

10

Statistical handling of data in economic analysis

Introduction

This chapter examines the use of data in economic analysis. The first part describes the statistical analysis of cost data. Before beginning this section you will need to understand the concepts covered in Chapter 3, and you will also need a basic understanding of statistical distributions (such as normal distributions), descriptive statistics (mean, median standard deviation and confidence intervals), inferential statistics (parametric tests such as t -tests and ANOVA tests, and non-parametric tests such as χ^2 tests).

The second part of the chapter describes the use of probabilistic sensitivity analysis to generate mean expected costs, outcomes and ICERs, and statistical measures of expected variance around the mean ICER. This method, also called *stochastic economic analysis*, or *probabilistic modelling*, allows estimation of the probability and extent to which uncertainty and variation in the data used affect the absolute and relative costs and outcomes. Before beginning this section, you will need to understand the concepts covered in Chapters 3, 4, 5 and 8. You will also need to understand the statistical concepts listed in the paragraph above.

Statistical analysis of costs

The shape of cost data distribution

When carrying out economic analysis, it is important to find out by using statistical tests whether there are differences in costs between different treatment alternatives. The objective of the statistical analysis is to test whether there are statistically significant differences in resource use and costs between the treatment groups. Cost is a continuous variable, so it would seem that the costs of treatment can be compared between groups using standard independent samples t -tests.

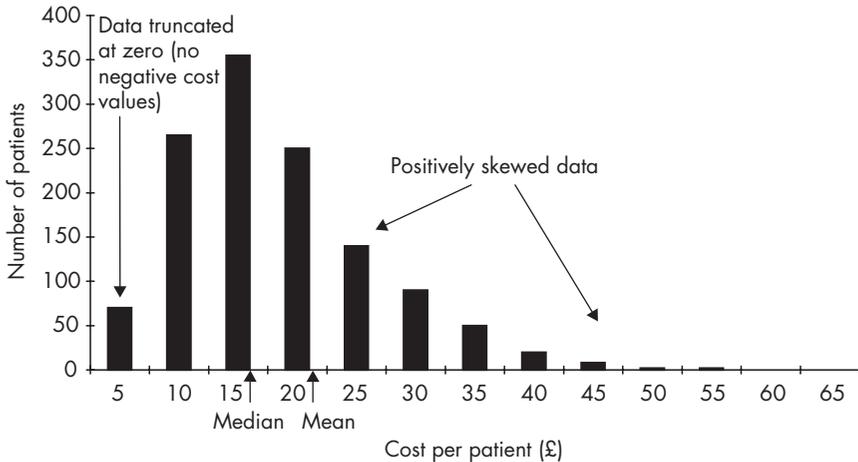


Figure 10.1 Distribution of cost data.

However, the distribution of cost data is typically *truncated* and *positively skewed*. This is caused by the presence of a relatively small number of patients with high treatment costs (patients developing complications after surgery, for example) and the absence of negative costs (Barber and Thompson, 1998). This can create a long right-handed tail in the distribution. A theoretical example of the distribution of patient-based cost data is shown in Figure 10.1. This shows the truncated and skewed distribution. It also shows that the median is lower than the mean (arithmetic mean). This is an indication of positively skewed data.

The mean is £21. This, however, is sensitive to outliers. The median is often used with skewed or non-normal data and where there are outliers. It is also used in non-parametric tests. The median for this dataset is £16. The median is useful because we can use it to describe a 'typical cost' for a patient (Thompson and Barber, 2000). However, using the median will not allow policy-makers to determine the total cost of treatment for a group of patients. For this, the mean is required because total cost for a group is the mean cost multiplied by number of patients in the group.

Furthermore, summarising the distribution of costs can be problematic. Because of skewed data, standard deviation alone is not an ideal way to present the spread of costs between individuals. Use of the range may give a distorted picture of variability, and so it is considered more useful to present the interquartile range.

This is not the same type of data distribution as the normal distribution upon which standard parametric independent samples *t*-tests are based. Standard parametric statistical tests assume that observations are

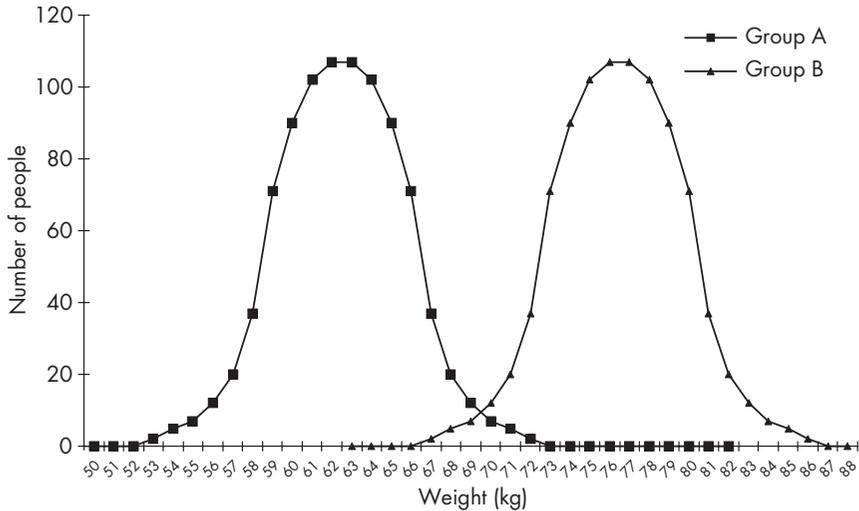


Figure 10.2 Determining the difference between two normally distributed data sets.

random samples from a normal distribution with the same variance. An example of this is shown in Figure 10.2. The distribution of weights of two groups of people have been plotted, and we can see that these data have normal distributions. Therefore, a *t*-test can be used to test whether the difference in weight between groups A and B is statistically significant.

When comparing two sets of cost data, the distributions are more likely to look like those in Figure 10.3.

When is a *t*-test the correct test to use?

We have said that the use of standard parametric statistical tests to look for differences between cost data may be inappropriate and the calculated means may be unreliable owing to the shape of distribution of cost data.

However, skewness may not be a problem for large samples, as the larger the sample size the more likely the central limit theorem will hold, and parametric assumptions will hold also. It has been reported that for samples larger than about 150, the *t*-test is generally robust (Barber and Thompson, 2000).

A published example of the appropriate use of *t*-tests for statistical analysis of cost data is an RCT of two different anaesthetic techniques in paediatric day surgery which obtained patient-based cost data (Elliott *et al.*, 2004). There were 159 and 163 patients, respectively, in each arm of the study. When sample distributions were examined, the skewness in

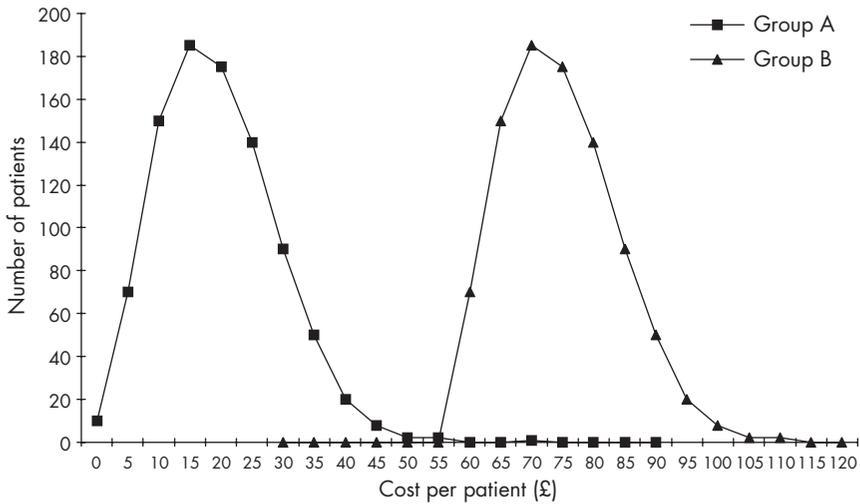


Figure 10.3 Determining the difference between two sets of cost data.

the samples was found to be sufficiently low to indicate normality for the sampling distribution of the mean. The *t*-test was therefore used for cost variables and a statistically significant difference was found.

Dealing with small sample sizes

Problems arise when analysing small samples. In summary, the *t*-test is robust for large samples but problems arise for small samples with skewness in the distribution. To judge whether skew in the data has important implications for the sampling of the mean, Briggs and Gray (1998) suggested that the skew in the sampling distribution of the mean (S_m) will be a factor:

$$S_m = S_s - \sqrt{n_s}$$

where S_s is the skewness of the original sample and n_s is the number of observations in the original sample.

There are three possible approaches to accommodate the problems of sample size and distribution in the analysis of cost data: non-parametric (rank sum) tests, transformation, and non-parametric bootstrapping. These methods are explained briefly here. For a more detailed handling of this subject, the reader is directed to Thompson and Barber (2000), Hart (2001), and Barber and Thompson (2000).

Non-parametric tests

For non-normal distributed data, non-parametric tests are recommended by conventional statistical guidelines. However, the non-parametric (rank sum) tests are better suited to hypothesis testing than to estimation. A p value is generated which shows the probabilities that the two samples are from the same distribution, and estimates of difference and confidence limits are based on assumptions that the shapes of the distributions are the same and only the location of the distribution is different (Hart, 2001). In addition, the estimates generated from rank sum tests effectively compare the difference in location between the two distributions, and thus compare medians. Therefore, these tests are inappropriate for cost data because economists are interested in the mean costs of treatment. The arithmetic mean is considered to be the most relevant measure for healthcare policy decisions, which should be based on information about the distribution of the costs of treating a patient group, as well as the average cost.

It is now generally accepted that using non-parametric tests may provide misleading conclusions (Thompson and Barber, 2000). However, many published studies of costs have 'revealed a lack of statistical awareness' by using these non-parametric tests (Barber and Thompson, 1998). It is important for the reader to examine statistical tests carefully in published studies.

Transformation

Data may be transformed to 'symmetrise' the distribution using natural log-transformation, square root transformations, or reciprocal transformation. This process reduces skewness and a normal distribution is approximated and similar size variances in the respective groups obtained. However, this means that geometric means are compared, and again arithmetic means cannot be compared. Positively skewed data such as costs will always produce a geometric mean that is lower than the arithmetic mean (Thompson and Barber, 2000). In addition, it is not possible to retransform cost differences to the original scale. This method has been recommended by some health economists (Rutten *et al.*, 1994) and used in some studies (Ellis *et al.*, 2000). However, it is now considered that these methods are inappropriate and are likely to produce misleading interpretation of results (Barber and Thompson, 1998, 2000; Thompson and Barber, 2000).

Non-parametric bootstrapping (Barber and Thompson, 2000; Briggs and Gray, 1998)

Non-parametric bootstrapping is a data-based simulation method for assessing statistical precision. Bootstrap simulation allows a comparison of arithmetic means without making assumptions about the cost distribution. Bootstrapping compares arithmetic means while avoiding distributional assumptions.

In reality, one sample is available and statistics of interest are calculated from that sample. Bootstrapping is based on how values of that statistic would vary if the sampling process could be repeated many times.

The validity of the bootstrapping approach is based on two assumptions. First, as the original sample size approaches the population size so the sample distribution tends towards the population distribution. Second, as the number of bootstrap replications approaches infinity, so the bootstrap estimate of the sampling distribution of a statistic approaches the true sampling distribution. In a bootstrap analysis, the observed sample is treated as an empirical distribution, a sample is taken from that distribution, and the statistic of interest is calculated.

Random values are selected from the original sample (size n_i) with replacement to yield a bootstrap sample of size n_i , and the statistic of interest is calculated. 'Replacement' means that once a random value has been used for the bootstrap resample, it is put back into the original sample. Therefore, a bootstrap sample may include the costs for some patients more than once, while excluding the costs for other patients. The process is repeated many times. Between 50 and 200 replications are necessary to provide an estimate of standard error (Briggs *et al.*, 1997). Usually at least 1000 reiterations are carried out. This simulation process creates a sample of bootstrapped means with a distribution. The mean and other parametric statistics may be calculated for the bootstrap distribution. Most spreadsheet and database packages can carry out bootstrapping very easily. This method of testing cost differences is now being used regularly and there are many published examples (Grieve *et al.*, 2003; Lambert *et al.*, 1998).

Bootstrapping can be used either for primary analysis or as a check on the robustness for using parametric tests, such as the t -test with non-normal data. This check was carried out on cost data in the UK Prospective Diabetes Study (UK Prospective Diabetes Study Group, 1998). Here, the bootstrapped confidence intervals were compared with parametric confidence intervals and found to be robust. This is partly due to the large sample size of 1148 patients.

Stochastic economic analysis

Statistical problems with ICERs

The results of an economic evaluation can be used to make decisions about how to allocate resources. In Chapter 8 we showed how uncertainty about cost and outcome data can affect how decisions are made. Therefore, it is important that we present uncertainty about the results of our economic evaluations – that is, about incremental cost-effectiveness ratios (ICERs). Effect sizes and cost differences are associated with uncertainty, and we use confidence intervals (usually 95%) to represent this. Any uncertainty in these parameters will lead to uncertainty in the ICER derived from them. Although derived from sampled data for both outcome and costs, the ICER is a ratio. In fact, it is a ratio of differences, so there is a finite chance that the effect size may be zero. In this case it will not be possible for that ICER to be generated, because dividing a number by zero necessarily produces a result of infinity. This causes a problem in standard statistical analysis of ICERs. Generating a 95% confidence interval for an ICER is not appropriate to test its precision, as the ratio of the two distributions does not necessarily have a finite mean, or, therefore, a finite variance (Barber and Thompson, 2000).

Methods for generating confidence intervals for ICERs

A number of approaches have been proposed for calculating confidence intervals, and Polsky *et al.* (1997) evaluated four methods: the Box method, the Taylor series method, the Fieller theorem method and the non-parametric bootstrap method. Confidence intervals for cost-effectiveness ratios resulting from the bootstrap method and the Fieller theorem method have been shown to be more dependably accurate than from the other two methods. It is beyond the scope of this book to describe all these methods in depth. For the interested reader, Briggs and Fenn (1998) also describe these four methods in more detail. For an example of the application of the Fieller theorem method in a published study see the UKPDS trial (UK Prospective Diabetes Study Group, 1998). Rather than bewilder the general reader and analyst, we confine ourselves to presenting the simplest, most appropriate method, which is the non-parametric bootstrap method. Therefore, the following section explains how this can be used for calculating quasi confidence intervals for ICERs.

Generating quasi confidence intervals for ICERs using the non-parametric bootstrap method

The non-parametric bootstrap method does not generate true statistical confidence intervals for an ICER, and so we refer to the intervals generated as *quasi confidence intervals*.

Generating bootstrapped ICERs is similar to generating bootstrapped costs, as described above. In this situation, the ICERs are generated with replacement in the following way:

1. Treatment group: sample with replacement n_t cost/effect pairs, where n_t is the sample size of the treatment group
2. Control group: sample with replacement n_c cost/effect pairs, where n_c is the sample size of the control group
3. Estimate the bootstrap ICER (Briggs *et al.*, 1997).

This method will produce a distribution of ICERs from which the mean can be generated. However, we cannot generate a true confidence interval.

Graphical representation of bootstrapped ICERs

In Chapter 5 the cost-effectiveness plane was introduced and explained. In the cost-effectiveness plane, the horizontal axis represents the difference in effectiveness between two interventions and the vertical axis represents the corresponding difference in costs. We can plot our bootstrapped sample of ICERs on this plane, which allows us to visualise the variation within the bootstrapped sample of ICERs. This is shown in Figure 10.4.

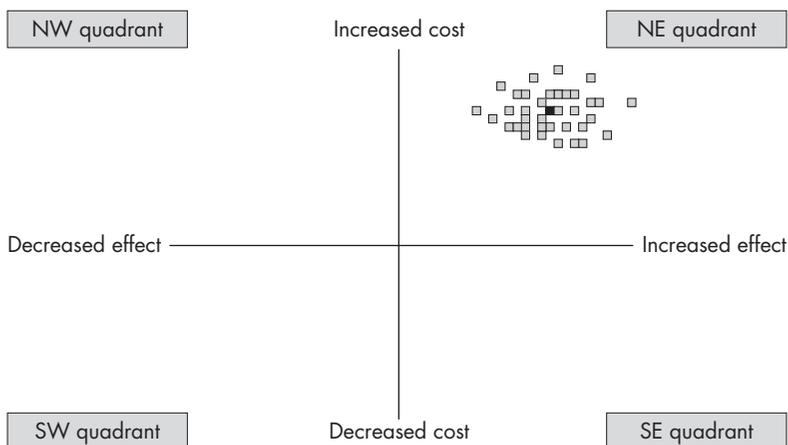


Figure 10.4 Cost-effectiveness plane showing bootstrapped ICERs for a treatment that is more effective and more costly than the comparator.

We can draw an ellipse around the bootstrapped sample of ICERs to show the distribution of data. This is shown in Figure 10.5. This ellipse includes all the ICERs, including the outliers.

We can also draw an ellipse around 95% of the bootstrapped ICERs. Within this ellipse we have 95% of the possible ICERs. An approximation of the 95% confidence interval is given by the rays from the origin of the CE plane that are tangential to the ellipse (Briggs and Fenn, 1998). This is shown in Figure 10.6.

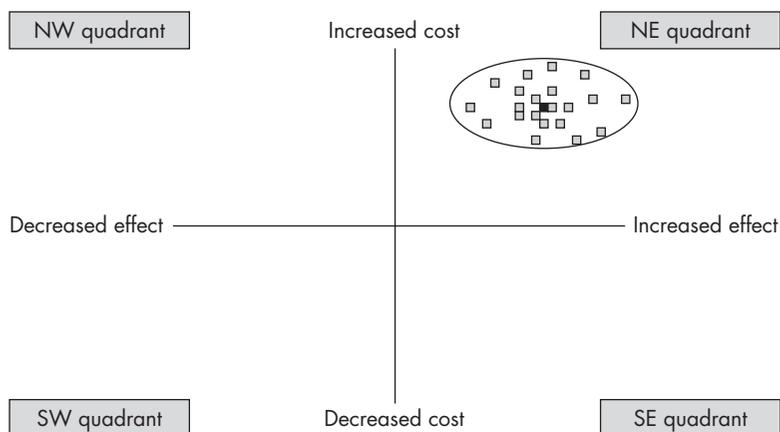


Figure 10.5 Cost-effectiveness plane showing the elliptical shape of the distribution of bootstrapped ICERs for a treatment that is more effective and more costly than the comparator.

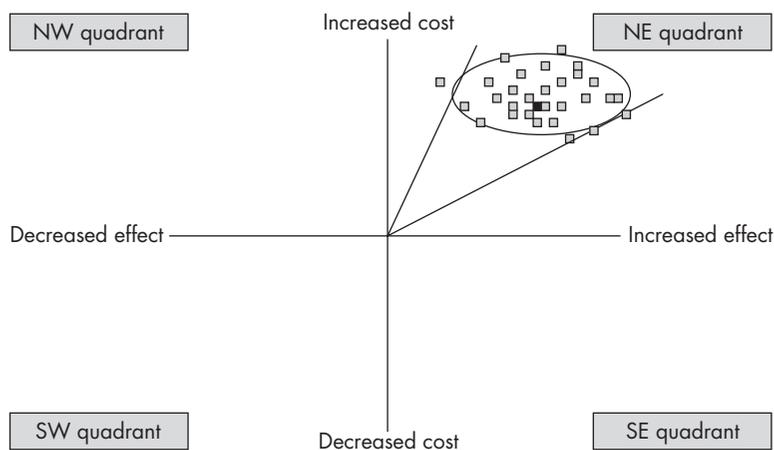


Figure 10.6 Cost-effectiveness plane showing the elliptical shape of the distribution of 95% of bootstrapped ICERs for a treatment that is more effective and more costly than the comparator.

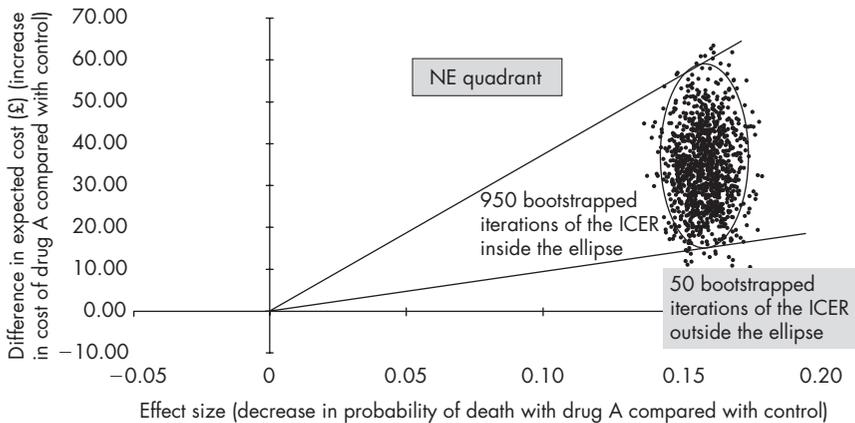


Figure 10.7 Cost-effectiveness plane showing 1000 bootstrapped ICERs for an intervention that is more effective and more costly than the comparator, with quasi 95% confidence intervals for the ICER.

In reality we will have at least 1000 iterations of the bootstrapped ICER, producing a plot similar to the one in Figure 10.7. The ICER distribution shown in this graph suggests that it is highly likely that selecting the new treatment will be more effective and more costly.

Using this method, it is also possible to more easily interpret negative ICERs, which can result if either the cost or effect difference is negative. We may have a positive mean ICER, but once the bootstrapped distribution has been derived it may be clear that uncertainties in cost or effect mean that there are also uncertainties in the ICER.

Uncertainties about difference in effect

We may have a situation where the mean effect size for an intervention is positive. However, if the 95% confidence interval around an effect size crosses zero, then we are not sure that the intervention is as effective as the mean effect size suggests. This uncertainty in effect size will lead to uncertainty in the ICER distribution. It may be that the mean ICER for the intervention suggests increased effect and increased costs. This is denoted by the dark square on the cost-effectiveness plane in Figure 10.8. However, once the bootstrapped distribution has been derived from the full range of effect and cost differences, it is clear that there are some instances where the ICER is negative (less effective and more costly). Figure 10.8 shows a typical bootstrap distribution for this type of ICER.

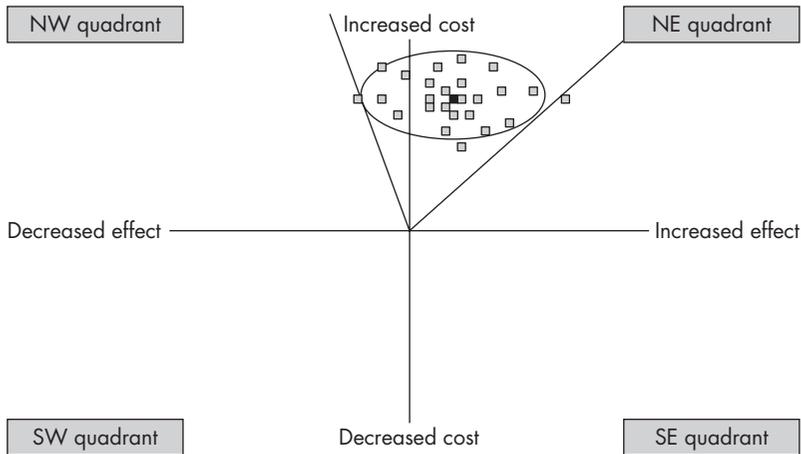


Figure 10.8 Cost-effectiveness plane showing the elliptical shape of the distribution of 95% of bootstrapped ICERs for a treatment that is more costly than the comparator and where the difference in effectiveness is not significant at 5% level.

This is important because ICERs in the northwest quadrant suggest that we should not use that intervention because it is more costly and less effective (is dominated).

Uncertainties about differences in cost

We may have a situation where the mean cost difference for an intervention is positive. However, if the 95% confidence interval around a cost difference crosses zero, then we are not sure that the intervention is more costly, as the mean effect size suggests. This uncertainty in cost difference will lead to uncertainty in the ICER distribution. It may be that the mean ICER for the intervention suggests increased effect and increased cost. This is denoted by the dark square on the cost-effectiveness plane in Figure 10.9. However, once the bootstrapped distribution has been derived from the full range of effect and cost differences, it is clear that there are some instances where the ICER is negative (more effective and less costly). Figure 10.9 shows a typical bootstrap distribution for this type of ICER.

This is important because ICERs in the southeast quadrant suggest that we should use that intervention because it is less costly and more effective (is dominant).

Graphical presentations of bootstrapped ICER distributions are very popular because they are straightforward to understand. Some published

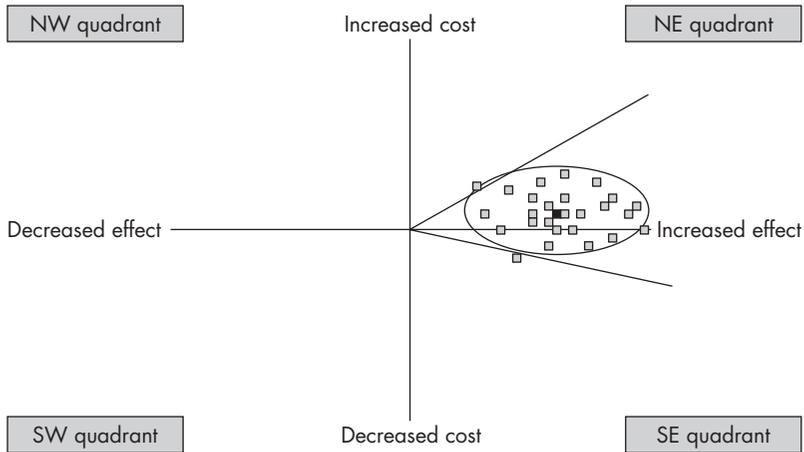


Figure 10.9 Cost-effectiveness plane showing the elliptical shape of the distribution of 95% of bootstrapped ICERs for a treatment that is more costly than the comparator and where the difference in cost is not significant at 5% level.

examples are listed at the end of the chapter. Also, 2.5% and 97.5% percentiles can be extracted directly from the simulated data to give a quantitative measure of uncertainty for the ICER. This can only be done if there have been sufficient iterations.

Estimating the probability of cost-effectiveness

We have seen that it is possible to represent the uncertainty concerning ICERs. Decision-makers often have problems with ICERs. We all agree that a lower ICER is more desirable than a high ICER because we are then paying less for a given improvement in effectiveness. However, we are not universally decided on what is the maximum size an ICER can be before we do not feel that we want to pay that amount for a given improvement in effectiveness. This means that there is no 'gold standard' cost per ICER above which we will not pay for a treatment. Therefore, if the ICER is below the decision-maker's own maximum willingness to pay for a given improvement in effectiveness, the treatment will be funded. However, if the ICER is above the decision-maker's own maximum willingness to pay for a given improvement in effectiveness, the treatment will not be funded. We call this maximum willingness to pay lambda (λ).

We can represent this graphically. In Figure 10.10 we have a cost-effectiveness plane showing a distribution of bootstrapped ICERs for a treatment that is more effective and more costly than the comparator.

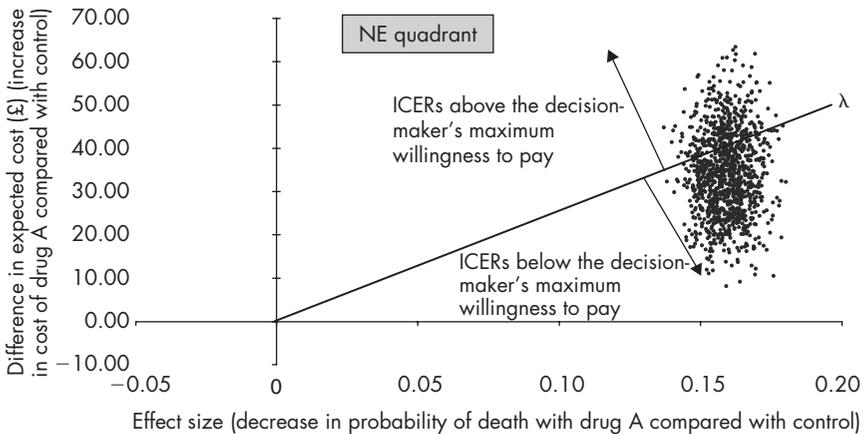


Figure 10.10 Cost-effectiveness plane showing the distribution of bootstrapped ICERs for a treatment that is more effective and more costly than the comparator, and the decision-maker's maximum willingness to pay.

The ICER distribution shown in this graph suggests that it is highly likely that selecting the new treatment will be more effective and more costly.

The decision-maker's maximum willingness to pay for an extra unit of outcome is shown by the line from the origin, labelled λ . ICERs above or to the left of this line are above the decision-maker's maximum willingness to pay. If the true ICER is one of these points the treatment must be rejected, as the increased effect is not considered to be worth the increase in cost. ICERs below or to the right of this line are below the decision-maker's maximum willingness to pay. If the true ICER is one of these points, the treatment will be accepted.

In Figure 10.10, 65% of the ICERs are to the right of λ . This means that there is a 65% probability that the true cost-effectiveness of the intervention is below the decision-maker's maximum willingness to pay. On the basis of this information, the decision-maker may decide not to introduce the new treatment, as there is a 35% chance that the true ICER is above his/her maximum willingness to pay.

However, as was said earlier, we do not know the societal value for λ . The higher a decision-maker's maximum willingness to pay, the higher the probability that the treatment will be considered to be cost-effective. This is shown in Figure 10.11.

From this graph we can see that as λ increases, the more are ICERs below the decision-maker's maximum willingness to pay. This means that as λ increases, the higher becomes the probability that the treatment will be considered to be cost-effective. We can express this graphically,

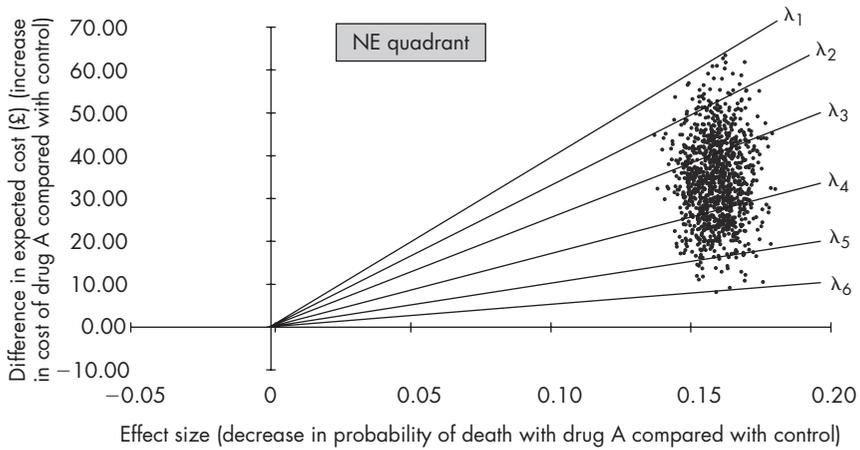


Figure 10.11 Cost-effectiveness plane showing the distribution of bootstrapped ICERs for a treatment that is more effective and more costly than the comparator, and the decision-maker's maximum willingness to pay for a range of values of λ .

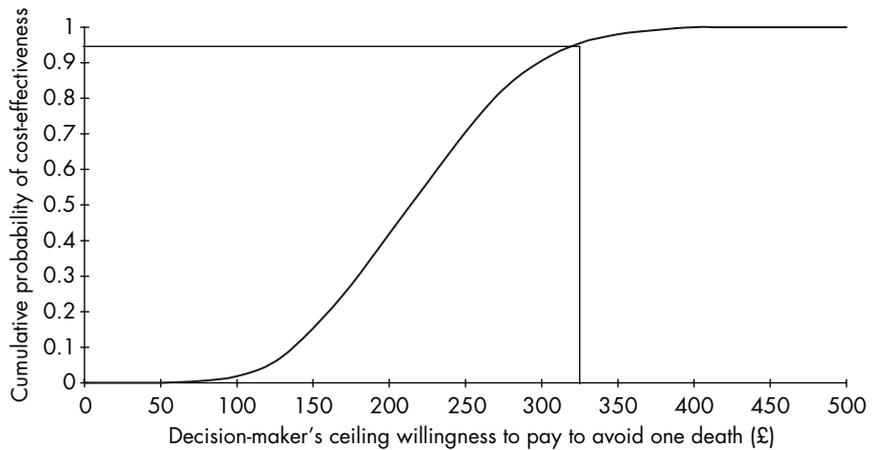


Figure 10.12 Cost-effectiveness acceptability curve for a treatment that is more effective and more costly than the comparator.

using a cost-effectiveness acceptability curve (CEAcc) (Van Hout *et al.*, 1994; Fenwick *et al.*, 2001).

Figure 10.12 shows that if the decision-maker's maximum willingness to pay is £325 or more, there is a 95% probability that the intervention will be cost-effective.

One of the advantages of CEAccs is their ability to quantify the probability that an intervention is cost-effective in a simple manner that

is understandable to decision-makers. It also takes away the traditional reliance on statistical rules about significance levels. Some published examples are given at the end of the chapter.

More advanced topics

In this chapter we have looked at the statistical handling of costs and ICERs. Further subjects that are related to this area are *net benefit* and *expected value of information*. Reading on these more advanced subjects is provided in the Further reading list.

.....

SELF-DIRECTED STUDY EXERCISES

EXERCISE 1: Explaining bootstrapped ICERs

Read Grieve R, Thompson S, Normand C, *et al.* (2003). A cost effectiveness analysis of rhDNase in children with cystic fibrosis. *Int J Technol Assessment Healthcare* 19: 71–79. In this paper, the authors present the use of bootstrapped ICERs for choice of treatment in children with cystic fibrosis. Figures 1a to 1c present the results of the incremental economic analysis for three comparisons. For each analysis represented in these figures, answer the following questions:

1. **What is the mean ICER for the analysis, and what does this tell you about the two treatments under comparison? (Hint: is one dominant?)**
2. **What is the level of uncertainty around this ICER, and will it affect the decision about which treatment to choose? (Hint: is the uncertainty large or small, and does the distribution of ICERs cover more than one quadrant on the cost-effectiveness plane?)**
3. **What is the added value of expressing the uncertainty here using bootstrapping, and do you think this will affect the decision making process?**
4. **What are the limitations associated with the outcome measure used?**

This exercise can be repeated with other published studies that present bootstrapped ICER distributions. Suggested examples are Griffiths *et al.* (2001), Korthals-de Bos *et al.* (2003) and Goodacre *et al.* (2004) (see Further reading).

EXERCISE 2: Explaining cost-effectiveness acceptability curves

Read Delaney BC, Wilson S, Roalfe A, *et al.* (2000). Cost effectiveness of initial endoscopy for dyspepsia in patients over age 50 years: a randomised controlled trial in primary care. *Lancet* 356: 1965–1969. In this paper, the authors present the use of cost-effectiveness acceptability curves for choice of treatment in patients over 50 years of age with dyspepsia. Table 4 and Figure 3 present the results of the

incremental economic analysis and cost-effectiveness acceptability analysis for initial endoscopy versus usual management. For the analysis represented in this figure answer the following questions:

1. **What is the mean ICER for the analysis, and what does this tell you about the two treatments under comparison? (Hint: is one dominant?)**
2. **How robust is this ICER, and will this affect the decision about which treatment to choose? (Hint: look at the results in Table 4.)**
3. **What do the cost-effectiveness acceptability curves in Figure 3 tell you about the two treatments under comparison? (Hint: what happens to the probability that one treatment is cost-effective as the ceiling willingness to pay is increased?)**
4. **What does it mean when a cost-effectiveness acceptability curve crosses the y-axis?**
5. **Can you explain the difference between the four cost-effectiveness acceptability curves in Figure 3?**
6. **What is the added value of expressing the uncertainty here using cost-effectiveness acceptability analysis, and do you think this will affect the decision-making process?**
7. **What are the limitations associated with the outcome measure used?**

This exercise can be repeated with other published studies that present cost-effectiveness acceptability analysis. Suggested examples are Sculpher *et al.* (2004) (see below).

.....

References

Barber JA, Thompson SG (1998). Analysis and interpretation of cost data in randomised controlled trials: a review of published studies. *Br Med J* 317: 1195–1200.

Barber JA, Thompson SG (2000). Analysis of cost data in randomised controlled trials: an application of the non-parametric bootstrap. *Statistics Med* 19: 3219–3236.

Briggs A, Fenn P (1998). Confidence intervals or surfaces? Uncertainty on the cost effectiveness plane. *Health Econ* 7: 723–740.

Briggs A, Gray A (1998). The distribution of healthcare costs and their statistical analysis for economic evaluation. *J Health Serv Res Policy* 3: 233–245.

Briggs AH, Wonderling DE, Mooney CZ (1997). Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Econ* 6: 327–340.

Elliott RA, Davies LM, Payne K, Moore JK, Harper NJN (2004). Costing day case anaesthesia: obtaining accurate patient-based costs for adults and children. *Int J Technol Assessment Healthcare* (in press).

Ellis SL, Carter BL, Malone DC, *et al.* (2000). Clinical and economic impact of ambulatory care clinical pharmacists in the management of dyslipidemia in older adults: the IMPROVE study. *Pharmacotherapy* 20: 1508–1516.

- Fenwick E, Claxton K, Sculpher M (2001). Representing uncertainty: the role of cost effectiveness acceptability curves. *Health Econ* 10: 779–787.
- Grieve R, Thompson S, Normand C, *et al.* (2003). A cost effectiveness analysis of rhDNase in children with cystic fibrosis. *Int J Technol Assessment Healthcare* 19: 71–79.
- Hart A (2001). Mann–Whitney test is not just a test of medians: differences in spread can be important. *Br Med J* 323: 391–393.
- Lambert CM, Hurst NP, Forbes JF, *et al.* (1998). Is day care equivalent to inpatient care for active rheumatoid arthritis? Randomised controlled clinical and economic evaluation. *Br Med J* 316: 965–969.
- Polsky D, Glick HA, Willke R, Schulman K (1997). Confidence intervals for cost-effectiveness ratios: a comparison of four methods. *Health Econ* 6: 243–252.
- Rutten-van Molken M, van Doorslaer EK, van Vliet RC (1994). Statistical analysis of cost outcomes in a randomized controlled trial. *Health Econ* 3: 333–345.
- Thompson SG, Barber JA (2000). How should cost data in pragmatic randomised trials be analysed? *Br Med J* 320: 1197–1200.
- UK Prospective Diabetes Study Group (1998). Cost effectiveness analysis of improved blood pressure control in hypertensive patients with type 2 diabetes. *Br Med J* 317: 720–726.
- Van Hout BA, Al MJ, Gordon GS, Rutten FF (1994). Costs, effects and C/E ratios alongside a clinical trial. *Health Econ* 3: 309–319.

Further reading

- Claxton K (1999). The irrelevance of inference: a decision-making approach to the stochastic evaluation of healthcare technologies. *J Health Econ* 18: 341–364.
- Claxton K (1999). Bayesian approaches to the value of information: implications for the regulations of new pharmaceuticals. *Health Econ* 8: 269–274.
- Claxton K, Neuman PJ, Araki SS, Weinstein MC (2001). The value of information: an application to a policy model of Alzheimer’s disease. *Int J Health Technol Assessment Healthcare* 17: 38–55.
- Grieve A (1998). Issues for statisticians in pharmaco-economic evaluations. *Statistics Med* 17: 1715–1723.
- O’Brien BJ, Drummond MF, Labelle RJ, Willan A (1994). In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in healthcare. *Med Care* 32: 150–163.

Examples of economic evaluations that present graphical bootstrapped ICERs

- Elliott RA, Payne K, Moore JK, *et al.* (2003). Clinical and economic choices in anaesthesia for day surgery: a prospective randomized controlled trial. *Anaesthesia* 58: 412–416.

- Goodacre S, Nicholl J, Dixon S, *et al.* (2004). Randomised controlled trial and economic evaluation of a chest pain observation unit compared with routine care. *Br Med J* 328: 1–6.
- Grieve R, Thompson S, Normand C, *et al.* (2003). A cost effectiveness analysis of rhDNase in children with cystic fibrosis. *Int J Technol Assessment Healthcare* 19: 71–79.
- Griffiths TL, Phillips CJ, Davies S, *et al.* (2001). Cost effectiveness of an outpatient multidisciplinary pulmonary rehabilitation programme. *Thorax* 56: 779–784.
- Korthals-de Bos IB, Hoving JL, van Tulder MW, *et al.* (2003). Cost effectiveness of physiotherapy, manual therapy and general practitioner care for neck pain: economic evaluation alongside a randomised controlled trial. *Br Med J* 326: 911–916.

Examples of economic evaluations that present cost-effectiveness acceptability curves

- Delaney BC, Wilson S, Roalfe A, *et al.* (2000). Cost effectiveness of initial endoscopy for dyspepsia in patients over age 50 years: a randomised controlled trial in primary care. *Lancet* 356: 1965–1969.
- Multi-centre Aneurysm Screening Study Group (2002). Multi-centre aneurysm screening study (MASS): cost effectiveness analysis of screening for abdominal aortic aneurysms based on four year results from randomised controlled trial. *Br Med J* 325: 1135–1138.
- Sculpher M, Manca S, Abbott J, *et al.* (2004). Cost effectiveness analysis of laparoscopic hysterectomy compared with standard hysterectomy: results from a randomised trial. *Br Med J* 328: 134–137.