

MAIMON RESEARCH LLC

**ARTIFICIAL INTELLIGENCE LARGE LANGUAGE
MODEL INTERROGATION**



LEGACY HTA DECONSTRUCTION

PAPER 2

PROCESS VALIDATION: THE IMPOSSIBLE UTILITY

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

LOGIT WORKING PAPER No 9085 SEPTEMBER 2026

www.maimonresearch.com

ABSTRACT

BACKGROUND Legacy health technology assessment (HTA) relies upon regression-derived health-state utilities as quantitative inputs to quality-adjusted life years (QALYs) and cost-per-QALY claims. Paper 1 established that time trade-off (TTO) scores originate in legitimate duration ratios but lack the demonstrated measurement properties required for their interpretation as proportional preference magnitudes. Paper 2 examines whether regression estimation and the subsequent EQ-5D utility algorithm can establish the missing measurement foundations.

METHODS Process validation applies the requirements of representational measurement sequentially to the TTO dependent variable, regression coefficients designated preference weights, and the arithmetic construction of EQ-5D utilities. The assessment examines attribute identification, unidimensionality, meaningful quantitative differences, commensurability, dimensional homogeneity, empirical additivity, invariance, non-arbitrary zero and proportional interpretation. Mandatory STOP decisions identify failures that statistical estimation or algorithmic calculation cannot remedy.

RESULTS The TTO dependent variable lacks an independently demonstrated ratio-scale preference attribute and an empirically validated mapping between preference space and duration space. Regression estimates numerical differences in predicted TTO scores but cannot establish that its coefficients represent invariant, commensurable preference magnitudes. The utility algorithm compounds these failures by combining estimated coefficients as decrements from an assigned reference value of unity. Neither empirical additivity nor a common unidimensional preference attribute is established. Assigning full health a value of one and death zero does not demonstrate a natural zero or meaningful proportional comparisons. Negative utilities further contradict their interpretation as nonnegative ratio-scale preference magnitudes. The resulting scores are calculable numerical indices, not demonstrated quantitative measures.

CONCLUSIONS Process validation identifies an unprecedented institutionalization of arithmetic without measurement. For more than four decades, legacy HTA has embedded unsupported quantitative interpretations within professional guidelines, reference cases, reimbursement assessments and graduate education. Statistical sophistication, predictive performance and computational reproducibility cannot establish measurement properties absent from the underlying constructs. The conventional EQ-5D utility is therefore an impossible proportional preference measure under the required standards of representational measurement. The mandatory STOP decision terminates its admission as a ratio-scale input to QALY construction. Paper 3 examines the additional requirement for an empirically demonstrated mapping from preference magnitudes to proportional durations, establishing why multiplication cannot recover the missing measurement foundations. Reconstruction requires attribute-based measurement, admissible arithmetic and empirically testable therapy-impact claims rather than continued refinement of legacy numerical models.

INTRODUCTION

The construction of EQ-5D utilities represents the second stage in the numerical architecture of legacy health technology assessment (HTA). Paper 1 demonstrated that time trade-off (TTO) produces a legitimate dimensionless ratio of durations but fails to establish either a ratio-scale preference attribute or an empirically validated mapping between preference space and time space¹. The resulting TTO preference score therefore has no independently demonstrated measurement properties as a preference magnitude. In representational-measurement terms, it is an imaginary quantitative construct: a number assigned a preference interpretation without establishing the attribute or scale properties necessary to support that interpretation.

Rather than resolving this foundational failure, legacy HTA proceeds directly to regression analysis. The TTO score is introduced as the dependent variable and treated as though it possesses the quantitative properties required for its intended interpretation. The fundamental question appears to have been overlooked: what exactly does the dependent variable measure, and what evidence establishes that it possesses the scale properties required for the quantitative claims derived from the regression? Instead of answering this question, the procedure accepts the assigned TTO number and proceeds to statistical estimation. An unvalidated numerical construct becomes the foundation for generating an entire system of purported preference measures.

A regression dependent variable must be appropriate to the statistical model employed and to the substantive interpretation placed upon its outputs. Generalized linear models can accommodate different response distributions and measurement types; they do not universally require ratio-scale dependent variables. However, where the objective is to estimate quantitative preference decrements and construct utilities interpreted as proportional preference magnitudes, the required measurement properties cannot simply be assumed. The dependent variable must represent an identified attribute with an empirically justified quantitative structure sufficient to support the proposed interpretation of coefficients and subsequent arithmetic. The TTO preference score has not been demonstrated to meet these requirements. Its numerical availability is mistaken for measurement validity.

The distinction between statistical estimation and representational measurement is fundamental²

³ Regression can estimate associations, generate coefficients and predict numerical outcomes without establishing that the dependent variable measures the attribute it purports to represent. Statistical fit cannot demonstrate a natural zero, invariant unit, unidimensionality or meaningful proportional comparisons. Nor can it establish that regression coefficients represent commensurable contributions to a single quantitative preference attribute. Statistical estimation cannot manufacture measurement properties absent from the dependent variable.

The remarkable feature of the legacy HTA framework is that this preliminary measurement question was never made an explicit condition for constructing utility algorithms. The methodological emphasis fell instead upon model specification, estimation procedures, coefficient interpretation and predictive performance. The prior scientific question whether the dependent variable constituted a lawful measure of preference was effectively bypassed. The TTO number was simply inserted into the regression, and the resulting statistical outputs were accepted as the

basis for constructing utilities. This is measurement inversion in its most direct form: arithmetic precedes the demonstration of measurement.

Paper 2 applies process validation to this second stage of legacy HTA. Its starting point is the measurement status of the dependent variable, not the computational performance of the regression. The paper asks whether the numerical preference scores inherited from TTO provide a defensible empirical foundation for interpreting estimated coefficients as quantitative preference weights and for constructing utility measures.

The central proposition is straightforward: legacy HTA takes an imaginary preference magnitude, subjects it to statistical manipulation and treats the resulting coefficients as though they possess empirical measurement meaning. The failure is not that regression cannot process the numbers. It is that no independent measurement warrant has been established for the preference quantities the regression is claimed to estimate.

PROCESS VALIDATION: THE REQUIRED PROPERTIES OF UTILITY

Process validation must begin by asking what measurement properties an anticipated utility would require to function as a proportional adjustment factor in the possible construction of a quality-adjusted life year (QALY). The intended outcome is expressed as $Q = U \times T$, where duration (T) is multiplied by utility (U) to produce a purported duration equivalent in perfect health. The first question is not how to calculate utility, but whether a numerical quantity can be established that legitimately performs this proportional adjustment. These requirements must be identified before any preference elicitation, regression estimation or utility algorithm is engineered.

Duration belongs to time space [T] and possesses the properties of linear ratio measurement. A legitimate proportional adjustment to duration can be expressed as T_1/T_2 , where both quantities measure the same attribute and are therefore commensurable. Their dimensions cancel, yielding a dimensionless proportion. For utility to perform the intended adjustment, it must possess an empirically justified proportional interpretation applicable to duration. Merely assigning a dimensionless numerical value between zero and one does not establish this property. Dimensional homogeneity is necessary, but dimensional cancellation alone cannot establish the empirical meaning of a proposed adjustment factor.

The difficulty arises when utility is derived from health-state preferences rather than independently measured durations. Preference and duration occupy different measurement spaces. Even if a unidimensional ratio-scale preference attribute could be established, its proportional magnitudes would not automatically correspond to proportional durations. The proposed equality between a preference ratio P_i/P^f and a duration ratio T_1/T_2 requires an independently demonstrated empirical mapping. This is precisely the measurement failure identified in Paper 1. The TTO score begins as a lawful duration ratio but is assigned a proportional preference interpretation without establishing either the required preference measure or the relationship between the two spaces. As a purported ratio-scale preference magnitude, the conventional TTO score is an impossible construct within its demonstrated measurement foundations.

A further obstacle concerns the anticipated utility itself. EQ-5D describes health through five heterogeneous dimensions, yet the intended utility is a single quantitative preference score. Process validation must therefore ask how these multidimensional classifications could establish one unidimensional attribute with an invariant unit, a non-arbitrary zero and meaningful proportional comparisons. It must also establish whether numerical contributions from different dimensions are commensurable and whether their combination is empirically additive. A scoring algorithm cannot demonstrate these properties merely by assigning one number to each health-state profile.

These are prior conditions, not problems to be resolved retrospectively through statistical estimation. The anticipated utility must have a demonstrated proportional meaning, a defensible unidimensional measurement structure and an independently validated relationship with duration if it is to perform its proposed role. Paper 1 established that TTO does not supply these warrants. Paper 2 begins by asking whether the subsequent utility-construction process has any measurement foundation capable of satisfying them. The purpose is not to presume that QALY construction is achievable, but to identify the requirements whose absence must terminate the process before arithmetic is mistaken for measurement.

PROCESS VALIDATION: REQUIRED PROPERTIES OF THE DEPENDENT VARIABLE

Having established the measurement properties that utility would require to function as a proportional adjustment factor to duration, process validation must now examine the dependent variable selected for the regression model. The purpose of this regression is to estimate preference weights from time trade-off (TTO) observations and use those weights to populate an EQ-5D utility algorithm. The critical question is whether the dependent variable possesses the measurement properties necessary to support the quantitative interpretation intended for the resulting coefficients and utilities. This question must be answered before regression estimation begins.

A dependent variable is not a measure merely because it consists of numerical observations. Its empirical meaning must be established independently of the statistical model. Generalized linear regression can accommodate different response types and distributions, but statistical admissibility does not establish measurement validity. Where the resulting coefficients are intended to represent quantitative preference decrements, the dependent variable must measure an identified attribute on a scale that supports that interpretation. If the coefficients are subsequently to be combined into purported proportional preference magnitudes, the necessary scale properties, commensurability and admissibility of the proposed arithmetic must be demonstrated.

Paper 1 established that the TTO preference score fails this preliminary requirement. Its numerical value is derived from a legitimate ratio of two commensurable durations within time space. However, no independent ratio-scale preference attribute has been established, and no empirical mapping has demonstrated that the duration ratio represents an equivalent proportion within preference space. The TTO score therefore has no demonstrated measurement status as a proportional preference magnitude. In the context of the proposed regression, it is an imaginary quantitative preference construct: a number with an assigned interpretation but without the empirical measurement properties required to justify that interpretation.

The decisive failure occurs when this number is accepted as the dependent variable without first questioning its measurement status. Regression can estimate relationships between the assigned TTO scores and EQ-5D classifications, generate coefficients and predict numerical outcomes. It cannot establish the missing preference attribute, create a natural zero or demonstrate meaningful proportional comparisons. Nor can it validate the relationship between preference and duration that Paper 1 found unsupported. The regression processes the TTO numbers as given; it does not transform them into lawful measures of preference.

This distinction exposes the fundamental inversion underlying legacy HTA. The methodological literature has concentrated on regression specification, estimation procedures, statistical fit and the interpretation of coefficients as preference weights. The prior question whether the dependent variable is an empirically established measure capable of supporting those interpretations has not been made a foundational requirement. The procedure moves directly from assigned numerical scores to statistical estimation, treating computational success as sufficient justification for quantitative meaning.

The process-validation assessment therefore identifies a fundamental measurement failure at the point where the dependent variable is selected. The TTO score has not been demonstrated to possess the properties required for its intended interpretation as a proportional preference magnitude. Under process validation, this finding would require the measurement foundation of the proposed regression to be reconsidered before proceeding. Legacy HTA proceeded regardless. The TTO score was accepted as the dependent variable, apparently without an independently validated alternative, and regression estimation became the mechanism for generating preference weights. Required measurement properties were never an issue of interest. regression estimates differences in predicted TTO scores. It does not establish that these differences represent measured preference magnitudes.

PROCESS VALIDATION: REGRESSION COEFFICIENT REQUIREMENTS FOR PREFERENCE WEIGHTS

The designation of regression coefficients as preference weights introduces a separate measurement claim. For coefficients to qualify as quantitative preference weights, they must represent measurable contributions to an explicitly defined preference attribute.

Unidimensionality and meaningful differences. The proposed weights must represent changes along a common preference continuum. Their differences must have an empirically established quantitative meaning. Estimating coefficients on the same numerical response scale does not demonstrate that they measure a single attribute.

Commensurability. The five EQ-5D dimensions describe heterogeneous aspects of health: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Coefficients associated with these dimensions can be compared numerically because they share the same dependent variable. However, numerical compatibility does not establish that they represent commensurable contributions to one preference attribute.

Invariance. Regression coefficients depend upon reference categories, model specification, included interactions and the valuation population. Changing the reference category can change individual coefficients without changing fitted predictions. These are legitimate statistical properties, but an individual coefficient cannot automatically be treated as an invariant preference magnitude. The empirical meaning of the proposed measure must be preserved under the relevant admissible transformations.

Process-validation finding. Regression can estimate coefficients, standard errors and predicted scores. It cannot independently establish unidimensionality, commensurability, meaningful preference differences or measurement invariance. The coefficients remain statistical parameters rather than demonstrated quantitative preference weights.

PROCESS VALIDATION: REQUIRED MEASUREMENT PROPERTIES OF THE UTILITY ALGORITHM

To examine the algorithm independently, assume provisionally that its inputs possess the required measurement properties. Process validation must then determine whether the arithmetic preserves those properties and produces a legitimate quantitative measure.

A simplified additive utility algorithm is:

$$U = 1 - (wM + wS + wA + wP + wD + C)$$

Here, U is the utility score; wM , wS , wA , wP and wD are decrements associated with the five EQ-5D dimensions; and C represents additional adjustments. Particular value sets may include intercepts, interactions or other penalties.

If the combined decrements equal 0.40:

$$U = 1 - 0.40 = 0.60$$

The arithmetic is valid. Whether the result represents a measured preference magnitude requires additional justification.

Dimensional Homogeneity and Commensurability

Dimensional homogeneity requires quantities added or subtracted to belong to the same dimensional class, while commensurability requires those quantities to represent contributions to the same measured attribute, expressed in compatible units.

The algorithm combines coefficients associated with five heterogeneous health dimensions. Even assuming that the coefficients are valid measurements, their combination requires evidence that they represent commensurable contributions to a single preference attribute. The production of one numerical score does not establish this relationship.

Empirical Additivity

The simplified additive algorithm assumes that the combined preference decrement can be calculated from its component decrements.

For mobility (M) and pain/discomfort (P), the additive proposition is:

$$D(M + P) = D(M) + D(P)$$

If the separate decrements are 0.10 and 0.20, the simplified additive model assigns their combination a decrement of 0.30.

Whether the empirical preference loss is actually 0.30 must be demonstrated independently. Adding coefficients does not establish empirical additivity. Some EQ-5D value sets include interaction terms or additional adjustments, but these modifications do not independently establish a lawful quantitative measurement structure. They alter the statistical specification without demonstrating that the resulting arithmetic corresponds to empirically established preference magnitudes.

Zero, Unity and Proportional Interpretation

The algorithm assigns full health a reference value of 1 and death a value of 0. These conventions do not establish an invariant unit or a non-arbitrary zero.

For $U = 0.60$ to represent 60% of a preference magnitude, the resulting utility must possess demonstrated ratio-scale properties. Subtracting numerical decrements from unity does not establish that proportional interpretation.

The algorithm also permits negative utilities:

$$U = 1 - 1.20 = -0.20$$

A negative score cannot represent a negative magnitude on a conventional nonnegative ratio scale with a natural zero. The capacity to generate such scores is incompatible with interpreting the entire resulting scale as nonnegative preference proportions.

Invariance of the Resulting Utility

The final utility must represent a position on an empirically established, invariant measurement continuum. Generating one numerical value for every EQ-5D health-state profile does not demonstrate that those values measure a single preference attribute.

The algorithm must establish that its unit, origin and quantitative relationships preserve their meaning under the admissible transformations of the claimed measurement scale. Numerical prediction cannot substitute for this evidence.

PROCESS VALIDATION: THE ALGORITHM AS A MEASUREMENT OPERATION

The process-validation assessment has identified failures in both the TTO dependent variable and the subsequent interpretation of regression coefficients as preference weights. To examine the utility algorithm independently, assume provisionally that these earlier failures have been resolved and that the estimated weights are legitimate quantitative measures. The question is now whether the algorithm combines these measures through empirically justified operations to produce a single proportional magnitude of health-state preference.

This distinction is fundamental. The possession of legitimate numerical inputs does not automatically make the arithmetic performed upon them a legitimate measurement operation. Each operation must be justified by the empirical properties of the quantities involved. The algorithm must demonstrate that its component weights represent commensurable contributions to the same preference attribute, that addition and subtraction preserve dimensional homogeneity, and that the assumed relationships between component and combined preference magnitudes satisfy empirical additivity.

Consider an algorithm that assigns a mobility decrement of 0.10 and a pain/discomfort decrement of 0.20. It may calculate a combined decrement of 0.30 and subtract this from unity to produce a utility of 0.70. These calculations are mathematically valid. They do not demonstrate that the combined health state represents an empirically measured preference loss of 0.30, or that the resulting score represents 70% of a preference magnitude. Both interpretations require independent measurement evidence.

The algorithm must also establish the measurement properties of its output. Assigning full health a value of unity and death a value of zero does not establish a non-arbitrary zero or an invariant measurement unit. Nor does generating a single numerical score demonstrate that the result represents one unidimensional preference attribute. The existence of negative utilities creates an additional incompatibility with interpreting the resulting scores as nonnegative proportional magnitudes.

The conventional EQ-5D algorithm specifies how coefficients are selected, combined and subtracted from a reference value. It does not independently demonstrate that these operations correspond to empirical relationships among measured preference magnitudes. Its statistical and arithmetic structure therefore cannot substitute for the measurement structure that must justify the calculations.

The process-validation finding is that the conventional utility algorithm is a scoring procedure, not a demonstrated measurement operation. Even provisionally accepting its inputs as legitimate measures, the algorithm has not established that its output is a lawful ratio-scale preference magnitude. This constitutes an independent measurement failure, additional to those identified in TTO and regression.

The further question of whether a legitimate preference proportion could be mapped into a proportional adjustment of duration is separate. That question belongs to the subsequent examination of QALY construction in Paper 3.

PROCESS VALIDATION: THE STOP AND REVIEW DECISIONS

Process validation has now examined the two critical stages through which conventional EQ-5D utilities are constructed: the estimation of regression coefficients from TTO preference scores and the subsequent combination of those coefficients within a utility algorithm. Each stage must satisfy its own measurement requirements before its numerical outputs can be admitted as quantitative measures. Failure at either stage is sufficient to terminate the process. Statistical estimation and arithmetic calculation cannot compensate for the absence of the empirical properties required to justify their intended measurement interpretations.

The assessment identifies two independent mandatory stop decisions.

STOP 1: REGRESSION COEFFICIENTS CANNOT BE ADMITTED AS QUANTITATIVE PREFERENCE WEIGHTS

The first STOP decision occurs when regression coefficients are interpreted as quantitative preference weights. The regression dependent variable consists of TTO scores derived from ratios of durations. Although these are legitimate numerical ratios within time space, Paper 1 established that their interpretation as proportional preference magnitudes has not been justified by an independently demonstrated preference attribute or an empirically validated mapping between preference and duration.

Regression does not resolve this failure. It estimates statistical relationships between the assigned TTO scores and the categorical descriptions of EQ-5D health states. The resulting coefficients represent estimated differences in predicted TTO scores, conditional upon the model specification. They do not automatically represent measured differences in preference magnitudes.

For the coefficients to qualify as quantitative preference weights, process validation would require evidence that they represent meaningful differences along a single preference continuum, that their contributions are commensurable across the five EQ-5D dimensions, and that their measurement interpretation is invariant under the relevant admissible transformations. These requirements must be demonstrated independently of statistical significance, goodness of fit or predictive performance.

The distinction is decisive. A coefficient of -0.20 may describe an estimated numerical decrement in a regression prediction. It does not establish that the associated health-state classification produces an empirically measured preference loss of 0.20 units. The numerical coefficient exists; the purported quantitative preference magnitude has not been independently established.

STOP 1 is therefore mandatory. Regression coefficients cannot be admitted as measured preference weights without demonstrating the measurement properties necessary for that interpretation. The appropriate response is to return to the definition and measurement of the

preference attribute, rather than proceed to utility construction. Legacy HTA nevertheless accepted these statistical parameters as the quantitative inputs to its utility algorithms.

STOP 2: THE UTILITY ALGORITHM CANNOT BE ADMITTED AS A MEASUREMENT OPERATION

The second STOP decision is independent of the first. To isolate the algorithm, assume provisionally that the regression-derived coefficients have satisfied all necessary measurement requirements and represent legitimate quantitative preference weights. It does not follow that combining them produces a legitimate measure of health-state preference.

The algorithm must independently demonstrate that its arithmetic corresponds to empirical relationships among the quantities being combined. Dimensional homogeneity requires the operations to preserve the dimensional meaning of the inputs. Commensurability requires the component weights to represent contributions to the same measured attribute. Empirical additivity requires evidence that combined preference magnitudes correspond to the proposed arithmetic relationships, rather than merely to the numerical specification of a regression model.

The algorithm must also establish the measurement properties of its resulting utility. A single numerical score does not demonstrate a single underlying preference attribute. Assigning full health a value of unity and death a value of zero does not establish a natural zero or invariant measurement unit. Nor does subtracting estimated decrements from unity demonstrate that a resulting utility of 0.60 represents 60% of an empirically established preference magnitude.

Negative utilities create an additional problem. An algorithm that generates values below zero cannot consistently interpret its entire output scale as nonnegative proportional preference magnitudes. Such values may be produced by the scoring convention, but their numerical existence does not establish the ratio-scale interpretation required for the proposed quantitative claims.

These requirements are not satisfied merely because the algorithm generates a reproducible score for every EQ-5D health-state profile. Reproducibility establishes that the same computational procedure produces the same numerical result. It does not establish that the result measures a preference magnitude with an invariant unit, meaningful zero and lawful proportional relationships.

STOP 2 is therefore independently mandatory. Even if the coefficients were legitimate measurements, the conventional algorithm has not demonstrated that its operations constitute a lawful measurement procedure or that its outputs possess the required ratio-scale properties. The algorithm remains a numerical scoring procedure rather than an independently validated quantitative measure of preference.

THE CONSEQUENCES OF THE TWO STOP DECISIONS

The significance of these findings extends beyond the technical construction of EQ-5D utilities. The two STOP decisions expose a fundamental failure in the sequence through which legacy HTA has attempted to establish quantitative health-state preferences.

At the first stage, numerical observations whose interpretation as preference magnitudes has not been established are introduced into regression models. Statistical coefficients are then designated preference weights without independently demonstrating their measurement properties. At the second stage, these coefficients are combined through an algorithm whose arithmetic has not been validated as an empirical measurement operation. The resulting utility is nevertheless treated as a quantitative preference magnitude suitable for subsequent multiplication by duration.

The failure is cumulative, but each STOP decision is independently sufficient to reject the claimed measurement interpretation. Improving regression estimation cannot establish the missing measurement properties of the dependent variable. Refining the utility algorithm cannot establish that its inputs represent measured preference magnitudes. Neither procedure can retrospectively create a natural zero, invariant unit, empirical additivity or meaningful proportional relationships.

There is also a fundamental distinction between validating a statistical model and validating a quantitative measurement claim. A regression may accurately predict TTO responses, and an algorithm may consistently reproduce its prescribed utility scores. Neither finding independently demonstrates that those scores measure the preference attribute they purport to represent. Without such a demonstration, there is no established quantitative preference magnitude against which the resulting utility can be independently evaluated as a measurement.

This raises a further scientific requirement: before a utility algorithm is implemented within an HTA reference case, its proposed quantitative claims must be capable of independent empirical evaluation and potential falsification. If a utility of 0.60 is interpreted as 60% of a preference magnitude, the measurement procedure must establish what that magnitude is, how its proportions are defined, and what empirical observations could contradict the claimed value. Agreement between statistical predictions and the numerical scores used to construct the model does not satisfy this requirement.

The conventional utility-construction process does not supply these independent measurement warrants. Instead, statistical estimation and algorithmic calculation have been accepted as sufficient grounds for treating assigned numerical scores as quantitative measures.

The final process-validation decision is therefore stop. Neither the regression coefficients nor the resulting EQ-5D utilities can be admitted as demonstrated ratio-scale preference measures. The pathway must terminate before the utilities are introduced into QALY construction.

The consequence is not that legacy HTA requires a better regression model or a more sophisticated utility algorithm. It requires a prior reconstruction of the measurement process itself. An attribute must first be defined, its measurement properties demonstrated, its permissible arithmetic established, and its quantitative claims made independently evaluable and falsifiable.

Process validation must precede implementation. Measurement must precede arithmetic. No HTA reference case should authorize a quantitative claim whose required measurement properties and empirical evaluation procedures have not first been established.

PROCESS VALIDATION: THE IMPOSSIBLE UTILITY

The conventional EQ-5D utility is not a quantitative measure whose accuracy remains uncertain. It is a numerical construct whose claimed measurement properties have never been demonstrated. Process validation exposes a succession of failures that cannot be corrected through improved statistical estimation, alternative regression specifications or more elaborate algorithms. The problem is foundational: arithmetic has been substituted for measurement.

The TTO procedure begins with a legitimate ratio of durations in time space [T] but assigns that ratio the meaning of a proportional preference magnitude in preference space [P]. Neither the required ratio-scale preference attribute nor the mapping between the two spaces has been established. The resulting TTO preference score is therefore an imaginary quantitative construct: a number to which an empirical measurement interpretation has been assigned without demonstrating its existence.

Regression compounds this failure. Coefficients estimated from the purported preference scores are designated preference weights, although their status as meaningful, commensurable and invariant measures of a single preference attribute has not been established. The utility algorithm then combines these coefficients as decrements from an arbitrarily assigned reference value of unity. The arithmetic is executable, but empirical additivity, dimensional homogeneity, a natural zero and proportional interpretation remain unsupported. Negative utilities provide further evidence that the resulting scores cannot represent nonnegative ratio-scale preference magnitudes.

What makes this failure unprecedented is the systematic application of arithmetic without measurement across an entire international discipline. For more than four decades, legacy HTA has institutionalized numerical procedures through reference cases, professional guidelines, reimbursement decisions and graduate education without first requiring the measurement warrants that give quantitative claims empirical meaning. Statistical sophistication has concealed rather than resolved the absence of measurement. The ability to calculate, predict, aggregate and replicate numerical scores has been mistaken for evidence that the quantities supposedly represented by those scores actually exist.

The utility is consequently a meaningless construct as a claimed quantitative measure of preference. It remains a calculable numerical index, but calculation does not confer measurement status. Neither statistical fit nor replication of the algorithm can establish the missing attribute, invariant unit, natural zero or lawful proportional relationships. The measurement failure is not a defect within an otherwise legitimate utility framework; it is a failure of the framework itself.

The mandatory process-validation decision is stop. The conventional utility cannot be admitted as a ratio-scale preference measure. Its subsequent multiplication by duration cannot restore the measurement properties it lacks. The unprecedented failure of legacy HTA is therefore not merely the production of an impossible utility, but the institutional acceptance of numerical storytelling as quantitative science. The next stage of deconstruction establishes why this failure necessarily carries forward into the impossible QALY.

CONCLUSION: THE TRANSITION TO THE IMPOSSIBLE QALY

Process validation has demonstrated that the conventional EQ-5D utility cannot be admitted as a lawful quantitative measure of health-state preference. The successive failures of TTO, regression-derived preference weights and the utility algorithm are not isolated methodological deficiencies. They expose a systemic failure in which arithmetic has substituted for measurement. Numerical scores have been constructed, manipulated and interpreted as quantitative magnitudes without establishing the empirical attributes and measurement properties required to justify those interpretations.

Legacy HTA represents an unprecedented measurement disaster. For more than four decades, an international framework of reference cases, professional guidelines, reimbursement assessments and graduate education has institutionalized arithmetic without measurement. The result is not simply an uncertain estimate of preference but a numerical construct whose claimed quantitative meaning has not been established. Statistical sophistication and computational reproducibility cannot compensate for the absence of lawful measurement.

The consequences extend directly to the QALY. In the equation $Q = U \times T$, duration is a legitimate ratio measure within time space [T], but the conventional utility is not an established proportional adjustment factor. Even if an independently valid ratio-scale preference measure were available, its proportional magnitudes could not automatically be interpreted as proportional durations. The required mapping between preference space [P] and time space [T] would still have to be demonstrated.

Paper 2 therefore terminates the utility pathway at the mandatory stop decision. Multiplication cannot recover measurement properties that utility never possessed. Paper 3 extends the analysis to the impossible QALY, examining the independent requirements for lawful proportional adjustment of duration and the absence of an empirically demonstrated mapping between preference and time spaces.

The challenge facing HTA is no longer to refine legacy numerical models but to replace them with an analytical framework in which measurement precedes arithmetic and therapy-impact claims meet the standards of normal science.

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* ⁴. 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^{5 6}. 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 Langley P. Process Validation: The Impossible Time Trade-Off. Logit Paper No 9080.
<https://maimonresearch.com/logit-working-paper-no-9080-october-2026/>

² Stevens S. On the Theory of Scales of Measurement. *Science*. 1946;103(2684):677-8

³ Krantz D, Luce R, Suppes P, Tversky A. Foundations of Measurement Vol 1: Additive and Polynomial Representations. New York: Academic Press, 1971

⁴ Maimon Research. *Reconstructing HTA Programs: Measurement before Arithmetic*.
<https://maimonresearch.com/hta-reconstruction-program-and-fees/>

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