

MAIMON RESEARCH LLC

**ARTIFICIAL INTELLIGENCE LARGE LANGUAGE
MODEL INTERROGATION**



**THE MEASUREMENT FUTURE IN HEALTH
TECHNOLGY ASSESSMENT**

**NONSENSE ON STILTS: WHY THE QALY NEVER
EXISTED AS A LAWFUL MEASURE**

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

LOGIT WORKING PAPER No 845 SEPTEMBER 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

The quality-adjusted life year (QALY) has occupied a central position in health technology assessment (HTA) for more than four decades, providing the principal outcome for cost-utility analysis and the denominator for the cost-per-QALY incremental cost-effectiveness ratio. Despite this extensive institutionalization, the QALY has rarely been subjected to the prior question required of any quantitative construction: does it satisfy the requirements of representational measurement?

This paper demonstrates that the QALY has never existed as a lawful measure. Its familiar construction, $QALY = T \times U$, multiplies ratio-measured time by a health-state preference value conventionally bounded by zero and one. For this operation to have the claimed proportional interpretation, U must first possess measurement properties permitting a value such as 0.5 to mean one-half. Conventional preference values do not meet this requirement merely by construction. They are valuations attached to multiattribute health-state descriptions and have not been demonstrated to constitute unidimensional ratio measures possessing equal units, a non-arbitrary zero, invariant ratios and the admissible transformations required for proportional arithmetic.

Dimensional homogeneity establishes that U must be dimensionless if the product is to retain the dimension of time. Dimensionless, however, does not establish proportional measurement. Nor does a numerical preference value of 0.5 establish that time should be reduced by one-half. Even if preference values hypothetically possessed lawful ratio properties, an additional measurement relationship would be required to establish that a proportion in preference space can be transferred unchanged to time space.

The failure has been obscured by 40 years of institutionalization. Increasing sophistication in preference elicitation, modelling, uncertainty analysis and decision methods proceeded downstream from the QALY while its prior measurement requirements remained largely unexamined. More than 200 interrogations of HTA knowledge bases demonstrate the resulting measurement inversion: arithmetic precedes measurement.

Rejecting the QALY creates no measurement vacuum. Therapy-impact assessment can return to defined attributes, employing linear ratio measurement for manifest attributes and Rasch measurement for latent attributes. The replacement for the QALY is therefore not another universal index. It is measurement.

INTRODUCTION

It has been made clear in over 200 interrogations of the measurement status of HTA knowledge bases that few practitioners appear to have a clear understanding of an attribute, what constitutes a measure, or why the axioms of representational measurement and the scales of measurement are fundamental to quantitative analysis¹. This matters because numbers do not become measures merely by being assigned to observations, health states or preferences. The measurement properties of a quantity determine what arithmetic can legitimately be performed upon it.

Few constructs have achieved the status in health technology assessment accorded to the quality-adjusted life year (QALY). For more than four decades it has provided the organizing outcome for cost-utility analysis, informed reference cases and reimbursement guidelines, occupied a central place in textbooks and graduate education, and supported the familiar cost-per-QALY incremental cost-effectiveness ratio. An extensive methodological literature has developed around the elicitation of health-state preferences, construction of utility values, choice of preference-based instruments, mapping, discounting, modelling, uncertainty and decision thresholds. The QALY has become so deeply embedded in contemporary HTA that its existence as a legitimate quantitative measure is rarely treated as a proposition requiring demonstration. It is ordinarily taken as a starting point.

This acceptance is undeserved. The QALY, as will be demonstrated in this paper, has never existed as a lawful measure. The reason is clear-cut: its construction proceeded without recognition of the requirements imposed by the axioms of representational measurement and the scales of measurement, and specifically the requirements for a numerical value to function as a proportional adjustment factor for time^{2 3}.

These elementary requirements must therefore be established before considering the QALY. Its construction, $QALY = T \times U$ requires more than a numerical preference weight and a measure of time. It requires that U first possess the measurement properties necessary to represent a lawful proportion and, second, that there be a demonstrated measurement basis for using that proportion to adjust time. Neither requirement follows from assigning a health-state preference a numerical value between zero and one. Understanding why these requirements are not met begins with the most basic question of all: what is a measure?

WHAT IS A MEASURE?

A measure is a numerical representation of the magnitude of a defined attribute that preserves relevant relationships observed for that attribute. Measurement therefore begins not with a number, but with the question: what is the attribute being measured?

Consider length. Length is an attribute of an object. If one rod is longer than another, a measurement system for length must preserve that empirical ordering in its numerical representation. If rod A is longer than rod B, its measured length must be greater. More demanding numerical statements require correspondingly more demanding measurement properties. To say that one rod is twice as long as another requires a measurement scale on which the numerical ratio represents the corresponding empirical ratio of lengths. This distinction is fundamental because assigning numbers is not the same as measurement.

Measurement requires three elements. First, there must be a defined attribute: length, mass, duration, temperature, pain, fatigue or some other property whose magnitude or possession is of interest. Second, there must be an empirical structure that permits relationships among instances of that attribute to be established. Third, the numerical representation must preserve the relationships required for the quantitative claims being made.

Different measures consequently permit different arithmetic. A numerical representation that establishes only order can support statements such as: $A > B$ but it does not necessarily support: $A = 2B$. The latter is a proportional statement and requires substantially stronger measurement properties. In particular, meaningful ratios require a ratio scale with equal units and a non-arbitrary zero, so that the statement that one quantity is twice another is preserved when the permissible unit of measurement is changed.

Time provides a familiar example. Duration can be measured on a ratio scale. If: $T_1 = 6$ months and $T_0 = 12$ months then $T_1 / T_0 = 0.5$ has a measurement interpretation: the first duration is one-half of the second. Expressing the durations in days rather than months does not change that relationship. The number 0.5 represents an invariant proportional relationship between two measured durations.

This provides the essential distinction needed for evaluating the QALY. A health-state preference may be assigned the numerical value $U = 0.5$ but the appearance of 0.5 does not establish that anything has been measured at one-half of anything else. To make that claim requires a demonstrated measurement structure capable of supporting proportional statements.

The question that must therefore precede the QALY construction $QALY = T \times U$ is not simply “what number has been assigned to U ?” it is “What has been measured, and what measurement properties establish that $U = 0.5$ means one-half?”

Only after that question has been answered can it be determined whether the number is capable of functioning as a proportional adjustment factor. Numbers are the products of measurement only when they represent demonstrated properties of an attribute. Numbers assigned without that demonstration remain numbers; arithmetic performed upon them cannot convert them retrospectively into measures.

A measure is therefore a numerical representation of the magnitude of a defined attribute that preserves the relevant empirical relationships of that attribute. Measurement therefore begins with the attribute, not with the number. Length, mass and duration are familiar attributes; pain, fatigue and physical function are latent attributes requiring a different measurement pathway. In each case, the question is whether observations can be represented numerically in a way that satisfies the requirements of measurement.

Merely assigning a number does not create a measure. Patient identification numbers, football jersey numbers and numerical codes for health states are numbers, but they do not measure an attribute. Measurement requires a demonstrated relationship between the attribute and its numerical representation. The properties of that relationship determine the scale of measurement and, consequently, the arithmetic that can legitimately be performed. The essential principle is therefore straightforward: measurement gives numbers their quantitative meaning; numbers alone do not create measurement.

RATIO MEASUREMENT

The distinction between scales of measurement was formalized for the modern behavioral and social sciences by Stevens in 1946³. Stevens distinguished nominal, ordinal, interval and ratio scales according to the empirical relationships represented and the transformations that leave those relationships unchanged. The importance of this classification is straightforward: the scale on which an attribute is measured determines the quantitative statements and arithmetic that can legitimately be interpreted. Numbers do not authorize arithmetic simply because arithmetic can be performed upon them.

The ratio scale provides the strongest of the familiar measurement structures. A unidimensional ratio measure possesses ordered magnitudes, equal units and a non-arbitrary zero. Its admissible transformations are similarity transformations: $X' = aX$, where $a > 0$.

Changing from meters to centimeters or from years to months changes the numerical values but does not change the relationships represented. Consequently, ratios are invariant: $X'_1 / X'_0 = aX_1 / aX_0 = X_1 / X_0$. If one duration is 6 months and another is 12 months then $T_1 / T_0 = 6 / 12 = 0.5$. The statement that the first duration is one-half of the second remains true whatever legitimate unit of time is employed. The number 0.5 therefore has a demonstrated proportional interpretation.

This is why ratio measurement is central to multiplication and to the QALY. The familiar construction is $QALY = T \times U$. Time can satisfy the requirements of linear ratio measurement. The critical question is therefore the measurement status of U . If $U = 0.5$ is to operate as a proportional adjustment factor, 0.5 must possess more than a numerical location between the conventional anchors zero and one. It must support the interpretation one-half. That interpretation cannot be created by calling U a utility, preference value or weight.

Conventional health-state preference values have not been demonstrated to possess the required ratio-scale structure merely by being elicited and placed on a 0-to-1 scale. Numerical anchors do not demonstrate equal units, a non-arbitrary zero, unidimensionality or invariant ratios. If $U = 0.8$ and $U = 0.4$, it cannot simply be assumed that the first represents twice the measured magnitude of the second.

The QALY therefore confronts a measurement requirement before multiplication begins. If U has not been demonstrated to support lawful proportional statements, it cannot function as the proportional factor required by $T \times U$. The ability to multiply the two numbers computationally does not establish that their product is a measure. This is why the ratio-scale requirements identified in the Stevens framework are not peripheral to the QALY debate: they determine whether the arithmetic that creates the QALY has quantitative meaning at all.

PREFERENCE VALUES AND RATIO MEASUREMENT

If a preference value is to enter the QALY as a proportional factor, it must possess substantially different properties from those ordinarily required of a health-state valuation. The conventional HTA preference value describes the relative desirability of a health state. The preference value

required for proportional multiplication must instead satisfy the requirements of ratio measurement. These are not interchangeable numerical constructs.

The first requirement is identification of the attribute being measured. Measurement requires a defined object of measurement. A conventional preference-based health-state instrument typically describes several attributes simultaneously. An EQ-5D health state, for example, combines mobility, self-care, usual activities, pain/discomfort and anxiety/depression. A preference elicited for the resulting health-state description is therefore a valuation of a multiattribute state. It does not, merely by being assigned a numerical value, become a measure of possession of a single defined attribute.

This distinction is made particularly important by descriptions of the QALY construction in which preference scores are used directly as adjustment factors for time⁴. Preferences for health states are elicited to permit quality-of-life adjustments to treatment outcomes. The QALY is then created by adjusting the length of time associated with an outcome by the preference weight assigned to the corresponding level of health status, conventionally represented on a scale from 0 to 1. Thus $QALY = T \times U$ with the preference value U is not merely reporting that one health state is preferred to another. It is being required to perform a much stronger quantitative function. It is treated as a proportional adjustment factor for time. A preference value of $U = 0.5$ is used as though it establishes $T_{adjusted} / T = 0.5$ so that $T_{adjusted} = 0.5T$.

This is the critical measurement step. Nothing about assigning a multiattribute health-state preference a value between 0 and 1 demonstrates that the value possesses the ratio properties required for this interpretation. In particular, the numerical value 0.5 does not establish that the health state possesses one-half of a defined attribute, still less that the corresponding duration should be represented as one-half of its unadjusted value.

The preference score has therefore moved from valuation of a multiattribute health state to proportional adjustment of a ratio measure of time. That transition requires a measurement demonstration. It cannot be established merely by defining the QALY as the product of time and the preference weight. To do so would assume, in the construction itself, the proportional property that must first be demonstrated before the multiplication is admissible.

This immediately raises the requirement of unidimensionality. A ratio measure must represent variation in the magnitude of the attribute it claims to measure. If the object being valued simultaneously changes in mobility, pain, anxiety and usual activities, there is no single health magnitude for which the resulting preference number has automatically demonstrated a ratio structure. Preference for a composite health state is a legitimate object of valuation, but valuation of a multidimensional description is not equivalent to measurement of a unidimensional attribute.

The second requirement is equal units. For a value such as $U = 0.5$ to have proportional meaning, the numerical continuum must support equal increments in the measured attribute. The difference represented by a given numerical interval must have the required invariant quantitative interpretation throughout the scale. Assigning numerical values through preference elicitation or statistical modelling does not itself establish equal measurement units.

Third, a proportional measure requires a non-arbitrary zero. The zero must represent absence of the measured quantity in the sense required for meaningful ratio comparisons; it cannot simply be a conventionally selected reference point. Assigning death = 0 and designated full health = 1 creates anchors for a valuation scale. It does not demonstrate that zero is the natural origin of a measured preference attribute. The possibility of negative values for states considered worse than death makes the distinction particularly conspicuous. A conventional anchor and a ratio-scale zero perform fundamentally different measurement functions.

Fourth, the preference values must permit meaningful ratios. If $U_1 = 0.8$ and $U_2 = 0.4$ then use of these values as proportional quantities requires that $U_1 / U_2 = 2$ have an empirical interpretation: U_1 represents twice the measured magnitude represented by U_2 . This cannot be inferred from the numerical values themselves. It must follow from the demonstrated measurement structure.

Fifth, these ratio statements must be invariant under admissible transformations. For a ratio scale, permissible changes of unit take the form $U' = aU$, where $a > 0$ so that $U'_1 / U'_2 = aU_1 / aU_2 = U_1 / U_2$. The proportional relationship is preserved. If transformations involving an arbitrary additive constant are permissible, such as $U' = aU + b$ then ratios are not preserved and the scale cannot support the proportional interpretation required for multiplication.

These requirements reveal the categorical distinction between the preference value conventionally supplied to HTA and the quantity required by the QALY construction. The former is a numerical valuation attached to a multidimensional description of a health state. The latter must be a demonstrated proportional measure capable of functioning as a dimensionless multiplier.

Even satisfying these requirements would establish only that the preference value possessed lawful proportional meaning within the preference attribute space. It would not yet establish that the resulting proportion could lawfully adjust time. If $P_1 / P_0 = 0.5$ were demonstrated as a valid preference ratio, it would establish $P_1 = 0.5P_0$. It would not by itself establish $T_{\text{adjusted}} / T = 0.5$. That requires the additional preference-to-time relationship considered separately.

The QALY therefore encounters failure before the multiplication begins. The conventional HTA preference value is a valuation of a multidimensional health-state description, not a demonstrated unidimensional ratio measure possessing equal units, a non-arbitrary zero, invariant ratios and the admissible transformations required of a proportional quantity. Calling that value a utility, bounding it between zero and one, or describing it as dimensionless cannot supply the missing measurement structure. Before U can proportionally adjust time, U must first be demonstrated to be a proportion. Conventional health-state preference values do not satisfy that requirement merely by construction.

CREATING A DIMENSIONLESS NUMBER

Creating a dimensionless number is straightforward. Demonstrating that the number is a lawful proportional measure is an entirely different matter. This distinction is critical to the QALY because the preference value U must be dimensionless if multiplication by U is to leave the resulting quantity in the dimension of time.

The theory of dimensions is the framework for checking whether the quantities in a mathematical relationship are dimensionally consistent. A dimension identifies the kind of quantity being measured, independently of the units used. Length has dimension [L], time [T], and mass [M]. Thus 6 months and 180 days have different numerical values and units but the same dimension: time.

Dimensions combine according to the arithmetic. For example, velocity is: $v = \text{distance} / \text{time}$ so $[v] = [L] / [T]$. A ratio of two quantities having the same dimension cancels T_1 / T_0 has $[T] / [T] = 1$ and is therefore dimensionless.

The central principle is dimensional homogeneity: a quantitative equation must be dimensionally consistent. Quantities being added or equated must have compatible dimensions, while multiplication and division produce the dimensions implied by their components.

The dimensional requirement follows immediately from $QALY = T \times U$. If the QALY is claimed to retain the dimension of time $[QALY] = [T]$ then $[T] = [T] \times [U]$ and therefore $[U] = 1$. U must be dimensionless. This is a necessary condition for dimensional homogeneity, but it says nothing by itself about whether U possesses proportional meaning.

A dimensionless proportional quantity can be constructed directly from two commensurable ratio measures. Suppose T_1 and T_0 are ratio measures of duration. Then $U = T_1 / T_0$ and $[U] = [T] / [T] = 1$. If $T_1 = 6$ months and: $T_0 = 12$ months then $U = 6 / 12 = 0.5$. Here 0.5 has a demonstrated proportional interpretation $T_1 = 0.5T_0$ with one duration is one-half of the other. The ratio is unchanged if both durations are expressed in days, hours or seconds. Its proportional meaning is invariant under admissible changes in the unit of time.

But dimensionless alone does not establish this property. Consider two ratio measures of distance: $R = 5 \text{ km} / 10 \text{ km} = 0.5$. Again: $[R] = [L] / [L] = 1$. The number is dimensionless and the ratio is legitimate. It establishes that one distance is one-half of another. It does not establish that a period of time should be reduced by one-half. The proportional meaning belongs to the relationship from which the ratio was derived.

The distinction becomes even more important when a number has simply been assigned to a health-state preference. A preference value: $U = 0.5$ is numerically dimensionless in the trivial sense that no physical unit such as years, kilograms or meters is attached to it. But this does not establish $P_1 / P_0 = 0.5$ nor does it demonstrate $P_1 = 0.5P_0$. Still less does it establish $T_{\text{adjusted}} / T = 0.5$. These are measurement propositions, not consequences of writing the number 0.5.

The QALY therefore requires substantially more than a unit-free preference value. U must first possess a demonstrated proportional interpretation and, if that proportion is derived from an attribute other than time, there must then be a demonstrated basis for applying that proportion to time.

The distinction can be summarized simply: dimensionless number $\neq$ dimensionless proportional measure $\neq$ proportional adjustment to time where each step requires additional justification.

Dimensional homogeneity establishes that U must be dimensionless. It does not establish that $U = 0.5$ means one-half, and it certainly does not establish that $U = 0.5$ lawfully makes time one-half.

CROSS-SPACE MAPPING: PREFERENCE SPACE AND TIME SPACE

The central error in the QALY construction is the failure to recognize that a numerical value defined in preference space cannot simply be transferred into time space and treated as a proportional adjustment factor. A path between the two spaces has to be demonstrated.

Time is measured on a ratio scale. If two time quantities are T_1 and T_2 then a proportional adjustment factor within time space can be written as $A_t = T_1 / T_2$. Because numerator and denominator are measures of the same dimension $[T_1] = [T_2] = \text{time}$ the dimensions cancel $[A_t] = [T] / [T] = 1$ and the resulting quantity is dimensionless. If $T_1 / T_2 = 0.5$ then the statement that T_1 is one-half of T_2 has a legitimate proportional meaning because both quantities are defined on the same ratio scale within the time space.

The QALY, however, does not derive its adjustment factor from time space. It imports a numerical value from a different space: a health-state preference value conventionally bounded by 0 and 1. The construction then treats $U = 0.5$ as though it were equivalent to $T_{\text{adjusted}} / T = 0.5$.

The missing step is the demonstration of the mapping Preference space $\rightarrow$ Time space or, more formally $f(P) = T_{\text{adjusted}} / T$.

The conventional preference space does not provide the required basis for such a mapping. Health-state preference values are generated from composite descriptions and have not been demonstrated to possess the properties of ratio measurement. Consequently, values such as 0.4, 0.5 or 0.8 cannot automatically be interpreted as proportions of possession of a defined attribute. There is therefore no legitimate proportional quantity in preference space available to be transferred to time space.

The problem remains even under the hypothetical assumption that the preference values could be established as ratio measures. Suppose $P_1 / P_2 = 0.5$ were a legitimate ratio in preference space. This would demonstrate only that P_1 bears a one-half proportional relationship to P_2 with respect to the preference attribute. It would not establish $T_1 / T_2 = 0.5$. The additional equality $T_1 / T_2 = P_1 / P_2$ would still require independent empirical and measurement demonstration.

The QALY therefore requires two things that are not supplied by simply assigning health-state values between 0 and 1: first, a legitimate proportional quantity within preference space; and second, a demonstrated mapping showing that this proportion can be transferred unchanged into time space.

Without both, the preference value is not a measured proportional adjustment factor for time. It is merely a numerical value that has been inserted into a multiplication with time and interpreted as though the required mapping had already been established. Absent an adjustment factor in time space for the attribute time the QALY is a meaningless construct.

INSTITUTIONAL HUBRIS

Why, if its required measurement properties were never demonstrated, did the QALY not merely survive but become the dominant outcome of health technology assessment? The answer lies less in the strength of its measurement foundations than in the institutional machinery that converted an unproved measurement assumption into normal science. That machinery appears to have operated with little understanding of representational measurement, the scales of measurement, ratio measurement and the constraints these impose upon quantitative operations. This absence of measurement knowledge provided fertile ground for ersatz measurement claims: numerical constructions were accepted as measures because they could be calculated, standardized and routinely applied, rather than because their measurement properties had first been demonstrated.

The initial assumption was deceptively simple. A health-state preference value U , conventionally represented on a numerical scale anchored by zero and one, was treated as though it possessed the proportional properties necessary to adjust time. Once this construction was accepted, the measurement question effectively disappeared from view. Research no longer had to establish whether U could lawfully perform the operation attributed to it. Instead, attention moved downstream: how should U be elicited, which preference instrument should be used, how should utilities be mapped when unavailable, how should QALYs be discounted, how should incremental QALYs enter economic models, and what cost-per-QALY threshold should guide decisions?

An institutional infrastructure then accumulated around the assumption. Textbooks taught it. Graduate programs reproduced it. Journals published applications of it. Professional organizations developed Good Practices around it. HTA agencies incorporated it into reference cases. Researchers developed increasingly sophisticated methods for generating and manipulating it. Consultants and pharmaceutical companies constructed submissions around it. Students entered the field learning not whether the QALY was a lawful measure but how to calculate and apply it. Even Wikipedia reproduced it without challenge.

This is institutional hubris in a specific scientific sense. It does not require arrogance or bad faith on the part of individual researchers. It arises when the accumulated authority of institutions becomes a substitute for demonstration. The QALY appeared secure because prestigious organizations endorsed it, leading textbooks taught it, agencies required it and an enormous literature applied it. Each layer of acceptance reinforced the others. Yet none could confer upon U measurement properties that had never been demonstrated.

The resulting system became self-validating. Widespread use was interpreted as maturity; methodological elaboration as scientific progress; standardization as legitimacy; and replication of the method as confirmation of the underlying quantity. But repeating an assumed measurement operation does not validate it. Twenty-seven thousand applications in PubMed cannot substitute for one demonstration that the required arithmetic is admissible.

The deeper hubris was therefore institutional rather than computational. HTA became extraordinarily sophisticated in what it did after accepting the QALY, while neglecting the elementary measurement question that had to be answered before the first QALY could legitimately be calculated. Increasingly elaborate decision models, probabilistic sensitivity

analysis, simulation, mapping, value-of-information analysis and cost-effectiveness thresholds created an appearance of quantitative maturity. Yet none addressed the prior question of whether the preference value being multiplied by time possessed the measurement properties required to perform that function.

ISPOR's international educational and methodological infrastructure subsequently helped globalize this architecture. Good-practice recommendations, educational programs, conferences, publications, task forces and regional networks reinforced an analytical sequence in which utilities, QALYs, ICERs and decision models were treated as established components of HTA rather than as constructions whose measurement foundations first required demonstration. Each generation of researchers could therefore inherit the QALY as accepted knowledge rather than encounter it as a measurement proposition requiring proof.

Measurement inversion consequently ceased to be simply an analytical mistake. It became embedded in the international knowledge base of HTA. More than 200 interrogations of HTA knowledge bases have repeatedly demonstrated the same underlying pattern: extensive recognition of contemporary economic evaluation, utilities, QALYs, cost-effectiveness and modelling accompanied by little or no recognition of the representational-measurement requirements that must precede their arithmetic.

Legacy HTA therefore occupies an extraordinary position for a putatively quantitative discipline. It promotes quantification while largely bypassing the axioms of fundamental measurement, the scales of measurement, admissible transformations and the constraints these impose upon arithmetic. Numbers are generated, combined, multiplied and divided without first demonstrating that they constitute measures capable of supporting those operations. The result is not simply imperfect quantitative science. It is a classic example of a non-science masquerading as a quantitative discipline: numerical sophistication without the measurement foundations required to give the numbers empirical meaning.

The inversion is complete. The downstream machinery became increasingly sophisticated while the measurement foundation on which it depended remained largely unexamined. Institutional acceptance substituted for measurement demonstration, repetition was mistaken for validation, and computational sophistication created the appearance of quantitative science where the elementary requirements for measurement had never been satisfied.

This explains both the survival and the dominance of the QALY. It was protected not by demonstrated measurement properties but by an increasingly powerful ecosystem whose methods, publications, curricula and decision frameworks presupposed its legitimacy. The institutional machinery converted an unproved measurement assumption into normal science and then used the resulting normal science as evidence that the assumption no longer required proof. That circularity is the essence of institutional hubris.

This is where the veneration of the QALY becomes scientifically consequential. The literature has overwhelmingly proceeded downstream from the multiplication rather than upstream to the measurement conditions that authorize it. Increasing sophistication in preference elicitation, economic modelling, probabilistic analysis and decision rules cannot compensate for a

measurement requirement that was never satisfied before the arithmetic began. Repetition, institutional endorsement and widespread application can establish the influence of a method; they cannot establish its measurement legitimacy.

The issue addressed here is therefore more fundamental than whether one preference instrument should be preferred to another or whether QALYs adequately represent social value. It concerns whether the QALY ever satisfied the measurement requirements necessary for it to exist as the quantitative measure it claims to be. If U has not been demonstrated to be a lawful proportional adjustment to time, the QALY fails at the operation that creates it.

REFUSING THE QALY

Refusing the QALY does not create a measurement vacuum. It removes an inadmissible construction and allows therapy-impact assessment to return to its proper scientific foundation: defined attributes and lawful measures of those attributes. The transition is straightforward and is developed in a nine-unit program *Reconstructing HTA Programs: Measurement before Arithmetic*⁵.

The starting point is the therapy claim. What attribute is the intervention expected to change? Once that attribute has been defined, only two measurement pathways are required. Manifest attributes such as survival time, hospital days, walking distance or treatment administrations require appropriate linear ratio measurement with a non-arbitrary zero, equal units and admissible transformations. Latent attributes such as pain, fatigue, physical function or need fulfilment require construction of a unidimensional measurement system. Rasch measurement provides the required conjoint person-item framework, invariance and an additive logit metric for evaluating possession and change in the latent attribute.

The nine-unit program develops this transition from attribute definition through scale requirements, manifest and latent measurement, admissible arithmetic, dimensional homogeneity, protocol design, therapy-impact claims, empirical evaluation, replication and falsification. Measurement once again precedes arithmetic.

Nothing scientifically valuable is lost by abandoning the QALY. There is no requirement to compress multiple attributes into a preference score and then pretend that this score can proportionally adjust time. Therapy impact can instead be reported directly for the attributes that treatment is intended to change, using measures capable of supporting the claims made for them.

The replacement for the QALY is therefore not another universal index. It is measurement.

CONCLUSION

The final defense of the QALY may appear disarmingly simple: that is how we define it. A QALY is defined as time multiplied by a health-state preference weight, and a preference weight of 0.5 is therefore defined to convert one year into 0.5 quality-adjusted years. End of story.

But convenience cannot manufacture measurement properties, any more than defining π as 3 for computational convenience would make π equal to 3. In the QALY case the problem is deeper: the preference weight has never been demonstrated to possess the proportional properties required of the time-adjustment factor in the first place.

Arbitrary definition cannot establish measurement. A numerical construction does not become a lawful measure merely because a professional community agrees to define, calculate and use it. Either quantitative HTA respects the requirements of representational measurement or it retreats to a numerical dark age in which numbers are manipulated without first establishing what those numbers are capable of representing.

Time provides no difficulty. Duration can be represented on a linear ratio scale with equal units and a non-arbitrary zero. Within that measurement space, proportional relationships are meaningful and invariant. If $T_1 = 6$ months and $T_0 = 12$ months then $T_1 / T_0 = 0.5$ is dimensionless and has an unambiguous empirical interpretation: T_1 is one-half of T_0 . Consequently, $T_1 = 0.5T_0$ is a lawful proportional transformation of time. Dimensional homogeneity is preserved because $[T_1 / T_0] = [T] / [T] = 1$.

The QALY requires its preference weight to perform precisely this proportional role. If $\text{QALY} = T \times U$ then U must function as a dimensionless proportional adjustment to T . Yet assigning a health-state preference the value $U = 0.5$ does not establish that anything has been measured as one-half of anything else. Conventional preference values have not been demonstrated to possess the unidimensional ratio-scale structure required for such an interpretation.

Even granting, for argument, that preference were lawfully measured on a ratio scale would not solve the problem. A preference ratio $P_1 / P_0 = 0.5$ would establish a proportional relationship in preference space. The QALY requires the further relationship $T_{\text{adjusted}} / T = P_1 / P_0$.

No rule of arithmetic, dimensional analysis or representational measurement establishes this equality merely because both ratios are dimensionless. The mapping from a proportion in preference space to an identical proportion in time space is a substantive measurement proposition that must be independently demonstrated.

It has not been demonstrated. Instead, the required proportional relationship is assumed by defining the QALY as $T \times U$. The construction is therefore being used to supply the measurement property required to justify the construction. A neatly circular argument.

That is insufficient. Definition cannot substitute for measurement, convention cannot establish admissible arithmetic, and decades of application cannot confer measurement properties that were never demonstrated. If the standards of representational measurement are respected, the conventional preference weight cannot perform the function required of it. Without a lawful proportional adjustment to time there is no lawful QALY and there never has been.

The QALY takes a preference score assigned to a composite, multiattribute health state and uses that score as though it were a proportional adjustment factor for ratio-measured time. The key questions for this absurdity are: What attribute does U measure? Is that attribute unidimensional?

Is U a ratio measure? Does $U = 0.5$ lawfully mean one-half? What measurement properties make $U = 0.5$ a lawful proportional multiplier of time? If not, on what measurement basis can 0.5 reduce 12 months to 6 months?

The answer cannot be: because $QALY = T \times U$ is how the QALY is defined.

The choice is therefore stark. We either require quantitative claims in HTA to satisfy the established standards of measurement, or accept numerical operations simply because convention has defined them to be so. The first is measurement science. The second is a numerical dark age.

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 Maimon Research Catalog of Measurement Interrogations www.maimonresearch.com

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

³ Stevens S. On the Theory of Scales of Measurement. *Science*. 1946;103(2684):677-80

⁴ Drummond M, Sculpher M, Claxton K et al. Methods for the Economic Evaluation of Health Care Programmes (4th Ed). New York: Oxford University Press, 2015

⁵ Maimon Research. Reconstructing HTA Programs: Measurement before Arithmetic
<https://maimonresearch.com/hta-reconstruction-program-and-fees/>