

MAIMON RESEARCH LLC
**ARTIFICIAL INTELLIGENCE LARGE LANGUAGE
MODEL INTERROGATION**



**REPRESENTATIONAL MEASUREMENT FAILURE IN
HEALTH TECHNOLOGY ASSESSMENT**
**THE THEORY OF ALLOWABLE TRANSFORMATIONS
AND THE QALY**

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

LOGIT WORKING PAPER No 887 JUNE 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

For almost five decades the quality-adjusted life year (QALY) has served as the dominant outcome metric in health technology assessment (HTA), providing the foundation for cost-effectiveness analysis, reference-case simulation modelling and reimbursement decision-making. Despite its widespread acceptance, little attention has been given to the measurement requirements necessary to support the arithmetic operation at its core. The QALY is constructed by multiplying survival time by a utility score. While survival time is an accepted ratio measure, the measurement status of the utility score remains unresolved. This paper argues that the central weakness of the QALY lies not in its implementation but in its foundations.

The analysis begins with the theory of allowable transformations and demonstrates that multiplication requires quantities possessing ratio measurement properties. A quantity can function as a proportion only if it has a fixed non-arbitrary zero and supports meaningful proportional comparisons. Utility scores are routinely interpreted as proportions of full health, yet there is no demonstration that they satisfy these requirements. The assignment of unity to full health does not establish ratio measurement, nor does the existence of bounded values between zero and one. The existence of negative utility values further undermines claims that utility scales possess a true zero.

The paper then examines an alternative defense of utilities as latent attributes. This route also fails. Latent attributes require a measurement model capable of estimating the extent to which an attribute is possessed. Rasch measurement provides the necessary and sufficient framework for constructing a logit ratio measure of latent attribute possession. However, a Rasch logit measures possession of an attribute rather than a proportion of time. Consequently, even if utility scores were transformed into valid latent measures, they could not function as multiplicative adjustment factors in the construction of QALYs.

The argument therefore identifies two independent reasons why the QALY cannot be sustained. First, utility scores have not been shown to possess the ratio properties required to support multiplication. Second, latent attribute measures and survival time belong to fundamentally different mathematical frameworks and cannot be merged into a single quantity through multiplication. The conclusion is that the QALY is an impossible construct. The failure lies not in modelling assumptions or implementation details but in the neglect of the measurement requirements that must precede arithmetic. Future HTA must therefore abandon the reference-case reliance on utilities and QALYs and return to evaluable claims grounded in valid measurement.

INTRODUCTION

For almost five decades the quality-adjusted life year (QALY) has occupied a central position in health technology assessment (HTA). It has become the preferred outcome measure for reference-case evaluations in many jurisdictions and is routinely employed to support resource allocation decisions, reimbursement recommendations and formulary assessments. The QALY is often presented as a self-evident quantity: survival time adjusted by the quality of that survival. The

arithmetic appears straightforward. A utility score multiplies a period of time or survival to generate quality-adjusted life or survival years.

Despite its widespread acceptance, remarkably little attention has been given to the measurement requirements necessary to support this multiplication. The debate surrounding the QALY has largely focused on issues of valuation methodology, preference elicitation, equity, discounting, uncertainty, modelling assumptions and decision thresholds. Critics and defenders alike have generally accepted the mathematical structure of the QALY and concentrated instead on the methods used to generate utility scores. The prior question has rarely been asked: what measurement properties must a quantity possess before it can function as a proportional adjustment factor in a multiplicative model?

This paper argues that the central weakness of the QALY lies not in its implementation but in its foundations. The construction of a QALY requires multiplication. Multiplication is not a neutral arithmetic operation. It imposes specific measurement requirements on the quantities involved. If those requirements are not satisfied, the resulting quantity cannot be regarded as a scientifically meaningful measure regardless of the sophistication of the modelling framework in which it is embedded.

The argument presented here proceeds in four stages. The first examines the theory of allowable transformations and its role in determining which arithmetic operations are admissible. The second reviews the defining properties of ratio measurement and explains why proportional reasoning depends upon those properties. The third demonstrates that the proportional adjustment required for QALY construction cannot be justified unless utility scores are ratio measures. The final section considers an alternative defense of utilities as latent attributes and shows that this route also fails.

1. THE QALY PROBLEM

The quality-adjusted life year (QALY) is the cornerstone of contemporary health technology assessment. It is presented as a measure that combines survival and health-related quality of life into a single quantity capable of supporting comparisons across diseases, therapies and patient populations. The appeal of the QALY lies in its apparent simplicity. Survival time is adjusted by a utility score representing the quality of that survival. The resulting quantity is then employed in cost-effectiveness analyses, reference-case simulations and reimbursement decisions.

Despite its central role, there is a fundamental question that has rarely been addressed. What measurement properties must a utility score possess before it discount time? This question precedes all issues concerning valuation methods, modelling assumptions, uncertainty analyses and decision thresholds. Before arithmetic can be undertaken, the quantities entering the arithmetic must be shown to possess the properties necessary to support the operation. Measurement precedes arithmetic.

The QALY is conventionally represented as the product of a utility score, denoted by u , and survival time, denoted by T , giving quality-adjusted survival $Q = uT$. The measurement status of time is not in dispute. Time is a ratio measure. It possesses a fixed non-arbitrary zero and supports

meaningful proportional comparisons. Ten years is twice five years. Twenty months is twice ten months. Multiplication and division are meaningful because time possesses ratio properties.

The challenge lies entirely with the utility score. In the QALY equation, the utility score is required to function as a proportional adjustment factor. A utility value of 0.8 is interpreted as representing 80 percent of full health. A utility value of 0.5 is interpreted as representing 50 percent of full health. The arithmetic therefore assumes that utility values are proportions.

This assumption is rarely examined. Utility scores are generally accepted as though their role in the QALY automatically establishes their measurement properties. Yet arithmetic cannot create measurement. The fact that a quantity is multiplied does not demonstrate that it can be multiplied. The scientific burden lies in demonstrating that the quantity possesses the measurement properties necessary to support the operation.

Indeed, the construction of utility scales points in the opposite direction. Utility systems begin by defining an upper anchor and then proceed through a sequence of decrements representing increasingly undesirable health states. The scale is built downward from unity. At no point is a true zero demonstrated. Instead, zero emerges as a consequence of scoring conventions rather than as an empirically established absence of a unidimensional attribute.

The problem becomes even more apparent when utility systems permit negative values. If states worse than death are assigned values below zero, then zero cannot simultaneously represent the complete absence of the attribute. The existence of negative utility values therefore undermines any claim that the scale possesses a true zero. The origin becomes a convention rather than a measurement property.

More fundamentally, utility systems rarely identify a single attribute whose absence could define such a zero. Utility instruments combine multiple dimensions of health and functioning into a composite description. Mobility, self-care, pain, anxiety and usual activities are aggregated into a single numerical score. The resulting quantity is then treated as though it represented a single measurable attribute capable of supporting proportional comparisons. Yet the scientific justification for this assumption is seldom considered.

The consequence is that the QALY rests upon an unproven measurement claim. The utility score is required to function as a proportion, but the ratio properties necessary to support that role have never been demonstrated. Instead, the proportional interpretation is assumed and subsequently embedded within the arithmetic structure of the model.

This observation shifts the debate concerning the QALY to a more fundamental level. The question is not whether utilities are useful, whether respondents can complete valuation exercises, or whether simulation models generate plausible results. The question is whether the quantity being multiplied by time possesses the measurement properties necessary to support multiplication.

Until that question is answered, the QALY remains an unsupported measurement construct. The issue is not the correctness of the arithmetic. The issue is whether the arithmetic is scientifically admissible. The remainder of this paper addresses that question by examining the theory of

allowable transformations, the requirements of ratio measurement, the role of proportions in quantitative science, and the implications of distinguishing between manifest and latent attributes for therapy assessment.

2. THE THEORY OF ALLOWABLE TRANSFORMATIONS

The debate over the QALY is usually framed in terms of valuation methods, utility instruments, preference elicitation techniques and simulation models. Yet these discussions overlook a more fundamental issue. Before any arithmetic operation can be justified, it must first be established that the quantities involved possess the measurement properties required to support that operation. This requirement is addressed by the theory of allowable transformations, one of the central concepts in representational measurement.

The theory of allowable transformations asks a simple question: what mathematical transformations can be applied to a numerical representation without changing the empirical information that the representation conveys? The answer determines the type of scale involved and, consequently, the arithmetic operations that may legitimately be performed.

This principle is familiar throughout the physical sciences. Length may be measured in inches, centimeters or meters. Weight may be measured in pounds or kilograms. Temperature may be measured in Celsius or Fahrenheit. Although the numerical values change when units change, the underlying empirical relationships remain the same. The purpose of measurement theory is to determine which transformations preserve those relationships and what conclusions may be drawn from them.

The implications are profound because different scales of measurement support different forms of analysis. A nominal scale permits only classification. An ordinal scale permits ranking. An interval scale supports meaningful differences between values. A ratio scale supports both differences and proportional comparisons. The arithmetic operations that are legitimate for one scale may be entirely inappropriate for another.

For nominal scales, any one-to-one relabeling is allowable. If patients are classified as male and female, the numerical labels assigned to those categories can be changed without altering the empirical content. Arithmetic operations are meaningless because the numbers function only as labels.

For ordinal scales, any monotonic transformation is allowable. Rankings may be altered while preserving order. A quantity ranked first remains above a quantity ranked second, regardless of the numerical values assigned. Ordinal scales therefore support statements concerning order but do not support statements concerning the magnitude of differences between values.

Interval scales represent a substantial advance because equal differences correspond to equal differences in the underlying attribute. The classic example is temperature measured in Celsius or Fahrenheit. The transformation between these scales takes the form: $x' = a + bx$ where (a) represents a shift in origin and (b) represents a change in units.

Because the origin is arbitrary, proportional comparisons are meaningless. Twenty degrees Celsius is not twice as hot as ten degrees Celsius. The existence of an arbitrary origin prevents ratio statements from having empirical meaning.

Ratio scales differ fundamentally from interval scales because the origin is fixed. The only allowable transformation is: $x' = bx$ where (b) represents a change in units. The absence of the additive constant is critical. Because the zero point is fixed and non-arbitrary, proportional comparisons become meaningful. Ten kilograms is twice five kilograms. Twenty years is twice ten years. Multiplication and division are admissible because the quantity possesses ratio properties.

The importance of allowable transformations lies in the fact that they determine which arithmetic operations are scientifically legitimate. Arithmetic is not justified merely because numbers are present. The scale type determines what may and may not be done with those numbers. A quantity admitting interval transformations cannot automatically be treated as though it were a ratio measure. The measurement properties must first be established.

This principle is directly relevant to the QALY. The construction of a quality-adjusted life year requires multiplication of time by a utility score. Time is a ratio measure. Consequently, the utility score must also possess ratio properties if it is to function as a proportional adjustment factor. The burden of proof therefore rests upon those who claim that utility scores support the allowable transformations of a ratio scale.

The challenge is that utility scores are routinely employed without any discussion of their allowable transformations. The central question is rarely asked: what transformations preserve the empirical information represented by utility values? If utility scores admit arbitrary shifts of origin, then they cannot function as proportions. If they cannot function as proportions, multiplication by time is inadmissible.

This observation reveals why the theory of allowable transformations is so important. It is not an abstract mathematical curiosity. It provides the rules that distinguish lawful arithmetic from unlawful arithmetic. It determines whether multiplication, division and proportional reasoning are meaningful. Most importantly, it establishes the principle that measurement precedes arithmetic. Before quantities can be manipulated mathematically, their measurement properties must first be demonstrated.

3. WHAT IS A RATIO MEASURE?

The theory of allowable transformations establishes that different scales support different forms of arithmetic. The next question is therefore straightforward: what properties must a quantity possess before it can function as a proportion? The answer lies in the concept of ratio measurement.

Ratio measurement occupies a unique position within quantitative science. It is the highest form of measurement because it supports all arithmetic operations, including multiplication, division and proportional reasoning. Whenever a quantity is interpreted as a proportion, percentage or multiplicative adjustment factor, ratio measurement is being assumed. The burden of proof

therefore rests on demonstrating that the quantity satisfies the conditions necessary for ratio measurement.

The defining characteristic of a ratio scale is the existence of a fixed non-arbitrary zero. This zero represents the complete absence of the attribute being measured. It is not a convention, a reference point or a convenient numerical anchor. It is an empirical property of the attribute itself. Because the origin is fixed, proportional comparisons become meaningful. Ten kilograms is twice five kilograms. Twenty years is twice ten years. The ratio between values remains invariant regardless of the units employed.

This invariance is reflected in the allowable transformations associated with ratio scales. If a quantity is measured in different units, the numerical values may change, but only through multiplication by a positive constant. A length measured in inches can be converted to centimeters. A weight measured in kilograms can be converted to pounds. The units change, but the proportional relationships remain unchanged. The critical feature is that no arbitrary constant can be added to the scale. Once an additive constant is introduced, the origin is no longer fixed and proportional comparisons lose their unique meaning.

The absence of an additive constant is therefore not a technical detail. It is the defining property that separates ratio scales from interval scales. A quantity that permits arbitrary shifts in origin cannot function as a proportion because proportional relationships change whenever the origin changes. Ratio measurement requires that the origin be fixed by the attribute itself rather than by convention.

The existence of a fixed zero, however, is only one requirement. Ratio measurement presupposes the existence of a clearly defined attribute. Before measurement can begin, the attribute must be identified. Measurement is not concerned with manipulating numbers. It is concerned with representing magnitudes of an attribute. If the attribute itself is poorly defined, then measurement becomes impossible.

The attribute must also be unidimensional. Measurement requires a single attribute whose magnitude varies across observations. This requirement is often overlooked because numerical scores can be generated without regard to dimensionality. Yet proportional comparisons presuppose that all observations refer to the same underlying phenomenon. To say that one quantity is twice another requires that both quantities represent different magnitudes of a single attribute. If several attributes are combined into a composite score, the resulting quantity may be useful as a summary index, but its status as a measure becomes problematic.

These requirements are familiar throughout the physical sciences. Length, weight, mass and time satisfy them. The attributes are clearly defined. They are unidimensional. Their zero points are fixed and non-arbitrary. Proportional comparisons are meaningful. Consequently, multiplication and division are permissible.

The significance of ratio measurement becomes particularly apparent when a quantity is required to function as a proportion. Consider a therapy assessment where survival time is adjusted by a factor intended to represent the proportion of an attribute that remains. The moment a quantity is

interpreted as representing 80 percent, 50 percent or 20 percent of some attribute, ratio measurement requirements are immediately imposed. Proportional comparisons must be meaningful. The quantity must possess a fixed non-arbitrary zero. The allowable transformations must preserve proportional relationships. Without these properties, the quantity cannot function as a proportion.

This observation exposes a common misunderstanding. A quantity does not become a ratio measure merely because it lies between zero and one. Nor does a quantity become a ratio measure because it is expressed as a percentage. A bounded scale is not the same as a ratio scale. The existence of an upper anchor tells us nothing about the existence of a true zero. Many numerical scales possess upper and lower bounds without satisfying the requirements of ratio measurement.

The critical question is therefore not whether values fall within a particular range, but whether the scale possesses a meaningful zero corresponding to the complete absence of the attribute and whether proportional comparisons are empirically justified. Without these properties, ratio measurement cannot be claimed.

4. PROPORTIONS AND RATIO TRANSFORMATIONS

The previous section argued that a quantity can function as a proportion only if it satisfies the requirements of ratio measurement. This conclusion follows directly from the theory of allowable transformations. The purpose of the present section is to apply this principle to the central arithmetic operation underlying the QALY and demonstrate why the reference-case framework depends upon an unproven measurement assumption.

The construction of a QALY requires multiplication of survival time by a utility score. In its simplest form, the relationship is expressed as: $T' = uT$ where T represents quality-adjusted survival time, u is the utility score and T' is survival time.

The role of the utility score is crucial. It is not merely another quantity entering an equation. It functions as a proportional adjustment factor. A utility value of 0.8 is interpreted as representing 80 percent of full health. A utility value of 0.5 is interpreted as representing 50 percent of full health. The arithmetic therefore assumes that utility values are proportions.

The validity of the QALY depends entirely upon this assumption. If the utility score cannot function as a proportion, then multiplication by time is inadmissible and the QALY loses its measurement-theoretic foundation.

To see why, consider a general proportional adjustment. Suppose survival time T is adjusted by a quantity x , giving:

$$T' = xT$$

For x to function as a proportion it must possess ratio measurement properties. In particular, it must possess a fixed non-arbitrary zero. If the scale lacks a true zero, then an arbitrary constant a

may be added to every value without changing the ordering of observations. The transformed quantity becomes:

$$x = x + a$$

Substituting this transformed quantity into the adjustment equation gives:

$$T' = (x + a)T$$

which expands to:

$$T' = xT + aT$$

The significance of this result is immediate. Adjusted survival is no longer determined solely by x . It also depends upon the arbitrary constant a . Rearranging the expression gives:

$$x = (T' - aT)/T'$$

or equivalently:

$$x = (T'/T) - a$$

The value of x is therefore no longer uniquely determined. Different choices of the arbitrary constant produce different values of the supposed proportion. A quantity that depends upon an arbitrary choice of origin cannot function as a genuine proportional adjustment factor.

This observation alone is sufficient to undermine any claim that a quantity lacking a fixed zero can function as a proportion. Proportions must be invariant. Their meaning cannot depend upon arbitrary shifts in the origin of the scale. Once the origin becomes arbitrary, the quantity ceases to support proportional reasoning and multiplication is no longer justified.

The same conclusion can be demonstrated in a more intuitive way. Suppose two utility values are interpreted as proportions:

$$x_1 = 0.8$$

$$x_2 = 0.4$$

The natural interpretation is that the first value represents twice the magnitude of the second because:

$$x_1/x_2 = 0.8/0.4 = 2$$

Now suppose an arbitrary constant of 0.2 is added to both values. The transformed quantities become:

$$x_1 = 1.0$$

$$x_2 = 0.6$$

The new ratio is:

$$x_1/x_2 = 1.0/0.6 = 1.67$$

Nothing about the underlying observations has changed. Yet the proportional relationship has changed from 2.0 to 1.67. A quantity whose proportional relationships vary when the origin shifts cannot function as a proportion. This is precisely why ratio measurement requires a fixed non-arbitrary zero.

The implications for utility scores are profound. Utility systems routinely assume that values represent percentages or proportions of full health. Yet the scientific justification for this interpretation has never been established. The existence of numerical values between zero and one does not create ratio measurement. A bounded scale is not a ratio scale. The assignment of unity to full health does not establish the existence of a true zero.

Indeed, the construction of utility scales points in the opposite direction. Utility systems begin by defining an upper anchor and then proceed through a sequence of decrements representing increasingly undesirable health states. The scale is built downward from unity. At no point is a true zero demonstrated. Instead, zero emerges as a consequence of scoring conventions rather than as an empirically established absence of a unidimensional attribute.

The difficulty becomes even greater when utility systems permit negative values. If states worse than death can be assigned values below zero, then zero cannot simultaneously represent the complete absence of the attribute. The existence of negative values demonstrates that the origin is determined by convention rather than by the attribute itself. This is incompatible with ratio measurement.

The problem extends beyond the final utility score to the entire process of utility construction. Consider the time trade-off procedure. Respondents are asked to trade a period of life in a specified health state for a shorter period in full health. The resulting value is immediately interpreted as a proportion. Yet this interpretation already assumes ratio measurement. The TTO score is treated as though it represents a percentage of full health despite the absence of any demonstration that the underlying quantity possesses ratio properties.

The same assumption reappears when utility algorithms are constructed from health-state valuations. It reappears again when utilities are inserted into the QALY equation. Finally, it becomes embedded within reference-case simulation models. At each stage quantities are treated as though they can function as proportions. At no stage is this assumption demonstrated.

The consequence is unavoidable. The QALY rests upon an unproven measurement claim. The quantity required to function as a proportion has never been shown to possess the ratio properties

necessary to support that role. The problem is not the arithmetic. The problem is that the arithmetic assumes measurement properties that have never been established.

This conclusion exposes the foundational weakness of the reference case. The failure does not occur at the level of modelling. It occurs much earlier, at the point where preference scores are first interpreted as proportional quantities. Once that assumption is accepted, every subsequent stage follows naturally. Utilities become proportions. QALYs become possible. Simulation models become feasible. Yet the entire structure rests upon a measurement assumption that has never been justified.

5. LATENT ATTRIBUTES AND RATIO TRANSFORMATIONS

The preceding sections demonstrated that the QALY requires utility scores to function as proportional adjustment factors and that such a role demands ratio measurement properties. A defender of the QALY might respond that this criticism misunderstands the nature of utility scores. Rather than viewing utilities as proportions, they might be regarded as representations of an underlying latent attribute. If utilities are measures of a latent construct, then perhaps the requirements of ratio measurement can be satisfied through a measurement model rather than through direct observation.

At first sight this appears to offer a possible escape from the difficulties associated with proportional interpretation. Closer examination, however, reveals that the latent attribute argument leads not to a solution but to a second and independent reason why the QALY cannot be defended.

The first challenge concerns the identification of the latent attribute itself. Measurement requires an attribute. Before a quantity can be measured, the attribute must be clearly specified. This is true for both manifest and latent attributes. Yet utility instruments do not identify a single latent attribute. Instead, they combine multiple dimensions of health and functioning into a composite description.

The EQ-5D provides an obvious example. Mobility, self-care, usual activities, pain or discomfort and anxiety or depression are represented simultaneously. Similar multidimensional structures are found throughout utility instruments. These dimensions may be important descriptors of health, but they are not a single attribute. They represent distinct phenomena. The requirement for unidimensionality therefore presents an immediate obstacle.

This point is critical because latent measurement is fundamentally different from the construction of summary indices. A latent attribute is not created by aggregating observations. It is discovered through evidence that observations reflect differing magnitudes of a single underlying phenomenon. Without unidimensionality there can be no latent measurement.

Suppose, however, that this difficulty is set aside. Assume for the sake of argument that a single latent attribute can somehow be identified. A second challenge immediately emerges. Latent attributes require a measurement model. The only model is the Rasch model¹. First proposed in 1960 it presents the necessary and sufficient rules for transforming observations to measurement.

The most important contribution of Rasch measurement is its recognition that latent attributes cannot be observed directly. They must be inferred from responses to carefully constructed items. The objective of Rasch analysis is not the creation of scores but the estimation of attribute possession. Individuals possess differing amounts of the latent attribute. Items represent differing thresholds along the same continuum. The interaction of persons and items yields a measure of possession.

The result of Rasch measurement is a logit scale. A logit is not a preference score, a ranking or a summary index. It is a measure of the extent to which an individual possesses a latent attribute relative to item difficulty. The resulting scale provides a measure of possession of an attribute.

This distinction is crucial because it changes the meaning of the quantity entirely. A Rasch logit does not represent a percentage of health. It does not represent a proportion of a maximum state. It does not function as a discount factor. It represents the extent to which a latent attribute is possessed. Consequently, even if utility scores could be transformed into a valid Rasch measure of a latent attribute, the fundamental QALY problem would remain.

The reason is straightforward. The QALY requires multiplication of time by a utility quantity: $Q = uT$, yet a Rasch logit is not a proportion of time. It is a measure of attribute possession. Time and attribute possession are fundamentally different quantities. Time measures duration. A Rasch logit measures possession of an attribute. These are not alternative representations of the same phenomenon. They belong to entirely different measurement domains.

The multiplication of a measure of attribute possession by a measure of time therefore lacks any obvious scientific interpretation. Consider the analogous case of intelligence. Suppose intelligence is measured through a Rasch model and expressed as a logit measure of possession. One would not multiply intelligence by years of survival and claim that the resulting quantity represented a meaningful measure. The fact that both quantities are numerical does not justify their multiplication.

The same observation applies to other latent attributes. Measures of pain, fatigue, anxiety, physical functioning, social participation or need fulfilment may all be constructed through Rasch measurement. Yet no one would claim that multiplying these measures by survival time automatically generates a scientifically meaningful quantity.

The problem is therefore deeper than the absence of ratio properties in utility scores. The problem is that latent measurement and proportional adjustment are fundamentally different concepts.

A Rasch logit ratio scale answers a specific question: how much of a latent attribute does an individual possess? The QALY requires a different quantity altogether: a proportional adjustment factor capable of modifying survival time. These are not equivalent roles. A measure of possession cannot automatically be transformed into a proportion of time.

The reference case therefore encounters a fundamental impasse. The first route fails because ratio measurement has never been demonstrated. The second route fails because latent attribute

possession cannot be interpreted as a proportion of time. The QALY cannot be rescued by moving from one interpretation to the other.

This conclusion has important implications for HTA. Once the distinction between manifest and latent attributes is recognized, the rationale for collapsing disparate outcomes into a single quality-adjusted survival metric largely disappears. Manifest attributes such as survival, hospital admissions, emergency department visits and treatment persistence can be measured directly through linear ratio measures. Latent attributes such as pain, fatigue, physical functioning and need fulfilment can be measured through Rasch logit ratio scales of attribute possession. Each attribute can then be evaluated through empirical protocols and subjected to replication and falsification.

The QALY attempts to combine fundamentally different dimensions into a single quantity. The measurement requirements necessary to justify that combination have never been demonstrated. Recognition of the distinction between manifest and latent attributes therefore provides a second and independent reason why the reference case cannot be defended as a scientific framework for therapy assessment.

CONCLUSION

The purpose of this paper has been to examine a question that is almost never asked in discussions of the QALY: what measurement properties must a quantity possess before it can be multiplied by time? This question precedes all debates concerning valuation methods, preference elicitation, utility instruments and simulation models. It is a question of measurement rather than economics.

The analysis has shown that the QALY depends upon a simple but demanding assumption. The quantity represented by the utility score must function as a proportion. Yet proportional adjustment requires ratio measurement. The theory of allowable transformations demonstrates why this is so. A quantity can function as a proportion only if it possesses a fixed non-arbitrary zero and supports meaningful proportional comparisons. Without these properties, multiplication is inadmissible.

This requirement has never been demonstrated for utility scores. The assignment of unity to full health does not establish ratio measurement. A bounded scale is not a ratio scale. The existence of negative utility values further undermines any claim that zero represents the complete absence of the attribute. The result is that the quantity required to function as a proportion has never been shown to possess the properties necessary to support that role.

More importantly, the problem originates at the very beginning of the reference-case framework. The time trade-off procedure treats preference scores as though they were proportional quantities. Utility algorithms treat those scores as though they possess ratio properties. The QALY then applies multiplication to quantities whose measurement status remains unresolved. Finally, reference-case simulation models embed these assumptions within increasingly complex analytical structures. The failure is therefore not confined to the QALY itself. It is present at every stage of reference case construction.

The latent attribute defense provides no escape. If utility scores are interpreted as latent attributes, then a measurement model is required. Rasch measurement provides the necessary and sufficient

framework for constructing measures of latent attribute possession. Yet a Rasch logit ratio scale measures possession of an attribute, not a proportion of time. A measure of attribute possession cannot automatically function as a multiplicative adjustment factor for survival duration. The QALY therefore fails under both interpretations. If utilities are preference scores, ratio measurement has not been demonstrated. If utilities are latent attributes, measures of possession cannot be interpreted as proportions.

The implications are unavoidable. The QALY is not merely a flawed measure, an imperfect approximation or a methodological convention in need of refinement. It is a construct that cannot satisfy the measurement requirements necessary for its existence. If utility scores are treated as preference scores, the ratio properties required for multiplication have never been demonstrated. If utility scores are treated as latent attributes, Rasch measurement yields measures of attribute possession that cannot be transformed into proportions of survival time. In neither case can the arithmetic operation at the heart of the QALY be justified. The conclusion is therefore uncompromising. The QALY is scientifically unsustainable and should be discarded from the HTA lexicon. Its continued use can only perpetuate the illusion that numerical outputs are equivalent to measurement. The future of HTA lies not in increasingly sophisticated attempts to rescue an impossible construct, but in the evaluation of separate manifest and latent attribute claims supported by valid measurement, empirical observation, replication and falsification. The task is not reform of the QALY. The task is its abandonment.

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

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