Does economic evaluation have anything to offer the rheumatologist?

Economic evaluation has become an increasingly important component of the evaluation of new drugs and technologies, or indeed any aspect of health care that involves significant resources. Whether from the perspective of the whole National Health Service, or simply that of a rheumatology service, the aim of economic evaluation is to estimate, using standard measures, both the health benefits and the resource consequences of health care interventions. The evidence thus provided allows direct comparison of the relative efficiency of competing programmes of care or, in the crude political vernacular, assessment of ‘value for money’.

The most important resource in rheumatology is the expertise and time of the rheumatology team. Along with variations in levels of service provision [1], case mix [1, 2] and referral rates of general practitioners (GPs) [3, 4], there is anecdotal evidence suggesting considerable variation in how rheumatology teams use their time. Furthermore, despite increases in consultant provision, and increased patient throughput, referral rates and waiting times continue to rise [5]. However, we do not know whether or how these differences in resource use affect the level of locomotor disability and unmet need in different populations around the country. Economic evaluation can address the significance of these variations in resource use and provide insight into what sort of rheumatology service should be provided at a local or regional level. For example, is it better to spend more time offering routine follow-up care of patients with inflammatory disease or should there be less of this and more clinic appointments ‘on demand’? What priority should be given to different types of clinical problem or does it not matter? Are ‘early arthritis clinics’ an effective use of resources? Time spent on one activity cannot be spent on another; hence, at their heart, these are economic choices which should hinge on consideration of ‘opportunity cost’ or the value of forgone opportunities. Perhaps there are even some activities undertaken by rheumatology teams which would be better left to others?

The basic methods and guidelines for undertaking economic evaluations have been described elsewhere [6]. Often these methods are applied in the context of a randomized clinical trial in which equal numbers of patients are treated in each arm, and comparisons are made of the health outcomes and resource consequences over a set time period. Equivalence trial designs can also be employed either to compare the resource consequences of interventions which have the same therapeutic effect or, alternatively, to assess the differences in health outcome between the alternative interventions while holding constant the resource input. For example, Sutton [7] has argued that a more relevant experimental approach from the perspective of the health service, where budgets are usually fixed and predetermined, is
that of ‘outcome maximization’. In this method, differences in the allocation of resources to each trial arm are minimized, and the emphasis is on identifying significant differences in health gain. The importance of the ‘outcome maximization’ method is that it is primarily focused on the health outcome rather than on differences in costs. Its main drawback, along with uncritical acceptance of largely historically determined levels of funding, is that it may too easily adopt the relatively narrow health service perspective and take no account of the indirect costs attributable to loss of work and productivity. Because these costs can often outweigh the direct costs of health care, any programme which returns patients to work, or indeed other daily activities valued by them, runs the risk of being poorly valued and underfunded from a societal perspective.

One of the main problems faced by rheumatologists in measuring health outcomes, and which is highlighted by economic evaluation, is the chronicity of rheumatic disease. While clinical intervention may produce both short- and medium-term benefits, these benefits are not infrequently outweighed by iatrogenic effects and, in the longer term, the health of the majority of patients with significant rheumatic disease tends to deteriorate despite our best efforts. Achieving the twin clinical goals of reducing both present and future disability requires a very wide range of interventions [8]. Measuring the net effect on health of these interventions requires instruments (Table 1) which capture broader facets of health than those measured traditionally, such as radiological erosions, joint counts, etc., and should include the psychosocial impact and attitudinal adaptations to health which patients undergo in the course of chronic disease. Although condition-specific outcomes are important and necessary, they are not sufficient measures of the health of our patients. Furthermore, instruments are required which, by using common or generic units of health, enable the health impact of disparate conditions to be compared. Without such instruments, it is difficult to compare the efficiency of therapies for different conditions, or to undertake intelligent audit of different teams serving different populations. This need has spawned a large number of generic health instruments, one set mainly derived from econometrics and the other from psychometric traditions.

The econometric construct of health, whose origins are largely utilitarian, measures health using common unidimensional units of ‘health utility’. This method values health as a subjective sense of ‘well-being’ or ‘quality of life’ which varies not only with ill health, but with experience, time, culture and socio-economic characteristics. It does not measure health in relation to some common health standard or norm. The possibility that perceived health may improve even though, in relation to population norms, health may have declined, is also allowed. These methods approach the goal of capturing ‘quality of life’, which Calman has defined as ‘the extent to which an individual’s hopes and expectations are matched and fulfilled by experience’ [9].

Two basic methods of measuring health utility used by health economists are the ‘time trade off’ (TTO) [10] and ‘standard gamble’ (SG) [11]. Both provide a measure of health utility on a scale which uses perfect health = 1 and death = 0 as anchor points. In the more widely used TTO method, the individual is asked to consider a health state (their own or a theoretical state) and then, for example, to trade some of the next 10 yr of life in exchange for a shorter life lived in perfect health; the units of utility are then calculated as years of healthy life/10. In the worst health states, the subject may be prepared theoretically to die immediately, giving a utility of zero, while in the best states the subject will give up little or no life at all, giving a utility equal or close to one. TTO or SG can be applied directly to individuals experiencing an illness to obtain their valuation of their own health. Alternatively, a societal valuation or health preference for theoretical health states can be obtained from members of the general population. This approach has been used to obtain valuations for the health states defined by a simple questionnaire, the EuroQol (EQ-5D), which was developed as a very simple descriptive instrument for public health use and as an econometric tool.

The EQ-5D questionnaire (part 1) has only five questions, each with only three responses. This gives rise to 3^5 (i.e. 243) possible responses, each of which represents a health state that has been assigned a health utility value or tariff. The tariff for EQ-5D, which was obtained using TTO on a sample of the general population [12], provides a societal view of the respondent’s health state. EQ-5D also contains a visual analogue scale (part 2) which offers the patient the opportunity to give a global valuation of their own health. These valuations, one from society and the other from the patient, should be viewed as complementary and it must not be assumed that they will necessarily agree. The methods of valuation are very different; the VA scale does not include death as an anchor point, and in chronic disease states there is evidence that the patient’s

Table 1. Examples of health outcome measures suitable for use in economic evaluation

<table>
<thead>
<tr>
<th>Type of health outcome scale</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition specific</td>
<td>Health Assessment Questionnaire—RA, OA ACR core disease activity set—RA EULAR core disease activity set—RA WOMAC—OA HAD—anxiety and depression</td>
</tr>
<tr>
<td>Generic health profiles</td>
<td>Nottingham Health Profile SF36 EuroQol Sickness Impact Profile</td>
</tr>
<tr>
<td>Generic summary scales</td>
<td>SF36 physical and mental summary scales SF12 physical and mental summary scales</td>
</tr>
<tr>
<td>Generic health indices and measures of health utility</td>
<td>EuroQol Rosser index Quality of Wellbeing Scale Patient Generated Index Time Trade Off Standard Gamble</td>
</tr>
</tbody>
</table>
valuation may be more optimistic than that of society [13, 14]. Normative data allowing comparison with the general population are now available for the EQ-5D, further improving its usefulness [15, 16].

The psychometric tradition has generated various health profiles which measure health in several distinct domains, e.g. physical, mental, social, etc. Normative values are available for some of these instruments and allow direct comparison of the ‘health status’ of disease populations with the general population. Superficially, the econometric and psychometric approaches are very different, but there is in fact significant common ground between them. For example, the SF36 health profile [17], which measures health in eight dimensions, has been shown using factor analysis to collapse into just two: a physical and a mental dimension each presented as z or T scores. The SF12, a ‘simple shorter daughter’ of SF36, which is also reported as two summary scores, significantly reduces the burden of data collection yet performs only marginally less well than SF36 [18]. Work is also now proceeding to produce a utility tariff for the SF questionnaires.

How does one choose between these apparently radically different approaches to health valuation, each of which clearly has significant advantages and disadvantages? Health profiles such as SF36 or SF12 enable a direct comparison to be made with population norms and they show which areas of health are most affected and where health change has occurred. Their disadvantages are that no overall assessment of the utility attached to health change can easily be made, trade-offs between survival and health-related quality of life cannot be explicitly assessed, and because they generate multiple endpoints, the statistical problems associated with multiple hypothesis testing are present. The temptation to search for significant results through multiple tests will to some extent be removed by adoption of the shorter SF12 with its two summary scales. The advantages of utility-based health indices are that firstly they include death as an anchor point, secondly they can be used to provide a societal valuation (health preference) for different health states and finally give a single value for health outcome. Their main disadvantage is that they may not identify explicitly which areas of health are affected, making it difficult to attribute any health change. Also, the EQ-5D (part 1), in particular, has relatively wide and uneven levels of separation between health states and it should perhaps be viewed more as a crude ‘desk ruler’ than an accurate ‘micrometer gauge’. Another concern raised in respect of all generic instruments is their responsiveness or sensitivity to change. Mostly this concern seems to be based on prejudice rather than empirical observation, and some of the available evidence suggests that health profiles (SF36) and indices (EQ-5D) may be at least as responsive, if not more so, than some traditional condition-specific measures [13, 17].

At present, those planning to embark on an economic evaluation need to consider carefully the requirements of their study and, if generalizability or comparisons between disparate conditions or different populations are at issue, a simple norm-based health profile (e.g. SF36 or SF12) and/or a utility-based measure (TTO or EQ-5D) should be included. When premature death is a potential outcome, a utility-based measure becomes even more relevant. Neither of these methods is free of problems, but the adage ‘it is better to measure imperfectly what is important than to measure perfectly what is unimportant’ is certainly relevant. Both methods give information that quantifies and adds considerably to our understanding of the impact of rheumatic disease. Equally important, they help to quantify the health return from investment in rheumatological services. If nothing else, economic evaluation and the inclusion of generic health measures alongside condition-specific measures will enhance rheumatology’s position in the competition for more resources and will improve the quality of the public health debate about future levels of funding for rheumatic disease.

N. P. Hurst and J. Forbes

Economic and Health Outcomes Group, Rheumatology Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU and Department of Public Health Sciences, University of Edinburgh, Teviot Place, Edinburgh, UK

References

Discussion paper 138. York: Publication Department, Centre for Health Economics, University of York.


