MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT AND HEALTH TECHNOLGY ASSESSMENT

LEGACY HTA: DESPITE ISPOR, THE PROOF FOR ABANDONMENT IS BEYOND QUESTION

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

LOGIT WORKING PAPER No 2008 SEPTEMBER 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

Legacy health technology assessment (HTA) has developed around utilities, quality-adjusted life-years (QALYs), cost-effectiveness ratios and reference-case modelling without first establishing the measurement properties required for their defining arithmetic. This paper subjects that framework to the requirements of representational measurement and asks whether there is any lawful measurement pathway capable of supporting its central quantitative claims. Measurement must begin with a defined unidimensional attribute, establish the required empirical and axiomatic structure, determine the scale achieved and restrict arithmetic to operations admissible for that scale. Legacy HTA reverses this sequence, allowing arithmetic requirements to determine the numerical constructs subsequently treated as measures: measurement inversion.

Two measurement pathways are identified for therapy-impact claims. Appropriate manifest attributes can support linear ratio measurement, permitting invariant ratios, multiplication and division where dimensional requirements are satisfied. Latent attributes such as pain, fatigue, physical function and need fulfilment require Rasch measurement, providing an additive logit continuum and invariant multiplicative comparisons through ratios of odds. These pathways cannot be interchanged, nor do they provide a scientific basis for collapsing different attributes into a universal outcome currency.

The QALY fails both pathways. Conventional health-state utility is a weighted valuation of a multidimensional health-state description, not a demonstrated unidimensional linear ratio measure. It therefore cannot provide the dimensionless proportional quantity required to adjust ratio-scaled time. Rasch measurement cannot rescue the QALY because a logit difference supports an invariant ratio of odds, not a linear proportion of latent-attribute possession capable of multiplying time. Consequently, there is no measurement-theoretic pathway from conventional utility or lawful latent-attribute measurement to the QALY.

Failure of the QALY removes the denominator of the conventional cost-per-QALY reference case, while the price-weighted composite cost numerator presents an independent measurement problem. Modelling, simulation and uncertainty analysis cannot repair these prior failures. Legacy HTA therefore fails both as a quantitative measurement framework and as a framework for establishing therapy impact.

Abandonment is consequently required rather than methodological refinement. The alternative is measurement-first HTA: define the therapy-relevant attribute, select the appropriate manifest or latent measurement pathway, establish admissible arithmetic, and construct attribute-specific claims capable of empirical evaluation, replication and falsification. A structured nine-unit program provides an immediately available pathway for this transition. The burden now rests upon those continuing to promote or teach legacy HTA to demonstrate, by measurement proof rather than institutional authority or convention, that its defining quantitative claims are scientifically admissible.

INTRODUCTION

Quantitative health technology assessment (HTA) rests upon a requirement that is logically prior to economic evaluation, modelling or decision analysis: before numbers can support a quantitative claim, the attribute represented by those numbers must be capable of lawful measurement [1].[2] Numerical assignment alone does not establish measurement, and mathematical operations do not acquire empirical meaning merely because they can be performed computationally.

The required sequence begins with a defined unidimensional attribute, followed by establishment of its measurement requirements and measurement scale, determination of admissible arithmetic and, only then, construction of a quantitative claim. This sequence places measurement before arithmetic. Legacy HTA has commonly reversed it. Numerical representations required for economic evaluation are generated first and subsequently subjected to multiplication, division, aggregation and subtraction without prior demonstration that their measurement properties permit those operations. This reversal is measurement inversion.

For therapy-impact claims, two distinct measurement pathways must be recognized. The appropriate pathway depends upon the nature of the attribute being measured. Manifest attributes are directly observable properties for which an appropriate unit and non-arbitrary zero can be established. Examples may include survival duration, hospital days, treatment administrations, walking distance and other directly observable quantities. Where the required empirical structure is demonstrated, these attributes can support linear ratio measurement.

Ratio status is critical because it determines the quantitative interpretations and arithmetic that are admissible. It cannot be inferred merely from the presence of numerical values or an assigned zero. The empirical and numerical structures required for ratio measurement must first be established. For a manifest therapy-impact claim, the sequence therefore proceeds from identification of the manifest attribute to establishment of linear ratio measurement, determination of admissible arithmetic and construction of a quantitative therapy-impact claim. This permits scientifically interpretable statements concerning the magnitude of therapy effects where the required measurement properties have been demonstrated.

Many attributes of central importance to therapy evaluation are not directly observable. Pain, fatigue, physical function, treatment burden and need fulfilment are examples of latent attributes. Their possession must be inferred from observable responses to appropriately constructed items.

A patient-reported outcome instrument does not itself solve this problem. A patient-reported outcome is a source of observations; it is not a measurement model. Assigning integers to response categories, summing responses or statistically transforming a score does not establish measurement of a latent attribute.

Rasch measurement provides the required pathway by establishing a defined unidimensional latent attribute through conjoint measurement of persons and items [3] [4] . The resulting logit metric provides an additive representation of differences in possession of the attribute, while its corresponding odds representation provides invariant multiplicative comparisons. The important point for HTA is not the detailed mathematics of the Rasch model, but the establishment of a lawful measurement structure before quantitative therapy-impact claims are constructed. The latent pathway therefore proceeds from identification of a latent attribute through Rasch measurement and demonstration of the required measurement properties to construction of an admissible quantitative therapy-impact claim.

The manifest and latent pathways are different, but they share the same scientific principle: measurement status must be demonstrated before arithmetic is undertaken. Recognition of these two pathways has an important consequence for HTA. Lawful measurement does not imply that different attributes become commensurable merely because each has been measured quantitatively. Survival, pain, fatigue, physical function and treatment burden remain different attributes. Each has its own empirical meaning and dimensional identity.

There is therefore no scientific requirement that therapy impacts be collapsed into a single universal outcome measure; a critical if not foundational failure in Legacy HTA. A treatment might simultaneously produce an increase in survival, a reduction in pain and an improvement in physical function. These are legitimate quantitative claims when each attribute has been lawfully measured. They do not have to be converted into a common preference-weighted currency before they can inform evaluation.

Preference weighting cannot manufacture commensurability between attributes that are intrinsically different. Nor can the assignment of a common numerical range transform distinct dimensions into possession of a single quantity. The appropriate outcome of measurement-first HTA is therefore an attribute-specific profile of therapy impact rather than necessarily a composite index.

The QALY illustrates why the distinction between measurement pathways matters [5] . Its construction requires a health-state value to operate as a proportional adjustment to elapsed time. This creates specific measurement requirements for the health component. It must possess the properties required to function as a legitimate proportional quantity.

Conventional health-state utilities do not acquire those properties merely because they are generated by preference-elicitation procedures or anchored numerically. They represent valuations of multidimensional health states rather than demonstrated ratio measurement of possession of a defined unidimensional attribute. The arithmetic required for the QALY therefore cannot be justified simply by calling these values utilities.

Replacing conventional utility with a Rasch measure does not rescue the construction. Rasch measurement provides lawful measurement of a particular latent attribute; it does not transform that attribute into a universal dimensionless proportional adjustment to time. A lawful measure of pain, fatigue or physical function remains a measure of that attribute. What Rasch provides, as a necessary and sufficient measure, is an assessment of the extent to which there is possession of the latent attribute.

The failure is therefore deeper than inadequate psychometrics. The QALY attempts to create a common quantitative currency by combining fundamentally different quantities without first establishing the measurement structure required for that combination.

The alternative is not abandonment of quantitative therapy evaluation. It is its reconstruction. HTA should begin by asking what therapy impacts matter, identifying the attributes represented by those impacts and determining the measurement pathway appropriate to each. Manifest attributes should be evaluated through lawful linear ratio measurement. Latent attributes should be evaluated through Rasch measurement. Quantitative claims should then be restricted to those supported by the demonstrated properties of the resulting measures.

three decades: the legacy HTA framework of utilities, QALYs, cost-effectiveness analysis and reference-case modelling. That framework is now challenged not merely because its defining quantitative operations fail the requirements of representational measurement, but because it does not provide lawful measures of the attributes through which therapies actually affect patients. It therefore fails at two levels: as a quantitative measurement framework and as a framework for establishing therapy impact. The alternative begins elsewhere with the identification of the relevant attribute, its lawful measurement and the construction of therapy-impact claims capable of empirical evaluation, replication and falsification.

Abandonment of legacy HTA and ISPOR’s promotion restores the required sequence. The attribute is defined first, the appropriate ratio-measurement pathway is selected, the required measurement properties are demonstrated, admissible arithmetic is determined, and only then is a therapy-impact claim constructed and subjected to empirical evaluation. Economic evaluation, statistical analysis and decision making can then proceed from quantities whose empirical meaning has first been established rather than assumed.

The central requirement is consequently uncomplicated. HTA does not need a universal numerical currency before it can evaluate therapy impact. It needs lawful measures of the attributes that therapies actually affect. There are two measurement pathways: linear ratio measurement for manifest attributes and Rasch measurement for latent attributes. The scientific task is to select the appropriate pathway, establish measurement first, and only then determine what quantitative claims the resulting measure permits for the selected attribute.

1. SCIENCE BEGINS WITH THE ATTRIBUTE

Quantitative analysis does not begin with numbers. It begins with a defined attribute: the particular property whose magnitude or possession is to be represented. Survival duration, hospital days, pain, fatigue and physical function are examples. Before numbers are assigned, the investigator must specify exactly what is being measured. An attribute cannot simply be nominated and numbers attached to it. Its empirical structure must support the proposed numerical representation.

The first requirement is unidimensionality. Measurement must concern one attribute at a time. A ruler can represent differences in length because length is the attribute being measured; a Rasch ruler can represent differences in possession of physical function only if the items conform to a single defined physical-function construct. If several distinct attributes are combined, the resulting number does not become a measure merely because an algorithm produces a single numerical output. One number is not evidence of one dimension.

This point is fundamental to the subsequent critique of legacy HTA. A multiattribute health-state description deliberately combines dimensions such as mobility, self-care, usual activities, pain/discomfort and anxiety/depression. These are different attributes. Applying weights to their levels and combinations may generate a composite utility number, but weighting cannot transform multiple attributes into one unidimensional attribute. The measurement failure therefore begins before the scale properties of the utility number are even considered.

Once a unidimensional attribute has been established, the investigator must determine whether its empirical structure satisfies the axiomatic requirements of the proposed measurement representation. These include, as appropriate, order, additivity, solvability, cancellation and the Archimedean condition. Only after these requirements have been addressed does it become meaningful to ask what scale of measurement has been achieved and, consequently, what arithmetic operations are admissible.

Scales of Measurement

Scales of measurement are important because numbers do not acquire identical mathematical properties merely by being numbers. Nominal, ordinal, interval and ratio scales represent different empirical structures and permit different transformations. An ordinal scale establishes order but not equal quantitative differences. An interval scale supports meaningful differences but permits an arbitrary origin. A ratio scale has the additional structure required for invariant ratio comparisons.

The permissible transformations identify what remains quantitatively meaningful. If an empirical conclusion changes merely because one permissible numerical representation has been replaced by another, that conclusion cannot be an invariant measurement claim. Scale type therefore determines the arithmetic that can legitimately be performed. Mathematical software may multiply or divide any numbers supplied to it; measurement science determines whether those operations have empirical meaning.

The Axioms of Representational Measurement

The classification of a numerical representation as a measurement scale cannot rest simply upon the appearance of the numbers assigned. Representational measurement requires a correspondence between the empirical relationships exhibited by the defined unidimensional attribute and the numerical relationships used to represent them. The numerical system must preserve the relevant structure of the empirical system. This is why the axioms of representational measurement are fundamental: they specify the conditions that the empirical attribute must satisfy before a particular numerical representation can legitimately be claimed. Measurement is therefore not achieved by assigning numbers and subsequently deciding what those numbers will mean. The empirical structure must justify the numerical structure.

For quantitative measurement, the relevant requirements include unidimensionality, order, additivity, solvability, cancellation and the Archimedean condition, as appropriate to the measurement structure being constructed, together with invariance under the permissible transformations of the resulting scale. Unidimensionality is indispensable: the relationships being represented must concern one defined attribute, not a weighted combination of different attributes. Order requires that greater and lesser possession can be consistently represented; additivity requires an appropriate concatenation or conjoint structure; solvability requires that intermediate quantitative relationships can, in principle, be represented; cancellation ensures consistency among combinations; and the Archimedean condition excludes structures in which finite comparisons cannot be established. The resulting numerical representation must then remain invariant under the transformations permitted for that scale. These requirements are not optional statistical tests added after numbers have been generated. They are the conditions that make measurement possible in the first place.

2. RATIO MEASUREMENT

Ratio measurement is the critical threshold for any analytical framework that depends upon multiplication, division or proportional interpretation. Failure to recognize this requirement virtually guarantees quantitative failure. This is precisely the problem encountered by legacy HTA. Its defining operations require utilities to act as proportional adjustments to time, QALYs to be aggregated and differenced, and incremental quantities to be divided to produce cost-effectiveness ratios. Yet the ability to perform these operations mathematically does not establish their admissibility as measurement operations. Arithmetic cannot confer ratio properties upon numbers that did not possess them before the arithmetic began**.** An analytical framework that starts by specifying the calculations it wishes to perform and only subsequently assumes that its inputs possess the necessary measurement properties has reversed the scientific sequence. Ratio measurement must precede ratio arithmetic.

Ratio measurement is therefore not defined simply by the presence of a numerical zero, by reporting positive numbers, or by the mathematical possibility of forming a quotient. A ratio measure must represent a defined unidimensional attribute whose empirical structure satisfies the axiomatic requirements necessary for quantitative representation. These include order, so that greater possession of the attribute is empirically distinguishable from lesser possession; additivity or an appropriate concatenation structure, so that quantitative differences have an empirical interpretation; the Archimedean property, ensuring that finite comparisons can be represented within the numerical system; solvability, allowing equivalent quantitative relationships to be established; and cancellation, ensuring consistency among comparisons and combinations of attribute magnitudes. The representation must also satisfy invariance, so that conclusions about the attribute do not depend upon arbitrary features of the numerical assignment, and the admissible transformations must preserve the ratio structure. For a conventional linear ratio scale this requires a meaningful non-arbitrary zero and transformations restricted to similarity transformations, $x' = ax$, $a > 0$. Consequently, statements such as *x is twice y* remain invariant under every admissible transformation.

These requirements establish a simple but uncompromising sequence: attribute → empirical relational structure → axiomatic requirements → ratio representation → admissible transformations → admissible arithmetic. Ratio status must therefore be demonstrated before multiplication, division or proportional interpretation is attempted; it cannot be assumed because an analytical framework requires those operations. If the attribute is undefined or multidimensional, if the empirical structure required by the representation has not been demonstrated, if the zero is arbitrary, or if transformations involving an additive constant are admissible, proportional statements are not invariant and ratio arithmetic is inadmissible. This is the point at which legacy HTA was effectively guaranteed to fail**.** Once utilities and other numerical constructs were permitted to enter multiplication and division without first demonstrating the required ratio properties, everything constructed downstream including the QALY, ΔQALY and the cost-per-QALY ICER became vulnerable to foundational measurement failure. The issue is therefore not whether the subsequent mathematics is sophisticated or correctly executed. It is whether the numbers entering that mathematics were measures of the required kind in the first place.

The notion of ratio measurement is absent from ISPOR's promotion of the legacy HTA framework. This omission is fundamental because multiplication and division require measurement properties that cannot simply be assumed from the assignment of numerical values. Yet utilities, QALYs and cost-per-QALY claims have been promoted without first establishing the ratio measurement foundations required for their defining arithmetic. The result is not a minor methodological omission but the absence of the measurement pathway required to justify the quantitative framework itself.

3. MANIFEST AND LATENT ATTRIBUTES

Once the attribute has been defined, its nature determines the appropriate measurement problem. The fundamental distinction is between manifest and latent attributes.

Manifest attributes are directly observable. Survival duration, hospital days, ICU hours and treatment duration are examples. Where the empirical structure satisfies the required axioms, these attributes can support conventional linear ratio measurement. There is a meaningful zero, units may be changed proportionally, and ratios remain invariant:

Latent attributes present a fundamentally different problem. Pain, fatigue, physical function and need fulfilment cannot be directly observed or physically concatenated. What can be observed are responses to items designed to provide evidence concerning possession of the specified attribute. Assigning integers to response categories or summing those integers does not create measurement. The latent attribute must first be explicitly defined and unidimensional. A measurement system must then establish an invariant continuum on which differences in possession can be represented. Rasch measurement provides this structure by locating persons and items jointly on a common logit continuum. Importantly, the observations must conform sufficiently to the requirements of the measurement model. The model does not confer measurement status automatically upon whatever responses are supplied to it.

HTA therefore has two distinct routes: (i) manifest attribute → axiomatic requirements → linear ratio measure (ii) latent attribute → axiomatic/model requirements → Rasch logit measure. This distinction has a major consequence. There is no measurement requirement that survival, pain, fatigue, mobility and physical function be collapsed into one overarching number. Each is an attribute in its own right and must retain the measurement structure appropriate to it.

If HTA is to assess therapy impact on latent attributes such as pain, fatigue, physical function or need fulfilment, it requires a measurement framework capable of establishing and quantifying attribute possession. For more than fifty years, Rasch measurement has provided the necessary and sufficient framework for constructing invariant measures of possession from ordinal response observations. There is no alternative route that satisfies the same measurement requirements simply by scoring, weighting or statistically modelling questionnaire responses. Rasch measurement begins with a defined unidimensional attribute and requires the data to conform to the measurement model before claims for possession and change in possession are accepted. For readers wishing to pursue the measurement of latent-attribute possession and its application to therapy-impact claims, the Maimon Research nine-unit program provides the necessary introduction to this framework (see Section 9 below).

4. TWO ROUTES TO RATIO MEASUREMENT

The two routes provide different forms of quantitative representation, and their ratio interpretations must not be confused.

For a manifest attribute satisfying the requirements of linear ratio measurement:

$X' = aX,\ a > 0$

Ratios of quantities are invariant. If:

$X_1 / X_2 = 2$

then after a permissible change of unit:

$aX_1 / aX_2 = X_1 / X_2 = 2$

This structure also explains why scale properties matter for multiplication. Suppose X and Y are appropriate ratio measures and:

$Z = XY$

Under permissible transformations:

$X' = aX$

$Y' = bY$

then:

$Z' = X'Y'$

$Z' = (aX)(bY)$

$Z' = abXY$

and since:

$Z = XY$

then:

$Z' = abZ$

The derived quantity changes only by the predictable change in units.

Contrast this with a quantity permitting an arbitrary origin:

$X' = aX + c$

Then:

$Z' = (aX + c)(bY)$

$Z' = abXY + bcY$

and therefore:

$Z' = abZ + bcY$

The additional term bcY means that the transformed product is no longer simply a rescaled version of Z. The arbitrary origin contaminates the product. The ability to perform the multiplication arithmetically does not make the resulting quantity measurement-theoretically admissible.

Rasch measurement provides a different ratio structure for a latent attribute. For the dichotomous Rasch model with questionnaire items defining the assessment instrument:

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

where P_{ni} is the probability that person *n* gives the specified response to item *i*, θ_n is the person's location on the latent attribute and β_i is the item's location on that same continuum. The expression:

$P_{ni} / (1 - P_{ni})$

is the odds of the specified response. The Rasch model therefore represents the logarithm of those odds, the log odds or logit as the difference between person and item locations.

If therapy changes a person's location on the latent continuum from θ_1 before treatment to θ_2 after treatment, the change is:

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

Because the Rasch continuum is expressed in log odds, exponentiating this difference returns the comparison to the odds scale. If the corresponding odds before and after treatment are O_1 and O_2, then:

$\theta_1 = \log(O_1)$

$\theta_2 = \log(O_2)$

Therefore:

$\Delta\theta = \log(O_2) - \log(O_1)$

Using the logarithmic identity:

$\log(A) - \log(B) = \log(A/B)$

gives:

$\Delta\theta = \log(O_2/O_1)$

Exponentiating both sides:

$\exp(\Delta\theta) = O_2/O_1$

Hence:

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

where OR is the ratio of the odds associated with the two locations.

For example, if:

$\Delta\theta = 1$

then:

$OR = \exp(1) \approx 2.72$

Equal logit differences have equal multiplicative interpretations on the odds scale regardless of where the difference occurs.

For example:

$-3 \rightarrow -2$: $\Delta\theta = -2 - (-3) = 1 \rightarrow OR = e^1 \approx 2.72$

and:

$2 \rightarrow 3$: $\Delta\theta = 3 - 2 = 1 \rightarrow \text{OR} = e^1 \approx 2.72$

Likewise, a two-logit change anywhere on the continuum gives:

$\text{OR} = e^2 \approx 7.39$

and a half-logit change gives:

$\text{OR} = e^{0.5} \approx 1.65$

The important qualification is that this does not mean that the *probability* increases by the same amount at every location. Odds and probabilities are related nonlinearly. The same one-logit change produces the same multiplicative change in odds, but different absolute changes in probability depending on the starting location.

The odds associated with the second location are therefore approximately 2.72 times the odds associated with the first. This is the ratio interpretation supplied by the Rasch logit structure. It is crucial, however, not to misinterpret it. It is a ratio of odds, not a ratio of quantities of the latent attribute itself. A change from 1 to 2 logits does not mean that the person has twice as much physical function, need fulfilment or another latent attribute. This distinction becomes decisive when considering whether a Rasch measure could ever provide the proportional adjustment to time required for a QALY.

The ratio is therefore an invariant ratio of odds. It is not a ratio of the numerical logit coordinates themselves. A person at 2 logits does not possess twice as much of the attribute as a person at 1 logit. This distinction is decisive. Linear ratio measurement permits direct ratios of quantities; Rasch logit measurement permits invariant differences and corresponding ratios of odds. Neither structure can be changed simply because a subsequent HTA calculation requires a particular type of number.

5. DIMENSIONLESS RATIOS AND DIMENSIONAL HOMOGENEITY

The preceding mathematics establishes what happens when quantities are multiplied. The next requirement is particularly important for the QALY: what kind of number can legitimately be used to adjust ratio-scaled time proportionally?

Consider a patient who spends 18 hours of a 24-hour observation period satisfying a precisely defined condition A. The proportion of time in A is:

D = time in A / total time

so:

D = 18 hours / 24 hours = 0.75

In dimensional analysis, square brackets denote the dimension of a quantity. Thus [T] denotes the dimension of time and [C] the dimension of cost. A ratio such as [X]/[X] is dimensionless because the same dimension cancels. By contrast, [X]/[Y] retains dimensions unless X and Y are dimensionally identical. This distinction is critical: a numerical ratio is not necessarily a dimensionless ratio.

As both numerator and denominator measure the same attribute time on the same ratio scale. The units cancel:

$[D] = [T] / [T] = 1$

D is therefore dimensionless. Its value is also invariant to the unit chosen. Expressing the same quantities in minutes gives:

$D = 1080 \text{ minutes} / 1440 \text{ minutes} = 0.75$

More generally, if:

$D = X / Y$

and X and Y are commensurable ratio measures of the same attribute, then under a permissible change of unit:

$X' = aX$

$Y' = aY$

and:

$D' = X' / Y'$

$D' = aX / aY$

$D' = X / Y = D$

The dimensionless ratio can therefore legitimately rescale time:

$T = D \times T^*$

with:

$[T] = [1][T] = [T]^*$

This is dimensional homogeneity. The calculation begins with time and ends with time because the multiplier is a genuine dimensionless proportion. This gives us a simple but demanding test. If a number such as 0.75 is to adjust time proportionally, we must be able to ask: 0.75 of what?

What is its numerator? What is its denominator? What single attribute do they measure? Are both ratio measures? What units cancel? A decimal between zero and one is not automatically a proportion. Its proportional meaning must arise from its measurement structure. This requirement provides the direct test for conventional utility.

6. THE QALY FAILS LINEAR RATIO MEASUREMENT

The conventional QALY begins not with a unidimensional ratio attribute but with a multiattribute health-state description. A utility algorithm applies weights to the dimensions and levels defining that description, conventionally anchoring full health at unity and applying decrements to generate the utility score. In simplified form: $U = 1 - \Sigma d_i$

The resulting U is therefore a numerical representation of a weighted multiattribute health-state description. It is not a demonstrated measure of a single unidimensional attribute satisfying the axiomatic requirements for linear ratio measurement.

The decrement structure may also generate: $U < 0$ with negative values conventionally representing health states worse than dead. Whatever preference interpretation is given to such numbers, this is not the structure of a conventional proportion constructed from a positive linear ratio attribute.

Yet the next step treats U precisely as a proportional adjustment:

$QALY = U \times T$

If:

$U = 0.75$

then one year is transformed into:

0.75×1 year = 0.75 QALYs

The critical question is now unavoidable:

What is $U = 0.75$ a proportion of?

A lawful dimensionless proportion would require:

$U = X / Y$

where X and Y are commensurable ratio measures of the same defined attribute. Then:

$[U] = [A]/[A] = 1$

But no such construction exists for conventional utility. There is no demonstrated ratio-scaled numerator, denominator or single attribute whose units cancel. $U = 0.75$ has not been demonstrated to mean 75% of anything.

The possibility of negative utilities makes the distinction especially transparent. If:

$D = X / Y$

with:

$0 \le X \le Y$

then:

$0 \le D \le 1$

A negative utility cannot represent a negative proportion of a positive ratio-scaled attribute. The utility number and a dimensionless ratio are different mathematical objects.

The QALY construction therefore depends upon three successive transitions, none of which has the required measurement foundation. The first is the movement from a multiattribute health-state description to a weighted utility score. Combining and weighting different health attributes may produce a single numerical value, but it does not establish a single unidimensional attribute possessing the empirical and axiomatic properties required for linear ratio measurement. The second transition occurs when this utility score is reinterpreted as a proportional adjustment to time**.** This requires the utility to be a dimensionless ratio constructed from commensurable ratio measures of the same attribute, yet no such ratio structure has been demonstrated. The third transition is the multiplication $U \times T$ and the declaration that the result is a QALY. Multiplication cannot supply the measurement properties missing from the utility input, nor can it create a new linear ratio attribute representing time adjusted for health-state experience. At each stage, the arithmetic proceeds without the measurement structure required to justify it.

Multiplication by time cannot manufacture the missing measurement properties. Under the requirements of linear ratio measurement and dimensional homogeneity, the conventional QALY cannot be constructed.

7. THE QALY FAILS RASCH LOGIT RATIO MEASUREMENT

The failure of conventional utility does not mean that health-related latent attributes cannot be measured. Rasch measurement provides a route to quantitative claims for defined unidimensional attributes such as pain, fatigue, physical function and need fulfilment. But Rasch measurement does not rescue the QALY. It closes the alternative route to its construction.

For the dichotomous Rasch model:

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

where θ_n represents person location and β_i item location on the latent continuum.

If therapy changes a person's location from θ_1 to θ_2:

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

and the corresponding ratio interpretation is:

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

Suppose:

$\theta_1 = 1$ logit

and:

$\theta_2 = 2$ logits

It would be incorrect to calculate:

$\theta_2 / \theta_1 = 2$

and conclude that the second location represents twice as much attribute possession. The zero of the logit continuum is not an absolute absence of the latent attribute. The legitimate quantitative statement is:

$\Delta\theta = 2 - 1 = 1$

and therefore:

$OR = \exp(1) \approx 2.72$

This is an invariant ratio of odds. It is not a statement that the person possesses 2.72 times as much physical function, pain relief or need fulfilment.

This distinction defeats any attempt to convert a Rasch measure into the proportional adjustment required by: $QALY = U \times T$

Rescaling a logit continuum to produce numbers between zero and one does not solve the problem. A transformed value of 0.75 does not thereby become 75% possession of the latent attribute. Numerical rescaling cannot create an absolute zero or convert a ratio of odds into a linear proportion.

Rasch measurement therefore permits rigorous claims concerning change in possession of a single latent attribute, but it cannot produce the dimensionless linear proportion required to adjust time. There is only one possible lawful route to such an adjustment. It must concern a single defined attribute already demonstrated to possess linear ratio properties, from which a genuine dimensionless ratio can be constructed: $D = X / Y$ where X and Y are commensurable measures of that same attribute. For latent attributes, Rasch measurement provides no such D. Nor can separate measures of pain, fatigue, physical function and other latent attributes be collapsed into a universal quality proportion.

The two legitimate measurement routes therefore close the argument. Linear ratio measurement rejects conventional utility because it is not a demonstrated dimensionless ratio. Rasch logit measurement rejects the conversion of latent-attribute possession into the linear proportion required to adjust time. There is no measurement-theoretic route from a multiattribute health-state description or a latent-attribute measure to the conventional QALY.

8. THE COLLAPSE OF THE REFERENCE CASE

The failure of the QALY removes the denominator of the conventional cost-effectiveness reference case. But the numerator presents a separate measurement failure. The familiar expression $ICER = \Delta C / \Delta QALY$ appears to represent the division of one quantitative measure by another. It does not. The denominator has already failed. Conventional utility cannot provide the dimensionless ratio required to adjust time, while Rasch measurement cannot convert latent-attribute possession into such a proportion. Consequently: $QALY = U \times T$ cannot provide the required measure, and neither aggregation nor subtraction can repair it: $\Delta QALY = QALY_1 - QALY_0$

The cost numerator is constructed differently but encounters the same requirement that measurement begin with a defined attribute. Conventionally:

$C = \Sigma(q_i \times p_i)$

where q_i may represent hospital days, ICU hours, physician visits, doses, laboratory tests or procedures, and p_i represents the corresponding unit price.

The resource quantities are heterogeneous attributes. A hospital day is not a physician visit; an ICU hour is not a dose. Multiplying each by a price and adding the monetary results may provide an accounting total, but it does not create a unidimensional ratio measure of resource utilization. The appropriate therapy-impact claims remain attribute-specific: Δ hospital days; Δ ICU hours; Δ physician visits or Δ doses

There is a further requirement. Even if numerator and denominator were independently ratio measures, this would not automatically make their quotient a meaningful measure. The attributes must be compatible with the derived quantity being claimed. For example:

distance / time = speed

has the coherent dimensional structure:

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

Similarly:

X / Y

for commensurable measures of the same attribute produces:

[A]/[A] = 1

But mathematical divisibility does not mean that any ratio measure can be divided by any other ratio measure to create a scientifically meaningful new attribute. The ICER therefore fails at every level. ΔC is a price-weighted composite of heterogeneous resource inputs; ΔQALY is not a defensible ratio measure; and their quotient does not establish a measured attribute called cost-effectiveness. The reference case cannot be repaired by changing utility instruments, discount rates, model structures or uncertainty methods. Its defining quantitative architecture has failed before those refinements begin.

9. THE COST-EFFECTIVENESS DEBACLE AND NORMAL SCIENCE

The measurement failure exposes a deeper problem. The reference-case cost-effectiveness result is not designed primarily as an empirically evaluable therapy-impact claim. It is constructed to provide a decision variable: modelled ΔC / modelled ΔQALY = cost per QALY. Yet normal quantitative science requires claims that confront observation. They should be capable of evaluation, independent replication and falsification.

Consider a claim that a therapy reduces hospital utilization by two days over six months in a defined population compared with a specified alternative. The attribute and unit are explicit. The claim can be tested prospectively. Another investigator can attempt replication. Most importantly, observations can demonstrate that the claim is false.

The same principle applies to a properly constructed Rasch claim concerning a latent attribute. A claim for improvement in physical-function possession can specify the ruler, population, comparator, timeframe and expected change:

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

with:

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

Again, the claim can confront evidence and fail.

The reference-case claim occupies a fundamentally different position from an ordinary empirical therapy-impact claim. Before questions of replication or falsification even arise, the model assumes that its numerous numerical inputs can legitimately be combined arithmetically. Health-state descriptions, utility scores, transition probabilities, event rates, resource quantities, prices, discount factors, survival estimates and other inputs are brought together within a common mathematical framework. Yet these inputs are not required to demonstrate the ratio-scale properties, dimensional compatibility or admissible transformations necessary for the operations performed upon them.

The reference case therefore fails before any question of empirical evaluation arises. Its inputs are combined arithmetically without first establishing the scale properties, dimensional relationships and permissible transformations required to justify those operations. The resulting QALY denominator is not a defensible ratio measure; the composite cost numerator is not a unidimensional measure of resource utilization; and dividing these two constructs does not create a measured attribute called cost-effectiveness.

The decisive question is therefore not whether ICER = $\Delta C / \Delta QALY$ can subsequently be replicated or falsified. The prior question is more fundamental: does the quotient have any defensible quantitative meaning? If neither its construction nor the relationship between numerator and denominator satisfies the requirements of measurement, the answer is no. Arithmetic can always generate a numerical quotient, but numerical calculability does not establish measurement meaning. A number can be calculated to any desired number of decimal places and remain scientifically meaningless.

This makes the conventional discussion of uncertainty secondary. Deterministic sensitivity analysis can substitute alternative input values, and probabilistic sensitivity analysis can attach distributions to those inputs and generate distributions of model outputs. Neither addresses the prior measurement question. Uncertainty analysis around an inadmissible construct cannot make the construct admissible. Greater sophistication in manipulating the inputs cannot supply the measurement properties that were absent before the model began.

The cost-effectiveness claim therefore fails at a more elementary level than replication or falsification. There is no scientifically established quantity to replicate or falsify or as a decision construct. The reference case produces a number intended to guide a decision, but it has not demonstrated that this number represents a measurable attribute with a coherent quantitative interpretation.

That distinction should be made explicit. Empirical evaluation, replication and falsification are requirements for meaningful scientific claims. They cannot rescue a numerical construct that has already failed the prior requirements of measurement. The first obligation of HTA is therefore to establish the attribute and its lawful measurement. Only after that has been achieved does it become meaningful to ask whether the resulting therapy-impact claim survives empirical evaluation, replication and attempts at falsification.

The implications are difficult to overstate. This is not an inconsequential academic construct confined to methodological exercises. The cost-per-QALY reference case is intended to influence whether a medicine is admitted to a formulary, whether patients obtain access to it, what restrictions are placed upon its use and, directly or indirectly, the price at which it is made available. Decisions with substantial consequences for patients, manufacturers and healthcare systems are therefore being informed by a numerical construct whose claimed quantitative meaning cannot be established under the required standards of measurement.

The extraordinary feature of legacy HTA is not simply that such a construct was devised, but that it acquired institutional authority and has been taught and applied for more than forty years without its defining operations first being required to demonstrate their measurement legitimacy. A decision maker may, of course, make judgements about access and price. What cannot be defended is conferring scientific quantitative authority upon an imaginary cost-effectiveness measure merely because an arithmetic model can generate a number. If the number has no defensible measurement meaning, it has no legitimate claim to guide formulary admission or pricing as a quantitative measure of cost-effectiveness.

10. TRANSITION TO MEASUREMENT IN HEALTH TECHNOLOGY ASSESSMENT

The transition to lawful measurement requires more than demonstrating the measurement failures of contemporary HTA. It requires a replacement framework grounded in representational measurement. The Maimon Research nine-unit transformation program provides that framework [6] . Beginning with the principle that measurement must precede arithmetic, the program progresses through the definition of attributes, scales of measurement, admissible arithmetic, dimensional homogeneity, manifest and latent attributes, linear ratio measurement, Rasch measurement, attribute possession, and the construction of empirically evaluable therapy-impact claims. The purpose is not to add measurement to the existing HTA paradigm while leaving its analytical structure intact. It is to replace that structure with one in which quantitative claims must first satisfy the requirements of lawful measurement and can then be evaluated, falsified and replicated.

At the same time, the nine units provide the substantive foundation for a graduate course in lawful measurement for HTA. Comprising approximately 45,000 words of structured instructional material, the program provides sufficient depth to move beyond a brief introduction to measurement theory and support a systematic transition from the legacy HTA framework to one grounded in representational measurement. Its sequence can be translated directly into a graduate curriculum in which measurement provides the organizing scientific foundation rather than appearing as an additional methodological topic downstream of economic evaluation. Faculty are therefore provided not simply with a critique of existing methods, but with the substantive material and sequence required to construct an alternative course.

As far as can presently be ascertained, no comparable program has been identified internationally that is designed specifically to support the transition from legacy HTA to an HTA grounded in the requirements of lawful measurement. That distinction is important. The program is not intended to add a measurement component to contemporary HTA while leaving utilities, QALYs and reference-case modelling undisturbed. It provides a structured pathway for reconsidering the scientific foundations of the discipline itself and for rebuilding quantitative therapy evaluation around measurement before arithmetic, ratio measurement and claims capable of empirical evaluation, replication and falsification.

The resulting course would replace the conventional emphasis on utilities, mapping, QALYs and reference-case modelling with an alternative scientific structure: defining the target attribute, determining whether it is manifest or latent, constructing the appropriate linear ratio or Rasch measure, and using that measure to formulate empirically evaluable therapy-impact claims. The nine-unit program therefore provides not simply an argument for abandoning the legacy framework, but the content and organizing structure from which its educational replacement can be built. The transition is consequently from one knowledge architecture to another: from arithmetic preceding measurement to measurement determining what quantitative claims and arithmetic are scientifically permissible.

Faculty and practitioners who enroll will have continuous access to question-and-answer support throughout the program, providing assistance as they work through the implications of representational measurement for teaching, research and HTA practice. This becomes particularly important where faculty are using the program as the foundation for curriculum development. Questions concerning course structure, individual topics, measurement concepts, examples, assignments and their application to existing HTA material can be addressed as they arise. Participants are therefore not expected to complete the nine units and then independently determine how an unfamiliar scientific framework should be translated into teaching or practice. Continuing support makes the movement from understanding to implementation an integral part of the transition.

Additional faculty and institutional support can also be arranged on an agreed basis. This may include Zoom seminars and workshops, assistance with graduate-course and curriculum redesign, development of teaching materials, recommended references and readings, and guest lectures addressing particular aspects of representational measurement and its application to HTA. Support can be tailored to individual faculty members, courses, departments, research centers or professional groups.

The nine-unit program should therefore be understood as the starting point for a managed transition to lawful measurement. It provides the replacement scientific framework, the substantive elements required to construct a graduate course around that framework, and continuing access to the expertise and educational resources required for implementation. The objective is not simply to demonstrate why contemporary HTA must change, but to provide faculty and practitioners with the foundations and practical means to build, teach and apply its replacement. Once the measurement failure of legacy HTA is recognized, there is no scientifically defensible route back.

11. CONCLUSION

The case for abandoning legacy HTA is impossible to avoid. Its defining quantitative framework was constructed without first demonstrating the measurement properties required for its arithmetic. Utilities, QALYs, incremental QALYs and cost-per-QALY claims cannot acquire scientific legitimacy through longevity, professional consensus, methodological guidelines or institutional acceptance. Once the required measurement sequence is applied, the conclusion follows directly: measurement must precede arithmetic, and quantitative constructs that cannot satisfy the measurement requirements necessary for their claimed operations must be abandoned.

There is no scientifically defensible intermediate position [7] The issue is not whether legacy HTA might benefit from refinement, better preference instruments, more sophisticated modelling or improved uncertainty analysis. These operate downstream of the measurement failure. Nor can continued use be justified on the grounds that decision makers require an established framework.

The necessity of making a decision cannot establish the scientific legitimacy of the quantities used to make it. If the measurement critique is wrong, its errors should be identified and demonstrated. If it is correct, continued use of the failed constructs cannot be defended simply because no institution wishes to abandon them.

This creates a particular responsibility for universities and faculty. No faculty member is required to accept the case for abandonment. Scientific disagreement is legitimate. A faculty member or program may decide to continue teaching utilities, QALYs, cost-per-QALY analysis and the legacy reference case. What is no longer defensible is presenting those methods to students as settled quantitative science without informing them that their measurement foundations have been formally challenged.

Students should therefore be told explicitly that there is an alternative scientific position: that legacy HTA fails the requirements of representational measurement and that abandonment rather than methodological modification is required. They should be introduced to the measurement principles necessary to understand the challenge and allowed to judge the arguments for themselves.

Faculty who reject the challenge assume a corresponding scientific responsibility. They should demonstrate where the measurement argument fails. If utility is a legitimate ratio measure, the required measurement properties should be established. If its proportional application to time is admissible, that operation should be defended. If the QALY is a legitimate quantitative measure, its measurement structure should be demonstrated. If incremental QALYs can support a cost-effectiveness ratio, that claim should similarly be established.

Appeals to established practice, professional authority, publication history or consensus are not proofs.

The choice is therefore straightforward. Faculty can accept the measurement failure and begin the transition to measurement-first HTA. Alternatively, they can reject that conclusion, continue teaching legacy HTA and demonstrate scientifically why the measurement critique is wrong. What they cannot reasonably do is ignore the challenge and continue teaching another generation as though it had never been made.

The burden has shifted. Legacy HTA is no longer entitled to acceptance because it is the established framework. Once its measurement foundations have been challenged, continued transmission requires a scientific defense. Students are entitled both to know that challenge exists and to see the proof upon which any decision to reject it rests.

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 (4th Ed). New York: Routledge, 2021

[5] Langley P. Legacy HTA: ISPOR – A forensic evaluation of representational measurement and the impossible QALY. Logit Working Paper No 2005. https://maimonresearch.com/logit-working-paper-no-2005-september-2026/

[6] Maimon Research. HTA Reconstruction. https://maimonresearch.com/hta-reconstruction-program-and-fees/

[7] Langley P. Legacy HTA: Is this the worst we could have done? Logit Working Paper No 7987 https://maimonresearch.com/logit-working-paper-no-7987-august-2026/