MAIMON WORKING PAPER No 29 DECEMBER 2025

MORE EPISTEMIC BULLSHIT: NICE INTRODUCES NEW THRESHOLDS AND VALUE SETS

Paul C Langley, Ph.D. Adjunct Professor, College of Pharmacy, University of Minnesota, Minneapolis MN

ABSTRACT

Recent announcements by the National Institute for Health and Care Excellence (NICE) introducing revised cost-effectiveness thresholds and an updated EQ-5D value set have been presented as technical reforms designed to modernize health technology assessment (HTA) and facilitate access to innovation. This Maimon Working Paper argues that these changes do not represent methodological progress but rather the intensification of a longstanding epistemic failure at the core of HTA. Drawing on representational measurement theory and Harry Frankfurt's analysis of epistemic bullshit, the paper demonstrates that NICE's evaluative framework remains indifferent to whether its central numerical constructs correspond to measurable attributes.

Cost-effectiveness thresholds can only be meaningful if they relate cost to a measurable outcome. The quality-adjusted life year (QALY), however, fails the axioms required for measurement: it lacks unidimensionality, invariant units, and a true zero, and it is constructed through arithmetic operations that are logically undefined. Raising or recalibrating thresholds therefore does not increase value or precision; it merely expands tolerance for paying for a non-measure. Similarly, updating the EQ-5D value set does nothing to resolve the conceptual incoherence of preference-based utilities. Population preferences remain ordinal judgements that cannot be transformed into interval or ratio measures, regardless of sample size or modelling sophistication.

The paper situates these developments within a broader question: whether NICE is institutionally capable of accepting the axioms of representational measurement. It argues that NICE's structural dependence on composite indices, simulation models, and policy-driven arithmetic makes such acceptance incompatible with its current role. Faced with pressure to adapt, NICE adjusts parameters rather than interrogating foundations.

The conclusion is stark. Until NICE abandons the QALY framework and rebuilds HTA on measurement-valid principles, its appraisals cannot support credible, evaluable, or replicable value claims. Threshold inflation and value-set revision merely perpetuate numerical storytelling under the guise of reform.

INTRODUCTION

In an earlier Maimon Working Paper, *The Reference Case Is Epistemic Bullshit*, I argued that the central evaluative architecture of health technology assessment (HTA) is not merely flawed, but

epistemically compromised [1]. Drawing on Harry Frankfurt's precise philosophical use of the term, bullshit, the charge was not rhetorical abuse but a classification: the reference case proceeds with indifference to whether its central quantities correspond to anything that can legitimately be measured [2]. Utilities, QALYs, and incremental cost-effectiveness ratios are manipulated as if they were grounded in empirical attributes, while the conditions required for measurement are neither examined nor satisfied. The persistence of this architecture for more than four decades reflects not isolated error but institutionalized epistemic neglect.

The extent to which NICE neglected measurement standards as demonstrated in Stevens' seminal paper on scales that allowed arithmetic published in 1946 and the formalization of the axioms of representational measurement by Krantz et al in 1971[3] [4]. This is made even more egregious by the neglect of Rasch rules for latent traits and Wrights formalization of them in 1977 [5] [6] Instead NICE opted for the path of measurement failure by endorsing the measurement of health states, multiattribute instruments and the reference case.

NICE's recent announcement introducing revised cost-effectiveness thresholds and a new EQ-5D value set must be understood in this context[7]. Presented as a technical update designed to modernize appraisal and support innovation, the announcement does not engage with any of the foundational questions raised in the earlier critique. There is no reconsideration of what a QALY is, whether it possesses the properties of a quantity, or whether arithmetic operations performed upon it are meaningful. Instead, the announcement treats thresholds as adjustable policy parameters and value sets as interchangeable inputs, to be revised as circumstances require. This is not methodological refinement; it is administrative adaptation.

Frankfurt's analysis is particularly apt here. The defining feature of bullshit is not falsity, but indifference to truth conditions. Error at least presupposes a standard against which claims might be judged. In contrast, NICE's revisions proceed without any apparent concern for whether the objects being priced, multiplied, or compared exist as measures at all. Raising the threshold does not address the absence of unidimensionality, the lack of a true zero, or the ordinal nature of preference-based utilities. Updating the value set does not transform population preferences into quantities. These moves simply recalibrate a system whose numerical outputs remain detached from measurement theory.

This Maimon Working Paper argues that NICE's latest reforms represent not a departure from the epistemic failures of HTA, but their intensification. When confronted with mounting pressure, economic, political, and industrial, NICE has chosen to adjust the parameters of its invalid framework rather than question its foundations. Thresholds are inflated, value sets refreshed, and the appearance of precision maintained, while the underlying impossibility of the QALY remains untouched.

The sections that follow examine this development in detail. First, I show why cost-effectiveness thresholds are meaningless in the absence of a measurable outcome. I then demonstrate that revising the EQ-5D value set does nothing to repair the conceptual incoherence of preference-based valuation. These analyses are situated within a broader account of forty years of methodological nonsense in HTA, culminating in a conclusion that argues that until

measurement is taken seriously, reform will remain cosmetic and epistemic bullshit will continue to masquerade as evidence in HTA.

MEANINGLESS THRESHOLDS

Cost-effectiveness thresholds are presented by NICE as if they represent a principled boundary between acceptable and unacceptable use of health system resources. In practice, they perform no such function. A threshold can only be meaningful if it relates cost to a measurable outcome. If the outcome lacks the properties of a quantity, then any ratio formed with cost is undefined. NICE's thresholds therefore do not regulate value; they merely assign prices to an artefact that has never satisfied the conditions for measurement.

For a threshold expressed as cost per unit of outcome to be coherent, the outcome must be measured on ast a ratio scale. It must be unidimensional, possess invariant units, and include a true zero that represents the absence of the attribute being measured. Only under these conditions does multiplication or division preserve meaning. The QALY satisfies none of these requirements. It is constructed by multiplying time by a preference-based utility score that is itself ordinal, bounded, and incapable of supporting arithmetic operations. No transformation of such a score, however sophisticated, can confer the properties that ratio measurement requires.

Raising the cost-effectiveness threshold does not alter this fundamental defect. Increasing the upper bound from one monetary figure to another does not make the QALY more coherent, nor does it increase the amount of "health" produced. It simply widens the range of prices the system is willing to tolerate for an outcome whose units are undefined. In this sense, threshold inflation is not a response to new evidence but a declaration of administrative flexibility. The quantity remains imaginary; only the tolerance for paying for it changes.

The recent NICE announcement makes this indifference explicit. Thresholds are no longer defended as reflections of opportunity cost or marginal productivity, claims that were always contestable, but are instead justified in terms of policy goals, including industrial strategy and innovation incentives. This reframing confirms that thresholds are not grounded in empirical measurement but are instruments of governance. Their function is to manage budgetary and political pressures while preserving the false appearance of quantitative decision-making.

The notion of a "willingness to pay per QALY" further exposes the problem. Willingness to pay presupposes a commodity with stable units that can be exchanged. Yet QALYs do not exist as commensurable units of health. They are composites of heterogeneous dimensions collapsed into a single index through population preference modelling. Treating them as tradable quantities is a category error. A higher willingness to pay per QALY does not correspond to greater health gain; it corresponds only to a higher price attached to a numerical charade.

From the perspective of representational measurement theory, NICE's thresholds are prices without quantities. They cannot be validated, falsified, or replicated because there is nothing against which they could be tested. Adjusting them in response to political or economic pressure does not constitute reform; it confirms that the system is indifferent to the truth conditions of its own numbers. In Frankfurt's terms, the thresholds exemplify epistemic bullshit: they are

produced and adjusted without regard to whether the ratios they express are meaningful. As long as HTA continues to treat non-measures as if they were quantities, no threshold, however calibrated, can escape this indictment.

MEANINGLESS EQ-5D VALUE SETS

NICE's decision to adopt a revised EQ-5D value set is presented as a technical improvement, reflecting more recent population preferences and updated valuation methods. Yet this change does nothing to address the fundamental question that should precede any such revision: whether the EQ-5D value set measures anything at all. Updating the preferences used to generate utility weights cannot repair a framework that lacks the defining properties of measurement.

The EQ-5D value set is not a measure of health, wellbeing, or quality of life. It is a statistical artefact derived from population-level preference elicitation exercises, most commonly time trade-off or related methods. These exercises ask respondents to rank or trade hypothetical health states under constrained assumptions. The resulting outputs are composite ordinal judgments, expressing orderings of preference, not quantities of an attribute. They have no link whatsoever to the standards for fundamental measurement. No amount of modelling sophistication can convert such orderings into interval or ratio measures without satisfying the axioms of representational measurement theory, axioms that the EQ-5D framework neither specifies nor meets.

The core defect lies in the conflation of preference with magnitude. Preferences do not possess units, cannot be meaningfully averaged, and do not admit a true zero. A health state assigned a value of zero in the EQ-5D system does not represent the absence of health; it represents an arbitrarily chosen anchor point, typically "dead," imposed by convention rather than discovered by measurement. The existence of negative values, often interpreted as states worse than death, further underscores the absence of ratio properties. A scale that permits negative values cannot support multiplication, yet such multiplication is central to the construction of QALYs. The entire exercise is just numerical storytelling; a judgement that would have been apparent in 1946.

Revising the value set does not resolve these issues. A newer dataset does not create unidimensionality where none exists. The EQ-5D descriptive system spans multiple domains, none of which are demonstrated to lie on a single latent continuum. Collapsing responses across these domains through preference weights does not produce a measure; it produces an index. Indices may be useful for classification or description, but they cannot support arithmetic claims about magnitude or change.

NICE's implicit justification for updating the value set is that more recent preferences better reflect contemporary social values. This is a political or ethical claim, not a measurement one. Social values may change; required measurement properties do not. If a construct lacks the formal structure required for quantification, updating the population from which preferences are drawn merely refreshes the appearance of legitimacy. It does not transform ordinal judgments into quantities.

By adopting a revised EQ-5D value set while leaving the underlying conceptual framework untouched, NICE reinforces the central epistemic failure of HTA. The revision acknowledges, implicitly, that earlier value sets were contingent and revisable, yet continues to treat the resulting numbers as if they possessed stable quantitative meaning. This is not methodological progress but epistemic recycling. In Frankfurt's terms, the indifference lies not in the choice of one value set over another, but in the refusal to ask whether the numbers generated by any value set can legitimately be used as measures. As long as preference-based utilities are treated as quantities, every update simply perpetuates the same category error under a new statistical veneer.

CAN NICE EVER ACCEPT THE AXIOMS OF REPRESENTATIONAL MEASUREMENT?

The preceding sections have treated NICE's revised thresholds and updated EQ-5D value set as manifestations of a deeper failure. The more fundamental question is whether NICE, as an institution, is capable of accepting the axioms of representational measurement at all. These axioms are not methodological preferences or philosophical niceties; they specify the conditions under which numbers can legitimately represent empirical attributes. Without them, quantitative claims collapse into numerology. Yet after more than forty years of HTA practice, there is no evidence that NICE has ever seriously engaged with these requirements.

Representational measurement theory establishes that measurement precedes arithmetic. Numbers cannot be manipulated simply because they are available; they must first be shown to preserve the empirical structure of the attribute being represented. This entails unidimensionality, invariance, and the correct identification of scale type. Ordinal scales support ordering, interval scales support addition and subtraction, and ratio scales alone support multiplication and division. These distinctions are not controversial. They have been standard in the physical and social sciences since 1946.. Nonetheless, NICE's methods documents, reference case, and appraisal guidance proceed as if such distinctions were optional.

The QALY framework directly violates these axioms. It treats preference-based utility scores as if they were interval measures and then multiplies them by time as if they were ratio measures. This arithmetic is not merely approximate; it is undefined. Accepting representational measurement theory would require NICE to acknowledge that its central outcome metric lacks the properties necessary to support its core calculations. Such an acknowledgment would invalidate not only current appraisals but the entire historical record of NICE decisions.

This institutional dependence helps explain the persistent resistance to measurement reform. NICE is not merely a user of QALYs; it is structurally committed to them. Guidance documents, modelling standards, appraisal committees, and stakeholder expectations are all built around the assumption that health can be reduced to a single quantitative index. Introducing representational measurement theory would force NICE to abandon this simplification and confront the heterogeneity of health outcomes. It would also require distinguishing between attributes that are directly observable and those that are latent, and adopting the Rasch rules of logit measurement for the possession of latent traits. To date, NICE has shown no willingness to make this distinction.

The recent threshold and value-set revisions underscore this point. Faced with criticism that the existing framework restricts access to innovation, NICE did not ask whether its evaluative currency was valid. Instead, it adjusted the exchange rate. This response reveals an institutional preference for parameter adjustment over conceptual examination. It is easier to move thresholds than to question whether the units to which they apply exist as measures.

Accepting the axioms of representational measurement would require NICE to make choices that are institutionally destabilizing. It would have to concede that preference-based utilities are ordinal, that QALYs are composite indices rather than measures, and that many of the arithmetic operations performed in appraisals are illegitimate. It would also have to accept that value claims must be protocol-driven, empirically evaluable, and potentially falsifiable, rather than inferred from simulation models. These concessions would not be incremental; they would be transformative.

The question, then, is not whether NICE understands representational measurement theory. The axioms are readily accessible, well documented, and uncontroversial outside HTA. The question is whether NICE can afford to accept them. The pattern of behavior over four decades, culminating in the latest announcement, suggests that it cannot. Measurement axioms are incompatible with a system that relies on composite indices, threshold tuning, and policy-driven arithmetic. Until NICE confronts this incompatibility, every reform will remain cosmetic, and epistemic indifference will continue to be institutionalized as methodological practice as the reference case takes center stage..

CONCLUSION

The analysis presented in this Maimon Working Paper leads to an unavoidable conclusion: NICE has no legitimate role in health technology assessment as long as it continues to reject the axioms of representational measurement. The recent introduction of revised cost-effectiveness thresholds and updated EQ-5D value sets does not represent progress, modernization, or reform. It represents the further entrenchment of a system that is indifferent to whether its numbers correspond to measurable attributes. In Frankfurt's terms, NICE is not merely mistaken; it is engaged in epistemic bullshit.

Health technology assessment, if it is to deserve the name, must be grounded in measurement. It must distinguish between attributes that can be directly observed and counted and those that are latent and require formal measurement models. It must respect scale-type constraints and limit arithmetic operations to those that are logically permissible. NICE does none of these things. Instead, it continues to treat preference-based indices as quantities, to multiply ordinal scores by time, and to impose thresholds on ratios that have no definable meaning. No amount of parameter adjustment can rescue a framework built on these violations.

By openly treating thresholds as policy levers and value sets as interchangeable inputs, NICE has abandoned even the pretense that its appraisals are constrained by empirical reality. Decisions are framed as quantitative while being driven by administrative convenience and political priorities. This is not assessment; it is numerical storytelling. The persistence of this practice over more than forty years cannot be attributed to ignorance. The axioms of measurement are

well established, widely taught, and routinely applied in other scientific disciplines. NICE's refusal to engage with them is a choice.

The consequences of this choice are profound. As long as NICE anchors access, pricing, and reimbursement decisions to non-measures, its outputs cannot be validated, replicated, or falsified. Claims of value cannot be empirically tested, and disagreement can only be resolved through negotiation or authority, not evidence. In such a system, thresholds become prices for imaginary quantities, and value sets become tools for legitimizing predetermined outcomes.

If NICE wishes to remain relevant, it must abandon the QALY framework, reject preference-based utilities as measures, and rebuild health technology assessment from the ground up on measurement-valid foundations. If it cannot or will not do so, then its role should be limited to administrative budgeting, not scientific evaluation. Health systems deserve assessment grounded in measurement, not institutions that substitute arithmetic ritual for empirical knowledge. Until that transition occurs, NICE will remain not a guardian of evidence, but a custodian of epistemic failure.

ACKNOWLEDGEMENT

I acknowledge that I have used OpenAI technologies, including the large language model, to assist in the development of this work. All final decisions, interpretations, and responsibilities for the content rest solely with me.

REFERENCES

[1] Langley P. The Reference Case is Epistemic Bullshit. Maimon Working Paper No 28 December 2025 https://maimonresearch.com/the-reference-case-is-epistemic-bullshit/

[2] Frankfurt H, On Bullshit. Princeton: Princeton University Press, 2005

[3] Stevens S. On the Theory of Scales of Measurement. *Science*. 1946;103(2684):677-80

[4]Krantz D, Luce R, Suppes P, Tversky A. Foundations pf Measurement Vol 1: Additive and Polynomial Representations. New York: Academic Press, 1971

[5] Rasch G, Probabilistic Models for some Intelligence and Attainment Tests. Chicago: University of Chicago Press, 1980 [An edited version of the original 1960 publication]

[6] Wright B. Solving measurement problems with the Rasch Model. *J Educational Measurement.* 1977;14(2):97-116

[7] National Institute for Health and Care Excellence. *Changes to NICE's cost-effectiveness thresholds confirmed*. NICE News (1 Dec 2025).