Measuring Outcomes in Palliative Care: Limitations of QALYs and the Road to PalYs

Charles Normand, BA, MA, DPhil, FFPH, FTCD
Centre for Health Policy and Management, School of Medicine, Trinity College Dublin, Dublin, Republic of Ireland

Abstract
Tools for measuring outcomes in health and social care have become key parts of the processes of evaluation and setting priorities. Measures of output that can be used in all settings and specialties have the advantage that they facilitate comparisons and choices between and within patient groups. However, the most commonly used composite measure of outcomes, the quality-adjusted life year (QALY) appears not to work well in complex interventions, such as palliative care, leading to the paradox that there is evidence that people would give priority to interventions and services that would be shown not to be cost-effective, using QALYs as an outcome measure. This article explores the possible reasons for this paradox, and looks at alternative approaches that may provide better tools for setting priorities within palliative care and for comparison of palliative and other care services.

Introduction
Outcome measures aim to meet three, sometimes conflicting objectives—to assess the success of treatment and care of individuals, to help in managing service delivery and monitoring quality, and to support priority setting by making easier the comparison of the effectiveness and cost-effectiveness of different interventions. It has long been recognized that outcomes of treatment are multidimensional and complex, and that, to varying degrees, any single index measure of outcome will be inadequate to capture important differences. However, it also has been argued that full comparability of different programs is greatly facilitated by simple measures of outcome. Economists often have made the case for comparability, because the overall optimization of the use of scarce health care resources requires that all potential interventions be compared with all others. The argument in this article is not against the desirability of outcome measures that allow full comparability of palliative care services and other health and social care, but rather that, in the case of many palliative services, the limitations of standard outcome measures makes the comparisons inappropriate, and tend to bias the use of resources away from services received at the end of life. It is argued that, faced with the complex and multidimensional objectives of palliative care, analysis of outcomes and cost-effectiveness

Key Words
End-of-life care, palliative care, outcomes, QALYs, economic evaluation, priority setting
needs to embrace the complexity, and be willing to accept that the argument for using resources for services near the end of life has to draw on a range of evidence, complex measurement tools, and a good understanding of context. This is not to argue against the need for rigor and clarity in the assessment of outcomes, but rather to argue that it is unhelpful to use partial measures, no matter how much these may facilitate simple comparisons between programs.

Combining Length of Life and Quality of Life in Outcome Measurement

Health service interventions generally have two related objectives— to increase the length of life and to improve health-related quality of life. Some interventions, such as hip replacements, are mainly concerned with reducing pain and increasing quality of life; some, such as heart treatments, hope to increase both length and quality of life; and some, such as serious trauma services or AIDS prevention, are mainly focused on preventing premature death. This range of objectives means that, in most circumstances, there is a need of evaluation to make comparisons of the combinations of improvement in each of the objectives, so that, for convenience, there should be a single system of measurement that combines both dimensions. The most commonly used measure is the quality-adjusted life year (QALY). There has been a lively debate on many aspects of the idea of QALYs and the measurement that goes into their derivation, but the broad approach has widespread acceptance within the health economics community. Although the early debates on QALYs are interesting, and some issues remain unresolved, this debate did not challenge some assumptions and principles that raise concerns about the appropriateness of the use of QALY-type measures in palliative care. The focus of this article is on some of the wider issues about the dimensions of outcomes in palliative care, and whether these can usefully be collapsed into a two-dimensional model. It questions whether it is reasonable in this context to ignore potentially important contextual factors.

The Quality-Adjusted Life Year Problem and Its Relevance to Palliative Care

A useful starting point for considering the applicability of standard outcome measures in palliative care is the small but interesting literature on what has been called the “QALY problem.”

At its simplest, the QALY problem is that there are uses of resources that appeal to public and decision makers’ preferences, but using conventional measurement of cost per QALY would be clearly outside the “cost-effective” range. Although it is generally accepted that measurement of QALYs is never precise, and that it may be wise to consider broad ranges within which the precise value might lie, in the case of some popular palliative interventions, the cost per QALY can lie far outside the range of cost-effective services.

Some authors identify context-specific problems in the use of QALYs. Some largely ignore the problem, which, in effect, is saying that the declared preferences for interventions that fail to achieve conventional cost-effectiveness thresholds represent lack of knowledge or understanding. Others reject the QALY approach altogether because of the limited and narrow base for evaluation provided by QALYs.

The great strength (and a major objective) of the QALY approach is comparability— health gains can be compared across a wide range of diseases and settings. Another advantage is their relative simplicity—time for any individual at any point in time has a constant value, which also has useful properties, such as being additive (i.e., it is possible to add up gains in years and quality of life, and in essence, these have cardinal properties). The difference between good time and less good time is simply the quality weighing, normally between 0 and 1, but formally allowed to be negative. The rules do not allow the quality of time spent as being more than perfect; hence, it is not possible to define any state as better than full quality of life.

It is easy to dismiss concerns, such as the QALY problem, as being special pleading. No system of measurement that aims to allow comparisons is ever perfect, and it is common to find those who do not like the outcome of
a set of rules or procedures arguing for the rules to be changed. Individuals and groups with interests in gaining access to services may not be very concerned about the overall efficiency in the use of health care resources, and have every reason to exaggerate the scale of benefits where they believe that reported benefits will increase the chance that an intervention will be provided. For users of service, the question is not whether a service is cost-effective, but rather, whether the benefits to them exceed the costs to them.

The QALY problem has been discussed mainly in the context of complex services, such as palliative care and in mental health. These tend to be holistic in intention, and intended outcomes tend to be multidimensional and complex. This may be the reason why the somewhat reductionist approach in the derivation of QALYs causes problems.

If we accept that there is a problem, then there are two possible explanations. First, it may be that there is nothing wrong with the measures used, but simply that some important benefits have not been measured. For example, we can argue that there are caring externalities, so that someone who gains from palliative care also increases the welfare of friends and relatives. If these caring externalities are sufficiently great, then including them might make the service conventionally cost-effective. Another possibility is that there are changes in quality of life that happen outside the observed time window. In this case, we might gain quality of life across our lifetime from the knowledge that we will receive effective and appropriate end-of-life care, and this might justify "loss-making" treatments.

The second and potentially more serious alternative is that there is a problem with adding up periods of time, even after adjusting along the quality dimension, because time is simply not additive. QALYs assume that time in a health state can be valued regardless of the context, but there may be circumstances where people put more or less value on time. We can think of simple examples. Time spent with a young child or time on a holiday may be regarded as having special value, and this may not be reflected in stated differences in the quality of life in any scoring system. Put another way, we can look at the quality of life in any given period, and score it on a scale such as 0–1, but the value put on any given period of time could also be different. This has serious implications, because this potentially undermines the use of QALY-type metrics across all health care settings and issues. In adding up QALYs, we are already assuming that we can add up benefits to different individuals, which violates some theorems in welfare economics, but it is more serious if periods of time cannot be added up at different points in time for individuals. The extent to which nonadditive time is a general problem for the use of QALYs may depend on the nature of particular health needs and interventions. The assumption of additive time may be acceptable in most circumstances, especially where the intervention has quite simple objectives, and these are reasonably assessed in terms of changes in quality-adjusted time.

Studies on the cost-effectiveness of palliative interventions suggest that the QALY problem does exist in this context and that alternative or additional measurement systems may be needed both to compare the uses of resources within palliative care and between palliative and other interventions. Drawing on experience in the evaluation of palliative care, there are several possible ways to proceed.

The Needs for Outcome Measures and Measurement in Palliative Care

There are two needs for economic evaluation of palliative interventions—to allow comparison between different palliative services for different needs, and to compare palliative interventions to others within health and social care (and potentially beyond). In the case of making comparisons outside palliative care, the need is for a common metric that covers all types of interventions. If the QALY problem is accepted, then the most likely way forward is a modified QALY process that allows in some way for the "problem." There would be a need to accept additional information and dimensions into any yardstick to provide a fair comparison between the different types of care. Developing the Palliative Care Yardstick (PaLY) would start with the normal QALY framework, but then add in items that may not be adequately measured. This might, for example, involve counting caring
externalities and benefits in the form of anticipation and reassurance that end-of-life care will be forthcoming. Some de facto acceptance of nonadditivity of time might be made by allowing a value to be put on components of a “good death,” which would be, in some way, separate from the days that led up to it. The PalY process could be more or less formal, and would have the fairly simple objective of providing a clearer framework for ensuring that the declared higher valuation by citizens for end-of-life interventions found expression alongside the more linear measurement in the use of QALYs. It might remain a process rather than a formal tool, and could be allowed to evolve into more formal metrics with use and experience.

Tools to allow comparisons within different aspects of palliative care do not, in principle, have to have the features of QALYs. An obvious starting point is to look at the existing generic and context-specific outcome tools. Generic quality-of-life measures, such as the EuroQol EQ-5D, do not really try to measure difference well within a particular disease or intervention category, but aim instead to provide the basis for comparing across categories. Researchers often choose to use EQ-5D alongside scales and measures that provide more sensitive measurement suitable for a particular disease or pattern of treatment. If the aim is to make better comparisons between palliative interventions, it is not obvious that there is much to learn from the use of such tools. Scales designed specifically for use in assessing outcomes in palliative care, such as the Palliative Care Outcome Scale, offer more insight into the dimensions of benefit that are of value to patients and families, and can give significant insight into what matters to recipients of care and their families. There has never been an intention that such scales would be used to derive single index measures of outcome, but they can, nevertheless, provide an input into understanding how the services are valued by the users.

There is an additional problem in evaluating palliative interventions—evidence suggests that benefits may vary between people who are apparently in the same circumstances. Where people have choice of which services to use, there is large variation that cannot be explained by disease and symptoms. Work on the evaluation of palliative day care suggests that users may place value on availability of services, even though they do not choose to use them. Outcome measurement may, therefore, need to include benefits from availability of services that are not in fact used. It may be in the nature of the period at the end of life that expressed preferences tend to be very strong. Evidence on day-care preferences suggests that users have almost lexicographic preferences. Put simply, they are only really interested in considering some questions when they have resolved others, and there are components that they are not willing to trade-off. For example, they do put a value on services such as massage and hair dressing, but are only interested in discussing such services if they are sure that good access to clinical care is assured. Economics has problems in handling such preferences, because they mean that utility and welfare functions are not well behaved and continuous. If the benefits of services are entirely contingent on the availability of a quite different service, it may be more sensible to consider only packages rather than individual components of care in any evaluation.

The measurement of outcomes remains very difficult, and at least in the short run, the best option might be not to try too hard, but rather to develop frameworks that facilitate direct comparisons of the different possible priorities. If direct measurement of outcomes is very difficult, but people have a good understanding of the experiences of individuals in receipt of the different services, it may be easier to make direct comparisons of costs and the different patient trajectories. This is analogous to some simplified forms of economic evaluation, where the question is not “Is this worth doing?” but rather “Which of these options is best value?” Some progress using this approach has been made in the use of choice experiments in the evaluation of palliative day care. Different ways of allocating scarce resources for palliative care were offered within a discrete choice framework, and respondents were able to make hypothetical choices that allowed the researchers to put a value on different aspects of palliative day care. It is clear that the tasks set were well understood by people in receipt of palliative care, and that they were happy to take part in the research. Service users have a diversity of preferences, and
overall, the key characteristics desired by users were surprising to some service providers. In particular, users were keen on the availability of access to specialist symptom management services and were not willing to trade these.

Eventually, it might be possible to use findings of such studies to provide better understanding of the components of palliative care that would allow the use of more conventional economic evaluation techniques.

**Conclusions**

Essentially, the best way to proceed with measurement of outcomes in palliative care needs to be pragmatic. Is the main problem that we are not making good judgments between palliative care and other health and social care, or is it that we do not have optimal use of palliative care resources? If it is the former, we need to find ways of combining the ability to make comparisons across areas of service with very different dimensions of outcome while ensuring that the measures used are appropriate to each context. If it is the latter, then it is not obvious that current standard methods of economic evaluation are currently well suited to the task, and that identifying the components of care packages that are most useful to users is more important. Stated preference techniques (including discrete choice experiments) offer a well-understood framework for selecting the components of care that are of most value to users. While aspiring to a more formal framework for comparison, unless the tools are fit for purpose, it may be better to accept more limited comparability but to use techniques that are more sensitive for measurement. The point is not to use bad versions of the perfect approach when better tools (that may not use quite the right approach) may improve resource allocation within palliative care.

**References**


