Bias adjustment and the nature of health-state utility

Paul T. Menzel*

Paciﬁc Lutheran University (emeritus), Tacoma, WA, USA
Corresponding author. E-mail: menzelpt@plu.edu

KEYWORDS: adaptation, health-state utility, patient ratings, QALYs, rating bias, value of life

SUBJECTIVE VALUE AND ACTUAL PERSONS

Ratings of health states are expressions of subjective value—the value of living in one state, relative to another state, to the person whose life it is. It is appropriate to focus on this kind of value. Why is it good, for example, that a person continues to live (presuming they can and want to)? Most importantly, because life has value to the person whose life it is. Paying attention to such subjective value of life and health is the right enterprise in discerning health-state utility.

Subjective value of life and health means nothing, of course, if it is not the value of life and health to actual persons. There simply is no subjective value without actual persons to whom it is value. When we apply this to health-state utility ratings, the implication is rather obvious: if we can, we should get ratings of the value of living in a given health state, or of moving from one state to another, from people who experience life in those states. That, of course, is actual patients, not merely people asked to imagine

* Paul Menzel taught philosophy and bioethics at the Paciﬁc Lutheran University from 1971 to 2012. Much of his scholarly work has focused on moral questions in health economics, including ‘Strong Medicine: The Ethical Rationing of Health Care’ and numerous articles, often with collaborating authors, on quality-adjusted life years. He has also written on distributive justice in the economics of insurance markets, on the relative priority of preventive care in ‘Prevention vs. Treatment: What’s the Right Balance?’ and most recently on a variety of end-of-life issues. Visiting scholar appointments have included the Kennedy Institute of Ethics (Georgetown), the University of York (England), the Rockefeller Study Center at Bellagio, the Brocher Foundation (Geneva), and the Chinese University of Hong Kong. He currently resides in Oakland, California.

© The Author(s) 2020. Published by Oxford University Press on behalf of Duke University School of Law, Harvard Law School, Oxford University Press, and Stanford Law School. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com. This is an Open Access article distributed under the terms of the Creative Commons Attribution NonCommercial-NoDerivs licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
themselves in such states.\(^1\) Nir Eyal’s discussion of health-state utility is placed squarely, and appropriately, within such a framework.\(^2\)

It is well known that actual patients often adapt to their condition. For most conditions, since the bulk of time people spend in them will be when they have adapted, the real subjective value of the health states will be affected by adaptation. Eyal notes carefully the various factors contributing to adaptation that favor using adapted patient values, and some that disfavor it. By and large he ends up sticking with the adaptation-influenced values of current patients if the alternative is ratings by general citizens who only imagine themselves to have the condition; after all, it is valuations by current patients that reflect real life with the condition. But then the further factor enters that occupies Eyal: patients whose condition has been cured or alleviated generally rate the condition as worse than patients who have not recovered do.

Thus we have another decision to make in discerning health-state utility: use current adapted patient ratings or the retrospective ratings of post-adapted, cured patients (‘past patients’). The simplest case will involve choosing between the ratings of current patients who could recover from their condition but have not yet recovered and ratings of that condition provided by past patients. For this kind of case, the question Eyal poses is powerful: ‘... if being on dialysis [e.g.] is so good, why do people whose longtime reliance on dialysis ends suddenly deem being on dialysis so bad?’\(^3\) Some sort of misperception by non-cured, adapted patients from their limited and limiting perspective would seem to be involved. Though both adapted and past patients have experienced life with the condition and without it, the past patient has experienced the better health more recently. Understandably, the current patient who thinks life with her condition just is the life she has, period, will adapt if she can, but her perspective may be unduly confined if she has lost sight of possible recovery.

Eyal addresses the situation with a provocative and constructive proposal: adjust current patient ratings for the bias of adapted patients (BAP), expressed as the ratio of their ratings to the retrospective ratings of former patients. I will pose three objections to the way he works this out. (i) There may be more bias in past patient ratings than he concludes. (ii) Using the BAP ratio on conditions that have no cure is questionable. (iii) The standard notion of health-state utility is ambiguous and possibly even incoherent, greatly complicating the use of BAP.

**BIAS APLENTY**

Eyal considers several potential biases that could lead us to reject past patient ratings in favor of even the admittedly biased ratings of current patients. His explanations for why several of them do not constitute good reasons to reject past patients’ ratings are persuasive, but with several potential biases, they are not.

‘Recall’ bias might be pervasive and general—not accurately remembering the previous experience before cure, omitting important details. But ‘as long as the mistakes...
do not have systematic trends,’ Eyal contends, ‘a large enough sample would overcome this complication.’ But what reason is there to think that the mistakes do not have systematic trends? Now that one is free of a condition to which one had adapted, but which even then one would have preferred not to have, how likely is one to remember the full positive effect of the adaptation? Might not the current relief—and of course it is relief, even if the previous adaptation was substantial—lead people to under-remember the adaptation? Until we have better empirical evidence, recall bias should not be dismissed.

Similarly, ‘investment’ bias—a reluctance to admit that after everything one sacrificed to get cured, the cure has achieved less than what was expected—may be more difficult to set aside than Eyal believes. His claim is that as far as we know, ‘there is no general correlation between the effort and burden that a medical intervention demands and patients’ retrospective reports that it made a big difference.’ Investment, though, includes more than effort made and trouble endured. An investment is made ‘just by having decided’ to go ahead with the treatment. For a decision-making being, any important, substantial decision is an investment of oneself. Decisions are ‘owned’ by the people who make them, and something of themselves is at stake in whether a decision turns out to be good to have made. People want it to be good and may tilt that way in any later assessment. Perhaps investment bias is not substantial enough to worry about, but in the absence of good analysis and empirical evidence, we do not know.

Evaluative ratings of a given condition by past and current patients are indeed different, and they are likely affected by various biases. By its very label, though, the ‘BAP’ approach focuses more on biases that affect the ratings of current adapted patients than on biases that may affect past patients. Given that both current and past patients have experienced life with and without the condition being evaluated, and given that both perspectives are likely vulnerable to different biases, opting for the past patient as having the more valid perspective would seem justified only on the assumption that a later judgment is better than the earlier one. Perhaps that is a sensible assumption in life generally; it is not obvious that it is here.

CONDITIONS EXPERIENCED BY REAL PATIENTS
My point in the previous section might well be accepted by Eyal, leading him not to shrink at all from his innovative BAP proposal but only to pursue and use it more cautiously. Another objection to his proposal, however, cuts more deeply.

Some conditions have no curative treatment. Eyal would adjust these, too, for bias in adapted patient ratings if the conditions are sufficiently similar to others which do sometimes have cures and cured patients. He has already refined any BAPs by taking an average across only similar conditions. Bias is likely to differ ‘across patients, health state dimensions, levels of severity . . . , the period for which patients had their health states, and . . . socioeconomic status and subcultures,’ so we should try ‘to identify categories of each condition within which results tend to cluster’ and ‘then calibrate only within those categories.’ On ‘further assumptions’ that he does not delineate, Eyal extends this

---

4 Id. at 9 (manuscript page).
5 Id. at 9 (manuscript page).
sensible refinement and applies BAP even to incurable health states for which there are no past (recovered) patients.\textsuperscript{6}

But even if current patient ratings of curable conditions should be adjusted for BAP, adjustment for incurable conditions is highly questionable. We need to remind ourselves of the nature of subjective value, the kind of value in health-state utility, the context here. All subjective value of a health state is value to actual persons who can experience that state. \emph{There is no subjective value of any health state without actual persons to experience the state.} People who do not have the condition being evaluated can imagine the condition and predict how they would rate it; they would be expressing subjective value, indeed—value \textit{to them}—but of an imagined, hypothetical state. Or if they are already experiencing a state, we might use BAP to predict what their ratings would be if their condition were curable—another expression of hypothetical subjective value. Hypothetical subjective values might be relevant in certain contexts, but in using health-state utility ratings in priority-setting, we aim to maximize the \textit{actual} gains or minimize the \textit{actual} losses from using or not using various treatments. If the condition that will be allowed or prevented is an incurable one, \textit{real} health-state utility gains and losses will accrue only to adapted patients who will likely never become ‘past patients.’\textsuperscript{7} If the real effects that people live with are effects only in the lives of adapted patients, then why should we adjust their ratings to account for the ratings of nonexistent recovered patients? That would not be maximizing any actual health-state utility for real people.

This consideration should not lead us to throw out all adjustment for BAP, but it does reveal a distinct limitation to BAP’s proper use: do not use it to adjust the ratings of incurable conditions.

\textbf{A DEEPER CHALLENGE FOR HEALTH-STATE UTILITY\textsuperscript{8}}

Reminding ourselves of the nature of the subjective value expressed in health-state utility ratings led us to see this limitation. Going back to the basics of health-state utility ratings and how they are used in the construction of a quality-adjusted life year (QALY) uncovers something else important, a deeper challenge, not a mere limitation.

The attention to health-state utility is driven by its role in the unit of benefit, the QALY, constructed for cost-effectiveness analysis. The aim is to discern, through comparison of the value of health benefits gained from treating (or preventing), different sorts of disabling and health-diminishing conditions, what investments produce the largest value in health benefit. To achieve this aim, it will not do just to compare various treatments and preventions for a given disease or condition, nor will it do just to compare one sort of health benefit (like life extension) for different conditions. A

\textsuperscript{6} \textit{Id.} at 5 (manuscript page).

\textsuperscript{7} Unless, of course, a cure comes along in the meantime. In some rare situations, it might be justifiable to count on that, but for most currently incurable conditions (or certainly many), a cure is not likely to come along soon enough in the lifetimes of those who will be impacted by the prioritizing judgments we make on the basis of health-state ratings.

\textsuperscript{8} I articulated a considerable portion of the substance of this section in \textit{Can Cost-Effectiveness Accommodate the Equal Value of Life?} 13 APA [American Philosophical Association] NEWSLETTER ON PHILOSOPHY AND MEDICINE (no. 1, fall 2013), 23–26, www.apaonline.org/resource/resmgr/medicine_newsletter/medicinev13n1.pdf. My discussion there focused more explicitly on QALYs than it does here and less directly on health-state utility.
common unit of health benefit is needed by which different sorts of health gains can be compared. Such a unit must bridge two fundamental kinds of health benefit, life extension and improvements in health-related quality of life (QOL). Without such a common unit of value, one could not compare, for example, kidney dialysis and its lifesaving benefit with hip replacement and its QOL improvement.9

The most plausible approach to constructing the QALY uses trade-offs people are willing to make between life itself (life extension/preservation) and improved QOL. How else but by discerning willing trade-offs between life itself and QOL improvement could a common unit for their subjective value be constructed? One such trade-off method is ‘time trade-off’ (TTO): how much shorter a life in full health does one think preferable to living longer with one’s disability or chronic illness?10 Health-state utility ratings are procured from people by asking them TTO or similarly functioning questions about the health conditions being rated.

In the conventional use of QALYs, life years of the disabled and chronically ill have lower value than the life years of those in full health. This offends widespread convictions, particularly on behalf of those with disabilities, that each person’s life has equal intrinsic value to them. A case can be made, though, that the use of QALYs with their non-equal values of life is not unfair. To be sure, within the QALY framework, saving a person with paraplegia is thought to produce less health-related value than saving someone for the same number of years of full health. This does disadvantage the person with paraplegia in competing for lifesaving resources. But suppose that in eliciting health-state utility ratings, the TTO questions are put not to members of the general public imagining themselves with paraplegia, but to people who actually live with it. Then a rating of such life expressed by a willingness to trade some time-in-life to regain the use of one’s limbs is a judgment that disadvantages that very person with paraplegia in any competition for lifesaving resources. Why would a person disadvantage herself this way by indicating a QOL less than 1.0?11 An answer is that by rating their QOL as less than 1.0, the person gains some advantage in competing for curative measures, which we could call the ‘QALY bargain’.12

Relative to a high 0.98 response, for example, responses by persons with paraplegia expressing a 0.9 rating do disadvantage them in competing for lifesaving resources, but

---

9 For a superb summary of the QALY, see Milton C. Weinstein, George Torrance, and Alistair McGuire, QALYs: The Basics, 12 VALUE IN HEALTH S5 (Supplement 1, 2009).
10 Suppose, for example, that people with paraplegia are typically willing to trade away 10% of their remaining lifetime to regain full limb function; they will then have rated their QOL and the relative subjective value to them (individual utility) of life with paraplegia at 0.9 out of a possible 1.0. Saving such a person for 10 years of life would achieve a 9.0 QALY gain. Restoring her to full function for 10 years would gain her 1.0 QALY (0.1 gain for each year). In full CEA, of course, costs get included, too. For a classic example of a full CEA using such methods, see the comparison of hemodialysis with hip replacement by Alan Williams in the course of reporting a larger study focused primarily on coronary bypass: Economics of Coronary Artery Bypass Grafting, 291 BRITISH MEDICAL JOURNAL 326 (1985).
11 Some are not willing to sacrifice any life expectancy to relieve their condition. See F. J. Fowler, P. D. Cleary, M. P. Massagli, et al., The Role of Reluctance to Give Up in the Measurement of the Value of Health States, 15 MEDICAL DECISION MAKING 195 (1995). The challenge this constitutes for CEA is noted by Erik Nord, Norman Daniels, and Mark Kamlet, QALYs: Some Challenges, 12 VALUE IN HEALTH S10 (supplement 1, 2009), at S10–S11.
12 The term is used by Menzel in Measuring Quality of Life, STRONG MEDICINE: THE ETHICAL RATIONING OF HEALTH CARE 79 (1990), at 86.
this lower rating advantages them in competing for QOL improvement care. That’s the ‘bargain’ people are involved in when they rate their own QOL as noticeably less than 1.0, and their society uses QALYs to help prioritize health services. If it is not an unreasonable bargain (and it doesn’t seem to be, if one is operating within the strictures of the QALY framework), one can claim that using QALYs does not create a net disadvantage at all for the disabled and chronically ill.

Though plausible, however, this argument for the fairness of QALYs conflicts with deeply rooted, widespread convictions about the equal value of life (EVL): that both for those in full health and those in all but the most difficult and despairing health conditions, ‘life itself’—the very business of being alive at all—has equal intrinsic value.13 Defenders of QALYs may dismiss such convictions as not facing up to the realities of the toll that illness and disability take on human well-being. The EVL claim, however, is sustained if we pay careful attention to what is, and is not, expressed in the very health-state utility ratings used in QALY computation.

Suppose the person with paraplegia said that saving her life for an additional 10 years has the same value to her as 9 years of additional life with restoration of full limb function. In saying that, she has not said that her life itself has any less value for her than the life itself of another person, without paraplegia, has for that person. Moreover, not only does she contend that life itself for her is as valuable as the other person’s life is for that person, but others, too, can quickly come to see her point and agree. It does not take much reflection on their part to do so.

This poses a huge problem for QALYs and the health-state utility ratings they involve: EVL (a 1.0 rating) seems impossible within the conventional QALY framework, for such a rating reduces the value of QOL improvement measures alleviating the condition to zero. Within any framework in which the value of life extension and of quality enhancement are brought under a common unit of value, one must choose either (i) stick with EVL and relinquish the value of cure or (ii) relinquish EVL and retain value for cure.15 Disabled or chronically ill persons will not relinquish either. They likely insist on both the EVL and a significant value for QOL improvement. And if we were in their shoes, would we not also? Moreover, even staying in our own shoes, if we understand their point, won’t we, too, insist on both EVL and significant value for QOL improvement?

It seems, then, that the conventional framework for QALYs cannot accommodate both EVL and significant positive value for restorative cure, two extremely basic and resilient convictions people have. This constitutes a fundamental challenge to conventional QALYs and the health-state utility ratings that go into them. What is being rated/evaluated in a person’s TTO responses (or in responses to any of the other methods of eliciting health-state ratings)? If it’s life itself, then chronically ill or disabled

---

13 The challenge for cost-effectiveness analysis and QALYs posed by EVL is noted by Erik Nord, Daniels, and Kamlet, supra note 11, at S10–S11.
14 That is, her life compared to death. The value of life itself—not the quality of improved life—can only be ascertained by comparison with life’s absence. For actual persons, that absence is death, since we are not talking about existence vs. not coming into existence.
15 This is the ‘QALY Trap,’ a phrase first used (to my knowledge) by Peter Ubel, Erik Nord, Marthe Gold, Paul Menzel, Jose-Luis Pinto-Prades, and Jeff Richardson, Improving Value Measurement in Cost-Effectiveness Analysis, 38 MEDICAL CARE 892 (2000).
persons will likely rate a year of life extension as 1.0 (or close), equal or nearly equal in subjective value to others’ valuation of their life years without illness or disability. On the other hand, if their focus is quality of life and any potential improvements (or diminishments) in it, they will typically rate their health-state utility as distinctly less than 1.0. ‘If we do not know which of these they are rating, we have no way of assigning a meaning to the ratings they express.’

Perhaps we should just make clear which of these we are asking them to value. Solicit these two valuations in different ways that eliminate ambiguity and confusion, and then note the ratings that get expressed by patients for these two different things and use them appropriately for different kinds of treatment. For lifesaving/life-extending purposes we would use one, for QOL improving purposes the other. Many treatments are one or the other, in which case separate use, though a bit messy, would still be feasible. For treatments that are both, the challenge is more difficult. And in a society’s prioritizing decisions, treatments with the two respectively different sorts of benefit often compete with each other. There may or may not be a way out of this problem.

What does this problem in the QALY/health-state utility framework mean for the proposal to use BAP? Until we know what it is that patients are rating, adjusting for BAP will be treacherous, to say the least. And how do we know that adapted patients have the same potential biases, in the same direction, for both these sorts of valuation? Similar questions apply to recovered patients. We will need thorough additional analysis of bias for both adapted and past patients, and for each, for both kinds of health-state utility.

This, as well as handling the limitation that I previously argued should be put on the use of BAP for incurable conditions, provides plenty of work for future refinement of any approach that would adjust for the bias of adapted patients. It is not clear to me that BAP adjustment has enough promise in the face of these challenges to warrant that effort. What is clear is that a fundamental modification of the way we conceive of, measure, and use health-state utility is in order. That should come first and refinements like adjustment for BAP later.

16 One such revision has been provided by Erik Nord, Beyond QALYs: Multi-criteria Based Estimation of Maximum Willingness to Pay for Health Technologies, 19 European Journal of Health Economics 267 (2018). DOI: 10.1007/s10198-017-0882-x.