Current practice in obtaining the "Q" in QALYs: a cautionary note

Roy A Carr-Hill, Jenny Morris

The economic evaluation of benefit from a course of treatment or an intervention depends on being able to measure the outcomes. One method currently popular with health economists and being considered by some health service managers is the quality adjusted life year (QALY).

The QALY is a cardinal number representing a unit of benefit which combines a measure of life expectancy with a measure of quality of life. One year of current life in perfect health is counted as one QALY, and one year of current life with a lower level of health, "q," (where $q < 1$) is counted as q QALYs; the value of a future year of life is adjusted to a current value by discounting at rate "r." The health profile of (prospective) patients over their expected future life span can be used to calculate the current value of their expected remaining QALYs and comparisons can then be made between different types of treatment on the costs per QALY gained. The assumptions underlying this procedure have been examined.¹

To determine the value of q a method is needed that provides a single value to represent the health related quality of life of a variety of patient groups and that is easy to administer in a clinical setting. Generic instruments providing profiles such as the Nottingham health profile² and the sickness impact profile³ are therefore unsuitable and the quality of well-being scale⁴ is lengthy and interviewer led. One suitable

instrument is the two dimensional classification system first described by Rosser and Watts (box),⁵ and subsequently used by Rosser and Kind to obtain valuations of the 29 health states derived from combining the disability and distress levels.⁶

This classification system and the matrix of valuations derived from the data reported by Rosser and Kind⁶ have been at the forefront of the work undertaken in the United Kingdom on the use of QALYs. A questionnaire (the York quality of life) has been developed for obtaining data on health related quality of life that can be reclassified into the Rosser-Watts system to calculate QALY values.⁷ Although no data on the validity and reliability of this questionnaire have been published, it has been recently used as the quality of life measure in at least one cost effectiveness study⁸ and one clinical trial.⁹ The authors of the clinical trial refer to the York quality of life as a "standard quality of life questionnaire," which suggests consensual agreement via peer review, although the only source document is an inhouse discussion paper.⁷

If this instrument is to be used in further trials before evidence is published of validity, reliability, and interrater reliability for the scoring system, potential users need to consider the content and mode of completion of the questionnaire and the derivation of the disability and distress scores for conversion into the Rosser-Watts classification system. We shall now examine these two issues.

Rosser-Watts classification of illness states

DISABILITY

- 1 No disability
- 2 Slight social disability
- 3 Severe social disability or slight impairment of performance at work, or both
Able to do all housework except very heavy tasks
- 4 Choice of work or performance at work severely limited
Housewives and old people able to do only light housework but can go out shopping
- 5 Unable to undertake any paid employment
Unable to continue any education
Old people confined to home except for escorted outings and short walks and unable to do shopping
Housewives able to perform only a few simple tasks
- 6 Confined to chair or able to move around in the house only with support from an assistant
- 7 Confined to bed
- 8 Unconscious

DISTRESS

- A None
- B Mild
- C Moderate
- D Severe

York quality of life questionnaire

The questionnaire is in two short parts and has been designed for self completion. The first, which is used to determine the Rosser-Watts disability rating, includes a self rating of general mobility (from no difficulty to bedridden); a rating of four categories of self care (washing, dressing, eating, using the toilet); a rating of whether or not usual activity is affected; and a rating of the extent to which social and personal relationships (social life, seeing friends or relations, hobbies or leisure activities, sex life) are affected.

The second part is used to derive the Rosser-Watts distress rating and asks respondents to rate 24 feelings firstly by ticking a box to indicate whether any of the feelings have been experienced and, secondly, by indicating how much distress has been caused by such feelings by marking a point on a 10 cm line anchored by "no distress at all" and "extreme distress." Scores from the ratings of the first 13 feelings are used to reclassify the data into the Rosser-Watts distress categories. The 13 feelings are feeling sad or depressed, feeling anxious or worried, pain, dissatisfaction with your weight, dissatisfaction with your appearance, embarrassment, uncertainty about the future, anger or resentment, guilt, loneliness, loss of self confidence, feeling dependent on other people, and feeling dependent on a machine for your health.

No information is provided about the origin, reliability, and validity of the items which comprise the

Centre for Health Economics, University of York, York YO1 5DD
Roy A Carr-Hill, DPHIL,
senior research fellow
Jenny Morris, PHD, research fellow

Correspondence to:
Dr Carr-Hill.

BMJ 1991;303:699-701

questionnaire, and it is unclear why certain items are seen as distinct from one another. For example, in the first part of the questionnaire under the heading, "social and personal relationships" it is not clear why "social life" is distinct from "seeing friends or relatives" and "your hobbies or leisure activities." Despite this the algorithm assigns one point for a yes to each of the items; similar assignment rules are used for the other dimensions (general mobility, self care, and usual activity). The dimension scores are then combined into disability categories according to a conversion table wherein the general mobility dimension takes the dominant role.⁷

Furthermore, though the form is designed for self completion, the respondent is required to move from the relatively simple task of ticking one of two boxes to a visual analogue scale without warning. Though this may not cause problems if an interviewer is present to explain what is required, such a format is likely to result in missing data if left for the patient to complete without instruction.

The assignment rules allow patients reporting different levels of disability on the York quality of life questionnaire to be reclassified into the same Rosser-Watts disability category. For example, patients would be assigned to Rosser-Watts disability category 3 (that is, severe social disability or slight impairment of performance at work, or both and able to do all housework except very heavy tasks) if their response to the York quality of life questionnaire was either (a) some difficulty on all four categories of self care or the four categories of social and personal relationships but usual activity only slightly affected and no difficulty in moving around or (b) no problems in self care, usual activity, and social and personal relationships but having to use a walking aid. Though example (a) corresponds well with disability category 3 it is not obvious why example (b) is categorised as such and, indeed, it is not clear how such a patient would fit into the Rosser-Watts classification system (box).

External validation

As indicated above the Rosser-Watts distress categories are derived from 13 visual analogue scales, each of which are scored from 0 to 100 and then combined according to the decision rules shown in table I. These rules are not explained or justified and could lead to some questionable classifications that would need to be verified by some method of criterion validity. For example, arthritic patients in severe pain are likely to give a rating of extreme distress on the pain scale, which would be scored between 90 and 100; this

TABLE I—Conversion table for distress categories^a

York quality of life distress scores	Rosser-Watts distress category
All scores ≤30	A No distress
Highest score 31-60	B Mild distress
Highest score 61-90	C Moderate distress
Highest score >90	D Severe distress

TABLE II—Correlations between components of York quality of life and other measures of health

	York quality of life		Nottingham health profile	General health questionnaire
	Disability	Distress		
Nottingham health profile	30**	10		
General health questionnaire	28**	29**		
Long standing illness	39**	27**	29**	12**
Admission to hospital	3	10**	20**	-6
Social class	15**	6	23**	6
Smoking	6	9*	6	3
Drinking	-4	-3	-5	0

*p<0.05, **p<0.01.

score would put them in category D (severe distress) even if they claimed no distress for the other 12 scales. In contrast, another patient group reporting some distress on all 13 scales, each being scored between 50 and 60, would be classified as category B (mild distress).

Though we accept that establishing external validity is difficult, the source data were collected as part of a study that deployed a wide range of measures of health status such as the general household survey measure of long standing illness and hospitalisation together with the Nottingham health profile² and the 12 item general health questionnaire.¹⁰ Table II shows the correlations between these more straightforward measures of health, the York quality of life questionnaire, social class, and indices of smoking and drinking behaviour.

The results from the disability scale seem to follow the same kind of pattern as the Nottingham health profile, whereas the distress score is nearest to the general health questionnaire; moreover, the disability and distress scales are intercorrelated ($r=0.4$, $p<0.001$). Of course, it is perfectly acceptable for the components or dimensions of a unitary scale to be intercorrelated, but the implications should be discussed.

Though these interrelations are interesting and plausible, the development and use of any new instrument (the Nottingham health profile and general health questionnaire are well established) requires more justification. This is especially important as the instrument might be used by people unfamiliar with the rigours of questionnaire design and research methods. Indeed, though the York quality of life questionnaire is designed to obtain data for inputting to the Rosser-Watts classification, the same general caveats about pilot testing and validation apply to any proposed procedure for obtaining the values of "q."

Health warning

Dean recently reported that 28 out of 76 district health authorities were considering using QALYs, 37 were not, and 11 were undecided.¹¹ In the United Kingdom, the usual tool for calculating QALYs is the matrix derived from the valuations of the health states first described by Rosser and Watts.⁵ The York quality of life questionnaire is proposed as a means of collecting health related data on quality of life from patients (rather than using clinicians' judgment) which can be reclassified into Rosser and Watts's system. Though development work of this kind is laudable, using an unvalidated instrument could have serious and far reaching consequences.

Although the comments above apply directly to the York quality of life questionnaire the same issues apply to the design of all questionnaires. Rigorous validation of health related quality of life measures is complicated, but using untested and unvalidated instruments in clinical trials and cost utility analyses (the results from which potentially aid decision making about allocation of resources) must be strongly discouraged.

We thank Kerry Atkinson for secretarial help and the respondents to the Wolverhampton health needs survey for providing the basic data. RCH was supported by the Economic and Social Sciences Research Council.

- 1 Carr-Hill R. Background material for the workshop on QALYs: assumptions of the QALY procedure. *Soc Sci Med* 1989;29:469-77.
- 2 Hunt S, McKenna S, McEwen J, Beckett E, Williams J, Papp E. A quantitative approach to perceived health status: a validation study. *J Epidemiol Community Health* 1980;34:281-6.
- 3 Bergner M, Bobbitt R, Carter W, Gilson B. The sickness impact profile: development and final revision of a health status measure. *Med Care* 1981;19:787-805.
- 4 Kaplan R, Anderson J. A general health policy model: update and applications. *Health Serv Res* 1988;23:203-35.
- 5 Rosser R, Watts V. The measurement of hospital output. *Int J Epidemiol* 1972;1:361-8.

- 6 Rosser R, Kind P. A scale of valuations of states of illness: is there a social consensus? *Int J Epidemiol* 1978;7:347-58.
- 7 Gudex C, Kind P. *The QALY toolkit*. York: Centre for Health Economics, University of York, 1988. (Discussion paper No 38.)
- 8 Leese B, Hutton J, Maynard A, eds. *The costs and benefits of the use of erythropoietin in the treatment of anaemia arising from chronic renal failure: a European study*. York: Centre for Health Economics, University of York, 1990.
- 9 Dixon AK, Southern JP, Teale A, Freer C, Hall L, Williams A, et al. Magnetic

resonance imaging of the head and spine: effective for the clinicians or patient? *BMJ* 1991;302:79-82.

- 10 Goldberg D. *Detection of psychiatric illness by questionnaire*. Oxford: Oxford University Press, 1972.

- 11 Dean M. Is your treatment economic, effective, efficient? *Lancet* 1991;337:480-1.

(Accepted 28 June 1991)

The Health of the Nation: responses

Coronary heart disease

Hugh Tunstall-Pedoe

The case for coronary heart disease being a key area

The Secretary of State for Health's consultative document *The Health of the Nation* has appropriately listed coronary heart disease first among 16 suggested key areas.¹ It is difficult to envisage a health strategy for England without including coronary heart disease at the top. The document's primary criteria for designating key areas are that they are major causes of premature death, of avoidable ill health, and of economic cost to the NHS and the community. Several diseases rank highly for one criterion alone, but coronary heart disease, the major contributor to cardiovascular disease, is at or near the top in all three. It is the leading cause of death in both sexes and is estimated to cost annually £500m in treatment and £1800m in lost production, besides having accounted for 11.6% of sick leave, in the mid-1980s.²

Other criteria are that the disease should be one in which effective interventions are possible, objectives and targets can be set, and progress can be monitored by means of indicators. Again, coronary heart disease is the chronic disease on which the greatest volume of research has been done: into causation and risk factors (prevention or risk reduction), as well as into monitoring trends in risk factors and the progression of the disease in cohort studies, randomised groups, and geographically defined populations.

A final criterion, for those concerned with reducing inequality, is that mortality from coronary heart disease shows large differences between nations (England scores badly), between English regions and districts (bad for the north and west and the traditional smoke stack areas), between different occupational social classes (bad for manual workers), between ethnic groups (bad for south Asians), and between the sexes (men die earlier and women suffer disproportionate bereavement thereby). Effective coronary prevention has great potential for reducing some of these differences.

What is the case against?

Many of the arguments against making coronary heart disease a key area for a national strategy for health would be those against having such a strategy at all and will not be discussed specifically. Over the past 10 years the case for England not joining other nations in designating coronary heart disease as a key target for a national strategy has been weakened by what has happened elsewhere. Other countries have got their act together and achieved a national consensus for action through collaboration among government agencies, professional bodies, and voluntary organisations. And mortality has fallen even faster. Britain seems to have been acting as the control for other English speaking

nations. Although Britain has produced some outstanding researchers and teachers in cardiovascular epidemiology and prevention, there has been national inertia on prevention. This has been justified by the mistaken argument that as long as one or more academics argue against the prevalent hypotheses, that there are two sides to the argument, the experts therefore disagree and it is safer, wiser, and cheaper to do nothing. The decline in coronary heart disease mortality in Britain has been delayed.

The association between these phenomena for Britain versus several European and English speaking countries is undoubted; the causal relation is not entirely straightforward. A declining mortality and the changing lifestyles and social attitudes against which it is occurring are far more conducive to a prevention programme than one in which mortality is static or rising. The comparative success of the Belgian versus the United Kingdom heart disease prevention project in the early 1970s is usually ascribed to more intensive intervention in the Belgian study and is correlated with the greater overall change in risk factors.³ In fact, the greatest study induced change in self reported cigarette smoking occurred in factories in Britain, where stopping smoking was well established. By contrast, there was virtually no impact of dietary counselling on blood cholesterol concentration in the British factories. It had no support or echoes in the men's previous knowledge or everyday experience. Perhaps its time had not yet come.

Coronary prevention therefore faces a paradox. When the disease is still rising or at its peak the task may be difficult and not widely accepted. When mortality and risk factor levels are declining already the purist could (and some do) argue that no purpose is served by interfering further. A cynic could claim that the guaranteed success of a prevention programme makes it politically desirable. A realist would claim that accelerating and generalising a beneficial trend may be the most cost effective point for concerted action. It may also be true that the general population is prepared to take greater efforts to avoid a disease that is seen to be declining and associated with life's losers than when it is claimed to be the price of a modern lifestyle and executive success.

Another argument used against making coronary heart disease a specific target is that a health strategy should concentrate on health rather than specific diseases. Though it is true that diseases cannot be considered in isolation, the health argument at worst can become metaphysical and untestable. Unless major causes of mortality and morbidity are improved by a health strategy it cannot be considered to be successful. Coronary heart disease should be an identified target but integrated, as in this document, with other diseases, and with lifestyles and risk factors of general relevance.

Cardiovascular
Epidemiology Unit,
University of Dundee,
Ninewells Hospital and
Medical School, Dundee
DD1 9SY
Hugh Tunstall-Pedoe, FRCP,
professor and director

BMJ 1991;303:701-4