Ratio

WILEY

SPECIAL ISSUE ARTICLE OPEN ACCESS

Cost-Effectiveness Analysis of Risky Health Interventions: Moving Beyond Risk Neutrality

Johanna Thoma

Kulturwissenschaftliche Fakultät, Universität Bayreuth, Bayreuth, Germany

Correspondence: Johanna Thoma (johanna.thoma@uni-bayreuth.de)

Received: 27 August 2024 | **Accepted:** 30 September 2024

Keywords: cost-effectiveness analysis | health economics | paternalism | public policy | risk

ABSTRACT

Cost-effectiveness analysis for health interventions is traditionally conducted in a risk-neutral way, insensitive to risk attitudes in the population, which are potentially non-neutral. While the standard outcome metric of quality-adjusted life years (QALYs) aims to be deferential to people's valuations of health states, cost-effectiveness analysis of risky interventions using the QALY metric is not similarly deferential to people's risk attitudes. I argue that there is no good justification for this practice. Non-neutral attitudes to risk, especially where they concern individually life-changing interventions need not be irrational, and so imposing neutrality is not justifiable as a way of debiasing preferences. Many common justifications for deference to health state preferences extend to risk attitudes. But even if reasons for deference do not extend, imposition of risk neutrality as opposed to any other rationally permissible risk attitude is under-motivated as default practice. Thus, either methods for measuring risk attitudes separately and incorporating them into cost-effectiveness analysis should be used more widely or a richer set of information should be presented to political decision-makers and the public to enable them to decide how to take into account the individual risks faced by members of the population, on top of aggregate effects on population health.

1 | Introduction

Most countries have institutions tasked with evaluating the cost-effectiveness of health technologies and public health interventions. For instance, the National Institute for Health and Care Excellence (NICE) plays this role in the United Kingdom, the Institute for Clinical and Economic Review (ICER) plays this role in the United States and the Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) plays this role in Germany. The evaluations produced by these bodies, and cost-effectiveness analyses more generally, often play a crucial role in health policy and setting standards for clinical practice. Cost-effectiveness analysis generally aims to establish to what extent we can expect health-related value for money from an intervention and compare that to possible alternatives. As such, it needs a metric of value. The most commonly used metric of value for cost-effectiveness analysis in the context of health is quality-adjusted life years (QALYs). One year spent in perfect health corresponds to one QALY. Time spent in less than perfect health is weighted down using weights representing degrees of health-related quality of life.

There are a number of methods for determining quality weights, some of which will be introduced below. What they have in common is that they display an ambition of deference to the relevant population's values. Quality weights are meant to measure health-related quality of life, *as judged by the potential recipients of healthcare themselves*. Methods to determine quality-weights thus all proceed by trying to elicit or infer the relevant judgements in the population. What I will argue is that this ambition of deference is in tension with another feature of cost-effectiveness analysis as commonly practiced: Under conditions of uncertainty, where we are not sure either about the cost or about the health-related effects of interventions, cost-effectiveness is

assessed in terms of the *expected* cost per QALY. Uncertainty is of course a prevalent and unavoidable feature of decision making in the health context. I am here especially concerned with uncertainty at the level of individual patients, where we do not know what the potential health outcomes for individuals will be as a result of the intervention, but can assign probabilities to the potential outcomes.

The practice of conducting cost-effectiveness analysis under conditions of uncertainty in terms of *expected* cost per QALY, I will show, imposes risk neutrality on patients who may not be risk neutral in their pursuit of health (Section 2). I will argue that there is no good justification for analysts to present policymakers, practitioners and the public only with a risk-neutral assessment (Section 3). I then explore a number of possible alternatives (Section 4). I will argue that methods for eliciting and incorporating non-neutral attitudes to risk in cost-effectiveness analysis should be further developed and more widely used and/or reports should be enriched with information on individual-level risks.

2 | Cost-Effectiveness Analysis For Health: The Risk-Neutral Default

Suppose you care about winning matches in your friends' mini-golf league. You can choose between either winning one match for certain or winning two matches with a 55% probability. If you go with the one safe win, you are *risk averse* in your pursuit of mini-golf wins. A *risk-neutral* contestant would maximise the expected number of wins, which is higher (1.1 instead of 1) when you attempt the double win. As a risk-neutral player, by contrast, you are willing to go for a lower expected number of wins for the sake of playing it safe.

The idea that people can be risk averse in their pursuit of goods like mini-golf wins or money, and that there is nothing irrational about this is entirely uncontroversial. Recent decision-theoretic literature has, however, also defended a stronger but related claim: people can be risk averse in their pursuit of subjective value, of what is valuable by their own lights. To illustrate this, suppose we make the further assumption in the mini-golf case that you value every win to the exactly same extent: They all bring you the same joy, pride, bragging rights, or whatever else you might find valuable in winning. If you are still risk averse once we have made this assumption, this implies you are not only risk averse with regard to mini-golf wins, you are also risk averse with regard to subjective value—you are not maximising its expectation either. Your risk aversion with regard to mini-golf wins thus cannot be explained in terms of there being decreasing marginal value of additional wins the more you have already won. The stronger claim is that even such 'pure' risk aversion, risk aversion in the pursuit of subjective value, is common as well as often rationally permissible. It reflects a commitment to the idea that rationality is permissive under risk (see also my Thoma 2023, 2024). Its proponents usually also defend the same claim regarding risk inclination. They need not hold that anything goes. Indeed, there may be tight constraints on the extent and the structure of the non-neutral risk attitudes we allow. But as long as we allow for more than one—the risk neutral—attitude to risk, we accept that rationality is permissive under risk.

Importantly for our purposes, it is intuitively very plausible that the permissiveness thesis holds in the case of health-related value. Take these two (highly) idealised cases of treatment choice:

- *Sick Patient*. Imagine a sick patient who disvalues every day spent in their current impaired health state to the same extent and would be sick for at least half a year without intervention. Sick Patient is offered this choice: Would you rather...
  a. shorten your ordeal by 45 days for certain, or
  b. have a 50% chance of shortening it by 100 days, and not at all otherwise?

- *Healthy Octogenerian*. Imagine a healthy octogenarian who values every additional month in good health to the same extent. Healthy Octogenerian is diagnosed with a condition that will eventually end her life after a short period of illness. The onset of illness can potentially be delayed, and in any case she will have enough time to bring her affairs in order. She is offered this choice: Would you rather have...
  a. 45 more months of good health for certain, or
  b. a 50% chance of 100 more months of good health, and none otherwise?

In both of these cases, if time spent in the relevant health state really is valued to the same extent across the potential outcomes (which may for many reasons not be strictly the case in more realistic settings), choosing option *a* would exhibit pure risk aversion. And in both of these cases, this does seem both like a choice that many people would make, and like one that we would not deem irrational.

Risk aversion (or risk inclination) with regard to subjective value can be formally accommodated in different ways by different decision theories.[1] Notably, it can also be accommodated by standard expected utility theory (EUT). As the name suggests, in EUT, agents maximise the expectation, that is, probability-weighted sum of the utilities of various potential outcomes. This implies risk neutrality with regard to *utility*. But some interpretations of EUT do not commit us to the claim that utility is a cardinal measure of the extent to which agents value a good. We could then coherently say that agents have decreasing marginal utility in subjective value, and that this is how we represent and accommodate pure risk aversion.

However, there are also alternative decision theories that more explicitly separate risk aversion from the subjective value of outcomes. One approach is to assign additional value or disvalue to a risky option, in accordance with its riskiness, over and above the expected value of its outcomes (see, e.g., Stefansson and Bradley 2019, Goldschmidt and Nissan-Rozen 2020 and Weirich 2020). Another is to use rank-dependent utility theories, which weight the utilities of outcomes not just by their probabilities, but also by a risk function expressing degrees of risk aversion (see Buchak 2013).

Whether we use a framework that accommodates pure risk aversion has special significance in contexts where we make decisions on behalf of other people—the patients—and aim to

defer to their interests as they themselves see them. Suppose we use a framework that imposes risk neutrality. It might then be that, even though we correctly identify how a patient evaluates the potential outcomes, we make a decision on her behalf that she would not have made and would not approve of. For instance, in our examples above, we would be imposing option *b* on a risk averse Sick Patient and Healthy Octogenerian, respectively. Elsewhere, I have called this phenomenon 'pure risk paternalism' (Thoma 2023).

What I want to argue now is that standard approaches in cost-effectiveness analysis for health interventions are guilty of precisely such pure risk paternalism: Even when ideally implemented, they impose risk neutrality on a population that may not be risk neutral regarding health-related value and without having taken reasonable steps to investigate and accommodate this potential non-neutrality. I will later consider whether this practice is problematic or not.

Standard approaches to cost-effectiveness analysis for health interventions display an ambition of deference to the population's values. This shows in particular when we look at the way that the quality weights in QALYs are conceptualised and measured. The most prominent methods for doing so all derive them from elicited preferences or judgements in the population—though, importantly, heterogeneity is rarely allowed for, so deference is to the average patient rather than the individual patient. Whether this lack of accommodation of heterogeneity already amounts to a problematic form of paternalism—one exemplifying an attitude we may no longer consider exhibiting an ambition of deference—depends on just how much heterogeneity there is and how difficult it would be to accommodate it.

The starting point are typically generic sets of health states, such as the EQ-5D instrument, used by many health bodies and research institutions around the world. It identifies five dimensions of health—mobility, ability to care for oneself, ability to do usual activities, pain/discomfort, anxiety/depression—as well as a different levels of impairment, most commonly three. A generic health state consists of a level of impairment assigned to each dimension. Preference or value elicitation is then used to produce weights comparing those generic states to perfect health. Ultimately, in order to produce a cost-effectiveness analysis relating to specific real-world interventions, the generic health states will need to be matched to the specific health states actually experienced in the population following an intervention.

The most commonly used methods to produce quality weights using the EQ-5D are visual analogue scale (VAS) and, even more commonly, time-tradeoff (TTO) methods. In the case of VAS, subjects are asked to evaluate their health states (first encoded using the EQ-5D) on a scale, for instance a sliding scale ranging from 0 ('the worst state you can imagine') to 100 ('the best state you can imagine'). In the case of TTO, subjects are asked to indicate their preferences between longer periods of time in an impaired health state, compared to shorter periods of time in perfect health. Points of indifference are then used to infer quality weights, under the (controversial) assumption that time spent in a particular health state is valued linearly.

Notably, neither of these methods takes into account attitudes to risk, as quality-weights are elicited under conditions of certainty (I will comment below on the less commonly used standard gamble (SG) method which elicits preferences under uncertainty). It is also generally agreed that if these methods achieve what they set out to do, they provide a measure of the subjective value of health states (see Broome 1993; Hausman 2015). Moreover, the measure they provide is a cardinal (or even ratio scale) one: It not only allows us to order health states in terms of their subjective value, we can also compare the sizes of differences between them. In other words, they allow us to measure the degree to which impaired health states are worse than perfect health. TTO methods and VAS methods in practice do not always agree, which is evidence that they may not in fact always succeed at measuring what they set out to measure. I will set aside these methodological worries here. What is important for us is that, because in both cases weights are elicited under conditions of certainty, attitudes to risk do not affect these cardinal weights.

In the standard approach, there is then also no further point at which the population's attitudes to risk affect cost-effectiveness analysis. When there is quantifiable uncertainty about the effects of a health intervention either on the individual or at population level, what is maximised are *expected* QALYs, that is, probability-weighted sums of potential QALYs (see Bilcke and Beutels 2022). If we combine this with QALYs whose quality-weights represent degrees of subjective health-related value, this amounts to implementing risk neutrality with regard to health-related value. Regarding risk attitudes, then, standard approaches do not show an ambition of deference: Risk neutrality is implemented whether the population is risk-neutral or not.

Sick Patient and Healthy Octogenarian already illustrate the plausibility of the claim that patients need not and do not always exhibit risk neutrality with regard to health-related value. There is, moreover, evidence supporting this claim. For one, applying rank-dependent frameworks yields better fit for non-neutral risk-weighting functions than the neutral one, both within health contexts (Attema et al. 2016) and more generally (Harrison and Swarthout 2023). Moreover, in studies employing an expected utility framework allowing for non-linear utility in subjective quality of health states (as measured by TTO methods), risk aversion is also frequently found (see Rosen, Tsai, and Downs 2003).

In theory, lack of deference to non-neutral risk attitudes could lead a policy-maker to choose interventions that are dispreferred by all patients affected by it. Take a stylised case along the lines of the ones given above, but this time in terms of TTO-elicited QALYs. A policy-maker can choose to fund either one of two treatments for patients suffering from a particular condition, which are equally costly. One of the treatments has uncertain outcomes at the individual level, where the probabilities are independent between different patients. The treatments have the following effects at the individual level:

- *Treatment A*: Certain gain of 2 QALYs.
- *Treatment B*: 50% chance of a gain of 5 QALYs, none otherwise.

The standard approach would recommend offering Treatment B, as it leads to more QALYs in expectation. However, suppose everybody in the population is risk averse to the extent that they would prefer Treatment A. In that case, the standard approach would recommend a treatment everybody disprefers, as a consequence of imposing risk neutrality on a risk-averse population.

This is a stylised and extreme example. But the general phenomenon can be observed in real case studies, resulting in policies that 'don't make sense at the bed-side' (Asch and Hershey 1995). Cher, Miyamoto, and Lenert (1997), for instance, compare a 'watchful waiting' (WW) approach to benign prostatic hyperplasia (BPH) with surgical intervention (transurethral resection [TUR], as practiced at the time). The advantages of WW are that an immediately risky surgery is avoided, while spontaneous remission may occur. TUR, on the other hand, immediately reduces the risks and discomforts associated with BPH if successful and obviates the potential need for an even riskier surgical intervention at a later time.

Using TTO-elicited QALYs for an otherwise healthy 70-year old sexually active man, a risk-neutral base case model finds that TUR is the better option in terms of expected QALYs, in line with a previous study. However, the authors then use a risk-adjusted (RA-)QALY model to see how the analysis changes once we assume mild to moderate risk aversion (more on these methods below). In fact, in that model, WW comes out ahead in terms of expected RA-QALYs, as the immediate risks of surgery are given greater relative weight. Of course, for a cost-effectiveness analysis, not only expected QALYs but also expected costs matter. But if WW is not much cheaper than TUR in expectation, a risk-neutral assessment would judge TUR to be more cost effective. And this may be used to justify a policy of recommending, or exclusively funding, TUR. But such a policy would impose a dispreferred option on any patient with mild to moderate or stronger risk aversion, which may be a significant part of the population. And it may do so even if WW is actually cheaper.[2] This constitutes pure risk paternalism as I outlined it above.

We have seen that the standard approach to cost-effectiveness analysis in health imposes risk neutrality on populations that may not be risk neutral. The question this raises is whether this practice is nevertheless justifiable. The next section explores and ultimately rejects a number of candidate justifications.

3 | Potential Justifications For Risk Neutrality And Their Failure

A first potential justification for a risk-neutral approach may appeal to the idea of debiasing people's preferences when they are irrational, in line with a prominent research programme in behavioural welfare economics. It is often argued that correcting for irrationality is paternalistic only in an unproblematic sense, as it ultimately helps people better achieve their objectives as they themselves see them.[3] The argument in case of risk aversion could be: QALYs capture patients' subjective values; the only rational way to pursue value under uncertainty is expected value maximisation; hence, QALY-based cost-effectiveness analysis is paternalistic only in an unproblematic sense: it helps people pursue their goals in the only rationally permissible way, and if this does not match their own preferences, this only goes to show that they are irrational.

The difficulty with this argument lies in establishing that risk neutrality is really rationally required. At times, authors appear to take risk neutrality to follow from EUT, and EUT to be supported as the correct theory of rationality by the plausibility of the axioms from which it can be derived. But this simply rests on a misunderstanding of EUT: the axioms of prominent representation theorems establish only that agents can be represented as expected utility maximisers, not that the utility function that represents them must be a cardinal measure of the degree to which they subjectively value outcomes (see, e.g. Dyer and Sarin 1982 on this regarding von Neumann-Morgenstern utility theory). In the health context, for instance, an expected utility maximiser who is risk averse, where her risk aversion can be represented as decreasing marginal utility in TTO-elicited QALYs, would still abide by all the standard EUT axioms.

This is not to say that those who have non-neutral pure attitudes to risk do not often violate the standard axioms of EUT—they do, which is part of the case for using rank-dependent frameworks to represent them for descriptive purposes at least. But, for one, some have argued that in fact, rank-dependent models are also normative, and agents are permitted to violate EUT as long as they abide by the requirements of a rank-dependent framework (most prominently Buchak 2013) in which case there is no case for correction. But even if we think that rationality requires agents to be expected utility maximisers, a model that corrects their preferences need not be risk neutral—risk-averse implementations of EUT may be more faithful to the preferences they started out with.[4]

What speaks against the idea that rationality requires pure risk neutrality is the intuitive rational permissibility of non-neutral attitudes in the kinds of examples we looked at above. Accordingly, as we saw above, more and more decision theorists have recently come around to the idea that rationality is permissive under risk. Against that, are there strong arguments that establish a requirement to be risk neutral over and above a requirement to follow EUT axioms?

The strongest arguments in my estimation are those that appeal to the long-run disadvantages of having non-neutral attitudes to risk, in cases where we take risks repeatedly (see my Thoma 2019, but also Wilkinson 2022). For instance, if you regularly refuse small monetary gambles with positive expected monetary value, you will almost certainly end up worse off over time. If that is not an outcome you want, you may have reason to adjust your risk-averse behaviour. The problem with appealing to such arguments for our purposes is that in many health applications we are dealing with the impacts of serious health conditions, as well as interventions that have a lasting impact on people's lives, of the type that people do not face many times in their lives. The applicability of long-run arguments is thus doubtful. Debiasing, in sum, does not seem to be a valid justification for the risk neutrality of the standard approach.

A second potential justification starts from the observation that, while individual patients often only face treatment for a

particular serious condition once, policy-makers get to decide on many such cases all at once. On that basis, it is often claimed that individual-level risks, as least, wash out in the aggregate and are irrelevant from the policy-maker's perspective (see, e.g. Drummond et al. 2015, 136 and Parkin and Devlin 2006). Indeed, due to the Law of Large Numbers, interventions with higher expected QALYs at the individual level, even if risky, are often extremely likely to also lead to more actual QALYs at the aggregate level. Suppose, for instance, that a policy-maker gets to choose whether to fund Treatment A or Treatment B above for 10,000 patients. Given independent probabilities, this would have the following results:

- *Treatment A*: Save 20,000 QALYs for sure.
- *Treatment B*: Save an expected 25,000 QALYs, where the chance of saving fewer than through Treatment A is virtually 0.

If the objective of public health policy is to bring about as much health-related value as possible (in a way that is deferential to people's subjective values), then Treatment B looks like a no-brainer once we take the aggregate perspective.

The problem is that all that such examples show is that high individual-level risk can be consistent with little social risk. There can be very little spread in potential outcomes for overall population health resulting from an intervention, while at the same time every individual experiences a large spread in the potential health outcomes they individually face. The fact that risk washes out at the social level does not do away with the possibility that every affected patient may reject the impact of a policy on them.

What proponents of the idea that individual-level risk becomes irrelevant at the social level would need to establish is that only the social-level risk matters.[5] Here we face a problem: Either we value population health only insofar as pursuing it means pursuing the health of individuals. To nevertheless argue that individual-level risk does not matter, we would then need to establish that a risk-neutral approach is a legitimate way to pursue an individual's health irrespective of her own risk attitude. But for that we would need to rely on one of the other potential justifications of risk neutrality considered in this section. Or otherwise, putting social-level risk first implies that the value of population health is *impersonal*—that one can pursue it even without pursuing any individual's health.

However, an impersonal notion of the value of population health is problematic. There are certain values that are plausibly impersonal, such as ex post equality.[6] Where this is the case, this often stems from these values being instantiated in social structures or relationships. But instrinsically, health is first and foremost of value to individuals. An impersonal perspective would view individuals as mere vessels for health, rather than it being the job of public health policy to make individuals healthy. Total population health may of course have important instrumental benefits for society as a whole.[7] But it should be equally uncontroversial that the intrinsic value of health to individuals should at least be an important policy objective, above and beyond what health may do, e.g. for economic productivity.

A third potential justification for risk neutrality is appeal to a kind of practical division of labour. Perhaps, one might think, cost-effectiveness analysis for health, and health economics more generally can focus exclusively on the goal of health, and accommodating non-neutral attitudes to risk could be taken care of at some other stage of policy analysis and decision making. There is plausibility to the general idea of a division of labour when it comes to policy analysis.[8] Health economics self-consciously looks only at one aspect of quality of life, namely health-related quality of life. Insofar as policy-making that impacts health involves tradeoffs with non-health-related quality of life, the presumption already is that on the one hand, cost-effectiveness analysis for health only delivers a *partial* assessment, and that thus, on the other hand, such tradeoffs will be made by policy-makers once also taking into account further partial assessments.

However, regarding risk attitudes in particular, leaving the incorporation of non-neutral attitudes to others to assess either does not make conceptual sense or at the very least is not practicable. First, there is one prominent way of thinking about what non-neutral attitudes to risk are on which such a division of labour does not make conceptual sense. The kind of division of labour envisioned, and that is common in policy evaluation, is division labour according to policy goals—e.g. health, sustainability, equality of opportunity, etc. But on this first interpretation of risk attitudes, the pursuit or more likely avoidance of risk is not a goal alongside such other policy goals. Attitudes to risk, rather, capture preferences over *how* such other goals are to be pursued: Do we pursue them in a more risk-inclined or a more risk-averse kind of way? This is the way in which, e.g., Buchak (2013) understands non-neutral attitudes to risk, and it is a natural interpretation within a rank-dependent framework like hers, which separates risk attitudes from the utility function. And this interpretation is also natural when accommodating pure risk aversion by allowing for decreasing marginal utility in subjective value within an expected utility framework.

Granted, there are other ways of thinking about what attitudes to risk are. We have seen that in some frameworks, non-neutral attitudes to risk are accommodated by assigning additional utility or disutility to risk itself, over and above the expected value of the outcomes (Stefansson and Bradley 2019; Goldschmidt and Nissan-Rozen 2020; Weirich 2020). Indeed, Stefansson and Bradley (2019) explicitly think of risk attitudes as a way of valuing or disvaluing risk itself. And so on this interpretation, it at least makes conceptual sense to separate out evaluation in terms of risk-related goals from evaluation in terms of other goals.

Still, even under this interpretation, division of labour would be impracticable. What is at issue are risks related to health outcomes, so it is unclear who but those carrying out cost-effectiveness analysis for health would supply the information and assessment necessary to take riskiness into account in a later all-things-considered policy choice. Perhaps this need not be as part of the cost-effectiveness analysis itself, but instead in the form of supplementary information, something I will explore in the next section. But either way, it should be the job of the very same people to supply both.

A final potential justification for risk neutrality in cost-effectiveness analysis for health could be an argument that deference to risk attitude is not required after all. One observation that might help here is that everybody seems to agree that risk neutrality is at least rationally permissible for individuals in the health domain, even if it is not rationally required. And perhaps policy-makers are just required to adopt one of the permissible risk attitudes, not to defer to the population's risk attitudes.

The prospects for this argument, too, depend on what exactly we think non-neutral attitudes to risk are. If they are, as Stefansson and Bradley (2019) and others argue, just ways of valuing or disvaluing risk itself, then any reasons for being deferential to the population's attitudes to health outcomes extend straightforwardly to risk attitudes. If we think deference is required regarding health-related quality of life at all—as most health economists agree and as is implemented in all the standard approaches—then it also seems to be required regarding risk attitudes.

If, on the other hand, we agree with Buchak (2013) and others that attitudes to risk are preferences over *how*, out of a number of permissible ways, the goal of health-related quality of life should be pursued, then the argument that deference is not required is potentially more promising.[9] Still, however, there would remain the question of what would justify the risk-neutral default. Healthcare is of central importance in people's lives. And while sometimes, risk attitudes are formed by people on the spot without deep commitment, when it comes to potentially life-changing interventions, how we evaluate the risks involved can be something we feel very strongly about. Moreover, we have seen that risk attitude can make a crucial difference in cost-effectiveness analysis. Given these observations, a merely arbitrary decision over what risk attitude to implement in cost-effectiveness analysis does not seem satisfactory. In the absence of a good positive reason to default to risk neutrality, deference would remain a good non-arbitrary default.[10]

The problem then is that, partly for the reasons already canvassed, we lack a compelling positive reason to default to risk neutrality. In addition to the failed justifications already discussed, one might perhaps think that, in a context where some people may be risk averse and others risk inclined, risk neutrality is a good, 'neutral', as it were, compromise. But for one, if rationality is permissive under risk, risk neutrality is not 'neutral' in the normative sense of being free of bias or distortion, in contrast to risk aversion and risk inclination. Moreover, whether risk aversion and risk inclination cancel each other out in the population is an empirical question and will likely also differ depending on domain. Most evidence suggests that risk aversion is more common than risk inclination, including in particular in the health context (see Attema et al. 2016; Mulligan et al. 2024).

Another justification for a risk-neutral default one could potentially appeal to is the greater simplicity of risk-neutral analysis. However, for one, if the assumption of risk neutrality makes a crucial difference for policy recommendations that have a profound effect on people's lives, simplicity does not seem a good enough reason for the analyst to impose that assumption. And second, as the next section will show, at least some ways of going beyond the risk-neutral default are not technically onerous and do not sacrifice much of the simplicity of risk-neutral analysis.

4 | Alternatives To Risk Neutrality

The last section discussed four potential justifications for the risk neutrality of the standard approach in health cost-effectiveness analysis and found them all wanting. How, then, could we do better? I will first look at one solution that does not work and then present two general approaches that would present improvements on the status quo.

I mentioned above that next to TTO and VAS, SG methods for eliciting QALY weights are also often advocated. These are much less frequently used in practice than the other two, even though they are often described as theoretically preferable. These methods ask individuals to compare a certain health outcome with a risky gamble that may either lead to a better or worse outcome. Points of indifference are then used to infer quality weights. Since SG methods elicit preferences in the context of risk, one might think that they thus do take account of attitudes to risk, thereby solving the problem I have been analysing.

Unfortunately, however, the more widespread use of SG methods would obfuscate rather than solve the problem of pure risk paternalism. For one, these methods assume that patients are expected utility maximisers and will not lead to reliable measurements for agents who are not. But, as we have seen, at least some of the ways of accommodating pure attitudes to risk that have been advocated in the literature involve accepting generalisations of or an abandonment of EUT. And we have seen evidence that alternative models fit actual population preferences better, including in the health setting.

Second, it can be shown that even for people who are expected utility maximisers, the SG method only provides a valid cardinal measure of health-related quality of life for agents who are also risk neutral (see Johannesson 1995). This makes sense given what we said above: Agents who are purely risk averse or risk inclined regarding subjective health-related value, but nevertheless abide by expected utility axioms can be accommodated by allowing for decreasing or increasing marginal utility in subjective health-related value, respectively. And then we cannot, as the SG method does, take the utility elicited through the method to be a cardinal measure of subjective health-related value. This means, at least, that the SG method should not be used to inform decision making in riskless contexts.

What about using the SG-elicited utilities to inform other choices under risk? The problem with inferring a cardinal measure of health-related quality of life using the SG method was essentially that the utility function thus elicited captures a mix of subjective health-related value as well as pure attitudes to risk, with no way to pull them apart. This is not necessarily a problem if the kind of risky health gambles evaluated in the cost-effectiveness analysis are very similar to those used to elicit the utility function. We could then be fairly confident we are in fact deferring to the combination of the patients' subjective health-related values and risk attitudes. But if the risky

gambles to be evaluated are different in their stakes or structure to those used in the elicitation, as they often are in applications, the inductive leap would be much bigger. Without separating out risk attitude and subjective health-related value, it is difficult to know whether we would really still be tracking what patients would prefer. Because of these ways in which the method can lead us astray—essentially by obfuscating the issue of potential non-neutral pure attitudes to risk—it seems preferable to adopt an approach that accommodates non-neutrality, while also explicitly separating pure attitudes to risk from subjective health-related value.

A first type of approach that would fit this bill is to use alternative formal frameworks that allow us to incorporate non-neutrality directly in the cost-effectiveness analysis.[11] This is the approach that was adopted in the Cher, Miyamoto, and Lenert (1997) study on treatment of BPH, as well as in Woodward, Schnitzler, and Kvols (1998). Their alternative analysis uses risk-adjusted (RA-)QALYs. They operate within an expected utility framework and incorporate different levels of risk aversion by allowing for decreasing marginal utility in health-related quality of life—which is itself elicited using standard TTO or VAS methods. A very similar approach, which they call generalised risk-adjusted cost-effectiveness (GRACE), is also advocated for by Lakdawalla and Phelps (2022), which they argue 'better aligns value assessment with the preferences of real human consumers' (445).

It is also conceivable to use non-expected utility frameworks, such as rank-dependent ones, in a similar way. Just as there are prospect theoretic models for eliciting subjective health-related value (see Attema et al. 2016), one could imagine using prospect theoretic models or other rank-dependent models like ones employing Buchak's (2013) risk-weighted EUT to introduce non-neutral attitudes to risk back into cost-effectiveness analysis while holding on to traditional riskless methods of eliciting subjective health-related value.

As far as I know, the latter approach has not been developed, while the former has had very little uptake in practice. But such alternative formal frameworks would allow us to do one of two things, both of which would seem to be an improvement over the status quo: On the one hand, we could investigate what the population's average pure attitudes to risk are, and then implement those. If feasible, we could also allow for some heterogeneity and look at what risk attitudes different groups in society tend to have. Heterogeneity could be expected, for instance, along age lines. But note that the standard framework does not usually take into account heterogeneity even for quality weights, where doing so seems desirable for similar reasons.

Alternatively, we could use such models to investigate how an assessment changes under a range of reasonable pure attitudes to risk. Where risk attitude crucially changes recommendations, we could allow for greater choice on the ground level, provided greater choice is not itself too expensive and does not lead to cognitive overload. E.g., we could allow for a range of treatments to be available without clear recommendation, and for the decision of which to use to be made at the bedside between health professional and patient. For those for whom a deferential approach to cost-effectiveness analysis is motivated by a commitment to patient autonomy, greater choice is more generally attractive on autonomy-grounds anyway. Moreover, it can go some way towards accommodating heterogeneity in patient values more generally. Makins (2023) makes the case for such greater choice for precisely the reason that there are differences in risk attitudes.

These alternative formal approaches are, however, much more complex to implement than the standard approach. Moreover, they make many of the other restrictive assumptions of the standard approach, and hence allow for non-neutral attitudes to matter only within the confines of this model. For instance, they do not incorporate, e.g. non-neutral attitudes regarding length of life.

An alternative approach that would both be less formally challenging and also has the potential to avoid this problem would be to pass on more risk-related supplementary information to both policy-makers and the public and make it explicit that this supplementary information may well provide reason to override the recommendation of a risk-neutral cost-effectiveness analysis. NICE in fact already requires cost-effectiveness analyses to be accompanied with estimates of social-level risk, that is, of how likely is it that an intervention is the most cost-effective on an aggregate level—which would result, for instance, from uncertainty about how generally effective a treatment is or how costly it will be. ICER, too, provides a range of supplementary information. However, there is no requirement in either case to report on individual-level risk. But doing so would be crucial to avoid pure risk paternalism.

Including such information would allow policy decision-makers to either potentially deviate from always choosing the most (risk-neutrally) cost-effective measure where individual-level risks seem like they would be rejected by a typical patient. Or such additional information could again provide reason to allow for more choice at the ground level and to issue more open-ended guidance. And this approach would be less complex to implement. It would also have the advantage that more politically and morally relevant decisions would lie with policy-makers rather than analysts, which is preferable on democratic grounds as well.

5 | Conclusion

I have argued that standard methods of cost-effectiveness analysis for health impose risk neutrality on populations that may not be risk neutral. This is at odds with the more general ambition of deference to population values embedded in these methods, and I have argued that there is no good justification for the practice. I have also considered feasible ways for going beyond the risk-neutral default. Three worries might remain.

First, we might question the commitment to the type of anti-paternalism standard methods for QALY-weight elicitation exemplify, which I have granted for the most part in my discussion. Maybe, when evaluating health states and health interventions, we should not be deferential to the way in which individuals evaluate these states and interventions for themselves. Perhaps, instead, we should defer to people's more

social-level preferences: How should we evaluate these health states for the purposes of policy? And should the relevant intervention be implemented or funded for everybody? People's preferences regarding their own health outcomes and regarding these social-level questions might in fact come apart. And when it comes to the social-level question, a larger set of stakeholders becomes relevant. Or, more radically, perhaps we should not aim for deference at all and instead seek to implement what we take to be objectively valuable about health and in the health policy domain.

I am sympathetic to the first of these alternatives in particular, even if it would require a more thorough revision of the methods used in cost-effectiveness analysis. However, my core argument would still apply to these alternative normative frameworks. Essentially, where they differ is on the question of whose preferences matter, and from what perspective—individual or social. But, however, we answer this question, there is a good case for thinking that the relevant preferences may exhibit non-neutrality regarding risk. And in each case, there is no compelling reason for thinking we should defer to preferences regarding health outcomes, while not deferring to risk attitudes: the two should come together.

A second worry could be that people exhibit all sorts of quirks in their preferences, from risk non-neutrality to ambiguity aversion to idiosynchratic context-dependence. We cannot possibly be deferential to all of these. Note, however, that few other 'non-standard' aspects of preference are as prevalent as non-neutral attitudes to risk, both across the population and by coming into play in the vast majority of health-related choice contexts. And they are bound up with, and non-separable from, the pursuit of health in a way in which not all preference 'quirks' might be. And not all preference idiosynchrasies are as plausibly rationally permissible as non-neutral attitudes to risk are. This is not to say that other non-standard aspects of preference may not also meet all of those conditions, and ambiguity aversion is a potential candidate for investigation here.

Finally, how does my argument relate to all of the other problems that have been raised for cost-effectiveness analysis in health (see, e.g. Brock 2004)? Have we not reached a point where we should fundamentally rethink its use? Here I actually want to end on a positive note. Some of the most persistent criticism has in fact led to promising changes in practice, and changes that are not inconsistent with the proposals I discussed here. In particular, we see this in the case of the concern for equity. There is a lot of theoretical attention as well as openness in practice to the idea that cost-effectiveness analysis should be either equity-weighted, or at least supplemented with information on effects on equity.[12] The upshot of my argument is that the same attention and openness should be granted to risk: Analysis should be risk weighted or at least supplemented with information on individual- as well as social-level risk.

Acknowledgements

I thank audiences at the Formal and Experimental Philosophy Workshop at the University of Pittsburgh, the Young Scholars Initiative Workshop ahead of the Economic Philosophy Conference in Reims and the Rutgers Center for Population-Level Bioethics Seminar for very helpful feedback, as well as Paul Schneider for crucial advice in the early stages of this project.

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

Data sharing not applicable–no new data generated.

Endnotes

[1] For a slightly more detailed overview of this, see my Thoma (2023).

[2] For a further, similar study demonstrating the difference incorporation of risk attitudes can make, see Woodward, Schnitzler, and Kvols (1998).

[3] See, e.g., Sunstein and Thaler (2003) or Beshears et al. (2008).

[4] Also see my Thoma (2024) for a critique of debiasing methods using cumulative prospect theory that impose risk neutrality.

[5] Although, even in the case of social-level risk, the question arises why we should adopt a risk-neutral approach. This would become relevant in cases where there is significant social-level risk, e.g. when we are uncertain about the general effectiveness of a treatment.

[6] And indeed, there are also sometimes argued to be ex post egalitarian reasons against individually risky policies, even where these are preferred by everybody (see Adler and Sanchirico 2006). My point here applies even if the individually risky treatment happens to create greater ex post equality.

[7] I take this to be the main takeaway from Hausman's (2015) discussion of health as a public value.

[8] Indeed, this is one aspect of the (multifarious and hard to define) idea of 'extra-welfarism' that many health economists have adopted, see Brouwer et al. (2008)—though other aspects of this idea go further than this, for instance in rejecting an approach to measuring the value of health that is deferential to the population's values.

[9] I discuss this question more thoroughly in Thoma (2023), where I argue that non-consequentialist reasons for deference may nevertheless still extend to risk attitudes.

[10] Deference would also avoid another kind of arbitrariness, which is that, by being deferential to people's health state valuations, we are already in some respects being deferential to attitudes to risk: Attitudes to risk are already baked into our evaluations of health states, because we partly value good health in order to achieve other goods that are uncertain, like having stable earnings. I thank Nir Eyal for this point.

[11] There are also some less formally challenging alternatives that allow to a limited extent to take into account non-neutral attitudes to risk, such as the idea to also represent the 'value of hope' in more traditional QALY-based cost-effectiveness analysis. See Peasgood et al. (2022).

[12] See, e.g. Round and Paulden (2018) and Paulden and McCabe (2021) for discussion of how this is done by NICE and could potentially be done better.

References

Adler, M. D., and C. Sanchirico. 2006. "Inequality and Uncertainty: Theory and Legal Applications." *University of Pennsylvania Law Review* 155: 279–377.

Asch, D. A., and J. C. Hershey. 1995. "Why Some Health Policies Don't Make Sense at the Bedside." *Annals of Internal Medicine* 122, no. 11: 846–850.

Attema, A. E., W. B. Brouwer, O. l'Haridonc, and J. L. Pinto. 2016. "An Elicitation of Utility for Quality of Life Under Prospect Theory." *Journal of Health Economics* 48: 121–134.

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian. 2008. "How Are Preferences Revealed?" *Journal of Public Economics* 92: 1787–1794.

Bilcke, J., and P. Beutels. 2022. "Generating, Presenting, and Interpreting Cost-Effectiveness Results in the Context of Uncertainty: A Tutorial for Deeper Knowledge and Better Practice." *Medical Decision Making* 42, no. 4: 421–435.

Brock, D. 2004. "Ethical Issues in the Use of Cost Effectiveness Analysis for the Prioritization of Health Resources." In *Handbook of Bioethics: Taking Stock of the Field From a Philosophical Perspective*, edited by G. Khushf. Dordrecht, Netherlands: Springer.

Broome, J. 1993. "QALYs." *Journal of Public Economics* 50, no. 2: 149–167.

Brouwer, W. B., A. J. Culyer, N. J. A. van Exel, and F. F. Rutten. 2008. "Welfarism vs. Extra-Welfarism." *Elsevier Journal of Health Economics* 27, no. 2: 325–338.

Buchak, L. 2013. *Risk and Rationality*. Oxford, UK: Oxford University Press.

Cher, D. J., J. Miyamoto, and L. A. Lenert. 1997. "Incorporating Risk Attitude Into Markov-Process Decision Models: Importance for Individual Decision Making." *Medical Decision Making* 17, no. 3: 340–350.

Drummond, M. F., M. J. Sculpher, K. Claxton, G. L. Stoddard, and G. W. Torrance. 2015. *Methods for the Economic Evaluation of Health Care Programmes*. 4th ed. Oxford, UK: Oxford University Press.

Dyer, J. S., and R. K. Sarin. 1982. "Relative risk aversion." *Management Science* 28, no. 8: 875–886.

Goldschmidt, Z., and I. Nissan-Rozen. 2020. "The Intrinsic Value of Risky Prospects." *Synthese* 198, no. 8: 7553–7575.

Harrison, G. W., and J. T. Swarthout. 2023. "Cumulative Prospect Theory in the Laboratory: A Reconsideration." In *Models of Risk Preferences: Descriptive and Normative Challenges*, edited by G. W. Harrison and D. Ross. Bingley, UK: Emerald.

Hausman, D. 2015. *Valuing Health: Well-Being, Freedom, and Suffering*. Oxford, UK: Oxford University Press.

Johannesson, M. 1995. "The Ranking Properties of Healthy-Years Equivalents and Quality-Adjusted Life-Years Under Certainty and Uncertainty." *International Journal of Technology Assessment in Healthcare* 11, no. 1: 40–48.

Lakdawalla, D. N., and C. E. Phelps. 2022. "A Guide to Extending and Implementing Generalized Risk-Adjusted Cost-Effectiveness (GRACE)." *European Journal of Health Economics* 23, no. 3: 433–451.

Makins, N. 2023. "Patients, Doctors and Risk Attitudes." *Journal of Medical Ethics* 49: 737–741.

Mulligan, K., D. Baid, J. N. Doctor, C. E. Phelps, and D. N. Lakdawalla. 2024. "Risk Preferences Over Health: Empirical Estimates and Implications for Medical Decision-Making." *Journal of Health Economics* 94: 102857.

Parkin, D., and N. Devlin. 2006. "Is There a Case for Using Visual Analogue Scale Valuations in Cost-Utility Analysis?" *Health Economics* 15, no. 7: 653–664.

Paulden, M., and C. McCabe. 2021. "Modifying NICE's Approach to Equity Weighting." *PharmacoEconomics* 39, no. 2: 147–160.

Peasgood, T., C. Mukuria, D. Rowen, A. Tsuchiya, and A. Wailoo. 2022. "Should We Consider Including a Value for "Hope" as an Additional Benefit Within Health Technology Assessment?" *Value in Health* 25, no. 9: 1619–1623.

Rosen, A. B., J. S. Tsai, and S. M. Downs. 2003. "Variations in Risk Attitude Across Race, Gender, and Education." *Medical Decision Making* 23, no. 6: 511–517.

Round, J., and M. Paulden. 2018. "Incorporating Equity in Economic Evaluations: A Multi-Attribute Equity State Approach." *European Journal of Health Economics* 19, no. 4: 489–498.

Stefansson, H. O., and R. Bradley. 2019. "What Is Risk Aversion?" *British Journal for the Philosophy of Science* 70, no. 1: 77–102.

Sunstein, C. R., and R. Thaler. 2003. "Libertarian paternalism." *American Economic Review: Papers and Proceedings* 93, no. 2: 175–179.

Thoma, J. 2019. "Risk Aversion and the Long Run." *Ethics* 129, no. 2: 230–253.

Thoma, J. 2023. "Taking Risks on Behalf of Another." *Philosophy Compass* 18, no. 3: e12898.

Thoma, J. 2024. "Merely Means Paternalist? Prospect Theory and 'Debiased' Welfare Analysis." *Philosophy of Science* 91, no. 1: 204–224.

Weirich, P. 2020. *Rational Responses to Risk*. New York, NY: Oxford University Press.

Wilkinson, H. 2022. "Can Risk Aversion Survive the Long Run?" *Philosophical Quarterly* 73, no. 2: 625–647.

Woodward, R. S., M. A. Schnitzler, and L. K. Kvols. 1998. "Reduced Uncertainty as a Diagnostic Benefit: An Initial Assessment of Somatostatic Receptor scintigraphy's Value in Detecting Distant Metastases of Carcinoid Liver Tumours." *Health Economics* 7, no. 2: 149–160.