MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

THE MEASUREMENT FUTURE IN HEALTH TECHNOLGY ASSESSMENT

LEGACY HTA: ISPOR – A FORENSIC EVALUATION OF REPRESENTATIONAL MEASUREMENT AND THE IMPOSSIBLE QALY

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

LOGIT WORKING PAPER No 2005 SEPTEMBER 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

Background: The quality-adjusted life year (QALY) has become the dominant outcome construct in health technology assessment (HTA), supporting cost-utility analysis, incremental cost-effectiveness ratios, reference cases and reimbursement decisions. This paper examines whether the QALY satisfies the requirements of representational measurement necessary to give its defining arithmetic empirical meaning.

Methods: The QALY is assessed against the historical and formal requirements governing quantitative measurement, from magnitude and proportion through dimensional homogeneity, scale theory and admissible transformations. The analysis proceeds through the required sequence: attribute → measurement → scale properties → dimensional structure → quantitative relationship → admissible arithmetic → derived quantity. Particular attention is given to linear ratio measurement, dimensionless proportional factors and the measurement properties required before a health-state value can legitimately multiply duration.

Results: The QALY fails at multiple prior measurement gates. If (Q=UT) is to retain the dimension of time, (U) must function as a dimensionless proportional factor. A genuine proportion requires an identifiable unidimensional attribute, lawful ratio measurement, a non-arbitrary zero, a defined reference quantity and invariant proportional interpretation. A value such as (U=0.5) must therefore represent one-half of a specified empirical quantity, not merely occupy a numerical position between assigned anchors. Health-state utilities provide no demonstrated structure of this kind. Their explicitly multiattribute construction describes or values combinations of distinct health attributes but does not demonstrate that these constitute a single measured attribute possessing equal units, a non-arbitrary zero and meaningful ratios. Additive or multiplicative preference algorithms cannot create these properties. Even if a lawful dimensionless proportional health measure existed, a second requirement would remain: an independently demonstrated empirical relationship establishing why proportional possession of that attribute should multiply elapsed duration and what derived empirical quantity (UT) represents.

Conclusions: The QALY is not an imperfect measure requiring improved elicitation, calibration or modelling. It is an impossible quantitative construct whose defining multiplication lacks the required measurement foundation. Subsequent subtraction, aggregation, cost-per-QALY calculation, modelling and uncertainty analysis cannot repair failure at the measurement gate. The global HTA infrastructure constructed around utilities and QALYs therefore rests upon inadmissible arithmetic. The appropriate scientific response is abandonment of the QALY and reconstruction of HTA from representational-measurement first principles.

INTRODUCTION

The purpose of this paper is to provide the definitive case for the impossible quality-adjusted life year (QALY). The standards of representational measurement provide the required analytical framework. Presented as the adjustment of time by a multiattribute index intended to represent quality of life, derived from a multiattribute health-state description, the QALY is mathematically and scientifically indefensible. The problem is not simply that existing utility instruments have undesirable psychometric properties or that alternative elicitation techniques might generate different values. The required quantitative structure does not exist.

The case, however, goes beyond demonstrating this impossibility. It is necessary to identify the properties that a numerical value must possess before it can legitimately operate as a dimensionless proportional adjustment to time. This requires examination of ratio measurement, admissible transformations, dimensional homogeneity, the construction of dimensionless proportional factors and, critically, the empirical relationship required between quantities before multiplication can produce a meaningful derived quantity. A number bounded between zero and one does not acquire these properties merely because it has been labelled a utility or preference weight.

The QALY must therefore be evaluated from the beginning rather than from within the methodological conventions that have grown around it. The relevant sequence is attribute → measurement → scale properties → dimensional structure → quantitative relationship → admissible arithmetic → derived quantity. If the QALY fails at any required stage, subsequent modelling cannot rescue it.

This assessment has a wider purpose. For decades ISPOR, academic research centers, HTA agencies and associated professional communities have promoted an international analytical framework in which the QALY occupies a central position. If the construct fails the elementary requirements of representational measurement that should have preceded its adoption, the issue is no longer how the QALY might be improved, recalibrated or defended. It is how an impossible quantitative construct became the foundation of a global HTA enterprise and what follows once that failure is recognized. The conclusion developed here is that there is no scientifically defensible route to rehabilitation. Once the required standards of representational measurement are applied, the QALY must be abandoned.

1. FROM EUCLID TO STEVENS: QUANTITIES BEFORE ARITHMETIC

The requirement that arithmetic respect the properties of the quantities upon which it operates is not a recent refinement of measurement theory. Its intellectual origins extend back more than two millennia. From the Greek theory of magnitude and proportion, through Newtonian mathematical physics and Fourier's dimensional analysis, to Helmholtz and twentieth-century measurement theory, quantitative science progressively clarified a fundamental distinction: the mathematical possibility of performing an arithmetic operation does not establish the empirical legitimacy of that operation. Numbers represent quantities only under specified conditions, and the properties of those quantities constrain the relationships that may meaningfully be expressed numerically.

Euclid provides the appropriate starting point. Book V of the *Elements*, drawing upon the theory of proportion associated with Eudoxus, treated ratios as relations between magnitudes rather than simply as quotients of numbers. Greek mathematics maintained an important distinction between number and magnitude. Length, area, volume and other magnitudes were not interchangeable collections of numerical values. A ratio expressed a relation between appropriate magnitudes, and proportional reasoning depended upon understanding what those magnitudes represented. The theory was sufficiently sophisticated to accommodate incommensurable magnitudes without pretending that measurement consisted simply of attaching convenient numbers to objects.

This distinction is fundamental. Modern arithmetic makes division computationally trivial: any two numerical values can be entered into a calculator and a quotient returned. Euclidean proportional reasoning imposed the logically prior question: what are the magnitudes whose relationship is being expressed? A numerical quotient was not self-validating. The quantitative relationship depended upon the nature of its terms.

Newton carried this discipline into mathematical physics. The *Principia* of 1687 begins with definitions of quantities and proceeds extensively through ratios and proportions. Newton did not work within modern representational measurement theory, nor did he possess the later formal language of dimensional analysis. Nevertheless, his mathematics did not treat empirical quantities as naked numbers. Mass, motion, time, distance and force possessed defined physical interpretations, and the mathematical relationships among them derived their significance from those interpretations.

Consider the familiar relationship between distance and time. Their quotient can represent velocity:

$v = d/t$.

This is not legitimate merely because d and t can each be represented numerically and division can therefore be performed. Distance and duration are measurable quantities, and their relationship defines another empirically interpretable quantity: distance traversed per unit time. The numerator and denominator retain their identities. Their division creates a derived quantity with a specified meaning.

This reveals an important principle sometimes lost when empirical quantities are reduced to columns of numbers. Even lawful measures cannot be combined arbitrarily simply because their scale properties permit calculation. The quantities must stand in a scientifically meaningful relationship. Dividing a measured body mass by an unrelated measured duration produces a numerical quotient, but unless the relationship represents something such as mass flow per unit time, the calculation does not thereby create a scientifically meaningful quantity. Arithmetic capability is not sufficient. The relationship among the quantities must itself possess empirical interpretation.

Fourier made another part of this discipline explicit. In his 1822 *Théorie analytique de la chaleur (The Analytical Theory of Heat)* dimensional homogeneity became a general requirement for physical equations. Quantitative relationships had to remain coherent when the units used to represent their component quantities were changed. A physical equation could not owe its apparent truth to an arbitrary choice of meters rather than feet, or seconds rather than minutes. The dimensions of the quantities constrained the equations in which they could legitimately participate.

This did not prohibit multiplication or division between different kinds of quantities. On the contrary, such operations generate many of the derived quantities fundamental to science. Distance divided by time gives velocity; mass multiplied by acceleration gives force. The requirement is deeper: the operation must preserve a coherent relationship among identified quantities, and its result must have an interpretation determined by those quantities rather than by numerical manipulation alone.

Helmholtz moved the question from dimensional consistency toward the foundations of measurement itself. In *Counting and Measuring* in 1887, he addressed the conditions under which relations among empirical objects could legitimately be represented by relations among numbers. This reverses the naïve conception that assigning numbers creates measurement. The empirical relational structure comes first. Numerical representation is justified only insofar as it preserves that structure. The central question becomes not "Can numbers be assigned?" but "What properties of the empirical attribute justify this numerical representation and the operations subsequently performed upon it?"

The development culminated, for present purposes, in twentieth-century scale theory. Stevens' 1946 classification of nominal, ordinal, interval and ratio scales made explicit that numerical representations differ in the transformations they permit and therefore in the quantitative statements that remain meaningful [1]. Nominal values classify. Ordinal values establish order. Interval scales possess equal intervals but permit translation of the origin:

$x' = ax + b, a > 0.$

Ratio scales possess a non-arbitrary zero and permit similarity transformations:

$x' = ax, a > 0.$

The distinction is decisive for proportional reasoning. If the origin may legitimately be shifted, ratios of numerical scale values are not invariant. A value of 20 is not meaningfully "twice" a value of 10 merely because 20/10 = 2. Under an admissible interval transformation those same empirical states might be represented by 30 and 20, and the ratio becomes 1.5. The arithmetic remains executable; the supposed empirical meaning disappears.

Stevens therefore formalized an insight whose ancestry was far older than his terminology. Arithmetic performed upon numerical representations is scientifically meaningful only when the empirical quantities, measurement structure and admissible transformations support the operation and its interpretation. Ratio measurement does not arise because investigators wish to multiply or divide. Its properties must be established independently before proportional arithmetic is interpreted empirically.

Euclid, Newton, Fourier, Helmholtz and Stevens should not be represented as advancing a single theory. They did not. What connects them is more fundamental: each belongs to the historical development through which quantitative science progressively constrained numerical operations by what the numbers represent. Magnitudes precede their ratios; physical quantities precede equations relating them; dimensional structure constrains quantitative relationships; empirical relations precede their numerical representation; and scale properties determine which numerical statements survive permissible changes of representation. The resulting principle is straightforward: quantities before arithmetic**.**

By the middle of the twentieth century, the scientific problem was no longer whether arithmetic could be performed upon assigned numerical values. Representational measurement asked whether the attributes represented by those values possess the properties required to give the proposed arithmetic invariant empirical meaning [2]. Any later quantitative framework built without imposing that requirement at its foundation was not waiting for measurement science to be invented. It was proceeding without a discipline whose foundations had been developing for more than two thousand years. This aptly describes legacy health technology assessment (HTA) with its readily accepted global distribution of the QALY over the past 40 years; but you can't put the clock back 2,000 years.

2. THE LINEAR RATIO SCALE

A linear ratio scale provides the strongest familiar form of measurement for manifest quantitative attributes. It does considerably more than assign numbers to observations. It establishes a numerical representation in which order, differences and ratios correspond to empirical relations in the attribute being measured. This is what permits statements such as one duration being twice another, one distance being half another, or one mass being three times another to have invariant quantitative meaning.

The starting point is always the attribute. A ratio scale cannot be created merely by declaring a set of numerical values to be ratio scaled. There must first be an identifiable empirical attribute whose structure supports the proposed numerical representation. Examples include duration, distance, mass and volume under appropriate measurement conditions. The scale represents possession of that attribute, not simply rankings, preferences or numerical labels assigned to observations.

A linear ratio scale has a number of defining properties:

- **A defined attribute.** The object of measurement must be explicitly identified. The numerical values must represent possession or magnitude of that attribute rather than an unspecified collection of characteristics.
- **Unidimensionality.** A single scale represents a single attribute. Movement along the scale must correspond to change in that attribute rather than an unknown combination of changes in several different attributes.
- **Order.** If ($x_A>x_B$), object (A) possesses more of the measured attribute than object (B). Numerical ordering therefore represents empirical ordering.
- **Equal units.** A unit has the same quantitative meaning throughout the scale. The difference represented by one centimetre does not change according to where the centimetre occurs on a ruler; one second represents the same duration wherever it occurs on a duration scale.
- **Additivity.** Appropriate quantities can be concatenated or combined so that their numerical representation corresponds to addition. Two adjacent lengths of three metres and two metres produce a total length of five metres. This additive structure provides the equal-unit foundation of extensive linear measurement.

- **A non-arbitrary or absolute zero.** Zero represents absence of the measured quantity in the relevant sense and is fixed by the empirical structure rather than selected for convenience. Zero duration represents no elapsed duration; zero length represents no extension along the dimension being measured. The origin cannot legitimately be translated while preserving the same ratio-scale representation.
- **Meaningful differences.** Because units are equal, differences between scale values have quantitative meaning. The difference between 10 and 15 units represents the same amount of the attribute as the difference between 20 and 25 units.
- **Meaningful ratios.** Because both equal units and a non-arbitrary zero are present, ratios of scale values possess empirical meaning. If one duration is ten seconds and another is five seconds, the first duration is twice the second. The statement is about the quantities themselves, not merely about their numerical labels.
- **Invariance of ratios under permissible changes of unit.** Changing the measurement unit does not change proportional relationships. Ten metres and five metres become 1,000 centimetres and 500 centimetres; their ratio remains two.

These properties determine the transformations that are admissible for a linear ratio scale. If (x) is one legitimate numerical representation of an attribute, another legitimate representation (x') may be obtained by changing the unit:

Admissible similarity transformation for a linear ratio scale

$x' = ax, a > 0$

where x is the original scale value, x′ is the transformed value, and a is a positive constant representing a change in the unit of measurement.

Example of a ratio

$x_1 = 10$ and $x_2 = 5$

$x_1/x_2 = 10/5 = 2$

Under an admissible change of unit:

$x_1' = ax_1$

$x_2' = ax_2$

Therefore:

$x_1'/x_2' = ax_1/ax_2 = x_1/x_2 = 2$

The ratio is invariant under every admissible positive similarity transformation.

Impermissible translation of the origin

For a ratio scale, the following transformation is not admissible:

$x' = ax + b, b \neq 0$

For example, if:

$x_1 = 10$ and $x_2 = 5$

then:

$x_1/x_2 = 2$

If an additive constant of 10 were permitted:

$x_1' = 10 + 10 = 20$

$x_2' = 5 + 10 = 15$

and:

$x_1'/x_2' = 20/15 = 4/3$

The original ratio of 2 is not preserved. The proportional statement therefore depends upon the choice of origin and cannot be an invariant empirical statement.

Contrast with an interval scale

The admissible transformation for an interval scale is:

$x' = ax + b, a > 0$

Differences are preserved up to the change of unit because:

$x_1' - x_2' = (ax_1 + b) - (ax_2 + b)$

$= a(x_1 - x_2)$

The additive constant cancels.

Ratios do not have this property:

$x_1'/x_2' = (ax_1 + b)/(ax_2 + b)$

which is not, in general, equal to:

x_1/x_2

Hence interval-scale values support meaningful comparisons of differences but not meaningful ratios of the scale values.

Ratio-scale multiplication

If X and Y are independently established ratio quantities with admissible transformations:

$X' = aX$

$Y' = bY$

then their product $Z = XY$ transforms as:

$Z' = X'Y'$

$= (aX)(bY)$

$= abXY$

$= abZ$

The product therefore changes only according to the changes in the units of X and Y. There is no contamination from an arbitrary origin.

If X were instead an interval-scale quantity:

$X' = aX + c$

while Y remained a ratio quantity:

$Y' = bY$

then:

$Z' = X'Y'$

$= (aX + c)(bY)$

$= abXY + bcY$

$= abZ + bcY$

The additional term bcY depends upon the arbitrary origin c. The product therefore does not possess the invariant proportional interpretation obtained when both component measures have the required ratio structure.

This establishes an important distinction. Demonstrating the ratio properties of X and Y establishes the measurement conditions required for proportional arithmetic. It does not, by itself, establish that XY represents a scientifically meaningful derived quantity. That requires the additional demonstration of an empirical quantitative relationship between the attributes represented by X and Y.

SECTION 3 DIMENSIONAL HOMOGENEITY

Establishing that a quantity possesses the properties of a linear ratio scale does not, by itself, authorize its combination with other quantities through arbitrary arithmetic. A further requirement is dimensional homogeneity. The dimensions of the quantities entering an equation must be consistent with the mathematical operations performed upon them and with the quantity represented by the result. This requirement, associated particularly with Fourier's *The Analytical Theory of Heat*, provides a fundamental constraint upon the construction and interpretation of quantitative relationships.

The principle is straightforward. Consider velocity defined as distance divided by time:

$$V = D/T$$

If distance has dimension length, $[D] = L$, and time has dimension T, then:

$$[V] = [D]/[T] = L/T = LT^{-1}.$$

The inverse relationship is:

$$D = VT$$

and therefore:

$$[D] = [V][T] = (LT^{-1})(T) = L.$$

The dimensions on each side of the equation are consistent. The arithmetic represents a defined quantitative relationship among distance, velocity and duration. Changing the units from meters to kilometers or seconds to hours changes the numerical representations but does not alter the underlying dimensional relationship.

Several consequences follow. Quantities retain their dimensional identities when subjected to arithmetic. Addition and subtraction require quantities that are dimensionally commensurable. Multiplication and division generate dimensions determined by the dimensions of their components. An equation purporting to represent an empirical relationship must therefore remain dimensionally coherent. The fact that numbers can be multiplied computationally does not establish that the resulting number represents a scientifically meaningful quantity.

This distinction is particularly important for the QALY. Its defining operation is:

$Q = UT$

where T represents duration and U is required to operate as a quality adjustment. If Q is claimed to retain the dimension of time, then:

$[Q] = [T].$

Where square brackets denote the dimension of a quantity. Thus, $[T]$denotes the dimension of time and $[Q]$the dimension of the QALY. If the QALY is claimed to retain the dimension of time, then $[Q] = [T]$. Because $Q = UT$, it follows that $[U] = 1$; utility must therefore function as a dimensionless factor.

But:

$[Q] = [U][T].$

It follows necessarily that:

$[U] = 1.$

Utility must therefore function as a dimensionless proportional factor. This requirement cannot be satisfied merely by assigning utility values between numerical anchors such as zero and one. Being dimensionless numerically is not equivalent to being a dimensionless measure. The proportional interpretation must itself have an empirical measurement foundation.

Nor is dimensional homogeneity sufficient to establish the scientific legitimacy of multiplication. Suppose P is a legitimate dimensionless proportion and:

$T^* = PT.$

Then:

$[T^*] = [P][T] = 1 \times T = T.$

The expression is dimensionally homogeneous. But this demonstrates only that the dimensions are consistent. It does not demonstrate why possession of the attribute represented by P should proportionally adjust duration or what empirical quantity PT represents. That relationship must be established independently.

Dimensional homogeneity therefore constitutes a necessary but not sufficient measurement gate. A proposed quantitative operation must satisfy both the dimensional requirements of the equation and the measurement properties of the quantities involved. Dimensional consistency cannot create measurement, and measurement alone cannot create an empirical relationship between otherwise unrelated quantities.

For the QALY, the next question is consequently unavoidable: what establishes U as a genuine dimensionless proportional quantity capable of adjusting time? Answering that question requires examination of the construction and properties of proportional factors.

SECTION 4 DIMENSIONLESS PROPORTIONAL FACTORS

The existence of a linear ratio scale does not by itself explain how one quantity can operate as a proportional adjustment to another. A further distinction is required between a ratio-scale measure and a dimensionless proportional factor. These are not the same thing.

Consider two quantities, X_1 and X_2, measured on the same linear ratio scale for the same attribute X. Their ratio is:

$$P = X_1/X_2$$

Because X_1 and X_2 measure the same attribute, they possess the same dimension. If the dimension of X is represented by [X], then:

$$[P] = [X_1]/[X_2]$$

$$= [X]/[X]$$

$$= 1$$

The dimensions cancel. P is therefore dimensionless. This is the familiar origin of a genuine proportion or relative quantity. It is not produced merely by assigning a number between zero and one. It is produced by taking the ratio of commensurable quantities possessing the required ratio-scale structure.

Suppose, for example, that one distance is 60 metres and a reference distance is 100 metres:

P = 60 metres/100 metres = 0.60

The value 0.60 has a clear empirical interpretation. The measured distance is 60% of the reference distance. If the measurement unit is changed from metres to centimetres:

P′ = 6,000 centimetres/10,000 centimetres = 0.60

The value is unchanged. The measurement units cancel because numerator and denominator refer to the same attribute and are expressed on the same ratio scale.

The same structure can apply to many relative quantities:

P = observed quantity/reference quantity provided that numerator and denominator are commensurable ratio-scale measures of the same attribute.

Several properties follow.

- **The numerator and denominator must represent the same attribute.** A meaningful dimensionless proportion cannot ordinarily be created by dividing quantities that measure different attributes. Distance divided by distance can produce a dimensionless relative distance. Distance divided by time produces velocity, not a dimensionless proportion.
- **Both quantities must possess the required ratio-scale structure.** If the origin of either scale is arbitrary, the resulting ratio depends upon the chosen numerical representation. A meaningful proportion therefore cannot be manufactured from ordinal or interval-scale values merely by division.
- **The quantities must be commensurable.** They must be expressible in compatible units and refer to quantitatively comparable manifestations of the same attribute.
- **The denominator must provide an empirically interpretable reference quantity.** A proportion is always a proportion relative to something. The denominator cannot be an arbitrary numerical anchor if the resulting P is claimed to have empirical proportional meaning.
- **Zero must have the appropriate interpretation.** If $X_1 = 0$, then $P = 0$ represents zero possession of the numerator quantity relative to the reference. This interpretation depends upon zero already being meaningful on the underlying ratio scale.
- **Unity has a defined interpretation.** If $X_1 = X_2$, then $P = 1$. Unity means equality with the reference quantity. The value 1 is therefore not an arbitrarily selected upper anchor; it follows from the relationship $X_1/X_2 = 1$.
- **Intermediate values have genuine proportional meaning.** $P = 0.5$ means that X_1 is one-half of X_2. $P = 0.8$ means that X_1 is four-fifths of X_2. These interpretations arise from the underlying measured quantities rather than from labels attached to points on a scale.
- **The proportion is invariant under a common change of measurement unit.** If $X_1' = aX_1$ and $X_2' = aX_2$, then:

$P' = X_1'/X_2'$

$= aX_1/aX_2$

$= X_1/X_2$

$= P$

The proportional factor is unchanged.

This last result is fundamental. A genuine dimensionless proportional factor survives a permissible change in the unit used to measure the underlying attribute. It therefore does not carry the arbitrary measurement unit into subsequent calculations.

4. PROPORTIONAL ADJUSTMENT OF ANOTHER QUANTITY

A dimensionless factor can, under appropriate conditions, operate proportionally upon another measured quantity. Suppose duration T is measured on a linear ratio scale and P is a genuine dimensionless proportional factor. We may propose:

$T^* = PT$

Because P is dimensionless:

$[T^*] = [P][T]$

$= 1 \times [T]$

$= [T]$

The adjusted quantity retains the dimension of time.

For example, if $P = 0.60$ and $T = 10$ years:

$T^* = 0.60 \times 10$ years $= 6$ years

Dimensionally, this is coherent. The multiplication has scaled the magnitude of duration without changing its dimension.

But dimensional coherence is necessary, not sufficient, for scientific meaning. We must still establish why P should adjust T. A dimensionless ratio derived from distance, for example, does not become a scientifically meaningful adjustment to survival time merely because its dimensions cancel. The empirical relationship between the proportional factor and the quantity being adjusted must also be specified.

The complete requirement is therefore stronger than dimensional homogeneity alone: lawful ratio measures → ratio of commensurable quantities → dimensionless proportional factor → demonstrated empirical relationship with the target quantity → legitimate proportional adjustment.

5. A BOUNDED PROPORTION

Where X_1 is constrained to lie between zero and a defined reference maximum Xmax, a bounded proportional factor can be written:

$P = X/X_{max}$

with:

$0 \leq P \leq 1.$

The endpoints now have explicit quantitative interpretations.

$P = 0$ means $X = 0$.

$P = 1$ means $X = X_{max}$.

$P = 0.5$ means $X = 0.5X_{max}$.

The ceiling therefore arises from the defined reference quantity. It is not imposed upon an otherwise arbitrary score. Likewise, the intermediate values acquire their meaning from the ratio relationship. The number 0.5 means one-half because the underlying quantity is one-half of the reference quantity, not because an investigator has placed the number 0.5 halfway between numerical anchors labelled zero and one.

This provides an important test for any proposed proportional adjustment factor. A numerical scale bounded by zero and one may look like a proportion without being one. To establish that it is a proportion, it must be possible to identify the underlying ratio quantities whose relationship generates the value.

The forensic questions are therefore straightforward:

What is the attribute X?

What is the ratio-scale measure of X?

What does zero possession of X mean?

What quantity provides the reference denominator?

Why does X/Xref constitute the proposed proportional factor?

Why should that factor operate proportionally upon the target quantity?

Without answers to these questions, a value between zero and one is simply a numerical value between zero and one. Its location within that interval does not establish that it represents a proportion.

This distinction becomes critical when duration is to be adjusted by a proposed health-state value. If time T is to be multiplied by U, with U claimed to act as a dimensionless proportional factor, it is not sufficient that U has been assigned values between zero and one. The required measurement structure must be demonstrated independently.

The logically prior question is therefore not yet whether U can be multiplied by T. It is whether U has been established as a genuine proportional quantity. A proportion must be a proportion of something. If U is represented as:

$U = X/X_{ref}$,

then X and Xref must be commensurable quantities measuring the same defined attribute on the same ratio scale. Their dimensions cancel, making U dimensionless:

$[U] = [X]/[Xref] = 1$.

A value such as $U = 0.60$ would then have a precise empirical interpretation: the measured magnitude of X is 60% of the defined reference magnitude Xref. The prior questions for the QALY are therefore: What is the attribute X? How is it measured? What is the reference quantity Xref? And what demonstrates that U represents the empirical ratio X/Xref?

Until that question is answered, the multiplication of U by time has no demonstrated foundation as a proportional adjustment.

6. A CALIBRATED PROPORTIONAL INSTRUMENT

Consider an instrument designed to measure a defined manifest attribute X and calibrated to report status over the range:

$0 \leq P \leq 1$.

The appearance of values between zero and one does not itself establish proportional measurement. For an instrument reading of, say, $P = 0.60$ to mean that the observed status represents 60% of a defined reference status, the instrument must satisfy the requirements of linear ratio measurement and the reported proportion must have an explicit measurement foundation.

The required properties are:

- **A defined attribute.** The instrument must measure a specified empirical attribute X. The object of measurement cannot be an undefined composite such as general health or quality of life.
- **Unidimensionality.** Variation in the instrument reading must represent variation in that single attribute. Changes in other attributes cannot be allowed to alter the reading while being interpreted as changes in X.
- **A defined measurement unit.** Equal numerical differences must represent equal differences in the measured attribute throughout the operating range of the instrument.
- **An empirically meaningful zero.** The value $P = 0$ must correspond to zero possession or zero magnitude of the defined attribute. Zero cannot simply be a convenient lower anchor selected for scoring purposes.
- **A defined reference quantity.** The value $P = 1$ must correspond to an explicitly identified reference magnitude Xref of the same attribute. Unity acquires meaning from equality with that reference quantity; it cannot simply be labelled “best,” “normal” or “full.”
- **A ratio construction.** The reported proportional status must be interpretable as:

$P = X/Xref$

where X and Xref are commensurable ratio-scale quantities measuring the same attribute.

- **Meaningful intermediate proportions.** A reading P = 0.50 must mean that X is one-half of Xref; P = 0.60 must mean 60% of Xref; and P = 0.80 must mean four-fifths of Xref. These are empirical ratio statements, not merely locations between numerical endpoints.
- **Invariance to permissible changes of unit.** If the underlying measurement unit is changed according to:

$X' = aX$

and:

$Xref' = aXref$,

then:

$P' = X'/Xref' = aX/aXref = X/Xref = P$.

The reported proportional status must therefore be independent of the particular permissible unit used to measure X.

- **Instrument calibration.** The relationship between the observable instrument response and the magnitude of X must be established empirically across the instrument's intended range. The numbers printed on the instrument cannot themselves create the required measurement structure.
- **Stability and reproducibility.** Repeated measurement under equivalent conditions must reproduce the same quantity within specified measurement error, and the calibration must remain invariant within the population and circumstances for which the instrument is intended.

Suppose these conditions have been satisfied and the calibrated instrument reports:

$P = 0.60$.

The statement now has a precise quantitative interpretation. The measured magnitude of X is 60% of the defined reference magnitude:

$X = 0.60Xref$.

This is fundamentally different from assigning the value 0.60 to a health state because respondents placed that state somewhere between numerical anchors labelled zero and one. In the calibrated instrument, 0.60 is the consequence of measurement. It is not a number whose proportional interpretation is assumed from its location on a bounded scale.

Only after these properties have been demonstrated can the further question be asked whether P may legitimately operate as a proportional adjustment to another quantity such as duration T:

$T^* = PT$.

Even then, measurement legitimacy does not establish the required empirical relationship between P and T. The instrument may lawfully measure X and P may be a genuine dimensionless proportion, while PT remains scientifically meaningless unless it is independently demonstrated why proportional possession of X should proportionally adjust duration.

There are therefore two separate gates:

First measurement:
X → lawful linear ratio measurement → calibrated instrument → X/Xref → dimensionless proportional status P.

Then relationship:
P + T → demonstrated empirical relationship → admissible proportional adjustment PT.

A calibrated 0–1 instrument can therefore provide a legitimate proportional factor, but only when the values between zero and one are generated by an underlying ratio measurement structure. The bounds do not create the proportion; the measurement structure creates the proportion.

7. PROPORTIONAL ATTRIBUTES AND QALY CLAIMS

The preceding requirements establish the conditions that must be satisfied before a numerical value can operate as a dimensionless proportional adjustment to time. The QALY can now be examined against those requirements. Its familiar construction is:

$$Q = UT$$

where T represents duration and U is variously described as a utility, health-state value, preference weight or quality adjustment. If Q is to retain the dimension of time, U must function as a dimensionless proportional factor. The forensic question is therefore not simply whether U is a number between zero and one. It is what measured attribute U represents and why its numerical value constitutes a proportion of that attribute.

A legitimate proportional factor has a recognizable measurement pedigree. Suppose X is a unidimensional attribute represented on a linear ratio scale and Xref is an empirically defined reference quantity of that same attribute. A dimensionless proportional factor can then be constructed:

$$P = X/X_{ref}$$

Because numerator and denominator measure the same attribute:

$$[P] = [X]/[X] = 1$$

and P is dimensionless. If X is bounded between zero and the reference quantity Xref, then:

$0 \leq P \leq 1$.

The endpoints and intermediate values now have empirical interpretations. P = 0 represents zero possession of X. P = 1 represents possession equal to Xref. P = 0.5 means that the measured quantity of X is one-half of the reference quantity. The numerical proportion follows from measurement; it is not created by assigning the numbers 0, 0.5 and 1.

For a QALY utility U to perform this role, the same questions must therefore be answered.

- **What is the attribute X?** A proportional factor must be a proportion of something. Terms such as health, health-related quality of life, utility and preference do not themselves demonstrate the existence of a defined quantitative attribute.
- **Is X unidimensional?** A single numerical magnitude requires a single attribute. If the health state is described through pain, mobility, anxiety, self-care, physical function and other distinct attributes, their collection does not become unidimensional merely because a single preference value is subsequently assigned to it.
- **How is X measured?** The underlying attribute must possess the required measurement structure before proportional values can be derived from it. Numerical assignment, scoring or preference elicitation cannot substitute for measurement demonstration.
- **Does X possess a non-arbitrary zero?** Zero must represent absence of the measured attribute in the sense required for proportional interpretation. Selecting death as the numerical anchor U = 0 does not by itself demonstrate that death represents zero possession of a defined quantitative attribute.
- **What is Xref?** If U = 1 represents the reference quantity, the quantity represented by that endpoint must be specified. Describing it as “full health” does not establish what measurable unidimensional attribute has reached its reference magnitude.
- **Do intermediate values represent genuine proportions?** If U = 0.5, there must be an empirical basis for saying that the individual possesses one-half of the quantity represented at U = 1. Its position halfway between two numerical anchors cannot establish that proposition.

These requirements expose a distinction between a bounded numerical index and a bounded proportional measure. Both may take values between zero and one. Only the latter has an underlying ratio structure that explains what those values mean.

The distinction can be illustrated with a single health attribute. Suppose, purely for argument, that a manifest status change is defined as a unidimensional attribute X and that lawful ratio measurement establishes a zero and a reference maximum Xmax. A proportional measure could then be defined as:

Pstatus = Xstatus/Xmax

If P = 0.6, the value would have an explicit interpretation: measured possession of status attribute is 60% of the defined reference quantity. A change from 0.6 to 0.5 would likewise represent a measured proportional change in that particular attribute.

If this factor were subsequently applied to duration:

$T^* = PT$

the resulting claim would, at most, concern duration adjusted with respect to the specified attribute, assuming that the relationship between that proportional possession and duration had also been empirically justified.

It would not be a measure of quality of life.

This distinction becomes decisive when several health attributes are involved. Pain, mobility, fatigue, anxiety, physical function, cognition and social functioning can each contribute to the description of a person's health. They need not be manifestations of a single quantitative attribute. Indeed, their clinical usefulness may depend precisely upon their representing different aspects of health.

Suppose each were measured perfectly: P_1, P_2, P_3, P_4, P_5

The existence of five lawful measures would not establish a sixth measure called “quality of life.” Nor would assigning weights and combining them establish one:

$U = f(P_1 \ldots\ldots\ldots P_5)$.

The result would be a composite index unless an independent measurement model demonstrated that these components constituted manifestations of a single underlying attribute with the required quantitative structure. The measurement properties of the components cannot simply be transferred to their weighted combination.

This gives the QALY problem two separate levels.

The first is measurement. U must represent a defined unidimensional attribute possessing the ratio structure necessary for proportional interpretation.

The second is relationship. Even if such an attribute were successfully measured, it must still be demonstrated why its proportional possession should multiply duration and what empirical quantity the product represents.

The second condition cannot be inferred from the first. Ratio-scale status establishes meaningful proportional comparisons within an attribute. It does not establish that the attribute can legitimately be multiplied by time.

Distance and duration illustrate the distinction. Their relationship can define velocity because velocity has an independently understood empirical interpretation as distance traversed per unit time. Similarly, a rate multiplied by duration may yield a meaningful accumulated quantity where the relevant process has been defined. The arithmetic represents an established relationship among quantities.

For the QALY the corresponding question is: What empirical quantity is created when the proportional possession of the attribute represented by U is multiplied by elapsed duration?

Saying that the answer is "quality-adjusted life-years" merely gives the product a name. Saying that one year at U = 0.5 equals 0.5 QALYs restates the arithmetic. Neither establishes the measurement structure of U nor the empirical relationship that makes UT a meaningful derived quantity. The forensic sequence required for a QALY claim is therefore: defined attribute X → lawful ratio measurement of X → defined reference quantity Xref → dimensionless proportion U = X/Xref → demonstrated empirical relationship between U and duration T → admissible multiplication UT → defined derived quantity Q.

Every step must be established independently. None can be inferred retrospectively from the desire to calculate a QALY.

This is the critical distinction between describing health and measuring a quantity capable of proportionally adjusting time. A multidimensional health profile may provide valuable information about a patient. A preference attached to that profile may provide information about how respondents value it. Neither fact demonstrates the existence of a unidimensional proportional quantity that can multiply duration.

The explicitly multiattribute construction of health-state utilities raises a further and logically prior question. Were the additive or multiplicative algorithms used to generate these values ever designed to construct a new, single empirical attribute possessing the measurement properties required of a dimensionless proportional factor? There appears to be no such objective. Multiattribute utility functions were developed to represent preferences over combinations of distinct attributes. Additive forms aggregate component utilities under specified independence assumptions, while multiplicative forms accommodate particular interactions among preferences for those attributes. Neither mathematical operation demonstrates that mobility, pain, cognition, anxiety, self-care and other health dimensions have become manifestations of a single unidimensional quantitative attribute. Nor does the resulting algorithm establish an empirical unit, a non-arbitrary zero or meaningful ratios for such an attribute.

The algorithm produces a preference index over multidimensional health states; it does not demonstrate the existence of a new ratio-scale quantity. Consequently, describing the resulting number as "overall health," "health-related quality of life" or "utility" cannot establish the measurement structure required for that number to operate as a dimensionless proportional adjustment to time. The critical question remains unanswered: what single empirical attribute does U measure, and what demonstrates that U represents proportional possession of that attribute?

The burden of proof therefore lies with the QALY claim. Before U can adjust T, it must be shown what U measures, why its zero is non-arbitrary, what its unit represents, what U = 1 represents quantitatively, why U = 0.5 represents one-half of that quantity, and why proportional possession of that attribute has the required multiplicative relationship with duration.

Without those demonstrations, U is not established as a dimensionless proportional measure. It is simply a numerical value being required to behave as one.

8. THE ASSESSMENT OF LATENT ATTRIBUTES

The preceding analysis has concerned linear ratio measurement and the conditions under which a dimensionless proportional factor might legitimately adjust a measured duration. These requirements are appropriate where the relevant quantities can be represented through linear ratio measurement. They do not, however, provide the measurement pathway required for latent attributes. Here the measurement problem changes fundamentally.

Many of the outcomes of greatest importance to patients are latent. Pain, fatigue, physical function, anxiety, treatment burden and need fulfilment cannot be measured by placing the patient against a physical ruler. They must be inferred from responses to appropriately constructed items representing increasing or decreasing possession of a defined attribute. The scientific objective remains measurement, but the measurement model must establish a quantitative continuum from the relationship between persons and items.

This is the role of Rasch measurement proposed in 1960 [3, 4]

The Rasch model provides the necessary measurement structure for transforming ordered categorical observations into locations on a unidimensional latent continuum. For dichotomous responses, the basic relationship may be written:

$$\log[P_{ni}/(1 - P_{ni})] = \theta_n - \beta_i$$

where P_{ni} is the probability that person n endorses item i, θ_n represents the person's possession of the latent attribute and β_i represents the location or difficulty of item i on that same attribute.

The significance of this formulation is fundamental. Persons and items are located jointly on a common continuum. The resulting logit represents the logarithm of the odds associated with the difference between person and item locations. Measurement is therefore not produced by summing arbitrary numerical labels attached to questionnaire responses. It is produced by testing whether the observed responses conform sufficiently to the requirements of the measurement model.

For a latent attribute to support a therapy-impact claim, the measurement structure must demonstrate, among other requirements:

- **A defined unidimensional attribute.** The items must represent one latent attribute rather than an uncontrolled mixture of distinct constructs.
- **Ordered item structure.** Items and, where applicable, response thresholds must behave consistently with increasing possession of the attribute.
- **Person-item conjoint measurement.** Persons and items must be located on the same latent continuum.
- **Specific objectivity.** Comparisons between persons should not depend upon the particular admissible items used, and comparisons between items should not depend upon the particular admissible persons used for calibration, subject to model requirements.

- **Invariance.** Item functioning must remain sufficiently stable across relevant groups and circumstances. Differential item functioning is evidence against the required invariance where it cannot be resolved.
- **Local independence.** Once possession of the latent attribute is accounted for, item responses should not retain unexplained dependence that would undermine the unidimensional measurement structure.
- **Model fit.** Observed responses must be evaluated against the requirements of the Rasch model. Measurement status is demonstrated through conformity to the model rather than conferred by the existence of questionnaire data.
- **A common logit metric.** Differences in possession of the attribute can then be represented on the resulting logit continuum, permitting assessment of change with associated measurement uncertainty.

This is a fundamentally different scientific objective from constructing a health-state utility. The objective is not to take several health dimensions, attach preference weights to them and compress them into a number intended to adjust survival time. It is to identify one attribute, construct a measurement system for that attribute and determine whether therapy changes the patient's possession of it.

Suppose the attribute is neck pain. The relevant therapy-impact question is not:

How many quality-adjusted life-years did the treatment produce?

It is: How did treatment change the patient's possession of the defined neck-pain attribute?

Likewise, if the attribute is physical function, fatigue or need fulfilment, the objective is to estimate change in possession of that particular attribute. Different attributes remain different attributes. There is no requirement to force them into a common preference-weighted composite and no scientific advantage in doing so merely to facilitate a cost-per-QALY calculation.

The Rasch approach therefore avoids the dimensional and proportional problems encountered by the QALY because it does not attempt to make the latent attribute a dimensionless adjustment factor for time. Duration remains duration. Neck pain remains neck pain. Physical function remains physical function. Need fulfilment remains need fulfilment. Therapy impact is evaluated separately for the attributes that the therapy is intended to change.

Within the framework adopted here, proportional interpretation is obtained through the odds structure underlying the logit. If θ_1 and θ_2 represent two locations on the latent continuum, their difference is:

$\Delta\theta = \theta_2 - \theta_1$

and exponentiation gives:

$OR = \exp(\Delta\theta)$.

A one-logit difference corresponds to an odds ratio of approximately 2.72. The therapy-impact claim can therefore be expressed in terms of change in possession of the latent attribute and the corresponding change in odds, rather than by pretending that the latent attribute can be converted into a dimensionless fraction of "full health" and multiplied by time.

This distinction is decisive for health technology assessment. The failure of the QALY does not imply that patient-reported outcomes cannot be measured or that therapy impact on latent attributes cannot be quantified. It implies precisely the opposite: if the objective is measurement, the latent attribute must be approached through a measurement model designed to establish possession of that attribute.

The replacement for utility measurement is therefore not a better utility instrument. Nor is it another preference algorithm, another set of population weights or a revised zero-to-one scale. Those approaches retain the premise that heterogeneous health experience should be compressed into a numerical factor capable of adjusting time. The alternative abandons that premise.

For manifest attributes, therapy impact should be evaluated using appropriate linear ratio measurement. For latent attributes, therapy impact requires Rasch measurement of possession of the defined attribute. The two pathways address different classes of attributes but share the same scientific discipline: the attribute is defined first, measurement is established second, and quantitative claims follow only after the measurement requirements have been satisfied.

There is consequently no need for the QALY to mediate therapy-impact claims. Survival can be measured as survival. Hospitalization can be measured as hospitalization. Walking distance can be measured as distance. Pain, fatigue, physical function and need fulfilment can be evaluated as defined latent attributes through Rasch measurement. Each claim retains its empirical identity and can be evaluated, replicated and potentially falsified.

The transition is therefore not from the QALY to another universal health index. It is from numerical compression to attribute-specific measurement. The scientific question is no longer how diverse manifestations of health can be forced into a common numerical currency. It is whether a therapy produces a demonstrable change in possession of the particular manifest or latent attributes that matter to the patient.

That is the measurement question the QALY displaced, and it is the question to which health technology assessment must return. This will be detailed in the successor paper

9. TRANSFORMATION TO LAWFUL MEASUREMENT

Once measurement inversion has been demonstrated, there is only one scientifically defensible avenue forward: reconstruction from measurement. Legacy HTA cannot be repaired by improving utilities, refining QALYs, elaborating reference cases or constructing more sophisticated decision models. These activities occur downstream of the foundational failure. If the numerical representations entering the arithmetic have not first been established as measures possessing the properties required for that arithmetic, additional arithmetic cannot correct the problem. Nor can professional consensus, institutional acceptance or decades of application provide the missing measurement demonstration. The required transformation is therefore not reform of legacy HTA but its replacement by a framework in which measurement precedes arithmetic.

That transformation is already operational. Maimon Research has developed a nine-unit program for rebuilding HTA around the requirements of representational measurement, designed both for independent study and as the foundation for a three-credit graduate course [5] . The program begins where quantitative science must begin: with the definition of the attribute and the requirements that must be satisfied before numerical operations are permitted. It then distinguishes the two measurement pathways required for therapy-impact claims: linear ratio measurement for manifest attributes and Rasch measurement for latent attributes.

Across the nine units, participants move from measurement inversion and the foundations of representational measurement through scale properties, admissible transformations and arithmetic, dimensional homogeneity, manifest and latent attributes, construction of Rasch measurement systems, possession and change, and the requirements for empirically evaluable therapy-impact claims. The sequence culminates in application, replication and falsification.

The transformation therefore requires neither another generation of methodological debate nor permission from an HTA agency or professional organization. The scientific sequence is straightforward: attribute → measurement → scale → admissible arithmetic → therapy-impact claim → empirical evaluation → replication → falsification. The program is available now. The remaining task is application.

10. CONCLUSION

The implications extend far beyond rejection of the QALY. The QALY was not an isolated methodological curiosity. It became the organizing quantitative construct of a global HTA enterprise. ISPOR, academic research centers, professional networks, journals and assessment agencies built an elaborate methodological infrastructure around utilities, QALYs, cost-effectiveness models, ICERs and reference cases. Researchers were trained to apply it, manufacturers were required to submit analyses within it, and agencies used its outputs to support reimbursement decisions.

Once the required standards of representational measurement are applied, the foundation of this edifice disappears. The QALY is not an imperfect measure requiring refinement. It is an impossible construct. Its multiattribute health-state values have not been demonstrated to constitute a unidimensional proportional measure of quality of life, and multiplication by time cannot create the measurement properties that are absent from the multiplicand. Everything subsequently constructed from the QALY inherits that failure.

The conclusion is therefore uncomfortable but unavoidable. An enormous global intellectual and institutional investment has been wasted. Decades have been devoted to refining methods, constructing models, estimating utilities, developing reference cases, debating thresholds and managing uncertainty around calculations whose foundational measurement requirements were never established. Greater sophistication downstream cannot rescue failure at the measurement gate.

This failure was avoidable. Had the elementary requirements of representational measurement been applied before the QALY took center stage, its impossibility would have been apparent before an international HTA infrastructure was constructed around it. Historical acceptance cannot now provide scientific legitimacy retrospectively. The appropriate response is abandonment and reconstruction from measurement first principles.

This paper has examined the first catastrophic failure of quantitative HTA: the impossible QALY. The next paper examines the second: the failure to distinguish the fundamentally different measurement requirements for manifest and latent attributes, and the consequences of that failure for claims of therapy impact.

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.

REFERENCES

[1] Stevens S. On the Theory of Scales of Measurement. *Science*. 1946;103(2684):677-80

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

[3] Rasch G, Probabilistic Models for some Intelligence and Attainment Tests. Chicago: University of Chicago Press, 1980 [An edited version of the original 1960 publication]

[4] Bond T, Zi Yan. Heene M, Applying the Rasch model: Fundamental measurement in the human sciences. New York: Routledge, 2021

[5] Maimon Research, Transformation Program. https://maimonresearch.com/hta-reconstruction-program-and-fees/