Utility Measurement in Healthcare
The Things I Never Got To

George W. Torrance
McMaster University, Hamilton, Ontario, Canada

Abstract

The present article provides a brief historical background on the development of utility measurement and cost-utility analysis in healthcare. It then outlines a number of research ideas in this field that the author never got to.

The first idea is extremely fundamental. Why is health economics the only application of economics that does not use the discipline of economics? And, more importantly, what discipline should it use? Research ideas are discussed to investigate precisely the underlying theory and axiom systems of both Paretian welfare economics and the decision-theoretical utility approach. Can the two approaches be integrated or modified in some appropriate way so that they better reflect the needs of the health field?

The investigation is described both for the individual and societal levels. Constructing a ‘Robinson Crusoe’ society of only a few individuals with different health needs, preferences and willingness to pay is suggested as a method for gaining insight into the problem.

The second idea concerns the interval property of utilities and, therefore, QALYs. It specifically concerns the important requirement that changes of equal magnitude anywhere on the utility scale, or alternatively on the QALY scale, should be equally desirable. Unfortunately, one of the original restrictions on utility theory states that such comparisons are not permitted by the theory. It is shown, in an important new finding, that while this restriction applies in a world of certainty, it does not in a world of uncertainty, such as healthcare. Further research is suggested to investigate this property under both certainty and uncertainty.

Other research ideas that are described include: the development of a precise axiomatic basis for the time trade-off method; the investigation of chaining as a method of preference measurement with the standard gamble or time trade-off; the development and training of a representative panel of the general public to improve the completeness, coherence and consistency of measured preferences; and the investigation, using a model of a very small society, of the conflict between the patient perspective and the societal perspective regarding preferences.

Finally, it is suggested that an important area of research, which the author never got to, would be to work closely with specific decision makers on specific decision problems, to help them formulate the problem, provide useful analyses, and to publish these as case studies to give the field a better understanding of the problems and the needs of decision makers.
This is an article about “the things I never got to.” Michael Drummond suggested the title. He must have decided that there were many things that I never got to. Of course, he was right. Hopefully, some of you will be inspired, or at least intrigued, by some of these ideas and will pursue some of those things I never got to.

One can think of my research career as consisting of three major themes. The initial theme concerned the development of the theory and methods for measuring health outcomes using utility theory, the use of the resulting utilities to compute QALYs, and the combination of QALYs with costs to perform cost-effectiveness analysis or, more specifically, the variant on cost-effectiveness analysis generally referred to as cost-utility analysis. A second major theme was the development and application, together with David Feeny, Bill Furlong and the rest of the team, of the preference-based health status measure, the Health Utilities Index (HUI). The third major theme in my research career consisted of various applications of economic evaluation and outcomes measurement, undertaken both as a faculty member at McMaster University and as a research consultant at Innovus Research Inc., now i3 Innovus.

This current paper focuses on the first research theme – the development of a coherent theory and methods for the measurement of health outcomes using utility theory, and the application of such measurements in cost-effectiveness or cost-utility analysis. There are also many things I never got to with regard to the other two themes but, fortunately, others will get to them. Specifically, David Feeny, Bill Furlong and my other HUI colleagues will pursue any unfinished business with the HUI. The matters not dealt with in terms of applications of economic evaluation and outcome measurement will be pursued by, among others, my i3 Innovus colleagues.

1. Background

It may be helpful to review briefly how I got into the field, and the source of some of the ideas. This may be interesting not only in its own right, but may also help to explain why certain things developed in the way they did.

In 1969, I was looking for a dissertation topic for my doctoral work in operations research at the State University of New York at Buffalo (USA). My professor suggested I talk to an academic colleague of his who had recently moved from Buffalo to McMaster University (Hamilton, ON, Canada) to set up a new department of Clinical Epidemiology and Biostatistics. Given that I was, at that time, a junior faculty member in the business school at McMaster University, this was particularly convenient.

So, I went across campus to see this person named David Sackett. It was an unforgettable, stimulating experience, to say the least. David discussed with me the problem of determining whether a particular intervention or programme did more good than harm. He then went on to explain that the healthcare system had no way of determining whether one programme did more good, on balance, than an alternative programme. In my operations research terminology, there was no objective function for the healthcare system. I found this incredible. How could decision makers allocate resources if they had no way to determine which activities and programmes did more good than others? Here was this huge healthcare system with no measures that allowed decision makers to choose wisely between programmes or to allocate scarce resources optimally.

As I mulled this over, several aspects of my background came together. One was my previous experience in manufacturing as an industrial engineer at Procter and Gamble. There I had learned the various techniques of project assessment, such as calculation of internal rate of return, return on investment and years to pay out. Basically, in all of these techniques the incremental costs and the incremental savings of a programme compared with the status quo are predicted over a relevant time horizon, discounted to their equivalent present values, and the discounted costs are subtracted from the discounted savings to give the net present value. If the net present value is positive (i.e. a savings), the programme is better than the status quo. If there are alternative programmes, the one with the greatest net present value is the best. This approach could obviously be applied in healthcare to compute and analyse the cost side of the cost-effectiveness equa-
tion. But what about the effect side? What could be used to measure the health outcomes? What could be used for the objective function?

There was consensus in the literature that the objective of healthcare was to maximise health by reducing both mortality and morbidity, i.e. to maximise both the quantity and quality of life. Fair enough. And it is certainly easy enough to measure the quantity of life, but how could one measure the quality of life? At the business school, I was then teaching a standard decision analysis course complete with decision trees and utilities measured using the standard-gamble approach. I was telling my business students that utilities were an integrative outcome that could be used in any decision problem with any content and with any number of seemingly disparate outcomes. It was natural for me to think of this approach for measuring the worth or value of the health outcomes produced by alternative healthcare programmes. Eureka! We could measure the preferences for the health outcomes using utility theory. Thus was born the general framework for cost-utility analysis.

2. Economics, Health Economics and Decision Science

I have always been struck by the criticism that health economics, specifically health economic evaluation, is the only application of economics that does not use the discipline of economics. If it did, it would use cost-benefit analysis, not cost-effectiveness or cost-utility analysis, and it would measure health outcomes using a monetary amount, on the basis of techniques such as willingness to pay and willingness to accept. Why are these methods not used? Gold and colleagues' book explains it as follows: “CBA [cost-benefit analysis] presumes to put a dollar figure on the value of human life and uses controversial methods to do so,” “… monetizing the price of life in these ways introduces ethical concerns that are avoided by CEA [cost-effectiveness analysis], albeit at some sacrifice of generalizability,” “Cost-benefit analysis’s primary valuation method is willingness to pay (WTP), an approach whose difficulty lies in its intrinsic favoring of the programs and diseases of the affluent over those of the poor.” But note, these are neither fundamental nor insurmountable difficulties. These may indeed be nothing more than post hoc rationalisations, developed after cost-effectiveness analysis became rooted and accepted in the field.

Why is it, then, that health economic evaluation does not use the discipline of economics? The reason may simply be historical. The basic framework for health economic evaluation was developed not by economists but by decision scientists. My formal training was in decision sciences, as was that of Milton Weinstein. Health economic evaluation was developed by decision scientists, and then adopted by economists. Is that the reason it does not conform to the discipline of economics? As an aside, I must confess that the lack of interest in health economic evaluation by my colleagues in decision sciences in the early years was one of my great disappointments. Thank goodness for the economists who saw some value in what was being proposed.

In any event, regardless of the historical rationales, it has always struck me that it is doctrinaire and unproductive to argue about whether or not the approach fits into a particular paradigm. What we should be doing, and what I had always hoped to do, but never got to, is to return to fundamentals and specify precisely the axioms or assumptions that underpin each approach. The two approaches could then be contrasted and compared. Economics has axioms about rational behaviour. Decision science has axioms about rational behaviour. Can these axioms systems be contrasted, compared and even tested for their appropriateness in the healthcare field? Are they intended to be prescriptive, or descriptive, or both? The axioms of decision sciences are certainly designed to be prescriptive. So, for them, the appropriate test would be whether or not decision makers wish their decisions to be consistent with the axioms.

I realise that some researchers have already made comparisons and explored linkages between cost-benefit and cost-utility analyses. But what I am suggesting is more – not just contrasting the two approaches or determining under what assumptions one approach fits into the paradigm of the other approach. No, what I am suggesting, and always wanted to do, but never got to, is a more fundamental investigation. What are the fundamental tenets of each approach? Based on that, what precisely is the interpretation of study results using each approach?
Finally, and most importantly, which approach is more useful for informing decision making in healthcare? This investigation can be undertaken at both the individual level and the systems level. It is unclear which level would lead to the most productive outcome, but each seems interesting and worthwhile.

At the individual level, we need to determine what people wish or desire in terms of health outcomes, and also what trade-offs they are willing to make to achieve desirable outcomes: i.e. trading dollars, which is really trading other consumption that is no longer affordable; trading years of life, either through time trade-off or through the probability of death in the standard gamble; or trading anything else of value. Can we integrate the axioms of welfare economics and decision sciences? Not just compare and contrast and determine under what conditions they are the same. But can we integrate them, or integrate and modify them so that they better reflect our concerns in the health arena?

Tsuchiya and Williams\(^3\) tackled this task, at least in part. They laid out the assumptions of Pareto welfare economics and contrasted that approach with the cost/QALY approach. Being economists, they did a superb job of describing and analysing the assumptions and implications of the Pareto welfare economics approach. But they did not do the same for the utility approach. I would like to see someone lay out precisely what assumptions or axioms underpin the utility approach and specify the implications of these assumptions. For example, there is much confusion about expected utility, and the fact that people typically don’t obey it. It is not generally understood that expected utility is not required to solve a decision problem using utility theory. Expected utility is just a convenient algebraic shortcut.

A decision problem under uncertainty can be solved directly using the axioms of von Neumann-Morgenstern (vNM) Utility Theory, without calculating expected utility. In his seminal book on decision analysis, Howard Raiffa introduces the concept of utility theory and solves example problems without ever using the word utility or calculating an expected utility.\(^7\) He called the unit of measure a BRLT (Basic Reference Lottery Ticket) – a ticket that entitled the owner to participate in a lottery with two outcomes, one good and one bad, with a specified probability \(p\) of obtaining the good outcome. With the BRLT procedure, the best possible outcome in the decision problem was set as the good outcome and the worst possible outcome was set as the bad outcome. All that is required to solve a decision problem without calculating expected utilities is that decision makers must agree that they would like the decision to be consistent with the axioms of utility theory, and they must be willing to express choices under uncertainty.

The axioms of utility theory can be stated in various ways, but in the system I prefer they are stated as follows:\(^8\)  

**Axiom 1.** The individual has transitive preferences. That is, if he/she prefers A to B and B to C, then he/she prefers A to C. A, B and C can be sure things or can be uncertain prospects.

**Axiom 2.** The individual agrees that the decision problem is unchanged by substituting one outcome with an equally preferred outcome. The outcomes can be sure things or can be uncertain prospects. What matters is that the outcome that is substituted is equally preferred to the one that is replaced.

**Axiom 3.** The individual is able to express choices under uncertainty.

If an individual agrees to these three things, the decision problem under uncertainty can be solved essentially by divide and conquer. The person answers small choice questions, such as, for a person who prefers A to B and B to C: “Do you prefer B for sure or a lottery with a 50% chance of A and a 50% chance of C?” The answers to these small choice questions are substituted directly into the large decision problem with equals replacing equals until the large problem is simplified down to the point where the answer is trivially obvious. More specifically, it is simplified down to the point where all the various choices have only two outcomes, the best and the worst, and the dominant choice is the choice with the highest probability of the best outcome.

---

1 For ease of understanding, the axioms are expressed here in simple words, rather than the usual mathematical statements.
therefore the lowest probability of the worst outcome. Nowhere in this process do you compute utilities, or calculate an expected utility, or even concern yourself with the person’s risk attitude. Computing utilities and calculating expected utility is a computationally equivalent and simpler approach, but it is not strictly necessary. Risk attitude is built in automatically, and does not need to be addressed in any special way.

Thus, the bottom line is that utility theory using the standard-gamble approach provides the right answer to people who wish their decision making to be guided by their own preferences, including their preferences for taking or avoiding risks, and by the axioms specified above. How does this compare at the individual level with the problem-solving guidance provided by welfare economics?

At the systems level how do the two approaches compare? For example, utility theory is developed only for an individual decision maker using his/her preferences. Additional assumptions are required to aggregate utilities across a group. One possible assumption is that the group should be treated as a single decision-making unit with utilities that are the mean of the utilities of the individuals in the group. This often seems to be the implicit assumption used by analysts, but it is seldom explicitly stated. If there are subsets of individuals within the group with widely differing utilities, the group could be divided into subgroups with more homogeneous preferences. But what are the implications of that for systems decision making? How does this compare and contrast with systems-level decision making using Paretian welfare economics?

Part of the problem with both of the existing paradigms, welfare economics and decision science, may be that they were initially developed for other applications, not for health. For example, welfare economics was presumably developed to help understand economic behaviour and to help those responsible for the economic welfare of a country, an industry, or an organisation to set economic policies that would achieve specified economic objectives. No doubt this is a gross simplification, but my defence is that I am not a card-carrying economist and thus, undoubtedly, do not fully understand all the complexities and subtleties. On the other hand, it perhaps gives me the opportunity to focus on the big picture. Similarly, decision science was developed to help clearly identified decision makers grapple with their specific decision. These decision makers had the authority and the responsibility to make the actual decision and, accordingly, were able to specify their objectives, their constraints, the alternatives, the possible outcomes, and their preferences for the possible outcomes. Thus, decision science assumes there is someone in charge, someone with the authority and responsibility to make the decision. It is the role of decision science to help this designated decision maker reach the best decision. But in healthcare, who really is in charge? Is it the payers, e.g. ministers of health? Is it the providers, e.g. the medical associations? Is it the patients, e.g. the advocacy groups?

Because the two existing disciplines of economics and decision science were developed for different arenas, and were then reapplied in health, they may not be completely suitable. We should start again, but at the beginning. What are we trying to achieve with the healthcare system? What do the patients want? What do the providers want? What do the funders want? What does the public want?

One approach that might help to address these issues more clearly would be to construct a model of a small simple society, a sort of ‘Robinson Crusoe’ island society. One would need only a few individuals, each with different healthcare needs, different preferences for outcomes, and a different willingness to pay for outcomes. A payer with a limited budget responsible for funding potential treatment programmes would also be required. It would be insightful to determine the results for the individuals and for the system as a whole under a variety of preference structures and using the two different approaches: Paretian welfare economics and cost/QALY analysis.

3. Interval Scale of Utilities

Every student in this field knows that the utilities required for the QALY calculation need to be on an interval scale. In teaching, we always emphasise that scales can be nominal, ordinal, interval or ratio, and that an interval scale is needed to support the kinds of parametric analyses and calculations that we perform with the data, e.g. means and standard devia-
tions. We also routinely mention that the standard- gamble method produces interval scale data.

However, there is a second sense in which the utilities and the resulting QALYs should be interval-based that is not well understood, and certainly not well tested. Given the use of the results to compare programmes according to the number of QALYs they produce, the utilities and the QALYs need to be on an interval scale such that a gain in utility from, e.g. 0.2 to 0.4 should be equally preferable to, not more or less than, a gain from 0.6 to 0.8. More specifically, a gain of equal utility increments anywhere on the scale should be equally preferable for the individual whose utilities are being represented, i.e. the intra-person interval property. For example, if an individual’s utilities are A = 0.2, B = 0.4, C = 0.6, D = 0.8, the person should be indifferent to whether the change is from A to B or from C to D. We will consider this issue in a world of uncertainty and a world of certainty.

Note that this is a different question from the one of equity, in which it is often argued that the social value of a change from 0.2 to 0.4 is not equal to the social value of a change from 0.6 to 0.8, because society may prefer to give QALYs to individuals in poor health. That is a different matter. The equity issue arises when the potential recipients are different individuals. The question we are concerned with here is that within the same individual, the one whose utilities are being represented by the utility scale, is a change from 0.2 to 0.4 equally preferable to one from 0.6 to 0.8? In other words, looking through a ‘veil of ignorance’, not knowing which outcomes and choices might occur to them in the future, would a person agree in advance that a change for them from 0.2 to 0.4 is equally preferable to a change from 0.6 to 0.8?

3.1 Uncertainty

Astute students of decision theory might comment at this point that utility theory does not permit the comparisons described above. Indeed, Luce and Raiffa, in one of the original seminal books on utility theory listed some common fallacies. They describe fallacy 3 as follows: [Note, the symbol > indicates ‘is preferred to’] Suppose that A > B > C > D and that the utility function has the property that \[ u(A) - u(B) > u(C) - u(D) \], then the change from B to A is more preferred than the change from D to C. They argue that this is a fallacy and that the theory does not permit such comparisons because the utility function is constructed from preferences between pairs of alternatives, not between pairs of pairs of alternatives. They are correct that the theory does not permit the comparison of a change from B to A with a change from D to C. To make this comparison using the theory, one would have to construct a decision problem where two of the alternative choices would be to change from B to A and to change from D to C. One cannot specify such a decision problem, so the theory cannot address the issue.

When I first came across this fallacy many years ago, I was devastated. It seemed to invalidate the whole foundation for cost-utility analysis. Interestingly, it has, to my knowledge, never been raised in print, and has only once been brought to my attention as a fundamental problem. The credit for that goes to Jerry Hurley, a McMaster University colleague, who brought this to one of our discussion groups and suggested that it represented a major, perhaps unsolvable, problem for cost-utility analysis.

However, there is a simple solution. The fallacy, as specified by Luce and Raiffa, applied to decisions under certainty. It is correct that utility theory does not permit the comparison of changing from B to A with certainty compared with changing from D to C with certainty. However, as we all know, healthcare decision making is decision making under uncertainty. So, consider the following problem under uncertainty. Alternative 3, the status quo, leads directly to a chance node with a 50/50 chance of outcomes B or D. There are two other alternatives. Alternative 1 leads directly to a chance node with a 50/50 chance of outcomes A or D, i.e. outcome B is replaced with A. Alternative 2 leads directly to a chance node with a 50/50 chance of outcomes B or C, i.e. outcome D is replaced with C. Now which is preferred? Utility theory does indeed apply to this version of the problem, and indicates that Alternative 1 would be preferred to Alternative 2, which would be preferred to Alternative 3. In other words, starting from the status quo, it would be preferred to change B to A rather than to change D.
to C. The mathematical proof is simple. This is a very important finding because it means that the health state utilities do indeed have the intra-person interval property required, i.e. a change from 0.2 to 0.4 is indeed equally preferred to a change from 0.6 to 0.8, in a world of uncertainty. It follows that this also applies to the QALYs calculated directly from these utilities, at least in the basic case of alternative health states all for the same life expectancy: i.e. a gain of x QALYs is equally preferable no matter where on the scale it occurs.

In the world of uncertainty, which surely includes the world of healthcare, the required interval property of utilities is thus directly built in when the utilities are measured with the standard gamble. Therefore, an individual who wishes his/her decisions to be consistent with the axioms of utility theory will indeed be indifferent to a change from 0.2 to 0.4 versus a change from 0.6 to 0.8 on the utility scale that represents his/her utilities, or on the elementary QALY scale built from these utilities, i.e. QALYs for chronic health states for a common time period.

Although the intra-person interval property is guaranteed prescriptively by the methods, it would still be interesting to determine whether the property also exists descriptively. Tests to investigate this are suggested in the following section for the case of certainty. Although they would be more difficult to use in the case of uncertainty, the same kinds of tests could be attempted.

3.2 Certainty

Although the healthcare system is not a world under certainty, one of the advantages of such a world would be the relative ease of testing the intra-person interval property. To my knowledge, this has not yet been done. For example, one could measure utilities for health states A, B, C and D, for a specified time period, e.g. a week, followed by a return to the person’s normal health. If an individual’s utilities were, e.g. A = 0.2, B = 0.4, C = 0.6, D = 0.8, one could then ask the person several types of follow-up questions. One simple one would be the maximum they would be willing to pay to have B instead of A, and the maximum they would be willing to pay to have D instead of C. Hopefully, those monetary amounts would be the same, within measurement error, for changes that had equal utility increments. For unequal utility increments, one would hope the willingness to pay amounts would be in the same direction, although they would not be expected to be proportional because of the non-linear utility of money. Alternatively, one could measure the respondent’s utility function for spending money, and make appropriate adjustments to determine whether the utility measured in these two different ways agreed, again within measurement error. A further variation on this project could be to ask respondents to make a direct comparison of the utility gain in progressing from A to B and from C to D. Although, as discussed in the previous section, such a comparison under certainty is not permitted under the vNM Utility Theory, the axioms supporting such a question have been developed as part of measurable value theory.[10]

Violations of the interval property in empirical testing could represent two kinds of problems. They could be simple measurement errors: i.e. one or more of the measurements might not be consistent with the person’s true feelings and might need to be revisited iteratively until the person uncovers their consistent set of preferences. Alternatively, the person could be following a different decision paradigm, i.e. vNM axioms do not hold descriptively for this individual. Bear in mind, however, that such an individual may wish them to hold prescriptively, even if they do not hold descriptively. Indeed, given the many lapses in human judgment that can easily be demonstrated in test problems, it seems eminently reasonable that many decision makers would want them to hold prescriptively.

This whole approach might be used to compare the pros and cons of the standard-gamble versus time trade-off versus the visual analogue scale (VAS) methods. Researchers, including myself, have commented that the VAS is not a good measure, that the standard gamble has the advantage of being based on solid theory, and that the time trade-off is also acceptable. It would be useful to examine each of these approaches in greater depth using fewer study subjects. In this regard, you might want to use subjects that receive special training, not to influence their preferences or their answers in any way, but in order that they gain some experience in making these kinds of judgments.
4. Axiomatic Foundation for Time Trade-Off

The standard-gamble method is based on a well defined set of axioms, thus ensuring a precise meaning for the measurement. Researchers, including myself, have argued that this makes the standard gamble a more desirable preference measurement tool. On the other hand, time trade-off is more directly related to the use of the results to compute QALYs. Indeed, time trade-off is essentially a QALY equivalence statement, and thus appears to have a natural advantage, given its intended use. Moreover, some researchers who have compared both have elected to use the time trade-off on the basis of their successful experience with it.\textsuperscript{[11]}

Why can we not develop a precise axiomatic basis for the time trade-off? I started down this road once with an obscure, seldom-cited paper that focused on an axiomatic basis for health outcomes over time.\textsuperscript{[12]} One approach to the time trade-off axiomisation would be to extend that paper to cover the time trade-off measurement approach. An alternative approach would be to start afresh with the time trade-off and the QALY and develop an axiomatic foundation. Indeed, Buckingham and Devlin\textsuperscript{[13]} have recently begun this process. Either approach seems to me to be eminently doable. It would then be interesting and probably enlightening to compare and contrast the axioms of the time trade-off with those of the standard gamble and those of traditional economics. This part of the project could become a further extension of the topic on economics, health economics and decision science described above.

Why all this interest in axioms? Because we need the ability to be precise about attaching meaning to our measurements. Just as we can state exactly what a standard-gamble result means and how it can be correctly interpreted, we need the same level of precision for the time trade-off.

5. Chaining, Double Links, Triple Links …

A simple little project that I always intended to do, but never got around to, is to investigate the pros and cons, both in theory and in practice, of chaining in utility measurement. Chaining is the method often used when the disability health states to be measured are all relatively mild and it seems unreasonable to compare them all to death. The situation arises most often with the standard gamble, but can also apply to the time trade-off. In chaining, all the states are measured relative to the worst of the living states, and then this state is measured relative to death in a second measurement.

The advantage is that the questions may be easier for the respondent to deal with for the following two reasons and may, therefore, produce less measurement noise. First, the questions, except for the last one, no longer involve death. Secondly, the indifference point – the indifference probability in the standard gamble, or the time traded in the time trade-off – is no longer near the end of the scale where measurement is difficult. Indeed, measurement is so difficult near the end of the scale that some researchers have reported time trade-off respondents who refused to trade at all. The disadvantage of chaining is that the extra measurement introduces additional measurement error, which propagates itself into the utilities of the states of interest. It would be useful to investigate this in theory and empirically, and to provide guidance to users – when to chain and when not.

6. Completeness of Preferences

Respondents do not arrive at an interview with a complete set of coherent preferences in mind. Preferences are formed, at least in part, during the measurement process. This of necessity, leads to considerable measurement noise and even reversals of preferences and other inconsistencies in the results. The typical approach when measuring preferences is to obtain as large a sample as possible, to increase the representativeness of the results and to reduce the variance in the group estimates. The interview itself is then relatively short, usually an hour or less, and rarely involves checking for inconsistencies and iterating back to make corrections.

As an alternative, we might consider selecting a representative panel of individuals and training them in the process. Care would be needed so that the preferences of the respondents were not influenced by the training, the focus being simply on helping them better understand what they were being asked and the implications of their answers. The preference interviews could then be conducted in...
much greater depth, including continuous consistency checking and iterating back to resolve inconsistencies as they arose. In addition, the respondents could be shown the explicit decision implications of their preferences to ensure that they agreed with these implications.

This approach could be used to design the interview process specifically to help the respondents formulate their true underlying preferences and agree not only to a coherent and consistent set of preferences but also to their decision implications. This process of measuring preferences in depth on a small number of individuals has been used successfully in applications of decision sciences in other fields. It would be worthwhile investigating this approach in healthcare.

7. Perspective for Preferences

The controversy between the patient perspective and the societal perspective has not been fully resolved. The current advice from others and myself is something like the following: use the patient’s perspective (i.e. the patient’s preferences) when making treatment decisions for the individual patient; use the collective societal perspective (i.e. the mean societal preferences) when making programme funding or policy decisions. To the extent that these differ, the latter, the societal perspective, limits treatment choices available to the patient for use in their decision making, using their individual preferences. It would be interesting to play this out in full form as a decision problem. At the patient level, you would need several patients for each disease, each with different preferences. At the societal level, you would need several diseases, alternative treatment programmes for each disease, and societal preferences for the potential outcomes.

Such a model could be an excellent teaching tool to display the issues for discussion. It could also be an excellent research tool to investigate alternative formulations and approaches and to determine the pros and cons of each approach. Perhaps, reminiscent of the first item (economics, health economics and decision sciences, see section 2), the situation could be investigated using both standard Paretian welfare economics and standard decision-theoretical utility theory. This could possibly help to shed light on the questions raised in section 2.

8. Decision Makers

Finally, one of the many things that I considered and never got around to was to take leave from the university and spend the time shadowing and helping a healthcare programme decision maker. Decision science, my discipline, is about helping decision makers. The basic approach is to work with the decision maker to help identify the objectives, constraints and the alternatives, and to help determine the best decision or strategy. I believe there is tremendous scope for research here. There are all kinds of decision makers, no doubt with different objectives, certainly different constraints, and different alternatives. I would like to see our journals full of case studies in which a decision maker together with an analyst worked through the case, much the way Steve Pauker and his colleagues used to do for clinical problems.

We consistently say that our analyses are aimed at informing decision making, but we seldom close the loop to determine to what extent we have actually helped and how we could do better.

9. Conclusions

The list of items in this paper is truly a list of things that have been on my agenda, some for a long time, and that I never got to. In addition, there are many other important questions regarding utilities and QALYs that never made it onto my list. One, for example, is the question of transnational differences in utility weights. This has become an important issue in recent years in the light of large multinational trials and the need for cost-utility analyses for each separate country. My observation is that most of the evidence for transnational differences in utilities is hopelessly confounded with differences in how the utilities are measured. To explore this issue properly, one would need to replicate exactly the same measurement process in the different countries. My hypothesis is that, in such an experiment, country would be a minor determinant of utility, probably no more important than age and sex. Health state would be the major determinant.

Another important item not on my list is the question of equity-weighted QALYs. Currently utilities and QALYs are equity-neutral, and equity considerations, if needed, must be factored in separately.
by the decision maker. Can equity weights be measured in some appropriate way and built into cost-utility analyses?

So, as you see, there is much to do and, particularly, much that I never got to. Hopefully, some of you will pick up on some of these themes. Indeed, you may already have done so. You may already be working on some of these ideas. If so, great. If not, I encourage you to consider them.

Acknowledgements

The author received no funding for the preparation of this article, and has no conflicts of interest with regard to the material herein. He would, however, like to acknowledge the subliminal help from all of his colleagues, with whom he has discussed and debated many of these issues over the decades. In addition, he is grateful for the helpful comments and suggestions from participants at the O’Brien Conference, as well as those from the referees.

References


Correspondence and offprints: Prof. George W. Torrance, 321 Markland Drive, Toronto, ON M9C 1R4, Canada.
E-mail: torrance@mcmaster.ca