

MAIMON RESEARCH LLC

**ARTIFICIAL INTELLIGENCE LARGE LANGUAGE
MODEL INTERROGATION**



**REPRESENTATIONAL MEASUREMENT FAILURE IN
HEALTH TECHNOLOGY ASSESSMENT**

**THE QALY: FORTY YEARS OF ARITHMETIC
WITHOUT MEASUREMENT**

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

LOGIT WORKING PAPER No 1510 OCTOBER 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

This paper challenges the scientific legitimacy of the quality-adjusted life year (QALY) and the global health technology assessment (HTA) curriculum built around cost–utility analysis and the reference case. For forty years, HTA has institutionalized arithmetic without demonstrating the measurement properties required to justify its central quantitative claims. The QALY fails two fundamental requirements of representational measurement: dimensional homogeneity and commensurability. Duration is a ratio measure within time space, whereas conventional health-state utilities are numerical preference assignments without demonstrated unidimensional ratio-scale properties. Even if preferences could be established as ratio measures, their proportional relationships would not automatically constitute proportional adjustments to duration. An independently demonstrated mapping between preference space and time space would remain essential. The QALY equation assumes precisely what measurement science requires it to establish.

The consequences extend beyond methodological error to graduate education, where successive generations have been trained to manipulate utilities, QALYs and cost-effectiveness ratios without first establishing lawful measurement. The paper calls for abandoning the QALY curriculum and reconstructing HTA around representational measurement. A nine-unit Maimon Research curriculum provides an immediately available alternative, emphasizing attribute definition, manifest and latent measurement, admissible arithmetic, empirical testing, replication and falsification. Measurement must precede arithmetic, not follow it.

INTRODUCTION

The global health technology assessment (HTA) curriculum, with its commitment to the quality-adjusted life year (QALY) and the reference case, has institutionalized four decades of arithmetic without measurement. What is presented to successive generations of graduate students as rigorous quantitative science is, at its foundation, a collection of numerical procedures that fail to satisfy the elementary requirements of representational measurement. The resulting cost-per-QALY claims are not merely questionable estimates. They are calculations whose purported measurement meaning has never been established.

Had the original architects of the QALY understood, or even seriously considered, the axioms of representational measurement, the proposal should have been rejected at inception^{1 2}. The initial mistake was to suppose that multidimensional descriptions of health states could be transformed into single numerical measures of value through preference elicitation. Neither the time trade-off nor the standard gamble establishes the required unidimensional ratio measure. The subsequent utility algorithms compound this failure by assigning arithmetic significance to numbers without demonstrating the measurement properties necessary to support that arithmetic.

Yet an even more fundamental objection has been overlooked. The QALY fails two indispensable requirements: dimensional homogeneity and commensurability. Duration belongs to time space and possesses ratio measurement properties. Health-state preferences occupy an entirely different conceptual space, with no demonstrated ratio measurement structure and no established mapping

into proportional adjustments of duration. The multiplication of these quantities cannot be legitimized by convention, statistical estimation, professional consensus or institutional endorsement.

The implications are devastating. NICE, PBAC, ISPOR and the wider HTA establishment have promoted an analytical framework in which measurement has been displaced by numerical storytelling. The reference case does not rescue the QALY; it institutionalizes its failure. Forty years of methodological refinement cannot repair an arithmetic operation that was never justified in the first place. The task is no longer to improve the QALY. It is to explain why it should never have been accepted as a measure.

1. SCIENCE DEMANDS REPRESENTATIONAL MEASUREMENT

Science demands measurement before arithmetic. This is not an optional methodological preference but a foundational requirement of quantitative inquiry formalized some 80 years ago. Numerical assignments become measures only when their properties correspond to an empirically demonstrable structure of the attribute being measured. Representational measurement theory establishes the conditions under which observations can be represented numerically and determines which arithmetic operations are admissible. Without those conditions, numbers remain labels, rankings or arbitrary assignments. Their manipulation may produce mathematically consistent results, but mathematical consistency does not establish empirical meaning.

The distinction is fundamental. Nominal scales classify objects; ordinal scales establish order; interval scales support meaningful comparisons of differences but lack a non-arbitrary zero; ratio scales possess equal units and a non-arbitrary zero, permitting meaningful ratios of magnitudes. Each scale admits a restricted class of transformations. For ratio scales, admissible transformations are similarity transformations of the form $x' = ax$, where $a > 0$. Arithmetic cannot be selected independently of these properties. The mere availability of numerical values does not authorize their multiplication, division or interpretation as proportions.

Representational measurement requires more than assigning scale labels. The proposed attribute must be identified, its unidimensionality established and the empirical relations supporting its numerical representation demonstrated. Order, additivity, invariance, cancellation, solvability and the relevant structural conditions must be addressed where applicable. For manifest attributes, such as duration, linear ratio measurement provides the appropriate foundation. For latent attributes, including patient-reported experiences and preferences, a defensible measurement model must establish the required structure. Rasch measurement offers a pathway through conjoint measurement of persons and items, with invariance and specific objectivity subject to empirical testing. Neither summated ordinal responses nor regression-generated preference weights acquire ratio properties simply through statistical manipulation.

Two additional requirements are decisive for the QALY: dimensional homogeneity and commensurability. Dimensional homogeneity requires quantitative expressions to preserve the dimensional meaning of their components. Addition and subtraction require compatible dimensions, while multiplication and division must produce quantities whose resulting dimensions are explicitly interpretable. A dimensionless multiplier applied to duration must have a

demonstrated meaning as the proportional adjustment it purports to impose. Being numerically dimensionless is not sufficient.

Commensurability imposes a separate requirement. Where a ratio is intended to represent a proportion of an attribute, its numerator and denominator must be measures of that same attribute. A legitimate proportional adjustment to duration can be expressed as T_1/T_2 , where both quantities belong to time space, dimensional class [T]. The dimensions cancel, producing a dimensionless factor with an empirically meaningful interpretation. A ratio of preference magnitudes, even if lawfully measured, does not thereby become a proportion of time. Any mapping from preference space to time space requires independent demonstration.

This is precisely where conventional HTA abandons measurement. The QALY equation $Q = U \times T$ treats a health-state utility as though it were an established proportional adjustment to duration. No demonstrated representational structure or commensurable mapping warrants that interpretation. The reference case simply assumes what measurement science requires it to prove.

The scientific obligation is therefore unambiguous: establish the attribute, demonstrate the scale, justify the arithmetic and only then calculate. Legacy HTA has reversed that sequence. It begins with the desired calculation and treats numerical convention as a substitute for measurement. The consequence is not an imperfect science awaiting methodological refinement, but an analytical framework whose foundational quantitative claims remain unsubstantiated.

2. THE IMPOSSIBLE QALY

Once arithmetic is divorced from the requirements of lawful representational measurement, it becomes open season on the manipulation of numbers. Nowhere is this more evident than in the construction of the QALY. The arithmetic is trivial. A utility score, conventionally anchored at zero and unity, is multiplied by survival duration to produce a number labelled a quality-adjusted life year. The utility is treated as a proportional adjustment factor that discounts duration according to the assigned value of a health state.

The apparent simplicity of this operation conceals its fundamental measurement failure. Assigning a utility score between zero and one does not establish that it represents a lawful proportion of duration. Nor does multiplication by time demonstrate that the resulting number measures adjusted survival. The arithmetic creates a number; it does not establish the existence of the quantity the number purports to represent.

The publication of tens of thousands of papers supporting the QALY and its application to therapy-impact claims is no defense of the QALY as a legitimate product of representational measurement. Far from it. The sheer volume of publications raises a far more disturbing question: how has a quantitative framework achieved such widespread acceptance without first demonstrating that its central outcome satisfies the required standards of representational measurement?

Publication, replication of numerical procedures and institutional endorsement cannot establish measurement legitimacy. If the QALY fails the requirements of representational measurement, the accumulation of papers employing it merely demonstrates the extraordinary extent to which an

unestablished measurement construction has become embedded in global HTA. The fact that a simple proof can make clear the failure of the QALY is a question those who have promoted legacy HTA for over 40 years will have to come to terms with. The QALY is the defining example of the failure of legacy HTA: forty years of numerical elaboration without first establishing the legitimacy of the measure upon which the entire QALY-dependent framework rests.

The QALY is not simply a convenient summary statistic. It makes a specific quantitative claim: that a period of time can be proportionally adjusted by the value assigned to a health state to create a new quantity, quality-adjusted time. This claim requires more than numbers. It requires measurement.

The critical distinction is between assigning a numerical value to a health state and demonstrating that the value possesses the measurement properties required to adjust time proportionally. A preference score can rank health states, summarize responses or represent the output of a statistical valuation model without possessing the ratio properties required for multiplication.

The measurement question is therefore prior to the familiar QALY calculation. Before U can multiply T , it must be established what U measures, what empirical structure it represents, what scale properties it possesses and why its numerical value can legitimately be interpreted as a proportion of duration. Without that proof, the arithmetic has no demonstrated quantitative meaning.

The conventional QALY combines duration with a numerical value intended to represent the quality of the health state experienced during that duration:

$$Q = U \times T$$

The interpretation is explicitly proportional. If $U = 1$, duration is unadjusted. If $U = 0.5$, one year in the health state is treated as one-half of a quality-adjusted year. The value U must therefore perform the role of a proportional adjustment factor.

Duration provides a simple demonstration of what a lawful proportional adjustment requires. Consider two durations T_1 and T_2 . Because both quantities are measures of duration, they belong to the same dimensional class, denoted by $[T]$, where the square brackets $[]$ indicate the dimension of the quantity rather than its numerical value or unit of measurement. Thus, $[T]$ denotes the dimension of time, irrespective of whether duration is expressed in seconds, days, months or years. In dimensional terms:

$$[A_t] = [T_1]/[T_2] = [T]/[T] = 1$$

The result is dimensionless because the time dimensions cancel. The notation $[A_t] = 1$ therefore means that A_t has no physical dimension. It does not, by itself, establish that A_t has proportional meaning.

If $T_1 = 6$ months and $T_2 = 10$ months:

$$A_t = 6/10 = 0.6$$

The interpretation is clear. Six months is 60% of ten months. The result has proportional meaning because T_1 and T_2 are commensurable ratio measures of the same attribute, duration.

The result is also invariant under a permissible change of unit. If duration is transformed according to:

$$T' = aT$$

then:

$$aT_1/aT_2 = T_1/T_2$$

Changing from months to years does not alter the ratio.

This demonstrates why describing U as dimensionless is insufficient. A dimensionless ratio has proportional meaning only when its numerator and denominator are commensurable ratio measures of the same attribute on a common measurement scale. Dimensional cancellation alone does not establish proportional meaning. If $U = 0.6$ is to mean that duration should be reduced to 60% of its original value, the corresponding question is unavoidable:

What two commensurable ratio measures of the same attribute generate the proportion represented by U ?

Conventional health-state preference systems do not provide such a construction. Their values typically originate in preferences expressed for multidimensional health states. Different aspects of health are described, valued and combined to generate a single numerical score. The resulting number may be useful as an ordinal index of preference, but assigning it a numerical value does not establish that it is a unidimensional ratio measure. The utility or preference score has no independently established interval- or ratio-scale measurement properties. It is a constructed numerical index, not a demonstrated quantitative measure of preference. At best it is a composite ordinal scale.

Anchoring a preference scale at zero and one does not resolve the problem. Reference points assigned by convention are not equivalent to demonstrating a non-arbitrary zero, equal units and meaningful ratios. The existence of negative utility values for health states considered worse than death creates a further difficulty. Negative values are incompatible with interpreting zero as the absence of a nonnegative magnitude on a linear ratio scale. The conventional utility scale therefore lacks the demonstrated structure required for meaningful proportional comparisons.

The QALY consequently encounters its first measurement failure before multiplication even begins. Utility has not been demonstrated to possess the unidimensional ratio-scale properties required for its interpretation as a proportion. Without these properties, U cannot legitimately be treated as the basis for a measured proportional adjustment factor for duration.

There is also a second, independent failure.

Suppose, purely for argument, that two preference quantities P_1 and P_2 were demonstrated to be commensurable ratio measures of the same preference attribute. We could then construct:

$$A_p = P_1/P_2$$

and:

$$[A_p] = [P]/[P] = 1$$

A_p could then have legitimate proportional meaning within preference space. That would still not make it a proportional measure of time.

The proportional adjustment required within time space remains:

$$A_t = T_1/T_2$$

A_p and A_t may both be dimensionless, but they arise from different attributes and different measurement spaces. Their common absence of physical dimensions does not establish their equivalence.

A separate mapping would therefore be required:

$$f: A_p \rightarrow A_t$$

This mapping would have to demonstrate why a particular proportional relationship in preference space corresponds to the same proportional adjustment in time space.

The conventional QALY does not demonstrate this relationship. It simply substitutes U into:

$$Q = U \times T$$

The multiplication assumes the mapping that measurement must first establish. The proportional relationship between preference and duration required to justify its calculation has never been demonstrated

The QALY therefore fails at two sequential points. First, conventional health-state preference values have not been demonstrated to possess the unidimensional ratio properties required for proportional interpretation. Second, even if such properties were provisionally granted in preference space, an independent mapping would still be required before that preference proportion could lawfully adjust ratio-measured duration. This has never been demonstrated.

The bottom line is clear: the QALY has no independent existence as an analytical construct.

3. ABANDONING THE QALY CURRICULUM

The traditional graduate curriculum in health technology assessment must be abandoned. Its central commitment to the QALY, cost–utility analysis and the reference case has produced generations of analysts trained in arithmetic without an understanding of the measurement requirements that make arithmetic scientifically admissible. Students are taught to calculate utilities, construct QALYs, estimate incremental cost-effectiveness ratios, populate probabilistic models and interpret cost-effectiveness acceptability curves. They are not required first to establish that the quantities entering these calculations are measures. Procedural competence has been substituted for measurement competence.

This is not a minor omission that can be corrected by adding a lecture on measurement theory. It is a fundamental failure in the intellectual architecture of the curriculum. The QALY is not an otherwise sound measure requiring improved estimation. Its foundational arithmetic lacks the demonstrated dimensional homogeneity and commensurability necessary to sustain its interpretation. Teaching increasingly elaborate methods for manipulating QALYs merely compounds the original error. Sensitivity analysis cannot rescue an inadmissible multiplication. Probabilistic modeling cannot confer measurement properties on arbitrary numerical assignments. A reference case cannot transform an impossible quantity into a scientific one through institutional endorsement.

The responsibility of university faculty is consequently unavoidable. Whether those teaching legacy HTA are familiar with representational measurement is an open empirical question. The absence of explicit measurement requirements from conventional course descriptions raises serious concerns about what faculty understand, what they teach and what students are expected to know. Faculty cannot indefinitely shelter behind established textbooks, professional guidelines or the assertion that these methods are universally accepted. Consensus is not evidence of measurement validity. Forty years of repetition cannot establish what was never demonstrated.

Universities have a duty to distinguish scientific measurement from numerical storytelling. If faculty elect to continue teaching the QALY and reference case, they should inform students explicitly that the measurement foundations of these procedures remain unestablished and that their arithmetic interpretation is contested on fundamental axiomatic grounds. They should not present cost-per-QALY results as though they were empirically warranted measurements of therapeutic value. Students deserve to know that mastering a computational procedure is not equivalent to acquiring the ability to make a defensible quantitative claim. Students will not wish to hear that their graduate course is nothing more than an exercise in numerical storytelling.

The consequences extend well beyond the classroom. Graduates carry these fictional methods into pharmaceutical companies, HTA agencies, health systems, consultancies and academic research. They become reviewers, decision advisers, journal referees and eventually teachers themselves. Each generation reproduces the same methodological assumptions, reinforcing a closed epistemic system in which professional acceptance substitutes for independent scientific validation. The curriculum has become a mechanism for transmitting measurement failure rather than correcting it.

The required response is reconstruction, not cosmetic reform. Graduate HTA education must begin with defined attributes, representational axioms, scale construction, admissible transformations, dimensional homogeneity and commensurability. It must distinguish manifest from latent measurement and require therapy-impact claims to be empirically testable, replicable and capable of falsification. The QALY and its reference case should be retained only as historical examples of what happens when arithmetic precedes measurement.

The central question for universities is no longer how to teach the QALY more effectively. It is why they continue to teach as science a quantitative framework whose foundational measurement claims have never been established or even recognized.

CONCLUSION: THE TRANSITION TO LAWFUL MEASUREMENT

The QALY represents forty years of arithmetic without measurement. Its failure is not a matter of imperfect preference elicitation, inadequate statistical estimation or uncertainty in model parameters. It is a foundational failure to satisfy the requirements of dimensional homogeneity and commensurability. No refinement of the reference case can overcome these defects. The QALY must be abandoned as a measure of therapeutic benefit, together with the cost-per-QALY framework built upon it.

The solution is already available. Maimon Research has developed a nine-unit curriculum program of some 60,000 words for reconstructing HTA around the requirements of representational measurement ³. The program replaces numerical storytelling with a scientific framework that begins with attribute definition and proceeds through manifest and latent measurement, admissible arithmetic, dimensional homogeneity, commensurability, empirical protocols, replication and falsification. It provides an immediately available curriculum foundation for graduate education, faculty development and the assessment of therapy-impact claims.

For institutions and faculty subscribing to the nine-unit curriculum package, Maimon Research offers continuing support through online questions and answers, on-line faculty seminars on request, and supplementary reference materials. These include resources on measurement inversion and related topics available through the Maimon Research website, together with findings from more than 140 interrogations of HTA knowledge bases across 27 countries.

Universities and HTA agencies therefore face a choice. They can continue teaching and endorsing calculations whose measurement foundations have never been established, or they can begin the transition to lawful measurement now. There is no scientific justification for another generation of students being trained in methods that mistake numerical manipulation for measurement.

The QALY curriculum has failed. Its replacement is not a future aspiration. The means of reconstruction exist today.

ACKNOWLEDGEMENT

I acknowledge that I have used OpenAI technologies, including the large language model, to assist in the development of this work. All final decisions, interpretations, and responsibilities for the content rest solely with me

ABOUT THE AUTHOR

Dr Paul Langley is a health economist. Originally from the UK, Dr Langley completed his graduate studies in Canada and has taught in the UK, Canada, Australia and the United States. After more than 20 years teaching at university in Australia, Dr Langley moved to take up academic positions in the United States. He now lives in Tucson, Arizona, to enjoy the Catalina Mountains, the Sonoran Desert, the absence of mosquitoes and the swimming pool.

REFERENCES

¹ 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. Volume I: Additive and Polynomial Representations. Academic Press, New York. 1971

³ Maimon Research. *Reconstructing HTA Programs: Measurement before Arithmetic*.
<https://maimonresearch.com/hta-reconstruction-program-and-fees/>