MAIMON RESEARCH LLC

RECONSTRUCTION OF HEALTH TECHNOLOGY ASSESSMEMNT



THE REFERENCE CASE: DESIGNED FOR MEASUREMENT FAILURE

Paul C Langley Ph.D Adjunct Professor, College of Pharmacy,

University of Minnesota, Minneapolis, MN

LOGIT WORKING PAPER No 2043 NOVEMBER 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

Background: The reference case has become the internationally accepted framework for economic evaluation in health technology assessment (HTA), particularly through the institutionalization of cost–utility analysis and the quality-adjusted life-year (QALY). Originating in the recommendations of the 1996 US Panel on Cost-Effectiveness in Health and Medicine and subsequently embedded in NICE guidance, the reference case standardizes numerical comparisons without first establishing their foundations in representational measurement. This paper challenges its scientific legitimacy as a framework for evaluating therapeutic benefit.

Methods: The NICE reference case is examined against the requirements of representational measurement, including specified empirical attributes, measurement axioms, scale properties, dimensional homogeneity, commensurability and admissible arithmetic. The assessment considers the construction of utility inputs, their combination with survival duration, the measurement requirements of simulation-model inputs, the aggregation of resource quantities and the interpretation of incremental cost-effectiveness ratios.

Results: Three fundamental failures are identified. First, preference-derived utility scores, typically generated through multiattribute valuation algorithms and decrements from unity, have not been demonstrated to constitute lawful ratio measures of a specified preference attribute. Second, their multiplication by survival duration lacks an empirically established mapping between preference space and time space. Consequently, QALYs cannot be interpreted as lawful measures of therapeutic benefit. Third, simulation models incorporate heterogeneous inputs without requiring independent demonstration that each possesses the measurement properties necessary for its intended arithmetic operations. Monetary cost aggregates further obscure the separate resource attributes that should constitute evaluable empirical claims. The resulting incremental cost-per-QALY ratio combines a composite monetary numerator with a denominator lacking established measurement legitimacy. Lifetime extrapolation, discounting, probabilistic sensitivity analysis and willingness-to-pay thresholds cannot correct these foundational deficiencies. Standardization has institutionalized numerical procedures rather than demonstrated their scientific validity.

Conclusions: The conventional reference case is unacceptable as a scientific foundation for quantitative HTA. Its continued endorsement has perpetuated a curriculum and decision framework in which arithmetic precedes measurement. The alternative is a return to normal science through the identification of clinically relevant manifest and latent attributes, demonstrated measurement properties and protocol-driven therapeutic claims that are credible, evaluable, replicable and falsifiable. Resource consequences should be measured independently in their appropriate units, leaving pricing and allocation decisions to health systems. Absent lawful measurement, neither the QALY nor the conventional cost–utility reference case has scientific legitimacy. The reconstruction of HTA requires measurement before arithmetic, replacing numerical storytelling with empirically testable knowledge.

INTRODUCTION

The outstanding feature of health technology assessment (HTA), as it became formalized into what is now recognized as legacy HTA during the early 1990s, was its extraordinary disregard for the foundations of quantitative measurement. At a time when the principles of representational measurement were already firmly established, HTA developed an analytical framework in which numerical assignment was treated as sufficient justification for arithmetic. Whether its early architects understood Stevens' classification of measurement scales or the more demanding axiomatic foundations developed by Krantz, Luce, Suppes and Tversky is secondary [1] [2] [3] [4] . What matters is that these foundations were not incorporated as prerequisites for the quantitative methods that came to define the discipline**.** Numbers were assigned to health-state preferences, combined with survival duration and subsequently incorporated into economic ratios without first demonstrating that the underlying attributes possessed the measurement properties required for those operations. Measurement failure was pre-ordained. HTA became the only quantitative subject where standards for lawful measurement were entirely absent.

The disregard of fundamental measurement requirements and the unqualified endorsement of arithmetic in legacy HTA have been examined in two preceding papers. The first, *The TTO: A Failure of Representational Measurement*, demonstrated that the time trade-off technique produces a lawful ratio of survival durations but fails to establish a corresponding measure of health-state preference [5] . The second, *The EQ-5D-3L: Numbers Without Measurement*, examined the subsequent transformation of TTO-derived numerical responses into regression-generated utility scores, demonstrating the absence of established measurement properties necessary to interpret these scores as lawful measures of preference or proportional adjustments to survival duration [6] .

The present paper continues the deconstruction of legacy HTA by examining the reference case, where unsupported utility scores and QALYs become foundational inputs to simulation models generating lifetime cost–utility estimates and incremental cost-effectiveness ratios. Rather than resolving the measurement failures inherited from TTO and EQ-5D valuation, the reference case institutionalizes and compounds them, transforming numerical constructions without established measurement properties into apparently authoritative evidence for reimbursement and resource-allocation decisions.

Legacy HTA reversed the fundamental sequence of quantitative science. Instead of beginning with specified therapeutic attributes, empirically justified measurement scales and admissible arithmetic, it adopted cost–utility analysis as its organizing principle [7] . Preference-derived utility scores were treated as proportional adjustments to survival duration, producing QALYs, while incremental costs were divided by incremental QALYs to generate cost-effectiveness ratios. Arithmetic preceded measurement, and numerical calculation became a substitute for scientific justification.

The 1996 US Panel on Cost-Effectiveness in Health and Medicine consolidated this approach through the reference case, establishing standardized conventions for valuation, modelling and economic comparison [8] [9] [10] [11] . NICE subsequently institutionalized the framework, while textbooks, academic curriculum, professional organizations and reporting standards reinforced its

international authority [12] [13] [14] [15] . Yet standardization could not establish the missing measurement foundations. An unsupported quantitative construct acquired the authority of a national decision standard without demonstrating its legitimacy under representational measurement.

This paper examines the measurement failures embedded in the reference case's foundations, inputs, mathematical operations and outputs. Its central argument is that the reference case was designed for measurement failure because it accepted cost–utility arithmetic without first establishing lawful measures of therapeutic benefit. Its continued acceptance perpetuates an educational and methodological inversion in which numerical procedures are taught and applied without questioning their measurement foundations. The appropriate response is not further refinement of the reference case but its abandonment in favor of an HTA founded on lawful measurement, credible empirical claims, independent replication and falsification.

THE COST-EFFECTIVENESS FAILURE

The 1996 Panel explicitly endorsed modelling as a valid and necessary scientific procedure for estimating effectiveness. Its recommendations identified regression analysis, Bayesian methods, life tables and state-transition models as legitimate techniques for extrapolating beyond clinical trial observations. Yet the discussion moved from life expectancy to quality-adjusted life expectancy without addressing the additional measurement requirements introduced by the utility coefficient. The sophistication of the model was treated separately from the scientific status of its central quantitative outcome.

This distinction matters because simulation modelling is not inherently incompatible with normal science. Models of physical processes, industrial production and supply chains commonly employ attributes with established measurement properties and mathematical relationships. Their assumptions and predictions may be uncertain, but their principal quantities retain defined empirical meanings. In legacy HTA, by contrast, health-state descriptions, preference valuations, survival estimates and resource costs are assembled into models without an overarching requirement to demonstrate the admissibility of the arithmetic used to combine them.

The available evidence does not establish that representational measurement was wholly absent from every discussion in the complete 1996 Panel report. It does, however, raise a fundamental question: if the requirements of representational measurement were understood and applied, why was the reference case not constructed around specified attributes, established measurement scales and demonstrably admissible arithmetic?

The fundamental failure of the reference case was its adoption of cost-effectiveness, particularly cost–utility analysis, as the organizing principle of quantitative health technology assessment rather than the lawful measurement of specified therapeutic attributes. From the perspective of representational measurement, cost–utility was already a failed concept. A scientifically meaningful quotient requires that its numerator and denominator possess established measurement properties, that their division satisfy dimensional homogeneity and that the resulting ratio have a defined empirical interpretation. Commensurability is additionally required when a quotient purports to represent a proportion of the same attribute. Although a derived ratio may legitimately combine quantities of different dimensions, its operands must be lawful measures appropriate to

the intended calculation. The reference-case cost numerator aggregates heterogeneous resource quantities through assigned monetary prices, producing a composite accounting total rather than a unidimensional measure of resource use. Its QALY denominator is constructed by multiplying survival duration by preference-derived utility scores that have never been established as unidimensional ratio measures of therapeutic benefit or demonstrated to provide lawful proportional adjustments to time. The resulting incremental cost-per-QALY ratio therefore lacks the measurement foundations required for its claimed interpretation as therapeutic cost-effectiveness.

The 1996 Panel reversed the scientific sequence by beginning with the desired economic quotient and constructing an analytical framework around its production. Health-state descriptions, preference valuations, survival estimates and monetary aggregates were assembled into models without first establishing the empirical axioms, scale properties and admissible operations required for meaningful quantitative claims. Methodological consistency, increasingly sophisticated simulation and standardized reporting could not remedy this foundational deficiency. The failure was not simply that particular inputs lacked validation, but that cost–utility analysis itself had been accepted as a legitimate quantitative concept before its measurement foundations were established**.** The reference case consequently institutionalized an arithmetic convention rather than a scientifically defensible measure of therapeutic cost-effectiveness. Its central output was a calculable ratio of numbers, not a lawful measurement claim. By making cost-effectiveness the starting point rather than a possible consequence of independently established attribute-specific measures, the reference case was designed for measurement failure.

This failure is particularly striking because the foundations of representational measurement were already well established by 1996. Stevens had distinguished nominal, ordinal, interval and ratio scales in 1946. Krantz, Luce, Suppes and Tversky subsequently developed the axiomatic foundations of numerical representation, empirical relational structures, uniqueness, admissible transformations and meaningful quantitative statements in the three volumes of *Foundations of Measurement* (1971–1990). Volume III (1990) made explicit the importance of the relationship between scale properties, admissible transformations and the meaningfulness of numerical operations. The assignment of numbers to observations or preferences could not, by itself, establish measurement. Quantitative claims required specified attributes, empirically justified representations and mathematical operations consistent with their demonstrated properties.

The established textbook by Drummond and colleagues, illustrates how thoroughly measurement failure became embedded in the intellectual and educational foundations of legacy HTA. From its opening chapter, economic evaluation is organized around comparisons of costs and consequences, with cost–utility analysis providing the framework for expressing health outcomes in QALYs. Although the textbook recognizes Stevens' classification of nominal, ordinal, interval and ratio scales, acknowledging these categories is not equivalent to incorporating the axiomatic foundations of representational measurement developed over the preceding seven decades and elaborated in the *Foundations of Measurement* volumes. The decisive substitution is that multidimensional health-state descriptions replace specified, unidimensional therapeutic attributes as the objects of quantitative assessment. Preference valuations attached to these descriptions are treated as measures of health benefit without demonstrating the required measurement properties. The textbook's discussion of health-state utility draws upon Torrance's conditional interval

interpretation, which does not establish the ratio-scale properties or the mapping into time space required for proportional adjustment of survival duration. The historical continuity is significant: Torrance contributed to the development of time trade-off methods, participated in the 1996 Panel's methodological work and was a coauthor of the Drummond textbook. The result is an educational framework in which students are taught to construct utility scores, QALYs and cost-effectiveness ratios without first establishing lawful measures of therapeutic attributes or demonstrating the admissibility of the arithmetic applied to them. Representational measurement remains peripheral to a curriculum organized around numerical calculation rather than scientific measurement.

The consequence has been a globally accepted false analytical framework in which measurement failure was not an incidental possibility but an inevitable consequence of the invalid utility foundations of cost–utility analysis. The QALY was constructed without the measurement properties required for a lawful measure of therapeutic benefit. The incremental cost-per-QALY ratio inherited that failure, while lifetime simulations perpetuated the resulting numerical constructs without establishing their empirical meaning. The reference case did not originate the measurement failure of cost–utility analysis. It institutionalized that failure, making it the organizing principle of legacy HTA.

THE NICE REFERENCE CASE: METHODOLOGICAL REQUIREMENTS

The National Institute for Health and Care Excellence (NICE) reference case specifies the preferred methods for conducting economic evaluations of health technologies in England [16] . Its purpose is to provide a standardized framework for comparing alternative interventions and informing decisions concerning NHS adoption and reimbursement. The preferred form of economic evaluation is cost–utility analysis, with therapeutic benefits expressed in quality-adjusted life-years (QALYs) and resource consequences expressed in monetary terms.

The reference case requires the identification of relevant comparators, normally including established NHS practice. The cost perspective encompasses the National Health Service (NHS) and Personal Social Services (PSS). Resource inputs include pharmaceutical acquisition, administration, monitoring, hospital admissions, outpatient visits, diagnostic procedures, management of adverse events and other relevant healthcare services. Resource quantities are estimated from clinical trials, observational studies, administrative databases and published evidence, then combined with appropriate unit costs to generate monetary estimates.

Clinical inputs include treatment effects, survival probabilities, disease progression, response rates, adverse events, treatment discontinuation and other outcomes relevant to the condition being evaluated. Evidence may be drawn from randomized controlled trials, systematic reviews, network meta-analyses, registries and real-world observational studies. Where direct comparative evidence is unavailable, indirect comparisons and evidence synthesis may be employed.

Health outcomes are expressed as QALYs, combining survival duration with preference-based health-state utility weights. NICE normally prefers EQ-5D for measuring health-related quality of life in adults, with valuations based on general population preferences. Utility weights are assigned to modelled health states and combined with the estimated duration spent in each state. Where EQ-

5D evidence is unavailable or unsuitable, alternative measures or mapping algorithms may be considered with appropriate justification.

Modelling is used to integrate these inputs and project costs and QALYs beyond the period of direct observation. Depending upon the clinical condition, models may employ decision trees, cohort state-transition or Markov models, partitioned survival models, or individual patient simulation. These structures represent alternative treatment pathways, transitions between disease states, survival, treatment discontinuation and subsequent healthcare utilization. Transition probabilities and other parameters are estimated from available evidence and may be extrapolated to represent longer-term outcomes.

The time horizon must be sufficiently long to capture all important differences in costs and health outcomes between interventions. Lifetime horizons are frequently adopted where treatments affect survival or have lasting consequences. Future costs and QALYs are normally discounted at 3.5% annually.

Model outputs include total expected costs and QALYs for each comparator, incremental costs, incremental QALYs and incremental cost-effectiveness ratios (ICERs). Uncertainty is assessed through deterministic sensitivity analyses, scenario analyses and probabilistic sensitivity analyses, with results commonly presented through cost-effectiveness planes and acceptability curves. Model assumptions, parameter sources and structural choices must be documented, with validation undertaken where feasible.

The resulting estimates are considered alongside NICE's cost-effectiveness thresholds, severity modifiers and other applicable decision considerations. The reference case thus combines clinical evidence, health-state valuations, resource-use estimates and mathematical modelling into a standardized framework for assessing the expected comparative costs and QALYs of alternative healthcare interventions.

THE ARITHMETIC TRIUMPH OF THE REFERENCE CASE

The NICE reference case, together with the numerous comparable economic evaluation frameworks endorsed internationally, represents the triumph of numerical storytelling over representational measurement. As demonstrated in the two preceding papers, beginning with the time trade-off (TTO) approach developed by Torrance and colleagues, the dominant approach to the quantitative assessment of therapy impact has been characterized by unconstrained arithmetic. Numerical values have been generated, transformed, combined and projected without first establishing the empirical attributes being measured, the axioms governing their numerical representation or the scale properties authorizing the operations performed. The original TTO duration ratio was reinterpreted as a preference utility; the subsequent EQ-5D valuation algorithms transformed heterogeneous numerical responses into regression-generated decrements from unity; and these resulting utility scores were adopted as proportional multipliers of survival duration. At each stage, arithmetic was permitted to precede measurement, with the quantitative legitimacy of the resulting values assumed rather than demonstrated.

The NICE reference case institutionalizes these assumptions within an elaborate framework of economic evaluation. Clinical outcomes, survival estimates, health-state utility weights, resource utilization, unit costs, transition probabilities and treatment effects are assembled into mathematical models projecting costs and QALYs over extended, frequently lifetime, horizons. The resulting estimates are discounted, compared, subjected to sensitivity analyses and expressed as incremental cost-effectiveness ratios. Yet the sophistication of these procedures cannot compensate for the absence of measurement foundations in their central health-benefit construct. The model does not establish the legitimacy of its inputs through the act of combining them, nor does statistical uncertainty analysis demonstrate that the arithmetic is admissible. The appearance of quantitative precision is therefore no substitute for demonstrating that the quantities entering and emerging from the model possess the measurement properties attributed to them.

The fundamental issue is not whether the reference case can generate numerical predictions. It plainly can. The question is whether those predictions constitute scientifically meaningful and empirically testable measurements of comparative therapeutic benefit. Representational measurement requires that attributes be identified, their measurement structures established and their admissible arithmetic demonstrated before quantitative claims are advanced. The conventional reference case reverses this sequence, accepting the QALY as a measure of benefit and constructing an extensive analytical apparatus around it. Absent a lawful QALY, the conventional cost–utility reference case loses its quantitative foundation. The following assessment therefore examines the prescribed inputs, modelling procedures and outputs of the NICE reference case to determine whether they satisfy the requirements of representational measurement or merely extend the original measurement failure through increasingly elaborate numerical constructions.

THE UTILITY INPUT: TWO FUNDAMENTAL MEASUREMENT FAILURES

The primary reference case inputs for estimating therapeutic benefit are health-state utility scores, typically generated by multiattribute valuation algorithms applied to responses describing health states. The objective is to combine these scores with survival duration, usually across successive disease states or stages of progression, to generate aggregate quality-adjusted life-years (QALYs) for competing therapies. Unfortunately, as demonstrated in the preceding paper, this procedure fails the requirements of representational measurement at two distinct stages. The first concerns the construction of the utility score itself; the second concerns its subsequent application as a proportional adjustment to survival duration. Neither failure is resolved by incorporating the resulting numerical values into a more elaborate economic model.

The first error arises from the application of a multiattribute valuation algorithm to categorical responses describing hypothetical or experienced health states. In the EQ-5D tradition, regression coefficients estimated from TTO-derived valuations are interpreted as additive decrements from unity, producing a numerical utility score for each combination of health-state descriptions. Yet the underlying dependent variable has not been demonstrated to measure a defined, unidimensional preference attribute, and the estimated coefficients have not been established as commensurable quantities representing additive losses of that attribute. Assigning full health the value of one and death the value zero does not demonstrate an absolute zero, equal measurement units or ratio-scale

properties. The resulting utility score is a numerical construction without an established status as a quantitative measure of preference. Its interpretation as a proportion of health or therapeutic benefit is therefore unsupported.

The second error is the incorporation of this score into the QALY calculation. Survival duration is a lawful ratio measure of time, but the utility score has no demonstrated standing as a proportional adjustment factor. Even if a ratio measure of preference could be established, its application to duration would require an independently demonstrated mapping from preference space to time space. No such mapping has been established. Multiplying an unsupported utility score by survival duration cannot create a lawful measure of quality-adjusted survival. The reference case nevertheless repeats this operation across disease states, treatment pathways and projected survival periods, accumulating the resulting products into aggregate QALYs. The accumulation of invalidly interpreted quantities does not repair their measurement deficiencies. Thus, not only is the utility input unsupported as a measure, but the central arithmetic operation used to generate the reference case's principal health-benefit output, the QALY, s equally without demonstrated measurement justification.

THE ADDITIONAL FAILURE: MEASUREMENT REQUIREMENTS FOR MODEL INPUTS

The NICE reference case fails not only because of the impossible utility score but also because of the impossible QALY. These are two distinct measurement failures: the absence of an established ratio measure of preference and the absence of a demonstrated mapping that would permit such a measure, even if it existed, to adjust survival duration proportionally. There is, however, a further and independent concern. Every input entering a simulation model must satisfy the measurement requirements of the arithmetic operations subsequently performed upon it. The construction of a mathematical model does not suspend the axioms of representational measurement. Each input must represent a defined empirical attribute, possess demonstrated scale properties and be subjected only to operations admissible for those properties.

This requirement applies throughout the reference case. Survival durations, event frequencies, transition probabilities, treatment effects, resource quantities, monetary costs and health-state valuations enter models through different mathematical relationships. Their numerical representations cannot be treated as interchangeable merely because they are expressed as numbers. Where multiplication or division is proposed, the operands must possess the properties required for that particular operation, including appropriate dimensional relationships and meaningful proportional interpretation where ratios of attribute magnitudes are claimed. Addition and subtraction similarly require commensurable quantities and justified interpretations of differences. Probabilities, for example, may legitimately enter probability-weighted calculations when their interpretation and the model structure are established; this does not authorize the same operations on arbitrary ordinal scores or unsupported preference valuations. Every model input must therefore be independently qualified for its intended quantitative use before simulation begins.

The implications are fundamental. Even if health-state utility scores could somehow be reconstructed as lawful ratio measures, and even if the necessary mapping into duration space

could be demonstrated, the reference case would not thereby acquire scientific legitimacy. Its remaining inputs, parameter transformations, structural assumptions and projected outputs would still require independent measurement and empirical justification. A lifetime simulation that combines quantities without demonstrating their measurement properties cannot establish the legitimacy of its results through model calibration, sensitivity analysis or apparent predictive precision. The reference case cannot be rescued by repairing one input or one equation. Its entire quantitative structure must be subjected to measurement qualification, operation by operation. This is the distinction between constructing a numerical simulation and establishing a scientifically defensible quantitative evaluation of therapy impact.

THE REFERENCE CASE OUTPUTS: ARITHMETIC WITHOUT MEASUREMENT

The principal outputs of the NICE reference case are estimated costs and quality-adjusted life-years (QALYs), reported separately or combined into average and incremental cost-per-QALY ratios. These results are subsequently compared with cost-effectiveness thresholds to inform reimbursement and formulary decisions. The framework fails to establish lawful measures of either therapeutic benefit or the separate resource consequences of treatment. The QALY is not an established measure of a specified therapeutic attribute, while the cost aggregate combines heterogeneous resource quantities through monetary weights rather than evaluating each resource attribute independently.

The central output of reference-case cost–utility analysis is the incremental cost-effectiveness ratio: $ICER = \Delta C / \Delta Q$ where ΔC represents incremental monetary cost and ΔQ represents incremental QALYs between competing interventions.

The numerator is constructed from heterogeneous healthcare resource inputs, including pharmaceutical quantities, physician visits, hospital days, diagnostic procedures and other services. These are assigned unit prices and aggregated: $C = \Sigma (p_i \times q_i)$ where p_i represents the assigned monetary price per unit of resource i and q_i represents the corresponding resource quantity. Although this operation can produce a monetary accounting total when the underlying quantities and prices are appropriately specified, it does not produce a measure of a single resource-use attribute. The constituent resource quantities belong to different empirical attributes and cannot be represented as a common physical resource measure merely through monetary conversion. Their aggregation also conceals the individual resource consequences of therapy that should be subjected to empirical evaluation.

For a measurement-based assessment of therapy impact, each resource attribute should instead be specified independently. Changes in hospital days, physician consultations, diagnostic procedures, pharmaceutical utilization and other resource quantities should be expressed in their appropriate units and evaluated as separate, testable claims. Each requires a defined attribute, lawful measurement, an appropriate comparator and an empirically testable hypothesis. The task of HTA is to establish what resources a therapy consumes, saves or reallocates, not to replace those separate empirical claims with a composite monetary score. Once the resource consequences have been established, the health system can apply its own unit prices, budget constraints and allocation priorities. Pricing is a subsequent decision-making activity, not a substitute for measurement.

The denominator presents a separate and equally fundamental failure. Incremental QALYs are differences between quantities constructed from preference-derived utility scores and survival duration. As demonstrated in the preceding papers, the utility scores lack demonstrated ratio-scale properties, while their use as proportional adjustments to time requires an empirically established mapping between preference space and time space. Neither condition has been satisfied. Subtracting two unsupported QALY totals cannot establish a lawful measure of incremental therapeutic benefit.

The ICER therefore combines two fundamentally different constructions: a monetary aggregate of heterogeneous resource inputs and an unsupported measure of quality-adjusted survival. The objection is not simply that the numerator and denominator possess different dimensions. Legitimate derived ratios can combine different measured attributes when the required measurement properties and interpretation of the quotient have been established. The objection is that the numerator suppresses the separate resource attributes relevant to therapy evaluation, while the denominator lacks the measurement foundation required for its claimed interpretation as therapeutic benefit. Their division cannot establish a scientifically meaningful measure of incremental therapeutic cost-effectiveness.

The subsequent application of cost-effectiveness thresholds does not resolve these deficiencies. A threshold establishes a decision rule for a numerical quotient; it does not validate the quantities from which that quotient was constructed. Nor can discounting, probabilistic sensitivity analysis, alternative model structures or lifetime extrapolation establish the missing measurement properties. These procedures manipulate numerical inputs and projected outputs without demonstrating that the underlying empirical claims have been measured lawfully.

A reconstructed HTA must therefore abandon the presumption that therapeutic benefits and resource consequences should be compressed into a single cost-per-QALY ratio. Therapy impact should be evaluated through defined manifest and latent attributes, with each claim subjected to appropriate measurement, statistical testing, replication and falsification. Resource consequences should likewise be evaluated as separate manifest attributes in their natural units. The health system may then assign monetary values and make allocation decisions using its own prices and priorities.

The essential distinction is between the scientific measurement of therapy impact and the subsequent economic decisions made using that evidence. HTA should establish the empirical consequences of therapy. The health system should determine what those consequences are worth.

THE RETURN TO NORMAL SCIENCE

The conclusion is clear-cut: the conventional reference case has no place in the scientific assessment of therapy impact. Once the principle that measurement must precede arithmetic is accepted, the accumulated baggage of legacy HTA, with its emphasis on valuing multidimensional health-state descriptions, collapses. The failures are not confined to the impossible utility score or the impossible QALY. They extend throughout the reference case, where numerical inputs, transformations and outputs are accepted without demonstrating the measurement properties required for their intended arithmetic. These failures have persisted because methodological

attention has focused on increasingly elaborate numerical procedures rather than the requirements of representational measurement. Unconstrained arithmetic has displaced measurement as the foundation of quantitative evaluation.

The alternative is straightforward: a return to the standards of normal science. Representational measurement demands defined empirical attributes, demonstrated scale properties, admissible arithmetic and testable hypotheses. For each therapy assessment, investigators should identify a limited number of clinically relevant manifest and latent attributes and develop protocol-driven claims concerning expected treatment effects. These claims must be credible, evaluable, replicable and open to falsification. Resource consequences should be measured separately in their appropriate units, leaving subsequent pricing and allocation decisions to the health system. There is no need for hypothetical utility scores, QALYs or lifetime cost–utility simulations. The reconstruction of HTA begins by abandoning numerical storytelling and restoring lawful measurement to its rightful place in quantitative science.

The ultimate objective must be the progressive growth of objective knowledge concerning therapy impact**.** Following Karl Popper, scientific progress depends upon conjectures formulated as empirically testable hypotheses, exposed to potential falsification and subjected to independent replication and critical scrutiny [17 18]. Knowledge advances through the identification and correction of error, not through the repeated production of model-generated assertions that cannot be independently tested. A reconstructed HTA must therefore establish a cumulative research program in which measurement-qualified therapeutic claims are evaluated, challenged, replicated and progressively refined or rejected. The purpose of HTA is not to manufacture numerical certainty about hypothetical futures, but to contribute to the evolution of objective knowledge concerning the actual impact of therapies**.**

CONCLUSION

The NICE reference case is unacceptable as a scientific foundation for health technology assessment. Its central construct, the QALY, combines preference-derived utility scores with survival duration without demonstrating the measurement properties required for that operation. The resulting incremental cost-per-QALY ratio cannot be interpreted as a lawful measure of therapeutic cost-effectiveness. Lifetime simulations, sensitivity analyses and decision thresholds cannot repair these foundational failures. The reference case substitutes numerical storytelling for the measurement, testing and falsification required by normal science.

The consequences extend beyond methodological disagreement. For decades, successive generations of students have been taught to manipulate numerical constructs without first establishing the attributes measured, their scale properties or the admissibility of the arithmetic performed. Academic institutions, professional societies and assessment agencies have perpetuated this inversion, institutionalizing numerical storytelling as quantitative evidence.

The alternative is already available. HTA must be reconstructed around specified manifest and latent attributes, demonstrated measurement properties and protocol-driven claims that are credible, evaluable, replicable and falsifiable. Resource consequences should be measured independently, leaving pricing and allocation decisions to health systems.

The conventional reference case has no legitimate place in a curriculum founded on representational measurement. Its appropriate role is as a warning of what happens when arithmetic replaces science. Forty years of institutional acceptance cannot transform numerical constructs into lawful measures. Measurement must precede arithmetic.

MEASUREMENT MUST PRECEDE ARITHMETIC.

There is no scientifically defensible alternative to representational measurement as the foundation for quantitative health technology assessment. Reconstructing HTA requires abandoning the legacy dependence upon multidimensional health-state preferences and returning to independently defined attributes, lawful measurement scales, admissible arithmetic and empirically evaluable claims. This cannot be achieved through marginal revisions to utility algorithms or reference-case guidance. It requires a fundamental reconstruction of the knowledge base taught to faculty, researchers and students. Representational measurement is not one methodological option among many; it establishes the scientific requirements against which quantitative claims must be judged.

Maimon Research has developed a comprehensive nine-unit distance-learning curriculum comprising approximately 60,000 words of instructional material to support this reconstruction [19]. The program begins with the foundations of representational measurement and the identification of empirical attributes, then develops the distinction between manifest and latent measurement, the properties of measurement scales, dimensional homogeneity, commensurability and admissible arithmetic. It addresses the construction and validation of linear ratio measures for manifest attributes and Rasch-based measurement for latent attributes, including requirements for unidimensionality, invariance and measurement fit. Subsequent units examine the formulation of therapy-impact claims, the measurement of possession and change, treatment compliance and persistence, resource utilization, prospective evaluation protocols, replication and falsification. The objective is not to replace one numerical convention with another, but to establish a coherent scientific framework in which every quantitative claim is supported by an explicitly defined attribute, a validated measure and an admissible mathematical interpretation. The program provides a practical foundation for reconstructing graduate HTA education, professional training and research assessment.

The curriculum is available now through Maimon Research, with structured instructional materials and continuing academic support. Participants can work through the nine units through distance learning, with opportunities for questions, clarification and discussion of measurement problems arising in their own research or teaching. Maimon Research provides continuing question-and-answer support, together with optional seminars and guided discussions to assist faculty, researchers and institutions in applying the framework. The program can support individual professional development, faculty preparation, graduate teaching and institutional curriculum reconstruction. Its significance is that the transition to lawful measurement does not depend upon waiting for established HTA organizations to revise their guidance or for another generation of methodological debate to conclude. The educational foundation and practical support already exist. The responsibility now rests with universities, professional organizations and HTA institutions to replace the inherited curriculum of numerical preference modelling with a curriculum grounded in the measurement of empirical attributes. The reconstruction can begin immediately.

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.

ABOUT THE AUTHOR

Dr Paul Langley is a health economist. Originally from the UK, Dr Langley completed his graduate studies in Canada and has taught in the UK, Canada, Australia and the United States. After more than 20 years teaching at university in Australia, Dr Langley moved to take up academic positions in the United States. He now lives in Tucson, Arizona, to enjoy the Catalina Mountains, the Sonoran Desert, the absence of mosquitoes and the swimming pool.

REFERENCES

[1] Stevens S. On the Theory of Scales of Measurement. *Science*. 1946;103(2684):677-8

[2] Krantz DH, Luce RD, Suppes P, Tversky A. Foundations of Measurement. Volume I: Additive and Polynomial Representations. New York: Academic Press; 1971

[3] Suppes P, Krantz DH, Luce RD, Tversky A. Foundations of Measurement. Volume II: Geometrical, Threshold, and Probabilistic Representations. San Diego: Academic Press; 1989

[4] Luce RD, Krantz DH, Suppes P, Tversky A. Foundations of Measurement. Volume III: Representation, Axiomatization, and Invariance. San Diego: Academic Press; 1990

[5] Langley P. The TTO: A Failure of Representational Measurement. Logit Working Paper No 2041. https://maimonresearch.com/logit-working-paper-no-2041-november-2026/

[6] Langley P. The EG-5D-3L: Numbers without measurement. Logit Working Paper No 2042. https://maimonresearch.com/logit-working-paper-no-2042-november-2026/

[7] Drummond M, Sculpher M, Claxton K et al. Methods for the Economic Evaluation of Health Care Programmes (4th Ed). New York: Oxford University Press, 2015

[8] Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996

[9] Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. *JAMA*. 1996;276(14):1172–1177

[10] Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-Effectiveness in Health and Medicine. *JAMA*. 1996;276(15):1253–1258

[11] Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. JAMA. 1996;276(16):1339–1341

[12] Rawlins MD, Culyer AJ. National Institute for Clinical Excellence and its value judgments. *BMJ*. 2004;329(7459):224–227

[13] Pearson SD, Rawlins MD. Quality, innovation, and value for money: NICE and the British National Health Service. *JAMA*. 2005;294(20):2618–2622

[14] Rawlins MD. NICE and the public health. *British Journal of Clinical Pharmacology*. 2004;58(6):575–580

[15] Rawlins M, Barnett D, Stevens A. Pharmacoeconomics: NICE's approach to decision-making. British *Journal of Clinical Pharmacology*. 2010;70(3):346–349.

[16] National Institute for Health and Care Excellence (NICE). NICE technology appraisal and highly specialised technologies guidance: the manual. NICE process and methods [PMG36]. London: NICE; 2022 (subsequently updated).

[17] Popper KR. The Logic of Scientific Discovery. London: Hutchinson; 1959.

[18] Popper KR. Objective Knowledge: An Evolutionary Approach. Oxford: Clarendon Press; 1972.

[19] Maimon Research. *Reconstructing HTA Programs: Measurement before Arithmetic*. https://maimonresearch.com/hta-reconstruction-program-and-fees/