

A “STANDARD NHS PRACTICE” : VALID TOOL OR SPURIOUS EXERCISE?

How does a National Health Service Strategic Health Authority (NHS SHA) consult individuals, within or without the NHS organisation, when it plans to make radical changes to the way a service is delivered in its area and when the changes involve closures, expansions and different options about arrangements for the future? It seems that, for this, there is a “standard NHS practice”. Before going into how we came to have direct knowledge of it, we exhibit a paradoxical inconsistency that emerged from the simplest possible application of the ‘standard practice’.

1. A disturbing scenario

—“disturbing” because the scenario developed, like that of any Hitchcock, from sweetness & light to something shadier. Suppose that a Strategic Health Authority (SHA) has carried out a consultation to discover how the general public in its area judges two options for cancer services, options that imply differences in the hospitals providing the services and differences in the travel patterns of patients from their catchment areas. The options (A and B) were judged on just two criteria (Quality and Access). In Stage 1 of the consultation, individuals were asked to give the personal *weights* they attach to the two criteria, expressing them as percentages adding to 100. In Stage 2, they were asked to give a single digit *score* (from 0 to 9) as their subjective assessment of how each option satisfies each of the criteria. The *standard NHS procedure* for weighting and scoring works with averages calculated over the whole group of individuals consulted, without any analysis of how those averages are constituted.

Suppose that the (rounded) averages of weights and scores were as in the table:

Access Wt	Quality Wt	Option	Access Score	Quality Score	Weighted Score
50	50	A	3	3	$3 \times 50 + 3 \times 50 = \mathbf{300}$
		B	5	5	$5 \times 50 + 5 \times 50 = \mathbf{500}$

The weighted scores calculated according to the standard NHS formula (as the simple sums of the products of average weights and average scores, for each option) are the two bold entries in the table. Option A gets **300** and Option B gets **500**. It may be concluded that Option B is the clear winner.

Actually, not so clear! Suppose that, when the results of the first stage of the consultation were collected,

someone had noticed that respondents could be split into two well-defined and roughly equal groups — individuals (call them *foraccs*) who gave roughly nine times the weight to access as to quality, and those (call them *forquals*) who did precisely the opposite. When the data were separated into these two groups, the (rounded) averages were those in our second table:

	Weight:		Option	Score:		Weighted
	Access	Quality		Access	Quality	
<i>foraccs</i>	90	10	A	5	1	$5 \times 90 + 1 \times 10 = \mathbf{460}$
			B	1	9	$1 \times 90 + 9 \times 10 = \mathbf{180}$
<i>forquals</i>	10	90	A	1	5	$1 \times 10 + 5 \times 90 = \mathbf{460}$
			B	9	1	$9 \times 10 + 1 \times 90 = \mathbf{180}$

This table revealed a serious inconsistency. When the formula was used to calculate weighted scores for the two groups separately, Option A was found to be the clear winner in *each* group — by **460** to **180**!

Of course, any formula can be expected to produce two different numbers when a group of responses is split into two subgroups. But a formula must be questioned that can deliver subgroup numbers that clearly contradict the numbers for the whole group — the numbers that appear to be used in the standard consultation process. The latter would be expected to lie somewhere between the relevant pairs of subgroup numbers. What is it about the figures in our table that gives rise to such a paradoxical inconsistency? The answer lies with that old statistical bugbear — *correlation*. For each option there is a marked correlation across the two groups between weight and score — one that is positive for A, but negative for B. The ‘foraccs’ think A is good for access and bad for quality, whereas the ‘forquals’ think the opposites.

Is the illustration so contrived that its message can be safely ignored? The answer to that is three-fold. Firstly, even if the illustration were considered extreme, it is both routine and good practice to test the logic of an arithmetical operation by the consideration of extreme cases — the logic should both recognise and respond appropriately the nature of the challenge. Secondly, a procedure that fails when thus challenged may well be failing in an unrecognisable way with the data that it actually uses. Thirdly, an apparent extremity of illustration may well be less contrived than first appears.

Suppose the physical proximity of the average geographical locations of our options and groups can be

faithfully represented on the single dimension:

foraccs

A

forquals B

The scenario may then go as follows. Further investigation reveals that the ‘foraccs’ live a long way from both A and B and understandably give great weight to access, whereas the ‘forquals’ are free to put their weight on quality. Naturally, both groups give access scores that reflect the relative proximities of the options. As far as the more speculative criterion of quality is concerned — both groups simply differ in their relative scoring. (Perhaps ‘foraccs’ have made little previous use of the services associated with either A or B and have had to rely on media stories, whereas ‘forquals’ may have been better informed.) That is how the strong correlations between weights and scores may have come about. Clearly, it would be unwise to make the implicit assumption of the formula that weights and scores are statistically independent (and therefore uncorrelated) — an assumption that is needed to justify the separate averaging of weights and scores in the two stages of the consultation process.

The arithmetic inconsistency revealed in the two tables casts a shadow over our hypothetical scenario. But we can keep the plot going a bit longer by imagining what the SHA’s reaction might have been. Suppose that the SHA’s political masters were pressing them to go for option B, and that the SHA then quite justifiably accused the foraccs of allowing their personal preference for Option A to bias their weights in favour of that option (overlooking the fact that the same accusation of bias could be levelled at the forquals for giving a weight to Quality that favoured Option B!).

All very confusing! What is clear, however, is that the application of the standard weighting and scoring formula that our scenario illustrates is a dubious one unless the overall averages are broken down (“analysed”) to reveal the inhomogeneity of their construction. The ultimate in analysis is to go down to the level of individual respondents, where two weighted scores (one for A, one for B) can be calculated for each individual. These numbers would be a better starting point for making any decisions influenced by the results of the consultation — assuming that the objective of consultation was to consult the public. It would be better still—and much more economical—simply to ask each individual which option he/she preferred. Questions of bias would then become a matter of honest judgement.

There is now a huge industry associated with the practices we are trying to understand and describe. Just now (February 2006) “weighting and scoring” gets 366 ‘hits’ in Google UK while “option appraisal” of which it is a major component gets 87,500. The ratio of these numbers is a measure of how many public bodies are responding to the H. M. Treasury’s Green Book on the matter: manuals are written and expensive courses are run to guide those who want to do things to HMT’s liking. The PA Consulting Group’s suggestion¹ that “there are hundreds of mathematical and statistical techniques that might be applied” in option appraisal is particularly and perhaps intentionally fearful. One of these techniques is Multi-criteria Decision Analysis (MCDA), which is just the pukka academic name for ‘scoring & weighting’. MCDA has pride of place in the manual from National Economic Research Associates (NERA) entitled *Multi-criteria Analysis*² (MCA). This manual (that often uses MCA as acronym instead of MCDA) is more encouraging than the fear-inducing PA advertisement:

“Many of the most important benefits from applying the MCA procedures do not depend upon sophisticated computer packages. In most applications, by far the most valuable benefit is the identification of a relevant set of criteria against which to judge options. From this follows the creation of an explicit framework to allow all options to be comprehensively evaluated in a consistent way. Typically this need not demand more than simple pencil and paper.”²

The work of the NERA team depended heavily, as does so much of government-inspired research in the UK, on the American penchant for rationalism through quantification. As MCDA, scoring & weighting came out of the Harvard Business School³ — just as the technique of Data Envelopment Analysis⁴ came from somewhat less inspired, even quango-frenetic, US sources. We can recommend the NERA manual to anyone interested in understanding the current wave of implementations of a method that has been around for three decades. It is already our impression that some not-too-simple thinking is needed as well as NERA’s “simple pencil and paper”. Some form or other of scoring and weighting now has an influential role in NHS decision making. In order to take a close look at the logic that goes or should go with it, we have to move from hypothetical scenario to real world. We start by documenting some informative items of public record.

2. On the record

The use of a simple ‘mathematical’ formula for option appraisal in the NHS started, along with the millennium, as an HMT imposition on NHS Trusts, and is now used more widely. Our direct knowledge of it stems from two particular instances. The first refers to the “Varley Review”⁵, set up by NHS Eastern Regional Office, which reported in 2002 on the future of the Mount Vernon Cancer Network & Centre. The second instance stems from the decision, based on the Varley review, that the Mount Vernon Cancer Centre is to be moved to Hatfield. It refers to the ongoing (as yet unreported) Radiotherapy Access & Capacity Review, led by consultants for the North West London SHA, which looks at options for future radiotherapy services across North West London.

Both reviews have raised matters of general concern relating to the *standard NHS practice* for evaluating the results of any internal or external consultation about options. We use the phrase only because we came across it in the following quotation from the SHA’s documents circulated to individuals consulted in the Radiotherapy Review:

“The option appraisal will take the form of a scoring and weighting exercise which is standard NHS practice for an exercise of this kind and will be familiar to most of the participants. This exercise will determine the relative non-financial merits of the options; in parallel the options will be subject to a separate financial appraisal to determine the relative costs.”

The descriptions we have seen of the NHS option appraisal process are not specific enough to warrant the epithet “standard”. Much of our analysis and concern about current implementations of scoring & weighting hinges on an apparently small matter — the order in which any exercise takes the weighting (W) and the scoring (S) and the extent to which the associated thinking about them is done separately or in an integrated fashion. In this quotation, S precedes W — and that is the order in which the two components were introduced and explained at length in the NERA manual. But it is not the order we find in the paper⁶ by Dr Galloway, a haematologist in the City Hospitals Sunderland NHS Trust. Dr Galloway’s illustrative exercise is not a public consultation but one that uses scoring & weighting to “write a business case for service development in pathology” by tapping into the knowledge and expertise

of NHS staff (users and providers) concerned with the quality of pathology services. In the interests of making decisions that are as objective as possible and are not biased by any particular option, Dr Galloway recommends “separating the exercises of weighting the benefit criteria and scoring the options.”

“The first step is to weight the benefit criteria. These can be given a relative rank and the most important criterion in the rank[ing] weighted at 100. Each of the other criteria should then be examined against the most important criterion and given an appropriate weighting; thus, if the first criterion is assigned a weight of 100, and the second criterion is considered to be half as important, then a weight of 50 is assigned to the second criterion. These steps are then repeated for the other criteria. The weights . . . are then scaled to total 100. . . . Each of the options is then considered in turn against the benefit criteria, and each option is then scored—for example, between 0 and 100 on each of the criteria.”

The products of scores and corresponding weights are totted up, as in our tables, to get the total weighted score for each option that is taken to be a measure of the benefit for that option in any cost-benefit analysis. Dr Galloway continues:

“It is important to recognise that the assigned weights of the scores are value judgements. To develop the assigned weights and scores, negotiation within the group that is carrying out this work will lead to compromises being made, and it is the number of people involved in the process and their expertise that gives credibility to these value judgements.”

So, for Dr Galloway, members of the group have to argue things out until they agree on objective weights. If averaging were deployed, it would only be to arrive at weights that are less subjective than those assigned by individual members of the group. Presumably the same thing may hold for an objective determination of a single set of scores for each option.

Another manual⁷ (with four options for the refurbishment of Scottish secondary schools) in which criteria are referred to as ‘objectives’ may be in agreement with that:

“41. The weights given to criteria cannot be decided by ‘experts’. Weightings inevitably involve the judgements of stakeholders and decision makers and can be decided upon through reasoned discussions at meetings/working groups. Using a weighting of 1-5 for each objective will usually be sufficient to ensure that an accurate result is achieved. It is generally good practice to weight individual objectives since it is unlikely that they are of equal importance within the project.”

The document goes on to express an interest in how the scores are quantified—but without explicitly linking that question to the apparently separate determination of weights:

“42. The scale [of integers] that is used to score the options should be wide enough to reflect the differences between different options, even if they are quite small. A scale of 0-10 will usually be appropriate, where a rating of 0 refers to a complete failure to deliver an objective, whilst 10 would indicate that an option delivers an objective in full.”

3. Not so simple, says NERA

What we need in all this is some standardisation for the NHS of the integrated thinking in the NERA manual¹ that underpins any scoring & weighting exercise. “We have separation!” is good for rocket lift-off but bad for scoring & weighting. Separation of the thinking about weights and scores is an easy trap to fall into, because it simplifies things. Just ask someone to assign relative weights to the two unquantified criteria of Access and Quality for a cancer network, and you may be surprised that quite intelligent people are willing to supply numbers.

Weights and scores enter the standard formula as *products*: their determinations are inextricably connected by this very fact. Which means that the ratios of the weights in the formula (whether those of individuals or of groups) make sense only if they do what the formula suggests they do. For example, if there were only two criteria A and B, putting twice as much weight on A as on B must mean that you would be indifferent between two options whose scores differ on criterion B by twice as much as they differ (in the opposite direction) on criterion A. More generally:

Suppose $S1A$, $S1B$, $S2A$, $S2B$ are the scores you give to options 1 and 2 on the two criteria A and B with weights WA and WB . Taken together, the two criteria make contributions to the weighted score totals:

$$WA \times S1A + WB \times S1B \text{ for option 1}$$

$$WA \times S2A + WB \times S2B \text{ for option 2.}$$

The contributions are equal (so that, as far as these two criteria are concerned, you would be indifferent between options 1 and 2) only if

$$\frac{WA}{WB} = \frac{S1B-S2B}{S2A-S1A}$$

i.e. if the ratio of the weight of A to the weight of B equals the ratio of the difference in scores for criterion B to the opposite difference for criterion A.

All of that means that weights make sense only when scores make sense too. Can this really be a ‘chicken and egg’ situation with ‘weights’ as chicken and ‘scores’ as egg? Surely it would be better if either weights or scores could be independently tied down. We can see that it cannot be done for weights without thinking about scores. But it can be done, at least conceptually, for the scoring— by objectively establishing some meaning for the scales of the criterion scores e.g. by refining the idea in the last quotation that a score of 0 represents ‘complete failure’ and 10 represents ‘full delivery’ of an ‘objective’ (criterion). Only when the score scales have such an interpretation for each criterion, can we talk meaningfully about weights. That is why the NERA manual made MCDA into ‘scoring’ *then* ‘weighting’.

The NERA view of the role of MCDA as a fairly technical matter was very far from the idea of public consultation that, as our scenario revealed, is paradoxically inconsistent with the MCDA formula. It is a view that underpins the exercise that Dr Galloway wrote about. The following eight quotes are from chapters 5 and 6 of the NERA manual²:

- (i) “We shall assume that MCA is being implemented by a small team of public sector staff with a sound general knowledge of the area in which they are working and that their role is to provide advice on appropriate courses of action to more senior staff or to politicians.”

The manual has a lot to say about scoring and weighting (note the order!). Here are a few extracts—with important messages for those who employ the ‘standard NHS procedure’ and with our *italics* and NERA’s **bold**. For scoring:

(ii) “Because the criteria serve as the performance measures for the MCA, they need to be *operational*. A measurement or judgement needs to specify how well each option meets the objectives expressed by the criterion. . . . a question to be borne in mind in developing a set of criteria is **“Is it possible in practice to measure or judge how well an option performs on these criteria?”**”

(iii) “It is important that each option can be judged against each criterion. The assessment may be objective, with respect to some *commonly shared or understood* scale of measurement, like weight or distance. Optionally it can be judgemental, reflecting the subjective assessment of an *expert*.”

(iv) “ Straightforward applications of MCA require that preferences associated with the consequences of the options are *independent* of each other from one criterion to the next. The key idea is simple: *can you assign preference scores for the options on one criterion without knowing what the options’ preference scores are on any other criterion?* . . . The condition has to be met if the sum of weighted averages is to be used to combine preference scores across criteria”

(v) “Public sector decisions can be particularly prone to double counting . . . *double counting should not be allowed in MCA, since double-counted effects are likely to be given more weight in the final overall decision than they deserve.*”

On how to establish the end-points for each scoring scale:

(vi) “One possibility (*global scaling*) is to assign a score of 0 to represent the worst level of performance that is likely to be encountered in a decision problem of the general type being addressed and 100 to represent the best level. Another option (*local scaling*) associates 0 with the performance level of the option in the currently considered set of options which perform least well and 100 with that which performs best.”

For weighting:

(vii) “The use of weights presents two kinds of *challenge*. One is the need for *exceptional care to ensure logical consistency between the ways in which weights and scores are constructed*. The other *challenge*, in some cases, is the *largely social problem of widely different value judgements of different contributors*.”

(viii) “Most proponents of MCDA now use the method of ‘swing weighting’ to elicit weights for the criteria . . . how does the swing from 0 to 100 on one [score] scale compare to the 0 to 100 swing on another scale?”

4. The post-NERA years

For the Varley Review to which we have referred, the separation of weighting and scoring that Dr Galloway recommended was carried as far as using different groups to assign weights and scores. Weights were predetermined by the steering group of experts that had chosen the criteria. The eight criteria and corresponding percentage weights were

Integration of services	14.4
Quality of clinical care	19.2
Access to services	14.4
Flexibility to change	9.6
Research and teaching	11.5
Staffing	13.5
Quality of physical environment	9.6
Timing and disruption	7.7

We do know that most of the scores were provided by another 59 individuals in seven “carefully and similarly balanced” groups, but we are unable to document the knowledge, expertise or special interests of those individuals. Each group was asked to agree, for each option, on a single score for each criterion — the averages of which (over groups) were then weighted by the steering group’s weights and added to get the total weighted score for each option. There were 12 options including the ‘Do Nothing!’ option of

leaving the cancer centre on the Mount Vernon site. The total weighted scores ranged from 4000 to 5700 with ‘Do Nothing!’ coming in fourth at 5100. The 5700 was for a greenfield site (Hatfield) that was the final decision when costings were introduced that took no account of the cost of a new acute hospital on the site, that would be able to support a new cancer centre.

For the Radiotherapy Review, however, there has been no separation of appraisers—just separation in *time*, first the weighting, then the scoring. In a postal exercise, individual “stakeholders”, including lay members of Community Voice (an umbrella group representing local people’s views on health services in NW London and SW Hertfordshire) were first asked for the weights (adding to 100) they would give to 11 different criteria. A week later, the same individuals were asked to assign a score (one of the digits 0, 1, 2,...,9) to each of four options, on each of the criteria. The 11 criteria were:

- Strategic Fit
- Clinical Quality & Safety
- Environmental Quality
- Access
- Clinical Adjacencies
- Capacity
- Workforce, Training, Education & Development
- New Technology
- Ease of Transition
- Research
- Patient Choice.

For both reviews, there were also weighting & scoring exercises for the members of steering groups, in workshops where definitions and criteria were further discussed, refined and amended.

The following paragraphs run the risk of labouring what is now fairly clear—that, at least in what the NHS is doing, we have come a long way and perhaps in the wrong direction, from the clear launching platform that NERA provided for realistic and scientific option appraisal.

a) *Expertise and knowledge concerning criteria* (c.f. the NERA quotes (ii), (iii) & (iv) above): For both the Varley and Radiotherapy Reviews, it seems that there was a diversity among participants without the common understanding that a small expert group might be able to achieve. Some participants may well have had clear stakeholder concerns to justify their inclusion in any consultation about impending changes and decision options (eg patients or clinicians or administrative staff)—yet have little personal knowledge to justify inclusion in the weighting and scaling exercise. For example, clinicians’ experience may be restricted to their own discipline; patients may know a lot about access but little about clinical issues; administrators may have knowledge of staffing problems but know nothing about research factors. If participants have little knowledge of the criteria they are considering, it is questionable whether the outcome of the exercise has any real value.

b) *Double counting* (NERA quotes (iv) & (v)): Looking at the list of criteria for the Varley Review, it can be argued that “Quality of clinical care” is largely dependent on “Staffing”, “Quality of physical environment”, “Research and teaching”. These variables are clearly not independent. Equally, “Quality of physical environment” will impact on “Staffing”, since staffing potential will be influenced not only by the workplace environment itself but also by other local differences such as availability of housing or property prices. Overlap of this kind may or may not be recognised by the individuals consulted but will almost certainly be interpreted differently by different participants, leading to variations in scoring that could undermine the validity of the process. NERA’s necessary strictures against double counting have force only when the criteria are well-defined and truly independent.

c) *Logical consistency of weights and scores* (NERA quotes (vi), (vii) & (viii)): Our descriptions of the recent literature and of what appears to have been done in both reviews gives us no confidence that such consistency has been achieved or even acknowledged. In the Varley Review scorers were not instructed as to whether they should adopt a *global* or a *local* scale (the distinction in NERA quote (vi))—in order to make sure that their scores were consistent with the weights predetermined by the Steering Group. One can imagine that such consistency might be attainable in a small working group of experts familiar with the fundamental literature and who thereby manage to integrate the weighting and the scoring in the way that NERA tells us to. Was such integration possible with the separations that have been part of the review protocols?

d) *Expertise and knowledge concerning options* (NERA quotes (i), (ii) & (iii)): When it comes to considering options, validity of outcome must again surely depend on the knowledge of the participants. Few participants will come to the exercise with equal knowledge of all options. In the North West London SHAs Radiotherapy Review, one of the site options was Wexham Park Hospital in Berkshire—a hospital outside the SHA's area of which few participants had even the most rudimentary knowledge. So they either left it out or engaged in a guessing game! In the Varley Review workshops, this problem was addressed by importing facilitators to outline merits and disadvantages of the various options, so participants had some information. However, both the quality of such facilitators and their impartiality may well be questionable. Such a method is open to manipulation of participants by the emphasis on criteria which others would consider less important.

(e) *Irresponsibility*: A general concern about using an opaque mathematical formula is that it spares individuals from taking responsibility for recommendations or decisions that are endorsed by the exercise. Anonymity is assured and no one is accountable. An irony of this is that if the exercise produces unpalatable results then those with authority can quietly shelve the outcome - for all the reasons quoted above! However if it produces a desired result, then blame can be attributed to other people if the decision is ultimately discredited—and the same people can go on using the same formula to make further decisions in future.

(f) *Personal bias*: It is tempting to suppose that involving a wide group of stakeholders guarantees objectivity and impartiality. But participants may come with clear and determined preconceptions, leading to deliberately skewed weighting or scoring in order to produce a desired outcome. For instance senior clinicians may be influenced by opportunities for empire building, patients by wanting services near their own home, administrators by personal career opportunities. Few people are dispassionate about decisions affecting their own lives and this process offers scope for manipulation simply by the composition of the participant group. An exercise using the same criteria, weights and options would be likely to produce different results if applied to groups of participants whose interests and concerns are dissimilar.

5. Conclusion

The above considerations all contribute to the conclusion that current NHS weighting and scaling exercises may be caricatures of the sound mathematical thinking that initially inspired H.M.T. to recommend them in its Green Book. The already available evidence for this judgement is now so strong that, if NHS managers have doubts about it, they should at least recognise that there is a case for a fuller independent inquiry that would be able to look at all the evidence, either way. What can be done to improve procedures if we are right in thinking that the “standard NHS practice” in its present implementations in the NHS is a confused version, if not actually an abuse, of the scientific approach that was intended by those who devised it? The confusion may have arisen because there have been two objectives, *wide (even public) consultation* and *knowledge-based option appraisal*, that should have been kept separate.

NERA has made the case for any option appraisal to be done by a small group of independent, knowledgeable and unbiased experts. But that smacks of elitism or authoritarianism —two ‘isms’ that are very much out of fashion. Hence the attempt to graft wider consultation onto the process to ensure greater political legitimacy — which, as we have seen, undermines the only legitimacy that really matters, namely, whether the practice delivers the option that serves the public interest when that is given an objective definition? Better than wider consultation would be transparency and full publication of the work of the option appraisal group, and the assurance that the members of that group are truly independent, knowledgeable and unbiased. A tall order, perhaps, and one that is difficult to accomplish — but working with that goal in mind is surely preferable to the pretensions of current practice.

References

1. PA Consulting Group (2006) Option appraisal and portfolio optimisation. At www.paconsulting.com/services/decision_sciences/option_evaluation/
2. NERA (2000) DTLR multi-criteria analysis manual. At www.odpm.gov.uk/pub/252/multicriteriaanalysismanualPDF1380Kb_id1142252.pdf.
3. Keeney, R. L. and Raiffa, H. (1976) Decisions with Multiple Objectives: Preferences and Value Trade-offs. Wiley (reprinted by Cambridge University Press in 1993).

4. Stone, M. (2002) Can public service efficiency measurement be a useful tool of government? The lesson of the Spottiswoode Report. *Public Money & Management, July-September 2002* pp.33-39.
5. Preliminary and Final (2002) Reports of the Long Term Review of Mount Vernon Cancer Network & Centre. NHS Executive Eastern Region.
6. Galloway, M. J. (2004) Best practice guideline: writing a business case for service development in pathology. Best Practice No.177. *J. Clin. Path.* **57** 337-343.
7. Scottish Executive (2006) Option Appraisal: Building our Future: Scotland's School Estate. www.scotland.gov.uk/library5/education/seoa-00.asp.

Joan Davis represented Hillingdon Community Health Council on the Varley Review, was a lay member of the Mount Vernon Ambulatory Radiotherapy Group and now represents Community Voice on the NW-London Sector Radiotherapy Review Group .

Mervyn Stone is emeritus professor of statistics at UCL.

February, 2006