

THE UNIVERSITY *of York*

CENTRE FOR HEALTH ECONOMICS
YORK HEALTH ECONOMICS CONSORTIUM
NHS CENTRE FOR REVIEWS & DISSEMINATION

*The Role of the EuroQol Instrument
in QALY Calculations*

Alan Williams

DISCUSSION PAPER 130

THE ROLE OF THE EUROQOL INSTRUMENT IN QALY CALCULATIONS

Alan Williams

March 1995

The Author

Alan Williams is Professor of Economics at the University of York, and founder of the EuroQol Group.

Acknowledgements

This is a slightly amended version of a paper given at a Quality-of-Life Workshop sponsored by the ESRC and the SOHHD, and held in Edinburgh in April 1993.

Further Copies

Further copies of this document are available (at price £5.00 to cover the cost of publication, postage and packing) from:

The Publications Secretary
Centre for Health Economics
University of York
Heslington
York, YO1 5DD.

Please make cheques payable to the University of York. Details of other papers can be obtained from the same address, or telephone York (01904) 433648 or 433666.

ABSTRACT

The EuroQol measure of health-related quality-of-life (HRQOL) has many purposes other than the calculation of QALYs, but it is upon that particular use that this paper concentrates. For that purpose a simple generic measure yielding a single index score for each health state is required. The development of any HRQOL measure requires some key strategic simplifications to be made. In the case of the EuroQol measure these are that a profile of health states can be divided into separate time segments, and that the value of being in any one state is independent of the states which precede or succeed it. Further simplification is required to keep the range of dimensions of HRQOL manageable. When it comes to the number of levels within a dimension, discrimination requires a large number, but practicability requires a small number.

The EuroQol Group wished from the outset to generate a measure which reflected the valuations of the general public, rather than the views of health care professionals. Thus the valuation process was as important as the descriptive content. For this purpose the instrument had to be capable of being used to elicit valuations in a postal questionnaire. This dictated the use of a rating scale in the form of a "thermometer" as the basic method. But more recent interview-based work has used the time-trade-off method as well.

The contribution of the EuroQol instrument to the measurement of HRQOL is:

- (1) as a simple way to generate descriptive data;
- (2) as a simple way to elicit people's own rating of their current health states;
- (3) as a preference based generic index of health benefits (for use alongside more detailed condition-specific or treatment-specific measures).

PREAMBLE

At the outset it has to be stressed that the Euroqol[®] instrument has other roles than the calculation of QALYs, and that there are ways of calculating QALYs which do not involve the use of the Euroqol[®] instrument, so we are here considering what is, in effect, the area of intersection of two much larger sets, each of which has its own *raison d'être*. Measures of health-related quality-of-life have an important role to play in the monitoring of treatment, in clinical audit, in clinical trials, and in measuring the health status of populations. Measures such as the EuroQol can be used to generate data for all of these purposes, quite independently of whether a QALY approach is adopted or not.

The *raison d'être* of the Euroqol[®] instrument was to provide a simple "abstracting" device, for use alongside other more detailed measures of health-related Quality of Life (henceforth QOL), to serve as a basis for comparing health care outcomes using a basic "common core" of QOL characteristics which most people are known to value highly. From the outset it was accepted that for such comparisons to be useful it would be necessary to go beyond generating such information in the form of a "profile" (though the Euroqol[®] data can be used in that way too), and therefore the issue of the relative valuation of different health states had to be confronted. It was further recognised that such information would be extremely valuable in a QALY-type context, but for that purpose it would be necessary to include a valuation for the state of being dead, otherwise it would be impossible to establish a cardinal index scale in which "dead" = 0 and "healthy" = 1, a property that

is required for QALY-type calculations.

The raison d'être of QALYs is to provide a measure of the benefits of health care that can be used in the economic evaluation of health care, where comparisons across a wide variety of conditions and/or treatments and/or patients and/or settings are involved. In such cases a simple, versatile, generic measure is required which is capable of being reduced to a single index number which has interval properties. This is the situation on which I will concentrate.

RESEARCH STRATEGY

People value both improvements in life expectancy and improvements in the quality of their lives, therefore any single index of the benefits of health care has to incorporate both. This is the fundamental justification for the QALY approach. Ideally we would like to know how every individual values every possible prospective time profile of health-related QOL, including the probabilities associated with each component in them. It need hardly be said that such an undertaking is impossible, so some strategic decisions have to be made concerning simplification of this research task. It is impossible to tell a priori which simplifications are best, so we are inevitably in the realm of intuition and scholarly judgement. It would therefore be most unwise for the research community to pursue only one strategy. What is called for is a broad range of different approaches, which need to be periodically reviewed to see what seems to be working and what does not.

THE DESCRIPTION OF HEALTH PROSPECTS

The particular simplifications involved in the approach adopted by the Euroqol[®] Group were as follows:

- (a) Each time profile of prospective health states is divided into separate time segments, such that within each segment the individual enjoyed the same QOL.
- (b) Initially it was to be assumed that the relative values attached to the different states were independent of the states that preceded or succeeded them, and of the length of time spent in each state.

These two assumptions led the group to reject the scenario approach to health state description, in favour of using composite health states set out in a standardised manner. It was anticipated that at a later stage assumption (b) could be relaxed, but in the meantime a common time duration was to be used (one year), with the subject told that what happened thereafter was not known and should not be taken into account. The relaxation of this assumption is now under way within the group, with experimental work currently in progress on the effect of varying duration and contextual states.

A second set of simplifications is required concerning the detailed description of the health states themselves. There is an understandable tendency in this kind of

enterprise to include everything that might be of any interest to anyone, and to work with fine enough gradations of "severity" within each "dimension" of QOL to pick up any effects of health care treatment that might be of interest to a discriminating practitioner. It is important here to recall that a ten-dimensional classification, with ten levels within each dimension, yields a classification system with 10^{10} , or 10,000 million different cells. Such detail is quite inappropriate in an abstracting device, and if there is to be a reasonable prospect that such summary data is to be collected readily, it must be very simple to collect. If, for some other purpose, more detailed data is required, it should be collected by an instrument designed for that specific purpose. Thus the general advice offered to prospective users is to use the (very simple) Euroqol[©] instrument alongside (not instead of) a more detailed specific measure, and, at this developmental stage in QOL measurement, preferably also alongside some more comprehensive generic measure (possibly one using the profile approach). This is now generating "calibration" data enabling systematic comparisons to be made between outcomes as measured by the Euroqol[©] instrument and outcomes as measured by other instruments.

The actual choice of descriptive content in the Euroqol[©] instrument originated from a review of existing instruments, the material so culled being reduced to manageable proportions through discussion between the original members of the group, who ranged across many disciplines and who drew collectively on a wide range of experience. Later there was an opportunity to test these judgements against the results of a survey of lay concepts of health, which suggested that a dimension of "energy-tiredness" ought to be added to the original 6 dimensions. To accommodate this, two

existing dimensions (one concerned with work activities and the other with other activities) were fused, since it had transpired that the other activities dimension added little to the overall valuations of states. It turned out, however, that "energy/tiredness" made little contribution to health state valuations, and in the pursuit of parsimony it was therefore dropped. This left us with the present 5-dimensional set, in which each dimension has 3 levels of "severity", generating 243 different cells [See page 2 of the Euroqol[®] Questionnaire, appended as Annexe A]. To these need to be added "unconscious" (because it cannot be regarded as a "composite" of the five dimensions) and "dead" (because it is required as a calibration point on the 0 to 1 scale), making 245 states in all.

THE VALUATION OF HEALTH PROSPECTS

Whilst in the process of establishing a workable descriptive system, the group had also been devoting a great deal of attention to valuation issues. Early on it had been agreed that relative valuations should be sought for composite (multi-dimensional) states, not for each dimension separately. This important decision complicated the valuation task, because it involved rejection of the multi-attribute utility scaling approach, in case there proved to be significant interaction between the dimensions. But since no one subject could be expected to value more than a dozen or so states, this meant that the choice of states to be valued had to be made in such a way that, if necessary, it would be possible to estimate the values of all the other states from that limited number of observations. Thus a standard minimum set of (14) states was chosen which were to be used in all valuation work by all members, though where

possible members were encouraged also to elicit values for a more extended set. From these states (plus a value for "being dead") the whole valuation space needs to be estimated. This is a task with which we are still experimenting with a variety of estimation techniques to see which uses our data most fully and produces the best fit.

For practical reasons the Euroqol^o Group imposed upon itself a very restrictive condition concerning the main body of data collection on relative valuations, namely that the questionnaire design should be so simple that it could be self-completed and conducted by a postal survey. This was essentially because none of us had research funds sizeable enough for any other alternative to be feasible, given that we wanted valuations from a general public, not from convenience samples. We quickly agreed that the only valuation method that would be practicable in that context was the visual analogue scale (VAS), and for this purpose we adopted a thermostat-like scale, the current form of which [see Annexe A] is the result of considerable experimentation (using shorter, longer, differently calibrated, differently labelled, and differently orientated versions). The associated problem, which was also subjected to a fair amount of empirical testing, was the layout within the questionnaire of the states to be valued. A complication here was that in order to standardise "framing" effects we had to repeat some states on subsequent pages, thereby reducing the number of observations available to us. But the most difficult issue with the visual analogue scale has been getting people to value the state of "being dead" alongside the other states, and this is still an active area of experimentation within the group.

Lately, with more research funds being devoted to this kind of research, it has

been possible to generate valuations for Euroqol^c states using valuation methods other than the VAS, and in particular comparing those valuations with ones derived from the Standard-Gamble (SG) and the Time-Trade-Off (TTO) methods. From preliminary work, there appears to be a systematic relationship between the VAS and each of the other two methods. It further appears, from DH-supported work done at York, that the TTO method yields somewhat better quality data than the SG method, if quality is judged by the internal consistency of the answers given by respondents, the sensitivity of valuations to parameters known to influence them, and the reliability of the responses when the valuation task is repeated by the same respondents some weeks later. For this reason the TTO method has been used alongside the VAS (Thermometer) in the latest round of our own work. But this is interviewer-generated data, and much more expensive to collect than data from postal questionnaires, so the latter is likely to remain the predominant method for getting large-scale data within the Euroqol^c group generally.

QALYs AND THE EUROQOL INSTRUMENT

The Euroqol^c instrument has two distinct contributions to make to the QALY enterprise. First, it offers a very convenient way of collecting descriptive data about QOL, and about people's own self-rating of their current health state (by using only pages 2 and 3 of the questionnaire in Annexe A). This descriptive data is needed if we are to fill the gaps in our knowledge about the QOL sequelae of many common health care activities. The second, and much more ambitious, role is that of supplying a tariff of social values of health states, to be used (alongside cost data) in a planning

context when determining priorities for health care.

Concerning the descriptive role of the data which can be collected using ONLY pages 2 and 3 of the Euroqol[®] questionnaire, this produces (from page 2) a simple description by patients of their own health-related quality of life, which can be used as a descriptive profile, either through time for a particular patient, or cross-sectionally comparing outcomes for different patients. Page 3 generates a self-rating of a patient's current health-related quality of life, which is probably more safely used in the former manner than in the latter manner, since optimistic people may rate all states higher than pessimistic people do, but each will rate the direction of change in their health accurately. But for the purpose of QALY calculation the descriptive material generated on page 2 needs to have applied to it a tariff representing the social valuation of the state in question, and this is the situation on which I am concentrating here.

This brings us to the second of the two contributions to be made by the Euroqol[®] instrument, which raises some further questions, not all of which are resolvable by research, though research may nevertheless help to clarify the issues involved. I will ignore here the objections people raise against the QALY approach in general (eg on grounds of ethics, equity, etc), not because they are unimportant or unanswerable, but because they are not unique to the relationship between Euroqol[®] and QALYs. I have set out my views on these matters in an earlier Discussion Paper (Number 121, 'Economics, QALYs and medical ethics. A Health Economist's Perspective'). Among the issues that I have set aside here are political questions about whose values should count, and how a set of social valuations is to be derived from

individual valuations. The implicit position of the Euroqol^c Group is that the views of the general public must be given some weight in the valuation of health outcomes, so it is important to know what they are. The Euroqol^c method could, however, be applied to any other group whose views were thought to be relevant in a policy-making context. As to the derivation of social values from individual values, the group has decided that its data should be reported giving arithmetic means and standard deviations on the one hand, and medians and interquartile ranges on the other, so that people can choose whichever measures of central tendency and of dispersion that they think most appropriate to the circumstances. More delicate weighting systems based on the socio-demographic characteristics of respondents would, of course, be possible in principle, as would equity-weighted valuations, but the Euroqol^c Group as such has understandably refrained from entering that territory.

Rather more central to the concerns of this paper are two other questions:

- (i) how valid and reliable are these measures of QOL?
- (ii) how can we be sure that we have generated a scale with the required interval properties?

Both are extremely difficult to answer.

VALIDITY AND RELIABILITY

In general, "validity" means establishing that your measure measures what it

purports to measure. But since there is no "gold standard" for the measurement of health-related quality of life, this seems an unanswerable question. So what people fall back on are appeals to plausibility, for instance whether the measure contains the kind of elements that we would expect such a measure to have, whether it goes up when we would expect it to go up and down when we would expect it to go down, and so on. These are all very subjective notions, and ultimately rely heavily on intuition and professional judgement, which is where we started! We know (i) that Euroqol^o self-rated health on the VAS declines (as expected) with age, (ii) that whilst pain and discomfort increase, anxiety and depression decrease, and (iii) that people's valuations are affected both by age and by experience of illness. My own personal view is that searching for "validity" in this field, at this stage in the history of QOL measurement, is like chasing a will o' the wisp, and probably equally unproductive. It would, however, be useful to find out whether the values elicited from particular individuals in particular circumstances are consistent with their actual behaviour when they are put in a situation in which those values should have been crucial. The devising, conduct and interpretation of the results of such a study would be a valuable contribution to this otherwise rather murky area.

Reliability refers to the issue as to whether the values elicited from an individual are stable, which is usually tested by (surreptitiously?) repeating a question at different stages within an interview, or by going back to the individual a short while later in order to see whether on the second occasion the same answers are obtained as on the first occasion. Within the context of a short self-completed questionnaire the former method does not seem appropriate, and if (as they usually are)

questionnaires are returned anonymously, the latter method is not possible. But in recent interviewer-led survey work we have been able to use the latter method, and it appears that the valuations elicited in that context are stable and reliable.

MEASUREMENT PROPERTIES

Establishing the measurement properties of the scale is perhaps a more tractable problem, but nevertheless quite difficult to resolve. It has been argued, for instance, by one member of the Euroqol^c group that the use of the visual analogue scale tends to compress valuations at one end of the scale (Nord, 1991b). On the other hand it appears that if the Euroqol^c data is treated purely as ordinal, and analysed as pairwise comparisons using a measurement model (Thurstone's) which is known to generate an interval scale, it yields a scale which is very close to the scale obtained when the VAS data are interpreted as cardinal data (see Figure 1). If it were not too burdensome for respondents, the matter might be resolved by asking for comparisons of magnitudes of intervals, as well as for simple ratings.

CHOICE OF VALUATION METHOD

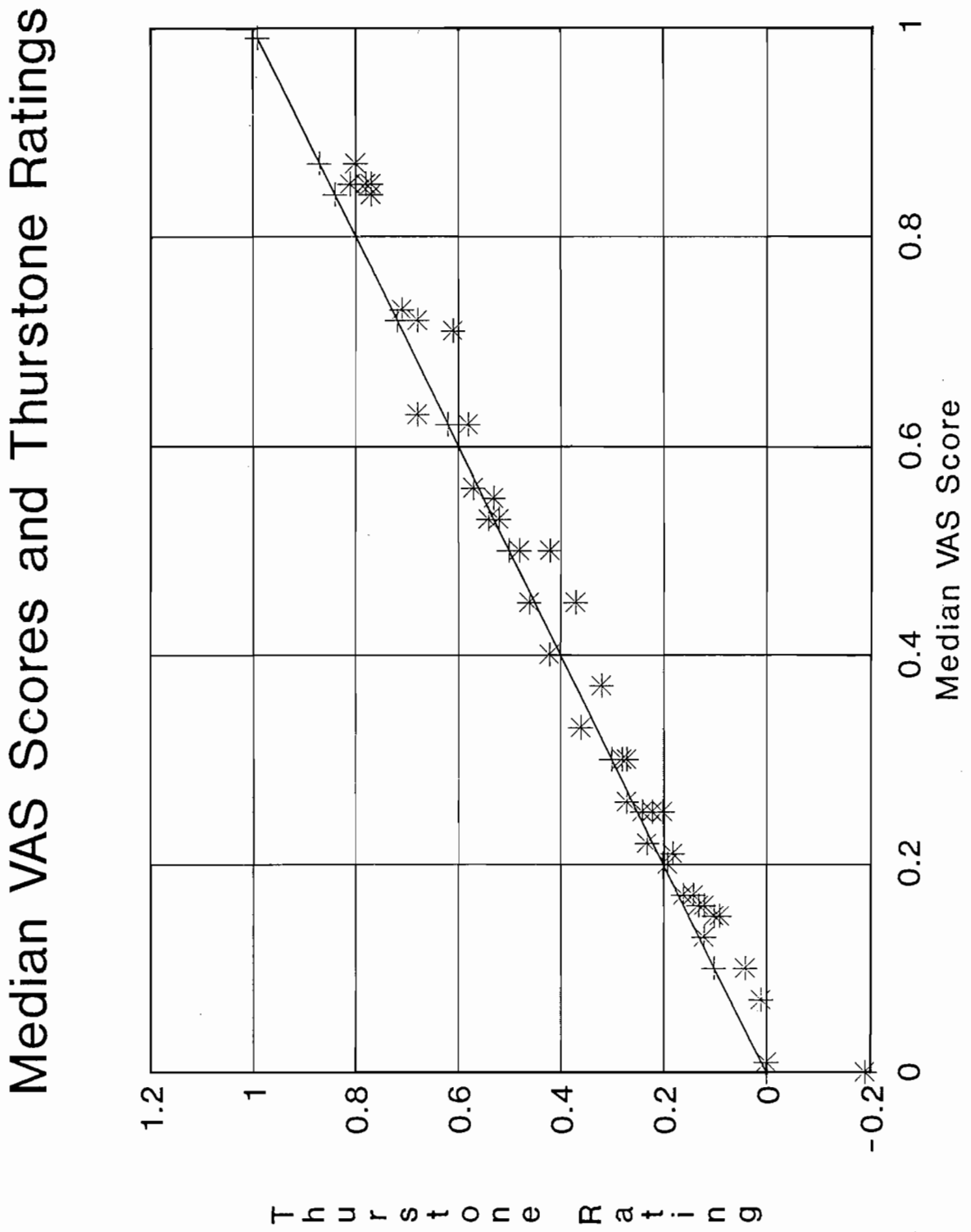
All this still leaves unresolved the fundamental question of which is the correct valuation method to use in this endeavour. As well as the VAS, SG and TTO methods, other possible contenders are Magnitude Estimation (ME), Equivalence of Numbers, and Willingness-to-Pay. At a theoretical level there seems to be no resolution possible in the general case, since each method has its flaws. It may

nevertheless be the case that a particular method might be superior to all others in a particular context. The Euroqol^o Group as such has simply pressed on with the VAS thermometer, for the practical reasons mentioned earlier, though the Group recognises the need to explore and understand the relationship between this and other methods, and individual researchers within the group are pursuing this part of the research agenda. At York we used ME earlier in connection with Rosser's Classification, simply because that was the method that Rosser had used originally, but we have never used it with the Euroqol^o descriptive system. Recently we have been using the TTO method alongside the VAS thermometer. If we can establish a strong and stable relationship between VAS valuations and TTO valuations, then the more user-friendly VAS method can be used in future experimental work, and the results "translated" into TTO values where appropriate.

CONCLUSION

I have to say, however, that there is a great danger here that we strain at gnats whilst camels pass by in the night. When one considers the sorry state of health care benefit measurement at both a descriptive and an evaluative level, I am inclined to the view that, rough and ready though our achievements to date may be, they represent a great step forward compared with the more common use of mortality data and physiological measures in the vast majority of evaluative studies in health care. We must not let the perfect become the enemy of the merely good.

Figure 1



REFERENCES

This contains, for ease of reference, the published material from the Euroqol^o Group to date, plus those major critical commentaries upon that work that we are aware of. If there are others that people know about that are not cited here, please draw them to our attention. It will be appreciated that with much of the work still ongoing, some of the conclusions cited in the text rely on material that has so far only reached publication in the form of Reports to research funders.

Björk S (1992) Euroqol Conference Proceedings, October 1991, IHE Working Paper 2, Lund.

Brazier J, Jones N and Kind P (1993) A Comparison of Two Health Status Measures: Euroqol Meets SF36, Paper to HESG, York.

Brooks R (1991) Euroqol: Health-Related Quality of Life Measurement. Results of the Swedish Questionnaire Exercise, Health Policy, 18: 37-48.

Carr-Hill R (1991) A Good Measure for Eurohealth? Health Service Journal, 101: 24-25.

Carr-Hill R (1992) A Second Opinion: Health Related Quality of Life Measurement - Euro Style, Health Policy, 20: 321-328.

Essink-Bot M (1990) Valuation of Health States by the General Public: Feasibility of a Standardised Measurement Procedure, Social Science and Medicine, 31: 1201-1206.

Euroqol Group (1990) Euroqol - A New Facility for the Measurement of Health Related Quality of Life, Health Policy, 16: 199-208.

Euroqol Group (1991) Not a Quick Fix, Health Service Journal, 101: 29. (Response to Carr-Hill).

Euroqol Group (1992) Euroqol - A Reply and Reminder, Health Policy, 20: 329-332. (Response to Carr-Hill).

Gafni A and Birch S (1993) Searching for a Common Currency: Critical Appraisal of the Scientific Basis Underlying Harmonization of the Measurement of Health Related Quality of Life (Euroqol), Health Policy, 23: 219-228.

Kind P (1990) Measuring Valuations for Health States: A Survey of Patients in General Practice, Discussion Paper No 76, Centre for Health Economics, University of York.

Kind P (1991) Measuring Health-Related Quality of Life - A European 'Lead' Standard THS Health Summary, 8: 6-7.

Nord E (1991a) Euroqol: Health-Related Quality of Life Measurement. Valuations of Health States by the General Public in Norway, Health Policy, 18: 25-36.

Nord E (1991b) The Validity of a Visual Analogue Scale in Determining Social Preference Weights for Health States, International Journal of Health Planning and Management, 6: 232-242.

Sculpher M (1992) An Economic Evaluation of Transcervical Endometrial Resection versus Abdominal Hysterectomy for the Treatment of Menorrhagia, Paper to Health Economics Study Group, Sheffield.

12	
----	--

October 1991

Page 1

HEALTH QUESTIONNAIRE

We are trying to find out what people think about health. We are going to describe a few health states that people can be in. We want you to indicate how good or bad each of these states would be for a person like you. There are no right or wrong answers. Here we are interested only in your personal view.

But first of all we would like you to indicate (on the next page) the state of your own health today.

By placing a tick (thus) in one box in each group below, please indicate which statements best describe your own health state today.

Mobility

- : I have no problems in walking about
- : I have some problems in walking about
- : I am confined to bed

Self-Care

- : I have no problems with self-care
- : I have some problems washing or dressing myself
- : I am unable to wash or dress myself

Usual Activities

- : I have no problems with performing my usual activities (e.g. work, study, housework, family or leisure activities)
- : I have some problems with performing my usual activities
- : I am unable to perform my usual activities

Pain/Discomfort

- : I have no pain or discomfort
- : I have moderate pain or discomfort
- : I have extreme pain or discomfort

Anxiety/Depression

- : I am not anxious or depressed
- : I am moderately anxious or depressed
- : I am extremely anxious or depressed

Compared with my general level of health over the past 12 months, my health state today is

PLEASE TICK ONE BOX

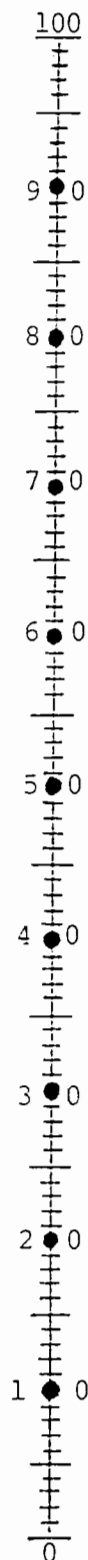
- : Better
- : Much the same
- : Worse

To help people say how good or bad a health state is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked by 100 and the worst state you can imagine is marked by 0.

We would like you to indicate on this scale how good or bad is your own health today, in your opinion. Please do this by drawing a line from the box below to whichever point on the scale indicates how good or bad your current health state is.

Your own health
state today

Best
imaginable
health state



Worst
imaginable
health state

- . We now want you to consider some other health states.

- . Remember, we want you to indicate how good or bad each of these states would be for a person like you.

- . They are described, on either side of the scale, on the page opposite.

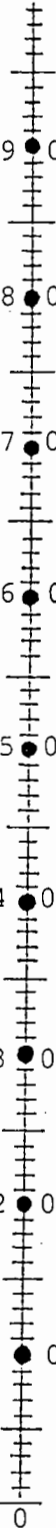
- . When thinking about each health state imagine that it will last for one year. What happens after that is not known and should not be taken into account.

- . Please draw one line from each box to whichever point on the scale indicates how good or bad the state described in that box is.

- . It does not matter if your lines cross each other.

Best
imaginable
health state

100



Worst
imaginable
health state

No problems in walking about
No problems with self-care
Some problems with performing usual activities (eg. work, study, housework, family or leisure activities)
No pain or discomfort
Not anxious or depressed

No problems in walking about
No problems with self-care
No problems with performing usual activities (eg. work, study, housework, family or leisure activities)
Moderate pain or discomfort
Not anxious or depressed

No problems in walking about
No problems with self-care
No problems with performing usual activities (eg. work, study, housework, family or leisure activities)
No pain or discomfort
Not anxious or depressed

Some problems in walking about
Some problems with washing or dressing self
Some problems with performing usual activities (eg. work, study, housework, family or leisure activities)
Extreme pain or discomfort
Extremely anxious or depressed

Some problems in walking about
No problems with self-care
Some problems with performing usual activities (eg. work, study, housework, family or leisure activities)
Extreme pain or discomfort
Moderately anxious or depressed

Confined to bed
Unable to wash or dress self
Unable to perform usual activities (eg. work, study, housework, family or leisure activities)
Extreme pain or discomfort
Extremely anxious or depressed

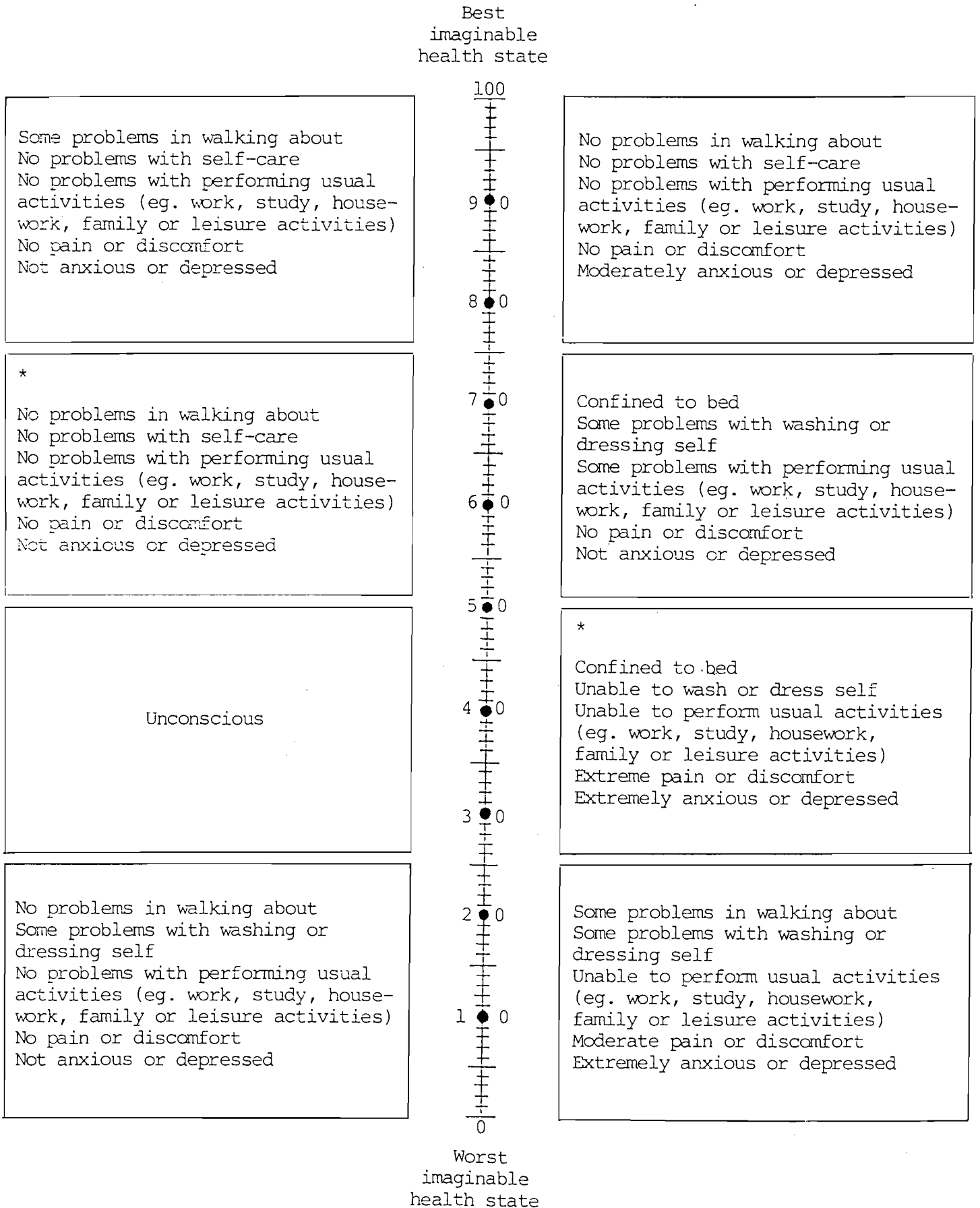
No problems in walking about
No problems with self-care
No problems with performing usual activities (eg. work, study, housework, family or leisure activities)
Moderate pain or discomfort
Moderately anxious or depressed

Confined to bed
Unable to wash or dress self
Unable to perform usual activities (eg. work, study, housework, family or leisure activities)
Moderate pain or discomfort
Not anxious or depressed

PLEASE CHECK THAT YOU HAVE DRAWN ONE LINE FROM EACH BOX
(that is, 8 lines in all)

In the same way as on the previous page, please indicate how good or bad these additional states are, by drawing a line from each box to a point on the scale.

You will find that 2 of these states (marked *) are repeated from the previous page.



PLEASE CHECK THAT YOU HAVE DRAWN ONE LINE FROM EACH BOX
(that is, 8 lines in all)

- . In the previous pages we asked you to say how good or bad various health states are in your view.

- . We would now like you to tell us how good or bad you feel the state 'dead' is, compared with being in the other states for one year.

- . Please turn back to pages 5 and 6 and draw a line across the thermometer at the point you would locate the state 'dead'.

- . Remember we would like you to do this on both pages 5 and 6.

Because all replies are anonymous, it will help us to understand people's answers better if we have a little background data from everyone, as covered in the following questions. (At the end there is space to add anything else you think may be helpful to us).

PLEASE TICK APPROPRIATE BOX

1. Have you experienced serious illness?

- in you yourself

Yes

No

- in your family

Yes

No

- in caring for others

Yes

No

2. What is your age in years?

3. Are you:

Male

Female

PLEASE TICK APPROPRIATE BOX

4. Are you: a current smoker

an ex-smoker

a never smoker

PLEASE TICK APPROPRIATE BOX

PLEASE TICK APPROPRIATE BOX

5. Do you now, or did you ever, work in health or social services?

Yes

No

If so, in what capacity?

6. Which of the following best describes your main activity?

- in employment or self employment

- retired

- housework

- student

- seeking work

- other (please specify)