MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT AND HEALTH TECHNOLGY ASSESSMENT

NORTH CAROLINA: LEGACY HTA - FORTY YEARS OF NUMERICAL STORYTELLING

The Absence of Representational Measurement in Teaching and Research

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

LOGIT WORKING PAPER No 6758 AUGUST 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

Health technology assessment (HTA) has for more than four decades relied upon utilities, quality-adjusted life years (QALYs), incremental cost-effectiveness ratios (ICERs) and reference-case simulation as foundations for quantitative decision making. This paper challenges that framework by applying a prior requirement of quantitative science: measurement must begin with a defined unidimensional attribute whose empirical structure satisfies the axioms required for the claimed measurement scale. Arithmetic follows measurement; it cannot create it.

Two measurement routes are distinguished. Manifest attributes satisfying the required axioms support linear ratio measurement, with permissible transformations of the form X' = aX. Latent attributes require a different solution, represented by Rasch measurement, where changes in attribute possession are expressed on a logit continuum and have an invariant ratio-of-odds interpretation. These structures impose strict requirements on multiplication, division and dimensional homogeneity.

The QALY fails both routes. Conventional utility is generated from a weighted multiattribute health-state description rather than a demonstrated unidimensional ratio attribute. It can assume negative values and is nevertheless treated as a proportional adjustment to time. A legitimate proportional adjustment must instead be a dimensionless ratio D = X/Y, constructed from commensurable ratio measures of the same attribute. Utility has no such construction and therefore cannot legitimately enter QALY = U × T. Rasch measurement provides no escape: OR = exp(Δθ) represents a ratio of odds, not a linear proportion of latent-attribute possession suitable for adjusting time.

The consequences extend to ICER = ΔC/ΔQALY. The QALY denominator is inadmissible, while the cost numerator aggregates heterogeneous resource attributes through prices rather than measuring a single resource-utilization attribute. Their quotient cannot establish a measured attribute called cost-effectiveness.

Legacy HTA should therefore be abandoned rather than repaired. Its replacement should generate attribute-specific, empirically evaluable, replicable and falsifiable therapy-impact claims grounded in representational measurement and the standards of normal science.

INTRODUCTION

This paper presents, as far as can be judged, a unique contribution to the critique of health technology assessment (HTA). For the first time, the arithmetic operations that define the analytical framework of legacy HTA are formally assessed against the scales of measurement and the required axioms of representational measurement. After more than forty years in which this framework has been taught, applied and institutionalized largely without challenge to these underlying measurement requirements, the question asked here is elementary but decisive: After more than forty years in which this framework has been taught, applied and institutionalized largely without challenge to its measurement foundations, the question asked here is elementary but decisive: can the analytical framework of legacy HTA withstand formal assessment against the required axioms of representational measurement and the scales of measurement [1] [2] ?

This question takes precedence over the familiarity, institutional acceptance or mathematical sophistication of the methods involved. Utilities, QALYs, incremental cost-effectiveness ratios and reference-case simulation have become embedded in HTA education and reimbursement decision making. Yet widespread application cannot establish measurement legitimacy. The relevant question is whether the numbers being manipulated represent defined attributes, whether those attributes possess the empirical structure required for the claimed scales, and whether multiplication, division, aggregation and subtraction are admissible operations for those scales. The standards against which legacy HTA is assessed here are not standards devised retrospectively to challenge it. The principles of representational measurement and the requirements of scale theory were established before the QALY/reference-case framework became institutionalized.

Quantitative science begins with an uncompromising requirement: before numbers can support a scientific claim, there must be a defined attribute capable of lawful measurement. Measurement does not begin with a questionnaire, a multiattribute health-state description, a score, a utility algorithm or a model. Nor can an attribute simply be nominated and assigned numbers. Its empirical structure must satisfy the requirements of representational measurement for the scale being claimed. The sequence is therefore: defined unidimensional attribute → axiomatic requirements → measurement scale → admissible arithmetic → quantitative claim

This paper applies that sequence to legacy HTA and distinguishes two routes to quantitative measurement: linear ratio measurement for appropriate manifest attributes and Rasch logit measurement for latent attributes. The decisive test is the quality-adjusted life year because it is the defining quantitative construct connecting multiattribute health-state modelling to the cost-effectiveness reference case: QALY = U × T where U is the utility score generated from the weighted multiattribute health-state description and T is the duration of time associated with that state. Utility is intended to act as a proportional adjustment to ratio-scaled time. The resulting QALYs are then aggregated and compared to generate ΔQALY, the denominator of the conventional incremental cost-effectiveness ratio: ICER = ΔC / ΔQALY.

Despite more than forty years of HTA application, the legacy analytical framework remains deeply entrenched. This is evident not only in the present analysis but across the preceding series of global logit interrogations of HTA agencies, academic research centers, curricula and program content. Taken together, these analyses reveal a striking and persistent fidelity to measurement inversion, curriculum inversion and program-content inversion: arithmetic and modelling precede measurement; established HTA methods precede instruction in the scientific conditions that would make those methods admissible; and programs comprehensively represent the techniques of legacy HTA while largely excluding the representational-measurement principles against which those techniques should first be judged. The persistence of this pattern across countries and institutions suggests that the problem is not an isolated methodological oversight. It is embedded in the intellectual architecture through which legacy HTA has been taught, reproduced and defended.

The present analysis takes the next and decisive step by applying the requirements of ratio measurement directly to that architecture. The conclusion is categorical. Conventional utility cannot provide the dimensionless ratio required to adjust time, while Rasch logit measurement cannot provide the linear proportional adjustment required for latent attributes. The QALY therefore cannot be repaired. Its failure removes the denominator of the conventional cost-effectiveness claim, while the composite cost numerator presents an independent measurement failure. The implications are difficult to overstate. For more than forty years, thousands of papers, reports, models, guidelines and teaching programs have been constructed around an analytical framework whose defining quantitative operations fail when formally assessed against the required standards of measurement. As far as can be judged from the legacy literature, curricula and institutional knowledge bases examined, this fundamental problem was neither identified nor treated as a threat to the validity of the enterprise.

Instead, increasingly elaborate models were constructed, uncertainty was quantified, reference cases were standardized and generations of students were trained to reproduce the framework. Yet the logically prior questions were effectively bypassed: what is the attribute; is it unidimensional; does its empirical structure satisfy the required axioms of representational measurement; what scale has been achieved; and does that scale permit the arithmetic subsequently performed? The extraordinary feature of legacy HTA is therefore not simply that mistakes were made. It is that methodological sophistication accumulated on top of measurement foundations that had never been secured.

In that sense, forty years of methodological development have been devoted to refining an analytical framework whose defining constructs cannot survive formal measurement scrutiny. The problem is not an imperfection at the margins, nor an anomaly that can be corrected by another utility algorithm, a revised reference case, different assumptions or a more sophisticated simulation. It reaches the foundations themselves: no unidimensional ratio attribute → no dimensionless utility ratio → no lawful proportional adjustment to time → no QALY → no ΔQALY → no cost-per-QALY reference case

The independent failure of the composite cost numerator closes the remaining route to the claimed cost-effectiveness quantity. Mathematical manipulation can produce a numerical quotient, but it cannot create the measurement meaning that its components and their relationship do not possess.

The consequence is therefore not reform, refinement or methodological accommodation. It is rejection. If the present analysis stands, the extraordinary conclusion is that an analytical framework taught and applied internationally for four decades fails measurement requirements that should have been addressed before the framework was adopted. The enormous intellectual, educational and institutional investment subsequently made in legacy HTA cannot constitute a scientific defense of it. Nor can forty years of acceptance compensate for forty years in which the prior requirements of measurement were overlooked.

The responsibility now is to stop reproducing that failure. Legacy HTA should no longer be taught to another generation of students as though its quantitative legitimacy were established. Its replacement must begin with defined unidimensional attributes, the axioms of representational measurement, appropriate linear or Rasch logit ratio measurement and only then admissible arithmetic. The endpoint should not be an imaginary universal cost-effectiveness number, but attribute-specific therapy-impact claims capable of empirical evaluation, replication and falsification and, through that process, contributing to the evolution of objective knowledge.

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.

Ratio Measurement

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 supports the required axioms and whose numerical representation possesses the invariance properties of the relevant ratio structure. Ratio status must be demonstrated before multiplication, division or proportional interpretation is attempted; it cannot be assumed because an analytical framework requires those operations.

2. 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 [3]. 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).

3. 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:

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

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

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

and a half-logit change gives:

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

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.

4. 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:

$D = \text{time in A} / \text{total time}$

so:

$D = 18 \text{ hours} / 24 \text{ hours} = 0.75$

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.

5. 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:

$0 \leq X \leq Y$

then:

$0 \leq D \leq 1$

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.

6. 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:

$$\theta_1 = 1 \text{ logit}$$

and:

$$\theta_2 = 2 \text{ logits}$$

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.

7. 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.

8. 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 = ΔC / ΔQALY 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.

9. TRANSITION TO LAWFUL MEASUREMENT

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 [4]. 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 progressing from measurement preceding arithmetic, attributes and scales of measurement through admissible arithmetic and dimensional homogeneity to manifest and latent attributes, linear ratio measurement, Rasch measurement, attribute possession and the construction of empirically evaluable therapy-impact claims.

As far as can presently be ascertained, this is the only program available internationally 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 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 or Rasch ratio measure, and using that measure to formulate empirically evaluable claims. The nine-unit program therefore provides not simply a critique of the legacy framework, but the content and organizing structure from which its educational replacement can be built.

Faculty and practitioners who enroll will have access to a 24/7 question-and-answer support service, providing continuing 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 is intended to make 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. There is no going back.

The issue facing graduate HTA education in North Carolina is no longer whether additional attention might usefully be given to measurement. The accumulated evidence from measurement inversion, curriculum inversion and program-content inversion indicates a more fundamental problem: representational measurement has not provided the scientific foundation for the quantitative methods taught and applied within HTA. The consequence has been the transmission across successive generations of students of an analytical framework in which arithmetic has been allowed to precede measurement.

The challenge is therefore one of transition rather than curriculum expansion. Faculty must acquire and integrate a body of measurement knowledge that may have been absent from their own graduate education. Students, in turn, should expect more than technical proficiency in inherited quantitative procedures. They should leave a graduate program capable of identifying an attribute, establishing its measurement requirements, selecting the appropriate measurement pathway and constructing a therapy-impact claim that is empirically evaluable, replicable and capable of falsification.

North Carolina universities now have an opportunity to define what graduate education in HTA should mean in a post-legacy environment. The state has a substantial and diverse academic health-sciences infrastructure spanning pharmacy, medicine, public health, health economics, pharmacoeconomics, outcomes research, comparative effectiveness research, measurement science and health-services research. Its universities and affiliated research centers therefore face a consequential choice. They can continue to reproduce an established analytical framework in which the requirements of representational measurement remain largely absent, or they can begin the transition toward a curriculum in which measurement provides the organizing scientific foundation for quantitative therapy evaluation.

This responsibility has become more immediate because the measurement failure is no longer hidden within the inherited HTA framework. The accumulated interrogations of HTA knowledge bases, curricula and program content, together with the analysis presented here, have exposed and documented the recurring absence of representational measurement. Faculty may reasonably argue that they were themselves educated within this framework and were not introduced to the measurement standards now being advanced. That explains how the legacy was transmitted; it cannot justify transmitting it again. Once the measurement failure has been identified, another generation of North Carolina students cannot simply be taught the same framework as though the problem had never been exposed.

The standard should therefore not be whether a North Carolina program reproduces established professional practice or prepares students to apply methods that have been accepted for decades. Institutional acceptance, methodological sophistication and widespread use do not establish scientific validity. Nor does the reputation of a North Carolina university transform an inadmissible quantitative operation into an admissible one. The appropriate standard is whether the program equips students to understand measurement, determine what arithmetic is lawful, construct defensible quantitative claims and thereby contribute to normal science and the evolution of objective knowledge.

For the prospective student investing substantial time, money and career opportunity, the question is correspondingly simple: does the North Carolina graduate program being considered provide this scientific education? Does it teach why measurement must precede arithmetic? Does it establish the role of ratio measurement? Does it distinguish the measurement requirements of manifest and latent attributes? Does it provide a pathway from attribute definition through lawful measurement to empirically evaluable, replicable and falsifiable therapy-impact claims?

If it cannot demonstrate that pathway, the final question is unavoidable: why enroll?

For North Carolina universities, the corresponding question is even more consequential. The state possesses major academic institutions that educate future pharmacists, health economists, outcomes researchers, measurement scientists, clinicians and health-policy professionals and contribute directly to the methods subsequently applied in healthcare decision making. This gives North Carolina a particular responsibility: its academic institutions do not merely transmit established methods; they help determine what the wider health-sciences community accepts as legitimate quantitative practice. Once the measurement failure has been exposed, another generation of students cannot reasonably be taught the same analytical framework without confronting that failure. The next generation of HTA should not attempt to repair the QALY. It should abandon the legacy framework and return quantitative assessment to the standards of normal science.

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):677–680

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

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

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