MAIMON RESEARCH LLC

RECONSTRUCTION OF HEALTH TECHNOLOGY ASSESSMEMNT



THE TIME TRADE-OFF: A FALSE TRAIL FROM HEALTHCARE PROGRAM VALUATION TO MEASUREMENT FAILURE

Paul C Langley PhD Adjunct Professor, College of Pharmacy,

University of Minnesota, Minneapolis, MN

LOGIT WORKING PAPER No 2031 NOVEMBER 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

Background: *The Time Trade-Off (TTO) procedure has provided a foundation for health-state utility valuation and quality-adjusted life-year (QALY) calculations in conventional health technology assessment (HTA). Introduced by Torrance and colleagues in the early 1970s, TTO generates numerical coefficients from respondents' indifference judgments between different combinations of health status and survival duration. These coefficients subsequently became inputs to multidimensional health-state utility tariffs and reference-case economic evaluations. Their measurement status, however, requires examination against the axioms of representational measurement.*

Objective: *To assess whether the transition from a TTO duration-equivalence ratio to a health-state utility coefficient satisfies the requirements of lawful measurement, and to consider the implications for HTA education and the evaluation of therapeutic benefit.*

Methods: *The analysis applies representational measurement theory, including attribute specification, representation and uniqueness theorems, scale properties, dimensional homogeneity, commensurability and admissible arithmetic. Particular attention is given to Torrance's 1976 conditional interval utility argument, its reliance upon expected-utility theory, and the additional assumptions connecting health-state preferences with survival duration.*

Findings: *The TTO coefficient, $h = t/T$, is a lawful dimensionless ratio derived from two measures of survival duration. Its ratio properties are established within time space. Reinterpreting this coefficient as health-state utility introduces a different attribute, preference, whose measurement properties are not established by the original arithmetic. Expected-utility theory permits conditional interval representations, but mathematical consistency does not demonstrate that the required empirical axioms and structural assumptions hold. Interval representations do not independently establish meaningful ratios between utility levels. Although additional assumptions may support a multiplicative health-duration decision model, they do not independently establish a ratio measure of preference or an empirically validated mapping between preference and duration. The proposed equality between a preference ratio, p/P, and the duration ratio, t/T, therefore remains an unresolved measurement claim. Regression-based utility tariffs and QALY calculations cannot independently remedy this deficiency.*

Conclusion: *TTO is a legitimate duration-equivalence procedure but not an independently validated measurement method for therapeutic benefit. Its continued presentation as a measurement foundation perpetuates the inversion of measurement and arithmetic within legacy HTA. A reconstructed curriculum must instead prioritize lawful measurement of specified manifest and latent attributes, validated scale properties, admissible arithmetic and falsifiable therapy-impact claims. TTO belongs in methodological history, not the core measurement curriculum. The governing scientific principle is measurement before arithmetic.*

INTRODUCTION

The Time Trade-Off (TTO) procedure occupies a foundational position in conventional health technology assessment (HTA). Introduced by Torrance and colleagues in the early 1970s as an alternative to the standard gamble, it was developed to elicit preferences for multiattribute health states rather than evaluate individual health attributes. Its subsequent adoption provided a numerical foundation for health-state utility algorithms, including those associated with the EQ-5D descriptive system. These algorithms supply the utility coefficients used in quality-adjusted life-year (QALY) calculations and reference-case cost-effectiveness models. Over the past five decades, this approach has become central to the economic evaluation of healthcare programs, an orientation reflected explicitly in Drummond and colleagues' *Methods for the Economic Evaluation of Health Care Programmes* [1] .

The development of representational measurement theory during the twentieth century established a fundamentally different foundation for quantitative evaluation. Across the physical sciences and increasingly within the more rigorous traditions of the social sciences, the scientific priority was the identification of specified attributes, the empirical demonstration of their measurement properties, and the establishment of mathematical representations that preserve the relevant empirical relationships. Representation and uniqueness theorems provided the formal basis for determining the properties of measurement scales and the admissibility of numerical operations. This framework received a major systematic formulation in 1971 with the publication of Volume I of *Foundations of Measurement* by Krantz, Luce, Suppes and Tversky, building upon earlier developments in measurement theory, including Stevens' influential 1946 classification of nominal, ordinal, interval and ratio scales [2 3] The subsequent publication of Volumes II and III in 1989 and 1990 further developed this theoretical foundation [4 5] .

Legacy HTA developed along a different trajectory. Rather than making the lawful measurement of individual attributes the foundation for evaluating therapy impact, it adopted an analytical framework centered on the valuation of multidimensional health states and the economic comparison of healthcare programs. The emphasis was placed on composite claims of program value rather than the demonstrated impact of interventions on specified manifest and latent attributes. The subsequent reliance on reference-case modelling reinforced this orientation, producing aggregate cost-effectiveness estimates whose empirical validity depended upon the measurement properties of their inputs and the assumptions governing their combination. The fundamental concern is that numerical modelling was allowed to assume a central role without first establishing the representational-measurement foundations required for the quantities being modelled. This marked a departure from the requirements of normal science, in which claims concerning therapy impact should be empirically testable, replicable and open to falsification.

The question confronting health economists during the 1970s and 1980s was how to obtain numerical valuations of health states that could be incorporated into economic evaluations of competing healthcare programs. The emerging framework required a method for assigning values to multidimensional health-state descriptions so that differences in survival and health status could be combined within a common analytical structure. Von Neumann and Morgenstern's 1944 formulation of expected-utility theory provided an influential theoretical foundation for representing preferences under uncertainty [6] . Its application to health-state valuation encouraged

the use of the standard gamble, in which respondents' express preferences between living in a specified health state with certainty and accepting a gamble involving alternative outcomes, commonly full health and death. The resulting probabilities were interpreted as health-state utility coefficients. However, the standard gamble imposed substantial cognitive demands upon respondents, requiring them to understand and evaluate uncertain outcomes. Torrance and colleagues therefore introduced the Time Trade-Off (TTO) in the early 1970s as a more straightforward alternative for eliciting health-state valuations [7].

In the TTO formulation, respondents are presented with a specified duration, T, in a described health state and asked to identify a shorter duration, t, in full health that they consider equivalent. The resulting indifference judgment permits calculation of the coefficient h = t/T. Both durations are measured on a ratio scale, possessing equal units and a non-arbitrary zero. Their quotient is therefore a legitimate dimensionless ratio in time space. For example, if a respondent considers six years in full health equivalent to ten years in a specified impaired health state, the resulting coefficient is 0.6. What has been directly established is a duration-equivalence judgment expressed as a ratio of measured durations**.** No independent measure of preference magnitude has been obtained.

The subsequent step was to interpret this duration-equivalence coefficient as a measure of health-state utility [8] . The numerical value was retained, but its proposed measurement interpretation changed. A ratio established in time space became a coefficient intended to represent preference for a multidimensional health-state description. This transition required a theoretical justification beyond the ratio properties of the underlying durations. Torrance and colleagues drew upon expected-utility theory and additional assumptions concerning preferences for health states and duration to support an interval utility interpretation. Under the relevant axioms, such a representation may be mathematically justified. The decisive question, however, is whether the required preference axioms and health–duration assumptions are empirically satisfied by the TTO responses. Without that demonstration, the numerical identity between the duration ratio and the proposed utility coefficient cannot establish the claimed interval measurement properties.

This distinction becomes particularly important when the TTO is incorporated into health-state utility algorithms and subsequently into QALY-based economic evaluation. The original duration-equivalence judgment is transformed into a general-purpose coefficient intended to represent the value of time spent in different health states. Yet the statistical estimation of utility algorithms cannot independently establish the empirical preference structure required to justify that interpretation. Nor does the resulting composite valuation provide measurements of the individual manifest and latent attributes affected by a healthcare intervention. The central issue is therefore not whether a numerical coefficient can be generated, but whether the coefficient possesses the measurement properties required for its intended scientific and decision-making purposes.

This question ix explored in Appendix A which examines the foundational measurement claim underlying the time trade-off (TTO): that a lawful ratio of survival durations can be interpreted as an equivalent measure of health-state preference. It distinguishes the dimensionally homogeneous and commensurable duration ratio, h = t/T, from the proposed preference ratio, k = p/P, and identifies the independent measurement requirements necessary to establish each. It then considers the further requirement for an empirically demonstrated mapping between preference space and

time space. The analysis demonstrates why numerical equality between two dimensionless ratios does not, by itself, establish measurement equivalence or justify the use of health-state utility scores as proportional adjustments to survival duration in QALY calculations.

The purpose of this paper is twofold. First, it examines the proposed transition from the TTO duration-equivalence coefficient to an interval representation of health-state utility and whether that transition satisfies the axioms and empirical requirements of representational measurement. Second, and potentially more controversially, it questions whether the TTO retains any legitimate methodological role in HTA curricula once representational measurement is accepted as the scientific foundation for evaluating therapy impact through individually measured manifest and latent attributes. The focus shifts from scales of measurement to precede arithmetic to the ratio scale to support, for manifest attributes, a linear ratio measure and for latent attributes a Rasch logit ratio measure [9] .

ATTRIBUTES AND REPRESENTATIONAL MEASUREMENT

The foundation of quantitative evaluation is the measurement of defined attributes. Representational measurement requires that numerical values correspond to empirically demonstrable properties of an attribute [10] . Measurement is not simply the assignment of numbers to observations, classifications or preferences. It requires an established relationship between an empirical attribute and a numerical representation that preserves the relevant structure of that attribute.

The starting point is therefore the identification of what is to be measured. Attributes may be manifest, such as duration, frequency or resource utilization, or latent, such as fatigue, functional ability or treatment burden. Manifest attributes may support linear ratio measurement where equal units and a non-arbitrary zero can be demonstrated. Latent attributes require an explicit measurement model, such as Rasch measurement, with evidence of unidimensionality, invariance and appropriate model fit.. In either case, the properties of the attribute determine the admissible measurement scale. Representational measurement distinguishes nominal, ordinal, interval and ratio scales. These distinctions are fundamental because the arithmetic permitted in quantitative evaluation depends on the measurement properties established for the attribute. Ordering does not establish equal intervals; equal intervals do not establish a non-arbitrary zero; and numerical values cannot legitimately be interpreted as proportions without the requisite ratio properties. Dimensional homogeneity and commensurability impose additional requirements on quantitative operations.

For HTA, this provides the scientific foundation for evaluating therapeutic interventions. Claims must identify the attribute affected, specify its measurement protocol and establish how possession or change will be quantified. Treatment effects can then be expressed as empirically testable differences, rates, proportions or other quantities consistent with the demonstrated scale properties. These claims can be evaluated prospectively, replicated across populations and subjected to falsification.

The significance of representational measurement is that it establishes the legitimacy of quantitative claims before arithmetic, statistical analysis or economic modelling begins. It distinguishes measured treatment effects from numerical scores whose measurement properties remain unproven. For HTA programs, this requires a fundamental change in educational priorities:

students must first understand attributes, measurement scales and admissible operations before learning to construct quantitative evaluations.

Without an identified attribute and a demonstrated measurement structure, numerical manipulation cannot substitute for measurement. The objective of HTA should therefore be to produce reproducible, falsifiable claims about defined attributes of therapeutic benefit and resource use, rather than to assume that the availability of numerical values is sufficient to establish their quantitative meaning.

THE AXIOMS AND REPRESENTATION THEOREMS OF MEASUREMENT

Representational measurement begins with the identification of an attribute whose possession or magnitude is to be measured. Numerical assignments are justified only when they preserve the empirical relationships governing that attribute. The required axioms specify those relationships; representation and uniqueness theorems establish the numerical scale that follows from them.

Volume III of *Foundations of Measurement* provides a systematic treatment of the axiomatic foundations of measurement, the classification of measurement scales and the conditions under which numerical statements are meaningful. Its central concern is the relationship between empirically specified relational structures and their legitimate numerical representations. Axiomatization identifies the formal conditions required for such representations, while representation and uniqueness results establish the existence of numerical scales and the transformations under which their empirical meaning is preserved.

The distinction between nominal, ordinal, interval and ratio scales is therefore not simply a classification of numbers but a statement about the mathematical properties justified by the underlying empirical structure. Ratio scales permit similarity transformations, $x' = ax$ $(a > 0)$, preserving meaningful ratios, whereas interval scales permit positive affine transformations, $x' = ax + b$ $(a > 0)$, under which numerical ratios are generally not invariant. Volume III further examines invariance and meaningfulness, emphasizing that quantitative claims must remain valid under the permissible transformations of their measurement scales. The implication for health technology assessment is fundamental: assigning numerical values to health-state preferences does not, by itself, establish the measurement properties required to interpret those values as ratio measures or to employ them as proportional adjustments.

Before numerical values can be assigned to an attribute and interpreted as measures, the attribute must be explicitly specified and the empirical relationships among its instances established. Representational measurement requires these relationships to satisfy axioms appropriate to the proposed measurement structure. Depending on the attribute and intended scale, these may include ordering, transitivity, consistency, solvability, Archimedean conditions and, where applicable, additive or concatenation relationships. These axioms are not interchangeable requirements imposed uniformly on every attribute; they define the empirical conditions necessary to justify particular numerical representations. Their satisfaction provides the basis for representation and uniqueness theorems, which determine whether measurement is possible, what scale properties are warranted and which arithmetic operations preserve empirical meaning. The essential principle is that the measurement properties of an attribute must be demonstrated before numerical operations are treated as scientifically meaningful**.**

The first requirement is unidimensionality. Measurement of a single attribute requires that the observations represent one identifiable construct rather than an undifferentiated collection of attributes. This is particularly important for latent attributes, where unidimensionality must be investigated empirically through an appropriate measurement model. However, unidimensionality is not a universal axiom of every mathematical representation theorem. Multidimensional preference systems may admit numerical utility representations under specified assumptions. The existence of such a representation does not establish that the underlying health attributes have been individually measured, nor does it establish that the resulting utility score possesses the properties of a validated measure of a single attribute.

The second requirement is ordering. The empirical structure must establish whether objects possess equal or different amounts of the specified attribute. Completeness and transitivity, where required, support consistent comparisons. Ordering may justify ordinal representation but does not establish equal intervals or meaningful ratios.

The third requirement concerns equal differences. An interval representation requires empirical relationships that justify comparisons between differences in attribute possession. Appropriate independence, consistency and cancellation conditions must be satisfied. Additive conjoint measurement provides one axiomatic framework for establishing an additive numerical representation. Numerical response categories and regression coefficients do not themselves demonstrate equal intervals.

The fourth requirement concerns meaningful ratios. Ratio measurement requires a non-arbitrary zero and empirical relationships supporting proportional comparisons. Extensive measurement theory specifies axioms governing concatenation, associativity, monotonicity and Archimedean conditions. Other ratio-scale structures require their corresponding axiomatic foundations.

The representation theorem establishes that, when the specified empirical axioms are satisfied, a numerical function exists that preserves the relevant empirical relationships. It provides the formal bridge between empirical observations and numerical measurement. Its conclusions remain conditional upon satisfaction of its premises.

The uniqueness theorem identifies the permissible transformations of that numerical representation. Ordinal scales admit strictly increasing transformations; interval scales admit positive affine transformations, $x' = a + bx$; and ratio scales admit similarity transformations, $x' = bx$, where $b > 0$. These transformations determine which numerical comparisons and arithmetic operations remain meaningful.

These requirements have direct implications for HTA. Manifest attributes require appropriate empirical foundations for ratio measurement. Latent attributes require validated measurement models, including evidence of unidimensionality, invariance and model fit. Statistical estimation cannot substitute for demonstrating the required measurement properties.

The TTO illustrates the distinction. Its coefficient, t/T, is a legitimate ratio of measured durations. The health states being valued, however, comprise multiple attributes. Although preferences over these multidimensional descriptions may admit a numerical utility representation, that possibility does not establish that TTO responses satisfy the required preference axioms or that the resulting coefficients are valid interval measures of health-state preference. Nor does it establish measurement of the individual attributes represented within those descriptions. This is detailed in Appendix A.

The scientific sequence for attribute measurement is therefore attribute identification, unidimensionality, specification and empirical evaluation of the applicable axioms, representation theorem, uniqueness theorem and admissible arithmetic. This provides the foundation for evaluating therapy impact through lawful measures of manifest and latent attributes rather than relying upon composite numerical valuations whose measurement properties have not been demonstrated.

THE TORRANCE CASE FOR CONDITIONAL INTERVAL UTILITY

The foundations of the Time Trade-Off (TTO) approach were established by Torrance, Thomas and Sackett in 1972, when they proposed a utility-maximization model for evaluating healthcare programs. Their objective was to generate numerical valuations of health outcomes that could support comparisons between interventions differing in their effects on health status and survival duration. The problem was to construct a common evaluative framework in which improvements in health and extensions of life could be represented numerically and incorporated into healthcare decision-making. This required assigning a single utility value to health states, rather than independently measuring the specific attributes affected by therapeutic intervention.

In 1976, Torrance sought to provide a formal theoretical foundation for health-status index models in *Toward a Utility Theory Foundation for Health Status Index Models*. The proposed approach drew upon the von Neumann–Morgenstern expected-utility framework, restating its axioms in terms of health outcomes and introducing additional assumptions appropriate to health-state valuation. The objective was to establish conditions under which individual preferences could be represented by a numerical utility function and used to evaluate healthcare programs. The resulting justification was conditional: the proposed representation depended upon the satisfaction of specified preference axioms and assumptions concerning the relationship between health status, survival duration and individual utility.

The starting point for TTO elicitation was an indifference judgment between two health-duration alternatives. A respondent might regard ten years in an impaired health state as equivalent to six years in full health. The observed coefficient would then be $h = t/T = 6/10 = 0.6$. The theoretical challenge was to explain why this duration-equivalence coefficient could also be interpreted as a numerical valuation of the impaired health state. Under an appropriate health-duration utility representation, the indifference relationship could be expressed as $V(H,10) = V(F,6)$. If the relevant assumptions permitted a multiplicative representation, $V(H,T) = T \times U(H)$, and full-health utility were normalized to one, the resulting coefficient would be $U(H) = 0.6$.

The significance of the conditional interval argument was its attempt to connect observed preference judgments with the formal properties of utility theory. Under the von Neumann–Morgenstern axioms, utility admits an interval representation, unique up to positive affine transformations, $U' = aU + b$, where $a > 0$. Additional assumptions concerning health-duration preferences were required to connect this theoretical representation to TTO elicitation. The resulting utility coefficient was therefore not presented merely as an arbitrary numerical label, but as a value whose interpretation could be justified within a specified decision-theoretic structure.

The axiom- and assumption-driven conditional interval argument imposed significant restrictions on the quantitative claims that could be supported. As detailed by Drummond et al (pp 130-131) a

ratio scale was considered unnecessary for economic evaluation for two reasons [1] . First, an interval scale, as a form of cardinal measurement, was regarded as sufficient to permit the application of parametric statistical operations. Second, because economic evaluations are inherently comparative, the relevant analyses concern differences between utility values rather than ratios of their absolute levels. An interval scale was therefore considered sufficient because it supports meaningful differences and ratios of differences, despite lacking an absolute, non-arbitrary zero. On this reasoning, ratio-scale measurement was not required for comparative economic evaluation.

The appeal to *cardinal measurement* introduces a further ambiguity. Cardinal is commonly used to encompass both interval and ratio scales, but these possess fundamentally different measurement properties. An interval scale permits meaningful differences but lacks a non-arbitrary zero, whereas a ratio scale additionally permits meaningful proportional comparisons between measured values. Describing TTO-derived utility as cardinal therefore fails to establish which measurement properties have been demonstrated. Nor does the designation justify the unrestricted application of parametric statistical operations or proportional adjustments to survival duration. The distinction is critical: the assertion that utility is cardinal cannot substitute for demonstrating that it possesses the ratio-scale properties required for its intended use. Cardinality is not a measurement warrant.

For Torrance and colleagues, the practical attraction was substantial. Conditional utility representations offered a theoretical basis for generating health-state valuations, comparing health outcomes and evaluating healthcare programs within a common numerical framework. The proposed approach connected utility theory, health-status indexing and economic evaluation. Its central claim was that, subject to the required axioms and structural assumptions, numerical health-state utilities could provide a consistent representation of individual preferences and support the evaluation of alternatives involving different combinations of health status and survival duration.

This puts HTA in a measurement straightjacket**.** An interval scale permits meaningful comparisons of differences and ratios of differences, but not proportional comparisons between absolute scale values. Without a non-arbitrary zero, it cannot establish that one health state possesses twice the preference of another or that a utility score of 0.6 represents 60% of full-health preference. Nor do interval-scale properties alone justify using utility levels as proportional adjustments to survival duration. The argument that economic evaluations are necessarily comparative does not remove these restrictions. It establishes the relevance of differences, not the legitimacy of proportional operations on utility levels. Additional assumptions may support a multiplicative decision model, but they do not transform an interval scale into an independently established ratio measure.

The terminology of *cardinal measurement* introduces a further ambiguity. Cardinal is not a separate measurement scale within the conventional representational taxonomy of nominal, ordinal, interval and ratio scales. In economics, cardinal utility generally denotes a representation with meaningful utility differences, commonly unique up to positive affine transformation. Describing an interval scale as cardinal therefore adds no measurement properties beyond those established for the interval representation itself. Nor does the designation authorize every parametric statistical operation or confer ratio-scale properties. The admissibility and meaningfulness of numerical operations must be determined by the demonstrated scale properties

and the invariance of the resulting claims under admissible transformations, not by the general label *cardinal*.

In practical terms, the conditional interval justification for TTO appears to have been largely disregarded in subsequent HTA applications. The coefficient $h = t/T$ has continued to be treated as a health-state utility value without consistent recognition of either the restrictions of an interval representation or its original meaning as a dimensionless ratio of two durations in time space. Its numerical value has been carried directly into utility tariffs, proportional adjustments to survival duration and QALY calculations, as though the ratio properties of the underlying durations had automatically transferred to health-state preference. The distinction between a lawful duration-equivalence ratio and an independently established measure of preference has effectively disappeared. The result is a measurement inversion in which arithmetic proceeds without demonstrating the scale properties or empirical mapping required to justify it.

THE FAILURE OF CONDITIONAL INTERVAL UTILITY AS A MEASUREMENT JUSTIFICATION

The proposed conditional interval interpretation of the Time Trade-Off (TTO) coefficient fails to establish what is required of a measurement procedure. The original coefficient, $h = t/T$, is a lawful dimensionless ratio of survival durations. Its measurement properties derive from time space, where duration is an established ratio attribute. The coefficient satisfies dimensional homogeneity and commensurability because both numerator and denominator are measures of the same attribute. The resulting ratio supports legitimate proportional comparisons of survival durations. The subsequent interpretation of this coefficient as health-state utility introduces a different attribute: preference. The numerical identity of the coefficient does not establish that the two attributes possess equivalent measurement properties. The transition from duration equivalence to preference measurement therefore requires a separate empirical justification.

The critical transition occurs when the duration-equivalence coefficient is reinterpreted, without numerical alteration, as a health-state preference score. A respondent identifies a shorter duration in full health considered equivalent to a longer duration in an impaired health state. The resulting coefficient is then assigned to the described health state. What has been established empirically is an indifference between two health-duration alternatives, expressed numerically through a ratio of measured durations. What has not been established by that calculation is the existence of an independently measured preference attribute possessing the numerical properties subsequently attributed to the coefficient. The distinction is fundamental: a lawful ratio derived from one attribute does not become a lawful measure of another attribute merely because the same numerical value is retained.

Torrance and colleagues sought to justify this transition through decision-theoretic utility theory, including the von Neumann–Morgenstern expected-utility framework, together with assumptions linking health-state preferences and survival duration. Under specified axioms, expected-utility theory establishes an interval representation of preferences, unique up to positive affine transformation, $U' = aU + b$ $(a > 0)$. However, the existence of a representation theorem does not demonstrate that the empirical preferences elicited through TTO satisfy its premises. Moreover, conventional TTO responses concern indifference between health-duration alternatives under

certainty, whereas expected-utility theory concerns preferences over uncertain outcomes. The proposed interval interpretation therefore requires additional assumptions connecting health-state preferences, survival duration and utility. These assumptions must be specified and their empirical applicability demonstrated. Neither the authority of expected-utility theory nor the mathematical existence of representation and uniqueness theorems establishes the measurement status of an observed TTO coefficient.

The decisive step concerns the proposed relationship between health-state utility and survival duration. A multiplicative representation, $V(H,T) = T \times U(H)$, permits the observed indifference between ten years in health state H and six years in full health to be expressed as $10U(H) = 6U(F)$. Normalizing $U(F) = 1$ produces $U(H) = 0.6$. This derivation is mathematically consistent within the specified model. However, the equality between the duration ratio and the utility coefficient follows from the assumed health-duration preference structure, including the proportionality conditions required by the model. Its derivation does not independently demonstrate that those conditions are empirically satisfied. The numerical equality $U(H) = h = 0.6$ is therefore a consequence of the proposed representation, not independent evidence that the coefficient measures preference. What began as a ratio between two measured durations has acquired a preference interpretation through assumptions about the relationship between health status and survival time.

This is the central weakness of conditional interval utility: the theoretical representation is being asked to perform the work of empirical measurement. The model specifies conditions under which a numerical preference representation may exist, but the specification of those conditions is not evidence that they hold. The distinction between mathematical possibility and empirical applicability is fundamental to representational measurement. A conditional theorem establishes the properties of a numerical representation when its required axioms are satisfied. It does not establish that the observed empirical relationships satisfy those axioms. The burden of proof is therefore not discharged by constructing a mathematically coherent utility function or demonstrating that the original duration ratio can be reproduced within that function. The required demonstration concerns the empirical preference structure itself, including the axioms and structural assumptions upon which the numerical interpretation depends.

The interval-scale interpretation introduces a further difficulty. Under representational measurement, an interval scale permits positive affine transformations and therefore does not preserve ratios between utility levels. Such scales support meaningful differences and ratios of differences, but they do not possess the non-arbitrary zero required for meaningful proportional comparisons of levels. Normalization between zero and one does not independently establish that zero represents the absence of the measured preference attribute. Consequently, interpreting $U(H) = 0.6$ as demonstrating that an impaired health state possesses 60% of the preference associated with full health is not justified by interval-scale uniqueness. The fact that the coefficient was originally calculated as a ratio of durations cannot confer ratio-scale properties upon its subsequent preference interpretation.

The paradox of conditional interval utility is therefore particularly revealing. The TTO begins with a coefficient, $h = t/T$, whose ratio properties are legitimate because both quantities are measured durations. The proposed theoretical justification then seeks to interpret this same coefficient as an

interval representation of health-state preference, whose admissible transformations do not generally preserve ratios between utility levels. Yet the coefficient retains its original numerical value and is subsequently employed as a proportional adjustment to survival duration. The ratio properties of the original duration coefficient have effectively survived the transition into preference space without being independently established for the preference attribute. The conditional interval argument does not, by itself, establish the measurement justification for this transition. Although additional structural assumptions may support a mathematically consistent multiplicative health-duration model, their empirical applicability must be demonstrated. Neither the interval representation nor the original ratio properties of duration independently establishes a ratio-scale measure of health-state preference.

An important qualification is necessary. An interval utility representation does not automatically preclude its incorporation into a multiplicative model of health status and survival duration. Under additional structural assumptions, including appropriate normalization and restrictions on preferences over health-duration combinations, such a model may be mathematically consistent. Indeed, a conditional decision-theoretic representation may be useful for comparing alternatives when its assumptions are empirically defensible. The criticism is not that interval utility and a multiplicative decision model are necessarily mathematically incompatible. Rather, the existence of such a model is not equivalent to demonstrating an independently measured ratio-scale preference attribute. Nor does it establish that the numerical relationship between preference and duration constitutes an empirically validated mapping between independently measured attributes. The distinction is between a model-dependent utility coefficient and a lawful measure of a specified preference attribute. The former may be justified conditionally within a decision model; the latter requires the empirical demonstration demanded by representational measurement theory.

This distinction also exposes the limitations of interpreting the TTO coefficient as a proportional adjustment to survival duration. Within time space, the ratio t/T is meaningful because t and T are measures of the same attribute, possessing a non-arbitrary zero and equal units. Their ratio is invariant under admissible similarity transformations and is therefore a lawful proportional measure of duration. Preference space presents a different problem. To establish a preference ratio, $k = p/P$, requires independently defined measures p and P of the same preference attribute, possessing ratio-scale properties, including a non-arbitrary zero, equal units and invariance under admissible similarity transformations. It would then be necessary to demonstrate the proposed correspondence between preference space and time space, $k = h = t/T$. The fact that both ratios would be dimensionless does not establish their numerical or empirical equivalence. Dimensional cancellation is a necessary property of a proportional adjustment but is not sufficient to justify a mapping between different attribute spaces.

The proposed mapping cannot be established simply by observing that the TTO coefficient lies between zero and one or by assigning conventional interpretations to the endpoints. The duration ratio possesses its numerical properties because the underlying durations are independently measured. By contrast, the preference interpretation depends upon the empirical adequacy of a model relating health-state descriptions, preference relationships and survival duration. A numerical coefficient bounded by zero and one is not necessarily a ratio measure. Nor does assigning zero to death and one to full health establish a non-arbitrary zero for a preference attribute. These reference points may define a normalized decision-theoretic representation, but

their designation does not independently demonstrate the ratio-scale properties required for interpreting preference levels as proportions. The distinction between normalization and measurement must therefore be maintained.

The problem is compounded when the coefficient is assigned to multidimensional health-state descriptions such as those generated by EQ-5D-3L. These descriptions combine problem levels across five distinct dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. They identify combinations of reported health problems but do not themselves constitute measures of a single preference attribute. A scalar utility representation of preferences over such alternatives may be mathematically possible under appropriate axioms, but this does not demonstrate that the constituent health attributes form a single empirical dimension or that the resulting coefficient is a validated measure of preference magnitude. Nor does it establish measurement of the individual manifest and latent attributes affected by therapeutic intervention. The construction of regression-based tariffs from TTO valuations extends numerical scores across health-state descriptions but cannot remedy an unestablished measurement foundation.

The conditional interval interpretation must therefore be evaluated against the requirements it claims to satisfy. Has the relevant preference attribute been specified? Have the empirical relationships among preferences been established? Have the axioms supporting an interval representation been demonstrated to hold for the respondents and alternatives under investigation? Have the additional structural assumptions linking health-state preferences and survival duration been independently validated? Has the proposed proportional interpretation of the resulting coefficient been justified? If the coefficient is interpreted as a ratio of preference levels, have the necessary ratio-scale properties and the proposed mapping to duration been demonstrated? These are empirical questions, not matters resolved by the mathematical form of a representation theorem. Without affirmative evidence, the TTO coefficient remains an observed duration-equivalence ratio assigned a model-dependent utility interpretation.

This distinction clarifies the limits of the 1976 conditional interval construction. The mathematical possibility of representing preferences through an interval utility function is not in dispute. What remains unestablished by that construction is whether the observed TTO coefficients possess the claimed measurement properties and whether the empirical conditions supporting their interpretation have been satisfied. A defender of the approach may argue that the additional health-duration assumptions are precisely what justify the numerical coefficient and that independent ratio measurement of preference is not necessary for every decision-theoretic model. That distinction is legitimate. However, it does not remove the obligation to demonstrate the empirical adequacy of the model's preference axioms and structural assumptions. A conditional decision model and an independently validated measurement instrument make different scientific claims. Neither claim is established merely by deriving $U(H) = t/T$.

The consequences for health technology assessment are substantial. The TTO was developed to generate numerical valuations of health-state descriptions for the economic evaluation of healthcare programs. Its subsequent incorporation into EQ-5D tariffs and QALY-based economic evaluation gave these coefficients a central role in claims concerning comparative healthcare value. Yet the statistical processing of preference scores does not establish the measurement properties of the underlying attribute. Regression algorithms, normalization procedures and

increasingly elaborate economic models cannot substitute for demonstrating that the original numerical valuations are lawful measures. The central methodological error is to treat the production of a numerical coefficient as sufficient justification for the arithmetic subsequently applied to it. The construction of a numerical value and the demonstration of its measurement properties are separate scientific tasks.

The implications extend directly to the QALY. The conventional expression $Q = U \times T$ treats health-state utility as a proportional adjustment to survival duration. However, the ratio properties of duration do not establish the ratio properties of the utility coefficient. If U is justified only as a conditional interval representation of preferences, its interpretation as a lawful proportional adjustment requires additional justification. A mathematically coherent health-duration decision model may assign a meaning to the resulting calculation under specified assumptions, but this does not independently establish that the product measures an adjusted quantity of survival time or a specified therapeutic attribute. The measurement status of the inputs, the admissibility of the arithmetic and the empirical meaning of the resulting quantity must each be demonstrated rather than presumed from the existence of a numerical formula.

The failure of conditional interval utility is not that the proposed mathematical representation is impossible, but that its conditional validity has been treated as sufficient justification for measurement. A representation theorem establishes what follows when specified axioms are satisfied; it does not establish that those axioms hold for observed TTO responses. Nor does it demonstrate that a duration-equivalence ratio is an independently established measure of health-state preference. The proposed interval interpretation therefore fails as a measurement justification unless its empirical premises and the required relationship between preference and duration are demonstrated. The mathematical consistency of a utility representation cannot be substituted for evidence that the proposed attribute has been measured. The paradox is that the coefficient begins with lawful ratio properties in time space, acquires a conditional interval interpretation in preference space, and is then employed as though its original proportional meaning had been preserved. That transition is precisely what requires empirical justification.

From the perspective of representational measurement, the TTO is a measurement dead end for evaluating therapeutic benefit. Its legitimate numerical output is a duration-equivalence ratio, not an independently established measure of health-state preference or of changes in the attributes affected by therapy. Further refinement of TTO elicitation procedures, preference models or utility tariffs does not, by itself, resolve this foundational deficiency. The TTO has no place as a measurement method in an HTA curriculum founded on normal science and representational measurement. Its legitimate educational role is historical and methodological: illustrating how a conditional numerical representation came to be treated as an established measure without the requisite empirical demonstration. A reconstructed curriculum must instead begin with the identification of specified manifest and latent attributes, the empirical demonstration of their measurement properties, the determination of admissible arithmetic and the evaluation of falsifiable therapy-impact claims. The scientific requirement is measurement before arithmetic, not the retrospective justification of numerical valuations.

CONCLUSION

The Time Trade-Off (TTO) procedure produces a lawful duration-equivalence ratio, h = t/T, derived from two ratio measures of survival duration. The coefficient satisfies dimensional homogeneity and commensurability within time space. What it does not independently establish is a measure of health-state preference. The subsequent interpretation of this coefficient as a health-state utility introduces measurement claims that cannot be justified by the arithmetic from which it was derived.

Torrance's conditional interval utility argument sought to provide a decision-theoretic foundation for this interpretation. However, the existence of a mathematically consistent utility representation does not establish that its empirical axioms and structural assumptions are satisfied. Nor does an interval representation independently establish the ratio properties required to interpret utility levels as proportional measures of preference. Additional assumptions may support a multiplicative health-duration decision model, but such a model is not equivalent to an independently demonstrated preference measure or an empirically validated mapping between preference and duration. The distinction between conditional mathematical representation and lawful measurement remains unresolved.

The significance for legacy HTA is substantial. As reflected in Drummond and colleagues' *Methods for the Economic Evaluation of Health Care Programmes*, TTO-derived utilities have become central to the numerical valuation of multidimensional health states and their incorporation into economic evaluation. Yet this framework begins with composite health-state valuations rather than the measurement of specified therapeutic attributes. Subsequent regression algorithms, utility tariffs and QALY calculations cannot, by themselves, establish measurement properties absent from the original coefficients. The result is a continuing inversion in which arithmetic and economic modeling take precedence over the demonstration of measurement.

The TTO is therefore a measurement dead end for the evaluation of therapeutic benefit. Its legitimate output is a duration-equivalence ratio, not an independently established measure of preference or therapeutic impact. Although TTO may retain historical and methodological interest as a preference-elicitation procedure, its conditional decision-theoretic interpretation does not justify presenting it as an established measurement method. Further refinement of elicitation procedures or utility tariffs cannot, by itself, overcome this foundational distinction.

A reconstructed HTA framework must begin instead with the identification and lawful measurement of manifest and latent attributes affected by therapeutic intervention. Manifest attributes require appropriate ratio measurement; latent attributes require empirically defensible measurement models demonstrating the relevant scale properties, invariance and unidimensionality. Claims of therapeutic benefit must then be evaluated through reproducible protocols, replication and falsification. Formulary committees and other decision-makers can assess profiles of measured benefits across relevant attributes without requiring their reduction to a composite preference score.

The conclusion is categorical: TTO has no place as a measurement method in an HTA curriculum founded on normal science and representational measurement. Its appropriate educational role is

to illustrate how conditional numerical representations came to be treated as established measures. The necessary transition is not to another health-state valuation technique but to a scientific framework in which attributes are specified, measurement properties demonstrated and therapy-impact claims subjected to empirical testing. The governing principle is measurement before arithmetic.

MEASUREMENT MUST PRECEDE ARITHMETIC.

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

Maimon Research has developed a comprehensive nine-unit distance-learning curriculum comprising approximately 60,000 words of instructional material to support this reconstruction [11]. The program begins with the foundations of representational measurement and the identification of empirical attributes, then develops the distinction between manifest and latent measurement, the properties of measurement scales, dimensional homogeneity, commensurability and admissible arithmetic. It addresses the construction and validation of linear ratio measures for manifest attributes and Rasch-based measurement for latent attributes, including requirements for unidimensionality, invariance and measurement fit. Subsequent units examine the formulation of therapy-impact claims, the measurement of possession and change, treatment compliance and persistence, resource utilization, prospective evaluation protocols, replication and falsification. The objective is not to replace one numerical convention with another, but to establish a coherent scientific framework in which every quantitative claim is supported by an explicitly defined attribute, a validated measure and an admissible mathematical interpretation. The program provides a practical foundation for reconstructing graduate HTA education, professional training and research assessment.

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

APPENDIX A

THE TTO MEASUREMENT PROBLEM: THE ABSENCE OF A DEMONSTRATED MAPPING BETWEEN TIME SPACE AND PREFERENCE SPACE

A.1. The Measurement Question

The fundamental measurement problem in the time trade-off (TTO) is not whether the procedure generates a legitimate numerical ratio. It does. The problem is whether that ratio, derived entirely from survival durations, can legitimately be interpreted as a ratio measure of health-state preference.

The conventional TTO procedure asks a respondent to identify a duration *t* in full health considered equivalent to a longer duration *T* in an impaired health state. The resulting numerical score is:

H = t/T

The score is conventionally interpreted as a health-state utility. However, this interpretation introduces a measurement claim that is not established by the arithmetic of the procedure. The TTO generates a ratio in time space; the proposed utility interpretation assigns that ratio to preference space. These are distinct attribute spaces, and the passage between them requires independent justification.

A.2. The Lawful Ratio in Time Space

Consider a respondent who expresses indifference between six years in full health and ten years in a specified impaired health state. The TTO produces:

H = t/T = 6/10 = 0.6

Both *t* and *T* are ratio measures of the same manifest attribute: survival duration. They possess a non-arbitrary zero, equal units and admissible similarity transformations of the form:

t′ = at and T′ = aT, where a > 0

The quotient is invariant under these transformations:

t′/T′ = at/aT = t/T

The ratio therefore satisfies the requirements of representational measurement for a ratio of durations.

Dimensional homogeneity is satisfied because both quantities belong to the same dimensional class, [T]:

[H] = [T]/[T] = 1

The result is dimensionless.

Commensurability is also satisfied because the numerator and denominator are measures of the same attribute expressed in compatible units. The value 0.6 therefore has an unambiguous interpretation: six years represents 60% of ten years.

This is a lawful statement about survival duration. It does not establish that the impaired health state possesses 60% of the preference, health or therapeutic benefit associated with full health.

A.3. The Proposed Ratio in Preference Space

To interpret the TTO score as a ratio measure of health-state preference, an equivalent measurement relationship would have to be established in preference space.

Suppose p represents the measured preference for an impaired health state and P represents the measured preference for full health. The proposed preference ratio would be:

k = p/P

For the TTO score of 0.6 to represent this preference ratio, the following equality would have to be justified:

k = p/P = 0.6

This requires more than assigning numerical values to expressed preferences. It requires demonstrating that preference constitutes a specified measurable attribute and that p and P are legitimate ratio measures of that attribute.

The necessary conditions include a clearly defined unidimensional attribute, a non-arbitrary zero, equal measurement units, and invariance under admissible ratio-scale transformations. The two quantities must also be commensurable, representing the same attribute in the same measurement system.

If these conditions were satisfied, the preference ratio would be dimensionless:

[k] = [P]/[P] = 1

However, the fact that both H and k would be dimensionless does not establish that they are numerically equivalent. Dimensional cancellation within one attribute space cannot demonstrate equivalence with a ratio constructed from a different attribute.

The TTO procedure does not independently establish the preference measurements p and P. It identifies an indifference between alternatives combining health-state descriptions and survival durations. Consequently, the proposed preference ratio remains an unverified measurement claim.

A.4. The Missing Mapping Between Attribute Spaces

Even if a lawful ratio scale for health-state preference could be established, an additional requirement would remain. It would be necessary to demonstrate a mapping between preference space and time space.

The proposed relationship is:

p/P = t/T

The left-hand side represents a proposed preference ratio. The right-hand side represents an established duration ratio. Although both would be dimensionless, they refer to different attributes and acquire their numerical meaning from different measurement systems.

A defensible mapping would require empirical evidence demonstrating that a particular preference ratio corresponds to a particular proportional adjustment of survival duration. In the present example, it would be necessary to establish that:

p/P = 0.6 corresponds to t/T = 0.6

The TTO does not provide this demonstration. The respondent's indifference between two health-duration alternatives is the observation from which the duration ratio is calculated. It is not an independent measurement of the preference ratio against which the duration ratio can be validated.

Conventional utility theory can derive the equality under specified assumptions about preferences and the separability of health status and duration. However, deriving a numerical relationship from an assumed utility representation is not equivalent to demonstrating that the resulting coefficient is a ratio measure of a preference attribute under representational measurement theory.

The missing mapping is therefore a distinct problem from the absence of a lawful preference measure. Establishing one would not automatically resolve the other.

A.5. Implications for Health-State Utility Tariffs

The measurement problem becomes more consequential when TTO scores are used to estimate health-state utility tariffs.

In systems such as the EQ-5D, respondents evaluate selected multidimensional health-state descriptions, and their elicited numerical scores provide inputs to statistical algorithms. Regression models are then used to estimate numerical values for additional health-state combinations.

These procedures do not independently establish the measurement properties of the underlying preference attribute. Regression may reproduce observed scores, estimate coefficients and generate predictions, but it cannot convert an unestablished preference measure into a lawful ratio measure merely through statistical estimation.

The resulting tariffs inherit the unresolved measurement status of the original TTO valuations. Assigning a tariff value of 0.6 to a health-state description does not establish that the state possesses 60% of the preference or health associated with full health.

The problem is therefore not confined to the original elicitation procedure. It extends to the numerical structure of the health-state utility algorithms derived from it.

A.6. Implications for the QALY

The conventional QALY expression is:

$$\mathbf{Q = U \times T}$$

where U represents a health-state utility score and T represents survival duration.

This expression treats U as a proportional adjustment factor for time. A legitimate duration adjustment can be constructed from two commensurable ratio measures in time space:

$$\mathbf{H = t/T}$$

Multiplying the duration T by this factor gives:

$$\mathbf{H \times T = (t/T) \times T = t}$$

This is lawful arithmetic involving survival durations. The ratio is dimensionless, and the result remains a duration measure.

The difficulty arises when the duration ratio is replaced by a numerical preference score U. For that substitution to be justified, U must have an established measurement interpretation that warrants its use as a proportional adjustment to duration.

The conventional QALY construction does not establish this merely by assigning a tariff value to a health state. It assumes that a numerical preference valuation can perform the same proportional function as a lawful duration ratio.

Even if an independently validated preference ratio existed, an empirically justified mapping to duration would still be required. The equality between preference and duration ratios cannot be inferred from their shared numerical value or dimensionless form.

The QALY therefore inherits both unresolved requirements: the establishment of a lawful preference measure and the demonstration of a mapping that authorizes its use as a proportional adjustment to time.

A.7. Conclusion: A Duration Ratio Is Not a Preference Measure

The TTO produces a mathematically legitimate dimensionless ratio of survival durations. Its arithmetic satisfies dimensional homogeneity, commensurability and the invariance requirements of ratio measurement in time space.

What the procedure does not establish is that the resulting numerical value constitutes a ratio measure of health-state preference.

The distinction can be summarized as follows:

Time space: H = t/T = 0.6 — a lawful duration ratio.

Preference space: k = p/P = 0.6 — a proposed ratio whose measurement properties have not been established.

Mapping: p/P = t/T — an asserted correspondence that requires independent justification.

The fundamental error is to treat the numerical equality as though it established measurement equivalence. A legitimate ratio in one attribute space does not become a legitimate ratio in another simply because the same number is assigned to both.

The TTO is therefore defensible as a procedure for eliciting indifference between health-duration alternatives and calculating a ratio of their durations. It is not, on that basis alone, validated as a procedure for measuring health-state preference on a ratio scale.

The decisive conclusion is that the TTO does not fail as arithmetic. It fails to establish the measurement claim subsequently attached to its arithmetic. The duration ratio is lawful; its reinterpretation as a preference measure remains unproven. This unresolved transition is inherited by health-state utility tariffs and the QALY framework constructed upon them.

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. (2015) Methods for the Economic Evaluation of Health Care Programmes (4th Ed). New York: Oxford University Press

[2] Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680

[3] Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of Measurement, Volume I: Additive and Polynomial Representations*. New York: Academic Press

[4] Suppes, P., Krantz, D. H., Luce, R. D., & Tversky, A. (1989). Foundations of Measurement, Volume II: Geometrical, Threshold, and Probabilistic Representations. New York: Academic Press

[5] Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (1990). Foundations of Measurement, Volume III: Representation, Axiomatization, and Invariance. San Diego: Academic Press

[6] von Neumann, J., & Morgenstern, O. (1944). Theory of Games and Economic Behavior. Princeton, NJ: Princeton University Press

[7] Torrance, G. W., Thomas, W. H., & Sackett, D. L. (1972). A utility maximization model for evaluation of health care programs. *Health Services Research, 7*(2), 118–133

[8] Torrance, G. W. (1976). Toward a utility theory foundation for health status index models. *Health Services Research*, 11(4), 349–369

[9] Bond T, Zi Yan, Heene M. (2021) Applying the Rasch Model: Fundamental measurement in the human sciences (4th Ed). New York: Routledge

[10] Luce, R. D., & Suppes, P. (2002). Representational measurement theory. In H. Pashler & J. Wixted (Eds.), *Stevens' Handbook of Experimental Psychology* (3rd ed., Vol. 4: *Methodology in Experimental Psychology*, pp. 1–41). New York: John Wiley & Sons.

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