MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

THE MEASUREMENT FUTURE IN HEALTH

TECHNOLGY ASSESSMENT

LEGACY HTA: ISPOR AND THE GLOBAL FAILURE TO RECOGNIZE LAWFUL MEASUREMENT

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

LOGIT WORKING PAPER No 2006 SEPTEMBER 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

Background: *Health technology assessment (HTA) presents itself as a quantitative discipline, yet quantitative legitimacy requires more than numerical assignment, statistical analysis and mathematical modelling. Measurement must begin with a defined attribute, establish an appropriate numerical representation of its empirical relational structure, and demonstrate the scale properties required for subsequent arithmetic. This paper assesses whether these foundations are recognized within the methodological knowledge base of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR).*

Methods: *The publicly available and institutionally attributable ISPOR knowledge base was subjected to three linked interrogations addressing: (1) the primacy of the attribute as the object of measurement; (2) recognition of representational measurement as the foundation for quantitative claims; and (3) recognition of scales of measurement as constraints upon admissible arithmetic. Each interrogation employed ten independently specified canonical statements. Institutional endorsement was assessed using ordered categorical probabilities with corresponding normalized logits. Probabilities and logits were not averaged; interpretation was based upon the pattern of categorical assignments.*

Results: *The findings were unequivocal. For attribute primacy, all 10 statements were assigned $p \leq 0.10$, including eight at $p = 0.05$. For representational measurement, all 10 were assigned $p \leq 0.10$, including seven at $p = 0.05$. For scales of measurement, all 10 were assigned $p \leq 0.15$, with the weakest endorsement occurring where scale properties were required to constrain arithmetic. The three interrogations identify a cumulative failure of the required quantitative sequence: defined attribute → representational measurement → scale properties → admissible arithmetic → therapy-impact claim. Legacy HTA instead permits multiattribute aggregation, utility construction, QALYs, composite scores and reference-case modelling without first demonstrating the measurement properties required for their quantitative interpretation.*

Conclusions: *ISPOR has developed and disseminated a sophisticated international framework for manipulating numerical outcomes without establishing lawful measurement as the mandatory foundation for those operations. As presently constituted, legacy HTA cannot support a scientifically defensible quantitative framework for therapy-impact assessment and should be abandoned. This does not require abandonment of quantitative therapy evaluation, but its reconstruction from measurement forward. Manifest attributes require appropriate linear ratio measurement; latent attributes require fundamental measurement, with Rasch conjoint measurement providing the pathway adopted here. A structured nine-unit Maimon Research program demonstrates that this transition is already operationally available. HTA has reached a watershed: continued numerical sophistication cannot substitute for measurement.*

INTRODUCTION

Every quantitative claim about therapy impact must begin with an attribute. A therapy does not increase or decrease health, benefit or value in the abstract. It changes identifiable properties of the patient: survival time, pain, fatigue, physical function, treatment burden, need fulfilment or other specifically defined attributes. Measurement therefore begins by identifying what is to be measured. Only then can the empirical properties of the attribute be established, an appropriate numerical representation constructed and the arithmetic admissible for that representation determined.

This sequence is fundamental to any discipline claiming quantitative scientific status. Assignment of numbers does not itself constitute measurement. Representational measurement requires a correspondence between the empirical relational structure of an attribute and an appropriate numerical relational structure. The resulting scale properties then determine what can legitimately be done with the numbers. Nominal, ordinal, interval and ratio representations are not interchangeable, and arithmetic operations cannot be justified merely because they are mathematically executable. Measurement must authorize arithmetic, not the reverse.

These requirements provide a direct means of assessing the quantitative foundations of health technology assessment. ISPOR is particularly important because its Good Practices, educational activities, publications and international networks have exercised substantial influence over the methodological development of health economics and outcomes research at a global reach. The relevant question is therefore not whether ISPOR promotes sophisticated quantitative methods. It plainly does. The question is whether those methods rest upon the measurement foundations required to give their numerical operations empirical meaning for the ISPOR analytical structure. This paper complements an earlier one on representational measurement and the QALY [1].

The assessment proceeds through three linked interrogations of the ISPOR knowledge base. The first asks whether ISPOR recognizes the primacy of the attribute: whether quantitative evaluation begins by defining the single property whose magnitude or possession is to be measured. The second asks whether ISPOR recognizes representational measurement: whether numerical representations must preserve an appropriate empirical relational structure and satisfy the requirements necessary for measurement [2]. The third asks whether ISPOR recognizes scales of measurement as constraints upon arithmetic: whether scale properties determine which mathematical operations and quantitative interpretations are admissible [3]. These are not three independent methodological preferences. They constitute a necessary scientific sequence:

Failure at any stage compromises everything: defined attribute → representational measurement → scale properties → admissible arithmetic → therapy-impact claim that follows.

Failure at the attribute stage leaves unclear what is being measured. Failure to establish representational measurement leaves numerical assignments without demonstrated quantitative status. Failure to respect scale properties permits arithmetic that the numbers cannot support. The three interrogations therefore provide a cumulative test of ISPOR's status as a quantitative discipline. The issue is not the sophistication of its models, statistical methods or computational

procedures, but something logically prior: has ISPOR established the measurement foundations required before those quantitative methods can legitimately begin?

THE ATTRIBUTE AS THE OBJECT OF MEASUREMENT

The starting point for measurement is the attribute. An attribute is a specifically defined property of an object, person or event for which empirical instances can be ordered or otherwise related according to the structure required for measurement. In therapy evaluation, the object is typically the patient and the attribute is the particular property whose magnitude or possession is claimed to be affected by treatment. Survival time, walking distance and duration of hospitalization are examples of manifest attributes; pain, fatigue, physical function and need fulfilment are examples of latent attributes.

An attribute must be distinguished from a description, domain, health state, instrument, score or index. A health state may describe a patient simultaneously in terms of mobility, pain, anxiety, self-care and usual activities. These remain different attributes. Their inclusion within the same descriptive system does not establish that they constitute a single attribute capable of measurement. Similarly, assigning numerical values to their response categories or combining them through a scoring or preference algorithm does not create a new measurable attribute.

The requirement is therefore attribute first. Before an instrument is selected, observations are assigned numbers, responses are aggregated, health states are valued or arithmetic is undertaken, the attribute whose change constitutes the proposed therapy impact must be identified. The measurement question then follows from the nature of that attribute.

This produces the fundamental division between the two measurement pathways. A manifest attribute is one whose realizations can be observed directly and for which the required quantitative relationships may be established through an appropriate empirical measurement operation. Where the required properties are demonstrated, therapy impact can be expressed through a linear ratio measure. A latent attribute cannot be observed directly. Its possession is inferred from observable manifestations or questionnaire items selected because they provide ordered evidence concerning possession of that defined attribute. Fundamental measurement must therefore be constructed through the relationship between those observations and the latent attribute; within the framework adopted here, this is accomplished through Rasch conjoint measurement and expressed on the Rasch logit ratio scale [4 5] .

The distinction has an immediate consequence. Measurement attaches to an attribute**.** It does not attach merely to a questionnaire, health-state description, numerical score or preference value. Different attributes retain their separate measurement identities and their therapy impacts must be reported separately. Measurement of survival does not provide measurement of pain; measurement of pain does not provide measurement of physical function; and measurement of several attributes does not authorize their aggregation into an overall measure of health or therapeutic benefit.

The first question for every quantitative therapy-impact claim is therefore not “What instrument should be used?”, “What score should be calculated?” or “What value should be assigned?” It is:

What is the attribute? Only after that question has been answered can the appropriate measurement pathway be determined.

EVALUATING ATTRIBUTE PRIMACY

For the present assessment, the ISPOR knowledge base is defined as the publicly available and institutionally attributable body of methodological guidance through which ISPOR communicates, teaches and legitimizes the methods of health economics and outcomes research. It includes ISPOR Good Practices Task Force reports, clinical-outcomes and patient-reported-outcome guidance, health-state utility and preference-measurement materials, educational resources, and relevant methodological material published or disseminated through ISPOR and *Value in Health*. The purpose is not to determine whether a particular term has ever appeared in an ISPOR publication, conference presentation or article, nor whether an individual author associated with ISPOR has discussed representational measurement. The relevant question is whether the proposition under examination can reasonably be regarded as part of the institutionally endorsed methodological knowledge base that investigators, students and decision makers are expected to apply. An isolated reference to an attribute, Rasch analysis, scale type or measurement theorem does not establish that the corresponding principle has been incorporated into ISPOR's governing methodological framework.

The knowledge base is interrogated against a set of canonical statements derived independently from the requirements being examined. In the present interrogation, the statements specify what recognition of the primacy of the attribute would require: the therapy-impact claim begins with a defined attribute; the attribute precedes instruments, scoring and valuation; different attributes retain separate measurement identities; combining attributes within a descriptive system does not create a new measurable attribute; manifest and latent attributes present different measurement problems; and quantitative therapy impacts remain attribute-specific unless a common measurement structure has independently been demonstrated. Each statement is then assessed against the defined ISPOR knowledge base. The interrogation deliberately asks more than whether ISPOR employs similar terminology. Evidence of endorsement requires that the substance of the statement be recognized as a methodological requirement. Conversely, established ISPOR practices that proceed contrary to a canonical statement constitute evidence against endorsement. This is particularly important where ISPOR may refer to concepts or attributes while simultaneously endorsing multiattribute health-state valuation, utility construction and QALY analysis that do not impose attribute primacy as a prior measurement requirement.

Endorsement is represented by categorical probabilities assigned to each canonical statement, with corresponding normalized logits used to display the strength and direction of the assessment. These probabilities are not estimated frequencies, sampling probabilities or measurements of an underlying continuous ISPOR characteristic. They are ordered evidentiary categories expressing the judged strength with which the ISPOR knowledge base endorses each proposition: $p = 0.05$ represents essential absence of endorsement or evidence of contrary practice; $p = 0.10$ very weak or isolated recognition; $p = 0.15–0.20$ limited recognition; $p = 0.30–0.40$ incomplete recognition; $p = 0.50$ a mixed or indeterminate position; and progressively higher categories indicate increasingly strong, explicit and systematic endorsement, reaching $p = 0.95$ where the proposition is foundational and consistently applied.

The corresponding normalized logits provide a common visual representation of these ordered categories but do not convert them into observations suitable for arithmetic aggregation. Neither the categorical probabilities nor their logits are averaged**.** Interpretation rests on the pattern of categorical assignments across the canonical statement for example, the number of statements falling within the lowest endorsement categories and on the documentary and methodological evidence supporting each assignment.

TABLE 1. ISPOR FAILED ENDORSEMENT OF THE PRIMACY OF THE ATTRIBUTE

No.	Requirement	*p*	Logit
1	Therapy-impact claims begin with a single defined attribute.	0.10	−2.20
2	The attribute is defined before instruments, scoring, aggregation or valuation.	0.10	−2.20
3	The attribute is distinguished from concepts, domains, instruments, scores and health states.	0.05	−2.50
4	Combining health characteristics in one instrument does not create a single attribute.	0.05	−2.50
5	Scoring or valuing a multiattribute health state does not establish measurement of one attribute.	0.05	−2.50
6	Different attributes retain separate measurement identities.	0.05	−2.50
7	Manifest and latent attributes are distinguished as different measurement problems.	0.05	−2.50
8	Measurement pathway follows attribute type, not reporter or instrument classification.	0.05	−2.50
9	Therapy-impact claims remain attribute-specific; numerical values do not establish commensurability.	0.05	−2.50
10	Different attributes cannot be aggregated without a demonstrated common measurement structure.	0.05	−2.50

ASSESSMENT

The interrogation presents a stark result. All ten canonical statements fall within the two lowest endorsement categories: eight at $p = 0.05$ and two at $p = 0.10$. Thus, 10/10 statements fall at $p \leq 0.10$, providing essentially no evidence that the ISPOR knowledge base recognizes the primacy of the attribute as a foundational requirement of quantitative therapy evaluation.

This does not mean that ISPOR never employs terms such as attribute, concept or domain. What is missing is the requirement that a quantitative therapy-impact claim begin with a single defined attribute whose magnitude or possession is the object of measurement. Representational

measurement begins with an attribute, not with a questionnaire, health-state description, preference value, scoring algorithm or statistical model.

Multiattribute preference-based instruments demonstrate the problem. The EQ-5D combines mobility, self-care, usual activities, pain/discomfort and anxiety/depression. These are different patient attributes. Incorporating them within one instrument does not transform them into quantities of a single attribute. Description permits plurality; measurement requires an identified attribute**.** No common unit or empirical relational structure is created merely by combining characteristics within a questionnaire.

Preference valuation cannot repair this failure. Preferences among multidimensional health states may be elicited and represented numerically, but it does not follow that the resulting value measures a single health attribute. Attribute primacy requires the prior question: what attribute does this number measure? Calling the result health, health-related quality of life, benefit, utility or value cannot provide the answer.

The connection with the impossible QALY is immediate. Before asking whether a health-state utility possesses the scale properties required for multiplication by time, it must first be established what attribute the utility measures. A preference number derived from heterogeneous characteristics cannot simply be assumed to represent a common measurable quantity. The QALY therefore inherits a measurement failure that exists before its arithmetic begins. Its demise is inevitable.

The same problem extends to patient-reported outcomes. Assigning integers to categorical responses and aggregating them into a score does not establish measurement. Reliability, validity and responsiveness cannot manufacture equal units or an invariant measurement structure. Nor does ISPOR's distinction between patient-, clinician-, observer- and performance-reported outcomes resolve the issue. These categories identify principally the source of observation, not whether the attribute itself is manifest or latent.

That distinction determines the measurement pathway. Manifest attributes such as survival time, hospital days and walking distance require demonstration of the properties necessary for linear ratio measurement. Latent attributes such as pain, fatigue, physical function and need fulfilment require a different pathway, with Rasch conjoint measurement providing the route adopted here. Therapy-impact claims must consequently remain attribute-specific. Increased survival, reduced pain and improved physical function are separate claims; numerical representation does not make them commensurable.

The interrogation therefore identifies failure at the first measurement gate. The required sequence is defined attribute → measurement pathway → demonstrated measurement properties → admissible arithmetic → attribute-specific therapy-impact claim. Legacy HTA reverses this sequence by permitting characteristics to be combined, valued and subjected to arithmetic without first establishing what attribute has been measured.

The origins of this inversion extend back more than four decades. Quantitative health economics began not with measurement of a defined attribute but with the valuation of health-state

descriptions. Time trade-off and standard gamble techniques assigned preference values to health states and subsequently became central to utilities, QALYs and cost-utility analysis. The starting question became what value should be assigned to this health state? rather than the logically prior question**:** what attribute is being measured?

The early health-economic literature, including work associated with institutions such as the UK Office of Health Economics, developed within this valuation tradition [6] .The problem predates ISPOR. ISPOR did not originate the measurement inversion; it inherited a framework in which health-state valuation had already displaced measurement as the starting point for quantitative evaluation. It subsequently institutionalized and disseminated that framework through methodological guidance, publications, education and international networks at the global level. ISPOR's uncritical acceptance of this flawed approach guaranteed 40 years of untenable legacy HTA with global uncritical acceptance.

This history explains the findings of the present interrogations. A global failure to recognize tribute primacy, representational measurement and scales of measurement as constraints upon arithmetic are not unrelated omissions. They follow from beginning quantitative HTA with valuation rather than measurement**.** Once multidimensional health-state valuation became the foundation, utilities, QALYs and reference-case models could become increasingly sophisticated without resolving the prior questions of what attribute was measured, whether lawful measurement had been established and what arithmetic the resulting scale permitted.

The inversion was therefore embedded from the beginning: health states were valued before attributes were measured, numbers were produced before their measurement status was established, and arithmetic followed from the numbers rather than from demonstrated measurement properties. ISPOR inherited this architecture and subsequently transformed it into an internationally accepted framework for quantitative HTA without establishing its required measurement foundation.

DOES ISPOR RECOGNIZE REPRESENTATIONAL MEASUREMENT?

Representational measurement is fundamental to any scientific discipline that makes quantitative claims. The assignment of numbers to observations does not, by itself, establish measurement. A quantitative discipline must be able to demonstrate why its numerical representations correspond to the empirical properties they purport to represent, what relationships among those properties are preserved by the numerical system, and what transformations and arithmetic operations remain meaningful. These requirements precede statistical analysis, modelling and inference. Without them, mathematical operations may be computationally correct while lacking the empirical interpretation attributed to their results.

This requirement is particularly important for health technology assessment because its conclusions depend extensively upon numerical operations. Differences are calculated, outcomes aggregated, health states valued, utilities multiplied by time, incremental outcomes constructed and costs divided by effects. Patient-reported observations are coded and aggregated into scores, treatment differences are estimated and quantitative claims concerning therapy impact are made. None of these operations establishes the measurement status of the quantities entering them. That

status must have been demonstrated beforehand. Representational measurement therefore cannot be regarded as a specialist interest peripheral to HTA methodology. It provides the scientific foundation upon which the interpretation of quantitative claims depends. A foundation which has been accepted for 50 years by the quantitative physical and the more sophisticated social sciences.

The issue is also distinct from the conventional psychometric use of the term *measurement properties*. Reliability, validity, responsiveness, interpretability and statistical performance may provide important information about an instrument or score, but they do not replace the requirements of representational measurement. A reliable numerical score is not necessarily a measure. Evidence of validity does not itself demonstrate equal quantitative units. Responsiveness does not establish an additive structure. Statistical association does not establish invariance, and precision of estimation cannot determine whether the arithmetic applied to a numerical representation is admissible. The prior question remains whether the empirical relational structure of the defined attribute has been represented by a numerical relational structure possessing the properties required for the quantitative claims being made.

For an international professional organization concerned with health economics and outcomes research, recognition of these requirements should therefore be foundational. ISPOR has exercised substantial influence over the methods, terminology, educational content and Good Practices of contemporary HTA. Its methodological knowledge base addresses outcome assessment, patient-reported outcomes, health-state utilities, preference measurement, economic evaluation and quantitative modelling. The relevant forensic question is consequently straightforward: does the ISPOR knowledge base recognize representational measurement as the necessary foundation for quantitative measurement and require its conditions to be satisfied before numerical representations are subjected to arithmetic, statistical analysis and claims of therapy impact?

To address this question, ten canonical statements were specified independently of the subsequent assessment of ISPOR. They represent the principal requirements that would be expected if representational measurement occupied a foundational position within the ISPOR knowledge base. The statements address the representation of an empirical relational structure by an appropriate numerical relational structure; the distinction between numerical assignment and measurement; demonstration that the empirical structure supports the proposed numerical representation; unidimensionality; preservation of empirical order; additive and equal-unit structure; the relevant axiomatic conditions required by applicable representation theorems; invariance under admissible transformations; the inability of conventional psychometric and statistical properties to substitute for measurement; and the requirement that measurement be established before arithmetic and quantitative therapy-impact claims are undertaken.

Each canonical statement was interrogated against the defined ISPOR knowledge base and assigned an endorsement probability from the categorical probability framework used throughout these assessments (Table 2). The probabilities represent ordered judgments of the strength of institutional endorsement rather than estimated frequencies or measurements of an underlying continuous characteristic. A probability of $p = 0.05$ indicates essential absence of endorsement or the presence of established practices contrary to the proposition; $p = 0.10$ indicates very weak or isolated recognition; $p = 0.15–0.20$ limited recognition; $p = 0.30–0.40$ incomplete recognition; $p = 0.50$ a mixed or indeterminate position; and progressively higher categories indicate increasingly

explicit, systematic and foundational endorsement, reaching p = 0.95 where the proposition is clearly embedded within the methodological framework and consistently applied.

The corresponding normalized logits provide a common representation of the ordered probability categories, ranging from −2.50 for p = 0.05 to +2.50 for p = 0.95. They do not convert the categorical judgments into measurements suitable for arithmetic aggregation. Neither probabilities nor logits are averaged. The evidentiary result is the pattern of categorical assignments across the ten statements. Importantly, the occurrence of terminology related to measurement, unidimensionality, invariance or other relevant concepts in an isolated ISPOR-associated publication does not by itself establish endorsement. The criterion is whether the principle is incorporated into the institutionally attributable methodological framework as a requirement governing quantitative analysis. The interrogation therefore distinguishes between the presence of measurement terminology and the recognition of representational measurement as a governing scientific standard.

TABLE 2. ISPOR FAILED ENDORSEMENT OF REPRESENTATIONAL MEASUREMENT

No.	Canonical statement	*p*	Logit
1	Measurement requires representation of an empirical relational structure by an appropriate numerical relational structure.	0.05	−2.50
2	Assignment of numbers, scores or values does not itself establish measurement.	0.10	−2.20
3	The empirical structure must support the proposed numerical representation.	0.05	−2.50
4	Measurement of a single attribute requires a unidimensional quantitative structure.	0.10	−2.20
5	Numerical order must preserve corresponding empirical order in the defined attribute.	0.10	−2.20
6	Equal numerical differences require a demonstrated additive, equal-unit structure.	0.05	−2.50
7	Applicable axiomatic conditions must be satisfied as required by the representation theorem.	0.05	−2.50
8	Measurement requires invariance under admissible transformations.	0.05	−2.50
9	Psychometric and statistical properties cannot substitute for demonstrated measurement.	0.05	−2.50
10	Measurement must precede arithmetic, statistical analysis and quantitative therapy-impact claims.	0.05	−2.50

ASSESSMENT

The interrogation produces an unequivocal result. All ten canonical statements fall within the two lowest categorical endorsement probabilities: seven are assigned p = 0.05 and three p = 0.10. Thus, 10/10 statements are at $p \leq 0.10$. The pattern provides essentially no evidence that representational

measurement has been incorporated into the ISPOR knowledge base as the governing scientific foundation for quantitative measurement.

This conclusion does not imply that ISPOR never uses the language of measurement or that relevant concepts cannot be found in individual publications. ISPOR materials discuss measurement properties, unidimensionality, instrument validity and related psychometric concepts. The issue is whether representational measurement is imposed as a prior methodological requirement. An isolated reference to Rasch analysis, invariance or unidimensionality does not establish institutional recognition if these requirements do not govern the quantitative methods ISPOR endorses.

The distinction between psychometric validation and measurement is central. Reliability, validity, responsiveness and interpretability may provide useful information about the performance of an instrument, but they do not establish the empirical relational structure necessary for quantitative measurement. Reliability does not demonstrate equal units. Responsiveness does not establish additivity. Correlation does not establish invariance. Statistical precision cannot determine whether the arithmetic applied to a score is admissible. A score can satisfy conventional psychometric criteria while failing to demonstrate the quantitative structure required for the interpretation assigned to it.

The categorical assignments reflect this fundamental absence. There is no endorsement of the requirement that measurement represents an empirical relational structure by an appropriate numerical relational structure, that the empirical structure supports the numerical representation, that equal numerical differences require demonstrated equal units, or that applicable axiomatic conditions and invariance must be established. The slightly higher $p = 0.10$ assignments for numerical assignment, unidimensionality and empirical order recognize limited elements within the ISPOR-associated outcomes literature. They do not amount to adoption of representational measurement as the required governing standard.

Particularly damaging is the absence of recognition that conventional psychometric properties cannot substitute for measurement. Within outcomes research, the language of reliability, validity and responsiveness has effectively come to stand in for measurement itself. Instruments are validated, scores generated and treatment effects calculated. Yet none of these procedures demonstrates that the resulting numerical representation possesses the structure required for the arithmetic subsequently performed upon it. Mathematical operations can be executed on assigned numbers without those operations possessing the empirical meaning attributed to them.

The consequences are visible throughout legacy HTA. Health-state utilities are constructed and multiplied by time; patient-reported responses are numerically coded, summed and differenced; treatment effects are calculated; and increasingly elaborate statistical and decision models are built from these quantities. None of this downstream sophistication can establish measurement properties that were not demonstrated before the arithmetic began. The QALY provides the most conspicuous example: multiplication by time cannot confer upon a health-state utility the measurement properties required for that multiplication. The QALY fails as a meaningful construct.

The same failure explains why the distinction between manifest and latent attributes has been obscured. Manifest attributes require demonstration of the properties necessary for linear ratio measurement. Latent attributes require a fundamentally different pathway in which observations concerning possession are used to construct measurement of a defined attribute. Within the framework developed here, Rasch conjoint measurement provides that pathway. Treating Rasch merely as another optional psychometric technique misses its significance as an operational realization of fundamental measurement requirements for latent attributes.

The categorical pattern therefore identifies considerably more than a gap in ISPOR terminology. The fatal weakness extends across the architecture of representational measurement: empirical and numerical relational structures, equal units, applicable axiomatic requirements, invariance and the requirement that measurement precede arithmetic. ISPOR has developed an elaborate methodological framework for generating, analyzing and manipulating numerical outcomes without incorporating these requirements as mandatory measurement gates. ISPOR fully endorses arithmetic without measurement. This is a framework which should be abandoned for any pretense at quantitative evaluation.

The conclusion is severe. The preceding interrogation found essentially no recognition of the primacy of the attribute; the present interrogation finds essentially no recognition of representational measurement as the governing quantitative standard. These are successive failures in the same scientific sequence. If the attribute is not primary and measurement is not demonstrated, the numbers entering subsequent arithmetic have no recognized measurement foundation.

This cannot be repaired by improving a utility algorithm, replacing one PRO instrument or modifying the QALY. The failure precedes those methods. Reconstruction must begin with the defined attribute, its empirical relational structure and demonstration of an appropriate measurement representation. Only after that gate has been passed can arithmetic, statistics and therapy-impact claim assessment legitimately follow**.** ISPOR has developed sophisticated standards for working with numbers without first imposing the standards required to establish that those numbers are measures. The ISPOR endorsement is for numerical storytelling.

DOES ISPOR RECOGNIZE SCALES OF MEASUREMENT

Recognition of representational measurement necessarily raises the question of what mathematical operations are admissible for the numerical representation established. The answer depends upon scale properties. Nominal, ordinal, interval and ratio scales preserve different empirical information and consequently support different quantitative operations. Scale of measurement therefore provides the gate between numerical representation and admissible arithmetic.

The implications for HTA are immediate. Ordinal values establish order but not equal intervals. Interval measures provide equal units but have an arbitrary origin and do not support invariant ratios of scale locations. Ratio measures provide the additional structure required for meaningful proportional comparisons. These distinctions determine whether differences, ratios, percentage changes, multiplication and division have the empirical interpretation claimed for them. Computational feasibility alone cannot establish that interpretation.

This is critical because legacy HTA routinely undertakes precisely these operations. Patient-reported scores are summed and differenced, health-state values are multiplied by time to create QALYs, incremental outcomes are calculated and costs are divided by effects. Each operation presupposes particular measurement properties. Subsequent statistical analysis, modelling or simulation cannot supply properties absent from the quantities before the arithmetic began.

The relevant question is therefore not whether the terms nominal, ordinal, interval and ratio occasionally appear in ISPOR materials, but whether scale properties operate as mandatory constraints on arithmetic. To asses this, the ISPOR knowledge base was interrogated against ten canonical statements addressing the distinctions among scale types, equal intervals, meaningful zero, numerical coding and aggregation, and the requirement that scale properties determine admissible arithmetic before quantitative therapy-impact claims are made.

Each statement was assigned an endorsement probability using the categorical framework applied in the two preceding interrogations. The probabilities represent ordered judgments of institutional endorsement, from $p = 0.05$ for essential absence or contrary practice to $p = 0.95$ for explicit, foundational and consistently applied recognition. The corresponding normalized logits provide a common representation of these categories. Neither probabilities nor logits are averaged. Interpretation rests upon the categorical pattern across the ten statements and the evidence supporting each assignment. The results are presented in Table 3.

TABLE 3. ISPOR FAILED ENDORSEMENT OF SCALES OF MEASUREMENT

No.	Canonical statement	*p*	Logit
1	Nominal, ordinal, interval and ratio scales are explicitly distinguished.	0.10	−2.20
2	Ordinal order does not establish equal intervals.	0.15	−1.75
3	Interval and ratio measurement are explicitly distinguished.	0.10	−2.20
4	Ratio measurement requires a non-arbitrary zero.	0.10	−2.20
5	Numerical coding of ordered categories does not establish interval measurement.	0.15	−1.75
6	Summing ordinal item scores does not create interval or ratio measurement.	0.10	−2.20
7	Equal intervals must be demonstrated before quantitative differences are interpreted.	0.10	−2.20
8	Scale type determines the arithmetic operations that are admissible.	0.05	−2.50
9	Ratios and percentage changes require demonstrated ratio properties.	0.05	−2.50
10	Scale properties must be established before quantitative therapy-impact claims.	0.05	−2.50

ASESSMENT

The interrogation identifies a consistent failure to recognize scales of measurement as a governing constraint on quantitative analysis. All ten canonical statements fall at $p \leq 0.15$: three are assigned $p = 0.05$, five $p = 0.10$ and two $p = 0.15$. Limited recognition occurs principally at the level of terminology and elementary scale distinctions. Endorsement falls to the minimum category when the decisive issue becomes whether scale properties actually determine admissible arithmetic.

ISPOR cannot be described as entirely unaware of distinctions between ordinal, interval and ratio representations. There is some recognition that ordered categorical responses should not automatically be interpreted as possessing equal intervals. What is largely absent is the requirement that these distinctions operate as mandatory gates controlling subsequent mathematical operations. Numbers do not acquire unrestricted mathematical properties merely because they have been assigned to observations. Ordinal values establish order but not equal intervals; summation does not manufacture interval structure; and interval measures, because they retain an arbitrary origin, do not support invariant ratios of scale locations. Ratio measurement requires the additional structure necessary for proportional interpretation.

The minimum endorsement probabilities assigned to the final three statements therefore expose the central failure. There is essentially no evidence that ISPOR systematically requires scale type to determine admissible arithmetic, requires demonstrated ratio properties before proportional operations, or requires scale properties to be established before quantitative therapy-impact claims are made. This is precisely where scales of measurement become scientifically consequential.

The contradiction with legacy HTA practice is immediate. Health-state utility values are multiplied by time to construct QALYs; QALYs are subtracted to produce incremental outcomes; and incremental costs are divided by incremental QALYs. Each operation presupposes measurement properties necessary to support it. Arithmetic cannot supply those properties. If the requirements are not satisfied before the first operation, additional arithmetic merely carries the original failure forward.

The same problem applies to patient-reported outcomes. Assigning integers to ordered response categories and summing them produces a numerical score, but does not establish equal units. Means, differences, percentage changes and treatment effects cannot retrospectively transform that score into the measurement scale required for their interpretation. Statistical sophistication cannot rescue an inadmissible numerical representation.

The categorical pattern therefore reveals more than incomplete knowledge of scale terminology. It indicates failure to incorporate the central consequence of scale theory into ISPOR's methodological architecture: measurement properties determine admissible arithmetic. ISPOR may recognize elements of the vocabulary, but its endorsed quantitative framework does not consistently accept their authority as an arithmetic gatekeeper.

This completes a disturbing sequence. The preceding interrogations identified essentially no recognition of attribute primacy or representational measurement as a governing standard. The present interrogation adds the third failure: scale properties do not function as mandatory constraints upon arithmetic. Legacy HTA has not merely neglected scales of measurement; it has proceeded as though their mathematical consequences could be ignored**.**

ISPOR is essentially a measurement wasteland; a wasteland with an uncritical global audience**.** By bypassing measurement science, ISPOR has undermined its claim to legitimacy as a quantitative discipline. Unless its methodological framework is reconstructed from the foundations of lawful measurement, abandonment of legacy HTA as a scientifically defensible quantitative framework is unavoidable. In doing so, ISPOR has created the conditions for its own rejection as an authority

on quantitative therapy evaluation. And because ISPOR's influence extends across countries, institutions, professional networks and generations of researchers and students, that rejection will have global repercussions.

CAN ISPOR AND LEGACY HTA SURVIVE AS A QUANTITATIVE DISCIPLINE?

The three interrogations leave a fundamental question: can ISPOR continue to present legacy HTA as a scientifically defensible quantitative framework once its failure to meet the requirements of lawful measurement has been demonstrated? Historical acceptance, professional consensus and methodological sophistication cannot answer that question. Quantitative legitimacy depends upon measurement.

The failures are cumulative. Attribute primacy is not established; representational measurement does not provide the governing scientific foundation; and scales of measurement do not operate as mandatory constraints upon arithmetic. The required sequence: attribute → representational measurement → scale properties → admissible arithmetic has been inverted. Numerical assignment, aggregation, modelling and arithmetic have been permitted to precede demonstration that the quantities involved are measures.

As it presently stands, the legacy HTA framework for quantitative claims of therapy impact must be abandoned. This does not mean abandoning quantitative therapy evaluation. It means reconstructing it from measurement forward: defined attributes, appropriate measurement pathways, demonstrated scale properties, admissible arithmetic and empirically evaluable therapy-impact claims.

The implications are immediate. Faculty and students should be made aware of these foundational limitations rather than being taught utilities, QALYs, composite scores and reference-case modelling as though their quantitative legitimacy were established. ISPOR has a corresponding responsibility to inform its membership, educational participants, regional networks and wider international constituency of its lack of understanding and application of lawful measurement..

The choice is no longer between alternative refinements of legacy HTA. It is between reconstruction upon lawful measurement and abandonment of the claim that legacy HTA provides a scientifically defensible quantitative assessment of therapy impact.

ISPOR is essentially a measurement wasteland; a wasteland with a global audience. By bypassing measurement science, ISPOR has undermined its claim to legitimacy as a quantitative discipline. Unless its methodological framework is reconstructed from the foundations of lawful measurement, abandonment of legacy HTA as a scientifically defensible quantitative framework is unavoidable. In doing so, ISPOR has created the conditions for its own rejection as an authority on quantitative therapy evaluation. As well, because ISPOR's influence extends across countries, institutions, professional networks and generations of researchers and students, that rejection will have global repercussions.

By bypassing measurement science, ISPOR has undermined its claim to legitimacy as a quantitative discipline. Unless its methodological framework is reconstructed from the foundations of lawful measurement, abandonment of legacy HTA as a scientifically defensible quantitative framework is unavoidable.

THE TRANSITION TO LAWFUL MEASUREMENT

Abandonment of legacy HTA does not imply abandonment of quantitative therapy assessment. It requires its reconstruction. A practical framework for making that transition has already been developed by Maimon Research in the nine-unit program *Reconstructing HTA Programs Measurement before Arithmetic* [7] . The program comprises approximately 45,000 words of structured instructional material and is designed both for faculty seeking to make the transition from legacy HTA and as the foundation for graduate-level instruction in lawful quantitative therapy assessment.

The program reverses measurement inversion. It begins with the attribute rather than the instrument, score, utility or model. The attribute is defined, its measurement requirements established and the appropriate pathway identified. Manifest attributes require demonstration of the properties necessary for linear ratio measurement. Latent attributes require construction of measurement from observations concerning possession, with Rasch conjoint measurement providing the pathway adopted in the program. Only after measurement has been established are scale properties, admissible arithmetic and quantitative therapy-impact claims considered.

Across its nine units, the program develops the transition from the foundations of representational measurement through attribute definition, manifest and latent measurement, admissible arithmetic, therapy-impact claims, empirical evaluation, replication and falsification. It is not simply a critique of QALYs or conventional economic evaluation. It provides the positive measurement framework required to replace them. It is a return to normal science.

Development, maintenance and delivery of a program of this scale require continuing resources. The program is therefore offered on a fee-supported basis. Fees support the instructional materials, continuing development of the measurement framework, delivery and faculty engagement, and the broader educational effort required to establish lawful measurement as an alternative to legacy HTA. The objective is to create a sustainable program capable of supporting a genuine disciplinary transition rather than another short-lived methodological initiative.

Universities, faculty and students need not wait for ISPOR or HTA agencies to reconstruct their methods. The transition can begin now. The destination is straightforward: define the attribute, establish the measure, determine the admissible arithmetic and only then make the therapy-impact claim. The approximately 45,000-word Maimon Research program demonstrates that this reconstruction is not merely possible in principle. It is already available for implementation.

CONCLUSION: A WATERSHED FOR QUANTITATIVE HTA

Legacy HTA has reached a watershed. Its claim to legitimacy as a quantitative discipline can no longer be separated from its failure to meet the required standards of representational measurement. For decades, legacy HTA has presented itself as a quantitative discipline while building its principal claims upon utilities, QALYs, composite scores, cost-effectiveness ratios and assumption-driven reference-case models where none have any claim for lawful measurement.

The three interrogations undertaken here expose a more fundamental problem. There is essentially no evidence that the ISPOR knowledge base establishes the primacy of the attribute, little evidence that representational measurement provides its quantitative foundation, and little evidence that scales of measurement operate as mandatory constraints upon arithmetic. These are not peripheral deficiencies. They concern the conditions that must be satisfied before legitimate quantitative analysis can begin. ISPOR has yet to recognize this requirement.

The implications extend well beyond ISPOR. Its methodological influence has reached universities, HTA agencies, manufacturers, consultants, journals and generations of students. The legacy framework has consequently become deeply institutionalized. But institutional acceptance cannot establish scientific legitimacy. Forty years of application cannot transform an inadmissible operation into an admissible one, and increasingly sophisticated modelling cannot manufacture measurement properties absent from its inputs. This must be abandoned.

The choice is therefore no longer between competing versions of legacy HTA. The legacy framework for quantitative claims of therapy impact must be abandoned. What must be preserved is quantitative therapy evaluation itself, reconstructed from its measurement foundations. The required sequence is straightforward: define the attribute; determine whether it is manifest or latent; establish the appropriate measurement structure; demonstrate the required scale properties; identify the arithmetic those properties permit; and only then construct an empirically evaluable therapy-impact claim.

This creates responsibilities that can no longer reasonably be deferred. ISPOR should address these findings directly and inform its membership and wider international constituency of their implications. Universities should ensure that faculty and students understand the measurement limitations of the methods they teach and apply. HTA agencies should reconsider analytical requirements that presuppose the legitimacy of utilities, QALYs and cost-per-QALY claims without first demonstrating their measurement foundations. Manufacturers and researchers should no longer assume that compliance with an established reference case is equivalent to compliance with the requirements of quantitative science.

There is no scientific reason to wait for these institutions to move collectively. A structured transition to lawful measurement is already available. The Maimon Research nine-unit program demonstrates that reconstruction can begin now and that the alternative to legacy HTA is not the abandonment of quantitative analysis but its restoration to defensible measurement principles. Research staff, academic instructors and HTA groups can move towards this immediately.

The watershed is therefore both destructive and constructive. On one side lies a legacy framework in which numerical assignment, aggregation and arithmetic have been allowed to precede measurement. On the other lies an established attribute-first discipline in which measurement determines what arithmetic and quantitative claims are permissible. Once that distinction has been exposed, returning to business as usual is no longer a scientifically defensible option.

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

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

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

[4] Rasch G, Probabilistic Models for some Intelligence and Attainment Tests. Chicago: University of Chicago Press, 1980 [An edited version of the original 1960 publication]

[5] Bond T, Zi Yan. Heene M, Applying the Rasch model: Fundamental measurement in the human sciences. New York: Routledge, 2021

[6] Langley P. United Kingdom: The Office of Health Economics – the inversion of representational measurement. Logit Working Paper No 25. https://maimonresearch.com/logit-working-paper-no-25/

[7] Maimon Research. Reconstructing HTA Programs: Measurement before Arithmetic. https://maimonresearch.com/hta-reconstruction-program-and-fees/