MAIMON RESEARCH LLC

RECONSTRUCTION OF HEALTH TECHNOLOGY ASSESSMEMNT



THE TIME TRADE-OFF: A FAILURE OF REPRESENTATIONAL MEASUREMENT

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

University of Minnesota, Minneapolis, MN

LOGIT WORKING PAPER No 2041 NOVEMBER 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

Background: *The Time Trade-Off (TTO) has occupied a foundational position in health technology assessment (HTA) for over five decades. Introduced to generate numerical valuations of multidimensional health-state descriptions, it became central to health-state utility algorithms, quality-adjusted life-year (QALY) calculations and reference-case economic evaluation. Its development prioritized numerical valuation over the scientific measurement of attributes affected by therapeutic intervention.*

Objective: *To demonstrate that TTO fails the requirements of representational measurement, that Torrance's 1976 conditional interval utility argument does not establish a preference measure, and that subsequent economic evaluation methods displaced measurement with arithmetic.*

Methods: *The TTO formulation is examined against Stevens' measurement-scale classification and the axiomatic requirements of representational measurement, including representation and uniqueness conditions, dimensional homogeneity, commensurability and admissible arithmetic. The analysis distinguishes manifest linear ratio measurement from latent Rasch linear measurement and evaluates the progression from Torrance's conditional interval argument through Drummond's treatment of cardinal utility to Dolan's EQ-5D-3L regression algorithm.*

Findings: *The original TTO coefficient, $h = t/T$, is a lawful, dimensionless ratio of two survival durations. Its interpretation as a measure of health-state preference is unsupported because no corresponding latent preference attribute, measurement structure or lawful mapping from preference space to time space has been established. Torrance's conditional interval argument supplies a mathematical representation dependent upon stipulated assumptions, not empirical evidence of preference measurement. Interval-scale utility does not support meaningful ratios between utility levels. Drummond's rejection of the need for ratio measurement, based upon cardinal utility and the comparative character of economic evaluation, disregards the measurement axioms and scale properties governing admissible arithmetic. Dolan's EQ-5D-3L algorithm subsequently treated TTO-derived valuations, including conventionally assigned negative values, as regression observations and generated numerical weights for 243 multidimensional health-state descriptions. Statistical estimation did not establish their measurement properties. The original measurement problem disappeared beneath increasingly elaborate numerical procedures.*

Conclusion: *TTO is a measurement dead end. Its legitimate output is a duration-equivalence ratio, not an established measure of preference or therapeutic benefit. The progression from Torrance through Drummond to Dolan illustrates how arithmetic banished measurement from conventional HTA. Reconstruction requires abandoning health-state valuation as a substitute for scientific measurement and establishing defined manifest and latent attributes, lawful measures and admissible arithmetic before evaluating therapeutic effects. An existing nine-unit educational program provides a practical foundation for this transition.*

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 multidimensional health states rather than independently measure the attributes affected by therapeutic intervention. Its subsequent adoption provided numerical inputs for health-state utility algorithms, including those associated with the EQ-5D descriptive system [1] . These algorithms supply the 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 in Drummond and colleagues' *Methods for the Economic Evaluation of Health Care Programmes* [2] .

The development of representational measurement theory during the twentieth century established the foundation for quantitative evaluation [3 4 5 6]. Scientific measurement requires the identification of specified attributes, empirical demonstration of their measurement properties and construction of numerical representations that preserve the relevant empirical relationships. Representation and uniqueness theorems establish the mathematical properties of measurement scales and determine which numerical operations preserve empirical meaning. This framework received a major systematic formulation with the publication of Volume I of *Foundations of Measurement* by Krantz, Luce, Suppes and Tversky in 1971, building upon earlier developments including Stevens' influential 1946 classification of nominal, ordinal, interval and ratio scales [7]. Volumes II and III, published in 1989 and 1990, further developed the axiomatic foundations, representation theory, invariance and meaningfulness of quantitative statements.

Legacy HTA developed along a fundamentally different trajectory, placing arithmetic before measurement. Rather than establishing the lawful measurement of individual manifest and latent attributes as the foundation for evaluating therapeutic impact, it adopted a framework centered on the numerical valuation of multidimensional health-state descriptions and the economic comparison of healthcare programs. These composite valuations became the basis for calculating aggregate measures of benefit despite their failure to satisfy the axioms and requirements of representational measurement or to establish the scale properties necessary for the arithmetic performed upon them. Arithmetic not only defeated measurement; the contest was never joined**.** The defining objective was to generate numbers for economic comparison, not to identify empirical attributes, demonstrate their measurement structures or establish the admissibility of quantitative operations. The result was a system of numerical analysis constructed without the scientific foundations required to give its calculations empirical meaning.

The purpose of this paper is to demonstrate that TTO fails the required standards of representational measurement. It examines Torrance's attempt to justify the preference interpretation through conditional interval utility theory and establishes why neither the axioms of expected-utility theory nor the ratio properties of duration provide the necessary measurement foundation. The subsequent development of regression-based utility tariffs compounds the original error by assigning further numbers to multidimensional health-state descriptions without demonstrating that they measure a specified attribute. Rather than beginning with attribute definition, lawful measurement and admissible arithmetic, the TTO tradition begins with health-

state descriptions, generates numerical coefficients and retrospectively seeks to justify their interpretation as measures. This inversion of measurement and arithmetic has become embedded in HTA methodology and education. The paper concludes that TTO is a measurement dead end: it produces lawful duration ratios but fails to establish measures of health-state preference or therapeutic benefit. Its continued presentation as a measurement method in HTA curricula is therefore indefensible. Reconstructing HTA requires abandoning the valuation of multidimensional health-state descriptions as a substitute for measurement and returning to the independent measurement of specified manifest and latent attributes.

WHAT CONSTITUTES MEASUREMENT

The fundamental distinction that must be established at the outset is between numerical valuation and scientific measurement. The TTO formulation, as will be made clear in this paper, is not a theory of measurement. As exemplified in Drummond and colleagues' *Methods for the Economic Evaluation of Health Care Programmes*, legacy HTA is a framework for eliciting preferences over multidimensional health-state descriptions and survival durations, assigning numerical valuations to those judgments and incorporating the resulting coefficients into comparative economic evaluation.

Representational measurement has a fundamentally different objective. Measurement is the construction of a numerical representation of a specified empirical attribute that preserves demonstrable relationships among instances of that attribute, subject to established representation and uniqueness conditions. It requires the prior identification of the attribute, the specification and empirical assessment of the relevant measurement axioms, and demonstration of the scale properties that determine admissible numerical comparisons and arithmetic operations. Manifest attributes such as duration may support linear ratio measurement, with equal units and a non-arbitrary zero. Latent attributes require a different measurement structure. Rasch measurement provides a framework for constructing invariant linear logit measures of a defined, unidimensional latent attribute, subject to empirical tests of model fit, invariance and other measurement requirements [8] . Yet conventional TTO-based HTA does not begin with this distinction. Instead, manifest and latent characteristics are combined in multidimensional health-state descriptions and assigned numerical valuations without establishing a common measured attribute. The fundamental difference between manifest ratio measurement and latent Rasch measurement is consequently absent from the foundations of conventional health-state utility valuation, despite the availability of Rasch measurement theory for more than six decades. The result is a quantitative framework in which numbers are assigned before the attributes they supposedly represent have been independently defined and measured.

The TTO formulation, as the subsequent analysis will demonstrate, does not undertake these tasks for a preference attribute. It begins with multidimensional health-state descriptions, elicits indifference judgments involving survival duration and interprets the resulting lawful duration ratio, $\mathbf{h = t/T}$, as health-state utility. The appeal to conditional interval utility theory seeks to justify a numerical preference representation, not to demonstrate that a defined preference attribute has been measured. The distinction is decisive: assigning numbers to preferences over health-state descriptions is not equivalent to measuring an empirical attribute. The central failure of the TTO tradition is therefore its substitution of numerical valuation for measurement, allowing arithmetic

and economic modelling to proceed without first establishing the measurement foundations of the quantities employed.

ATTRIBUTES AND REPRESENTATIONAL MEASUREMENT

The foundation of quantitative evaluation is the measurement of defined attributes. Representational measurement requires numerical values to correspond to empirically demonstrable properties of the attribute being measured. Measurement is not simply the assignment of numbers to observations, classifications, judgments or preferences. It requires an established relationship between an empirical attribute and a numerical representation that preserves the relevant structure of that attribute. The scientific sequence begins by identifying what is to be measured, establishing the empirical relationships that characterize its possession or magnitude, and demonstrating the mathematical properties of the resulting measurement scale. Only then can admissible transformations and legitimate arithmetic operations be determined.

For HTA, the distinction between manifest and latent attributes is fundamental. Manifest attributes include survival duration, event counts, resource quantities and other directly specified empirical quantities. Where the necessary measurement properties are established, these attributes support linear ratio measurement, possessing equal units and a non-arbitrary zero. Their permissible transformations are similarity transformations of the form $\mathbf{x}' = \mathbf{ax}$, where $a > 0$, preserving meaningful ratios between measured quantities. Survival duration illustrates these properties. A duration of ten years is twice a duration of five years, regardless of whether the quantities are expressed in years, months or days. Zero duration represents the absence of elapsed time, and proportional comparisons are meaningful. These properties authorize the calculation of ratios between durations and their interpretation as proportions of the same attribute.

Latent attributes, such as fatigue, functional ability, treatment burden or a suitably defined preference construct, require a different measurement construction. Their possession cannot be measured by simply counting responses or assigning numbers to descriptive categories. The scientific task is to define a single attribute and construct an instrument capable of producing invariant measures of that attribute. Under the Rasch framework, this requires empirical demonstration of unidimensionality, appropriate response-category functioning, item fit, local independence, invariance and the absence of consequential differential item functioning. Where these requirements are satisfied, Rasch measurement produces linear person measures calibrated in logits. The measures are estimated through the response structure in relation to calibrated item locations, rather than by treating counts of affirmative responses or objects possessed as direct quantities of the latent attribute.

The linearity of Rasch measurement is essential. Equal logit differences represent equal differences on the calibrated latent continuum, and a difference of one logit corresponds to an odds ratio of approximately 2.718. However, the logit scale has an arbitrary origin. Ratios between numerical person locations do not represent proportions of the latent attribute, and a person located at +2 logits does not possess twice the attribute of a person located at +1 logit. The distinction is therefore not between linear and nonlinear measurement but between two lawful linear measurement structures with different mathematical properties. Manifest linear ratio measures support meaningful proportional comparisons, while latent Rasch linear measures support meaningful

differences and associated odds-ratio interpretations, not ratios of logit locations interpreted as proportions of attribute possession.

Stevens' classification of nominal, ordinal, interval and ratio scales identifies fundamental differences in measurement properties. Nominal scales classify, ordinal scales establish order, interval scales possess equal intervals but lack a non-arbitrary zero, and ratio scales additionally possess a non-arbitrary zero supporting meaningful proportional comparisons. These distinctions determine which numerical statements remain meaningful under admissible transformations. Dimensional homogeneity and commensurability impose further requirements. Quantitative operations must preserve the dimensions and attribute meanings of the quantities involved, while ratios intended to express proportions must be constructed from measures of the same attribute. The existence of numerical values in different measurement systems does not establish that their arithmetic is interchangeable.

For HTA, these principles require therapeutic claims to begin with the identification of the attribute affected, the specification of its measurement protocol and the empirical demonstration of its measurement properties. Treatment effects may then be expressed through lawful differences, rates, proportions or other admissible quantities appropriate to the established scale. The availability of numbers does not establish measurement, and statistical analysis cannot confer measurement properties that have not been demonstrated. A scientific framework for evaluating therapeutic benefit must therefore place attribute definition and measurement before the arithmetic used to describe, compare or model treatment effects.

THE AXIOMS AND REPRESENTATION THEOREMS OF MEASUREMENT

Representational measurement requires numerical assignments to preserve empirically established relationships. The relevant axioms specify those relationships, while representation and uniqueness theorems determine the numerical structures justified by them. Volume III of *Foundations of Measurement* provides an extensive treatment of axiomatic measurement, scale classification, invariance and meaningfulness. Its central concern is the relationship between empirical relational structures and their legitimate numerical representations. A representation theorem establishes the existence of a numerical function preserving specified empirical relationships when the required axioms are satisfied. A uniqueness theorem determines the transformations under which that representation retains its empirical meaning. Neither theorem independently demonstrates that the empirical premises required for its application have been satisfied.

The requirements vary according to the attribute and proposed measurement structure. They may include ordering, transitivity, consistency, solvability, Archimedean conditions, cancellation or concatenation relationships. These are not interchangeable conditions imposed uniformly on every measurement problem. They define the empirical requirements appropriate to particular representations. For interval measurement, permissible transformations take the form $\mathbf{x' = ax + b}$, where $a > 0$. For ratio measurement, permissible transformations take the form $\mathbf{x' = ax}$, where $a > 0$. Under interval transformations, numerical differences are preserved up to a common positive multiplier, but ratios between scale locations are generally not invariant. Under ratio

transformations, ratios between measured quantities remain invariant. These distinctions establish the quantitative claims that can legitimately be made from each measurement structure.

The implications for health-state utility are immediate. A numerical representation of preferences over alternatives may exist under specified decision-theoretic axioms, but such a representation does not automatically constitute measurement of a single latent attribute. Nor does it establish that the resulting numerical values possess the properties required for proportional comparisons. The distinction between a scalar representation of preferences over multidimensional alternatives and the measurement of a defined attribute must be maintained. The former may assign numbers consistently with an assumed preference structure; the latter requires an empirical measurement instrument whose properties have been demonstrated.

TIME TRADE OFF: THE FOUNDATION OF NON-MEASUREMENT

The foundations of the Time Trade-Off (TTO) were established by Torrance, Thomas and Sackett in 1972 through a utility-maximization approach to 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 central problem was economic evaluation. A common numerical framework was sought in which different combinations of health status and survival could be compared and incorporated into healthcare decision-making. The approach did not begin with the independent measurement of therapeutic attributes but with preferences over health-duration alternatives. Its significance was the attempt to convert judgments concerning multidimensional health outcomes into numerical coefficients suitable for comparative economic analysis.

The fundamental error in the Torrance et al. TTO technique was its decision to begin with the valuation of a health-state description rather than the measurement of specific attributes relevant to a therapeutic intervention. A health-state description is a composite representation of multiple characteristics of health. It is not itself a single empirical attribute possessing established measurement properties. In the TTO procedure, a respondent, initially guided through the valuation exercise by an interviewer, is asked to identify the duration in perfect health that would be considered equivalent to a specified longer duration in the described health state. If T represents the duration in the described health state and t the shorter duration in perfect health, the respondent identifies an indifference point between the two alternatives. The resulting numerical coefficient is expressed as $\mathbf{h = t/T}$.

Applying the standards of representational measurement, including Stevens' distinction between nominal, ordinal, interval and ratio scales, establishes precisely what this coefficient represents. Both t and T are measures of survival duration on a linear ratio scale, possessing equal units, a non-arbitrary zero and admissible similarity transformations. Their quotient, $\mathbf{h = t/T}$, is therefore a lawful, dimensionless ratio within time space. It satisfies dimensional homogeneity because both quantities have the same time dimension, [T], which cancels in division. It also satisfies commensurability because the numerator and denominator are measures of the same attribute: duration. For example, if a respondent expresses indifference between ten years in a described health state and six years in perfect health, the resulting coefficient is $h = 6/10 = 0.6$. This

establishes a ratio between two durations. It does not establish that the respondent's preference for the described health state is 60% of a measured preference for perfect health.

The decisive error occurs when this lawful duration ratio is transferred from time space into an assumed preference space without demonstrating that a corresponding preference attribute has been measured. The indifference judgment establishes a relationship between two health-duration alternatives, but it does not establish a numerical representation of a single latent preference attribute. No empirical preference scale has been constructed, no non-arbitrary zero for that attribute has been demonstrated, and no representation or uniqueness theorem has established that the numerical coefficient possesses the properties required of a preference measure. The ratio properties of t and T cannot simply be inherited by an alleged preference quantity. The fact that h is dimensionless does not resolve this problem: a dimensionless number derived from duration remains a duration ratio unless an independently justified measurement interpretation is established.

This distinction exposes the difference between a theory of numerical valuation and a theory of measurement. As exemplified by the subsequent presentation of TTO in Drummond and colleagues' *Methods for the Economic Evaluation of Health Care Programmes*, the objective is to elicit preferences, generate utility coefficients and use those coefficients in comparative economic evaluation. Representational measurement requires something fundamentally different. Measurement is the construction of a numerical representation of a specified empirical attribute that preserves demonstrable relationships among instances of that attribute, subject to established representation and uniqueness conditions. The attribute must first be defined, its relevant empirical structure identified, and the axioms necessary for its proposed scale assessed. Only then can the admissible numerical transformations and arithmetic operations be determined. Manifest attributes such as duration may support linear ratio measurement, while appropriately defined latent attributes may support invariant linear measurement through Rasch calibration. Neither possibility authorizes the conversion of an observed duration ratio into a preference measure merely because the respondent has expressed indifference between two alternatives.

The significance of the original TTO formulation is therefore not that it discovered a method for measuring health-state preference, but that it introduced a procedure for generating numerical valuations from choices involving multidimensional health-state descriptions and survival durations. The measurement question was bypassed at the outset**.** Instead of asking which specific attributes of therapeutic benefit were to be measured and what measurement properties they possessed, the procedure assumed that a numerical coefficient derived from a duration trade-off could represent the value of a health state. Torrance's subsequent conditional interval utility argument sought to provide a theoretical justification for this interpretation.

THE CONDITIONAL INTERVAL SCORE

The foundations of the Time Trade-Off (TTO) were established by Torrance, Thomas and Sackett in 1972 through a utility-maximization approach to evaluating healthcare programs. Their objective was to generate numerical valuations of health outcomes for comparing interventions with different effects on health status and survival duration. The starting point was not the measurement of attributes affected by therapeutic intervention but the valuation of

multidimensional health-state descriptions. This was the foundational error. Scientific measurement requires the prior definition of an empirical attribute, demonstration of its measurement properties and determination of admissible arithmetic. A multidimensional health-state description is not itself a single measured attribute.

The TTO asks respondents to identify a shorter duration, t, in perfect health considered equivalent to a longer duration, T, in a specified impaired health state. The resulting coefficient, **h = t/T**, is a lawful ratio of survival durations. Under Stevens' classification, both t and T are linear ratio measures possessing equal units and a non-arbitrary zero. Their quotient satisfies dimensional homogeneity and commensurability because both quantities measure duration. If ten years in an impaired health state are considered equivalent to six years in perfect health, h = 6/10 = 0.6. This is a dimensionless ratio in time space, not evidence that the health state possesses 60% of a measured preference for perfect health. The measurement error occurs when this duration ratio is reinterpreted as a preference measure.

In 1976, Torrance attempted to justify this interpretation through conditional interval utility theory, drawing upon expected-utility theory and additional assumptions concerning health-state preferences and survival duration. Under a stipulated multiplicative health-duration model, the indifference between ten years in an impaired state and six years in perfect health produces a utility coefficient of 0.6. However, this result follows from the assumed preference structure, not from independent measurement of a latent preference attribute. Moreover, an interval representation does not establish meaningful ratios between utility levels. Torrance's conditional argument therefore failed to demonstrate the measurement properties required to interpret h = t/T as a proportion of preference.

The subsequent treatment of utility in Drummond and colleagues' *Methods for the Economic Evaluation of Health Care Programmes* illustrates a further departure from measurement principles. Their argument that ratio-scale measurement is unnecessary because economic evaluations are comparative and cardinal utility permits numerical analysis effectively dismisses the original measurement problem. Cardinal is not a separate scale within Stevens' classification of nominal, ordinal, interval and ratio measures. An interval representation permits meaningful differences but not ratios of absolute levels. Neither the comparative purpose of economic evaluation nor the designation of utility as cardinal establishes that utility coefficients represent proportions of preference or can lawfully adjust survival duration. The issue is not whether numbers can be compared or manipulated, but whether the empirical axioms and measurement properties required for those operations have been demonstrated. Drummond's argument is emblematic of a failure to recognize that arithmetic must follow measurement, not substitute for it.

THE DEPENDENT VARIABLE TO ESTIMATE THE EQ-5D-3L ALGORITHM

The subsequent development of regression-based utility algorithms completed the transition from conditional measurement justification to numerical estimation. Dolan's 1997 UK EQ-5D-3L algorithm provides the clearest example. The EQ-5D-3L describes health states through five dimensions—mobility, self-care, usual activities, pain/discomfort and anxiety/depression—each

with three ordered levels, producing 243 possible descriptions. These are multidimensional categorical classifications, not measures of a single empirical attribute.

Respondents valued a selected subset of these descriptions using TTO. For states judged better than death, the observations were based on the duration ratio **h = t/T**. States judged worse than death were assigned negative values through a modified TTO procedure and scoring convention; negative values cannot arise from the original non-negative duration ratio. Both sets of numerical valuations were nevertheless treated as observations of a common dependent variable in regression models. The explanatory variables were categorical indicators of the EQ-5D-3L dimension levels, supplemented by terms representing severity patterns.

The regression algorithm estimated coefficients for the dimension levels and applied numerical decrements to the full-health reference value, generating weights for all 243 health-state descriptions. The dependent variable was therefore a collection of TTO-derived numerical valuations, not an independently established measure of preference. Regression estimated relationships between categorical health-state descriptions and assigned numbers; it did not demonstrate the measurement properties of those numbers. Nor did combining positive duration ratios with conventionally assigned negative values establish a common latent preference scale.

The historical sequence is revealing. Torrance attempted in 1976 to justify the duration coefficient through conditional interval utility theory. Drummond and colleagues subsequently dismissed the need for ratio measurement by appealing to cardinal utility and the comparative character of economic evaluation, without addressing the empirical axioms and evidence required to establish lawful measures. Dolan's EQ-5D-3L regression algorithm then accepted TTO-derived valuations as utility observations and extended them across the 243 health-state descriptions. At no stage was a single latent preference attribute independently measured, its measurement properties demonstrated or a lawful mapping from preference space to time space established.

The measurement question had effectively disappeared. As will be demonstrated in a subsequent paper, the EQ-5D-3L valuation algorithm was essentially an exercise in statistical estimation and arithmetic, with no evident recognition of, or concern for, the requirements of representational measurement. Its objective was to generate numerical weights from categorical health-state descriptions, not to establish lawful measures of specified empirical attributes. The resulting coefficients were accepted as utilities and subsequently employed in QALY calculations without demonstrating the measurement properties required for their interpretation or use. The progression from Torrance through Drummond to Dolan thus marks the displacement of scientific measurement by numerical modelling. What remained was arithmetic without measurement.

CONCLUSION: TTO ARITHMETIC BANISHES MEASUREMENT

The Time Trade-Off represents a fundamental failure of scientific measurement in health technology assessment. Its origins lay in the economic objective of assigning numerical values to multidimensional health-state descriptions rather than identifying and measuring the specific manifest and latent attributes affected by therapeutic intervention. The original TTO coefficient, h = t/T, is a lawful ratio of survival durations. Its interpretation as a measure of health-state preference, however, requires an independent measurement justification that the procedure does

not provide. The foundational error was to assume that a number derived from an indifference judgment involving duration could acquire the measurement properties of a different, undefined attribute.

Torrance's 1976 conditional interval utility argument did not resolve this problem. It sought to justify a numerical preference representation through expected-utility theory and additional assumptions concerning health states and survival duration. Yet a conditional mathematical representation is not evidence that an empirical attribute has been measured. Even if the stipulated conditions are accepted, an interval utility representation does not establish meaningful ratios between utility levels or demonstrate that a coefficient of 0.6 represents 60% of a measured quantity of preference. The original duration ratio remained intact; its interpretation as preference remained unsupported by the required measurement evidence.

The subsequent treatment of TTO in Drummond and colleagues' economic evaluation framework marked a more consequential departure. Their dismissal of the need for ratio measurement, on the grounds that economic evaluations are comparative and that cardinal utility permits numerical analysis, exemplifies a failure to recognize the foundational role of measurement axioms, scale properties and empirical evidence. The distinction between interval and ratio measurement is not optional, nor can it be set aside to accommodate preferred analytical procedures. The arithmetic must be justified by the properties of the measure; the intended arithmetic cannot determine those properties. By reversing this requirement, conventional economic evaluation effectively released utility coefficients from the scientific obligation to demonstrate what they measured.

The development of the EQ-5D-3L regression algorithm completed this transition. TTO-derived numbers were accepted as dependent-variable observations, including positive duration-equivalence coefficients and negative values introduced through an alternative scoring convention. Regression models then generated weights for the 243 multidimensional health-state descriptions without establishing a single measured preference attribute. As will be demonstrated in a subsequent paper, this was an exercise in statistical estimation and arithmetic rather than representational measurement. The resulting numbers acquired authority through their incorporation into economic evaluation, not through empirical demonstration of their measurement properties.

The consequence was the banishment of measurement from the quantitative foundations of conventional HTA. What began as an attempt to justify a numerical utility coefficient became a system in which assigned values, regression estimates, QALYs and cost-effectiveness ratios could be generated without first establishing lawful measures. The absence of a defined attribute, tested measurement axioms and demonstrated scale properties ceased to be treated as barriers to quantitative analysis. Numerical production became its own justification, while increasingly elaborate models concealed rather than corrected the original measurement failure.

The TTO tradition is therefore a measurement dead end. Its legitimate numerical output is a duration-equivalence ratio, not an established measure of health-state preference or therapeutic benefit. Refinements to elicitation procedures, utility algorithms or economic models cannot repair a failure that occurs before measurement begins. The scientific reconstruction of HTA requires abandoning the valuation of multidimensional health-state descriptions as a substitute for

measurement and restoring the necessary sequence: define the attribute, establish the measure, demonstrate its properties and only then undertake admissible arithmetic. Until that sequence is restored, conventional HTA will remain committed to the production of numbers whose empirical meaning has never been established. Arithmetic will continue to displace measurement rather than serve it.

MEASUREMENT MUST PRECEDE ARITHMETIC

The implications extend beyond the TTO procedure. The continued reliance upon multidimensional health-state valuations reflects a wider inversion within the knowledge base of conventional HTA. Numerical methods have been developed and disseminated without consistently requiring prior demonstration of the measurement properties of their inputs. Reconstructing HTA requires a return to the foundations of representational measurement, beginning with specified empirical attributes, lawful measures and admissible quantitative operations. The objective is to evaluate therapeutic interventions through measurable changes in manifest and latent attributes rather than reducing multidimensional treatment effects to algorithmically generated numbers. This requires a fundamental change in the knowledge base transmitted to faculty, researchers and students, replacing the inherited emphasis on numerical preference modelling with the scientific requirements of attribute measurement.

Maimon Research has developed a comprehensive nine-unit distance-learning curriculum comprising approximately 60,000 words of instructional material to support this reconstruction [9]. The program begins with the foundations of representational measurement and the identification of empirical attributes, then develops the distinction between manifest and latent measurement, scale properties, dimensional homogeneity, commensurability and admissible arithmetic. For manifest attributes, it addresses the construction and validation of linear ratio measures. For latent attributes, it introduces Rasch measurement, including unidimensionality, item calibration, ordered response thresholds, invariance, model fit and the interpretation of linear logit measures. 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 a defined attribute, a validated measure and an admissible mathematical interpretation.

The curriculum is available through Maimon Research as a structured distance-learning course, with opportunities for questions, clarification, faculty seminars and guided discussions concerning the application of representational measurement to research and teaching. It provides a practical foundation for individual professional development, faculty preparation, graduate education and institutional curriculum reconstruction. Its significance is that the transition to lawful measurement does not require waiting for established HTA organizations to revise their methodological guidance or for another generation of debate over utility algorithms. The educational foundation already exists, together with a practical route for applying it to the evaluation of therapeutic benefit.

Universities, professional organizations and HTA institutions must now decide whether to continue transmitting a framework in which numerical valuation precedes measurement or to

reconstruct their curricula around the lawful measurement of empirical attributes. The issue is not whether conventional utility algorithms can be improved, but whether the scientific foundations of quantitative evaluation should be established before arithmetic is undertaken. Measurement must precede arithmetic.

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] 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

[2] 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

[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] 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

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

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

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