

MAIMON RESEARCH LLC
ARTIFICIAL INTELLIGENCE LARGE LANGUAGE
MODEL INTERROGATION



LEGACY HTA DECONSTRUCTION
PAPER 3
PROCESS VALIDATION: THE IMPOSSIBLE QALY

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

LOGIT WORKING PAPER No 9082 OCTOBER 2026

www.maimonresearch.com

ABSTRACT

BACKGROUND The quality-adjusted life-year (QALY) has dominated health technology assessment (HTA) for over four decades without demonstrated foundations in representational measurement. This third paper in a four-part process-validation series examines whether conventional health-state utilities can legitimately function as proportional adjustments to survival duration. The central question is whether $Q = U \times T$ produces a lawful measure of quality-adjusted time or merely an arithmetic construct lacking empirical measurement meaning.

METHODS Process validation applies the requirements of representational measurement, including unidimensionality, commensurability, dimensional homogeneity, invariance, non-arbitrary zero and admissible arithmetic. Two independent requirements are assessed: whether conventional utility possesses the ratio-scale properties necessary for meaningful proportional comparisons, and whether an empirically validated mapping establishes equivalence between preference proportions and duration proportions. Mandatory STOP decisions identify failures that prevent further quantitative interpretation.

RESULTS Conventional EQ-5D utilities fail the first requirement. TTO scores originate in legitimate duration ratios but are assigned proportional preference interpretations without demonstrating a corresponding ratio-scale preference attribute. Regression coefficients designated preference weights do not establish commensurable preference magnitudes, while utility algorithms fail to demonstrate empirical additivity, an invariant unit or a natural zero. Consequently, conventional utility scores remain numerical indices rather than lawful preference proportions. The second requirement fails independently. Even if a legitimate preference ratio P_1/P_2 existed, its equivalence to a duration ratio T_1/T_2 would require empirical demonstration. Both ratios may be dimensionless without representing equivalent proportional relationships. Conventional QALY construction supplies no such mapping. Process validation therefore identifies two mandatory STOP decisions before utility can legitimately adjust duration.

CONCLUSIONS The conventional QALY cannot be admitted as a lawful quantitative measure under the requirements of representational measurement. Its historical institutionalization illustrates the consequences of allowing arithmetic to precede measurement. The elaborate progression from TTO elicitation through regression, utility algorithms and cost-per-QALY modelling represents an unnecessary encumbrance to HTA. Therapy impact can instead be assessed through attribute-specific profiles employing linear ratio measurement for manifest attributes and validated Rasch measurement for latent attributes. Claims must be empirically evaluable, replicable and falsifiable. The appropriate response is abandonment of the QALY-dependent framework rather than further numerical refinement. Paper 4 extends process validation to legacy reference case simulation models and their quantitative claims.

INTRODUCTION

The concept of the quality-adjusted life-year (QALY) emerged from developments in health-status measurement and decision analysis during the 1960s and 1970s. Its central proposition was that numerical valuations of health states could be combined with duration to express health outcomes in units equivalent to years in full health. The attraction was straightforward: preference elicitation and multiattribute health-state questionnaires would generate utilities that could serve as proportional adjustments to time. The QALY subsequently became central to legacy health technology assessment (HTA), providing the denominator for cost-per-QALY claims¹. Yet its scientific legitimacy depends upon a fundamental requirement that has never been demonstrated: the existence of a lawful utility measure capable of providing the required proportional adjustment to duration.

The preceding two papers established why this requirement fails. Paper 1 examined the time trade-off (TTO), demonstrating that its purported preference score is an imaginary quantitative construct². TTO produces a legitimate ratio of two commensurable durations within time space [T], but interprets that ratio as a proportional preference magnitude without establishing either an independent ratio-scale preference attribute or an empirical mapping between duration and preference spaces. The duration ratio exists as a lawful measure; the corresponding preference ratio has not been established. Simple arithmetic produces a numerical preference score, but representational measurement rejects its claimed quantitative interpretation.

Paper 2 examined the subsequent attempt to construct EQ-5D utilities³. It demonstrated that regression cannot establish the preference weights required to populate a utility algorithm when its dependent variable is an imaginary TTO preference score. The estimated coefficients are statistical parameters, not independently demonstrated preference magnitudes. Their designation as preference weights does not establish unidimensionality, invariance or commensurability. The utility algorithm introduces further failures by combining these coefficients without demonstrating empirical additivity, a non-arbitrary zero or meaningful proportional comparisons. Paper 2 therefore concluded that the conventional utility is an impossible measurement construct: a numerical output that has not been established as a ratio-scale preference measure capable of serving as a proportional adjustment factor.

This conclusion provides the starting point for Paper 3. The conventional QALY is expressed as:

$$Q = U \times T$$

Here, T is duration, a manifest attribute possessing the properties of linear ratio measurement. Within time space, proportional adjustments are entirely legitimate. Given two commensurable durations, T₁ and T₂, their ratio T₁/T₂ is dimensionless and represents a lawful proportional relationship. For example, a duration ratio of 0.6 establishes that one measured duration is 60% of another. Such a factor can support a proportional adjustment within time space because its numerator and denominator are measures of the same attribute.

The problem is that the conventional utility does not provide an equivalent proportional quantity. A utility of 0.6 is not an independently demonstrated preference magnitude representing 60% of a

full-health preference measure. It is the numerical outcome of the failed measurement process examined in Papers 1 and 2. There is therefore no established preference ratio, P_1/P_2 , corresponding to the legitimate duration ratio T_1/T_2 . The proposed preference measurement space lacks the independently demonstrated ratio-scale structure required to construct such a proportion.

Even if an independently validated ratio-scale preference attribute could be established, a further problem would remain. The proposed equality:

$$P_1/P_2 = T_1/T_2$$

would require an empirical mapping between preference space [P] and time space [T]. Both ratios could be dimensionless and internally commensurable without being empirically equivalent. Dimensional cancellation within one measurement space does not establish a proportional relationship with another. The missing mapping cannot be supplied by assigning numerical values to health states or by multiplying those values by duration.

The QALY therefore fails at two distinct levels. First, the conventional utility has no established existence as the proportional preference measure required by the calculation. There is consequently no lawful preference ratio available to map into time space. Second, even if such a ratio existed, its correspondence with a duration ratio would still require independent empirical demonstration. The conventional QALY establishes neither requirement.

This third paper develops the extended proof of the impossible QALY. Its purpose is to demonstrate that the multiplication of a numerical utility by a legitimate duration measure cannot produce a lawful measure of quality-adjusted time. The problem is not computational. The expression $Q = U \times T$ can be evaluated arithmetically. The problem is that its proposed empirical interpretation requires a proportional adjustment factor whose measurement existence has not been established and a mapping between preference and duration spaces that has not been demonstrated.

The distinction explains why representational measurement is indispensable^{4 5}. It establishes whether numerical values correspond to empirical quantities, whether their scale properties support the proposed arithmetic and whether relationships between different measurement spaces have been independently validated. Without these requirements, arithmetic can produce numerical results that possess no demonstrated status as measures.

The impossible QALY is therefore not an isolated defect in the final multiplication. It is the culmination of the imaginary TTO preference score, the impossible utility and the absence of an empirically established mapping between preference and duration. The conventional framework cannot demonstrate the proportional adjustment it claims to perform. What remains is numerical calculation without lawful measurement.

PROCESS VALIDATION: REPRESENTATIONAL MEASUREMENT AND THE QALY

The purpose of representational measurement is to establish the conditions under which numbers can legitimately represent empirical quantities. Without such conditions, numerical assignments

remain conventions, classifications or computational devices. They may be useful, but their existence does not establish measurement. The fundamental scientific problem is therefore not how to assign numbers to observations, but how to demonstrate that the relationships among those numbers correspond to relationships in the empirical world.

If the QALY is to be accepted as a valid quantitative construct supporting decisions and resource allocation within health systems, it must be demonstrated that its proposed adjustment to duration satisfies the requirements of representational measurement. Under the conventional legacy HTA framework, the adjustment is based on preferences for multidimensional health-state descriptions rather than independently measured individual health attributes. The first requirement is therefore to establish that these preferences can be represented by a lawful ratio-scale measure capable of supporting meaningful proportional comparisons. The second is to demonstrate an empirically validated mapping through which a proportional preference magnitude can be translated into an equivalent proportional adjustment within time space. The purpose of process validation is to determine whether these two indispensable measurement requirements have been satisfied before the QALY calculation is admitted as a measure. Numerical operations cannot establish the existence of a preference measure or demonstrate its proportional equivalence to duration. Without these independent measurement warrants, the QALY remains an arithmetic construct rather than an empirically established measure of health outcomes.

Representational measurement provides the bridge between empirical observations and mathematical operations. It requires an identifiable attribute, an empirical relational structure and a numerical representation that preserves the relevant relationships. The properties of that representation determine which arithmetic operations are admissible. Measurement must therefore precede arithmetic. Numbers cannot be multiplied, divided or interpreted as proportions merely because they have been assigned to health states, preferences or other observations.

Stevens' classification of nominal, ordinal, interval and ratio scales made the distinction between numerical representation and permissible operations explicit. Subsequent developments in representational measurement theory formalized the empirical and mathematical requirements underlying these distinctions. A ratio scale, for example, requires more than a numerical zero and positive values. It requires an empirical structure supporting meaningful ratios, a non-arbitrary zero and transformations that preserve proportional relationships. These properties cannot be established through statistical estimation or numerical normalization alone.

The scientific importance of these requirements is that they make quantitative claims open to empirical challenge. A proposed measurement model must specify what is being measured, how the attribute is represented, which relationships are expected to hold and what observations would contradict the model. Replication, invariance and falsification become meaningful because the measurement claim has an explicit empirical foundation. Without that foundation, numerical results may be reproducible as calculations while remaining unsupported as measurements.

This distinction is particularly important in HTA. The objective is not simply to produce numerical comparisons of therapies but to establish empirically defensible claims about their impact. A statistical model may predict assigned scores accurately without demonstrating that those scores represent quantities. A simulation may generate thousands of numerical outcomes without

establishing that its inputs possess the properties required for the arithmetic performed upon them. Computational sophistication cannot compensate for missing measurement foundations.

Representational measurement exists precisely to prevent this substitution of calculation for measurement. It provides the scientific discipline necessary to distinguish an empirical quantity from a numerical fiction. Its requirements are not obstacles to quantitative analysis; they are the conditions under which quantitative analysis acquires empirical meaning.

The QALY provides a decisive test. If a utility is to function as a proportional adjustment to duration, its legitimacy must be demonstrated through the appropriate measurement properties and an empirically established relationship to time. Without these warrants, the expression $Q = U \times T$ remains an arithmetic construction rather than a demonstrated measure of health benefit.

PROCESS VALIDATION: UNDERSTANDING PROPORTIONAL ADJUSTMENT

It is important to understand the critical role of proportional adjustment in ensuring the failure of legacy HTA with its focus on arithmetic and an absence of representational measurement. 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 utility preference measurement 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 adjustment factor exists as a lawful ratio-scale measure and whether its application to another attribute space is empirically justified. These are separate measurement requirements that must be satisfied before the resulting numerical values can be admitted as measures.

A fundamental principle of representational measurement is that a proportional adjustment factor must have an empirically established meaning within the attribute space in which it operates. It may be constructed directly from two commensurable ratio measures within that space, as in T_1/T_2 for duration within time space $[T]$, or transferred from another attribute space through an independently validated empirical mapping. A dimensionless ratio is not automatically transferable between spaces. Thus, 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: Paper 2 established that utility has no demonstrated existence as a lawful ratio-scale preference measure, while no validated mapping establishes that its purported preference proportion corresponds to a proportional adjustment within time space. These failures provide the sequential stopping rules for process validation.

The reason for this restriction is that a proportional adjustment expresses an empirical relationship between measured magnitudes, not simply an arithmetic relationship between numbers. A ratio has quantitative meaning because its numerator and denominator are commensurable measures of the same attribute, with 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. Their numerical equality does not establish their empirical equivalence. Applying a preference proportion to duration therefore requires evidence that the preference relationship corresponds to the proposed proportional relationship in time space. Without that evidence, multiplication remains computationally possible but lacks the empirical justification required for measurement. Representational measurement imposes this restriction because arithmetic must preserve established empirical relationships; it cannot create relationships between attributes that have never been demonstrated.

At the risk of repetition, it is important to recognize that legacy HTA has developed its quantitative methods without systematically applying the standards of representational measurement or, more specifically, the implications of Stevens' 1946 classification of measurement scales. Stevens distinguished nominal, ordinal, interval and ratio scales, establishing that the empirical properties of a measurement scale determine which numerical transformations and arithmetic interpretations are admissible. This distinction should be fundamental to any discipline claiming to produce quantitative measures of health outcomes. Yet legacy HTA has proceeded as though the assignment of numerical values to health-state preferences is sufficient to justify subsequent arithmetic operations. Preference scores are subjected to regression analysis, regression coefficients are combined into utilities, utilities are multiplied by duration, and the resulting QALYs become denominators in cost-effectiveness calculations.

At no stage does the conventional framework require independent demonstration that the numerical quantities possess the measurement properties necessary to support their claimed interpretation. The restrictions imposed by scale type and admissible arithmetic have consequently remained outside the effective epistemic knowledge base governing legacy HTA. This is not a minor methodological omission. It represents a fundamental inversion of the scientific sequence in which measurement must precede arithmetic. The computational procedures have become accepted conventions, while the empirical requirements that would establish their legitimacy as measurement have been displaced. Process validation reverses this inversion by requiring each proposed quantitative operation to satisfy the relevant measurement conditions before its results can be admitted into subsequent calculations. The central failure of legacy HTA is therefore not an inability to perform arithmetic, but its institutional acceptance of arithmetic without first establishing the lawful measures that would give the resulting numbers empirical meaning.

PROCESS VALIDATION: THE ABSENT MEASUREMENT PROPERTIES OF UTILITY AS A PROPORTIONAL ADJUSTMENT FACTOR

The first question in assessing the QALY is whether the conventional health-state utility possesses the measurement properties necessary to function as a proportional adjustment factor. Before considering the multiplication of utility by duration, or any proposed mapping between preference and time spaces, it is necessary to establish that utility exists as a lawful quantitative ratio measure of preference. The ability to assign numerical values to health states does not establish that those values represent measurable preference magnitudes or that they support proportional comparisons. As we have emphasized the failure of legacy HTA as evidence by the global application of

reference case modeling is that it ignores representational measurement. The reference case is built on number crunching arithmetic with no consideration of the constraints imposed by lawful measurement. A position that has held without challenge for 40 years.

The legacy HTA conventional QALY is expressed as $Q = U \times T$, where U is utility and T is duration. The intended interpretation is that utility represents a proportion of full health, permitting duration to be adjusted accordingly. Thus, a utility of 0.6 is treated as though it represents 60% of the preference magnitude associated with full health and can therefore reduce a duration of ten years to six quality-adjusted years. This interpretation makes two distinct measurement claims. First, utility must exist as an empirically established ratio-scale preference measure capable of supporting meaningful proportional comparisons. Second, the resulting preference proportion must be demonstrably equivalent to a proportional adjustment within time space. Even if a preference ratio $P_1/P_2 = 0.6$ were independently established, this would not demonstrate that the corresponding duration ratio T_1/T_2 must also equal 0.6. That equivalence requires an independently validated mapping between preference space $[P]$ and time space $[T]$.

The conventional QALY establishes neither requirement. Paper 2 demonstrated that the conventional EQ-5D utility has no existence as a lawful ratio-scale preference measure once the standards of representational measurement are applied. Its numerical values originate in TTO duration ratios that have been assigned an unsupported preference interpretation. These purported preference scores are subsequently used as dependent variables in regression models, whose estimated coefficients are designated preference weights without demonstrating that they measure a single preference attribute. The coefficients are then combined through a utility algorithm that assumes, rather than establishes, commensurability, empirical additivity, invariance and a non-arbitrary zero. The resulting utility scores are therefore model-generated numerical indices, not independently established measures of preference magnitude. A utility of 0.6 is simply an output of this scoring procedure; it has not been demonstrated to represent 60% of the preference magnitude associated with full health. It ranks health state descriptions. The numerical scale is constructed by convention and statistical estimation, not by establishing the empirical properties required for ratio measurement. Consequently, there is no lawful preference proportion P_1/P_2 available to serve as a putative proportional adjustment factor.

This leads to a further question: where do conventional utility scores fit within Stevens' scales of measurement? The classification proposed by Stevens in 1946 distinguished nominal, ordinal, interval and ratio scales according to their measurement properties and the transformations and arithmetic operations they support. Nominal scales permit classification, ordinal scales establish order, interval scales support meaningful differences, and ratio scales possess equal units and a non-arbitrary zero, permitting meaningful proportional comparisons. These distinctions are fundamental because assigning numbers to observations does not establish that the numbers possess the properties required for quantitative arithmetic.

Conventional EQ-5D utilities are numerical indices derived from preference elicitation, regression estimation and an additive scoring algorithm. Their construction does not establish equal empirical units, a non-arbitrary zero or meaningful ratios of preference magnitude. Nor does the assignment of numerical values between zero and one demonstrate interval or ratio properties. At most, these scores may support an ordinal interpretation of health-state preferences, insofar as their ordering

corresponds to empirically established preference judgments. Even that interpretation requires evidence that the scoring algorithm preserves the relevant empirical order. They cannot be admitted as interval or ratio measures merely because regression coefficients have been estimated and combined arithmetically. The conclusion is decisive for QALY construction: a numerical utility score without demonstrated ratio-scale properties cannot represent a proportional preference magnitude and therefore cannot supply the proportional adjustment factor required to multiply duration. The fact that a utility of 0.6 can be multiplied by ten years to produce six numerical units does not establish that six quality-adjusted years have been measured. The arithmetic is computationally possible, but the claimed empirical interpretation fails process validation.

PROCESS VALIDATION: TWO FAILURES IN THE PROPORTIONAL ADJUSTMENT OF DURATION

The proposed use of utility as a proportional adjustment to duration confronts two distinct measurement requirements. The first concerns the measurement status of utility itself. The second concerns the empirical relationship required to translate a legitimate preference proportion into an equivalent duration proportion. Both must be satisfied before utility can perform its intended role in QALY construction. These should have been apparent before the QALY was proposed.

First: The absence of a lawful ratio-scale preference measure. A utility intended to represent a proportional preference magnitude must be grounded in an identified, unidimensional attribute possessing an invariant unit, a non-arbitrary zero and meaningful proportional relationships. Paper 2 demonstrated that conventional EQ-5D utility construction does not establish these properties. TTO preference scores, regression-derived preference weights and the subsequent utility algorithm generate numerical values without demonstrating the required ratio-scale measurement structure. Consequently, a utility of 0.60 cannot be admitted as an empirically established 60% preference magnitude. The purported preference space [P] lacks the demonstrated properties necessary to support a meaningful ratio P_1/P_2 .

Second: The absence of an empirically demonstrated mapping to time space. Even if an independent ratio-scale preference measure were established, its proportions would not automatically correspond to proportions of duration. Preference space [P] and time space [T] concern different attributes. A legitimate proportional adjustment to duration can be expressed as T_1/T_2 , where both quantities are commensurable ratio measures of time. To use a preference proportion as an equivalent duration proportion would require an independently demonstrated empirical relationship establishing $P_1/P_2 = T_1/T_2$. Neither numerical agreement, stated indifference nor the fact that both ratios are dimensionless establishes this equivalence.

These failures are logically separate. The first establishes that conventional utility has not been demonstrated to possess the measurement properties required of a proportional preference magnitude. The second establishes that even a legitimate preference proportion could not be applied to duration without an independently validated relationship between the two measurement spaces.

Process validation therefore requires two STOP decisions. Conventional EQ-5D utility fails the preliminary requirement for ratio-scale preference measurement. Independently, the proposed transformation from preference proportions to duration proportions lacks an established empirical mapping. The QALY calculation $Q = U \times T$ cannot acquire a lawful quantitative interpretation

merely because the multiplication is mathematically executable. Its proposed proportional adjustment requires measurement warrants that have not been demonstrated.

PROCESS VALIDATION: THE FORMAL IMPOSSIBLE QALY PROOF

The publication of tens of thousands of papers supporting the QALY and its application to therapy-impact claims is no defense of the QALY as a legitimate product of representational measurement. Far from it. The sheer volume of publications raises a far more disturbing question: how has a quantitative framework achieved such widespread acceptance without first demonstrating that its central outcome satisfies the required standards of representational measurement?

Publication, replication of numerical procedures and institutional endorsement cannot establish measurement legitimacy. If the QALY fails the requirements of representational measurement, the accumulation of papers employing it merely demonstrates the extraordinary extent to which an unestablished measurement construction has become embedded in global HTA. The fact that a simple proof can make clear the failure of the QALY is a question those who have promoted legacy HTA for over 40 years will have to come to terms with. The QALY is the defining example of the failure of legacy HTA: forty years of numerical elaboration without first establishing the legitimacy of the measure upon which the entire QALY-dependent framework rests.

The QALY is not simply a convenient summary statistic. It makes a specific quantitative claim: that a period of time can be proportionally adjusted by the value assigned to a health state to create a new quantity, quality-adjusted time. This claim requires more than numbers. It requires measurement.

The critical distinction is between assigning a numerical value to a health state and demonstrating that the value possesses the measurement properties required to adjust time proportionally. A preference score can rank health states, summarize responses or represent the output of a statistical valuation model without possessing the ratio properties required for multiplication.

The measurement question is therefore prior to the familiar QALY calculation. Before U can multiply T, it must be established what U measures, what empirical structure it represents, what scale properties it possesses and why its numerical value can legitimately be interpreted as a proportion of duration. Without that proof, the arithmetic has no demonstrated quantitative meaning.

The conventional QALY combines duration with a numerical value intended to represent the quality of the health state experienced during that duration:

$$Q = U \times T$$

$A_t = T_1/T_2$ The interpretation is explicitly proportional. If $U = 1$, duration is unadjusted. If $U = 0.5$, one year in the health state is treated as one-half of a quality-adjusted year. The value U must therefore perform the role of a proportional adjustment factor.

Duration provides a simple demonstration of what a lawful proportional adjustment requires. Consider two durations T_1 and T_2 . Because both quantities are measures of duration, they belong to the same dimensional class, denoted by $[T]$, where the square brackets $[]$ indicate the dimension of the quantity rather than its numerical value or unit of measurement. Thus $[T]$ denotes the dimension of time, irrespective of whether duration is expressed in seconds, days, months or years. In dimensional terms:

$$[A_t] = [T_1]/[T_2] = [T]/[T] = 1$$

The result is dimensionless because the time dimensions cancel. The notation $[A_t] = 1$ therefore means that A_t has no physical dimension. It does not, by itself, establish that A_t has proportional meaning.

If $T_1 = 6$ months and $T_2 = 10$ months:

$$A_t = 6/10 = 0.6$$

The interpretation is clear. Six months is 60% of ten months. The result has proportional meaning because T_1 and T_2 are commensurable ratio measures of the same attribute, duration.

The result is also invariant under a permissible change of unit. If duration is transformed according to:

$$T' = aT$$

then:

$$aT_1/aT_2 = T_1/T_2$$

Changing from months to years does not alter the ratio.

This demonstrates why describing U as dimensionless is insufficient. A dimensionless ratio has proportional meaning only when its numerator and denominator are commensurable ratio measures of the same attribute on a common measurement scale. Dimensional cancellation alone does not establish proportional meaning. If $U = 0.6$ is to mean that duration should be reduced to 60% of its original value, the corresponding question is unavoidable:

What two commensurable ratio measures of the same attribute generate the proportion represented by U ?

Conventional health-state preference systems do not provide such a construction. Their values typically originate in preferences expressed for multidimensional health states. Different aspects of health are described, valued and combined to generate a single numerical score. The resulting number may be useful as an index of preference, but assigning it a numerical value does not establish that it is a unidimensional ratio measure. As demonstrated in Paper 2, the utility or

preference score has no independently established interval- or ratio-scale measurement properties. It is a constructed numerical index, not a demonstrated quantitative measure of preference.

Anchoring a preference scale at zero and one does not resolve the problem. Reference points assigned by convention are not equivalent to demonstrating a non-arbitrary zero, equal units and meaningful ratios. The existence of negative utility values for health states considered worse than death creates a further difficulty. Negative values are incompatible with interpreting zero as the absence of a nonnegative magnitude on a linear ratio scale. The conventional utility scale therefore lacks the demonstrated structure required for meaningful proportional comparisons.

The QALY consequently encounters its first measurement failure before multiplication even begins. Utility has not been demonstrated to possess the unidimensional ratio-scale properties required for its interpretation as a proportion. Without these properties, U cannot legitimately be treated as the basis for a measured proportional adjustment factor for duration.

There is also a second, independent failure.

Suppose, purely for argument, that two preference quantities P_1 and P_2 were demonstrated to be commensurable ratio measures of the same preference attribute. We could then construct:

$$A_p = P_1/P_2$$

and:

$$[A_p] = [P]/[P] = 1$$

A_p could then have legitimate proportional meaning within preference space. That would still not make it a proportional measure of time.

The proportional adjustment required within time space remains:

$$A_t = T_1/T_2$$

A_p and A_t may both be dimensionless, but they arise from different attributes and different measurement spaces. Their common absence of physical dimensions does not establish their equivalence.

A separate mapping would therefore be required:

$$f: A_p \rightarrow A_t$$

This mapping would have to demonstrate why a particular proportional relationship in preference space corresponds to the same proportional adjustment in time space.

The conventional QALY does not demonstrate this relationship. It simply substitutes U into:

$$Q = U \times T$$

The multiplication assumes the mapping that measurement must first establish. The proportional relationship between preference and duration required to justify its calculation has never been demonstrated

The QALY therefore fails at two sequential points. First, conventional health-state preference values have not been demonstrated to possess the unidimensional ratio properties required for proportional interpretation. Second, even if such properties were provisionally granted in preference space, an independent mapping would still be required before that preference proportion could lawfully adjust ratio-measured duration. This has never been demonstrated.

The bottom line is clear: the QALY has no independent existence as an analytical construct.

CONCLUSION: THE QALY IS THE UNNECESSARY ENCUMBRANCE TO HTA

The survival of the QALY raises a fundamental historical question: why was arithmetic allowed to precede measurement? Health economists in the 1960s and 1970s sought a common numerical framework for comparing healthcare interventions and allocating resources. The QALY offered an apparently convenient solution by combining survival duration with health-state valuations. Yet the requirements of representational measurement were already established. The critical failure was to accept numerical preference scores as proportional adjustment factors without first demonstrating their ratio-scale properties or an empirical mapping into time space. Once institutionalized through reference cases, professional guidelines, journals and graduate education, the QALY became an accepted convention rather than a measurement hypothesis requiring independent validation. Increasing statistical sophistication reinforced this inheritance without resolving its foundational defects.

Process validation identifies two decisive failures. First, conventional EQ-5D utilities have not been established as lawful ratio-scale preference measures. They lack a demonstrated unidimensional attribute, invariant unit, non-arbitrary zero and meaningful proportional interpretation. Second, even if legitimate preference proportions existed, their equivalence to duration proportions would require an independently validated mapping between preference space [P] and time space [T]. Neither requirement has been satisfied. Regression estimation, alternative utility algorithms and reference case simulations cannot establish the missing measurement properties. The conventional QALY cannot exist as the lawful quantitative measure it purports to represent. Its continued use illustrates the consequences of institutionalizing arithmetic without measurement.

Do we need the QALY? No. It is an unnecessary encumbrance and an embarrassment to HTA. The elaborate progression from TTO elicitation through regression-derived preference weights, utility algorithms, QALYs and cost-per-QALY simulations consumes substantial scientific effort without establishing a legitimate measure of therapy impact. It can be replaced by direct measurement of identifiable attributes through two pathways: linear ratio measurement for manifest attributes and Rasch measurement for latent attributes. An attribute-specific profile

permits quantitative therapy-impact claims expressed in admissible units, with uncertainty, independent replication and opportunities for falsification. There is no scientific requirement to collapse heterogeneous health outcomes into an imaginary quantity of quality-adjusted time.

The appropriate response is therefore not further refinement of the QALY but its abandonment as the organizing principle of HTA. Graduate education, professional practice and reimbursement assessment must be reconstructed around representational measurement, process validation and empirically evaluable therapy-impact claims. Procedural competence in manipulating inherited numerical constructs cannot substitute for measurement competence. Measurement must precede arithmetic, and every quantitative claim must be capable of independent empirical evaluation. Paper 4 completes the deconstruction by examining the legacy reference case simulation model, where unsupported utilities and QALYs are incorporated into purportedly quantitative claims about therapy impact and resource allocation.

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* ⁶. 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 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

² Langley P. Process Validation: The Impossible Time Trade-Off. Logit Paper No 9080.
<https://maimonresearch.com/logit-working-paper-no-9080-october-2026/>

³ Langley P. Process Validation: The Impossible Utility. Logit Paper No 9085.
<https://maimonresearch.com/logit-working-paper-no-9085-september-2026/>

⁴ Stevens S. On the Theory of Scales of Measurement. *Science*. 1946;103(2684):677-8

⁵ Krantz D, Luce R, Suppes P, Tversky A. *Foundations of Measurement Vol 1: Additive and Polynomial Representations*. New York: Academic Press, 1971

⁶Maimon Research. *Reconstructing HTA Programs: Measurement before Arithmetic*.
<https://maimonresearch.com/hta-reconstruction-program-and-fees/>