Some guidelines on the use of cost effectiveness league tables

James Mason, Michael Drummond, George Torrance

Decisions to allocate resources in health care are increasingly influenced by relative cost effectiveness. To warn decision makers of some of the pitfalls currently found in cost effectiveness league tables and to suggest how meaningful comparisons may be made between health care technologies a published league table was scrutinised by examining its sources. This showed some of the methodological problems surrounding such tables and how such difficulties could be reduced in future. The source studies in the table featured different years of origin, discount rates, health state valuations, settings, and types of comparison programmes; all of these differences may raise problems for meaningful comparison.

Decision makers need to assess the relative value for money of competing health care interventions. In the absence of systematic comparisons such assessments are likely to take place informally. This will probably have a worse risk-benefit trade off than the formalised use of league tables.

In recent years comparisons between health care interventions in terms of their relative cost effectiveness in cost per life year or per quality adjusted life year (QALY) gained have become fashionable. The principal motivation for this comparison is to help decisions about the allocation of health care resources. For example, Williams calculated the cost per QALY of a range of health care interventions and divided them into "strong candidates for expansion" and "less strong candidates for expansion" to illustrate what could be achieved by using available data and to argue for further refinements.[1] Recently, Laupacis et al proposed that the adoption and use of new health technologies should be classified into five grades of recommendation based on their cost per QALY, although they also acknowledged that many other issues other than cost effectiveness, such as ethical and political considerations, affect the implementation of a new technology.[2]

Centre for Health Economics, University of York, Heslington, York YO1 5DD
James Mason, *research fellow*
Michael Drummond, *professor of economics*

Centre for Health Economics and Policy Analysis and Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada L8N 3Z5
George Torrance, *professor*

Correspondence to:
Dr Mason.

Some extracts of costs for quality adjusted life years (QALY) of competing treatments[7]

Treatment	Cost/QALY (£ Aug 1990)
Cholesterol testing and diet therapy only (all adults aged 40-69)	220
Neurosurgical intervention for head injury	240
Advice to stop smoking from general practitioner	270
Neurosurgical intervention for subarachnoid haemorrhage	490
Antihypertensive treatment to prevent stroke (ages 45-64)	940
Pacemaker implantation	1 100
Hip replacement	1 180
Valve replacement for aortic stenosis	1 140
Cholesterol testing and treatment	1 480
Coronary artery bypass graft (left main vessel disease, severe angina)	2 090
Kidney transplant	4 710
Breast cancer screening	5 780
Heart transplantation	7 840
Cholesterol testing and treatment (incrementally) of all adults aged 25-39	14 150
Home haemodialysis	17 260
Coronary artery bypass graft (one vessel disease, moderate angina)	18 830
Continuous ambulatory peritoneal dialysis	19 870
Hospital haemodialysis	21 970
Erythropoietin treatment for anaemia in dialysis patients (assuming 10% reduction in mortality)	54 380
Neurosurgical intervention for malignant intracranial tumours	107 780
Erythropoietin treatment for anaemia in dialysis patients (assuming no increase in survival)	126 290

Cost effectiveness league tables have recently attracted the interest of decision makers in health care, such as those concerned with the allocation of public resources in the state of Oregon.[3 4] The role of league tables has also been discussed in the British Department of Health (H Neuberger, personal communication), although a final decision on the appropriateness of their use has yet to be made.

Many argue that a league table of costs per QALY, if properly constructed, can provide comprehensive and valid information to help decisions on the allocation of resources. The potential dangers of the unthinking use of league tables, however, as currently presented have been pointed out by several authors.[5 6] Despite such reservations, comparisons of cost effectiveness between health care programmes will inevitably be made by both analysts and decision makers, either formally or informally. We report some inappropriate comparisons and make some recommendations for constructing future league tables. We examined a recently reported league table and identified methodological deficiencies. We have presented guidelines to help decision makers interpret league tables as currently presented.

Problems with cost effectiveness league tables

The paper by Maynard contains a discussion typical of league tables.[7] He presents a table containing 21 health care interventions (see table) and writes:

The methods used to produce such estimates are crude but do facilitate the production of a "league table" which ranks the cost of producing a QALY by investing in competing treatments. An example of these "guesstimates" is set out in Table 1. The implication of these data is that resources should be invested by purchasers in treatments which produce QALYs at low cost.

Note that there is not necessarily presupposition that activities should be funded in the order they appear in the list nor that the same ranking holds for every location. There is a risk, however, that this inference may be drawn. The mere act of assembling data on a range of interventions gives greater prominence to the cost effectiveness data than does the reporting of the studies individually. It also implies a certain degree of homogeneity in study methods. That is, the reader is invited to believe that the ranking of interventions in the table is a function of their relative value for money, independently of the methods of the original economic evaluation study or the quality of the clinical evidence on which it was based.

Closer examination of published league tables shows that the methods of the source studies are rarely homogeneous. We found that 20 of the estimates originated from six source studies, with one being of unknown origin.[18-12] In interpreting the league table a number of methodological features are particularly important; these are discussed in turn.

YEAR OF ORIGIN

The years of analysis in the source studies ranged from 1982 to 1989, although all estimates were converted to 1990 prices for the comparison. The analysis in each source study reflects the then current state of knowledge about costs and benefits. Technological change may mean it is inappropriate to use results for

which the "sell by" date has expired (the reflating of older values may be inaccurate). In addition to technological change, shifts in the relative prices of different interventions over time may affect the ranking.

DISCOUNT RATE

The discounting of costs and benefits occurring in the future was mostly consistent at 5% a year in real terms.[13] This reflects the consensus in reports on economic evaluation and, as all the source studies were from the United Kingdom, the discount rate recommended by the Treasury. One of the six sources, however, responsible for three of the 21 estimates of cost per QALY did not discount either costs or benefits. As this study was concerned with preventing death, or as much as 30 years of long term serious disablement, by neurosurgical intervention the effect of this omission is considerable. In some of the other source studies discounting was quite appropriately not applied as all costs and benefits occurred within a year.

The issue of discounting has been debated recently after a Treasury decision to apply a zero rate in the case of health benefits.[14 15] Therefore there may not be as much agreement on this point in future economic evaluations. The onus is on authors of the original studies to provide cost effectiveness ratios with a range of discount rates, a minimum requirement being 5% and 6% a year on costs and 0%, 5%, and 6% on benefits.

PREFERENCE VALUES FOR HEALTH STATES

Preference values for health states were most commonly derived from the matrix developed by Kind et al.[16] Other of the source studies, however, used patient values, clinician values, or educated guesses. Different approaches to the estimation of utilities for health states generate different values; this has the potential to reduce greatly the comparability of the six source studies. The degree of homogeneity in this league table, however, is greater than many because of the relative prominence of the "Rosser matrix" in cost utility analyses in the United Kingdom. If league tables from North America were being reviewed these would probably show greater variation in analytical approach as there has been considerably more research into, and validation of, measures of health state preference.[17] (See, for example, the table reported in Torrance and Zipursky.[18]) At present it is difficult for users of league tables to satisfy themselves of the validity of the estimates of benefit used once these estimates have been concealed in a league table. Limiting the studies included to those that use the same measure of quality of life might lend a table some face validity. If in future, however, reporting of each estimate is extended to include descriptive profiles of quality of life—that is, health states and the valuations placed on them for the new and comparison treatment—then this may enhance credibility and allow reworking in alternative settings. Writing in the context of quality of life measurement, Cox et al pointed out that in reporting data the analyst has to serve unknown future users with unknown future interests.[19]

RANGE OF COSTS

Detailed investigation of the range of costs considered in each of the six source studies was hampered by inadequate descriptions of methods of costing. This problem is likely to be encountered often by decision makers wishing to assess the methodological quality of published studies. In most of the source studies, however, only direct health care costs were included in the ratio of cost per QALY. Only one estimate[10] included consideration of patients' time, which biased the ranking of this procedure downwards in the league table. If more detailed information were available on costing methods further discrepancies in analytic

approach would probably be exposed, with further consequential shifts in the league table. If future league tables report disaggregated profiles for resources and costs for each study then the costs may be validated or recalculated for the decision maker's own setting.

COMPARISON PROGRAMME

The source studies reported in the table differed in respect of the type of comparison programme from which the incremental cost per QALY is assessed. The "baseline" for comparison may have a major impact on these costs. If a comparison programme of no formal intervention is chosen it is important to establish that the costs and consequences of this have been valued credibly. "Doing nothing" is unlikely to have zero costs and zero benefits. Even in the case of a screening programme (when doing nothing is a real option) some cases would be found even in the absence of any organised service and there would also be the costs of treating disease as and when it occurred. Thus, for most economic evaluations, the notion of an option with zero costs and zero benefits is not relevant.

A central problem for the choice of a comparison programme lies in the perception of the "margin." In other (non-health) areas of social policy economic evaluations usually start with the presumption that all current activities are efficient—that is, worthwhile when compared with all alternatives. Additional money is allocated (at the margin) to the next most worthwhile activity. No such margin exists for health care provision, however, much of which has never been scientifically evaluated.[20] Instead, health care provision has a political margin with budgets allocated on criteria other than cost effectiveness. Decision makers should satisfy themselves that current practice is itself worth having before using it as a comparison for a new treatment. If the comparison programme is inefficient the analysis will be misleading.

Some of the source studies used to create the table compared with an alternative of "doing nothing," others compared with a minimum intervention (which may differ among programmes), and yet others considered the incremental cost per QALY of expanding services to other groups of patients. In such a case, with restricted and non-standardised comparisons, interpretations across the programmes in the league table are fraught with difficulty.

SETTING OF STUDY

All the studies reported by Maynard originated in Britain. Additional complexities are involved if league tables contain studies from several countries. Gray pointed out the need to make adjustments for exchange rates by using appropriate purchasing power parities for medical goods and services (Health Economists Study Group, Aberdeen, 1991). A comparison of this approach with simple exchange rate conversions changed the rankings. More importantly, however, differences also exist between countries in clinical practice, the availability and relative prices of health care resources, and the incentives given to health professionals and institutions. In a recent analysis that compared economic evaluations of the same intervention (a new drug) performed by identical methods in four countries, Drummond et al showed that although the drug was 36% more expensive in the United States than in the three European Community (EC) countries considered, it was also more cost effective in the United States because of the relative price and use of other health care resources compared with those in the EC (such as the higher charges for surgical operations).[21] Considerable care should be exercised when extrapolating the results of cost effectiveness from one country to another or including them in the same league table.

The descriptions of treatments listed in league tables are sometimes inadequate. Cholesterol testing and treatment seems good value at £1480 per QALY. This value, however, presumes opportunistic screening by general practitioners and the use of a series of laboratory tests on blood samples. It says nothing about the setting up of programmes of screening by invitation or the use of desk top analysers.

Many authors have used estimates generated by others in their own league table; did they check the original analyses? In the table the value for continuous ambulatory peritoneal dialysis is not attributable to any of the source studies. The whole table has been reported here and elsewhere.[22] Given the authoritative "halo" effect that such tables evidently acquire, accurate reporting is a most basic requirement.

Finally, the representativeness and quality of contributions to the league table may be questionable. For instance, Teeling Smith's value for antihypertensive treatment to prevent stroke[10] is clearly an outlier, with other analysts giving much higher values.[23] Unfortunately, often it is not possible to comment on a given source study's assumptions as only summary calculations are presented. Studies in which essential information is not available should not be included in league tables.

Some guidelines for using current league tables

Users of league tables should compare their own setting with that for the source studies included. In particular, they should pose three related questions. Firstly, in my setting could the interventions listed be applied with the same likely success as in the time period and setting reflected in the league table? (For example, are the outcomes generalisable to my local need and clinical practice, and are the same service structures in place?) Secondly, are the comparisons embodied in the various estimates of cost effectiveness relevant to my local situation? (For example, if a comparison of a new intervention is made with current practice, minimum practice, or doing nothing do I agree that in my setting the same comparison programme would apply?) Thirdly, are the resource implications and consequent costs similar to those pertaining in my setting? (For example, how does my local availability of health care resources compare with that in the source study?)

It may reduce contention if comparison in a league table is restricted to those interventions targeted at one condition (such as cardiovascular disease) or at one group of patients (such as elderly people). Although, inevitably, broader comparisons need to be made when allocating the whole health care budget, the homogeneity of the underlying studies is undoubtedly greater within more restricted boundaries.

Currently league tables present only point estimates; this conveys a false sense of precision and encourages unwarranted conclusions. Possible changes would include the reporting of sensitivity analyses in league table documents or, as more economic evaluations of health care interventions are undertaken alongside controlled clinical studies,[24] the use of confidence intervals around estimates. Even a methodologically sound table is unlikely to feature estimates with a high degree of precision; they should be viewed in this light.

Cost effectiveness estimates should not be used in a mechanistic fashion; at best they provide a useful aid for decision making. Other factors—for example, equity—legitimately influence decisions. In particular a given league table is unlikely to contain all the relevant comparisons of programmes of the appropriate intensity or scale to enable a budget to be allocated. To

approach this in a more formal sense would also require mathematical programming techniques.[25][26]

Conclusions

Although league tables of costs per QALY may help decision makers in the allocation of resources, potentially they can also mislead. One approach would be to abandon the construction and use of league tables altogether. Whether or not league tables are reported, however, there is a natural tendency for evaluators to compare their study results with those for other health care programmes. Also, there is a need on the part of decision makers to assess the relative value for money from competing health care interventions. In the absence of systematic comparisons, as embodied in league tables, these assessments of relative value for money are likely to take place informally. This will probably have a worse risk-benefit trade off than the formalised use of league tables.

Against this background we have outlined some of the problems with current league tables and made some suggestions for improving their construction and interpretation. Even if our recommendations were adopted by those undertaking economic evaluations and those constructing league tables, however, decision makers should still exercise the appropriate caution, care, and intelligence.

1 Williams AH. Economics of coronary artery bypass grafting. *BMJ* 1985;291: 326-9.
2 Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. *Can Med Assoc J* 1992;146:473-81.
3 Hadorn DC. Setting health care priorities in Oregon: cost-effectiveness meets the rule of rescue. *JAMA* 1991;265:2218-25.
4 Eddy DM. Oregon's methods: did cost-effectiveness analysis fail? *JAMA* 1991;266:2135-41.
5 Drummond MF. Output measurement for resource allocations in health care. *Oxford Review of Economic Policy* 1989;5:59-74.
6 Gerard K, Mooney G. *QALY league tables: three points for concern—goal difference matters*. Aberdeen: Departments of Public Health and Economics, University of Aberdeen, 1992. (Health Economics Research Unit discussion paper 04/92.)
7 Maynard A. Developing the health care market. *The Economic Journal* 1991;101:1277-86.
8 Department of Health. *Blood cholesterol testing: the cost-effectiveness of opportunistic cholesterol testing*. London: Standing Medical Advisory Committee, 1990.
9 Pickard JD, Bailey S, Sanderson H, Rees M, Garfield JS. Step towards cost-benefit analysis of regional neurosurgical care. *BMJ* 1990;301:629-35.
10 Teeling Smith G. The economics of hypertension and stroke. *Am Heart J* 1990;119(suppl):725-8.
11 Department of Health and Social Security. *Breast cancer screening*. London: HMSO, 1990. (Forrest report.)
12 Leese B, Hutton J, Maynard A. *The cost-effectiveness of the use of erythropoietin in the treatment of anaemia arising from chronic renal failure*. York: Centre for Health Economics, University of York, 1990. (Occasional paper.)
13 Drummond MF, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press, 1987.
14 Parsonage M, Neuberger H. Discounting and health benefits. *Health Economics* 1992;1:71-6.
15 Cairns J. Discounting and health benefits: another perspective. *Health Economics* 1992;1:76-9.
16 Kind P, Rosser RM, Williams AH. Valuation of quality of life: some psychometric evidence. In: Jones-Lee MW, ed. *The value of life and safety*. Amsterdam: North-Holland, 1982:159-70.
17 Torrance GW. Measurement of health state utilities for economic appraisal. *Journal of Health Economics* 1986;5:1-30.
18 Torrance GW, Zipursky A. Cost-effectiveness of antepartum prevention of Rh immunization. *Clin Perinatol* 1984;11:267-81.
19 Cox DR, Fitzpatrick R, Fletcher AE, Gore SM, Spiegelhalter DJ, Jones DR. Quality of life assessment: can we keep it simple? *J R Stat Soc A* 1992;155:353-93.
20 Smith R. Where is the wisdom . . . ? The poverty of medical evidence. *BMJ* 1991;303:798-9.
21 Drummond MF, Bloom BS, Carrin G, Hillman AL, Hutchings HC, Knill-Jones RP, et al. Issues in the cross-national assessment of health technology. *Internat J Tech Assess Health Care* 1992;8:671-82.
22 Akehurst R, Godfrey G, et al. *The health of the nation. An economic perspective on target setting*. York: Centre for Health Economics, University of York, 1992. (Discussion paper 92.)
23 Drummond MF, Coyle D. Assessing the economic value of antihypertensive medicines. *J Human Hypertens* (in press).
24 Drummond MF, Davies LM. Economics analysis alongside clinical trials: revisiting the methodological issues. *Internat J Technol Assess Health Care* 1991;7:561-73.
25 Torrance GJ, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programmes. *Health Services Research* 1972;7: 118-33.
26 Birch S, Donaldson C. Applications of cost-benefit analysis to health care: departures from welfare economic theory. *J Health Econ* 1987;6:211-25.