

MAIMON RESEARCH LLC

**ARTIFICIAL INTELLIGENCE LARGE LANGUAGE
MODEL INTERROGATION**



LEGACY HTA DECONSTRUCTION

PAPER 1

**PROCESS VALIDATION: THE IMPOSSIBLE TIME
TRADE-OFF**

**Paul C Langley PhD Adjunct Professor, College of Pharmacy, University of
Minnesota, Minneapolis, MN**

LOGIT WORKING PAPER No 9080 OCTOBER 2026

www.maimonresearch.com

ABSTRACT

BACKGROUND: *This paper initiates a four-paper deconstruction of legacy health technology assessment (HTA) using representational measurement and process validation. It examines whether time trade-off (TTO) scores can legitimately be interpreted as ratio-scale measures of health-state preference. The central proposition is that conventional HTA has allowed arithmetic and statistical estimation to precede measurement.*

METHODS: *Process validation applies sequential requirements for attribute identification, unidimensionality, empirical relational structure, scale properties, invariance, admissible transformations, dimensional homogeneity and commensurability. Each indispensable requirement constitutes a potential mandatory STOP decision. The analysis distinguishes multidimensional EQ-5D-3L health-state descriptions from quantitative measurement, evaluates the duration ratios generated by TTO indifference judgments, and examines the proposed mapping between preference and duration measurement spaces. It also considers implications for regression-derived utilities, QALYs and the reference case.*

RESULTS: *EQ-5D-3L classifies health states across five heterogeneous dimensions and three ordered levels, generating 243 profiles. These classifications do not independently establish a unidimensional preference attribute, invariant unit or non-arbitrary zero. TTO records indifference between specified health-duration alternatives. For example, six years in full health compared with ten years in another health state yields a lawful duration ratio of 0.6. This establishes a proportion within time space, not that preference for the health state is 60% of preference for full health. The proposed equality between a preference ratio and a duration ratio requires independently established ratio-scale preference magnitudes and an empirically validated mapping between measurement spaces. Neither requirement is demonstrated by the TTO procedure. Regression cannot confer missing measurement properties upon TTO scores, and subsequent utility estimation, QALY multiplication and economic modelling cannot repair the failure. Process validation therefore imposes a STOP decision before TTO scores are admitted as quantitative preference measures.*

CONCLUSIONS: *TTO may document hypothetical choices, but its duration ratios have not been established as proportional measures of health-state preference. Refinements in elicitation, statistical fit or model validation cannot substitute for the missing empirical warrants. The paper concludes that TTO should be abandoned as a source of quantitative preference measures for utility and QALY construction. Therapy-impact assessment should instead begin with explicitly defined attributes, linear ratio measurement for appropriate manifest attributes and Rasch measurement for latent attributes, supporting replicable and falsifiable quantitative claims.*

INTRODUCTION: THE DECONSTRUCTION OF LEGACY HTA

This paper is the first in a four-paper series devoted to the systematic deconstruction of legacy health technology assessment (HTA) through the application of the standards of representational measurement with the theory of measurement scales ^{1 2} . The series examines the successive

quantitative procedures through which health-state descriptions are transformed into preference valuations, time trade-off (TTO) scores, utility algorithms and, ultimately, quality-adjusted life years (QALYs). Its central proposition is that these procedures were developed and institutionalized without demonstrating the measurement properties required to justify their quantitative interpretations and subsequent arithmetic. The resulting framework has allowed arithmetic to precede measurement.

The origins of this failure can be traced to the development and institutionalization of the cost–utility reference case during the 1990s³. The foundational methodological literature concentrated on preference elicitation, utility estimation, statistical modelling, QALY construction and the standardization of economic evaluation. What is conspicuously absent from the reference-case framework is a systematic recognition of the established theory of measurement scales and the axioms of representational measurement. There was no demonstrated requirement to identify the attribute being measured, establish its scale properties or determine whether the proposed arithmetic was admissible. The reference case was constructed primarily as a framework for statistical estimation and economic calculation rather than as a program of empirical measurement. The omission was fundamental: numerical procedures were accepted without first demonstrating that their outputs possessed the measurement properties attributed to them.

This series introduces process validation as a systematic framework for evaluating whether each quantitative procedure satisfies the scientific requirements of representational measurement. Process validation translates these requirements into an explicit assessment of attribute identification, unidimensionality, scale properties, admissible transformations, invariance, dimensional homogeneity, commensurability and the justification of arithmetic operations. The distinction between manifest and latent attributes determines the appropriate measurement pathway. A manifest attribute requires an empirically established measurement scale with properties appropriate to the quantitative claim. A latent attribute requires a measurement model capable of demonstrating the relevant structural properties, including unidimensionality and invariance. Numerical assignments, statistical estimation and computational procedures cannot substitute for establishing these requirements.

Process validation examines the measurement warrants underlying each successive stage of the legacy HTA pathway. It distinguishes between the mathematical feasibility of a calculation, the statistical properties of an estimated model and the scientific validity of the quantitative interpretation assigned to its outputs. A procedure may be computationally correct, statistically well fitted and widely accepted while failing to demonstrate that its numerical results measure the attribute they purport to represent. The critical question is whether the required measurement properties have been established independently of the arithmetic and statistical procedures that depend upon them.

The four papers apply this framework sequentially to health-state valuation and TTO preference scores, regression-derived utility algorithms, QALY construction and the legacy HTA reference case. Each identifies the quantitative claim, specifies the applicable standards of representational measurement and examines whether the empirical evidence supports the proposed measurement interpretation. Particular attention is given to the transformation of legitimate duration ratios into

purported preference magnitudes, the statistical construction of multiattribute utility algorithms and the subsequent multiplication of utility weights by survival duration.

The consequence is fundamental: a failure to establish the measurement properties required at an indispensable stage of the quantitative pathway undermines the validity of the claims constructed from its outputs. Subsequent statistical refinement, model validation, sensitivity analysis or institutional endorsement cannot retrospectively establish missing measurement properties. The purpose of process validation is therefore not to improve the statistical performance of legacy HTA but to determine whether its foundational quantitative procedures satisfy the requirements of empirical measurement. The deconstruction begins with TTO, where a legitimate ratio of durations is assigned a quantitative preference interpretation without an independently demonstrated measurement warrant ⁴.

PROCESS VALIDATION: REPRESENTATIONAL MEASUREMENT BEFORE ARITHMETIC

For eight decades, measurement theory has established that the legitimacy of quantitative operations depends upon the properties of the measurement scale. Stevens distinguished nominal, ordinal, interval and ratio scales, while representational measurement subsequently formalized the empirical conditions required for meaningful numerical representations and admissible arithmetic. These principles preceded contemporary health technology assessment (HTA), yet the legacy reference case developed without systematically incorporating them. Preference elicitation, utility estimation and QALY calculation proceeded without first establishing the measurement properties required by their quantitative claims. Arithmetic was allowed to precede measurement.

The legacy HTA framework follows a sequence in which multidimensional health-state descriptions receive TTO preference scores, regression algorithms transform these into utilities, utilities are multiplied by duration to construct QALYs, and simulation models generate incremental cost-per-QALY estimates. At each stage, numerical outputs become inputs to further calculations without independent demonstration of the measurement properties required for their interpretation.

The initial failure occurs with TTO. Indifference between hypothetical health-duration alternatives produces a legitimate dimensionless ratio of durations within time space. However, interpreting that ratio as a proportional preference magnitude requires an independently established preference attribute, demonstrated ratio-scale properties and an empirically validated mapping between preference and duration spaces. A lawful duration ratio does not establish a lawful preference ratio.

Regression-derived utility algorithms compound this failure. Statistical estimation can describe relationships, estimate coefficients and predict numerical scores, but it cannot establish that the dependent variable possesses the empirical properties of ratio measurement. The resulting utilities inherit the unresolved measurement status of the original TTO scores. Multiplication by duration cannot repair this deficiency or establish the QALY as a measured quantity.

The consequences extend to the reference case itself. Computational accuracy, sophisticated simulation, sensitivity analysis and institutional endorsement cannot confer measurement

legitimacy upon numerical constructs whose required empirical properties remain unestablished. Reproducible cost-per-QALY calculations do not demonstrate that their denominators represent measured health benefits.

Process validation reverses this sequence. It requires each proposed attribute, measurement scale and arithmetic operation to satisfy the relevant empirical requirements before its numerical output can enter the next stage as a measure. Where a required measurement warrant fails, a mandatory STOP decision terminates that quantitative pathway. The deconstruction of legacy HTA therefore begins at its first unsupported transformation: the interpretation of a ratio of durations as a proportional measure of health-state preference.

PROCESS VALIDATION: APPLICATION

Developed systematically by Krantz and colleagues in *Foundations of Measurement* (1971 onward), the axiomatic framework of representational measurement establishes the conditions under which empirical relationships can be represented by numerical structures. Measurement requires more than assigning numbers to observations. It requires demonstrating that the relationships among numerical values preserve the relevant relationships among empirical objects or attributes. Depending upon the empirical measurement structure, the applicable axioms include order, additivity, cancellation, solvability and the Archimedean property of the relationship between magnitudes. These provide the scientific foundation for determining whether numerical representations support the quantitative interpretations and arithmetic operations attributed to them.

Process validation translates these requirements into a systematic assessment of the measurement warrants underlying a proposed or established quantitative procedure. Its purpose is to determine whether the procedure has demonstrated the empirical properties necessary to support its intended claims. The initial requirement is identification of the attribute to be measured, followed by a distinction between manifest and latent attributes. Where proportional comparisons are intended for manifest attributes, an empirically established linear ratio scale is required. For latent attributes, the framework adopted here employs Rasch measurement to evaluate whether a defensible unidimensional measurement structure can be established^{5 6}. Rasch locations are expressed in logits, with differences interpreted on an interval scale and exponentiated differences expressed as odds ratios. Neither numerical assignment nor statistical model fit alone establishes the required measurement properties.

Process validation examines the empirical and mathematical requirements applicable to each measurement claim. These include unidimensionality, the relevant representation and scale properties, invariance, admissible transformations, dimensional homogeneity and commensurability. The requirements are not interchangeable. Unidimensionality concerns the attribute being measured; invariance concerns the stability of measurement relationships; admissible transformations determine which numerical interpretations are preserved; and dimensional homogeneity and commensurability govern the legitimacy of proposed arithmetic operations. Each requirement must be evaluated in relation to the particular quantitative claim rather than assumed from the existence of numerical outputs.

The framework applies both to the development of new measurement procedures and to the examination of established ones. For a proposed procedure, process validation identifies the attribute, the empirical measurement protocol, the scale properties required by the intended claim and the evidence needed for replication and potential falsification. For an established procedure, it examines whether those properties have actually been demonstrated and whether subsequent quantitative operations are justified by them. The scientific standard is identical in both circumstances: the measurement warrant must precede the quantitative interpretation and arithmetic that depend upon it. Statistical estimation, computational accuracy and institutional acceptance cannot substitute for that warrant.

The defining feature of process validation is the sequential application of measurement requirements as mandatory stopping rules. Each stage must establish the empirical properties necessary to support its proposed quantitative interpretation before its output can enter the next stage as a measured quantity. Where a required attribute, empirical relationship, scale property or admissible operation cannot be demonstrated, process validation imposes a stop and review decision. The proposed measurement pathway terminates at that point. Subsequent arithmetic, statistical transformation or simulation cannot retrospectively establish the missing measurement properties. A stop decision does not prohibit further investigation of the underlying observations or numerical scores; it rejects their admission as measures supporting the particular quantitative claim under examination.

Applied to legacy HTA, process validation provides a sequential examination of the transformation of health-state descriptions into TTO preference scores, regression-derived utility algorithms, QALYs and cost-per-QALY claims. Each transformation introduces a distinct quantitative proposition requiring its own empirical justification. The mathematical legitimacy of a duration ratio does not establish its interpretation as a proportional preference magnitude. Statistical estimation of preference scores cannot independently confer ratio-scale properties upon them. The multiplication of utility by duration requires both a lawful proportional factor and an empirically justified relationship between preference proportions and proportional duration. Subsequent QALY construction cannot resolve measurement properties left unestablished at earlier stages, and incorporation into a reference case simulation model cannot restore them.

Process validation therefore replaces the presumption that numerical outputs constitute measurements with a systematic requirement to demonstrate measurement before arithmetic. Its contribution is to identify precisely where quantitative interpretations lack the empirical foundations required by representational measurement, impose the corresponding stop decisions and establish their consequences for successive calculations. The objective is not merely to criticize legacy HTA procedures, but to distinguish numerical constructions from independently established quantitative measures and ensure that therapy-impact claims remain empirically evaluable, replicable and capable of falsification.

PROCESS VALIDATION: COMMENSURABILITY AND DIMENSIONAL HOMOGENEITY

Dimensional homogeneity requires quantitative operations to preserve the dimensional meaning of the quantities involved. Addition and subtraction require compatible dimensions, while

multiplication and division must produce results whose dimensions follow from those of their operands. Commensurability requires quantities entering a ratio intended to express a proportion to be measures of the same attribute, expressed on a common ratio scale.

For duration, two ratio measures T_1 and T_2 belong to time space $[T]$. Their quotient, T_1/T_2 , is a lawful dimensionless proportion because the dimensions cancel. A corresponding preference ratio, P_1/P_2 , would require independently established ratio measures within preference space $[P]$. Even if both ratios were lawful, their dimensionless status would not establish empirical equivalence between them.

This distinction is decisive for time trade-off (TTO). The ratio of two durations cannot automatically be interpreted as a proportional preference magnitude. Such an interpretation requires a demonstrated mapping from preference space to time space, establishing that P_1/P_2 corresponds to T_1/T_2 .

Process validation therefore imposes two independent STOP decisions: failure to establish ratio-scale preference measurement and failure to demonstrate the required mapping between attribute spaces. Numerical equality cannot substitute for empirical equivalence. Without both warrants, TTO scores cannot legitimately enter utility estimation as proportional preference measures.

PROCESS VALIDATION: UNDERSTANDING PROPORTIONAL ADJUSTMENT

Proportional adjustment plays a central role in the arithmetic of legacy HTA, particularly in the construction of time trade-off (TTO) preference scores and quality-adjusted life-years (QALYs). In the TTO, respondents' express indifference between alternative combinations of health states and survival durations. A ratio of two durations is then interpreted as a proportional measure of health-state preference. In the QALY, the direction is reversed: a utility derived from purported preference measurements is applied as a proportional adjustment to duration to produce quality-adjusted time. Both procedures depend upon transferring proportional relationships between distinct attribute spaces. Process validation must therefore establish whether the proposed proportional measure possesses the required ratio-scale properties and whether its transfer between attribute spaces is empirically justified. These are separate measurement requirements, neither of which can be established by the numerical calculation itself.

A fundamental principle of representational measurement is that a proportional adjustment factor must have an empirically established meaning in relation to the attribute being adjusted. For duration, a lawful proportional factor can be constructed directly from two commensurable ratio measures within time space $[T]$, as in T_1/T_2 . Alternatively, a proportion originating in another attribute space would require an independently validated empirical mapping establishing its equivalence to the intended duration proportion. A dimensionless ratio is not automatically transferable between attribute spaces. Even if P_1/P_2 were a lawful preference ratio within preference space $[P]$, its equivalence to T_1/T_2 would still require demonstration. The conventional QALY fails both requirements: conventional utility has not been established as a lawful ratio-scale preference measure, and no validated mapping establishes that its purported preference proportion

corresponds to a proportional adjustment within time space. These constitute independent stop decisions under process validation.

The restriction follows from the meaning of proportional adjustment. A proportion expresses an empirical relationship between measured magnitudes, not merely an arithmetic relationship between numbers. A ratio has quantitative meaning when its numerator and denominator are commensurable measures of the same attribute, possessing an invariant unit and a non-arbitrary zero. Dividing one duration by another therefore establishes a meaningful proportion of time. Dividing two independently established ratio-scale preference magnitudes would establish a proportion of preference. Although both ratios are dimensionless, they represent relationships between different empirical attributes. Numerical equality does not establish empirical equivalence. Applying a preference proportion to duration therefore requires independent evidence that the preference relationship corresponds to the proposed proportional relationship within time space. Without that evidence, multiplication remains computationally possible but lacks the empirical justification required for measurement. Arithmetic must preserve established empirical relationships; it cannot create relationships between attributes that have never been demonstrated.

PROCESS VALIDATION: ATTRIBUTES AND HEALTH STATES

The initial departure from representational measurement in legacy HTA occurs when the valuation of multidimensional health-state descriptions is treated as sufficient to establish a quantitative measure without first identifying and validating the attribute being measured. Representational measurement begins with an explicitly defined attribute whose empirical relationships can be represented numerically. The reference case instead begins with descriptions of health and assigns them single numerical values. The fundamental question of what is being measured is displaced by the procedural question of how a health state should be valued. Valuation is not, by itself, measurement.

Why begin with health states? The motivation is to establish a common quantitative expression for heterogeneous health outcomes across diseases, treatments and patient populations. Compressing multiple dimensions of health into a single score promises a basis for comparing disparate interventions. However, the demand for a common measure does not establish the existence of a common measurable attribute. Nor does the convenience of a single score demonstrate the scale properties required for quantitative comparisons.

The EQ-5D-3L illustrates the problem ^{7 8}. It describes health across five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each has three ordered response categories, generating 243 possible profiles. A profile such as 12321 identifies a combination of reported problems, not a quantitative measurement of a single attribute. The numerical codes classify responses; they do not establish equal intervals, a common unit, a non-arbitrary zero or meaningful ratios. A multidimensional health-state description does not establish a position on a single quantitative measurement continuum.

Representational measurement distinguishes between manifest and latent attributes. Manifest attributes intended to support proportional comparisons require an appropriate linear ratio scale. Latent attributes require a validated unidimensional measurement continuum, with Rasch

measurement providing the approach adopted here. The EQ-5D dimensions concern different aspects of health and cannot be presumed to share a common quantitative unit. If the proposed attribute is preference for the overall health state, it must be independently defined and investigated. Multidimensional health descriptions do not preclude a unidimensional preference attribute, but neither do they establish one.

This exposes the first critical measurement failure in the reference-case construction. A descriptive classification becomes the basis for assigning a numerical preference value without demonstrating the measurement properties of the proposed attribute. Elicited choices, rankings or stated indifference do not independently establish ratio-scale magnitudes. Attribute identification, unidimensionality, invariance and scale properties remain essential. Process validation imposes a stop decision when a health-state valuation is admitted as a quantitative measure without the required empirical warrant.

The consequences extend throughout the reference case. Health-state descriptions generate preference scores; regression models transform these into utility algorithms; utility weights are multiplied by duration to construct QALYs; and simulation models produce incremental cost-per-QALY estimates. Each stage inherits the unresolved measurement status of the preceding stage. Statistical estimation and computational sophistication cannot establish properties that were never demonstrated.

Health-state descriptions and preference elicitation may have legitimate descriptive, behavioral and decision-making uses. The failure arises when their numerical outputs are interpreted as lawful ratio measures without demonstrating the required empirical properties. A health-state description identifies a combination of reported conditions; a measurement scale represents an empirically established attribute. Assigning a preference score to the former does not establish the latter.

The deconstruction of legacy HTA therefore begins with attribute identification. Without independently demonstrated measurement properties, the transformation from health-state description to quantitative utility lacks the warrant required for subsequent QALY arithmetic. Measurement must begin with an attribute, not with the assumption that a numerical valuation of a health state constitutes a measure.

PROCESS VALIDATION: CAN TIME TRADE-OFF MEASURE PREFERENCE?

Time trade-off (TTO) as a preference-elicitation procedure is widely used in health technology assessment (HTA) to assign numerical scores to health-state descriptions. These scores subsequently become observations in regression models used to estimate multiattribute utility algorithms. The fundamental measurement question arises before statistical estimation: can a ratio of two durations legitimately be interpreted as a quantitative measure of preference for a health state? This is the first transformation examined through process validation.

The conventional TTO presents respondents with two hypothetical alternatives: living for a specified duration in a particular health state or living for a shorter duration in full health. The duration in full health is adjusted until the respondent reports indifference. For example, a

respondent may consider ten years in a specified health state equivalent to six years in full health. The conventional TTO score where T is duration is:

$$\text{TTO score} = T_1/T_2 = 6/10 = 0.6$$

Both durations belong to the same dimensional class [T]. They are commensurable ratio measures, and their quotient is a legitimate dimensionless proportion. The value 0.6 expresses the ratio of six years to ten years. It does not, by itself, establish that preference for the specified health state is 60% of preference for full health.

The measurement failure arises when this duration ratio is reinterpreted as a proportional preference magnitude. The respondent's indifference judgment identifies the durations entering the calculation but does not establish a unidimensional preference attribute, an invariant unit, a non-arbitrary zero or meaningful proportional relationships within preference space. Conventional anchors of zero for being dead and one for full health do not demonstrate these properties. The mathematical legitimacy of the duration ratio cannot be transferred to preference merely by assigning it a different interpretation.

Could preference nevertheless be measured independently? In principle, this is an empirical question. A single preference attribute would first have to be defined and an appropriate ratio-scale measurement structure demonstrated. Suppose independently established preference magnitudes were available for health state *i* and full health, represented by P_i and P^f . Their ratio, P_i/P^f , would be a legitimate dimensionless proportion within preference space [P], just as T_1/T_2 is a legitimate proportion within time space [T].

However, two lawful dimensionless ratios do not necessarily represent the same empirical relationship. The proposition

$$P_i/P^f = T_1/T_2$$

would require independent empirical demonstration. Dimensional cancellation within each attribute space does not establish equivalence between them. Even a lawful ratio-scale preference measure would therefore be insufficient to validate the conventional interpretation of TTO without an independently demonstrated mapping between preference and duration.

The decisive test would be whether independently measured preference ratios correspond to TTO duration ratios across health states, respondents and elicitation conditions. The proposed relationship would require prior specification, empirical evaluation and independent replication. Systematic disagreement would challenge the conventional TTO interpretation, even if both ratios were legitimate measures within their respective spaces. Any reproducible relationship would require demonstration of its mathematical form and conditions of validity.

This distinction is fundamental to the legacy HTA reference case. TTO scores are not retained merely as records of hypothetical choices. They are interpreted as quantitative preference magnitudes and used as dependent variables in regression models that generate health-state utility algorithms. Statistical estimation cannot establish the measurement properties missing from the

original scores. Nor can subsequent QALY calculations supply an empirical relationship between preference and duration that has never been demonstrated.

The consequences for subsequent utility estimation are immediate. TTO scores become dependent-variable observations in regression models used to estimate multiattribute utility algorithms. However, if these scores have not been established as quantitative measures of preference, what exactly is the regression estimating? Its dependent variable is a dimensionless ratio of durations assigned an unvalidated preference interpretation. Regression may identify statistical relationships between these ratios and EQ-5D health-state classifications, but it cannot establish that the resulting coefficients are measurements of preference.

PROCESS VALIDATION: THE IMPORTANCE OF ATTRIBUTES IN THERAPY-IMPACT ASSESSMENT

The fundamental requirement of representational measurement is that a quantitative claim must concern an explicitly defined attribute. A measurement scale represents the empirical relationships of that attribute, and the arithmetic permitted upon its numerical values depends upon the properties of the scale. Numerical assignments to heterogeneous health-state descriptions do not establish a common measurable attribute. Measurement must begin with attribute identification, not with the construction of a numerical score. Unidimensionality, empirical relational structure and admissible transformations are prerequisites for interpreting numerical values as measures of therapy impact.

For ratio-based therapy-impact assessment, this framework recognizes two measurement pathways. The first concerns manifest attributes, whose empirical properties support linear ratio measurement. These require an invariant unit, a non-arbitrary zero and admissible similarity transformations, permitting meaningful differences and proportional comparisons. The second concerns latent attributes, which require an appropriate measurement model. Rasch probabilistic conjoint measurement provides the adopted pathway for establishing a unidimensional latent continuum, with attribute locations expressed in logits. Changes are evaluated through logit differences, with exponentiation providing corresponding odds ratios. Each pathway requires independent empirical validation, invariance and justification of the proposed quantitative interpretation.

The distinction is fundamental. A measured duration, distance or frequency belongs to the dimensional class of its particular manifest attribute. A Rasch logit measure represents a location on a latent continuum. These quantities are not interchangeable. A duration cannot simply be added to a logit measure, and independently measured latent attributes cannot be aggregated merely because each has been assigned numerical values. Dimensional homogeneity and commensurability determine which operations are legitimate. Numerical representation does not make different attributes quantitatively interchangeable.

The legacy HTA reference case, the one for the National Institute for Health and Care Excellence (NICE) for example, does not establish this manifest–latent distinction as the foundation of its quantitative framework ⁹ . It does not require separate measurement pathways or demonstration of the relevant scale properties before preference scores are admitted into utility estimation. Rasch

measurement, including its requirements for unidimensionality, invariance and conjoint measurement, is not incorporated into the foundational reference-case approach. Instead, health-state descriptions receive preference scores without first establishing the measurement structure of the proposed preference attribute. The omission is more fundamental than the absence of a particular technique: the reference case does not make attribute-based measurement the organizing principle of therapy-impact assessment.

The EQ-5D descriptive system illustrates the problem. It classifies health states across mobility, self-care, usual activities, pain/discomfort and anxiety/depression. These represent heterogeneous aspects of health, involving observable manifestations and latent experiences. Each dimension is described through ordered response categories rather than an independently established quantitative continuum. Combining these classifications into a health-state profile does not create a common unit, a non-arbitrary zero or a single underlying attribute. A health-state profile is a multidimensional description, not a quantitative measure of therapy impact.

Why should therapy-impact assessment begin with such a description? The legacy reference case seeks a common numerical expression through which heterogeneous health outcomes can be compared across diseases and interventions. Health-state profiles are assigned preference scores, which are subsequently transformed into utility values. However, the demand for a universal summary value does not establish the existence of a universal measurable attribute. Representational measurement requires the reverse sequence: identify the attribute, establish its measurement properties and determine the admissible arithmetic.

A respondent may legitimately prefer one multidimensional health state to another. Such judgments provide information about choices between alternatives, but they do not establish that the constituent attributes are commensurable or that preference for the overall health state possesses ratio-scale properties. If preference is proposed as a latent attribute, it must be independently defined and its measurement structure demonstrated. Even if lawful proportional preference comparisons could be established, their equivalence to TTO duration ratios would require separate empirical validation. Preference cannot inherit the mathematical properties of duration merely because duration was used in its elicitation.

The implications for therapy-impact assessment are substantial. A scientifically defensible claim must identify the specific attribute affected by treatment, establish the appropriate measurement pathway and quantify change according to the properties of that scale. For a manifest attribute, this may involve a difference or proportional change in a linear ratio measure. For a latent attribute, it may involve a change in Rasch logit location, accompanied by its standard error and an odds-ratio interpretation. These claims are attribute-specific, empirically evaluable, replicable and potentially falsifiable. They do not require a hypothetical multidimensional health-state preference score as an intermediary.

The decisive distinction is between describing a health state and measuring therapy impact. EQ-5D may provide useful classifications, and TTO may record hypothetical choices, but neither procedure establishes the measurement properties required for the quantitative claims attributed to its outputs. Process validation therefore rejects the admission of health-state preference scores as ratio-scale measures without independent empirical warrants.

For therapy-impact measurement, multidimensional health-state valuation is an unnecessary intermediary. The scientific task is to measure change in explicitly defined manifest and latent attributes through validated measurement procedures. Legacy HTA reverses this requirement by constructing preference scores before demonstrating measurement. Once attribute-based measurement becomes the foundation of quantitative assessment, the claimed necessity of a universal health-state utility disappears.

CONCLUSION: THE CASE FOR ABANDONING TIME TRADE-OFF

Time trade-off (TTO) has been challenged for decades over its theoretical foundations, framing effects, duration dependence, respondent consistency and measurement properties. These criticisms have encouraged attempts to improve elicitation procedures and statistical estimation. However, the present process-validation analysis identifies a more fundamental problem. TTO has not established the measurement foundations required for its interpretation as a ratio-scale measure of health-state preference.

Two independent requirements remain unmet. First, overall preference for a multidimensional health-state description has not been independently established as a single attribute possessing the required ratio-scale properties. Second, no empirically validated relationship has been demonstrated between ratios of preference magnitudes and the dimensionless duration ratios calculated through TTO. The legitimacy of the duration arithmetic cannot establish the legitimacy of its preference interpretation.

These are not deficiencies that can be corrected by changing elicitation formats, increasing sample sizes, improving regression fit or introducing more sophisticated utility algorithms. Such modifications may improve the performance of the procedure, but they cannot substitute for demonstrating the required measurement properties. The failure concerns what TTO claims to measure, not how efficiently its scores can be generated.

Process validation therefore requires a stop decision before TTO scores enter utility estimation. Continuing to construct preference weights, QALYs and cost-effectiveness ratios from scores whose quantitative interpretation remains unsupported perpetuates the original measurement failure¹⁰. Institutional acceptance and decades of application cannot discharge the burden of empirical proof.

TTO should be abandoned as a method for generating quantitative health-state preference measures for utility and QALY construction. Therapy-impact assessment should instead begin with explicitly defined attributes and validated measurement procedures: linear ratio measurement for appropriate manifest attributes and Rasch measurement for latent attributes. The objective is not to improve the numerical valuation of hypothetical health states but to measure the actual attributes affected by treatment.

The appropriate response to foundational measurement failure is abandonment, not further refinement of an unsupported quantitative claim.

NOTE: THE TRANSITION TO REPRESENTATIONAL MEASUREMENT

The reconstruction of health technology assessment (HTA) does not require the development of an entirely new analytical framework. A structured program for replacing legacy HTA with representational measurement has already been developed by Maimon Research: *Reconstructing HTA Programs: Measurement before Arithmetic* ¹¹. The nine-unit program provides an immediately available foundation for reconstructing HTA around attribute-based measurement and process validation. Its purpose is not to add a measurement component to conventional HTA while preserving its existing analytical structure. Its purpose is to replace that structure.

The program reverses the organizing principle of legacy HTA. Instead of beginning with a desired economic calculation and identifying numerical inputs to populate it, assessment begins with an explicitly defined attribute and the empirical requirements for measuring it. Only after the measurement properties have been established can admissible arithmetic and quantitative therapy-impact claims be considered. Measurement precedes arithmetic, and process validation determines whether each proposed quantitative operation is justified.

Across nine units, the program addresses the foundations of representational measurement; attribute identification and measurement claims; scale properties and admissible arithmetic; dimensional homogeneity and commensurability; manifest attributes and linear ratio measurement; latent attributes and Rasch measurement; attribute possession and change; process validation; and the construction, empirical evaluation, replication and falsification of therapy-impact claims. Two measurement pathways provide the foundation: linear ratio measurement for appropriate manifest attributes and Rasch measurement for latent attributes. Each requires independently demonstrated measurement properties, invariance and admissible quantitative interpretation.

The legacy framework of multiattribute utilities, QALYs and cost-per-QALY claims is not retained as the organizing basis for quantitative assessment. Where these procedures fail process validation, their claimed measurement status cannot be rescued by statistical estimation, model validation, sensitivity analysis or computational sophistication. A mandatory STOP decision terminates the unsupported quantitative pathway. The objective is not to refine the legacy reference case but to replace arithmetic without measurement with empirically defensible therapy-impact claims.

The scientific object of assessment becomes the attribute-specific therapy-impact claim. A clinically or patient-relevant attribute is identified, an appropriate measurement protocol is established, and a quantitative hypothesis is formulated for empirical evaluation. The resulting claim must be open to independent replication and potential falsification. Its scientific standing derives from demonstrated measurement properties and empirical testing, not from incorporation into a hypothetical economic simulation.

Maimon Research supports this transition through its existing nine-unit program, published papers and technical notes, worked teaching materials, class notes, curriculum and knowledge-base assessments, and faculty and professional-development resources. The program provides a practical foundation for graduate education, research and professional retraining. There is no

requirement to wait for another international consensus, professional task force or generation of methodological debate. The framework for reconstruction is available now.

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 D, Luce R, Suppes P, Tversky A. Foundations of Measurement Vol 1: Additive and Polynomial Representations. New York: Academic Press, 1971
- ³ Dolan P. Modelling valuations for EuroQol health states. *Medical Care*, 35(11), 1997; 1095–1108.
- ⁴ Torrance G. Social preferences for health states: An empirical evaluation of three measurement techniques. *Socio-Economic Planning Sciences*, 10(3), 1976; 129–136.
- ⁵ Rasch, G. (1960/1980). Probabilistic Models for Some Intelligence and Attainment Tests. Expanded edition, with foreword and afterword by Benjamin D. Wright. Chicago: University of Chicago Press
- ⁶ Bond T, Zi Yan, Heene M. Applying the Rasch Model: Fundamental measurement in the Human Sciences (4th Ed). New York: Routledge, 2021
- ⁷ The EuroQol Group. (1990). EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy*, 16(3), 199–208
- ⁸ Brooks R. EuroQol: The current state of play. *Health Policy*, 37(1), 1996; 53–72
- ⁹ National Institute for Health and Care Excellence (NICE). (2026). NICE technology appraisal and highly specialised technologies guidance: the manual. NICE process and methods [PMG36]. Originally published 31 January 2022; updated 13 October 2026.
- ¹⁰ 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

¹¹ Maimon Research. *Reconstructing HTA Programs: Measurement before Arithmetic*.
<https://maimonresearch.com/hta-reconstruction-program-and-fees/>