MAIMON RESEARCH LLC

MEASUREMENT IN HEALTH TECHNOLOGY ASSESSMENT

THE TRANSITION TO REPRESENTATIONAL MEASUREMENT IN HEALTH TECHNOLOGY ASSESSMENT: THE EUROPEAN UNION

GERMANY: FROM MEASUREMENT INVERSION TO ADMISSIBLE CLAIMS IN THERAPY IMPACT ASSESSMENT

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

LOGIT WORKING PAPER No 1309 APRIL 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

This paper presents a structural critique of contemporary health technology assessment (HTA) within the European Union, with particular reference to Germany, and proposes a transition to a measurement-based framework for evaluating therapy impact claims. Interrogation of HTA-related knowledge bases through logit analysis demonstrates a consistent pattern: the absence of adherence to the axioms of representational measurement alongside strong endorsement of constructs such as utilities, QALYs, and reference case outputs as if they possessed admissible measurement properties. This asymmetry is diagnostic of measurement inversion, where arithmetic operations are applied to quantities whose status as measures has not been established.

The implications are fundamental. Claims generated within the reference case framework cannot support empirical evaluation, replication, or falsification. They do not represent approximate measures; they are internally inconsistent constructions that lack evidentiary standing. The persistence of this framework reflects institutional convention rather than scientific validation, resulting in a closed system that cannot support the accumulation of objective knowledge.

The paper argues that once the axioms of representational measurement are recognized, the range of admissible claims is constrained. Only two forms of measurement are permissible: linear ratio measures for manifest attributes and Rasch-based logit ratio measures for latent attributes. These provide the only foundation for therapy impact claims that can be tested and potentially refuted.

A structured transition is outlined, distinguishing between training and implementation. This includes the reorientation of claims toward single, unidimensional attributes, the introduction of measurement as a gatekeeper for admissible operations, and the development of protocol-driven submissions that support empirical evaluation. The transition is framed not as an alternative methodological option, but as a requirement imposed by the standards of normal science. The conclusion is that meaningful HTA can only be achieved through the re-establishment of measurement as the basis for all claims.

1. UNDERSTANDING THE NEED FOR TRANSITION

1.1 Measurement Before Arithmetic: The German and EU Context

The interrogation of HTA-related knowledge bases relevant to Germany uncovers a pattern that is consistent with findings across all assessed jurisdictions: a marked divergence between the requirements of representational measurement and the constructs employed in therapy impact assessment. The logit profiles derived from these assessments do not suggest isolated inconsistencies, but rather reveal a coherent and systematically reinforced structure. Statements reflecting the axioms of measurement, unidimensionality, dimensional homogeneity, and the requirement that measurement must precede arithmetic, receive low levels of endorsement, while constructs such as utilities, QALYs, and reference case outputs are treated as though they possessed the properties necessary for arithmetic operations. This contrast demonstrates that measurement inversion is not a peripheral weakness, but a fundamental and defining feature of the knowledge base.

Within the German context, this pattern reflects the incorporation of European HTA conventions within a national framework that places considerable emphasis on methodological conformity. This alignment has not resolved the prior requirement of measurement. Instead, it has consolidated a structure in which claims for therapy impact are generated through arithmetic operations applied to quantities whose measurement properties have not been established. As a consequence, such claims cannot support empirical evaluation, replication, or refutation. They remain internally consistent within the framework, but do not meet the conditions necessary for the accumulation of objective knowledge.

Recognition of the axioms of representational measurement removes any ambiguity. The range of admissible claims is constrained by the requirements of measurement itself. Only those claims grounded in linear ratio measures for manifest attributes and Rasch-based logit ratio scales for latent attributes can support arithmetic operations and empirical evaluation. There is no alternative pathway within the current HTA framework capable of producing evaluable claims. The issue is therefore not one of methodological variation, but of admissibility.

With two logit papers posted to the Germany section of the Maimon Research website, there have been approximately 574 visits to these assessments. While the sources of this engagement are not identifiable, the scale and persistence of activity suggest extensive and active engagement within the German HTA and research community. This document is presented in response to that engagement, setting out a structured transition framework through which measurement-valid, single-attribute claims can be introduced.

1.2 The Consequences of Measurement Inversion

The starting point for any transition away from the current health technology assessment framework is recognition that it is not merely imperfect, but fundamentally incoherent. The accumulated evidence from repeated knowledge base interrogations demonstrates a consistent pattern across the European Union: the absence of adherence to the axioms of representational measurement. This is not a marginal technical failure. It is systemic. The reference case, widely treated as the methodological gold standard, is revealed instead as a closed system of assumptions in which arithmetic operations are applied to constructs that do not meet the requirements for measurement. This inversion—where arithmetic precedes, and substitutes for, measurement—lies at the heart of the problem.

The consequences are profound. Claims generated within the reference case are not approximations or rough estimates. They are not even pseudo-measurements. They are internally contradictory constructions in which each step violates the conditions necessary for meaningful numerical operations. Utilities derived from instruments such as the EQ-5D lack unidimensionality and do not possess interval or ratio scale properties. Time, a ratio measure, is multiplied by these composite scores to generate the quality-adjusted life year. Costs, themselves often constructed as aggregates of heterogeneous resource inputs without dimensional homogeneity, are then divided by these outputs. At no point in this chain is there a defensible measurement structure. The cost-per-QALY is not a degraded measure; it is an imaginary claim.

This is why the language of "quantitative analysis" must be rejected in this context. The presence of numbers does not confer legitimacy. Arithmetic operations are only meaningful when applied to quantities that satisfy the axioms of measurement. When those axioms are ignored, the outputs are not approximate truths; they are artifacts of invalid operations. In this sense, the current HTA framework does not fail because it is insufficiently refined. It fails because it violates the logical conditions required for any claim to be evaluable, replicable, or falsifiable.

The persistence of this framework can be understood as a form of institutionalized numerical storytelling. The reference case provides a structure within which assumptions can be layered, parameters adjusted, and outputs generated that appear precise but are insulated from empirical challenge. Because the claims are not anchored in measurable attributes, they cannot be subjected to falsification. Competing models can produce different results without any means of determining which, if any, corresponds to reality. Decision-making is therefore guided not by evidence, but by the internal coherence of competing constructions.

From the perspective of the evolution of objective knowledge, this represents a dead end. Science advances through claims that can be tested and potentially refuted. The current HTA paradigm substitutes this process with outputs that are immune to such testing. There is no cumulative learning. Each model adds to the structure without resolving its internal inconsistency.

1.3 Logit Confirmation and the Case for Transition

With the availability of large language models, it is now possible to undertake a structured interrogation of HTA knowledge bases in a way that was previously impractical. By applying a canonical set of true–false statements grounded in representational measurement, endorsement probabilities can be derived and transformed into normalized logits. The significance of this lies not in the technology, but in what it reveals. For the first time, it is possible to demonstrate, systematically and reproducibly, the extent to which the global HTA framework fails to recognize the axioms of measurement.

Across interrogations of national agencies, academic centers, and teaching programs, a consistent pattern emerges. Statements reflecting measurement requirements receive low endorsement probabilities (0.05–0.20), while statements endorsing the measurement properties of HTA constructs receive high probabilities (0.80–0.95). The resulting logit profiles are highly polarized. Clusters of strongly negative logits correspond to the rejection or absence of measurement axioms, while strongly positive logits correspond to the acceptance of non-measurement constructs as if they were measures. This is the empirical signature of measurement inversion.

Importantly, this pattern is not jurisdiction-specific. It is observed across all knowledge bases, indicating that the issue is systemic. The attempt by EUnetHTA and subsequent European Commission HTA coordination efforts to impose a common framework based on these assumptions is therefore a futile exercise. Harmonization of a structure that lacks a foundation in measurement cannot produce evaluable claims.

What the logit confirmation provides is not a new critique, but a definitive demonstration. The framework is internally inconsistent. The numbers produced are coherent within their own

assumptions but do not correspond to measurable attributes. They cannot be validated, cannot be falsified, and cannot support the accumulation of knowledge. The logits do not suggest measurement failure; they demonstrate it.

The implication is unavoidable. The existing HTA framework does not produce evidence. It produces numerical constructions that simulate evidence while violating the conditions required for its existence. The issue is not one of refinement, but of admissibility. Once this is recognized, the need for transition is no longer a matter of preference. It is imposed by the requirements of measurement itself.

2. THE EUROPEAN UNION TEMPLATE: INVITATION TO A CHARADE

2.1 Invitation to a Charade

The EUnetHTA template, effectively the NICE reference case writ large, has no status as an analytical framework. As demonstrated in the logit assessments, it is a structured expression of measurement inversion. Member states of the European Union are, in effect, being invited to set aside the axioms of representational measurement in favor of a framework that applies arithmetic where measurement has not been established.

The reference case invites participation in a process that presents itself as analytical rigor while denying the conditions required for measurement. It offers a structure in which numbers are generated, compared, and debated without establishing whether those numbers correspond to measurable attributes. It is therefore not merely a flawed methodology. It is an invitation to engage in a charade: a system in which the appearance of evidence substitutes for its existence.

Participants are not compelled to accept this framework. They are drawn into it through institutional expectation and professional convention. Submissions must conform to reference case requirements; journals expect cost-effectiveness analyses; teaching programs train students in accepted methods. Each step reinforces the next. The result is a closed system in which compliance is equated with competence, and where questioning the framework places the participant outside accepted practice.

The charade is sustained by its apparent internal coherence. Models are constructed with care, assumptions are documented, and sensitivity analyses are performed. Within the framework, these activities are considered entirely consistent. Yet they do not address the prior question: whether the constructs being manipulated are measurable. That question is not asked because the framework does not require it. Arithmetic proceeds without demonstration that it is admissible. Outputs are produced that cannot support empirical test, yet are treated as if they were evidence.

This creates a paradox for participants. On the one hand, they are engaged in activities that appear to meet the standards of scientific inquiry. On the other, the absence of measurement means that those standards are not satisfied. The distinction is obscured by the language of analysis. Terms such as “estimate,” “model,” and “uncertainty” suggest approximation to a measurable reality. In the absence of measurement, there is no such reality to approximate.

The invitation is reinforced by apparent utility. The framework produces determinate outputs and allows decisions to be framed as evidence-based. These benefits are not trivial. They address administrative and policy needs. Yet they do so at the cost of scientific validity. A system that delivers conclusions without the possibility of refutation does not generate knowledge. It never could, because each step in constructing a cost-effectiveness claim denies the requirements of measurement. The logit deconstruction is definitive: the NICE analytical model has no admissible meaning. It is sustained through institutional convention.

2.2 The Exclusion of Measurement

The issue is not simply that contemporary HTA denies the requirements of measurement. More precisely, it operates within a framework that excludes them. The reference case is not a neutral analytical structure that has failed to incorporate measurement principles; it is constructed in such a way that those principles cannot be applied. At each stage of the modeling process—definition of utilities, construction of composite outcomes, and derivation of cost-effectiveness ratios—the conditions required for measurement are set aside. The result is not an imperfect approximation to measurement, but a framework in which measurement is structurally absent.

This is why the reference case cannot be regarded as a competing approach. A competing framework would offer an alternative means of generating measurable claims. The reference case does not do this. It proposes a simulation-based structure in which claims for cost-effectiveness are derived through a sequence of operations that deny the axioms of representational measurement. Utilities lack unidimensionality, composite outcomes lack dimensional homogeneity, and arithmetic operations are applied without regard to scale type. These are not correctable features. They define the framework itself.

The consequence is that the outputs of the reference case do not stand alongside measurement-based claims as alternative forms of evidence. They occupy a different category altogether. They are internally consistent constructions that simulate evidence while excluding the conditions required for its existence. There is therefore no meaningful sense in which they can be compared with, or substituted for, claims grounded in measurement. The issue is not one of methodological pluralism, but of admissibility.

Once this is recognized, the apparent choice between frameworks disappears. There is no competition between the reference case and measurement-based approaches, because only one of these can generate claims that are capable of empirical evaluation. The other cannot. The transition is therefore not a matter of selecting between alternatives, but of restoring the conditions under which claims can be said to have meaning.

2.3 The Only Legitimate Claims

Meaningful claims for therapy impact can be expressed through two, and only two, forms of legitimate claim. The first concerns manifest attributes: observable events, durations, or resource units. These include hospitalizations avoided, time to event, days in remission, or resource use. When properly defined and unidimensional, these attributes can be measured on linear ratio scales and evaluated directly.

The second concerns latent attributes: symptoms, functioning, need fulfillment, or patient experience. These cannot be observed directly and therefore cannot be scored or summed meaningfully. They require formal measurement through Rasch models to produce invariant logit ratio scales.

These two forms are sufficient. They are also exclusive. Each claim must be tied to a single attribute, defined by population, comparator, and timeframe, and supported by a measurement

structure that satisfies the axioms of representational measurement. Most importantly, each claim must be capable of failure. A meaningful claim is one that can be shown to be wrong.

Composite constructs such as QALYs do not fail in this sense. They persist regardless of outcome because they are insulated by assumptions. They are recalculated rather than tested. As a result, they cannot support learning. The evolution of objective knowledge regarding therapy impact is therefore excluded from the framework.

By re-centering formulary evaluation on single-attribute claims, health systems regain control of assessment. Decisions become grounded in observable or measurable change rather than model-generated outputs. Evidence becomes something that can accumulate through replication and refutation, rather than being reconstructed for each submission.

While this framework does not compel acceptance, the constraint imposed by measurement standards is unavoidable. Attributes are either manifest or latent. Each requires a specific form of ratio measurement. These conditions are not optional. They determine whether a claim can be evaluated.

2.4 The Futility of a Common Template

The insistence on a common template within the European Union does more than standardize procedure; it eliminates the conditions necessary for meaningful evaluation. By imposing a uniform framework built on constructs that do not satisfy the axioms of representational measurement, the process effectively discards the possibility of generating empirically evaluable claims. In seeking harmonization, the framework does not preserve what is essential and remove what is extraneous; it removes the essential altogether. The result is that, in attempting to impose methodological consistency, the system abandons the very conditions required for evidence to exist.

There is no defensible justification for this form of standardization. A common template might be appropriate where the underlying measurement framework is sound and where shared methods enhance comparability across jurisdictions. That is not the case here. The reference case does not provide a measurement foundation. It provides a set of conventions that apply arithmetic to non-measurable constructs. To standardize such a framework is not to promote comparability, but to replicate a structural error across all member states. Harmonization, in this context, does not advance knowledge; it constrains it.

The consequence is a loss of analytical sovereignty. Individual health systems are deprived of the ability to define, measure, and evaluate therapy impact in terms that reflect their own clinical priorities and resource structures. Decisions are instead framed within a template that dictates both the form of claims and the methods by which they are derived, irrespective of whether those claims are measurable. This is not coordination; it is substitution. It replaces locally grounded, potentially measurable assessments with a centralized structure that cannot support empirical test.

A framework grounded in measurement would achieve the opposite. It would allow for diversity in the selection of attributes while maintaining consistency in the standards applied to their measurement. Member states could define claims relevant to their populations and health system objectives, provided those claims meet the requirements of unidimensionality, invariance, and dimensional homogeneity. Comparability would arise not from uniformity of method, but from adherence to shared measurement principles.

The present approach reverses this logic. It enforces uniformity of method while disregarding the conditions required for measurement. In doing so, it not only fails to deliver comparable evidence, but prevents its development. The imposition of a common template is therefore not a neutral administrative choice. It is a constraint that limits the capacity of health systems to generate, evaluate, and act upon measurable claims.

3. THE TRANSITION TO MEANINGFUL THERAPY IMPACT CLAIMS

It has been made clear in each of the logit assessments that if health technology assessment is to have a meaningful future, it must abandon the pursuit of reference case frameworks and the continued reliance on false measurement. The imposition of a standardized approach, most notably associated with the National Institute for Health and Care Excellence (NICE) has created the appearance of methodological rigor while embedding a fundamental error. This is not a matter of interpretation or refinement. It is a matter of admissibility. A framework that applies arithmetic to constructs that have not been demonstrated to be measurable cannot be rescued through improved data, more sophisticated modeling, or greater transparency. The rejection of the reference case is therefore not controversial once the axioms of representational measurement are accepted. It is unavoidable.

Those axioms have been established for more than half a century. They define the conditions under which numerical representations correspond to empirical attributes and under which arithmetic operations are meaningful. If these conditions are not met, numbers do not approximate reality; they cease to have any meaning whatsoever. The reference case framework violates these conditions at every stage. It assumes unidimensionality without demonstration, imposes interval or ratio properties where none exist, and aggregates constructs that lack dimensional homogeneity. Each of these steps is individually sufficient to invalidate a claim. Taken together, they constitute a system that is internally incoherent. The case against the reference case is therefore not incremental but decisive. It cannot be reformed because its defining operations are inadmissible.

The solution is, by contrast, straightforward. Where measurement precedes arithmetic, there are only two logically acceptable forms of claim. The first concerns manifest attributes: events, counts, durations, and resource use that are directly observable and, when properly defined, possess the properties required for linear ratio measurement. These attributes support addition, multiplication, and comparison because they have a true zero and invariant units. The second concerns latent attributes: symptoms, functioning, need fulfillment, and patient experience, which cannot be observed directly and therefore cannot be reduced to scores through summation or weighting. Their measurement requires a formal model capable of producing invariant units. The Rasch framework provides this capability, yielding logit ratio scales that satisfy the conditions required for comparison and transformation.

These two forms are sufficient. They are also exclusive. A claim for therapy impact must be expressed in terms of a single attribute, defined by population, timeframe, and comparator, and supported by a measurement standard appropriate to that attribute. It must be defensible, in that its measurement properties are demonstrated; replicable, in that it can be reproduced under defined conditions; and falsifiable, in that it can be shown to fail. These are not new requirements. They are the standards that have governed empirical inquiry since the scientific revolution of the seventeenth century. Their absence from contemporary HTA is not a matter of evolution, but a shift to valuing composite health state descriptions. This represents a movement away from the measurement of attributes toward the construction and valuation of multidimensional representations that do not satisfy the conditions required for measurement.

The reference case framework has achieved only one substantive outcome: it has delayed the application of these standards. For more than forty years, HTA has operated within a structure that substitutes composite, model-driven constructs for measurable claims. This has created a closed system in which conclusions are derived from assumptions rather than tested against observation. Evidence is generated once, through simulation, rather than accumulated over time through replication and refutation. The consequence is not merely methodological weakness, but the absence of a mechanism for learning.

This does not imply that the transition to measurement should be accepted uncritically. On the contrary, it invites scrutiny of a different kind: scrutiny directed at the measurement properties of claims rather than the internal consistency of models. If the response to the Maimon Research website is any guide, with substantial engagement across European HTA communities, there is already recognition that the current framework offers no pathway forward. The level of interest suggests that the constraints imposed by a standardized reference case are increasingly seen as limiting rather than enabling.

The implication is clear. The future of HTA does not lie in the continued refinement of a framework that fails the conditions of measurement. It lies in the re-establishment of those conditions as the foundation of all claims. This is not a radical departure. It is a return to normal science. It restores the requirement that claims be measurable, that evidence be testable, and that knowledge evolve through the possibility of failure.

3.1 The Attribute is Foundation

Once the standards of normal science are recognized as the only legitimate basis for evaluating therapy impact claims, the attribute becomes the indispensable foundation for assessment. This is not a matter of analytical preference or methodological style. It follows directly from the requirement that measurement must precede arithmetic. Unless a claim is anchored in an attribute that can be defined, measured, and tested, there is no basis upon which numerical operations can be justified. The attribute is therefore not simply a component of evaluation; it is the condition that determines whether evaluation is possible.

An attribute is a property or characteristic that can be meaningfully ascribed to a defined population over a specified timeframe. It is not an abstraction created through aggregation, nor a composite constructed to summarize multiple effects. It must be unidimensional by construction, representing a single aspect of therapy impact that can be isolated, defined, and subjected to measurement. This distinction is critical. Contemporary HTA has moved away from attributes toward composite constructs—most notably utilities and QALYs—that combine disparate dimensions into a single value. These constructs are ordinal, multidimensional, and lack the properties required for measurement. They are representations, not measures. The attribute, by contrast, is the point at which representation becomes measurement.

The discipline imposed by the attribute requirement is immediate and consequential. Every claim must identify a specific attribute of interest, justify its relevance to the clinical or policy question, and define how it is to be measured. This includes explicit linkage to a target population, a comparator, and a timeframe for evaluation. Without this specification, claims remain vague and

cannot be subjected to empirical assessment. Statements such as “overall benefit,” “value,” or “quality-adjusted outcomes” are not attributes. They are interpretive summaries that obscure rather than clarify the underlying phenomena. They cannot be measured because they do not correspond to a single, coherent property.

Attributes must also be classified according to their observability. Manifest attributes such as events, durations, and resource use are directly observable and, when properly defined, can support linear ratio measurement. Their defining characteristics include a true zero and invariant units, allowing for addition, multiplication, and comparison. Latent attributes such as symptom burden, functional status, or need fulfillment cannot be observed directly. They require construction through a measurement model that enforces unidimensionality and invariance. The Rasch framework provides the only established method for achieving this, yielding logit ratio scales that support meaningful comparison. This distinction between manifest and latent attributes is not merely descriptive. It determines the form of measurement that is admissible and the operations that can be performed.

The rejection of composite constructs follows necessarily from this classification. Dimensional homogeneity is a precondition for arithmetic. Where attributes differ in kind, they cannot be combined without violating the conditions required for measurement. The aggregation of multiple attributes into a single index does not create a more comprehensive measure; it creates a quantity that lacks measurement properties altogether. This is the defining error of the current HTA framework. By attempting to collapse multiple dimensions into a single value, it abandons the conditions required for any dimension to be measured.

Each attribute must therefore stand alone, with its own measurement structure and its own protocol for evaluation. Claims must be formulated in terms of these individual attributes, not as components of a composite outcome. This restores clarity to the assessment process. It allows each claim to be defined, measured, and tested independently, and it permits evidence to accumulate through replication and potential refutation. The attribute becomes the unit of analysis, replacing the composite construct as the basis for decision-making.

In this sense, the attribute is the dividing line between representation and measurement. On one side lie constructs that summarize, combine, and simulate, but cannot be evaluated empirically. On the other lie attributes that can be defined, measured, and subjected to test. The transition to measurement-based HTA is therefore, at its core, a transition from composite representation to attribute-based measurement. Without this shift, the conditions required for meaningful claims cannot be established. With it, the framework for evaluation is restored to the standards that define scientific inquiry.

3.2 From Understanding to Implementation

A transition to measurement-based health technology assessment requires more than recognition of analytical failure. It requires a structured reorientation of both conceptual understanding and operational practice. The distinction between training and implementation is not a matter of presentation; it is fundamental. Without a clear understanding of measurement, implementation becomes inconsistent and collapses into existing conventions. Without implementation,

understanding remains theoretical and has no impact on decision-making. The transition must therefore proceed in parallel across both domains.

The starting point is the recognition that numerical outputs are not, in themselves, evidence. Numbers may summarize, describe, or simulate, but only measurement can support inference. This requires a disciplined understanding of what constitutes a measurable attribute. An attribute must be defined, unidimensional, and capable of supporting operations that are admissible given its scale type. The failure of contemporary HTA lies in the assumption that constructs such as utilities, QALYs, and model-generated outputs meet these conditions. They do not. Training must therefore begin with the distinction between representation and measurement, and with the consequences of applying arithmetic to quantities that do not satisfy the axioms of representational measurement.

The second requirement is mastery of those axioms. Unidimensionality, invariance, and dimensional homogeneity are not optional properties; they are preconditions for the meaningful use of numbers. Scale type determines permissible operations. Linear ratio scales, characterized by a true zero and constant units, support addition, multiplication, and comparison. Latent attributes cannot be observed directly and require construction through a measurement model capable of enforcing these conditions. The Rasch framework provides the only established approach for achieving this, yielding invariant logit ratio scales. Without an understanding of these principles, there is no basis for determining whether a claim is admissible.

The third requirement is the reorientation of HTA toward attribute-based claims. The current framework is characterized by the construction of composite outputs that combine multiple attributes into a single value. This must be reversed. Each claim must be defined in terms of a single attribute, specified by population, comparator, and timeframe. Manifest attributes such as events, durations, and resource us can be measured directly on linear ratio scales when properly defined. Latent attributes symptoms, functioning, and patient experience require formal measurement through Rasch models. The critical restriction is that these claims cannot be combined. Dimensional homogeneity is a precondition for arithmetic. Without it, aggregation is not meaningful.

Implementation proceeds through the restructuring of submissions. Rather than presenting model outputs as primary evidence, submissions must be organized around a limited number of explicit claims. Each claim must identify the attribute, define the population and comparator, specify the timeframe, and demonstrate the measurement properties of the attribute. This shifts the focus from model construction to claim definition. Models may still have a role as exploratory or descriptive tools, but they cannot serve as the basis for claims unless they are anchored in measurable attributes.

A further requirement is the introduction of measurement as a gatekeeper. No arithmetic operation is admissible unless the measurement properties of the underlying quantity have been demonstrated. For manifest attributes, this requires confirmation of ratio properties and unit consistency. For latent attributes, it requires demonstration of Rasch model fit, invariance, and scale functioning. Where these conditions are not met, outputs may be reported descriptively, but they cannot support claims.

Finally, implementation requires the development of protocols for empirical assessment. Each claim must be treated as a proposition to be tested. This requires specification of data sources, observation periods, and evaluation criteria. Claims must be capable of replication and potential refutation. This introduces an iterative process in which evidence accumulates over time, replacing the one-time generation of model outputs with continuous empirical evaluation.

The transition does not require immediate abandonment of existing HTA frameworks. A staged approach is possible. Measurement-valid claims can be introduced alongside current submissions, providing a parallel track for evaluation. Over time, as familiarity and confidence develop, these claims can displace composite constructs that lack measurement validity. The endpoint is a framework in which all claims are grounded in measurable attributes and subject to empirical test.

3.3 Formulary Submission Standards

A formulary submission guideline grounded in measurement has a single purpose: to ensure that claims presented to decision-makers meet the conditions required for empirical evaluation. This requires the explicit integration of measurement standards and protocol standards within the submission process. Without these, submissions revert to the presentation of non-evaluable constructs, regardless of their analytical complexity.

The first requirement is the enforcement of measurement standards. Only two forms of measurement are admissible. For manifest attributes \claims must be expressed on linear ratio scales. These attributes must be defined with a true zero, invariant units, and clear operational definitions that permit addition, multiplication, and comparison. For latent attributes such as symptoms, functioning, and patient experience claims must be supported by Rasch-based measurement, yielding invariant logit ratio scales. Summation, weighting, or scoring of questionnaire responses is not acceptable. Unless the measurement properties of an instrument have been demonstrated, the resulting values cannot support claims.

The implication of this requirement is immediate. Composite constructs are disallowed. Claims must be restricted to single attributes that satisfy the conditions for measurement. Where measurement is not demonstrated, arithmetic is not permitted. The guideline therefore functions as a gatekeeper. No claim enters the evaluative process unless its measurement status has been established.

The second requirement concerns protocol standards. A claim is not an output of a model; it is a proposition to be tested. Each submission must therefore include a protocol that defines the claim in explicit terms. This includes identification of the attribute, the target population, the comparator, the timeframe over which the claim is to be evaluated, and the method of measurement. The protocol must specify how data will be collected, how the attribute will be measured, and how the claim will be assessed.

Protocols must be prospective and transparent. Retrospective justification of claims through modeling exercises is not acceptable. Where existing data are used, their relevance and measurement properties must be demonstrated. Where new data are required, the protocol must define the conditions under which they will be collected. The emphasis is on reproducibility. A

claim that cannot be reproduced under defined conditions cannot contribute to the accumulation of knowledge.

A further requirement is that evaluation be iterative. Claims are not accepted as final outputs. They are subject to ongoing assessment, with results reported over defined intervals. This allows claims to be confirmed, modified, or rejected based on evidence. It replaces the static presentation of model outputs with a dynamic process of empirical learning over the lifetime of a product.

Finally, the guideline must ensure that decision-makers are equipped to enforce these standards. This requires familiarity with representational measurement and the ability to distinguish between measurable and non-measurable claims. Without this capability, the guideline cannot function as intended. The evaluation process becomes vulnerable to the reintroduction of non-measurable constructs under the guise of analytical sophistication.

In summary, a formulary submission guideline grounded in measurement replaces the submission of composite, non-evaluable outputs with a disciplined framework for assessing therapy impact. It restricts claims to measurable attributes, requires protocols that support empirical test, and enforces standards that ensure reproducibility and falsifiability. In doing so, it restores the conditions under which formulary decisions can be informed by evidence rather than by internally consistent but empirically unsupported constructions.

3.4 Implementation Support and Capability Development

The transition outlined in this document imposes requirements that extend beyond conceptual acceptance. It demands the development of a level of capability that is not currently embedded within HTA practice. Participants must be able to distinguish between measurable and non-measurable claims, understand the implications of scale types for admissible operations, and apply these principles consistently in the construction and evaluation of therapy impact claims. These are not competencies that can be assumed. Their absence is the reason the current framework has persisted.

This creates a capability gap. The existing HTA knowledge base, as demonstrated through repeated logit interrogations, does not provide the foundations required to support measurement-based evaluation. Training programs emphasize modeling, statistical analysis, and data interpretation, but do not address the prior question of whether the quantities being analyzed are measurable. As a result, participants are equipped to operate within the reference case framework, but not to replace it. The transition therefore requires a structured approach to capability development.

Maimon Research has developed a set of materials designed to address this requirement. These materials are not supplementary to the framework presented here; they are its operational extension. They provide a systematic treatment of representational measurement, including the distinction between manifest and latent attributes, the role of unidimensionality and invariance, and the application of Rasch measurement to latent constructs. They also address the design of protocol-driven claims and the restructuring of submissions to support empirical evaluation.

The significance of these materials lies in their scope and integration. They do not present isolated concepts or techniques. They define a coherent framework that links measurement theory to HTA practice, from claim definition through to evaluation. In this respect, they represent a structured implementation pathway rather than a set of optional resources. At present, there is no alternative body of work that provides this level of integration between measurement standards and HTA application.

Access to these materials is provided through a sequence of distance education programs available via the Maimon Research website [www.maimonresearch.com]. These programs reflect the progression required for transition: from understanding the principles of measurement, to applying those principles in the construction of claims, and finally to embedding them within institutional processes. They are intended to support individual researchers, academic programs, and HTA organizations in developing the competencies required for implementation.

This support should not be interpreted as a substitute for institutional change. It is a prerequisite. Without the development of measurement capability, the transition outlined in this document cannot be realized in practice. The framework has been defined. The constraints are clear. The remaining requirement is the ability to operate within those constraints.

The transition to measurement-based HTA is therefore not limited by theory or by the availability of data. It is limited by capability. Until that capability is established, the existing framework will persist, not because it is defensible, but because there is no structured alternative that can be implemented. The materials developed by Maimon Research are intended to provide that alternative in operational form.

MAIMON RESEARCH LLC

DISTANCE EDUCATION PROGRAMS IN THE THEORY OF MEASUREMENT

Three programs are available: two short 5-module programs and a 12-module program that is structured as a senior level course on the transition from the current HTA belief system to a new paradigm for HTA

The two short programs are (i) **NUMERICAL STORYTELLING: SYSTEMATIC MEASUREMENT FAILURE IN HEALTH TECHNOLOGY ASSESSMENT** and (ii) A **NEW START IN MEASUREMENT FOR HEALTH TECHNOLOGY ASSESSMENT.** They are designed to complement the 12-module course program. They can be accessed through the **DISTANCE EDUCATION** section of the website with URL https://maimonresearch.com/distance-education-programs/

The senior level course **HEALTH TECHNOLOGY ASSESSMENT REBUILT: EVIDENCE AND VALUE** is accessed through the EVIDENCE AND VALUE section of the website or URL link https://maimonresearch.com/evidence-and-value/.

CONCLUSION

The argument presented in this paper leads to a single, unavoidable conclusion. The current framework for health technology assessment, centered on reference case modeling and composite cost-effectiveness constructs, does not meet the conditions required for measurement. As a result, it cannot generate claims that are empirically evaluable, reproducible, or falsifiable. Its outputs do not constitute evidence. They are numerical constructions that simulate evidence while violating the axioms upon which meaningful measurement depends.

This is not a matter of methodological preference or incremental refinement. It is a matter of admissibility. Once the axioms of representational measurement are recognized, the status of existing HTA practice is no longer ambiguous. The application of arithmetic to non-measurable quantities cannot be justified, regardless of the sophistication of the models employed or the quality of the data inputs. The reference case framework cannot be repaired. It must be set aside.

The alternative is both limited and sufficient. Therapy impact claims must be expressed in terms of single attributes that satisfy the conditions for measurement. For manifest attributes, this requires linear ratio scales with a true zero and invariant units. For latent attributes, it requires Rasch-based logit ratio measurement that ensures unidimensionality and invariance. These two forms of measurement define the boundary of admissible claims. There is no competing framework that can produce evaluable outcomes.

The transition to this framework is not optional. It is imposed by the requirements of scientific inquiry. Claims must be capable of empirical test. Evidence must be subject to replication and potential refutation. Without these conditions, there is no mechanism for learning and no pathway for the evolution of objective knowledge. The continued use of composite constructs and model-generated outputs offers only the repetition of internally consistent but empirically meaningless results.

For the European Union, the implications are direct. Harmonization of HTA methods is both desirable and necessary. However, harmonization must be grounded in measurement. The current trajectory, standardizing a framework that lacks a foundation in representational measurement, appears designed to deny a viable system for evaluating therapy impact. It can only perpetuate the appearance of consistency without its substance.

The transition outlined in this paper provides a practical alternative. It replaces composite, non-measurable constructs with single-attribute claims that can be defined, measured, and tested. It introduces protocols that support empirical evaluation and establishes measurement as the gatekeeper for admissible operations. It also recognizes that implementation requires capability, and that this capability must be developed systematically.

The choice is therefore clear. Continue with a framework that produces results that cannot be known to be true or false, or adopt one that restores the conditions under which claims can be evaluated and knowledge can evolve. The latter is not a reform of HTA. It is its reconstruction on scientific foundations.

What remains is not a question of theory, but of implementation. The transition to measurement-based HTA is constrained not by the availability of data or analytical tools, but by the capability to apply the principles outlined in this paper. The existing knowledge base does not provide this capability. It must be developed. This places responsibility squarely with institutions, agencies, and academic programs to ensure that participants are equipped to distinguish measurable from non-measurable claims and to operate within a framework that supports empirical evaluation. The alternative is continued reliance on a system that cannot produce evidence. The framework has been defined. The requirements are clear. The only remaining question is whether they will be acted upon.

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.