	120.00	476.00	3.24	599.24
499	120.00	833.00	8.47	961.47
500	120.00	238.00	1.62	359.62
501	0.00	586.00	1.17	587.17
502	1428.00	1181.00	4.16	2613.16
503	0.00	238.00	1.58	239.58
504	0.00	943.00	4.74	947.74
505	0.00	833.00	0.00	833.00
506	0.00	238.00	1.58	239.58
507	0.00	1053.00	6.05	1059.05
508	0.00	348.00	0.39	348.39
509	0.00	586.00	2.37	588.37
510	0.00	586.00	2.43	588.43
511	0.00	238.00	2.42	240.42
512	0.00	238.00	1.58	239.58
513	0.00	833.00	4.34	837.34
514	0.00	952.00	12.32	964.32
515	0.00	238.00	1.04	239.04
516	0.00	595.00	2.00	597.00
517	0.00	476.00	3.24	479.24
518	0.00	357.00	1.50	358.50
519	0.00	1428.00	6.24	1434.24
520	0.00	1062.00	2.73	1064.73
521	0.00	357.00	2.43	359.43
522	500.00	1071.00	10.26	1581.26
523	0.00	476.00	3.40	479.40
524	0.00	357.00	3.75	360.75
525	0.00	833.00	5.67	838.67
526	0.00	119.00	1.21	120.21
527	0.00	1547.00	20.80	1567.80
528	0.00	357.00	2.07	359.07
529	0.00	1776.00	9.36	1785.36
530	0.00	1309.00	10.12	1319.12

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
531	0.00	0.00	0.00	0.00
532	620.00	238.00	0.78	858.78
533	0.00	0.00	0.00	0.00
534	500.00	714.00	7.68	1221.68
535	500.00	3718.00	9.48	4227.48
536	0.00	238.00	0.20	238.20
537	120.00	238.00	1.62	359.62
538	0.00	1401.00	60.48	1461.48
539	120.00	3828.00	189.42	4137.42
540	120.00	357.00	1.50	478.50
551	0.00	705.00	3.16	708.16
552	0.00	833.00	10.08	843.08
553	0.00	476.00	3.24	479.24
554	500.00	0.00	3.42	503.42
555	0.00	943.00	9.24	952.24
556	0.00	348.00	1.41	349.41
557	1428.00	119.00	0.10	1547.10
558	0.00	4870.00	36.00	4906.00
559	0.00	467.00	4.00	471.00
560	500.00	0.00	0.00	500.00
561	0.00	238.00	2.42	240.42
562	0.00	833.00	7.84	840.84
563	0.00	467.00	1.00	468.00
564	0.00	238.00	1.58	239.58
565	500.00	1776.00	36.52	2312.52
566	1548.00	833.00	3.50	2384.50
567	120.00	119.00	0.79	239.79
568	0.00	238.00	1.00	239.00
569	0.00	943.00	21.24	964.24
570	0.00	824.00	19.00	843.00
571	0.00	119.00	1.21	120.21
572	0.00	0.00	6.16	6.16
573	0.00	595.00	6.25	601.25
574	0.00	357.00	0.30	357.30
575	1428.00	238.00	0.20	1666.20
576	0.00	238.00	2.30	240.30
577	0.00	1547.00	6.50	1553.50
578	0.00	595.00	93.80	688.80
579	1428.00	476.00	9.32	1913.32
580	0.00	348.00	1.21	349.21
581	0.00	1053.00	3.95	1056.95
582	1548.00	1053.00	6.05	2607.05
583	0.00	2124.00	49.94	2173.94
584	0.00	119.00	0.43	119.43
585	0.00	357.00	0.87	357.87
586	1428.00	0.00	0.00	1428.00
587	0.00	1428.00	20.88	1448.88
588	1428.00	1071.00	10.89	2509.89

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
589	1548.00	952.00	9.68	2509.68
590	120.00	2023.00	15.64	2158.64
591	0.00	3076.00	8.58	3084.58
592	0.00	467.00	1.62	468.62
593	0.00	943.00	4.74	947.74
594	0.00	357.00	4.62	361.62
595	0.00	696.00	2.28	698.28
596	0.00	1181.00	22.98	1203.98
597	0.00	595.00	2.50	597.50
598	0.00	348.00	1.14	349.14
599	0.00	238.00	1.62	239.62
600	0.00	833.00	5.67	838.67
601	1428.00	476.00	3.24	1907.24
602	0.00	934.00	3.16	937.16
603	0.00	696.00	2.24	698.24
604	0.00	1309.00	21.23	1330.23
605	0.00	934.00	39.16	973.16
606	0.00	348.00	1.21	349.21
607	0.00	1282.00	2.50	1284.50
608	0.00	595.00	5.70	600.70
609	0.00	348.00	0.92	348.92
610	0.00	595.00	4.05	599.05
611	0.00	0.00	12.32	12.32
612	0.00	0.00	0.39	0.39
613	0.00	0.00	1.95	1.95
614	0.00	0.00	4.05	4.05
615	0.00	0.00	2.43	2.43
616	0.00	0.00	1.84	1.84
617	0.00	0.00	2.00	2.00
618	0.00	0.00	5.53	5.53
619	0.00	0.00	0.50	0.50
620	0.00	0.00	4.86	4.86
621	0.00	0.00	2.37	2.37
622	0.00	0.00	5.37	5.37
623	0.00	0.00	7.70	7.70
624	0.00	0.00	3.66	3.66
625	0.00	0.00	6.00	6.00
626	0.00	0.00	0.58	0.58
627	120.00	0.00	0.00	120.00
628	0.00	0.00	4.00	4.00
629	0.00	0.00	0.79	0.79
630	0.00	0.00	0.58	0.58
631	1548.00	0.00	3.16	1551.16
652	0.00	0.00	0.00	0.00
653	500.00	1071.00	12.69	1583.69
654	0.00	595.00	7.05	602.05
655	0.00	2023.00	6.63	2029.63
656	0.00	1657.00	14.52	1671.52

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
657	1428.00	1401.00	2.34	2831.34
658	0.00	4394.00	10.15	4404.15
659	0.00	1062.00	3.50	1065.50
660	2810.00	2133.00	29.14	4972.14
661	500.00	1785.00	26.40	2311.40
662	500.00	1181.00	32.74	1713.74
663	0.00	238.00	0.78	238.78
664	0.00	357.00	1.17	358.17
665	500.00	357.00	3.63	860.63
666	500.00	714.00	12.00	1226.00
667	0.00	119.00	0.50	119.50
668	0.00	2005.00	9.23	2014.23
669	1928.00	4870.00	60.96	6858.96
670	0.00	1062.00	28.40	1090.40
671	2810.00	2142.00	14.22	4966.22
675	0.00	357.00	2.37	359.37
676	0.00	714.00	9.24	723.24
677	0.00	1172.00	43.56	1215.56
678	1428.00	4165.00	152.90	5745.90
692	0.00	357.00	2.07	359.07
693	120.00	1300.00	30.90	1450.90
694	0.00	1648.00	44.40	1692.40
695	1928.00	2618.00	29.26	4575.26
696	0.00	586.00	0.30	586.30
697	0.00	2261.00	57.95	2318.95
698	120.00	1071.00	13.86	1204.86
699	0.00	943.00	25.74	968.74
700	0.00	595.00	5.05	600.05
701	0.00	1062.00	20.00	1082.00
702	0.00	238.00	0.80	238.80
703	0.00	943.00	4.86	947.86
704	500.00	595.00	7.05	1102.05
705	0.00	476.00	2.00	478.00
706	0.00	1181.00	22.82	1203.82
707	0.00	238.00	1.00	239.00
708	0.00	595.00	5.70	600.70
709	0.00	1071.00	14.22	1085.22
710	0.00	714.00	7.26	721.26
711	120.00	952.00	11.28	1083.28
712	0.00	357.00	4.23	361.23
713	0.00	824.00	8.80	832.80
714	120.00	714.00	4.74	838.74
715	0.00	1886.00	106.96	1992.96
716	0.00	943.00	10.98	953.98
717	0.00	238.00	1.00	239.00
718	0.00	1410.00	17.28	1427.28
719	0.00	238.00	1.00	239.00
720	0.00	476.00	8.88	484.88

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
721	0.00	357.00	5.22	362.22
722	0.00	238.00	2.82	240.82
723	0.00	833.00	9.87	842.87
724	0.00	705.00	3.24	708.24
725	0.00	714.00	12.96	726.96
726	0.00	595.00	9.15	604.15
727	0.00	1776.00	13.13	1789.13
728	0.00	476.00	3.16	479.16
729	0.00	952.00	6.48	958.48
730	0.00	476.00	4.84	480.84
732	0.00	1547.00	25.09	1572.09
733	120.00	238.00	1.58	359.58
734	0.00	0.00	0.00	0.00
735	0.00	119.00	0.79	119.79
736	0.00	1071.00	10.89	1081.89
737	0.00	705.00	3.16	708.16
738	0.00	348.00	0.71	348.71
739	0.00	467.00	1.62	468.62
740	0.00	238.00	1.58	239.58
741	0.00	119.00	0.79	119.79
742	0.00	833.00	5.67	838.67
743	0.00	119.00	0.81	119.81
744	120.00	119.00	0.79	239.79
745	0.00	595.00	3.95	598.95
746	0.00	595.00	7.05	602.05
747	0.00	348.00	1.21	349.21
748	0.00	348.00	0.79	348.79
749	0.00	1062.00	8.47	1070.47
750	0.00	348.00	1.21	349.21
751	0.00	119.00	2.00	121.00
752	0.00	357.00	3.42	360.42
753	0.00	952.00	6.48	958.48
754	0.00	476.00	7.64	483.64
755	0.00	476.00	4.84	480.84
756	120.00	238.00	2.22	360.22
757	120.00	1190.00	17.00	1327.00
758	120.00	238.00	2.24	360.24
759	0.00	0.00	0.00	0.00
760	0.00	0.00	0.00	0.00
761	0.00	1172.00	7.08	1179.08
762	0.00	696.00	2.42	698.42
763	120.00	1190.00	12.10	1322.10
764	500.00	1895.00	38.06	2433.06
765	0.00	119.00	0.39	119.39
766	0.00	119.00	0.81	119.81
767	120.00	238.00	1.58	359.58
768	0.00	2499.00	17.01	2516.01
769	0.00	595.00	4.05	599.05

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
770	0.00	833.00	5.04	838.04
771	0.00	0.00	12.32	12.32
772	0.00	0.00	11.22	11.22
773	0.00	0.00	2.68	2.68
774	0.00	0.00	5.10	5.10
775	0.00	0.00	5.60	5.60
776	0.00	238.00	2.82	240.82
777	120.00	0.00	1.00	121.00
778	0.00	0.00	8.82	8.82
779	0.00	0.00	2.16	2.16
780	0.00	0.00	5.80	5.80
781	0.00	0.00	5.22	5.22
782	120.00	0.00	2.82	122.82
784	0.00	0.00	0.00	0.00
785	0.00	0.00	30.24	30.24
786	1428.00	0.00	0.00	1428.00
787	0.00	0.00	4.56	4.56
788	0.00	0.00	8.80	8.80
789	1428.00	0.00	7.32	1435.32
790	1428.00	0.00	10.80	1438.80
791	0.00	238.00	1.62	239.62
792	120.00	238.00	0.58	358.58
793	0.00	595.00	8.50	603.50
794	0.00	467.00	1.44	468.44
795	0.00	943.00	6.72	949.72
796	0.00	824.00	2.60	826.60
797	1428.00	467.00	2.42	1897.42
798	0.00	119.00	0.10	119.10
799	0.00	1895.00	11.06	1906.06
800	0.00	467.00	1.00	468.00
801	0.00	943.00	7.86	950.86
802	0.00	595.00	7.90	602.90
803	0.00	476.00	4.04	480.04
804	0.00	1172.00	41.58	1213.58
805	0.00	2838.00	53.20	2891.20
806	120.00	1428.00	20.88	1568.88
807	0.00	357.00	3.03	360.03
808	0.00	0.00	0.00	0.00
809	0.00	476.00	4.04	480.04
810	0.00	952.00	8.96	960.96
811	0.00	467.00	3.66	470.66
812	0.00	2142.00	32.94	2174.94
813	0.00	595.00	7.15	602.15
814	120.00	943.00	24.96	1087.96
815	0.00	1987.00	81.66	2068.66
816	0.00	2380.00	63.20	2443.20
817	0.00	238.00	0.20	238.20
818	0.00	119.00	0.72	119.72

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
820	0.00	3204.00	35.75	3239.75
821	0.00	238.00	2.02	240.02
822	0.00	357.00	3.03	360.03
823	0.00	833.00	12.81	845.81
824	0.00	1529.00	43.08	1572.08
825	0.00	238.00	1.42	239.42
826	1428.00	357.00	4.23	1789.23
827	0.00	824.00	25.65	849.65
828	0.00	833.00	7.07	840.07
829	0.00	0.00	0.00	0.00
830	0.00	833.00	7.84	840.84
831	0.00	815.00	5.79	820.79
832	0.00	357.00	3.63	360.63
833	500.00	238.00	2.42	740.42
834	0.00	357.00	3.63	360.63
835	0.00	357.00	1.50	358.50
836	120.00	238.00	0.20	358.20
837	0.00	357.00	3.42	360.42
838	0.00	1547.00	25.09	1572.09
839	0.00	824.00	6.05	830.05
840	0.00	119.00	1.14	120.14
841	0.00	833.00	97.81	930.81
842 .		815.00	1.17	#VALUE!
843	0.00	934.00	2.00	936.00
844	0.00	238.00	2.42	240.42
845	0.00	1062.00	8.47	1070.47
846	0.00	0.00	0.00	0.00
847	0.00	2719.00	15.39	2734.39
848	0.00	586.00	2.43	588.43
849	0.00	119.00	0.50	119.50
850	0.00	595.00	6.40	601.40
851	0.00	595.00	2.60	597.60
852	0.00	0.00	0.00	0.00
853	0.00	348.00	0.79	348.79
854	0.00	348.00	1.14	349.14
855	0.00	119.00	0.79	119.79
856	0.00	348.00	0.39	348.39
857	0.00	0.00	0.00	0.00
858	0.00	0.00	0.00	0.00
859	0.00	348.00	1.21	349.21
860	500.00	1785.00	1.50	2286.50
861	0.00	119.00	0.10	119.10
862	0.00	1181.00	3.12	1184.12
863	0.00	238.00	0.20	238.20
864	0.00	833.00	10.78	843.78
865	0.00	2609.00	32.00	2641.00
866	0.00	0.00	0.00	0.00
867	500.00	3727.00	170.06	4397.06

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
868	0.00	119.00	0.39	119.39
869	0.00	1776.00	1.30	1777.30
870	0.00	348.00	0.79	348.79
871	0.00	357.00	1.17	358.17
872	0.00	0.00	0.00	0.00
873	0.00	1666.00	32.48	1698.48
874	500.00	943.00	33.48	1476.48
875	0.00	238.00	0.20	238.20
876	120.00	714.00	8.46	842.46
877	2810.00	1419.00	34.80	4263.80
878	0.00	952.00	8.08	960.08
879	0.00	1062.00	7.07	1069.07
880	120.00	357.00	3.03	480.03
881	620.00	1071.00	13.86	1704.86
882	500.00	586.00	4.23	1090.23
883	500.00	714.00	10.98	1224.98
884	2810.00	2600.00	25.38	5435.38
885	0.00	1071.00	9.09	1080.09
886	0.00	1062.00	3.50	1065.50
887	0.00	833.00	12.81	845.81
888	2810.00	2252.00	32.30	5094.30
889	120.00	943.00	4.32	1067.32
890	500.00	1300.00	9.09	1809.09
891	500.00	1053.00	45.05	1598.05
892	0.00	0.00	6.72	6.72
893	0.00	0.00	0.00	0.00
894	0.00	1785.00	15.60	1800.60
895	620.00	595.00	8.65	1223.65
896	0.00	0.00	0.00	0.00
897	0.00	0.00	15.84	15.84
898	0.00	1071.00	0.90	1071.90
899	0.00	0.00	11.90	11.90
900	0.00	0.00	3.66	3.66
901	0.00	0.00	8.58	8.58
902	0.00	0.00	4.05	4.05
903	0.00	0.00	1.01	1.01
904	0.00	2142.00	27.72	2169.72
905	0.00	0.00	0.00	0.00
906	0.00	0.00	0.00	0.00
907	0.00	0.00	0.00	0.00
908	0.00	0.00	18.48	18.48
909	120.00	0.00	4.29	124.29
910	120.00	0.00	0.00	120.00
911	0.00	0.00	0.00	0.00
912	120.00	2499.00	38.43	2657.43
913	0.00	824.00	25.65	849.65
914	1428.00	595.00	7.35	2030.35
915	0.00	1291.00	45.32	1336.32

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
916	120.00	595.00	3.60	718.60
917	0.00	2380.00	36.80	2416.80
918	0.00	476.00	6.96	482.96
919	0.00	476.00	6.16	482.16
920	0.00	2499.00	45.36	2544.36
921	0.00	476.00	1.56	477.56
922	0.00	833.00	10.29	843.29
923	0.00	1428.00	18.48	1446.48
924	0.00	1181.00	3.12	1184.12
953	120.00	0.00	11.28	131.28
954	0.00	0.00	1.04	1.04
955	0.00	0.00	0.00	0.00
956	120.00	952.00	11.28	1083.28
957	0.00	0.00	4.29	4.29
958	0.00	0.00	11.00	11.00
959	0.00	0.00	0.00	0.00
960	0.00	0.00	3.08	3.08
961	0.00	0.00	0.00	0.00
962	0.00	0.00	0.00	0.00
963	0.00	0.00	5.28	5.28
964	120.00	0.00	0.00	120.00
965	0.00	0.00	0.00	0.00
966	1548.00	0.00	6.72	1554.72
967	0.00	0.00	9.50	9.50
968	0.00	0.00	7.15	7.15
969	0.00	0.00	0.00	0.00
970	0.00	0.00	0.00	0.00
971	0.00	0.00	0.00	0.00
972	120.00	0.00	0.00	120.00
973	0.00	2811.00	11.34	2822.34
974	120.00	1181.00	30.10	1331.10
975	500.00	2014.00	12.15	2526.15
976	0.00	467.00	0.58	467.58
977	120.00	815.00	1.56	936.56
978	0.00	586.00	2.37	588.37
979	0.00	0.00	0.00	0.00
980	620.00	943.00	23.76	1586.76
981	0.00	1172.00	41.58	1213.58
982	0.00	1767.00	8.69	1775.69
983	120.00	1300.00	10.26	1430.26
984	120.00	1172.00	39.84	1331.84
985	0.00	705.00	19.34	724.34
986	0.00	586.00	19.86	605.86
987	2930.00	5712.00	38.88	8680.88
988	500.00	1181.00	25.62	1706.62
989	0.00	476.00	3.24	479.24
990	0.00	1767.00	7.92	1774.92
991	0.00	2582.00	76.08	2658.08

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
992	0.00	586.00	3.63	589.63
1002	120.00	833.00	10.01	963.01
1007	500.00	1419.00	27.90	1946.90
1009	0.00	238.00	2.28	240.28
1010	0.00	357.00	1.17	358.17
1011	0.00	238.00	2.08	240.08
1012	120.00	952.00	6.48	1078.48
1013	120.00	1181.00	6.48	1307.48
1014	620.00	6080.00	343.48	7043.48
1015	0.00	2133.00	45.78	2178.78
1016	120.00	943.00	21.36	1084.36
1017	0.00	476.00	4.84	480.84
1018	0.00	586.00	3.75	589.75
1019	0.00	934.00	4.84	938.84
1020	0.00	357.00	4.62	361.62
1021	620.00	3085.00	77.46	3782.46
1022	0.00	1062.00	10.78	1072.78
1023	500.00	2719.00	36.67	3255.67
1024	0.00	1190.00	11.40	1201.40
1025	500.00	4147.00	70.51	4717.51
1026	0.00	476.00	3.24	479.24
1027	1428.00	1767.00	12.54	3207.54
1028	0.00	1190.00	15.40	1205.40
1029	0.00	348.00	0.79	348.79
1030	0.00	943.00	104.50	1047.50
1031	0.00	467.00	0.58	467.58
1032	0.00	815.00	2.43	817.43
1033	120.00	1163.00	52.66	1335.66
1034	120.00	467.00	2.42	589.42
1035	0.00	357.00	2.43	359.43
1036	0.00	714.00	10.26	724.26
1037	0.00	357.00	2.43	359.43
1038	0.00	586.00	2.37	588.37
1039	0.00	952.00	3.12	955.12
1040	0.00	586.00	1.17	587.17
1041	0.00	1053.00	1.45	1054.45
1042	0.00	1044.00	2.13	1046.13
1043	0.00	934.00	36.24	970.24
1044	0.00	348.00	0.40	348.40
1045	0.00	705.00	3.24	708.24
1046	0.00	952.00	6.32	958.32
1047	0.00	1071.00	18.27	1089.27
1048	0.00	238.00	0.78	238.78
1049	0.00	238.00	2.42	240.42
1050	0.00	238.00	1.00	239.00
1051	0.00	3671.00	54.87	3725.87
1052	0.00	833.00	5.67	838.67
1053	0.00	1529.00	16.20	1545.20

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
1054	0.00	586.00	0.30	586.30
1055	0.00	238.00	4.00	242.00
1056	0.00	348.00	0.81	348.81
1057	0.00	1904.00	20.48	1924.48
1058	0.00	3094.00	10.14	3104.14
1059	0.00	119.00	1.21	120.21
1060	0.00	238.00	1.62	239.62
1061	0.00	952.00	10.48	962.48
1062	0.00	348.00	0.50	348.50
1063	0.00	119.00	0.79	119.79
1064	0.00	476.00	4.84	480.84
1065	0.00	476.00	87.88	563.88
1066	0.00	119.00	0.79	119.79
1067	120.00	476.00	2.08	598.08
1068	0.00	467.00	2.62	469.62
1069	0.00	0.00	0.50	0.50
1070	120.00	0.00	8.28	128.28
1071	0.00	0.00	0.00	0.00
1072	0.00	0.00	14.21	14.21
1073	0.00	0.00	0.81	0.81
1074	0.00	0.00	0.50	0.50
1075	0.00	0.00	0.00	0.00
1076	0.00	0.00	0.00	0.00
1077	0.00	0.00	0.20	0.20
1078	0.00	0.00	0.00	0.00
1079	0.00	0.00	0.00	0.00
1080	0.00	0.00	2.43	2.43
1081	0.00	0.00	0.00	0.00
1082	120.00	0.00	3.00	123.00
1083	0.00	0.00	1.16	1.16
1084	0.00	0.00	2.34	2.34
1085	0.00	0.00	1.54	1.54
1086	0.00	0.00	1.56	1.56
1087	0.00	0.00	0.00	0.00
1088	0.00	0.00	0.00	0.00
1089	0.00	357.00	6.99	363.99
1090	0.00	595.00	3.95	598.95
1091	500.00	952.00	14.64	1466.64
1092	0.00	119.00	0.39	119.39
1093	0.00	238.00	2.42	240.42
1094	500.00	357.00	3.63	860.63
1095	0.00	595.00	4.05	599.05
1096	1428.00	238.00	1.44	1667.44
1097	0.00	119.00	0.39	119.39
1098	0.00	714.00	7.08	721.08
1099	0.00	119.00	0.39	119.39
1100	0.00	2802.00	122.16	2924.16
1401	0.00	1053.00	1.45	1054.45

Patient study no.	Procedures cost	Hospital stay cost	Drug cost	Admission cost
1402	0.00	348.00	0.71	348.71
1403	0.00	1172.00	7.26	1179.26
1404	0.00	238.00	1.62	239.62
1405	0.00	1410.00	36.12	1446.12
1406	0.00	1181.00	9.68	1190.68
1407	0.00	1071.00	7.11	1078.11
1408	0.00	824.00	21.55	845.55
1409	0.00	238.00	1.62	239.62
1421	0.00	357.00	3.63	360.63
1422	0.00	714.00	2.34	716.34
1423	0.00	705.00	4.84	709.84
1424	0.00	4414.00	11.06	4425.06
1425	0.00	0.00	1.12	1.12
1426	120.00	705.00	20.18	845.18
1427	0.00	1181.00	3.12	1184.12
1428	0.00	357.00	2.37	359.37
1429	500.00	1300.00	7.11	1807.11
1430	0.00	1172.00	40.74	1212.74
1431	0.00	586.00	3.63	589.63
1432	0.00	476.00	6.16	482.16
1441	0.00	1190.00	8.10	1198.10
1442	0.00	815.00	3.63	818.63
1443	500.00	714.00	3.00	1217.00
1444	500.00	3397.00	130.11	4027.11
1445	2930.00	0.00	0.00	2930.00
1446	120.00	476.00	3.24	599.24
1447	120.00	1181.00	23.86	1324.86
1448	0.00	357.00	3.63	360.63
1449	500.00	476.00	4.04	980.04
1450	0.00	238.00	1.58	239.58
1451	0.00	1309.00	17.60	1326.60
1452	500.00	1419.00	30.60	1949.60
1453	120.00	1053.00	2.50	1175.50
1454	0.00	1053.00	6.05	1059.05
1455	620.00	952.00	14.08	1586.08
1456	0.00	1062.00	0.70	1062.70
1457	120.00	119.00	0.50	239.50
1458	0.00	476.00	3.68	479.68
1459	0.00	934.00	36.16	970.16
1460	0.00	357.00	4.74	361.74

Appendix 2

Results of simulations of ECMO data set

Tables 1 – 10 Simulation statistics

Figures 1 - 10 Histograms of simulation results.

Data tables of the ECMO study

Subgroup 1	**Birth weight < 3300 grams**
Subgroup 2	**Birth weight > 3300 grams**
Subgroup 3	**Gestation < 275 days**
Subgroup 4	**Gestation > 275 days**
Subgroup 5	**Diagnosis CDH**
Subgroup 6	**Diagnosis IPFC**
Subgroup 7	**Diagnosis PPH**
Subgroup 8	**OI category I (<60)**
Subgroup 9	**OI category II (>60)**

Table 1: Descriptive statistics of probability distribution – subgroup1

Name	Ee - Ec	R	effect E	effect C	Ce - Cc
Minimum =	- 0.03	- 5,772,243	0.28	0.02	7,875
Maximum =	0.68	3,197,265	0.88	0.46	29,973
Mean =	0.30	75,847	0.54	0.24	19,711
Std Deviation =	0.09	99,133	0.07	0.06	3,003
Variance =	0.01	9,827,286,000	0.00	0.00	9,017,402
Skewness =	0.01	- 10	0.01	- 0.01	- 0
Kurtosis =	3.06	1,517	3.04	2.97	3
Mode =	0.19	54,404	0.46	0.17	18,020
5% Perc =	0.15	39,639	0.42	0.14	14,731
10% Perc =	0.18	44,183	0.45	0.16	15,871
15% Perc =	0.20	47,349	0.47	0.18	16,575
20% Perc =	0.22	50,424	0.48	0.19	17,153
25% Perc =	0.24	52,877	0.49	0.20	17,700
30% Perc =	0.25	55,517	0.50	0.21	18,151
35% Perc =	0.26	58,028	0.51	0.22	18,562
40% Perc =	0.28	60,416	0.52	0.23	18,941
45% Perc =	0.29	62,982	0.53	0.23	19,325
50% Perc =	0.30	65,716	0.54	0.24	19,735
55% Perc =	0.31	68,795	0.55	0.25	20,104
60% Perc =	0.32	71,967	0.56	0.25	20,478
65% Perc =	0.34	75,418	0.57	0.26	20,879
70% Perc =	0.35	79,531	0.58	0.27	21,286
75% Perc =	0.36	84,395	0.59	0.28	21,730
80% Perc =	0.38	90,397	0.60	0.29	22,237
85% Perc =	0.39	98,819	0.61	0.30	22,815
90% Perc =	0.42	111,543	0.63	0.32	23,569
95% Perc =	0.45	137,317	0.65	0.34	24,637

Figure 1: Probability distribution of $R = (C_E - C_C) / (E_E - E_C)$ – subgroup 1

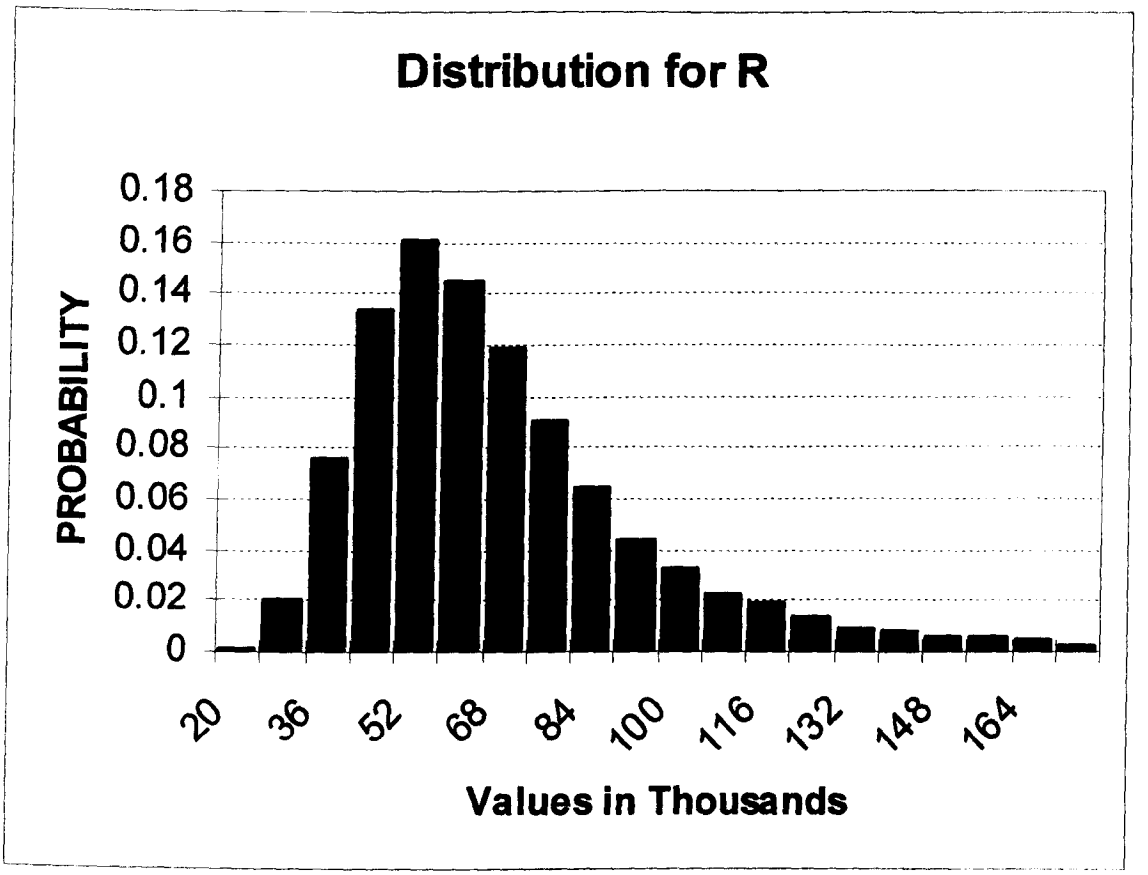


Table 2: Descriptive statistics of probability distribution – subgroup 2

Name	Ee - Ec	R	Ce - Cc	effect E	effect C
Minimum = -	0.11	- 4,211,902	- 4,369	0.55	0.29
Maximum =	0.60	50,114,870	18,954	1.04	0.80
Mean =	0.26	38,628	7,430	0.79	0.53
Std Deviati	0.09	506,517	2,847	0.06	0.07
Variance =	0.01	256,559,800,000	8,107,124	0.00	0.00
Skewness	0.01	97	0	- 0.02	0.01
Kurtosis =	2.97	9,554	3	3.05	3.03
Mode =	0.25	24,940	6,506	0.71	0.53
5% Perc =	0.11	9,498	2,736	0.69	0.42
10% Perc =	0.14	13,273	3,754	0.71	0.44
15% Perc =	0.16	15,808	4,484	0.73	0.46
20% Perc =	0.18	17,821	5,028	0.74	0.47
25% Perc =	0.20	19,716	5,524	0.75	0.48
30% Perc =	0.21	21,457	5,953	0.76	0.49
35% Perc =	0.22	23,165	6,334	0.77	0.50
40% Perc =	0.24	24,984	6,735	0.78	0.51
45% Perc =	0.25	26,714	7,090	0.78	0.52
50% Perc =	0.26	28,488	7,432	0.79	0.53
55% Perc =	0.27	30,451	7,782	0.80	0.54
60% Perc =	0.28	32,448	8,166	0.81	0.55
65% Perc =	0.29	34,897	8,553	0.81	0.56
70% Perc =	0.31	37,493	8,923	0.82	0.57
75% Perc =	0.32	40,582	9,332	0.83	0.58
80% Perc =	0.34	44,434	9,801	0.84	0.59
85% Perc =	0.35	49,899	10,350	0.85	0.60
90% Perc =	0.38	59,009	11,053	0.87	0.62
95% Perc =	0.41	75,952	12,100	0.89	0.64

Figure 2: Probability distribution of $R = (C_E - C_C) / (E_E - E_C)$ – subgroup 2

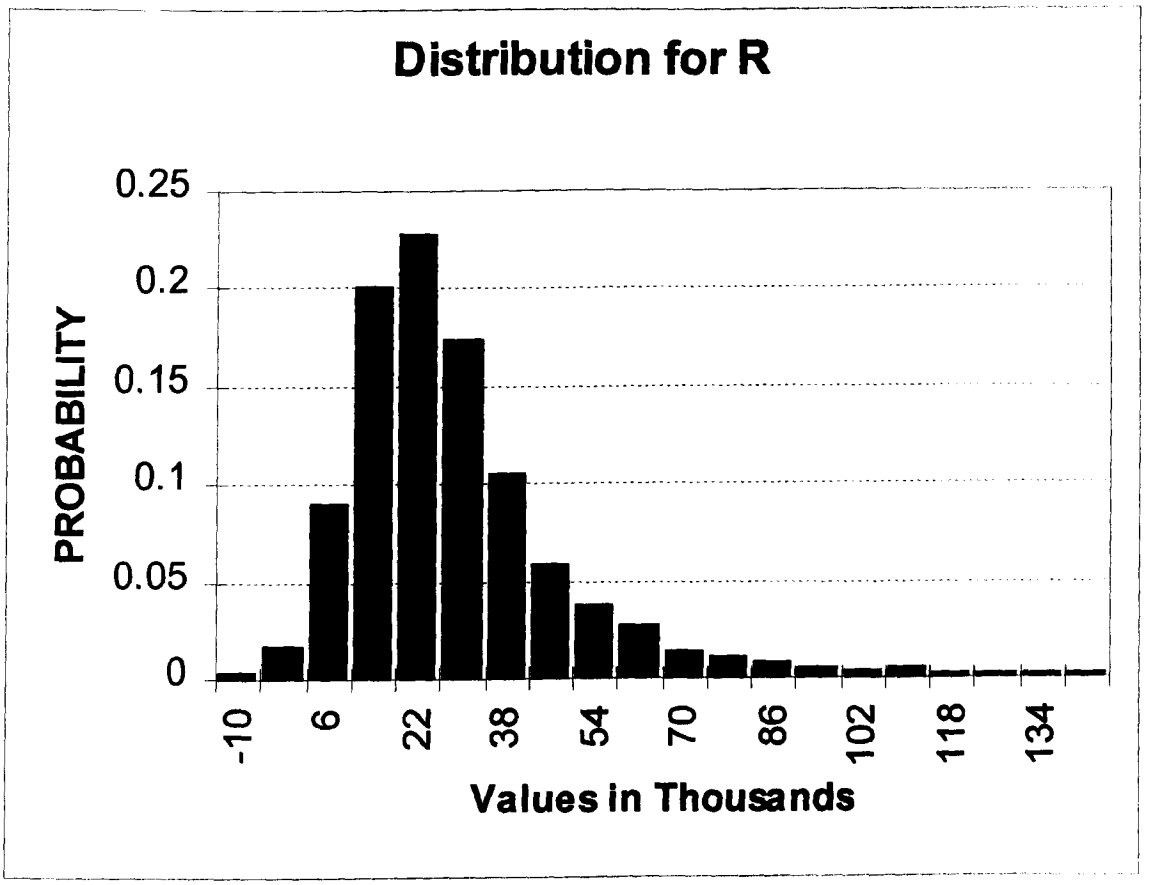


Table 3: Descriptive statistics of probability distribution – subgroup 3

Name	Ee - Ec	R	Ce - Cc	effect E	effect C
Minimum =	0.15	- 186,882,300	1,342	0.26	0.06
Maximum =	0.62	106,551,000	29,496	0.84	0.65
Mean =	0.24	63,290	15,271	0.58	0.34
Std Deviati	0.11	2,424,596	3,636	0.07	0.08
Variance =	0.01	5,878,665,000,000	13,220,320	0.00	0.01
Skewness	0.02	- 39	0	0.00	- 0.02
Kurtosis =	2.98	4,024	3	3.02	3.03
Mode =	0.20	45,575	11,752	0.52	0.25
5% Perc =	0.07	28,638	9,241	0.47	0.21
10% Perc =	0.10	34,349	10,659	0.49	0.24
15% Perc =	0.13	38,753	11,507	0.51	0.26
20% Perc =	0.15	42,270	12,208	0.52	0.27
25% Perc =	0.17	45,642	12,842	0.53	0.29
30% Perc =	0.19	48,733	13,344	0.54	0.30
35% Perc =	0.20	51,977	13,866	0.55	0.31
40% Perc =	0.21	55,395	14,340	0.56	0.32
45% Perc =	0.23	58,804	14,816	0.57	0.33
50% Perc =	0.24	63,015	15,242	0.58	0.34
55% Perc =	0.25	67,156	15,697	0.59	0.35
60% Perc =	0.27	71,653	16,171	0.60	0.36
65% Perc =	0.28	77,036	16,633	0.61	0.37
70% Perc =	0.29	83,127	17,148	0.62	0.38
75% Perc =	0.31	90,862	17,665	0.63	0.39
80% Perc =	0.33	100,698	18,306	0.64	0.41
85% Perc =	0.35	115,578	19,039	0.65	0.42
90% Perc =	0.38	141,539	19,953	0.67	0.44
95% Perc =	0.42	206,888	21,340	0.70	0.47

Figure 3: Probability distribution of $R = (C_E - C_C) / (E_E - E_C)$ – subgroup 3

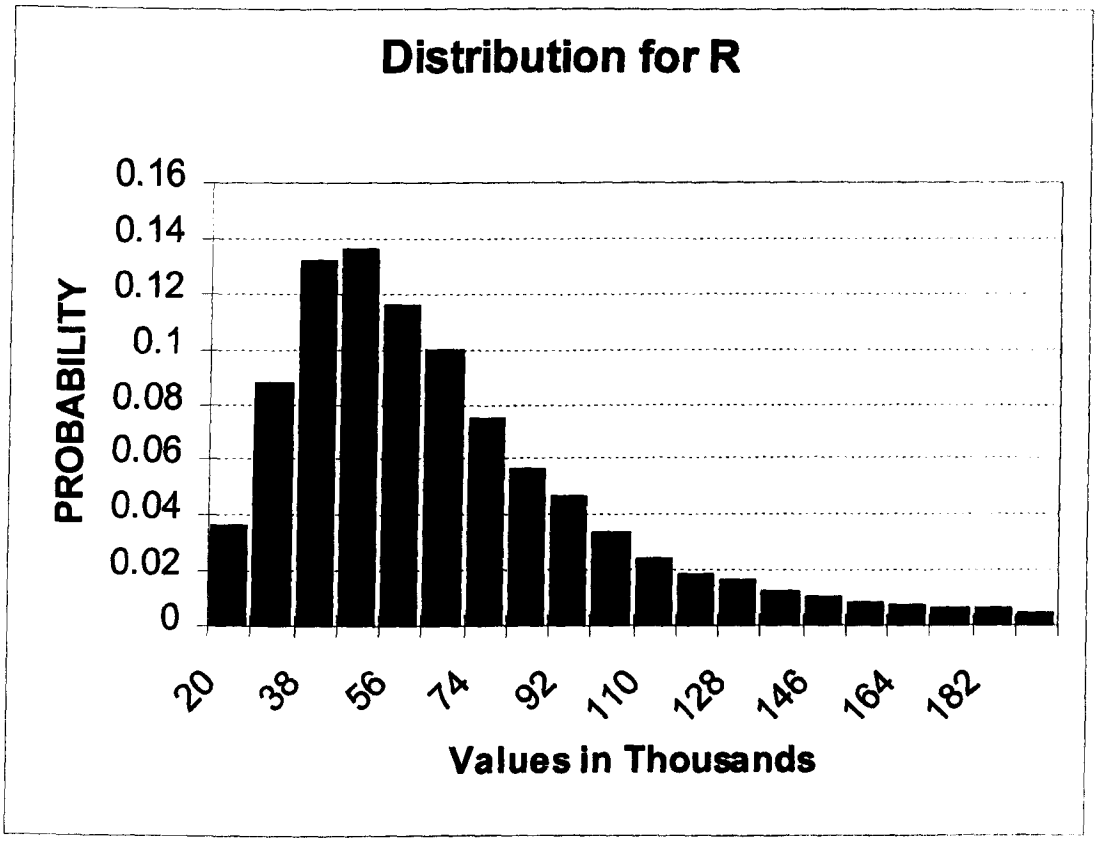


Table 4: Descriptive statistics of probability distribution – subgroup 4

Name	Ee - Ec	R	Ce - Cc	effect E	effect C
Minimum =	-0.08	-5,921,415	3,987	0.42	0.14
Maximum =	0.66	1,148,895,000	22,214	0.99	0.70
Mean =	0.30	163,231	12,229	0.73	0.43
Std Deviati	0.10	11,488,450	2,413	0.07	0.07
Variance =	0.01	131,984,500,000,000	5,820,236	0.00	0.00
Skewness =	0.00	100	0	0.01	-0.01
Kurtosis =	2.99	9,996	3	3.04	3.08
Mode =	0.23	35,917	10,764	0.65	0.36
5% Perc =	0.14	22,637	8,247	0.61	0.31
10% Perc =	0.17	25,691	9,143	0.64	0.34
15% Perc =	0.20	28,186	9,721	0.66	0.36
20% Perc =	0.22	30,167	10,203	0.67	0.37
25% Perc =	0.23	31,826	10,603	0.68	0.38
30% Perc =	0.25	33,525	10,943	0.69	0.39
35% Perc =	0.26	35,302	11,286	0.70	0.40
40% Perc =	0.27	37,024	11,593	0.71	0.41
45% Perc =	0.29	38,831	11,898	0.72	0.42
50% Perc =	0.30	40,650	12,218	0.73	0.43
55% Perc =	0.31	42,823	12,535	0.74	0.44
60% Perc =	0.33	44,936	12,859	0.75	0.45
65% Perc =	0.34	47,456	13,168	0.76	0.46
70% Perc =	0.35	50,287	13,515	0.77	0.47
75% Perc =	0.37	53,707	13,860	0.78	0.48
80% Perc =	0.38	57,918	14,284	0.79	0.49
85% Perc =	0.40	63,594	14,747	0.80	0.50
90% Perc =	0.43	72,668	15,336	0.82	0.52
95% Perc =	0.46	91,789	16,175	0.85	0.54

Figure 4: Probability distribution of $R = (C_E - C_C) / (E_E - E_C)$ – subgroup 4

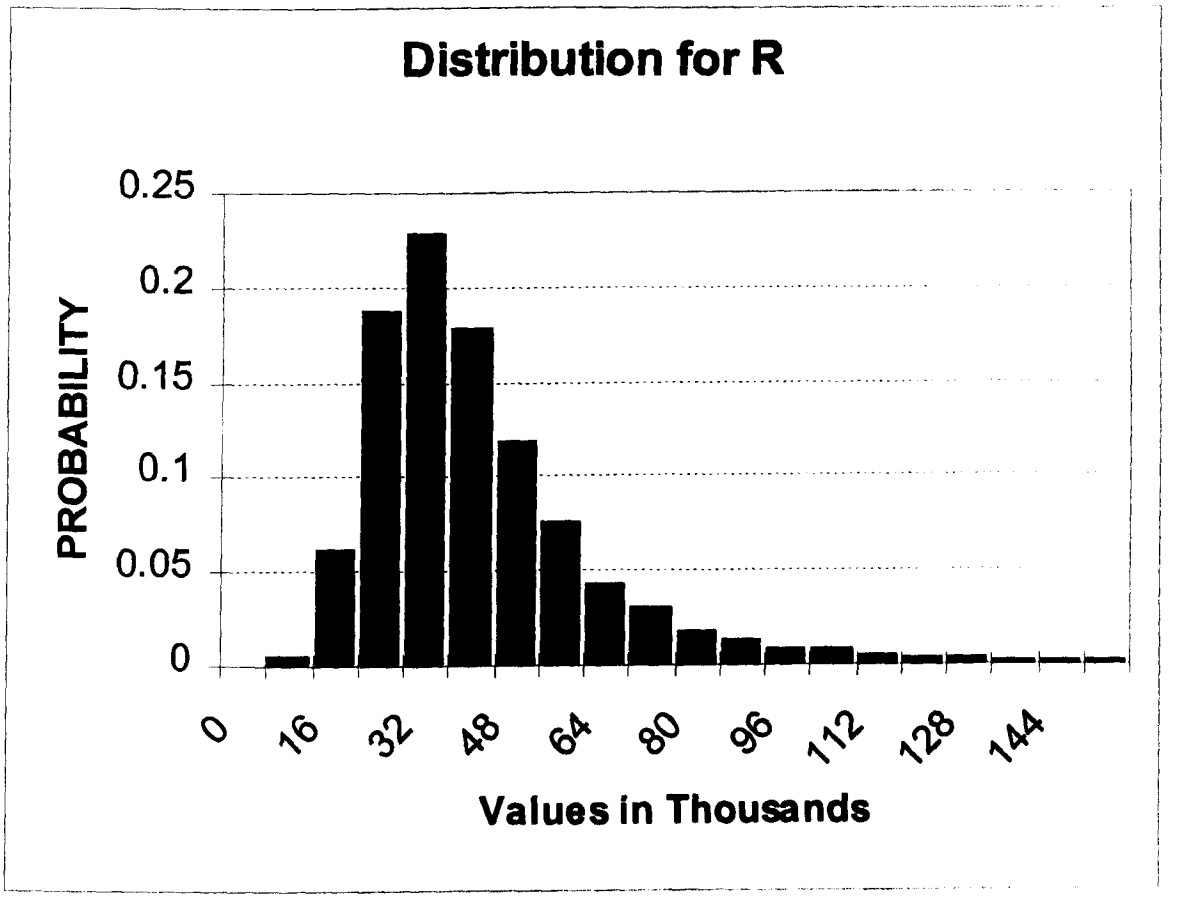


Table 5: Descriptive statistics of probability distribution – subgroup 5

Name	Ee - Ec	R	Ce - Cc	effect E
Minimum =-	0.15	- 55,851,440	10,571	- 0.15
Maximum =	0.65	370,262,000	62,112	0.65
Mean =	0.22	250,358	36,201	0.22
Std Deviati	0.10	3,995,778	6,319	0.10
Variance =	0.01	15,966,240,000,000	39,933,830	0.01
Skewness	0.00	80	- 0	0.00
Kurtosis =	3.03	7,370	3	3.03
Mode =	0.07	128,624	35,654	0.07
5% Perc =	0.05	81,035	25,530	0.05
10% Perc =	0.09	94,586	28,042	0.09
15% Perc =	0.12	104,707	29,675	0.12
20% Perc =	0.14	113,024	30,974	0.14
25% Perc =	0.15	120,561	31,990	0.15
30% Perc =	0.17	128,391	32,927	0.17
35% Perc =	0.18	136,066	33,791	0.18
40% Perc =	0.19	144,085	34,645	0.19
45% Perc =	0.21	152,598	35,445	0.21
50% Perc =	0.22	162,146	36,234	0.22
55% Perc =	0.23	171,819	37,029	0.23
60% Perc =	0.25	183,749	37,821	0.25
65% Perc =	0.26	197,361	38,658	0.26
70% Perc =	0.27	213,739	39,511	0.27
75% Perc =	0.29	233,056	40,491	0.29
80% Perc =	0.30	261,027	41,560	0.30
85% Perc =	0.32	300,422	42,764	0.32
90% Perc =	0.35	371,911	44,285	0.35
95% Perc =	0.38	539,475	46,581	0.38

Figure 5: Probability distribution of $R = (C_E - C_C) / (E_E - E_C)$ – subgroup 5

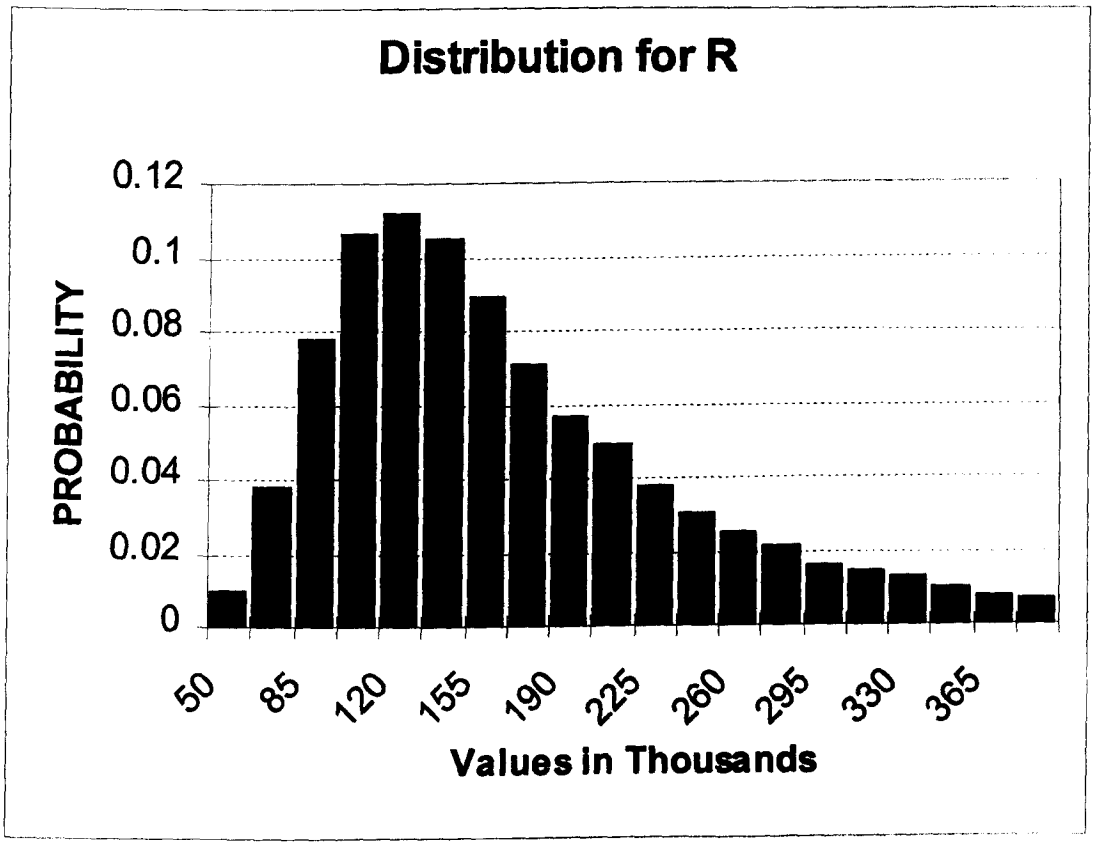


Table 6: Descriptive statistics of probability distribution – subgroup 6

Name	Ee - Ec	R	Ce - Cc	effect E	effect C
Minimum =	-0.05	-1042598	-8816.766	0.52	-0.09
Maximum =	1.11	3441217	23747.1	1.21	0.80
Mean =	0.49	20105.52	8235.314	0.87	0.38
Std Deviati	0.16	42040.23	4274.482	0.09	0.13
Variance =	0.02	1.77E+09	1.83E+07	0.01	0.02
Skewness	0.04	54.75468	-6.83E-02	-0.04	-0.00
Kurtosis =	2.98	4471.903	2.985778	3.00	2.95
Mode =	0.50	12585.84	11473.76	0.80	0.30
5% Perc =	0.23	1957.928	967.4439	0.72	0.17
10% Perc =	0.29	5339.917	2747.382	0.75	0.22
15% Perc =	0.32	7587.487	3842.766	0.78	0.25
20% Perc =	0.36	9139.431	4672.26	0.79	0.27
25% Perc =	0.38	10540.44	5371.251	0.81	0.29
30% Perc =	0.41	12005.91	6017.742	0.82	0.31
35% Perc =	0.43	13259.48	6637.041	0.84	0.33
40% Perc =	0.45	14495.47	7218.548	0.85	0.35
45% Perc =	0.47	15747.94	7751.034	0.86	0.37
50% Perc =	0.49	16927.39	8303.717	0.87	0.38
55% Perc =	0.51	18283.21	8795.231	0.88	0.40
60% Perc =	0.53	19654	9346.19	0.89	0.41
65% Perc =	0.55	21322.99	9918.157	0.90	0.43
70% Perc =	0.57	22950.97	10506.01	0.92	0.45
75% Perc =	0.59	24867.26	11165.74	0.93	0.47
80% Perc =	0.62	27231.59	11853.43	0.94	0.49
85% Perc =	0.65	30425.55	12682.44	0.96	0.51
90% Perc =	0.69	35061.55	13708.04	0.99	0.55
95% Perc =	0.75	44429.67	15148.89	1.02	0.59

Figure 6: Probability distribution of $R = (C_E - C_C) / (E_E - E_C)$ – subgroup 6

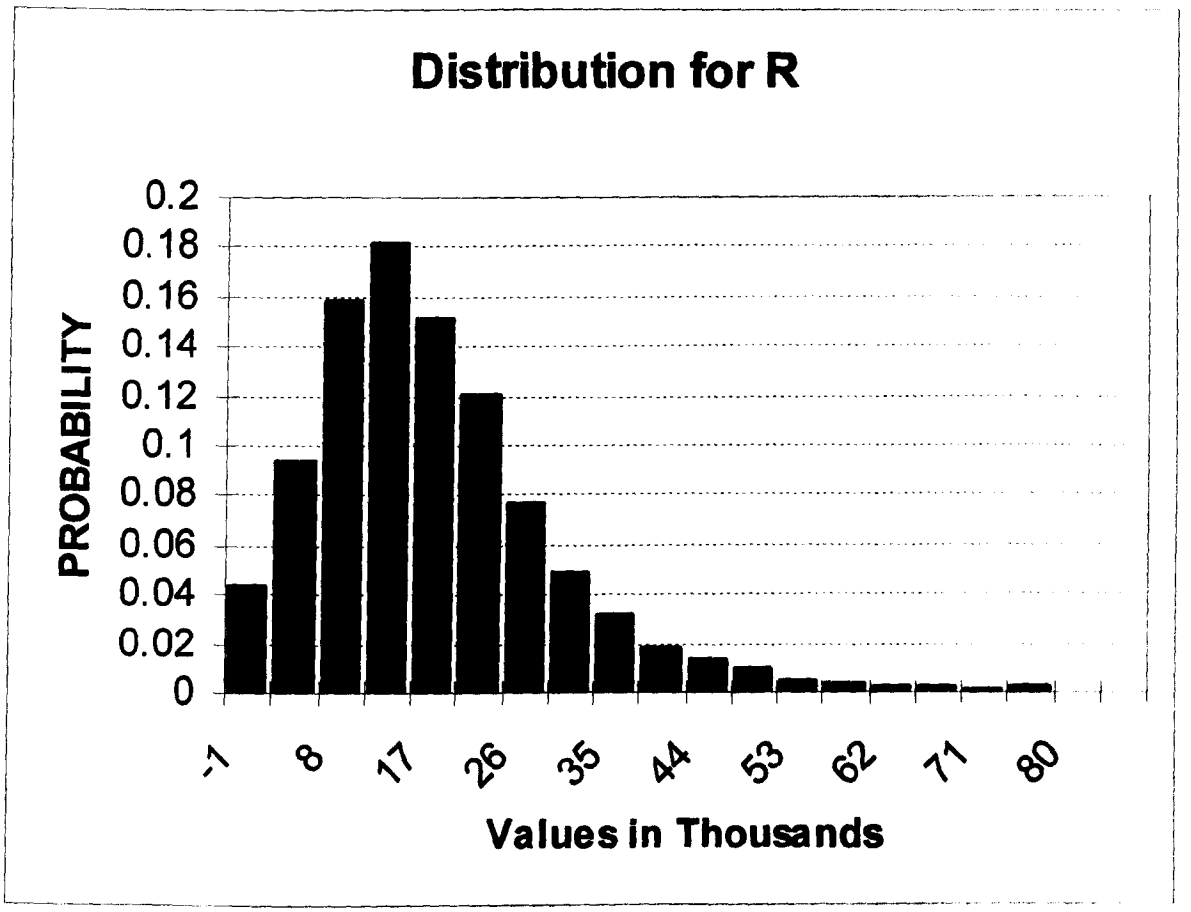


Table 7: Descriptive statistics of probability distribution – subgroup 7

Name	Ee - Ec	R	Ce - Cc	effect E	effect C
Minimum =	- 0.05	- 2,078,959	575	0.52	0.25
Maximum =	0.57	4,609,683	15,838	0.97	0.72
Mean =	0.25	39,502	8,325	0.75	0.50
Std Deviati	0.08	72,574	2,120	0.06	0.06
Variance =	0.01	5,266,931,000	4,492,929	0.00	0.00
Skewness	0.06	25	0	0.01	- 0.04
Kurtosis =	3.00	1,905	3	2.96	3.02
Mode =	0.26	20,599	8,603	0.66	0.41
5% Perc =	0.11	16,568	4,919	0.65	0.40
10% Perc =	0.14	19,319	5,665	0.67	0.42
15% Perc =	0.16	21,522	6,159	0.69	0.44
20% Perc =	0.18	23,129	6,519	0.70	0.45
25% Perc =	0.19	24,891	6,865	0.71	0.46
30% Perc =	0.21	26,504	7,165	0.72	0.47
35% Perc =	0.22	28,036	7,459	0.73	0.48
40% Perc =	0.23	29,629	7,751	0.73	0.48
45% Perc =	0.24	31,267	8,035	0.74	0.49
50% Perc =	0.25	33,037	8,306	0.75	0.50
55% Perc =	0.26	34,966	8,563	0.76	0.51
60% Perc =	0.27	37,013	8,847	0.77	0.51
65% Perc =	0.28	39,341	9,142	0.77	0.52
70% Perc =	0.30	41,758	9,442	0.78	0.53
75% Perc =	0.31	44,932	9,762	0.79	0.54
80% Perc =	0.32	48,455	10,120	0.80	0.55
85% Perc =	0.34	53,762	10,548	0.81	0.56
90% Perc =	0.36	61,427	11,106	0.83	0.58
95% Perc =	0.40	77,879	11,833	0.85	0.60

Figure 7: Probability distribution of $R = (C_E - C_C) / (E_E - E_C)$ – subgroup 7

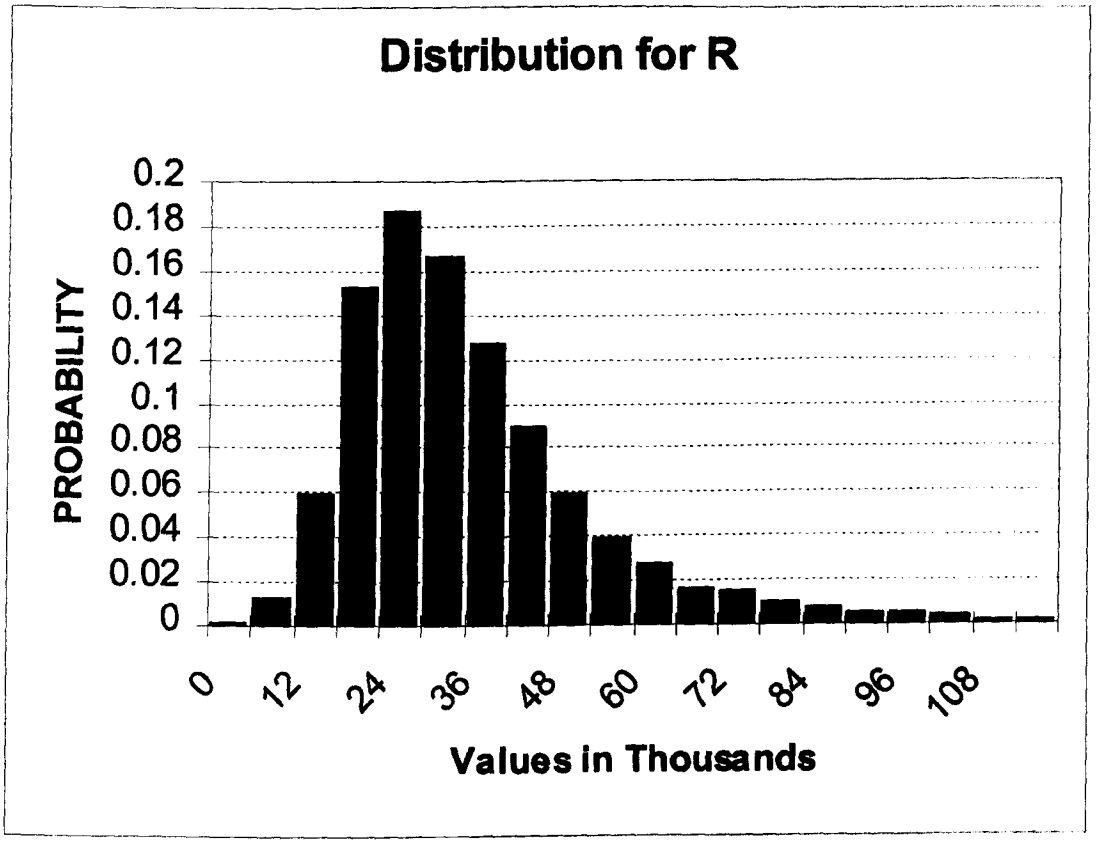


Table 8: Descriptive statistics of probability distribution – subgroup 8

Name	Ee - Ec	R	Ce - Cc	effect E	effect C
Minimum =	- 0.05	- 1,934,004	1,956	0.54	0.17
Maximum	0.68	32,194,780	22,127	0.97	0.73
Mean =	0.30	49,793	12,242	0.75	0.45
Std Deviati	0.09	325,215	2,647	0.06	0.07
Variance =	0.01	105,764,500,000	7,007,751	0.00	0.00
Skewness	- 0.00	97	0	- 0.02	- 0.01
Kurtosis =	3.03	9,545	3	2.93	3.02
Mode =	0.27	20,859	12,184	0.66	0.37
5% Perc =	0.15	22,334	7,907	0.65	0.33
10% Perc	0.18	25,766	8,843	0.67	0.36
15% Perc	0.21	28,142	9,511	0.69	0.38
20% Perc	0.22	30,081	10,048	0.70	0.39
25% Perc	0.24	31,896	10,492	0.71	0.40
30% Perc	0.25	33,733	10,884	0.72	0.41
35% Perc	0.26	35,528	11,235	0.73	0.42
40% Perc	0.28	37,265	11,562	0.73	0.43
45% Perc	0.29	39,018	11,866	0.74	0.44
50% Perc	0.30	40,861	12,212	0.75	0.45
55% Perc	0.31	42,859	12,551	0.76	0.46
60% Perc	0.32	44,928	12,902	0.77	0.47
65% Perc	0.33	47,235	13,275	0.77	0.48
70% Perc	0.35	49,698	13,624	0.78	0.49
75% Perc	0.36	52,953	14,028	0.79	0.50
80% Perc	0.38	57,061	14,472	0.80	0.51
85% Perc	0.39	62,324	14,973	0.81	0.52
90% Perc	0.42	70,571	15,651	0.83	0.54
95% Perc	0.45	86,210	16,624	0.85	0.56

Figure 8: Probability distribution of $R = (C_E - C_C) / (E_E - E_C)$ – subgroup 8

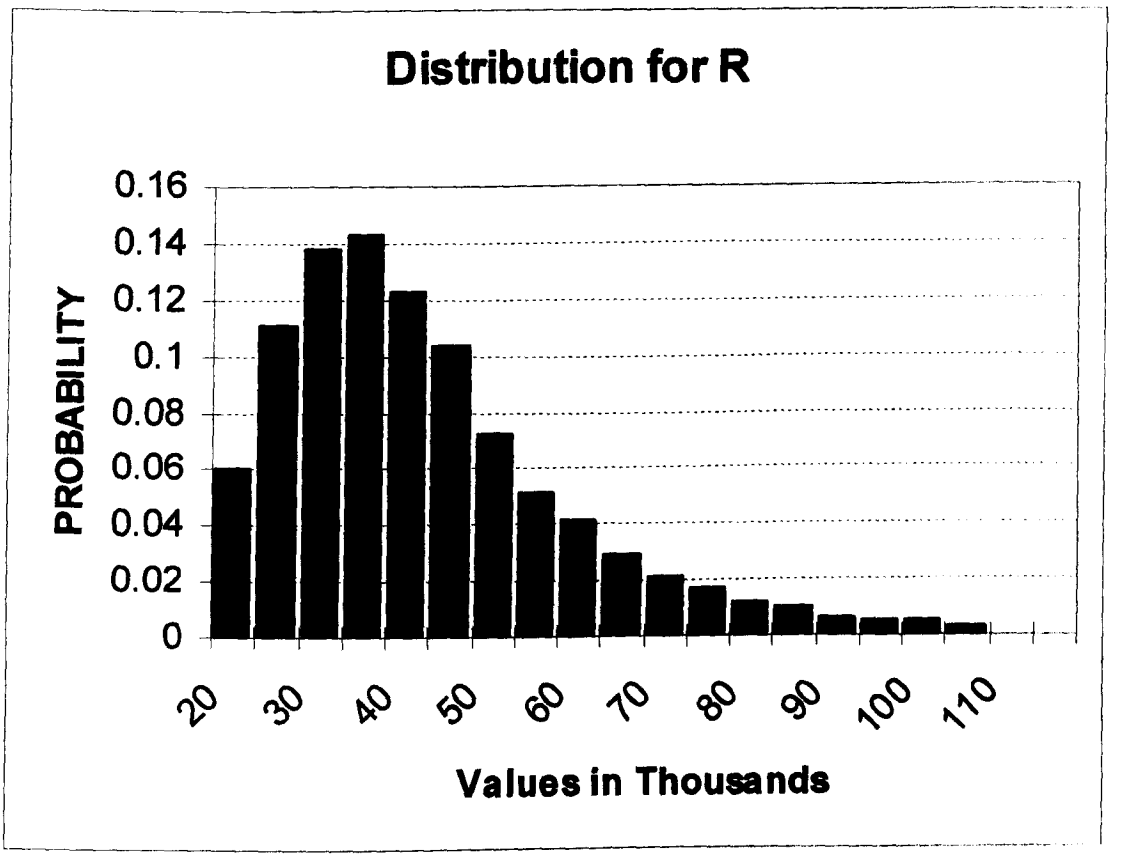


Table 9: Descriptive statistics of probability distribution – subgroup 9

Name	Ee - Ec	R	Ce - Cc	effect E	effect C
Minimum =	- 0.18	- 15,253,970	- 65	0.23	- 0.00
Maximum	0.67	324,426,500	29,937	0.81	0.57
Mean =	0.21	153,227	16,278	0.51	0.30
Std Deviati	0.11	3,447,760	3,530	0.08	0.08
Variance =	0.01	11,887,050,000,000	12,459,730	0.01	0.01
Skewness	- 0.01	85	- 0	0.02	0.01
Kurtosis =	3.00	7,850	3	3.05	2.96
Mode =	0.13	51,353	12,219	0.41	0.21
5% Perc =	0.03	29,945	10,494	0.38	0.17
10% Perc	0.07	38,990	11,745	0.41	0.20
15% Perc	0.09	43,778	12,633	0.43	0.22
20% Perc	0.12	48,048	13,301	0.44	0.23
25% Perc	0.13	52,221	13,905	0.46	0.25
30% Perc	0.15	56,124	14,468	0.47	0.26
35% Perc	0.17	60,166	14,933	0.48	0.27
40% Perc	0.18	64,804	15,401	0.49	0.28
45% Perc	0.20	69,441	15,839	0.50	0.29
50% Perc	0.21	74,564	16,277	0.51	0.30
55% Perc	0.22	79,487	16,720	0.52	0.31
60% Perc	0.24	85,513	17,157	0.53	0.32
65% Perc	0.25	92,461	17,637	0.54	0.33
70% Perc	0.27	101,976	18,087	0.55	0.34
75% Perc	0.29	113,751	18,648	0.56	0.35
80% Perc	0.30	129,637	19,221	0.58	0.37
85% Perc	0.33	154,528	19,916	0.59	0.38
90% Perc	0.35	196,578	20,836	0.61	0.40
95% Perc	0.39	318,934	22,160	0.64	0.43

Figure 9: Probability distribution of $R = (C_E - C_C) / (E_E - E_C)$ – subgroup 9

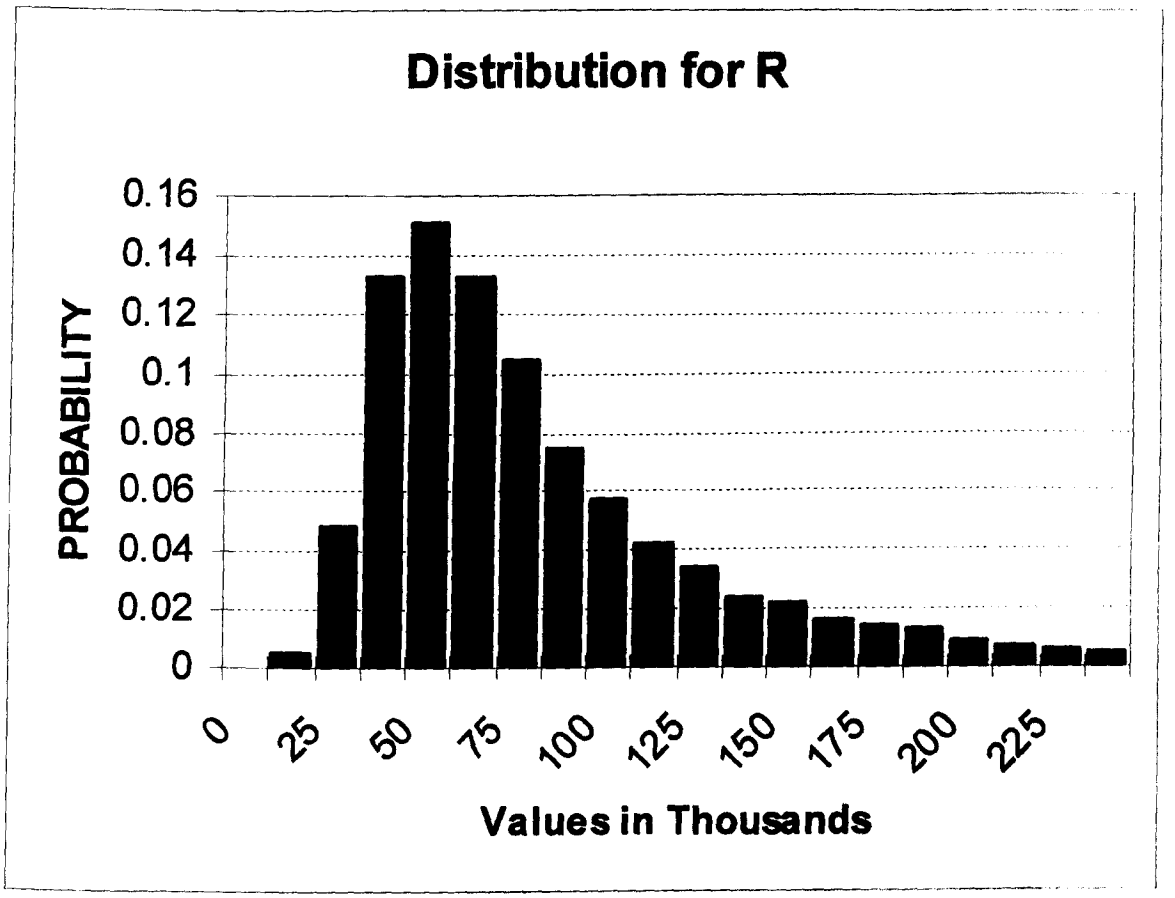
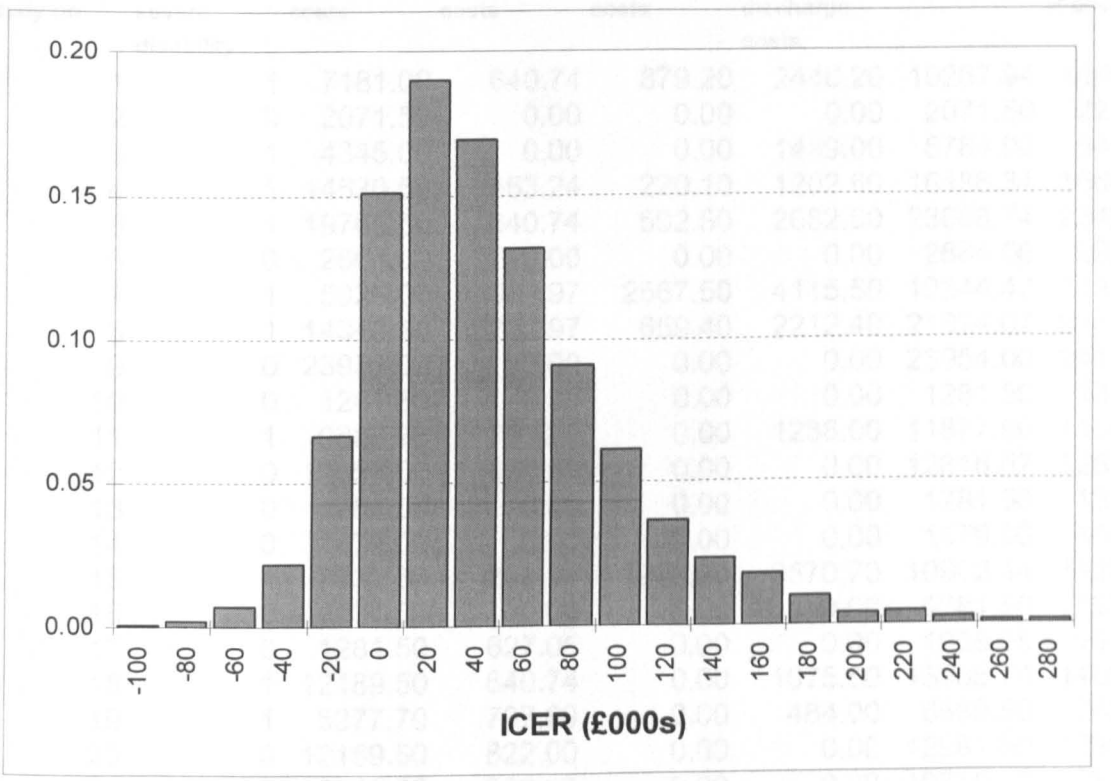


Table 10: Descriptive statistics of probability distribution – ECMO study

Name	(Et-Ec)	R	Et	Ec	(Ct-Cc)
Minimum =	0.01	16,949	0.44	0.19	6,271
Maximum	0.54	2,214,339	0.85	0.58	22,597
Mean =	0.27	56,757	0.66	0.39	13,868
Std Deviati	0.07	33,675	0.05	0.05	2,110
Variance =	0.01	1,134,020,000	0.00	0.00	4,454,027
Skewness	- 0.01	31	- 0.01	0.03	0
Kurtosis =	3.05	1,821	3.02	3.02	3
Mode =	0.19	38,791	0.58	0.35	13,497
5% Perc =	0.15	32,351	0.57	0.31	10,397
10% Perc	0.18	35,939	0.59	0.33	11,181
15% Perc	0.20	38,329	0.61	0.34	11,692
20% Perc	0.21	40,357	0.62	0.35	12,104
25% Perc	0.22	42,431	0.62	0.36	12,441
30% Perc	0.23	44,238	0.63	0.36	12,760
35% Perc	0.24	45,961	0.64	0.37	13,051
40% Perc	0.25	47,754	0.65	0.38	13,319
45% Perc	0.26	49,543	0.65	0.38	13,585
50% Perc	0.27	51,383	0.66	0.39	13,848
55% Perc	0.28	53,459	0.67	0.40	14,122
60% Perc	0.29	55,687	0.67	0.40	14,391
65% Perc	0.30	58,234	0.68	0.41	14,652
70% Perc	0.31	60,997	0.69	0.42	14,955
75% Perc	0.32	64,365	0.69	0.42	15,284
80% Perc	0.33	68,154	0.70	0.43	15,663
85% Perc	0.34	73,187	0.71	0.44	16,078
90% Perc	0.36	81,462	0.72	0.45	16,612
95% Perc	0.38	95,137	0.74	0.47	17,325

Figure 10: Probability distribution of $R = (C_E - C_C) / (E_E - E_C)$ – ECMO study



ECMO Study data tables

Patient Study no.	Survival w/o severe disability	Admission costs	Transport costs	Readmission costs	Post-discharge costs	Total costs	Total costs at one year
1	1	7181.00	640.74	879.20	2446.20	10267.94	10404.20
2	0	2071.50	0.00	0.00	0.00	2071.50	2233.35
3	1	4345.00	0.00	0.00	1439.00	5784.00	6496.14
4	1	14630.50	653.24	220.10	1202.60	16486.34	16630.17
5	1	19765.50	640.74	592.50	2682.50	23088.74	23429.39
6	0	2664.00	0.00	0.00	0.00	2664.00	2922.96
7	1	5925.00	303.97	2567.50	4115.50	10344.47	11315.57
8	1	14349.30	5262.97	659.40	2212.40	21824.67	21494.68
9	0	23954.00	0.00	0.00	0.00	23954.00	24135.68
10	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
11	1	9888.70	753.20	0.00	1236.00	11877.90	11562.23
12	0	12357.00	459.67	0.00	0.00	12816.67	12831.81
13	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
14	0	1479.00	0.00	0.00	0.00	1479.00	1543.74
15	1	7033.60	298.14	1978.20	3570.70	10902.44	11224.68
16	0	1281.50	0.00	0.00	0.00	1281.50	1289.07
17	0	1281.50	627.05	0.00	0.00	1908.55	1940.92
18	1	12189.50	640.74	0.00	1075.50	13905.74	14094.99
19	1	5377.70	737.89	0.00	484.00	6599.59	7428.38
20	0	12159.50	822.00	0.00	0.00	12981.50	12989.07
21	0	10346.50	369.67	0.00	0.00	10716.17	10723.74
22	0	1281.50	338.07	0.00	0.00	1619.57	1651.94
23	0	13921.50	0.00	0.00	0.00	13921.50	16025.55
24	1	3555.00	1126.49	879.80	1614.80	6296.29	6432.55
25	0	1303.80	450.56	0.00	0.00	1754.36	1764.43
26	1	7019.00	640.74	0.00	1749.00	9408.74	9469.30
27	1	97086.40	1095.19	0.00	0.00	98181.59	93451.24
28	0	25832.50	634.50	8887.50	10203.00	36670.00	36965.23
29	0	1281.50	366.33	0.00	0.00	1647.83	1655.40
30	0	6811.50	0.00	0.00	0.00	6811.50	7750.23
31	0	1303.80	0.00	0.00	0.00	1303.80	1313.87
32	0	1281.50	0.00	0.00	0.00	1281.50	1289.07
33	0	1479.00	280.66	0.00	0.00	1759.66	1824.40
34	1	3077.20	0.00	3160.00	4502.50	7579.70	7720.68
35	1	6122.50	0.00	1580.00	3676.50	9799.00	10802.47
36	1	7019.00	1190.19	440.20	1691.20	9900.39	9960.95
37	0	1479.00	0.00	0.00	0.00	1479.00	1543.74
38	1	8132.60	0.00	0.00	1501.50	9634.10	10006.69
39	1	13482.40	1423.33	0.00	1452.00	16357.73	15901.92
40	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
41	0	1281.50	0.00	0.00	0.00	1281.50	1289.07
42	0	19411.50	5000.00	0.00	0.00	24411.50	24419.07
43	1	12422.50	6694.94	0.00	1471.00	20588.44	20717.13
44	1	14433.00	6732.58	0.00	3360.50	24526.08	24662.34
45	1	3752.50	1811.05	0.00	1563.50	7127.05	7742.08

Patient Study no.	Survival w/o severe disability	Admission costs	Transport costs	Readmission costs	Post-discharge costs	Total costs	Total costs at one year
46	1	3297.00	0.00	879.20	3098.20	6395.20	6546.25
47	0	1304.10	0.00	0.00	0.00	1304.10	1313.87
48	1	16876.30	728.21	1320.60	4069.10	21673.61	21417.27
49	0	2906.10	295.23	0.00	0.00	3201.33	3448.06
50	1	11794.50	0.00	0.00	954.00	12748.50	12922.61
51	1	5424.90	280.66	0.00	1024.50	6730.06	7281.78
52	0	2071.50	0.00	0.00	0.00	2071.50	2233.35
53	1	9108.70	815.68	0.00	1062.50	10986.88	10803.37
54	1	9723.50	1267.23	0.00	2874.50	13865.23	13682.13
55	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
56	0	1874.00	0.00	0.00	0.00	1874.00	2003.48
57	1	6193.50	6693.27	0.00	1312.00	14198.77	14297.18
58	1	11931.50	646.99	0.00	1408.50	13986.99	13819.03
59	1	10395.30	12524.65	660.30	2033.80	24953.75	24886.50
60	0	1304.10	0.00	0.00	0.00	1304.10	1313.87
61	1	13188.00	0.00	220.10	1473.10	14661.10	15265.30
62	1	37412.20	331.66	11652.50	14218.50	51962.36	52539.32
63	0	45070.10	977.25	0.00	0.00	46047.35	46964.14
64	1	8995.60	722.34	0.00	1393.50	11111.44	11770.38
65	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
66	0	1281.50	0.00	0.00	0.00	1281.50	1289.07
67	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
68	1	5201.90	1164.87	0.00	1790.00	8156.77	8286.69
69	0	1874.00	403.80	0.00	0.00	2277.80	2407.28
70	0	37475.00	814.83	0.00	0.00	38289.83	39198.23
71	1	12800.70	1032.25	0.00	1598.50	15431.45	15042.13
72	1	24172.90	1534.81	10775.00	13238.00	38945.71	38644.77
73	0	1281.50	342.60	0.00	0.00	1624.10	1656.47
74	1	10017.00	640.74	197.50	1088.50	11746.24	11852.22
75	0	2071.50	0.00	0.00	0.00	2071.50	2233.35
76	0	1479.00	0.00	0.00	0.00	1479.00	1543.74
77	1	17664.00	922.25	0.00	2645.50	21231.75	21352.87
78	1	5816.70	298.14	0.00	1610.00	7724.84	8574.37
79	1	2962.50	765.68	0.00	999.50	4727.68	4841.23
80	1	5495.00	0.00	0.00	2233.00	7728.00	7979.75
81	1	18885.00	280.66	879.20	2501.70	21667.36	21140.77
82	1	80777.50	0.00	0.00	1868.00	82645.50	95884.83
83	0	1281.50	344.87	0.00	0.00	1626.37	1658.74
84	1	17988.00	280.66	0.00	1106.50	19375.16	19647.68
85	0	6517.50	0.00	2567.50	4791.00	11308.50	12376.71
86	1	10392.90	709.47	0.00	1063.00	12165.37	12100.52
87	1	4740.00	280.66	2765.00	4343.50	9364.16	10141.04
88	1	9191.50	634.50	0.00	1714.50	11540.50	11684.33
89	1	3383.00	634.27	1318.80	2768.80	6786.07	7080.99
90	1	3580.50	359.34	0.00	2092.00	6031.84	5937.53
91	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
92	1	12645.50	4444.22	0.00	672.50	17762.22	17667.91

Patient Study no.	Survival w/o severe disability	Admission costs	Transport costs	Readmission costs	Post-discharge costs	Total costs	Total costs at one year
93	0	13662.60	6955.82	6278.60	6278.60	26897.02	26563.97
94	1	12110.40	856.17	0.00	548.00	13514.57	14736.63
95	1	7967.20	1280.03	660.30	1741.30	10988.53	10921.17
96	1	3516.80	0.00	0.00	1096.00	4612.80	4773.92
97	0	4046.50	0.00	0.00	0.00	4046.50	4532.05
98	0	13972.50	4099.00	0.00	0.00	18071.50	18079.07
99	0	24617.50	0.00	0.00	0.00	24617.50	24685.63
100	1	8679.10	418.63	0.00	2483.50	11581.23	11589.62
101	1	8690.00	0.00	0.00	954.50	9644.50	11068.78
102	1	12817.50	499.67	0.00	921.00	14238.17	14382.00
103	0	3454.00	492.08	0.00	0.00	3946.08	4334.52
104	0	45710.00	904.79	0.00	0.00	46614.79	47391.67
105	1	15634.40	709.47	1758.40	2982.90	19326.77	19254.35
106	1	3736.60	0.00	1318.80	2577.80	6314.40	6485.59
107	1	7110.00	0.00	3297.30	4281.80	11391.80	12557.12
108	1	27215.00	6327.55	0.00	1452.00	34994.55	35342.77
109	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
110	1	34697.30	1203.26	1758.40	3812.90	39713.46	39561.05
111	1	17628.50	646.99	0.00	1688.50	19963.99	20153.24
112	0	41167.50	4069.99	0.00	0.00	45237.49	45245.06
113	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
114	0	1303.80	0.00	0.00	0.00	1303.80	1313.87
115	0	11263.00	709.36	0.00	0.00	11972.36	12154.04
116	1	10110.80	0.00	0.00	1001.50	11112.30	11575.52
117	0	30289.50	0.00	0.00	0.00	30289.50	30297.07
118	0	1479.00	0.00	0.00	0.00	1479.00	1543.74
119	0	66270.30	1487.03	0.00	0.00	67757.33	64814.42
120	1	11217.60	1167.10	0.00	1030.50	13415.20	13378.56
121	0	47310.00	906.61	0.00	0.00	48216.61	48807.07
122	0	1479.00	0.00	0.00	0.00	1479.00	1543.74
123	1	12360.60	1886.42	0.00	1043.50	15290.52	15276.04
124	1	3425.30	562.45	219.80	1205.30	5193.05	5675.54
125	0	28836.00	1283.64	0.00	0.00	30119.64	30210.48
126	0	1281.50	5000.00	0.00	0.00	6281.50	6313.87
127	0	1303.80	0.00	0.00	0.00	1303.80	1313.87
128	1	9768.10	878.82	2417.80	4480.80	15127.72	14900.02
129	1	3357.50	0.00	0.00	1234.50	4592.00	5142.29
130	1	5537.30	790.69	0.00	1169.00	7496.99	7431.32
131	1	52031.40	1138.23	1538.60	5037.10	58206.73	57862.26
132	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
133	0	1303.80	280.66	0.00	0.00	1584.46	1594.53
134	1	2765.00	0.00	879.20	1988.20	4753.20	4859.18
135	0	2466.50	516.95	0.00	0.00	2983.45	3210.04
136	0	1281.50	0.00	0.00	0.00	1281.50	1289.07
137	0	10687.60	280.66	0.00	0.00	10968.26	12693.64
138	0	18118.20	653.24	0.00	0.00	18771.44	18331.18
139	1	7216.50	6096.59	0.00	951.50	14264.59	14332.72

Patient Study no.	Survival w/o severe disability	Admission costs	Transport costs	Readmission costs	Post-discharge costs	Total costs	Total costs at one year
140	1	10270.00	286.49	0.00	1154.00	11710.49	13393.73
141	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
142	1	9207.00	568.99	0.00	2264.00	12039.99	11790.81
143	1	13558.50	765.61	0.00	1846.50	16170.61	16156.02
144	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
145	1	6912.50	280.66	220.10	2454.60	9647.76	10780.71
146	1	32871.10	667.30	197.50	1541.50	35079.90	34996.66
147	1	4587.70	347.69	0.00	1972.50	6907.89	7607.20
148	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
149	0	1676.50	902.41	0.00	0.00	2578.91	2601.62
150	1	8295.00	0.00	1320.60	3052.60	11347.60	12707.14
151	1	22996.10	810.69	4615.80	7629.30	31436.09	30974.35
152	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
153	1	16392.10	709.65	0.00	2001.50	19103.25	18920.97
154	0	60107.50	6337.04	0.00	0.00	66444.54	66830.61
155	0	3651.50	0.00	0.00	0.00	3651.50	4072.31
156	0	1479.00	0.00	0.00	0.00	1479.00	1543.74
157	1	9464.30	0.00	1320.60	3813.10	13277.40	13697.51
158	1	18780.10	412.10	0.00	1513.00	20705.20	20563.13
159	1	12027.50	646.99	0.00	968.50	13642.99	13756.54
160	1	4345.00	1002.24	0.00	1419.50	6766.74	7478.88
161	0	55671.50	0.00	0.00	0.00	55671.50	55679.07
162	1	15811.70	790.69	0.00	2060.50	18662.89	18618.24
163	1	13627.50	0.00	5135.00	7426.50	21054.00	23287.53
164	1	2962.50	0.00	592.50	1391.00	4353.50	4839.05
165	1	8006.50	634.50	0.00	1267.00	9908.00	10006.41
166	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
167	0	25245.50	8444.33	0.00	0.00	33689.83	33712.54
168	1	10143.50	665.73	0.00	688.00	11497.23	11747.04
169	0	2269.00	283.57	0.00	0.00	2552.57	2746.79
170	1	27736.50	659.90	0.00	813.50	29209.90	29717.09
171	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
172	1	2567.50	0.00	395.00	1711.00	4278.50	4699.31
173	1	11652.50	280.66	0.00	2317.50	14250.66	16160.49
174	1	6547.10	878.82	0.00	750.00	8175.92	8125.80
175	1	9857.70	6734.82	0.00	767.00	17359.52	17247.29
176	1	13410.00	280.66	0.00	1047.50	14738.16	14904.70
177	0	55020.30	2951.77	0.00	0.00	57972.07	57955.13
178	0	1303.80	0.00	0.00	0.00	1303.80	1313.87
179	1	13379.10	306.89	2417.80	3670.80	17356.79	18459.50
180	0	1281.50	0.00	0.00	0.00	1281.50	1313.87
181	0	13901.50	0.00	0.00	0.00	13901.50	14045.33
182	0	3256.50	343.31	0.00	0.00	3599.81	3955.88
183	0	1281.50	0.00	0.00	0.00	1281.50	1289.07
184	0	1281.50	359.67	0.00	0.00	1641.17	1648.74
191	0	1303.80	0.00	0.00	0.00	1303.80	1313.87

STATISTICS IN MEDICINE
Statist. Med. **18**, 3245–3262 (1999)

CONSTRUCTING CONFIDENCE INTERVALS FOR COST-EFFECTIVENESS RATIOS: AN EVALUATION OF PARAMETRIC AND NON-PARAMETRIC TECHNIQUES USING MONTE CARLO SIMULATION

ANDREW H. BRIGGS[1]*, CHRISTOPHER Z. MOONEY[2] AND DAVID E. WONDERLING[3]

[1] *Health Economics Research Centre, Oxford Institute of Health Sciences and Nuffield College, University of Oxford, U.K.*
[2] *Department of Political Science, West Virginia University, U.S.A*
[3] *Health Promotion Sciences Unit, London School of Hygiene and Tropical Medicine, U.K.*

SUMMARY

The statistic of interest in most health economic evaluations is the incremental cost-effectiveness ratio. Since the variance of a ratio estimator is intractable, the health economics literature has suggested a number of alternative approaches to estimating confidence intervals for the cost-effectiveness ratio. In this paper, Monte Carlo simulation techniques are employed to address the question of which of the proposed methods is most appropriate. By repeatedly sampling from a known distribution and applying the different methods of confidence interval estimation, it is possible to calculate the coverage properties of each method to see if these correspond to the chosen confidence level. As the results of a single Monte Carlo experiment would be valid only for that particular set of circumstances, a series of experiments was conducted in order to examine the performance of the different methods under a variety of conditions relating to the sample size, the coefficient of variation of the numerator and denominator of the ratio, and the covariance between costs and effects in the underlying data. Response surface analysis was used to analyse the results and substantial differences between the different methods of confidence interval estimation were identified. The methods, both parametric and non-parametric, which assume a normal sampling distribution performed poorly, as did the approach based on simply combining the separate intervals on costs and effects. The choice of method for confidence interval estimation can lead to large differences in the estimated confidence limits for cost-effectiveness ratios. The importance of such differences is an empirical question and will depend to a large extent on the role of hypothesis testing in economic appraisal. However, where it is suspected that the sampling distribution is skewed, normal approximation methods produce particularly poor results and should be avoided. Copyright © 1999 John Wiley & Sons, Ltd.

1. INTRODUCTION

The purpose of the economic appraisal of health care interventions is to inform public health decision makers of the relative value for money (or cost-effectiveness) of funding alternative interventions. Where the intervention in question generates improved health outcomes for

* Correspondence to: Andrew Briggs, Health Economics Research Centre, Institute of Health Sciences, University of Oxford, Headington, Oxford OX3 7LF, U.K. E-mail: andrew.briggs@ihs.ox.ac.uk

Contract/grant sponsor: U.K. Department of Health
Contract/grant sponsor: Nuffield College Goodhart fund
Contract/grant sponsor: Office of Health Economics

CCC 0277–6715/99/233245–18$17.50

patients, but at increased overall cost, the appropriate summary measure of cost-effectiveness is the incremental cost-effectiveness ratio (ICER). The ICER measures the additional cost of one intervention over another (say treatment A and treatment B) per unit difference in effectiveness. Where data have been obtained from two samples of patients receiving the different treatments, the ICER is calculated as in equation (1):

$$\hat{R} = \frac{\bar{C}_A - \bar{C}_B}{\bar{E}_A - \bar{E}_B} = \frac{\Delta\bar{C}}{\Delta\bar{E}} \tag{1}$$

where $\bar{C}_A$ and $\bar{E}_A$ are the mean costs and effects for the sample receiving treatment A and $\bar{C}_B$ and $\bar{E}_B$ are the mean costs and effects for the sample receiving treatment B.

When sample data on costs and effects are available it is natural to consider the use of statistical techniques to calculate confidence intervals around such point estimates. Unfortunately, the calculation of confidence intervals for a ratio is far from straightforward since the probability of obtaining a zero or near zero value on the denominator of the ratio is non-negligible, which suggests that the moments of the ICER may be undefined. In practice, this is a very real problem since it is common for clinical trials to be designed to detect the smallest meaningful clinical difference between treatments and is likely to lead to a large number of studies showing differences in treatment effects which are close to zero.[1] Clearly, this presents a problem for the use of standard parametric statistical methods. Recent research has focused on parametric approximations to the confidence interval for the ICER.[2-5] In addition, several commentators have also proposed the non-parametric approach of bootstrapping as a method for estimating confidence intervals,[1,2,6,7] and this approach has been successfully demonstrated using clinical trial data.[5,8] In the face of all the possible methods, one question quickly surfaces. Which of these methods is the most appropriate?

In this paper we present the results of a Monte Carlo simulation exercise designed to evaluate the alternative methods of calculating confidence intervals for the ICER statistic, under a variety of different conditions. The experiments require a massive number of iterations which, even a few years ago, would have put this exercise beyond our reach, but which is now possible thanks to the increasing power of personal computers.

2. METHODS

This section is split into three. The first two parts describe in some detail the alternative parametric and non-parametric methods that have been proposed for estimating the confidence limits for the ICER. The third part of this section presents the overall Monte Carlo simulation experiment designed to evaluate each of the methods.

2.1. Parametric approaches to estimating the ICER confidence interval

Three main methods, based on the parametric approach for calculating confidence intervals for an ICER, have recently appeared in the literature.[2-5] Each of these methods is explored in turn, highlighting the assumptions on which it is based.

2.1.1. The confidence box approach

A number of commentators have advocated the cost-effectiveness plane (CE plane) for presenting the results of economic evaluation and for aiding policy decisions.[9,10] O'Brien and colleagues

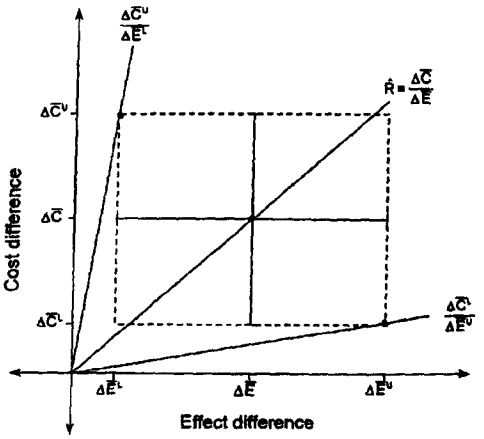


Figure 1. Confidence limits on the cost-effectiveness plane and the 'confidence box' approach to estimating confidence limits for the ICER

showed how the CE plane could also be used to present the confidence limits for the estimate of incremental cost-effectiveness.[2] Figure 1, which is based on the representation by O'Brien and colleagues, shows the results of a hypothetical prospective economic evaluation on the CE plane. The difference in effect between two therapies is shown on the horizontal axis with mean effect difference $\Delta\bar{E}$ and upper and lower confidence limits for the effect difference ($\Delta\bar{E}^{U}$ and $\Delta\bar{E}^{L}$) represented by the horizontal 'I' bar. Similarly, the difference in cost between two therapies is shown on the vertical axis with mean cost difference $\Delta\bar{C}$, and upper and lower confidence limits for the cost difference ($\Delta\bar{C}^{U}$ and $\Delta\bar{C}^{L}$) represented by the vertical 'I' bar. These 'I' bars intersect at point ($\Delta\bar{E}$, $\Delta\bar{C}$), hence the ray that connects this point of intersection to the origin has a slope equal to the value of the ICER. O'Brien and colleagues argue that combining the limits of the confidence intervals for costs and effects separately gives natural best and worst case limits on the ratio; that is, the upper limit of the cost difference over the lower limit of the effect difference ($\Delta\bar{C}^{U}/\Delta\bar{E}^{L}$) gives the highest values of the ratio (worst case) and the lower limit of costs divided by the upper limit of effects ($\Delta\bar{C}^{L}/\Delta\bar{E}^{U}$) gives the lowest (best) value of the ratio.

2.1.2. The Taylor series approximation

Rather than use these extreme limits, which are likely to overestimate the true interval, O'Brien and colleagues argue that it is possible to use the Taylor series approximation of the variance of a function of two random variables to estimate the variance of a ratio.[2] The advantage of this method is that it accounts for the covariance between the numerator and denominator. Having approximated the variance of the ICER statistic in this way, assuming the sampling distribution of the ICER to be normal allows the confidence interval to be estimated in the traditional manner.

The Taylor approximation shows that where y is a function of two random variables x_1 and x_2, the variance of y can be expressed in terms of the partial derivatives of y with respect to x_1 and x_2, weighted by the variances and covariance of x_1 and x_2. The Taylor series formula is

$$\mathrm{var}(y) \approx \left(\frac{\partial y}{\partial x_1}\right)^2 \mathrm{var}(x_1) + \left(\frac{\partial y}{\partial x_2}\right)^2 \mathrm{var}(x_2) + 2\left(\frac{\partial y}{\partial x_1}\right)\left(\frac{\partial y}{\partial x_2}\right)\mathrm{cov}(x_1, x_2). \quad (2)$$

Equation (2) can now be solved for the case of the ICER presented in equation (1) by substituting $\Delta\bar{C}$ for x_1 and $\Delta\bar{E}$ for x_2.* Hence equation (3) gives the Taylor series approximation of the variance of the ratio estimator, using the sample estimates of the means and variances (since by definition, the population values cannot be observed):

$$\text{var}(\hat{R}) \approx \frac{1}{\Delta\bar{E}^2}\text{var}(\Delta\bar{C}) + \frac{\Delta\bar{C}^2}{\Delta\bar{E}^4}\text{var}(\Delta\bar{E}) - 2\frac{\Delta\bar{C}}{\Delta\bar{E}^3}\text{cov}(\Delta\bar{C}, \Delta\bar{E}). \tag{3}$$

Factoring $\hat{R}^2 = \Delta\bar{C}^2/\Delta\bar{E}^2$ from the right-hand side simplifies (3) to

$$\text{var}(\hat{R}) \approx \hat{R}^2\left[\frac{\text{var}(\Delta\bar{C})}{\Delta\bar{C}^2} + \frac{\text{var}(\Delta\bar{E})}{\Delta\bar{E}^2} - 2\frac{\text{cov}(\Delta\bar{C}, \Delta\bar{E})}{\Delta\bar{C}\Delta\bar{E}}\right].$$

Noting that the coefficient of variation for a random variable x is defined $\text{cv}(x) = \sqrt{\text{var}(x)}/\bar{x}$ and that the correlation coefficient between two random variables x and y is defined $\rho_{xy} = \text{cov}(x, y)/\sqrt{\{\text{var}(x)\,\text{var}(y)\}}$ further simplifies the exposition:

$$\text{var}(\hat{R}) \approx \hat{R}^2[\text{cv}(\Delta\bar{C})^2 + \text{cv}(\Delta\bar{E})^2 - 2\rho\text{cv}(\Delta\bar{C})\text{cv}(\Delta\bar{E})]. \tag{4}$$

Employing standard parametric assumptions gives the confidence interval as

$$(\hat{R} - z_{\alpha/2}\sqrt{\text{var}(\hat{R})}, \quad \hat{R} + z_{\alpha/2}\sqrt{\text{var}(\hat{R})})$$

where $z_{\alpha/2} = \Phi^{-1}[1 - \alpha/2]$, Φ^{-1} is the inverse of the cumulative distribution of the standard normal function, and $100(1 - \alpha)$ per cent is the confidence level.

O'Brien and colleagues recognize that although the assumption of a normal distribution may be justified in the case of large samples, it is unlikely that the distribution of a ratio will follow a well-behaved distribution in general.[2] However, even if samples are large, the distribution is likely to be skewed where the coefficient of variation of the denominator of the ICER (effect difference) is high.[5]

2.1.3. Fieller's method

An alternative method of calculating confidence intervals around ratios has been described by Fieller.[11] This approach has been advocated for use in calculating confidence intervals around ICERs by both Willan and O'Brien[3] and Chaudhary and Stearns.[5] The method is described in general terms by Cochran.[12]

The advantage of Fieller's method over the Taylor series expansion is that it takes into account the skew of the ratio estimator. The method assumes that the numerator and denominator of the ratio follow a joint normal distribution function such that (in the case of the ICER) $\Delta\bar{C} - R\Delta\bar{E}$ is normally distributed. Hence, dividing through by the standard deviation equation (5) follows the standard normal distribution:

$$\frac{\Delta\bar{C} - R\Delta\bar{E}}{\sqrt{\{\text{var}(\Delta\bar{C}) + R^2\,\text{var}(\Delta\bar{E}) - 2R\,\text{cov}(\Delta\bar{C}, \Delta\bar{E})\}}} \sim \text{N}(0, 1). \tag{5}$$

* The partial derivatives of the ICER with respect to $\Delta\bar{C}$ and $\Delta\bar{E}$ are $1/\Delta\bar{E}$ and $\Delta\bar{C}/\Delta\bar{E}^2$, respectively.

Setting this expression equal to $z_{\alpha/2}$ and rearranging gives the following quadratic equation in R (using the simplified notation introduced in equation (4)):

$$R^2[1 - z^2_{\alpha/2}\,cv(\Delta\bar{E})^2] - 2R\hat{R}[1 - z^2_{\alpha/2}\,\rho cv(\Delta\bar{E})cv(\Delta\bar{C})] + \hat{R}^2[1 - z^2_{\alpha/2}cv(\Delta\bar{C})] = 0 \qquad (6)$$

where $\hat{R}$ is defined from equation (1).

Solving equation (6) for R using the standard quadratic formula* gives the confidence interval as

$$\hat{R}\left[\frac{1 - z^2_{\alpha/2}\,\rho cv(\Delta\bar{C})cv(\Delta\bar{E})}{1 - z^2_{\alpha/2}cv(\Delta\bar{E})^2}\right]$$

$$\pm z_{\alpha/2}\hat{R}\left[\frac{\sqrt{\{cv(\Delta\bar{C})^2 + cv(\Delta\bar{E})^2 - 2\rho(\Delta\bar{C})cv(\Delta\bar{E}) - z^2_{\alpha/2}(cv(\Delta\bar{C})^2cv(\Delta\bar{E})^2 - \rho^2cv(\Delta\bar{C})^2cv(\Delta\bar{E})^2)\}}}{1 - z^2_{\alpha/2}cv(\Delta\bar{E})^2}\right].$$

Where the sampling distribution of the ICER is skewed, this confidence interval will not be symmetrically positioned around the point estimate. This method has been criticized on the grounds that the assumption of joint normality may be hard to justify where sample sizes are small.[5]

2.2. Bootstrap approaches to estimating the ICER confidence interval

Given the unknown nature of the ICER's sampling distribution, there is reason to be cautious of the parametric approaches to confidence interval estimation. A number of commentators have suggested the non-parametric approach of bootstrapping as a possible method of estimating confidence limits for the ICER.[1,2,6,7] The advantage of such intervals is that they do not depend on parametric assumptions concerning the sampling distribution of the ICER.

The bootstrap approach for the simple one sample case is straightforward. Suppose a particular population has a real but unobserved probability distribution F from which a random sample x of n observations is taken, and the statistic of interest $s(\mathbf{x})$ is calculated. The concern of inferential statistics is to make statements about the population parameter θ based on the sample drawn from that population. In the 'bootstrap world', the observed random sample x is treated as the empirical estimate of F by weighting each observation in x by the probability $1/n$. Successive random samples of size n are then drawn from x *with replacement*† to give the bootstrap samples (re-samples from the original sample). The statistic of interest is calculated for each of these samples and these bootstrap replicates of the original statistic make up the empirical estimate of the sampling distribution for that statistic. This estimated sampling distribution can be used in a variety of ways to construct confidence intervals.

In principle, the bootstrap estimate of the ICER sampling distribution can be obtained in a very similar way to that of the simple one sample case. However, since the ICER is estimated on the basis of four estimators from two samples (equation (1)) care must be taken to bootstrap each sample appropriately. For data structures which are more complicated than a one sample structure, Efron and Tibshirani advocate that the bootstrap mechanism for the observed data

* The solution formula for a quadratic equation of the form $ax^2 + bx + c = 0$ is $-b \pm (\sqrt{(b^2 - 4ac)})/2a$.

† Clearly, sampling from x without replacement would simply yield x itself. Hence it is the sampling with replacement which provides the variability through the chance that some observations will appear in the bootstrap sample more than once while others will be omitted altogether.

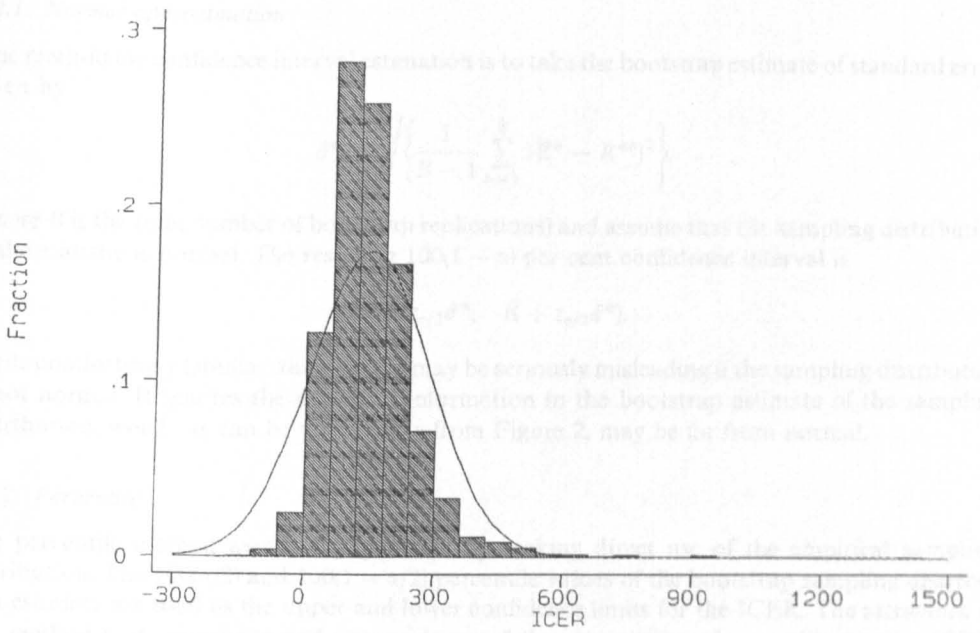


Figure 2. Bootstrap estimation of the sampling distribution of the ICER calculated from clinical trial data (overlaid is a normal distribution with the same mean and variance)

mirror the mechanism by which those original data were obtained.[13] In the case of the ICER, where data on resource use and outcome exists for two groups of patients of size n_A and n_B receiving treatments A and B, respectively, this will involve a three-stage process:

1. Sample with replacement n_A cost/effect pairs from the sample of patients who received treatment A and calculate the bootstrap estimates $\bar{C}_A^*$ and $\bar{E}_A^*$ for the bootstrap sample.
2. Sample with replacement n_B cost/effect pairs from the sample of patients receiving treatment B and calculate the bootstrap estimates $\bar{C}_B^*$ and $\bar{E}_B^*$ for the bootstrap sample.
3. Calculate the bootstrap replicate of the ICER given by the equation

$$R^* = \frac{\bar{C}_A^* - \bar{C}_B^*}{\bar{E}_A^* - \bar{E}_B^*} = \frac{\Delta\bar{C}^*}{\Delta\bar{E}^*}. \tag{7}$$

Repeating this three-stage process many times gives a vector of bootstrap estimates, which is an empirical estimate of the sampling distribution of the ICER statistic. For example, the histogram in Figure 2 shows the estimated sampling distribution from a previously reported study which used the bootstrap to estimate the sampling distribution of the ICER calculated from data generated by an economic evaluation conducted alongside a clinical trial.[8]

Once the sampling distribution of the ICER has been estimated in this way, several approaches exist to estimate confidence limits using the bootstrap estimate of the sampling distribution.

2.2.1. Normal approximation

One method for confidence interval estimation is to take the bootstrap estimate of standard error, given by

$$\hat{\sigma}^* = \sqrt{\left\{\frac{1}{B-1}\sum_{b=1}^{B}(\bar{R}^* - R^{*b})^2\right\}} \tag{8}$$

(where B is the total number of bootstrap replications) and assume that the sampling distribution of the statistic is normal. The resulting $100(1-\alpha)$ per cent confidence interval is

$$(\hat{R} - z_{\alpha/2}\hat{\sigma}^*, \quad \hat{R} + z_{\alpha/2}\hat{\sigma}^*).$$

While comfortingly familiar, this method may be seriously misleading if the sampling distribution is not normal. It ignores the wealth of information in the bootstrap estimate of the sampling distribution, which, as can be seen clearly from Figure 2, may be far from normal.

2.2.2. Percentile

The percentile method avoids this problem by making direct use of the empirical sampling distribution. The $100(\alpha/2)$ and $100(1-\alpha/2)$ percentile values of the bootstrap sampling distribution estimate are used as the upper and lower confidence limits for the ICER. The attraction of this method is its simplicity and its avoidance of the assumption of normality for the ICER. However, it has received considerable criticism from some commentators; for example, Hall (reference 14, p. 36) describes the percentile method as equivalent to '... looking up the wrong statistical tables backwards'.[14] That is, skewed estimation can cause trouble for the percentile method. In particular, in this context, the percentile method assumes that the bootstrap replicates of the ICER are unbiased, whereas it is known that ratio estimators are biased and that bootstrap replicates will magnify the bias of the sample estimate.[15]

2.2.3. Bias-corrected and accelerated

Efron[16] suggests a modification of the percentile method, which seeks to adjust for the bias and skew of the sampling distribution. This is the bias-corrected and accelerated (BCa) percentile method, which involves algebraic adjustments to the percentiles selected to serve as the confidence interval endpoints. The adjusted percentiles are given by

$$\alpha_1 = \Phi\left(\hat{z} + \frac{\hat{z} + z_{\alpha/2}}{1 - \hat{a}(\hat{z} + z_{\alpha/2})}\right)$$

$$\alpha_2 = \Phi\left(\hat{z} + \frac{\hat{z} + z_{(1-\alpha/2)}}{1 - \hat{a}(\hat{z} + z_{(1-\alpha/2)})}\right) \tag{9}$$

where $\Phi(.)$ is the standard normal cumulative distribution function and z_α is the 100α percentile point of the standard normal distribution. Two adjustments to the percentiles are incorporated into equation (9): $\hat{z}$ adjusts the sampling distribution for the bias of the estimator, while $\hat{a}$ adjusts for the skew of the sampling distribution. Setting $\hat{a} = 0$ yields the adjustment for bias on the percentiles chosen to serve as endpoints, and is equivalent to the bias-corrected method

advocated by Chaudhary and Stearns:[5]

$$\alpha_1 = \Phi(2\hat{z} + z_{\alpha/2})$$
$$\alpha_2 = \Phi(2\hat{z} + z_{(1-\alpha/2)}). \tag{10}$$

This bias correction, $\hat{z}$, is given by $\hat{z} = \Phi^{-1}(Q)$ where Q is the proportion of bootstrap replicates which are less than the sample estimate, $\hat{R}$. Therefore, if the bootstrap sampling distribution has median $\hat{R}$, $Q = 0{\cdot}5$ which gives $\hat{z} = 0$ and (in the absence of a skew adjustment) the percentiles from equation (10) correspond to those from the straightforward percentile method. However, where the sampling distribution is not centred on $\hat{R}$ a correction is made for this bias. Notice that the non-linear relationship between the z-score and its probability results in the percentile end points being shifted at unequal rates. It is also worth noting that the bias correction adjustment of the BCa method, while not employing distributional assumptions concerning the sampling distribution of the ICER itself, does make use of parametric assumptions concerning the distribution of the observed bias. This reliance on parametric assumptions has been cited as a potential weakness of the BCa method.[17]

The acceleration constant adjusts for the skew of the sampling distribution. Efron and Tibshirani suggest using a jack-knife estimate for $\hat{a}$:[13]

$$\hat{a}^{**} = \frac{\sum_{i=1}^{n}(\bar{R}^{**} - \hat{R}_i^{**})^3}{6\left[\sum_{i=1}^{n}(\bar{R}^{**} - \hat{R}_i^{**})^2\right]^{3/2}} \tag{11}$$

where $\hat{R}_i^{**}$ is the jack-knife replicate of the ICER with the ith observation removed, $\bar{R}^{**} = \sum \hat{R}_i^{**}/n$ for $i = 1$ to n and $n = n_C + n_T$. In terms of the adjustments to the percentiles given in equation (9), in the absence of a bias correction adjustment, the skew adjustment is given by

$$\alpha_1 = \Phi\left(\frac{z_{\alpha/2}}{1 - \hat{a}z_{\alpha/2}}\right)$$
$$\alpha_2 = \Phi\left(\frac{z_{(1-\alpha/2)}}{1 - \hat{a}z_{(1-\alpha/2)}}\right). \tag{12}$$

Equation (11) shows that if the sampling distribution is symmetric, $\hat{a} = 0$ and equation (12) shows that no adjustment to the percentile interval endpoints is made.

2.2.4. *Parametric bootstrap*

Efron and Tibshirani outline a simulation-based method of confidence interval estimation that they refer to as a parametric bootstrap approach.[13] Notice that from equation (1), the difference in costs on the numerator and the difference in effects on the denominator of the ICER are both simply the difference between two normally distributed variables, the two sample means* The difference of two means is also normally distributed. The parametric bootstrap approach involves using this property of the distribution of the numerator and denominator in combination with the observed means, variance and covariance to estimate the parameters of the sampling distribution of the cost and effect differences. Sampling from each of these two distributions, while allowing for

* They are normally distributed if the sample sizes are large enough to invoke the central limit theorem or if both costs and effects are normally distributed.

the estimated covariance between them, gives an estimate of the ICER. Repeating this process many times generates an empirical estimate of the sampling distribution of the ICER. The $100(\alpha/2)$ and $100(1-\alpha/2)$ percentiles of this estimated distribution are used as estimates for the upper and lower limits of the confidence interval, as with the percentile method.

2.3. The Monte Carlo simulation experiments

A simulation experiment was designed to test the coverage properties of each method for calculating confidence intervals in terms of the percentage number of times the true parameter falls outside the interval. Recall that a precise $100(1-\alpha)$ per cent confidence interval will contain the true population parameter $100(1-\alpha)$ per cent of the time in repeated sampling. Therefore, the expectation is that in 100α per cent of samples, the true population parameter lies outside of the interval. In deciding the levels of power and significance to accept, analysts trade off between type I and type II errors. If $\hat{\alpha}$, the observed proportion of Monte Carlo trials where the true population parameter lies outside of the interval, is greater than α, too many type I errors are committed. If $\hat{\alpha}$ is less than α, too many type II errors are committed. Clearly, if an analyst has specified an acceptable rate of error in advance, the method employed should deliver that chosen rate of error.

The Monte Carlo experiments employed the same population parameter values for the average costs and average effects of two hypothetical treatments A and B as those used in the experiments conducted by Wakker and Klaassen.[4] The population mean cost for the group receiving treatment A was set at 40,000 and for group B was set at 30,000; the population mean effects for groups A and B were set as 60 and 50, respectively. Hence the population value of the ICER can be calculated as

$$R = \frac{C_A - C_B}{E_A - E_B} = \frac{40{,}000 - 30{,}000}{60 - 50} = \frac{10{,}000}{10} = 1000. \tag{13}$$

However, in the experiments conducted by Wakker and Klaassen, the standard deviations specified for the population parameters were unrealistically low.[4] Recall that the coefficient of variation for a random variable x is defined as $\mathrm{cv}(x) = \sqrt{\mathrm{var}(x)}/\bar{x}$. Employing the standard deviations and population values specified by Wakker and Klaassen[4] suggests that the average observed coefficient of variation on the numerator of the ratio (the difference in costs) in their simulation experiments was 0·12, while the average observed coefficient of variation of the denominator (effect difference) was approximately 0·02. Low coefficients of variation such as these are likely to give a sampling distribution for the ICER that is very close to a normal sampling distribution.[5] However, we believe this is unrealistic and that many economic evaluations will have much higher coefficients of variation on both the numerator and denominator of the ratio leading to sampling distributions which are significantly skewed. For example, the coefficients of variation for the original data on which Figure 2 is based were 0·55 for the numerator and 0·27 for the denominator. As Figure 2 shows, the estimated sampling distribution was far from normal. Hence the standard deviations of the individual population parameters employed in the Monte Carlo experiments were set such that they generated a range of specified levels of coefficient of variation in the numerator and denominator (details of these calculations are given in the Appendix).

The problem with a single Monte Carlo experiment is that it will be valid only for the chosen parameters and conditions set in that experiment. Hence we designed a series of experiments which systematically varied the underlying conditions most crucial to the shape of the ICER

Table I. Overall performance of the different methods across the 480 experiments

Confidence interval	Lower alpha	Upper alpha	Overall alpha	Low error	Upper error	Overall error
Taylor	0·0051	0·0615	0·0665	−0·0199	0·0365	0·0165
Fieller	0·0205	0·0139	0·0524	−0·0045	0·0069	0·0024
Confidence box	0·0019	0·0047	0·0066	−0·0231	−0·0203	−0·0434
Norm approx	0·0047	0·0507	0·0554	−0·0203	0·0257	0·0054
Percentile	0·0185	0·0376	0·0561	−0·0066	0·0216	0·0061
BCa	0·0229	0·0364	0·0593	−0·0021	0·0114	0·0093
BC	*0·0252	0·0416	0·0668	*0·0002	0·0166	0·0168
Paraboot	0·0155	0·0342	*0·0497	−0·0095	0·0092	*−0·0003

* Non-significant at t-ratio < 2, employing an estimated standard error of a proportion of $\text{se}(p) = \sqrt{\{p(1-p)/n\}} = 0{\cdot}0007$ (for lower/upper alpha/error) and 0·0010 (for overall alpha/error)

sampling distribution. Five different correlation coefficients for the covariance between the costs and effects in the two groups were set: −0·90; −0·45; 0; 0·45, and 0·90. Coefficients of variation for the numerator and denominator were independently specified as 10, 20, 30 and 40 per cent. Six sample sizes were tested: 10; 30; 50; 60; 80, and 100. Population cost and effect data are rarely normally distributed; in particular, cost data is often significantly skewed. Hence, we set the underlying cost and effect data for groups A and B generated in the Monte Carlo experiments to be log-normally distributed.

Each experiment involved taking a random sample of values from one of the specified populations described above. On the basis of the values obtained in these samples, confidence intervals were calculated by each of the seven methods described in Sections 2.1 and 2.2. In addition, the straightforward bias-corrected (BC) bootstrap interval, as employed by Chaudhary and Stearns,[5] was estimated by simply ignoring the accelerator adjustment described in Section 2.2.3. The estimated intervals were then compared to the true ICER from equation (13). Where the true value lay outside of the calculated interval, this result was recorded. This process was repeated 1000 times for each experiment. Hence the number of times the true ICER lay outside the interval divided by the 1000 simulations was the estimated alpha level for that experiment. The upper alpha level recorded the number of times the true ICER lay above the interval, the lower alpha recorded the number of times the true ICER lay below the interval, and the overall alpha was the addition of the upper and lower alphas. Varying all of the conditions above represents 480 different experiments (5 correlation coefficients × 4 coefficients of variation for the numerator × 4 coefficients of variation for the denominator × 6 sample sizes) for which eight confidence intervals were calculated, giving a total of 3840 data points.

3. RESULTS

The overall results across the 480 experiments are presented in Table I. For each of the eight methods the estimated upper, lower and overall alpha rates are shown. To aid interpretation 'error rates' are also shown. These are simply the value of $(\hat{\alpha} - \alpha)$, the estimated value of alpha less the nominal value of alpha chosen for the experiments. The nominal value of alpha appropriate for the upper and lower results is 0·025 and for the overall results is 0·05. Each of the estimated

alpha/error rates was tested for significance using the binomial approximation for the standard error of a proportion.* All were significantly different from the nominal levels except for the lower estimate of the BC bootstrap method and the overall estimate for the parametric bootstrap method.

Care must be taken when interpreting the results of the overall error values. Systematic overestimation in one tail of the distribution combined with underestimation in the other tail can lead to a small overall error generated by large upper and lower errors of opposite sign. This effect is most noticeable in the parametric bootstrap method where the overall alpha is not significantly different from the nominal alpha level. However, it is clear that this is a result of the errors in the upper and lower alpha values cancelling each other out. Similar, although not so dramatic, effects are also apparent in the results for the Fieller and Taylor series methods and the normal approximation, percentile and BCa bootstrap methods.

On the basis of the results from Table I, Fieller's method appears to be performing most consistently across experiments, since it has the lowest upper error, the second lowest lower error and the lowest overall error. The BCa and parametric bootstrap performing best of the non-parametric methods and also outperforming the Taylor method and the box approach. However, since these results are based on summing across the 480 separate experiments, the results presented in Table I could potentially mask underlying variation in the estimated errors between experiments if overestimates in some experiments cancel out with underestimates in other experiments. These variations may be systematically related to experimental factors, which would have significant practical importance.

In order to analyse the effect of these experimental factors on the overall accuracy of the eight confidence interval methods, a technique known as response surface analysis was used.[18,19] This is a technique based on simple OLS regression employing dummy variables for each of the methods. Since the technique requires a single dependent variable we constructed a performance index based on the upper alpha value, defined as $-|\hat{\alpha}^U - \alpha^U|$. The negative sign was included for interpretative purposes – the greater the number, the better the performance. The upper alpha value was chosen partly due to the problem with the overall alpha value detailed above, but mainly due to the fact that economic analysts are more interested in deciding whether an observed ICER is below some threshold value used for decision making and might therefore be more interested in the upper alpha value being close to its chosen nominal level.[4]

The results of the response surface analysis are presented in Table II. The reference interval chosen was the Fieller method since it seemed to perform best from the results presented in Table I. The natural log of the sample size was chosen as an explanatory variable since it was hypothesized that performance would improve with sample size asymptotically to some limit. The seven confidence interval dummies are presented on separate rows, with the first row being the reference (Fieller) interval. The majority of the coefficients were significant at the standard levels, indicating that there are important differences between the Fieller interval and other methods across the different experiments.

The key to interpreting the results of the analysis lies in the sign of the coefficient; positive coefficients indicate an improvement in performance relative to Fieller's method for that variable and negative coefficients indicate worsening performance relative to the Fieller method. However,

* The binomial approximation for the standard error of a proportion p is given by $se(p) = \sqrt{\{p(1-p)/n\}}$ where n is the sample size.

Table II. Results of the response surface analysis: estimated coefficients*

Dummy variables	Intercept	Interactions			
		Log of sample size	Coefficient of variation for numerator	Coefficient of variation for denominator	Correlation
Ref(Fieller)	−0·0217†	0·0035†	−0·0106†	0·0108†	0·0013†
Taylor	−0·0159	0·0027†	0·0433	−0·1362†	0·0159†
Confidence box	0·0043†	−0·0041†	0·0088†	−0·0123†	−0·0089†
Norm approx	−0·0338†	0·0052†	0·0299†	−0·0487†	0·0142†
Percentile	−0·0257†	0·0051†	−0·0134†	0·0178†	0·0031†
BCa	−0·0198†	0·0044†	−0·0101†	0·0079†	0·0012†
BC	−0·0348†	0·0069†	−0·0200†	0·0192†	0·0056†
Para boot	−0·0147†	0·0031†	−0·0024	0·0056	0·0001

* Adjusted R^2 for the model = 0·74, n = 3840
† t − ratio > 2

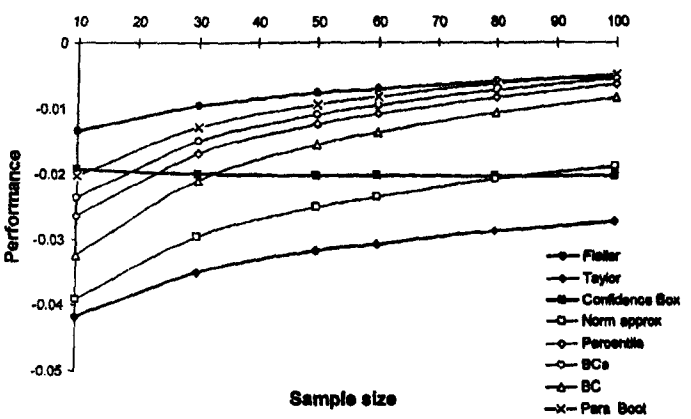


Figure 3. Predicted effect of sample size on the performance variable using results of the *response surface analysis* (coefficients of variation = 0·2, correlation = 0). Performance variable is defined as $-|\hat{\alpha}^U - \alpha^U|$

due to the different intercept values, it is not easy to see the relative performance of each method. In order to demonstrate better the relative performance, the results of the response surface analysis presented in Table II were used to generate predicted performance values for each of the methods. By holding three of the four quantitative variables constant, it was possible to examine the effect of the fourth on the performance of each method.

Figure 3 shows the predicted performance of each method for increasing sample size between 10 and 100, holding the coefficients of variation constant at 0·2 and the correlation coefficient constant at 0. The parametric methods are shown with the weightier lines and solid symbols. At low sample sizes, Fieller's method performs best and the Taylor series performs worst. The confidence box approach appears largely unaffected by sample size. All the other methods

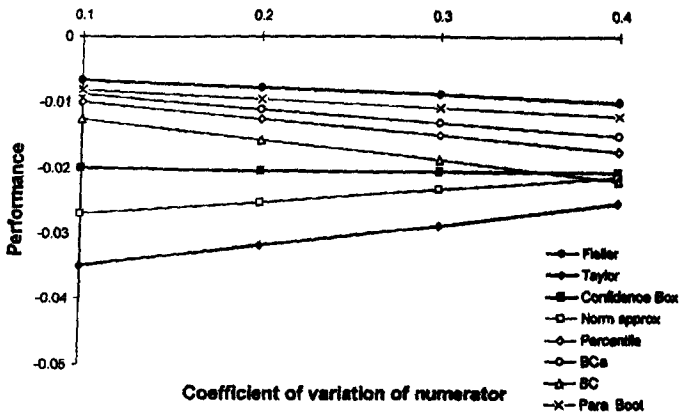


Figure 4. Predicted effect of coefficient of variation of the numerator on the performance variable using results of the *response surface analysis* (sample size = 50, coefficient of variation of the denominator = 0·2, correlation = 0). Performance variable is defined as $-|\hat{\alpha}^U - \alpha^U|$

improve with sample size, and by $n = 100$ there is little to choose between Fieller's method and the bootstrap methods with the exception of the normal approximation bootstrap method, which, like the Taylor series method, performs poorly.

Figure 4 shows the predicted performance of each method for values of the coefficient of variation of the numerator between 0·1 and 0·4, holding the coefficient of variation of the denominator constant at 0·2, the correlation coefficient constant at 0 and the sample size constant at 50. Again, Fieller's method performs best for all values of the coefficient of variation of the numerator and the Taylor series method performs worst. The normal approximation and Taylor series methods improve in performance as the coefficient of variation of the numerator increases while the performance of the other methods decrease, with the exception of the confidence box method which again appears largely unaffected by changes in the coefficient of variation of the numerator.

Figure 5 shows the predicted performance of each method for values of the coefficient of variation of the denominator between 0·1 and 0·4, holding the coefficient of variation of the numerator constant at 0·2, the correlation coefficient constant at 0 and the sample size constant at 50. A similar picture emerges in that Fieller's method performs best overall, Taylor series performs worst for all but the lowest coefficients of variation of the denominator and the confidence box method seems largely unaffected. This time, however, the performance of the Taylor series and normal approximation bootstrap methods worsen as the coefficient of variation of the denominator increases while the other methods improve in performance.

Figure 6 shows the predicted performance of each method for values of the correlation coefficient between $-0{\cdot}9$ and $+0{\cdot}9$, holding the coefficients of variation constant at 0·2 and the sample size constant at 50. Fieller's method again performs best and appears unaffected by variation in the correlation coefficient of the underlying data. All methods improve with increasing correlation with the exception of the confidence box method, which worsens dramatically as correlation increases. At the very highest correlation, the confidence box method performs worse than the Taylor series method.

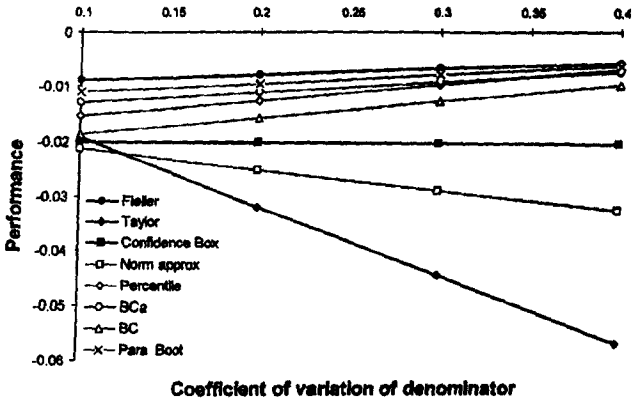


Figure 5. Predicted effect of coefficient of variation of the denominator on the performance variable using results of the *response surface analysis* (sample size = 50, coefficient of variation of the numerator = 0·2, correlation = 0). Performance variable is defined as $-|\hat{\alpha}^U - \alpha^U|$

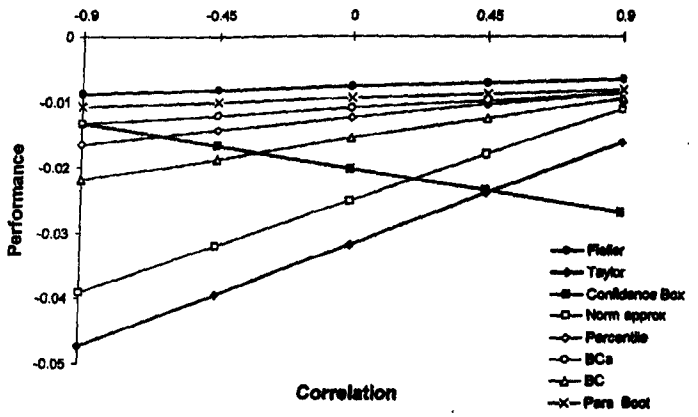


Figure 6. Predicted effect of correlation in the underlying data on the performance variable using results of the *response surface analysis* (sample size = 50, coefficients of variation = 0·2). Performance variable is defined as $-|\hat{\alpha}^U - \alpha^U|$

4. DISCUSSION AND CONCLUSIONS

The purpose of this paper was to compare a number of parametric approximations of the confidence limits around the incremental cost-effectiveness ratio with non-parametric bootstrapping methods. By devising a series of experiments that represented a realistic range of statistical conditions, we are able to make general conclusions about the relative performance of these approaches, and the factors affecting their relative performance.

No single method dominated (or was dominated by) all other methods across all of the experiments. However, as is shown by Figures 3–6, Fieller's method consistently performed well under a wide variety of assumptions, including small sample sizes, where its assumption of joint

normality between the cost and effect differences has been questioned.[5] Of the bootstrap methods, a clear pattern emerged in terms of the rank ordering of performance. The parametric bootstrap performed the best under most circumstances, closely followed by the BCa and then the straightforward percentile method. The normal approximation method performed most poorly of the bootstrap methods. It is clear that the 'accelerator' adjustment presented by Efron as a refinement to the straightforward bias corrected approach does improve the performance of the method. Although the term 'parametric bootstrap' has been adopted by Efron and Tibshirani,[13] this method is simply a straightforward Monte Carlo simulation of the numerator and denominator of the ratio on the basis of parametric assumptions and the observed means and variances of the data. To what extent this constitutes 'bootstrapping' as the term is commonly applied is an open question.

The predicted effect of increasing sample size is of particular interest. Both the parametric and bootstrap methods rely on asymptotics, and it appears from Figure 3 that the asymptotics of the parametric methods come in to play more quickly than those of the bootstrap methods. In many ways this is a surprising result since bootstrapping has often been linked to the analysis of small samples where standard parametric assumptions are thought to be violated.[17]

As the correlation between the cost and effect in the underlying data increased, the performance of the confidence box method worsened. This is due to the fact that the combination of limits for the confidence box approach is consistent with an assumption of perfect negative covariance between cost and effect. For all other methods, performance increased with increasing correlation. Although there was little to choose between the methods when correlation was high, in practical application it would be unusual to observe extremely high positive or negative correlations. In the data from which Figure 2 was generated, the correlation between cost and effect in the treatment arm of the trial was 0·19, while in the control arm of the trial it was $-0{\cdot}05$.[6] These figures translate into very little covariance between the numerator and denominator of the ICER.

The predicted effect of the coefficient of variation of the numerator and denominator was interesting in that each seemed to influence the methods in the opposite direction. The methods based on an assumption of the normal distribution worsened as the coefficient of variation of the denominator increased, but improved as the coefficient of variation of the numerator increased. For the other methods, the converse was true, with the exception of the confidence box method, which seemed largely unaffected by either coefficient of variation.

One very clear result from these experiments was the inadequacy of methods based principally on the assumption of a normal sampling distribution. Both the parametric based Taylor approximation method and the bootstrap normal approximation were consistently poor performers, although the bootstrap normal approximation seemed to outperform the Taylor series method in general. It is our belief that the sampling distribution of the ICER will almost certainly exhibit an element of skewness in most practical applications, which makes the normal distribution assumption rather limiting.

Recent reviews of economic evaluations have suggested that many authors present only point estimates of cost-effectiveness without any representation of the uncertainty associated with their estimates,[20,21] which suggests that any method of interval estimation is preferable to point estimates alone. However, we have shown that there are substantial differences in the accuracy of the methods advocated in the recent health economics literature. We believe that the nominal error rates accepted by analysts when calculating confidence intervals should be reflected by the actual rates of error that would occur in repeated application of the method. Of course, these error rates will only occur if in practice analysts begin to test hypotheses on the basis of the results

of prospective economic evaluation. We believe the time is now ripe for an analysis of the role of hypothesis testing in economic appraisal.

APPENDIX

This Appendix lays out the method for generating variances of the cost parameters; the same method also applies to effects. Suppose the underlying population cost parameters for treatments A and B are known to be C_A and C_B, respectively. Define the mean difference in cost between two groups sampled from these populations as $\Delta\bar{C} = \bar{C}_A - \bar{C}_B$. In terms of the Monte Carlo experiments, we want to set the coefficient of variation of the difference in costs, $\text{cv}(\Delta\bar{C})$, since it is this which is assumed (in tandem with the effect difference) to determine the shape of the ICER sampling distribution. Hence, the problem is to work backwards from this coefficient of variation to define values for the population cost variance for patients receiving treatments A or B, σ_A^2 and σ_B^2, which will generate the desired $\text{cv}(\Delta\bar{C})$.

We know that the coefficient of variation for the cost difference is defined as

$$\text{cv}(\Delta\bar{C}) = \sigma_{\Delta C}/\Delta\bar{C},$$

hence

$$\sigma_{\Delta C} = \Delta\bar{C}\,\text{cv}(\Delta\bar{C}). \tag{14}$$

Assuming that random samples of size n_A and n_B are sampled from the population for treatments A and B, respectively, then the treatment costs in each group should be independent. Thus it is possible to relate the variance of the cost difference to the variances of the underlying treatment and control group cost data:

$$\sigma_{\Delta C}^2 = \frac{\sigma_{CA}^2}{n_A} + \frac{\sigma_{CB}^2}{n_B}. \tag{15}$$

Combining equations (14) and (15) gives

$$[\text{cv}(\Delta C)\,\Delta\bar{C}]^2 = \frac{\sigma_{CA}^2}{n_A} + \frac{\sigma_{CB}^2}{n_B}$$

and rearranging

$$n_A n_B[\text{cv}(\Delta C)\,\Delta\bar{C}]^2 = n_B\sigma_{CA}^2 + n_A\sigma_{CB}^2. \tag{16}$$

Further suppose that the coefficients of variation of the underlying costs are the same, that is, $\sigma_{CA}/C_A = \sigma_{CB}/C_B$ or equivalently that

$$\sigma_{CA} = \sigma_{CB}\frac{C_A}{C_B}. \tag{17}$$

Combining equations (16) and (17) gives

$$n_A n_B[\text{cv}(\Delta C)\,\Delta\bar{C}]^2 = n_B\sigma_{CB}^2\left(\frac{C_A}{C_B}\right)^2 + n_A\sigma_{CB}^2$$

$$= \sigma_{CB}^2\left[n_A + n_B\left(\frac{C_A}{C_B}\right)^2\right]$$

and rearranging

$$\sigma^2_{CB} = n_A n_B \frac{[\mathrm{cv}(\Delta C)\Delta\bar{C}]^2}{n_A + n_B(C_A/C_B)^2}. \tag{18}$$

Clearly, σ^2_{CA} can now be calculated from equation (17).

In our experiments, the sample size in each group was the same, that is, $n_A = n_B = n$, therefore equation (18) simplifies to

$$\sigma^2_{CB} = n \frac{[\mathrm{cv}(\Delta C)\Delta\bar{C}]^2}{1 + (C_A/C_B)^2}.$$

ACKNOWLEDGEMENTS

Support of the U.K. Department of Health, the Nuffield College Goodhart fund (AB) and the Office of Health Economics (AB and DW) is gratefully acknowledged. Thanks must go to Dr. James Carpenter for helpful comments on an earlier draft, to Michael Wadley for help transcribing data from the results of the experiments and to the anonymous reviewers who provided valuable input to the final paper. Of course, the views expressed in this document are those of the authors alone, together with the responsibility for any errors.

REFERENCES

1. Mullahy, J. and Manning, W. 'Statistical issues in cost-effectiveness analysis', *in* Sloan, F. (ed.), *Valuing Health Care*, Cambridge University Press, Cambridge, 1994, pp. 149–84.
2. O'Brien B. J., Drummond, M. F., Labelle, R. J. and Willan, A. 'In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care', *Medical Care*, **32**, 150–163 (1994).
3. Willan, A. R. and O'Brien, B. J. 'Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem', *Health Economics*, **5**, 297–305 (1996).
4. Wakker, P. and Klaassen, M. 'Confidence intervals for cost-effectiveness ratios', *Health Economics*, **4**, 373–381 (1995).
5. Chaudhary, M. A. and Stearns, S. C. 'Estimating confidence intervals for cost-effectiveness ratios: an example from a randomized trial', *Statistics in Medicine*, **15**, 1447–1458 (1996).
6. Mullahy, J. 'What you don't know can't hurt you? Statistical issues and standards for medical technology evaluation', *Medical Care*, **34** (12 Suppl.), DS124–DS135 (1996).
7. Manning, W. G., Fryback, D. G. and Weinstein, M. C. 'Reflecting uncertainty in cost-effectiveness analysis', *in* Gold, M. R., Siegel, J. E., Russell, L. B. and Weinstein, M. C. (eds.), *Cost-effectiveness in Health and Medicine*, Oxford University Press, New York, 1996.
8. Briggs, A. H., Wonderling, D. E. and Mooney, C. Z. 'Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation', *Health Economics*, **6**, 327–340 (1997).
9. Anderson, J. P., Bush, J. W., Chen, M. and Dolenc, D. 'Policy space areas and properties of benefit-cost/utility analysis', *Journal of the American Medical Association*, **255**, 794–795 (1986).
10. Black, W. C. 'The CE plane: A graphic representation of cost-effectiveness', *Medical Decision Making*, **10**, 212–214 (1990).
11. Fieller, E. C. 'Some problems in interval estimation', *Journal of the Royal Statistical Society, Series B*, **16**, 175–183 (1954).
12. Cochran, W. G. *Sampling Techniques*, 3rd edn, Wiley, New York, 1977.
13. Efron, B. and Tibshirani, R. *An Introduction to the Bootstrap*, Chapman and Hall, New York, 1993.
14. Hall, P. *The Bootstrap and the Edgeworth Expansion*, Springer-Verlag, New York, 1992.
15. Stinnett, A. 'Adjusting for bias in C/E ratio estimates', *Health Economics*, **5**, 469–472 (1996).
16. Efron, B. 'Better bootstrap confidence intervals', *Journal of the American Statistical Association*, **82**, 171–200 (1987).

17. Mooney, C. Z. and Duval, R. D. *Bootstrapping: A Nonparametric Approach to Statistical Inference*, Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-095, Sage, Newbury Park, CA, 1993.
18. Hendry, D. F. 'Monte Carlo experimentation in econometrics', *in* Griliches, Z. and Intriligator, M. D. (eds), *Handbook of Econometrics*, vol. II, Elsevier, Amsterdam, 1994.
19. Mooney, C. Z. *Monte Carlo Simulation*, Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-116, Sage, Newbury Park, CA, 1997.
20. Udvarhelyi, S., Colditz, G. A. and Epatein, A. M. 'Cost-effectiveness and cost-benefit analyses in the medical literature: are the methods being used correctly?', *Annals of Internal Medicine*, **116**, 238–244 (1992).
21. Briggs, A. and Sculpher, M. 'Sensitivity analysis in economic evaluation: a review of published studies'. *Health Economics*, **4**, 355–371 (1995).

