

Discovering the QALY

Or how Rachel Rosser changed my life

Alan Williams

Prologue

During the first twenty years of my professional life I was a public finance economist, interested primarily in alternative models of fiscal federalism as they might apply to intergovernmental financial relations in the UK. I wrote a textbook on tax incidence and incentive problems. But I was also interested in how to make cost-benefit analysis work better when appraising different public services, paying special attention to the water industry (becoming a member of the Yorkshire Water Authority and later the National Water Council). My odd conclusion that properly sited reservoirs were potentially more valuable as recreational resources than as sources of water did not go down well.

Later, when health and health care occupied most of my professional life, my activities were so varied that trying to cover them all would have meant that I would have been unable to get to grips with any of them satisfactorily. So I decided to take one strand only, and to follow it through in a coherent manner. This still left me with plenty of options. I could have concentrated on how I helped to develop the sub-discipline of health economics, or how, starting from scratch, York became a world-famous centre for research and training in that field, or how membership of the Royal Commission on the NHS precipitated my mid-life crisis, or the great educational experience I got from the crash course on the perils and problems associated with multi-disciplinary health services research that came with membership of Douglas Black's Chief Scientist's Research Committee. But I decided to ditch all of these and concentrate on the topic with which I have primarily become associated in other people's minds, namely the Quality-Adjusted Life Year (or QALY) as a measure of health.

This particular story starts when I found myself part of Harold Wilson's "white hot technological revolution". Between 1966 and 1968 I was seconded from the University of York to Her Majesty's Treasury, where my official designation was Director of Economic Studies at the Centre for Administrative Studies, the forerunner of the Civil Service College. My principal task was devising and teaching on economics courses for senior civil servants, with the aim of ensuring that they became economically literate (and a few of them even economically numerate). But this was not regarded as a full-time occupation, so I was also designated an "Economic Consultant to the Treasury", and in that capacity became an odd-job man in the Government Economic Service, available on loan to other departments who had little jobs that a part-time economist might usefully do. I have already recounted elsewhere where one of these odd jobs led me (Williams, 1997a), and I am about to recount where another one led me. But first of all I must set the scene by telling you about my brief encounter with Roy Jenkins.

Roy Jenkins had recently been made Home Secretary, and was surprised to find that the Home Office had no economists amongst its complement of in-house advisers and researchers. So I was drafted in for a few weeks to identify elements in the Home Office's portfolio of responsibilities that might be susceptible to economic analysis. I leave you to imagine the frosty reception that this Treasury spy got as he toured the higher echelons of a very conservative department of state, which was exceedingly (some would say excessively) proud of its achievements, and a jealous guardian of its professional self-esteem. But I soldiered on, ignoring the frequent citation of the catchphrase "Let justice be

done though the heavens fall”, which was taken to mean that the pursuit of the objectives of the Home Office was not to be constrained by sordid economic considerations, like trying to calculate whether the benefits were worth the costs. Moreover, I was often told that there was very little in their portfolio that could possibly be of any interest to an economist (and this at a time when one of the major issues was immigration policy). At the end of the day I did have to admit that there were a few things that were of little interest to an economist, the most prominent of which was the protection of birds. Everything else seemed to constitute suitable grist for an economist’s mill, and especially the cost-effectiveness of the police, prisons and the whole criminal justice system. I so reported, but future progress down that route was severely jeopardised when Roy Jenkins moved on and Jim Callaghan took over. Callaghan’s enthusiasm for looking at the cost-effectiveness of the police was distinctly less strong than that of his predecessor.

After my return to York, my interest in the Home Office waned and my interest in the NHS waxed, and I found myself trying to devise a measure of NHS effectiveness that focused on the intrinsic value of health as a good in its own right, rather than merely as something enabling people to contribute better to the welfare of society (especially by maintaining their productivity). This attempt to get away from “human capital” type thinking, which had dominated the economics literature, was partly motivated by the work of one of my former students, Ken Wright, who, as a researcher, was then grappling with the problem of how to evaluate the care of the elderly. When people reach the stage in their lives when they are going to consume more than they produce, then on naïve “human capital” grounds a quick cheap death is the optimal solution. While this was obviously the wrong framework within which to conduct policy analysis in this field, it was not obvious what the right one should be. The solution being explored was to take the widely used measures of Activities of Daily Living (ADL), and use them for evaluative purposes as well as for assessment purposes (Wright, 1974). If enabling people to remain independent as long as possible was the main objective of policy, then assessment of their capacity to perform well on an ADL scale should be a good outcome measure. But how should it be weighted? It could be by the predictive value of each element in indicating when a person needs to be taken into a care home, but this might not be the right set of weights if what you wanted was how serious each element is in affecting the individual’s sense of well-being. For instance, pain may be a far more distressing element for an individual than physical disability, yet it could be the latter that would be used as the indicator that a person can no longer live independently in their own home.

In the middle of all this I remembered that when in the Home Office I had encountered the work of Sellin and Wolfgang (1964), two American criminologists who had created an index of crime seriousness using the views of the general public rather than the views of the police or of the courts. So I got in touch with Vincent Watts, who had been one of the very helpful operational research analysts to whom I had talked quite a lot during my stint at the Home Office, to see if he knew of any further developments of the Sellin and Wolfgang work, especially in a UK setting, that might be adapted for use in health care. My plan was to draw on this and see how best to get the valuations of the general public into the ADL index. He suggested that I should meet his wife, a psychiatrist who had been developing such a measure for use in the evaluation of medical treatments. His wife was Rachel Rosser, and meeting her was the start of a voyage of discovery that has lasted the rest of my life.

The Rosser index

Rachel was interested in measuring “the sanative output of a hospital”; i.e. the extent to which an episode of hospital treatment improved someone’s health, and how far this fell short of returning them to full health. It is still the \$64,000 question, and for most hospital treatments we still do not know the answer (a matter I shall come back to towards the end). She had constructed a classification system based only on two dimensions, disability and distress. It generated 29 different health states for which she and her husband had sought a set of relative social values to act as weights in a scoring system. Rosser and Watts had started by using Court Awards as their source of social valuations (Rosser and Watts, 1974), but Rachel had gone on to interview a convenience sample of doctors, nurses, patients and members of the general public, which was of more direct interest to me.

Rachel was linked into an international network of health status index developers, so through her I discovered key actors such as Bush in the USA and Torrance in Canada (Berg, 1973). In the UK, some early pioneers were Grogono and Woodgate (1971), but they were not making much impact amongst clinicians, and were totally unknown to social scientists involved in the evaluation of health and social care.

It slowly dawned on me that an important drawback with many of these indexes was that their measurement scales were rather idiosyncratic, and difficult to interpret. For instance, the Rosser scale used one of the mild states in her classification system as the basic unit of value, and then all other states were rated according to how many times worse they were than that state. This scale had an extreme upper value of 497, a number that had no obvious quantitative interpretation. I knew that we needed to integrate these measures of health-related quality of life with measures of life expectancy if we were to capture the essence of a person’s healthiness. When estimating life expectancy the convention is that dead is rated at zero and alive is rated at one. Since what we were doing was essentially saying that some people are more “alive” than others, then we should be working with a scale in which dead = 0 and healthy = 1, and in which states of less than full health would be rated at less than one. These weights could then be used to calculate quality-adjusted life expectancy (i.e. healthiness). So I suggested that she rescale her data in that way, a task undertaken by one of her senior researchers, Paul Kind, who was to play a dominant role in later developments when he joined my group at York (Kind *et al.*, 1982).

The main policy interest in the social sciences at that time was not in micro-evaluation but in “social indicators”, with a big international programme of work co-ordinated by the OECD, in which the SSRC (which later became the ESRC) was a major player. As part of that work, Tony Culyer, Bob Lavers and I had devised a simple analytical scheme for thinking more clearly about what was involved in generating social indicators of health, as a by-product of which we also devised an analytical scheme for conducting economic appraisals of health care (Culyer *et al.*, 1972). We stressed the inescapable role of preferences in any outcome measure that purports to compare the value of one prognosis with that of another, and the need to make these preferences explicit. Since both length of life and quality of life are valued by people, we used a graphical representation which had length of life on the horizontal axis and quality of life on the vertical axis. And that was the formulation I used many years later in my analysis of the economics of coronary artery bypass grafting (Williams, 1985), of which more anon.

The scene changes

By the mid-1970s I felt that these matters were sufficiently clear in my own head to start trying to convert others to this way of thinking. My basic position, stripped of all

qualifications, was that the NHS should be setting priorities in health care by creating a cost per QALY league table, and then, starting with the interventions with the lowest cost per QALY, should work its way through the table until its budget was exhausted. I advocated calculating QALYs using the Rosser Index, since that was simple and it had a UK-based scoring system. Strangely for an economist, I was less interested in the cost side, though I always stressed its importance as an indicator of the sacrifices that were being imposed on others. In order to enable people to go down my recommended route I got Michael Drummond (formerly a graduate student at York) to produce a couple of resource books to support the more adventurous at a practical level (Drummond, 1981a; b).

In the late 1970s I was presented with many opportunities to pursue this crusade, not all of which proved to be very successful. My most important single convert was Douglas Black, who had become Chief Scientist at the Department of Health. He recognised the potential of this way of thinking, and indeed of the role of health economics in general, in the evaluation of health care. He was one of the people mainly responsible for creating in the UK a cadre of professionally trained health economists unrivalled elsewhere in the world. He invited me to serve on many of his advisory committees, the most challenging being the Research Liaison Group on the Elderly, which took me back to my starting point and Ken Wright's work. I only made a little headway. The same is true of my later membership of the Royal Commission on the NHS, where I found myself totally at loggerheads with the Chairman, Alec Merrison, over the Commission's role. I saw this as doing for the NHS what the Robbins Report had done for Higher Education, but he seemed to see it as some kind of holding operation in which all we had to do was re-state basic principles and hold the line at a general strategic level. After a couple of years the tension got too much for me and I quit, with a strong sense of inadequacy and personal failure.

Then, in the middle of the 1980s, I was invited by Bryan Jennett to make a presentation at a Consensus Development Conference on the role of coronary artery bypass grafting (CABG) in the treatment of angina. This was an opportunity to roll out the cost per QALY approach in all its glory, though with data that was, shall I say, somewhat problematic. But, with help from colleagues in the Department of Health and elsewhere, I did manage to produce some cost per QALY league tables. Much to my surprise, I did not get annihilated by the critics, and my findings proved to be quite influential. I was shocked that no-one had data on many of the key issues, and that they too were filling the gaps by making heroic assumptions. It was a turning point as far as I was concerned, and emboldened me sufficiently to take what proved to be a most important step in broadening my approach to quality of life measurement.

EQ5D and the MVH Group

In comparing the cost-effectiveness of CABG with that of rival contenders for NHS funds, I had once more used the Rosser Index, but I was becoming increasingly aware of the fragility of its valuation base, and the limitations of its classification system. I wondered whether the time had come for us to have another look at the entire field and see whether, fifteen years later, we could not do better. So in 1987, with Rosser and Kind, I gathered together some interested researchers in Europe to see whether, by pooling our knowledge and expertise, we could identify a "common core" of some three or four key elements that most health status indexes contained, and that were presumably regarded as important in determining whether someone's health was getting better or not. The goal of this "Common Core Group" was to identify an essential minimum data set, not to produce a comprehensive measure containing every item that might be considered relevant by

someone or other. Being both multi-national and multi-disciplinary, and having in mind the different purposes to which such a measure might be put, our early discussions were difficult and often tense, but we stuck it out and slowly learned to understand and accept each other's conceptual frameworks and modes of thought and expression. The miracle is that in so doing we did not fight shy of forthright argument, and often made quite severe criticisms of each other's positions, a tradition which persists within the group, now known as the EuroQol Group. From that difficult gestation period emerged the EQ5D descriptive measure of health-related quality of life,¹ which is now widely used worldwide, and is the most commonly used measure in QALY calculations for the National Institute for Clinical Excellence (NICE).

But I am jumping too far ahead. The EQ5D descriptive system is only half the story. The other half is the scoring system, and when used by the NHS I wanted this to be based on the valuations of a representative sample of the UK population. So there began, in 1990, the lengthy negotiations and scientific interchanges with the Research and Development (R&D) arm of the Department of Health, which funded a large research programme carried out under my supervision at York. It led to the creation of the Measurement and Valuation of Health (MVH) Group, the essence of our task being to gather from a large representative sample of the UK population a set of valuations for the 245 health states generated by the EQ5D classification system. We had to adopt a complicated research design to meet the constraints we faced in this ambitious task, and it took us several years to work our way through the intellectual and logistical difficulties involved. But the methodology devised by the MVH Group (Williams, 1995) has been treated as a model replicated by others in their own countries, with gradual modification of detail as we have discovered more streamlined ways of getting to the desired end-product. This was the most important period in my entire life as an empirical researcher, since it required a very high level of commitment to carry it through in the face of a great deal of scepticism and even of outright hostility. I depended very heavily on the support of colleagues, and especially on Paul Kind, who became a key player in a very difficult enterprise. What carried us through was a shared vision about the importance of what we were doing. Fortunately our R&D support was unswerving, and in 1995 we delivered the goods, namely a scoring system for use with the EQ5D descriptive system, which rested on firmer foundations than anyone had previously achieved in this field, so that we now had an instrument specifically designed for bodies like NICE to use for priority-setting in the NHS.

But at the time there was no body like NICE, and the public health community were, to put it mildly, unenthusiastic about this newcomer on the block, and most clinicians did not want to be associated with health care rationing in any shape or form. So, apart from being picked up by a few enthusiasts, our work languished for years. Rosser herself could not stomach the use of the measure for priority-setting at a population level, still being locked in to a clinical perspective in which evaluating alternative treatments for an individual patient was OK, but choosing between treatments for different patients was unacceptable. In her view, no-one should be denied effective treatments on cost-effectiveness grounds. This is an attitude still shared by many clinicians, despite the fact that they have always sought to be efficient in the allocation of their own time and skills, which, if they considered their own behaviour carefully, they would have to admit was guided by "cost-

¹ This has five dimensions (mobility; self-care; usual activities; pain/discomfort; anxiety/depression), each of which can be reported at one of 3 levels (no problems, some problems, severe problems, though the wording differs slightly from dimension to dimension). There is a scale of preference for each of the resulting 243 health states, including some negative values for some very severe states (indicating that being in them is regarded by the majority of the UK population as being worse than dead). For more information about EQ5D and the EuroQol Group in general, go to www.euroqol.org

effectiveness” thinking. But by calling it “clinical priority-setting” instead of “rationing” they manage to avoid the cognitive dissonance that explicitness and honesty would have created. It is still a prominent source of tension between those who take a population perspective and those who take an individual patient perspective when evaluating health care, but every individual clinician is responsible for more than one patient, and has to balance their respective needs and their consideration of what if anything can be or should be done about them. So they all face, and have to resolve, this tension between an individual perspective and a population perspective, whether they realise it or not, and whether they acknowledge it or not.

A QALY is a QALY is a QALY. Or is it?

At this stage I decided that, having delivered the tools, perhaps I could leave others to get on with the job, so in the mid-1990s I turned my attention increasingly to the issue of whether the NHS has objectives other than QALY maximisation, and if so how they might be addressed systematically in the priority-setting process.

Looking back I realised that when developing the methodology of economic appraisal in health care we had perhaps accepted too readily the frame of reference of the clinical trial, where the sole objective is to determine which intervention maximises health. Clinical trials do not look at distributive justice, either regarding the distribution of benefits or the distribution of costs. Indeed, most of them do not collect data about costs at all. I still find it strange that cost per QALY maximisation is so heavily criticised for ignoring equity, when using the results of clinical trials for NHS decision-making seems exempt from such criticism. Perhaps it is because we provide a frame of reference within which it is easy to pose the question, “what is a healthy life-year worth to different people?” (since that will affect how healthy life-years should be added together). It may seem odd to ask, “what is a reduction in tumour size worth to different people”, yet that is an equally important question if you are concerned with equity. Instead, any reduction is regarded as of equal value no matter who gets it, and reductions can be added together, and the means and standard deviations calculated, all quite unproblematically it seems.

In the pursuit of QALY maximisation I took the same stance, namely that a QALY is to be regarded as of equal social value no matter who gets it. Note carefully the terms used. The assertion is not that everyone values a QALY equally, because that is patently false. The assertion is that, from a public policy perspective we *assume* that the value of a QALY is the same no matter who gets it. Anyone is free to challenge that assumption and propose another, and we could then examine the differential consequences for priority setting. But some assumption has to be made and justified. In such a dialogue, a strong ethical case can be made for the “equal social value” position, since it reflects the well-established principle that a doctor should not sit in judgment on the economic, social, moral or personal worth of the person being treated, but seek only to improve their health in a caring but detached manner.

What eventually persuaded me to abandon this position were two apparently disconnected observations: firstly, that public opinion surveys indicated that most people (including the old) thought that the young should have priority over the old when a choice had to be made between them, and, secondly, that the main justification for having a public health care system rather than relying on the private sector was because it would help to reduce inequalities in health.

Paradoxically, it was through the writings of John Harris that I came to see that there was a link between these two observations. John Harris had been one of the strongest ethical objectors to the cost per QALY approach, arguing that the correct ethical position was to accord equal social value to what remained of every individual's life, irrespective of whether it was expected to be long, prosperous and healthy, or nasty, brutish and short. There should no calculation of expected life years, and certainly not of expected *quality-adjusted* life years, since that would imply that someone else could legitimately sit in judgement on the quality of a person's life, whereas Harris's view was that this should only be done by the individuals themselves. This view was held despite the fact that it is other people's money that people are claiming should be spent on them rather than anyone else. In his book, *The Value of Life*, Harris (1985, p. 93) wrote:

“What the fair innings argument needs to do is capture and express in workable form the truth that while it is always a *misfortune* to die when one wants to go on living, it is not a *tragedy* to die in old age; but it is on the other hand both a tragedy and a misfortune to be cut off prematurely.”

This was a challenge I decided to accept, and thinking about the distinction between a misfortune and a tragedy, I saw that the fair innings argument was the missing link between favouring the young and wishing to reduce inequalities in health. The fair innings argument essentially says that each of us is entitled to a certain span of years² and those who fail to make it may justifiably feel unfairly treated by life, whilst those who exceed it should consider themselves lucky. But I wanted to go beyond mere years of life and to include such considerations as whether someone's life has been fit and healthy, or disabled and racked with pain, since that too is surely a relevant attribute when considering social justice. So, in 1997, I advocated amending the QALY maximising rule to make it a rule about maximising *equity-weighted* QALYs (Williams, 1997b), where the equity weights would vary inversely with the probability that a person or group would achieve a fair innings. Obviously, the poor as a group would be favoured by this, which was the main purpose of the exercise, but I also observed that men would be favoured compared with women if the same “fair innings” were applied to both, and there were some interesting implications for smokers and non-smokers too. So I started collecting data (mainly from health care professionals) on the sacrifices in total population health that they would be prepared to accept in order to reduce these various inequalities. They show broadly that the median person would make quite significant sacrifices to reduce social class differences in life expectancy, none to reduce differences between smokers and non-smokers, and modest sacrifices to reduce inequalities between men and women (Smith *et al.*, 2004).

But the really controversial implication of all this is that no matter which subgroup you belong to, as you get older your chances of achieving a fair innings improve, so your equity weight goes down. Old people like me have, on any reasonable interpretation of the concept, already enjoyed a fair innings, so my equity-weight would now be less than one. So the obvious logical conclusion from all this is that if you wish to reduce inequalities in people's lifetime experience of health, you have to discriminate against the old. If you won't discriminate against the old, you do not really care about such inequalities. But before considering that matter any further, I must recount what had been going on in another bit of my life.

The rationing agenda

² In Psalms it is suggested that this is three-score years and ten.

Back in 1995, I had been getting increasingly frustrated by the unwillingness of influential and knowledgeable people to face up publicly to the inevitability of the “rationing” of health care, and to acknowledge the fact that it had been going on since time immemorial. This reluctance was preventing any discussion taking place with the general public about the principles on which priority setting should be conducted, and comparing them with the principles on which it had historically been conducted, so that we could reach some informed view about how best to proceed in future.

The opposition came from various quarters. The most radical were the romantic idealists who believed that if we only got our act together there would be no such thing as scarcity (Roberts, 1996). Then there were (and still are) those trapped by the emotive connotations of the term “rationing”, which to me (having been a wartime schoolboy) was benignly synonymous with distributing scarce resources equitably. Then there were those who did see the need for priority setting, but objected to “mechanistic” solutions (by which I think they meant solutions that are explicit, systematic and evidence-based, and consequently leave little room for those in charge to exercise their own individual discretion). I think a strong argument can be made for publicly accountable decisions to be subjected to such a discipline. Any departure from the rules requires very strong and explicit justification, otherwise priority setting becomes idiosyncratic and capricious. Such departures should not simply be accepted uncritically, or, worse still, welcomed as a militant assertion of clinical freedom. The sole justification for clinical freedom is that it is in the public interest. Whenever and wherever it ceases to be so, it must be (and has been) constrained (Williams, 1988).

Just when I was about to despair of ever finding a constructive way forward, I discovered that Robert Maxwell at the King’s Fund and Richard Smith at the *BMJ* shared my frustration, so together we created the Rationing Agenda Group to help fill the void. The members were selected for their known interest in and experience of the priority setting problem, but also so as to minimise the chances that we would rapidly agree on what the actual priorities should be. Indeed, our explicit objective was to identify differences of opinion and get them out in the open, and if some well-known difference of opinion did not emerge within the group, we invited someone from outside the group to express a view opposite to ours. The outcome of this endeavour (which took a couple of years to bring to fruition) was a book containing both our own document (“The Rationing Agenda”) and the commissioned think-pieces laying out controversial issues from opposing viewpoints (New, 1997).

One of these controversial issues was my view that in order to reduce inequalities in people’s lifetime experience of health, it is necessary to discriminate against the old. Predictably this did not go down well with those crusading for better treatment of the old, the most articulate of whom was John Grimley Evans. The Rationing Agenda Group commissioned a piece from each of us and placed our opposing views side by side, leaving the reader to form a judgment (New, 1997, pp. 108-123). This same issue arose again more recently when NICE’s Citizens’ Council considered the role of age in priority setting. Grimley and I were both given an opportunity to convince the Council of the merits of our respective viewpoints (NICE, 2003). I only convinced about one-third of them. But it is an issue that has not yet been resolved, and it is not going to go away. The fair innings argument reverberates strongly with people, and although its link with the inequality-reduction objective is not yet properly understood, when it is people will have to face up to an ethical dilemma they have so far not had to address. No wonder economics is regarded as the dismal science.

What next?

For me the most important future development in QALY measurement is the derivation of equity-weights, as I suggested earlier in connection with the fair innings argument. If we are to move beyond QALY maximisation and embrace also the demands of distributive justice, a QALY going to a deprived person must be given more weight than one going to a less deserving person. This should also help to sharpen up the trade-off issues at present left implicit in policy discussions about inequalities in health within society. But that is likely to take a few years to bring to fruition in a practical way, and in the meantime there are other things that we could usefully get on with in parallel.

I had an opportunity to outline my strategy for doing this when invited, in 2004, to give the Annual Lecture established some years ago by the Office of Health Economics in London. Since that organisation is particularly interested in public policy issues at the interface between the Government and the pharmaceutical industry, I thought a particularly relevant topic would be NICE, so I chose as my subject “What could be Nicer than NICE?” (Williams, 2004). Starting from the observation that NICE currently only applies cost per QALY thinking systematically to its technology appraisals, I suggested that in future all clinical guideline development should also adopt a cost per QALY perspective. Traditionally, clinical guidelines are formulated with the objective of increasing the likelihood that *a patient with a particular condition* will obtain the maximum possible benefit from medical treatment. NICE formally acknowledges the presence of resource constraints, and seeks to work within them in a systematic manner, so its clinical guidelines need to pursue the objective of increasing the likelihood that *all patients* will obtain the maximum possible benefit from medical treatment. Therefore, the recommended treatment for patients with a particular condition must take into account the consequences for other patients’ health, and therefore the relative costs. So clinical guidelines need to be designed for cost-effectiveness and not simply for effectiveness. The implication is that sometimes it will be necessary to deny a patient something a little more effective than the alternative, because the extra costs would be quite disproportionate and inflict too great a health sacrifice on those who are competing for the same resources. When deciding what level of extra costs was acceptable for the NHS, NICE was applying an upper limit of about £30,000 per QALY in its technology appraisal system, which I regarded as far too high, given that UK national income per head was only about £18,000 per year. To be spending far more than this on medical care alone to provide someone with an extra year of healthy life seemed to me excessive, so I suggested that £18,000 be the limit.

My second proposal was designed to counteract the fact that NICE concentrated its scarce analytical and managerial talents on a very small fraction of the NHS’s clinical activities. It was obviously impossible for NICE to subject the entire realm of clinical work to such detailed scrutiny, yet it seemed odd to work so thoroughly over innovations whilst leaving the great mass of routine activity unevaluated. So I suggested that each year about thirty clinical interventions, spread across all medical specialties, should be screened for their likely cost-effectiveness. Each specialty would be expected to convene a small expert group to provide evidence about the costs and effectiveness of the selected intervention. To simplify matters at the first stage, on the cost side reference costs could be used. But things look rather more difficult on the effectiveness side, since estimates of QALY gains are needed, which are only rarely possible from existing data sources. Low in scientific status though it is, I suggested using *expert opinion* to fill the gap, and getting a panel of expert clinicians to estimate length and quality of life profiles for patients. By these means a rough

cost per QALY ratio could be estimated for each patient subgroup in receipt of the intervention. With luck, most of these would be below my benchmark of £18,000. But those with cost per QALY ratios higher than this would be designated as “on probation”, and enter the second stage, during which simple monitoring of costs and outcomes would be required of a large enough representative sample of treated patients in the relevant category to be able to check whether the suspicion that it falls beyond the cost-effectiveness threshold is justified or not. If not, well and good, but if so, we enter a third phase, when negotiation takes place as to whether this treatment is really appropriate for this class of patient, or whether some alternative might be better. If this cannot be resolved, then it is time for this intervention to be referred to NICE for a full-scale evaluation, possibly preceded by the setting up of a proper trial so that it is only permissible to continue treating such patients as part of the trial.

Curiously, by this means we might finally get around to fulfilling Rachel Rosser’s original ambition to “measure the sanative output of a hospital”. It takes time, patience and hard work to move the system slowly forward, and we have not quite got there yet. Maybe in another ten years?...

Epilogue

So, in my view, the QALY story is only half finished, and I doubt whether I shall live long enough to see the second half played out. I am hoping that, when it is, people will be mildly amused, or perhaps even amazed, about what all the fuss was about. But the most amusing sequel to all of this is that in 2003 the Home Office began to explore ways in which QALY-type thinking, which it saw as being so successful in health care priority-setting, might be adapted for use in the evaluation of the criminal justice system, and I found myself participating in a series of multidisciplinary seminars exploring this possibility. Sellin and Wolfgang please step forward once more. It seems that even in the world of health service research there is a place for poetic irony.

References

- Berg, R. L. ed. 1973. *Health Status Indexes*. Chicago: Hospital Research and Educational Trust.
- Culyer, A. J., R. J. Lavers and A. Williams. 1972. Health Indicators. In: Shonfield, A., and S. Shaw. eds. *Social Indicators and Social Policy*. London: Heinemann Educational Books for the SSRC, 94-118.
- Drummond, M. F. 1981a. *Principles of Economic Appraisal in Health Care*. Oxford: Oxford University Press.
- Drummond, M. F. 1981b. *Studies in Economic Appraisal in Health Care*. Oxford: Oxford University Press.
- Grogono, A. W., and D. J. Woodgate. 1971. Index for Measuring Health. *The Lancet* II:1024-1026.
- Harris, J. 1985. *The Value of Life*. London: Routledge and Kegan Paul.

- Kind, P., R. Rosser and A. Williams. 1982. Valuation of Quality of Life: Some Psychometric Evidence. In: Jones-Lee, M. W. ed. *The Value of Life and Safety*. Amsterdam: North Holland.
- New, W. ed. 1997. *Rationing: Talk and Action in Health Care*. London: King's Fund and BMJ Publishing Group.
- NICE. 2003. *Citizens Council Report on Age*. London: National Institute for Clinical Excellence.
- Roberts, C. J. 1996. The Anti-rationing Group Also Wants to Contribute to the Debate. *BMJ* 313:557.
- Rosser, R. M. and V. C. Watts. 1974. The Development of a Classification of the Symptoms of Sickness and its Use to Measure the Output of a Hospital. In: Lees, D., and S. Shaw. eds. *Impairment Disability and Handicap*. London: Heinemann Educational Books for the SSRC, 157-170.
- Sellin, T. and M. Wolfgang. 1964. *The Measurement of Delinquency*. New York: Wiley.
- Smith, P. C., L. Ginnelly and M. Sculpher. eds. 2004. *Health Policy and Economics: Opportunities and Challenges*. Oxford: Oxford University Press.
- Williams, A. 1985. The Economics of Coronary Artery Bypass Grafting. *BMJ* 291:326-329.
- Williams, A. 1988. Health Economics: The End of Clinical Freedom? *BMJ* 297:1183-1186.
- Williams, A. 1995. *The Measurement and Valuation of Health: A Chronicle*. Discussion Paper 136. Centre for Health Economics, University of York.
- Williams, A. 1997a. All Cost-Effective Treatments Should Be Free...Or How Archie Cochrane Changed My Life! The Cochrane Lecture. *Journal of Epidemiology and Community Health* 51:116-120.
- Williams, A. 1997b. Intergenerational Equity: An Exploration of the Fair Innings Argument. *Health Economics* 6:117-132.
- Williams, A. 2004. What Could be Nicer than NICE? London: Office of Health Economics.
- Wright, K. G. 1974. Alternative Measures of the Output of Social Programmes: The Elderly. In: Culyer, A. J. ed. *Economic Policies and Social Goals*. London: Martin Robertson, 239-272.