How to read a paper
Papers that tell you what things cost (economic analyses)

Trisha Greenhalgh

What is economic analysis?
An economic analysis can be defined as an analysis that uses analytical techniques to define choices in resource allocation. This article is based largely on a short booklet by Professor Michael Drummond and two of the forerunners to the “Users’ Guides to the Medical Literature” series. A recent book, Elementary Economic Evaluation in Health Care, is also useful.

Measuring costs and benefits of health interventions
Not long ago, I was taken to hospital to have my appendix removed. From the hospital’s point of view, the cost of my care included my board and lodging for five days, a proportion of doctors’ and nurses’ time, drugs and dressings, and investigations (blood tests and a scan). Other direct costs (see box) included my general practitioner’s time for attending me in the middle of the night and the cost of the petrol my husband used when visiting me (not to mention the grapes and flowers).

In addition to this, there were the indirect costs of my loss in productivity. I was off work for three weeks, and my domestic duties were temporarily carried out by various friends, neighbours, and a hired nanny. Also, from my point of view, there were several intangible costs, such as discomfort, loss of independence, and a cosmetically unsightly scar. As the box shows, these direct, indirect, and intangible costs constitute one side of the cost-benefit equation. On the benefit side, the operation greatly increased my chances of staying alive and I had a nice rest from work.

In this example, few patients (and even fewer purchasers) would perceive much freedom of choice in deciding to opt for the operation. But most health interventions do not concern definitive procedures for surgical emergencies. At some stage, almost all of us will be forced to decide whether having a routine operation, taking a particular drug, or compromising our lifestyle to treat a chronic but not immediately life threatening condition is “worth it.”

It is fine for informed individuals to make choices about their own care by gut reaction (“I'd rather live with my hernia than be cut open,” or “I know about the risk of thrombosis but I want to continue to smoke and stay on the pill”). But when the choices are about other people's care, subjective judgments are the last thing that should enter the equation. Most of us would want the planners and policymakers to use objective, explicit, and defensible criteria when making decisions such as “No, this patient may not have a kidney transplant.”

One important way of addressing the “what's it worth?” question for a given health state (such as having poorly controlled diabetes or a flare up of rheumatoid arthritis) is to ask someone in that state how they feel. A number of questionnaires have been developed which attempt to measure overall health status, such as the Nottingham health profile, the SF-36 general health questionnaire, and the McMaster health utilities index questionnaire.

In some circumstances, disease specific measures of wellbeing are more valid than general measures. For example, answering “yes” to the question, “Do you get very concerned about the food you are eating?” might indicate anxiety in someone without diabetes but normal self care attitudes in someone with diabetes. There has also been an upsurge of interest in patient specific measures of quality of life, to allow different patients to place different values on particular aspects of their health and wellbeing. Of course, when quality of life is being analysed from the point of view of the patient, this is a sensible and humane approach. However, the health economist tends to make decisions about groups of patients or populations, in which case patient specific, and even disease specific, measures of quality of life have limited relevance.

The authors of standard instruments (such as the SF-36) for measuring quality of life have often spent years ensuring they are valid, reliable, and relevant.

Summary points
An economic analysis should be based on a primary study or meta-analysis that is scientifically valid, reliable, and relevant.

When deciding whether an economic analysis has been done correctly, you should not simply check the arithmetic but consider whether all direct, indirect, and intangible costs and benefits have been included.

In the allocation of limited resources, the comparison of different health states is unavoidable, but instruments for measuring health related quality of life are not as objective as they seem.
Examples of costs and benefits of health interventions

<table>
<thead>
<tr>
<th>Costs</th>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct:</td>
<td>Economic:</td>
</tr>
<tr>
<td>“Board and lodging”</td>
<td>Prevention of illness that is</td>
</tr>
<tr>
<td>Drugs, dressings, etc</td>
<td>expensive to treat</td>
</tr>
<tr>
<td>Investigations</td>
<td>Avoidance of admission to hospital</td>
</tr>
<tr>
<td>Staff salaries</td>
<td>Return to paid work</td>
</tr>
<tr>
<td>Indirect:</td>
<td>Clinical:</td>
</tr>
<tr>
<td>Work days lost</td>
<td>Postponement of death or disability</td>
</tr>
<tr>
<td>Value of “unpaid” work</td>
<td>Relief of pain, nausea, breathlessness, etc</td>
</tr>
<tr>
<td>Intangible:</td>
<td>Improved vision, hearing, muscular strength, etc</td>
</tr>
<tr>
<td>Pain and suffering</td>
<td>Quality of life:</td>
</tr>
<tr>
<td>Social stigma</td>
<td>Increased mobility and independence</td>
</tr>
<tr>
<td></td>
<td>Improved wellbeing</td>
</tr>
<tr>
<td></td>
<td>Release from sick role</td>
</tr>
</tbody>
</table>

sick person would place on a return to health. There are three main methods of assigning such values:

- Rating scale measurements—the respondent is asked to make a mark on a fixed line, labelled, for example, “perfect health” at one end and “death” at the other, to indicate where he or she would place the state in question (for example, being confined to a wheelchair by arthritis of the hip);
- Time tradeoff measurements—the respondent is asked to consider a particular health state (for example, infertility) and estimate how many of their remaining years in full health they would sacrifice to be “cured” of the condition;
- Standard gamble measurements—the respondent is asked to consider the choice between living for the rest of their life in a particular health state and taking a “gamble” (such as having an operation) with a given odds of success which would return them to full health if it succeeded but kill them if it failed. The odds are then varied to see at what point the respondent decides the gamble is not worth taking.13

The quality adjusted life year (QALY) can be calculated by multiplying the preference value for that state with the time the patient is likely to spend in that state. The results of cost-benefit analyses are usually expressed in terms of “cost per QALY,” some examples of which are shown in the second box.14

The use of QALYs is controversial. Any measure of health state preference values is, at best, a reflection of the preferences and prejudices of the individuals who contributed to its development. Indeed, it is possible to come up with different values for QALYs, depending on how the questions from which the health state preference values are derived were posed.15 Furthermore, it is virtually impossible to combine different QALYs to measure the effect of more than one serious or disabling condition on a patient.16 As medical ethicist John Harris has pointed out, QALYs are, like the society that produces them, inherently ageist, sexist, racist, and loaded against those with permanent disabilities (since even a complete cure of an unrelated condition would not restore the individual to “perfect health”).

Furthermore, QALYs distort our ethical instincts by focusing our minds on years of life rather than people’s lives. A disabled premature infant in need of an intensive care cot will, argues Harris, be allocated more resources than it deserves in comparison with a 50 year old woman with cancer, since the infant, were it to survive, would have so many more life years to quality adjust.17

Other authors have come up with the HYE (healthy years equivalent) measure, which incorporates the individual’s likely improvement or deterioration in health status in the future and is said to avoid some, but not all, of the disadvantages of the QALY.18 Given that the critics of QALYs and HYEs have offered no alternative, all encompassing measure of health status, these utility based units are set to remain in the health economist’s toolkit for the foreseeable future. For a more detailed discussion of these issues by a multi-disciplinary panel, see Anthony Hopkins’s booklet Measures of the Quality of Life.19

There is, however, another form of analysis which, although it does not abolish the need to place arbitrary numerical values on life and limb, avoids the buck stopping at the unfortunate health economist. This approach, known as cost-consequences analysis, presents the results of the economic analysis in a disaggregated form. In other words, it expresses different outcomes in terms of their different natural units (something real such as months of survival, legs amputated, or babies taken home), so that individuals can assign their own values to particular health states before calculating whether the intervention is “worth it.”

Ten questions to ask about an economic analysis

The checklist which follows is based on the sources mentioned earlier,12 as well as suggestions made by a working party set up by the BMJ to produce guidelines for journal editors on appraising economic evaluations (M Drummond, personal communication).

**Question 1:** Is the analysis based on a study that answers a clearly defined clinical question about an economically important issue?

Before pursuing any of the economic arguments, make sure that the trial being analysed is scientifically relevant and capable of giving unbiased and unambiguous answers to the clinical question posed in its introduction.

**Results of cost-benefit analysis for some medical procedures**

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Cost per QALY (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cholesterol testing and diet therapy</td>
<td>220</td>
</tr>
<tr>
<td>Advice to stop smoking from patient’s own doctor</td>
<td>270</td>
</tr>
<tr>
<td>Hip replacement for arthritis</td>
<td>1 180</td>
</tr>
<tr>
<td>Kidney transplant</td>
<td>4 710</td>
</tr>
<tr>
<td>Breast cancer screening</td>
<td>5 780</td>
</tr>
<tr>
<td>Cholesterol testing and drug therapy if indicated (ages 25-39)</td>
<td>14 150</td>
</tr>
<tr>
<td>Neurosurgery for malignant brain tumours</td>
<td>107 780</td>
</tr>
</tbody>
</table>

BMJ VOLUME 315 6 SEPTEMBER 1997
Question 2: Whose viewpoint are costs and benefits being considered from?
From the Treasury's point of view, the most cost effective health intervention is one which returns all citizens promptly to taxpayer status and, when this status is no longer tenable, causes immediate sudden death. From the drug company's point of view, it would be difficult to imagine a cost-benefit equation which did not contain one of the company's products, and from a physiotherapist's point of view, the removal of a physiotherapy service would never be cost effective. Almost all economic analyses have some funding, and all have been inspired by someone with a vested interest; the paper should say which.

Question 3: Have the interventions being compared been shown to be clinically effective?
In general, the intervention that "works out cheaper" should not be substantially less effective in clinical terms than the one which stands to be rejected on the grounds of cost.

Question 4: Are the interventions sensible and workable in the settings where they are likely to be applied?
Too many research trials look at intervention packages which would be impossible to implement in the non-research setting (they assume, for example, that general practitioners will own a state of the art computer and agree to follow a protocol, that infinite nurse time is available for the taking of blood tests, or that patients will make their personal treatment choices solely on the basis of the trial's conclusions). Remember that standard current practice, which may be to do nothing, should almost certainly be one of the alternatives compared.

Question 5: Which method of analysis was used, and was this appropriate?
This decision can be summarised as follows:
- Cost minimisation analysis would be most appropriate if the interventions produced identical outcomes;
- Cost effectiveness analysis would be most appropriate if the important outcome is unidimensional;
- Cost utility analysis would be most appropriate if the important outcome is multidimensional;
- Cost benefit analysis would be most appropriate if the cost benefit equation for this condition needs to be compared with cost benefit equations for different conditions;
- Cost consequences analysis would be most appropriate if a cost benefit analysis would otherwise be appropriate but the preference values given to different health states are disputed or likely to change.

Question 6: How were costs and benefits measured?
Consider an economic evaluation of a trial comparing the rehabilitation of stroke patients into their own homes, including attendance at a day centre, with a standard alternative intervention (rehabilitation in a long stay hospital). The economic analysis must take into account not just the time of the various professionals involved, the time of the secretaries and administrators who help run the service, "overheads" (such as heating and lighting), and the cost of the food and drugs consumed by the stroke patients, but also a fraction of the capital cost of building the day centre and maintaining a transport service to and from it.

In a cost effectiveness analysis, changes in health status will be expressed in natural units. But just because the units are natural does not automatically make them appropriate. For example, the economic analysis of the treatment of peptic ulcer by two different drugs might measure outcome as "proportion of ulcers healed after a six week course." Treatments could be compared according to the cost per ulcer healed. However, if the relapse rates on the two drugs were very different, drug A might be falsely deemed "more cost effective" than drug B. A better outcome measure here might be "ulcers that remained healed at one year."

Question 7: Were incremental, rather than absolute, benefits considered?
This question is best illustrated by a simple example. Let's say drug X, at £100 per course, cures 10 out of every 20 patients. Its new competitor, drug Y, costs £120 per course and cures 11 out of 20 patients. The cost per case cured with drug X is £200 (since you spent £2000 curing 10 people), and the cost per case cured with drug Y is £218 (since you spent £2400 curing 11 people).

The incremental cost of drug Y—the extra cost of curing the extra patient—is not £18, but £400, since this is the total amount extra that you have had to pay to achieve an outcome over and above what you would have achieved by giving all patients the cheaper drug. This striking example should be borne in mind the next time a pharmaceutical representative tries to persuade you that his or her product is "more effective and only marginally more expensive."

Question 8: Was the "here and now" given precedence over the distant future?
A bird in the hand is worth two in the bush: in health as well as money terms, we value a benefit today more highly than we value a promise of the same benefit in five years' time. When the costs or benefits of an intervention (or lack of the intervention) will occur some time in the future, their value should be discounted to reflect this. The actual amount of discount that should be allowed for future, as opposed to immediate, health benefit is fairly arbitrary, but most analyses use a figure of around 5% per year.
Economic analysis alongside controlled trials.

Thanks to Professor Mike Drummond and Dr Alison Tonks for drawing pretty well for his age.

A cost-utility or cost-benefit analysis which gives a composite measure of cost and effectiveness, and, more commonly, the reader is faced with a cost-consequences analysis, in which the reader of the paper can attach his or her own values to different utilities, was introduced earlier. In practice, this is an unusual way of presenting an economic analysis, and, more commonly, the reader is faced with a cost-utility or cost-benefit analysis which gives a composite score in unfamiliar units which do not translate readily into exactly what gains and losses the patient can expect. The situation is analogous to the father who is told “your child’s IQ is 115” when he would feel far better informed if he were presented with the disaggregated data: “Johnny can read, write, count, and draw pretty well for his age.”

Thanks to Professor Mike Drummond and Dr Alison Tonks for advice on this chapter.


When I use a word...

Sausages

I recently learnt about a rare type of neuropathy called hereditary neuropathy with liability to pressure palsies (abbreviated to HNPP) or hereditary pressure sensitive neuropathy (HPSN), also known as tomaculous neuropathy. In many cases it is associated with a deletion in chromosome 17p11.2. However, the term is not exclusive, and tomaculous changes have been described in other neuropathies, such as type 1B Charcot–Marie–Tooth syndrome and hereditary neuralgic amyotrophy.

Tomaculous refers to the sausage shaped swellings of myelin that occur along the affected nerves—Latin: tomaculum, a diminutive of tomus, a sausage. Latin had several other words for a sausage: hillae, which referred to the sausage itself (originally salsicia) comes from the Latin: sal, ´(tome) meaning a cut. From sal came sal on which is based the Greek word salami, for a type of sausage. The bacterium Clostridium botulinum is so called because it looks sausage shaped, not because it gets into sausages or affects the stomach. And intestinal sausages feature in other languages too: spiced (intestines again); and botulus, a word for the stomach.

Spices are another important ingredient. The currently popular pepperoni, not surprisingly, contains peppers. The word sausage itself (originally salcia) comes from the Latin: sal, meaning salt, as does salami. Because you slice a salami thinly, the word has been used for metaphorical slicing: salami tactics, the gradual whittling away of the members of an organisation; salami technique, a fraud involving the deduction of tiny amounts of money from innumerable sources (like Richard Pryor’s scam in Superman III); and salami publication, when you get several papers out of a single piece of work, slicing it up as finely as you can.

Which brings us back to tomaculous neuropathy. The tomaculum was a sausage that could be served sliced, from the Greek word τομή (tome) meaning a cut. beim dr. jeff aronson, clinical pharmacologist, oxford

The articles in this series are excerpts from How to read a paper: the basics of evidence based medicine. The book includes chapters on searching the literature and implementing evidence based findings. It can be ordered from the BMJ Publishing Group: tel 0171 383 6185/6245; fax 0171 383 6662. Price £13.95 UK members, £14.95 non-members.

2 Drummond MF, Richardson WS, O’Brien BJ, Levine M, Heyland D. Users’ guides to the medical literature XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? JAMA 1997;277:3352-7.
3 O’Brien BJ, Heyland D, Richardson WS, Levine M, Drummond MF. Users’ guides to the medical literature XIII. How to use an article on economic analysis of clinical practice. B. What are the results and will they help me in caring for my patients? JAMA 1997;277:1802-6.
17 Greenhalgh T. Papers that summarise other papers (systematic reviews and meta-analyses). BMJ 1997 (in press).
Statistics notes

Trials randomised in clusters

J Martin Bland, Sally M Kerry

In most randomised trials each subject is individually assigned at random to an intervention group. The intervention is applied directly to the subject and observations are made on each individual to determine the outcome of the intervention. Sometimes subjects cannot be allocated independently, or they may interact with one another during the treatment period.

In trials of population screening, for example, screening centres may be set up in some districts and not in others. This may be necessary because widespread publicity is needed to encourage subjects to come for screening, or because members of the screening group might pass on information to neighbours who have been allocated to the control arm (no screening), leading them to demand screening.

In the Swedish two county trial of breast cancer screening the county of Kopparberg was divided into seven geographical areas.1 2 Each was then subdivided into three units, either parishes or municipalities, two of which were randomly allocated to screening and the other to control. The county of Östergötland was divided into 12 areas, each of which was subdivided into two experimental units, one allocated to screening and the other to control. The subjects within a unit are called a cluster, and the trial used cluster randomisation. Cluster randomisation is used especially in public health and general practice research.

A price must be paid

There is a price to be paid for this design at the analysis stage. We cannot think of our trial subjects as independent individuals but must do the analysis at the level of the experimental unit.3 In the two county trial women within a parish will be more alike than a random sample of women from the two counties. We have two sources of variation: that between people in a parish and that between parishes. The variability between parishes must be taken into account in the analysis.

The effect of cluster randomisation is to increase the size of standard errors and hence widen confidence intervals and increase P values compared with a study of the same size using simple randomisation. The effective sample size is reduced and power is lost. The larger and fewer the clusters are, the more important and greater the effect becomes.

Many cluster randomised trials ignore this design effect in the analysis. Early reports of the two county trial4 5 did this, although more recent analyses have taken it into account.6 In a review of 16 non-therapeutic intervention trials employing cluster randomisation only eight allowed for the clusters in the analysis.7 Ignoring the correct unit of analysis in this way may lead to spurious positive findings.8

Health promotion is another area where cluster randomised designs are common. For example, in the evaluation of a health education programme schools may be randomly allocated to receive the education programme or to act as control. The subsequent behaviour and knowledge of the children can be compared. As children may influence each other children within a school cannot be regarded as independent of one another and the school should be the unit of analysis.

Use the right unit

Patients are often allocated so that all the patients of one general practitioner receive the same treatment. In a trial of terminal care coordination, for example, general practitioners were allocated into two groups and the patients of doctors in one group were offered the extra intervention.9 All the patients needing terminal care in a practice formed a single cluster. In this example the treatment was applied directly to the patient, who received visits from the care coordinators. Sometimes the treatment is applied to the provider of care rather than to the patient directly; and here the effect of the clustering may be much larger. For example, to improve the treatment of asthma in general practice general practitioners were allocated randomly to three groups.10 The first group was given an intensive programme of small group education, the second a lesser intervention, and the third no intervention at all. A sample of each general practitioner's asthmatic patients was selected. These patients received questionnaires about their symptoms, and the prevalence was compared between the groups. The experimental unit was the general practitioner, not the patient. The proportion of patients who reported symptoms was used as a measure of the general practitioner's effectiveness and the three groups of doctors compared by analysis of variance.11

We shall discuss the design and analysis of cluster randomised trials in future statistics notes.