MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

UNITED STATES: PHARMACY COLLEGES AND SCHOOLS - MEASUREMENT INVERSION AND ARITHMETIC CHAOS IN HTA PROGRAMS

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

LOGIT WORKING PAPER No 337 JUNE 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

For the first time, large language model interrogations of 71 Colleges and Schools of Pharmacy in the United States have transformed concerns regarding the scientific foundations of health technology assessment (HTA) into a measurable observation. The findings reveal a pervasive pattern of measurement inversion, where arithmetic routinely precedes measurement. This paper argues that the implications are profound. Measurement inversion undermines the scientific legitimacy of the reference-case framework that has dominated HTA teaching, research and policy for more than four decades.

The analysis focuses on six leading pharmacy colleges and schools and evaluates endorsement probabilities for canonical statements drawn from the scales of measurement and the axioms of representational measurement. The results demonstrate a remarkably stable pattern. False propositions central to utilities, QALYs and reference-case modelling receive consistently strong endorsement, while true propositions concerning unidimensionality, ratio measurement, measurement before arithmetic, Rasch measurement and latent trait possession receive consistently weak endorsement. The findings suggest that the HTA knowledge base embedded within PharmD education normalizes arithmetic operations while neglecting the measurement requirements necessary to justify them.

This pattern is interpreted as evidence of arithmetic chaos. Beginning with ordinal health-state descriptions and preference scores, the reference-case framework proceeds through utility construction, QALY estimation and simulation modelling without establishing the lawful measurement properties of the entities involved. The result is a chain of inadmissible arithmetic operations that culminates in cost-effectiveness claims lacking recognized quantitative meaning. Increasing model sophistication, sensitivity analysis and computational complexity do not resolve this problem because the failure is structural rather than technical.

The paper argues that the reference-case framework has reached closure. Closure refers not to the disappearance of HTA programs but to recognition that the framework can no longer justify its claims according to the standards of quantitative science. The proposed alternative is a return to normal science through lawful measurement. Manifest attributes require linear ratio measures, while latent attributes require Rasch logit ratio measures. Both support empirical evaluation, replication and falsification. The conclusion is that PharmD education in HTA should be reconstructed to underwrite a transition from reference-case modelling to a measurement-based framework capable of supporting credible, evaluable and scientifically defensible claims regarding therapy impact.

INTRODUCTION

For the first time, interrogation of 71 Colleges and Schools of Pharmacy has transformed what was previously difficult to demonstrate into a measurable observation: a pervasive pattern of measurement inversion. The implications are immediate. HTA programs in the US has reached the stage of closure.

Closure does not imply that HTA programs disappear, journals cease publication or simulation models cease to be constructed. Closure has a more precise meaning. It refers to final recognition of the lack of scientific legitimacy of the existing HTA teaching framework. Once arithmetic routinely precedes measurement, the pathway through which quantitative claims acquire meaning is disrupted. The framework becomes closed because it no longer possesses the capacity for correction from within its own assumptions.

Interrogation of the respective HTA knowledge bases reveals a structural failure that prevents the existing framework from continuing as a credible quantitative science. For more than four decades HTA evolved around reference-case methodologies built upon utilities, QALYs, cost-per-QALY thresholds and simulation models. Increasing analytical sophistication created an appearance of scientific rigor. Larger Markov models, expanded sensitivity analyses and increasingly elaborate assumptions reinforced confidence that quantitative uncertainty was being addressed. Yet a prior question remained largely absent: did the attributes supporting these numerical operations satisfy the requirements of representational measurement?

Once this question is asked, the consequences become unavoidable. Ordinal composite utilities derived from health-state descriptions, ordinal preference scores and composite measures fail to demonstrate lawful measurement properties. Arithmetic operations therefore become detached from measurable attributes, yet these same numerical constructions continue to support PharmD programs with cost-effectiveness ratios, pricing decisions and policy recommendations. The result is measurement inversion: arithmetic creating conclusions before measurement has been established.

This process created closure because the reference-case framework of HTA evolved into a self-reinforcing ordinal analytical system. Assumptions generated evidence; simulations substituted for empirical challenge; sensitivity analyses examined parameter uncertainty without questioning whether the measurement foundations themselves were lawful. Internal refinement continued, but criticism of underlying assumptions remained external to the framework itself. Closed systems protect themselves from falsification and cease to contribute to the evolution of objective knowledge.

Recognition of closure immediately raises a second question: what replaces it? Incremental revision cannot solve the problem. More sophisticated utilities, revised assumptions and larger simulations merely reproduce the same error. Replacement becomes necessary because the problem lies not with specific models but with the sequence itself: measurement must precede arithmetic.

The proposed replacement is a return to the standards of normal science. Manifest attributes require linear ratio measures; latent attributes require Rasch logit ratio measures. Attribute claims become explicit propositions linked to protocols and exposed to empirical challenge. Closure therefore describes the end of a framework built upon measurement inversion; replacement describes transition to a framework capable of restoring scientific accountability.

THE EVIDENCE FOR MEASUREMENT INVERSION

A series of HTA knowledge interrogations was undertaken involving, in the first instance 71 Colleges and Schools of pharmacy followed by a detailed assessment reported on here of the extent to which the xix leading Colleges and Schools in the US endorsed true and false statements to capture the understanding of representational measurement and the scales of measurement. These findings were important because they transformed what had previously been criticism and suspicion into a measurable observation. Measurement inversion appeared not as an isolated methodological weakness but as a defining characteristic of the US pharmacy HTA knowledge structure. Foundational propositions concerning representational measurement, lawful arithmetic and Rasch measurement received weak endorsement while false measurement assumptions received strong support.

Importantly, the US pharmacy findings proved not to be unique. The endorsement patterns observed in these HTA knowledge bases represented the mirror image of what would be expected under normal science: false propositions received strong support while true propositions were repeatedly sidelined. More importantly still, identical endorsement profiles emerged across a large number of international interrogations. Interrogations involving twenty-four countries, together with agencies, journals, educational institutions and professional organizations, produced remarkably stable findings. Across more than 230 assessments the same pattern repeatedly appeared.

The implications extend beyond HTA itself. Interrogations of COSMIN and the Cochrane Collaboration suggested that measurement inversion may be embedded not only within HTA but within the broader architecture of evidence generation and evidence synthesis. Instrument standards, systematic reviews and meta-analysis appeared to inherit assumptions concerning measurement that largely ignored representational measurement. If so, the issue extends beyond reimbursement frameworks and reaches into the institutional assumptions supporting evidence itself.

At the same time, separate interrogations of the American Association of Colleges of Pharmacy (AACP), the American Foundation for Pharmacy Education (AFPE) and the National Pharmaceutical Council (NPC) all demonstrate measurement inversion.

Table 1 presents a comparison of endorsement probabilities for six US pharmacy colleges and schools based on the results of published interrogations, It should be noted the findings relate the HTA knowledge base, curated beliefs, and not to faculty or staff within those institutions, Stable inversion across institutions is different. It indicates the presence of a reproducible knowledge structure. Examined collectively, the pattern is remarkably stable and reveals a profile consistent with what has been described throughout these assessments as measurement inversion: the reversal of the scientific requirement that measurement must precede arithmetic.

TABLE 1

MEASUREMENT INVERSION: FAILURE TO MEET REQUIRED STANDARDS

CANONICAL STATEMENT	CATEGORICAL PROBABILITY ENDORSEMENT					
	Eshelman	Minnesota	Michigan	Florida	Kentucky	Ohio
FALSE STATEMENTS						
1, Ratio measures can have negative values	0.90	0.90	0.90	0.90	0.95	0.95
2.The QALY is a ratio measure	0.75	0.90	0.95	0.90	0.95	0.95
3, Summations of subjective instrument responses are ratio measures	0.90	0.90	0.90	0.85	0.90	0.90
4. The QALY is a dimensionally homogeneous measure	0.95	0.90	0.95	0.85	0.95	0.95
5. Reference case simulations generate falsifiable claims	0.85	0.85	0.90	0.85	0.95	0.90
TRUE STATEMENTS.,						
6. Measures must be unidimensional	0.15	0.10	0.25	0.15	0.20	0.20
7. Multiplication requires a ratio measure	0.10	0.10	0.15	0.10	0.10	0.10
8. There are only two classes of measurement: linear ratio and Rasch logit ratio	0.05	0.10	0.10	0.05	0.05	0.05
9.Measurement precedes arithmetic	0.15	0.10	0.20	0.10	0.10	0.15
10. The outcome of interest for latent traits is the possession of that trait	0.05	0.10	0.30	0.25	0.30	0.30

The colleges are : University of North Carolina Eshelman School of Pharmacy; University of Minnesota College of Pharmacy; University of Michigan College of Pharmacy; University of Florida College of Pharmacy; University of Kentucky College of Pharmacy and Ohio State University College of Pharmacy

The results in Table 1 provide a striking and highly consistent profile of measurement inversion across six leading United States pharmacy colleges and schools: Eshelman, Minnesota, Michigan, Florida, Kentucky and Ohio State. The pattern is not subtle. Statements that are false under the standards of scales of measurement and representational measurement receive consistently high endorsement, while statements that are true and foundational for lawful quantitative claims receive consistently weak endorsement. The implication is that the relevant knowledge bases appear to normalize the arithmetic structures of contemporary health technology assessment while failing to recognize the measurement conditions necessary to justify them.

The false statements are especially revealing because they map directly onto the core assumptions supporting utilities, QALYs, reference-case modelling and the cost-effectiveness framework. The proposition that ratio measures can have negative values receives endorsement probabilities of 0.90 to 0.95 across all six institutions. This is a foundational error. Ratio measures require a meaningful non-arbitrary zero and cannot take negative values. Yet much of the QALY framework accommodates negative utilities for states considered worse than death. The high endorsement probabilities therefore signal acceptance of a measurement structure incompatible with ratio measurement.

The claim that the QALY is a ratio measure is also strongly endorsed, ranging from 0.75 at Eshelman to 0.95 at Michigan, Kentucky and Ohio State. This is central to the diagnosis of measurement inversion. If the QALY is treated as a ratio measure, then it becomes available for aggregation, comparison, division and cost-per-QALY calculation. Yet the QALY is derived from utility structures that do not possess demonstrated ratio properties. Treating it as a ratio measure licenses precisely the arithmetic operations that should first be disallowed.

The endorsement of summated subjective instrument responses as ratio measures is similarly high, ranging from 0.85 to 0.90. This points to a broader failure beyond the QALY alone. It suggests that ordinal response scores are treated as if they automatically become measures when summed. This is false. Summation of ordinal responses does not create interval or ratio measurement. It merely produces another ordinal structure unless a lawful measurement model, such as Rasch measurement, is used to construct an invariant latent scale.

The statement that the QALY is dimensionally homogeneous receives even stronger endorsement, with probabilities of 0.85 to 0.95. This is troubling because dimensional homogeneity is a basic requirement for meaningful quantitative operations. The QALY combines a utility construct with time, yet there is no demonstrated dimensional compatibility between an ordinal or composite ordinal utility and a ratio measure of time. Multiplication by time cannot transform an inadmissible utility into a lawful outcome measure.

The fifth false statement, that reference-case simulations generate falsifiable claims, also receives high endorsement across all six schools, from 0.85 to 0.95. This indicates acceptance of simulation outputs as if they were empirical scientific claims. Yet reference-case simulations typically generate projected estimates based on assumptions, model structures and hypothetical future pathways. They may be internally elaborate, but that does not make them falsifiable in the normal scientific sense. A claim insulated within assumptions is not equivalent to an empirically evaluable proposition.

The true statements show the inverse pattern. The requirement that measures must be unidimensional receives weak endorsement, from 0.10 at Minnesota to 0.25 at Michigan. This is remarkable because unidimensionality is one of the most basic requirements for lawful measurement. If a measure is not unidimensional, then arithmetic operations applied to it lack clear quantitative interpretation. The weak endorsement suggests that the construct requirements for measurement are not central to the knowledge base.

The proposition that multiplication requires a ratio measure receives even weaker endorsement, clustered at 0.10 with only Michigan at 0.15. This is perhaps the clearest signal of measurement inversion. Multiplication is central to QALY construction, where utilities are multiplied by time. If the requirement for ratio measurement is not recognized, then the entire arithmetic chain of cost-utility analysis proceeds without the necessary measurement foundation.

The statement that there are only two lawful classes of measurement for therapy impact—linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes—receives near-floor endorsement, between 0.05 and 0.10. This indicates that the distinction between manifest and latent attributes is largely absent. It also shows that Rasch measurement, despite its role in transforming subjective responses into lawful latent measures, is effectively invisible.

The statement that measurement precedes arithmetic receives endorsement probabilities between 0.10 and 0.20. This is the core result. If leading pharmacy colleges and schools weakly endorse the foundational principle that measurement must precede arithmetic, then the conditions for measurement inversion are present at the curriculum and knowledge-base level. Arithmetic is accepted before the entities entering the arithmetic are shown to be measures.

The final true statement, concerning possession of latent traits, receives mixed but still weak endorsement. Eshelman is at 0.05, Minnesota at 0.10, Michigan at 0.30, Florida at 0.25 and Kentucky and Ohio State at 0.30. These are the highest among the true statements but remain weak. They suggest partial recognition that latent traits may be relevant, but not sufficient recognition that latent attributes require a distinct measurement architecture.

The common theme across all six institutions is therefore clear. False propositions that enable the existing reference-case framework are strongly endorsed. True propositions that would challenge that framework are weakly endorsed. This is the signature of measurement inversion. The knowledge bases appear to support arithmetic operations while sidelining the measurement requirements that should govern those operations.

The implications for PharmD education are substantial. These are leading institutions. If their knowledge bases reproduce the same endorsement pattern, then the problem cannot be dismissed as marginal or confined to weak programs. Students are likely being introduced to utilities, QALYs, cost-effectiveness models, ICERs and simulation structures without an equally rigorous grounding in scales of measurement, representational measurement, dimensional homogeneity, Rasch measurement and falsification.

The result is a professional curriculum in which arithmetic is normalized but measurement is not. That is the core danger. Graduates may learn how to interpret or even construct cost-effectiveness claims while lacking the tools to ask whether those claims are quantitatively meaningful in the first place. In this sense, the table points not merely to methodological error but to curriculum failure.

Overall, the table demonstrates a remarkably stable inversion profile. Across six prominent pharmacy colleges and schools, the same structure appears: high endorsement of false QALY and simulation assumptions, weak endorsement of true measurement requirements and near invisibility of Rasch measurement. The conclusion is difficult to avoid. The PharmD-related HTA knowledge base appears to reproduce the same arithmetic chaos found in the broader reference-case framework. Measurement does not precede arithmetic. Arithmetic has become the curriculum.

. ARITHMETIC CHAOS AND THE DECISION MODEL PROCESS

Measurement inversion is the signal for arithmetic chaos in the development and application of the closed decision model or reference case that defines the PharmD belief in meaningful claims for therapy impact assessment. More importantly, inversion signals that the PharmD closed decision framework is not fit for purpose. The reason is straightforward. Beginning in the late 1980s and early 1990s, HTA adopted the belief that all available evidence relating to a new therapy, clinical observations, health-state descriptions, utilities, resource estimates and future assumption, could be collapsed into a single analytical structure capable of generating one definitive claim for comparative cost-effectiveness. While this possessed obvious attractions for health systems seeking simplified decision rules for resource allocation, the framework was disallowed from the outset because it violated the requirements governing lawful quantitative inference.

If we accept, which the PharmD framework does not, that measurement must precede arithmetic, then analysis must begin with measurement theory rather than with model construction. This immediately imposes constraints. Quantitative claims must satisfy the requirements imposed by scales of measurement, particularly the ratio scale, together with the further constraints established by the axioms of representational measurement. These standards are not optional methodological preferences. They determine whether arithmetic operations possess recognized meaning.

Once these requirements are applied, the consequences for therapy impact assessment become straightforward. There are only two lawful forms of measurement. Manifest attributes such as hospitalizations, persistence or switching behavior require linear ratio measures. Latent attributes such as symptom burden, emotional well-being or need fulfilment require Rasch logit ratio measures constructed through transformation of ordinal observations into invariant latent rulers. The two forms of measurement cannot be merged into a single aggregate score or composite claim because they represent fundamentally different quantitative structures. The notion that disparate attributes can be collapsed into a single summary utility or cost-effectiveness estimate is therefore not merely problematic; it is quantitatively incoherent.

The PharmD commitment to simulated decision models is accurately described as arithmetic chaos because no attention is given to how the various numerical components were combined through successive stages of model development. The problem begins immediately with the use of time trade-off valuations to generate utility scores from health-state descriptions. These utilities are then applied to responses from multiattribute health-state questionnaires through scoring algorithms intended to estimate utility values. The resulting utilities are presented as though they represent quantitative measures suitable for arithmetic manipulation. Yet the underlying entities possess no demonstrated measurement properties. At best they remain composite ordinal scores reflecting aggregated preferences across multiple dimensions of health-state description. They ensure measurement failure for all subsequent processes.

Composite ordinal scores cannot support econometric modelling because econometric analysis assumes that the entities entering the model possess lawful quantitative properties. At a minimum, inputs must represent unidimensional measures where arithmetic operations preserve meaning. Ordinal composites fail this requirement. They merely rank observations; they do not establish constant intervals, ratio properties or admissible multiplication. The issue is therefore not whether ordinal scores can be assigned numbers, but whether those numbers support the mathematical manipulations imposed upon them. They do not.

The entire econometric exercise is therefore meaningless from the standpoint of measurement theory. Numerical outputs can certainly be generated, but the claim that these outputs support lawful ratio weights capable of constructing utility algorithms is indefensible. All that is achieved is transformation of one composite ordinal structure into another. For every health-state description the resulting utility score remains a composite ordinal construct lacking demonstrated unidimensionality, invariance and ratio properties. There is no possibility of transformation to any other scale or measure.

The implications are substantial. These utility structures are neither unidimensional nor ratio measures. Indeed, the presence of negative scores for health states described as “worse than death” demonstrates immediately that they cannot represent lawful ratio measures because ratio scales require a meaningful non-arbitrary zero and cannot support negative values. Yet despite lacking the properties necessary for multiplication, these composite ordinal utilities are multiplied by time to construct QALYs.

This operation is disallowed. Time itself is a lawful ratio measure possessing a meaningful absolute zero. However, multiplication requires that both entities entering the operation possess admissible ratio measurement properties. If discounting is applied, then dimensional homogeneity between the discount factor and time must also be preserved. An ordinal or composite ordinal utility cannot be transformed into a lawful quantity merely because it is multiplied by time. The resulting QALY therefore inherits the inadmissibility of the utility structure upon which it is based. By the standards of measurement theory and admissible arithmetic operations, the QALY is an impossible construct.

Yet the decision model process proceeds regardless. The QALY becomes a central input to simulation modelling together with a further bundle of assumptions whose own measurement properties are rarely, if ever, examined. Cost inputs are typically composite aggregates combining unlike resource elements into single numerical structures without regard to dimensional homogeneity or admissible aggregation. Transition probabilities, persistence assumptions and projected future events are incorporated into Markov simulation frameworks to generate estimates of lifetime costs and QALYs.

The final stage compounds the contradiction. Estimated costs are related to estimated QALYs to generate incremental cost-effectiveness ratios presented as though they possess quantitative meaning sufficient to support therapy choice. Yet both numerator and denominator are themselves constructed from entities lacking established measurement properties. The resulting ratio therefore inherits the inadmissibility embedded throughout the analytical chain. What appears as a sophisticated quantitative framework is, in reality, arithmetic chaos reproduced systematically through each stage of the decision model process.

To place this in perspective, it is precisely this sequence of disallowed arithmetic operations that forms the core methodological structure of HTA teaching and research. Utilities, QALYs, Markov simulation models, discounting procedures and cost-effectiveness ratios are presented routinely within HTA courses as though they represent accepted quantitative science. Yet the prior question is never asked: do the entities entering these analytical structures satisfy the requirements imposed by scales of measurement and representational measurement? The issue is not disagreement over assumptions or model specification. The issue is that the arithmetic itself is disallowed.

Questions concerning nominal, ordinal, interval and ratio scales are effectively absent from the curriculum. The axioms of representational measurement receive little or no attention. The distinction between manifest and latent attributes disappears. Rasch logit ratio measurement is ignored. Instead, arithmetic operations proceed independently of measurement constraints as though assigning numbers to observations automatically creates lawful quantities suitable for multiplication, aggregation and ratio construction.

This condition has persisted for more than four decades. Successive generations of HTA faculty, researchers, practitioners and policy analysts have inherited the reference-case model as a settled methodological framework while remaining largely unaware of the measurement contradictions embedded within its structure. Debate has remained confined within the architecture itself, comparators, assumptions, utilities, discounting and sensitivity analyses, while the prior question disappeared entirely: does the arithmetic possess recognized quantitative meaning?

Only now, through systematic interrogation of HTA knowledge bases using large language models, has the scale and consistency of this failure become fully visible. What emerges is not isolated methodological weakness but a pervasive pattern of measurement inversion and arithmetic chaos embedded at the core of the HTA reference-case framework itself. It is not clear what HTA students are being trained for.

As Table 1 indicates and the remaining 65 individual interrogations, there is no understanding of the requirement that measurement precedes arithmetic. The result is arithmetic chaos in the process of analysis. The end result, claims based on the reference case are just numbers. They have no meaning in terms of the standard for representational measurement .

The result is institutional reproduction of closure. Analytical sophistication increases while the standards necessary to interrogate the quantitative legitimacy of the framework remain absent. What is reproduced is not merely a methodological preference but a professional culture in which measurement inversion and arithmetic chaos are normalized as standard HTA practice.

CLOSURE AND TRANSITION

Recognition of closure immediately raises a practical question: what standards are required if Australian HTA is to transition from a framework defined by measurement inversion and arithmetic chaos toward one consistent with quantitative science? Criticism alone is insufficient. Once arithmetic detached itself from lawful measurement, closure became inevitable. The issue now is reconstruction. The following standards do not represent refinements to the existing reference-case architecture. They define the conditions necessary for replacement.

1. Measurement must precede arithmetic
This principle is the point from which all reconstruction begins. For decades HTA largely assumed that attaching numbers to observations automatically created quantitative evidence. Yet arithmetic operations cannot create measurement properties where they do not already exist. Multiplication, averaging and aggregation acquire meaning only after lawful measurement structures have been demonstrated. The failure to recognize this sequence created the conditions for measurement inversion itself.

2. Scale classification becomes a prerequisite for evidence generation
Not all numerical scales possess the same properties. Nominal, ordinal, interval and ratio scales support different forms of interpretation and different arithmetic operations. Historically these distinctions often disappeared within HTA practice as scores and utilities were treated as quantities without first asking what type of scale existed. Reconstruction requires explicit recognition of scale structure before claims can be advanced.

3. Ratio measures occupy a unique role in therapy evaluation
Only ratio measures support unrestricted arithmetic because they preserve meaningful quantitative relationships and possess identifiable origins. Under conventional HTA, ratio assumptions were frequently imposed upon structures lacking these properties. Future therapy claims therefore require lawful ratio measures rather than numerical approximations presented as quantities.

4. All therapy impact claims must concern attributes
Therapies do not act upon utilities, scores or simulation outputs. Therapies influence attributes within patients and populations. Hospital utilization, persistence, symptom burden and treatment satisfaction represent examples. Reconstruction therefore begins by asking a simple question largely absent from conventional HTA: what attribute is expected to change?

5. Manifest and latent attributes must be distinguished
One of the major consequences of measurement inversion was failure to distinguish between directly observed and inferred phenomena. Manifest attributes such as physician visits or hospitalization differ fundamentally from latent constructs such as symptom burden or need fulfilment.

6. Manifest attributes require linear ratio measures
Observable phenomena often possess natural quantitative structures. Hospital days, persistence behavior and switching rates provide examples where arithmetic may be meaningful. Manifest

claims therefore require explicit ratio measurement structures. Quantitative interpretation follows from the properties of the attribute itself rather than numerical conventions imposed upon observations.

7. Latent attributes require Rasch logit ratio measures

Latent constructs create a different challenge. Responses to questionnaires are not measures simply because they have numbers attached to them. Scores are observations, not quantities. Reconstruction requires recognition that latent attributes demand Rasch procedures capable of creating lawful logit ratio structures and evaluating possession. The outcome of interest becomes possession of the latent attribute rather than change in arbitrary scores.

8. Attribute claims require explicit justification

Under conventional HTA, claims frequently emerged as outputs from analytical systems rather than propositions requiring prior explanation. Utilities generated QALYs and models generated cost-effectiveness estimates, yet the rationale for selecting particular outcomes often remained unclear. Reconstruction requires a different approach. Each attribute claim should provide an explicit justification explaining why therapy is expected to influence the attribute, what evidence supports this expectation and why the anticipated change possesses clinical relevance. Claims should therefore emerge from a coherent explanatory framework rather than analytical convenience or historical precedent.

9. Attribute claims require protocols

Protocols are the mechanism through which claims acquire scientific standing. Under reference-case systems assumptions frequently disappeared within simulation structures and remained insulated from direct challenge. Reconstruction requires every attribute claim to be linked prospectively to a protocol specifying the target population, observation period, expected treatment effects and evaluation procedures. Protocols therefore become scientific instruments rather than administrative documents. They establish the framework through which claims move from expectation to evidence.

10. Attribute claims must be empirically evaluable

Claims cannot remain hypothetical projections extending decades into uncertain futures. A claim acquires scientific meaning only when evidence capable of supporting or challenging it can be generated. Reconstruction therefore requires attribute claims to specify measurable endpoints, defined populations and observation schedules capable of prospective assessment. Evaluation becomes a continuing process rather than a one-time analytical exercise completed at product launch.

11. Falsification becomes a requirement

Scientific progress requires more than confirmation; it requires vulnerability to challenge. Attribute claims should therefore specify the conditions under which they may fail. If anticipated effects are absent or treatment outcomes diverge substantially from expectations, claims should be modified or rejected. Exposure to challenge replaces protection of assumptions. Claims incapable of failure cease to function as scientific propositions.

12. HTA returns to the standards of normal science

Taken together these standards do not create a new ideology. They restore principles long familiar within quantitative science: measurable attributes, lawful measurement structures, explicit claims and continuing empirical challenge. Numerical storytelling gives way to measurable propositions. The objective is straightforward: reconnect HTA with the continuing evolution of objective knowledge. HTA therefore re-enters a framework where evidence grows through replication, criticism and empirical learning rather than through increasingly elaborate assumptions.

CONCLUSION

For more than four decades health technology assessment evolved within a framework increasingly organized around utilities, QALYs, reference-case methodologies and assumption-driven simulation structures. During this period analytical sophistication expanded steadily. Models became larger, assumptions more detailed and sensitivity analyses increasingly complex. Yet a more fundamental question remained largely absent: did the attributes supporting these numerical operations satisfy the requirements necessary for lawful quantitative claims?

The findings presented here suggest a troubling answer. Interrogation reveals not isolated methodological weaknesses but a stable endorsement structure consistent with measurement inversion. Foundational propositions concerning representational measurement receive weak support while propositions known to be false repeatedly receive strong endorsement. More importantly, these findings mirror more than 230 international interrogations demonstrating remarkable stability across agencies, journals, academic institutions and professional organizations.

The implications are difficult to avoid. Measurement inversion does not simply introduce analytical error. It changes the status of the framework itself. Arithmetic chaos is inevitable. Once arithmetic routinely precedes measurement, quantitative claims become detached from measurable attributes and the conditions necessary for scientific correction disappear. Reference-case systems increasingly evolve into closed analytical structures where assumptions generate evidence and internal refinement substitutes for empirical challenge.

Closure means closure. It does not mean agencies disappear, journals cease publication or simulation models are abandoned. Closure has a more precise meaning. It refers to a recognized lack of scientific legitimacy. A framework becomes closed when it no longer possesses the capacity for correction from within its own assumptions. The evidence presented here suggests that HTA has reached precisely that point.

One final implication should therefore be stated directly. The findings do not support a program of internal reform. The reference-case structure of HTA is beyond redemption. This is not because individual assumptions are weak, particular models incomplete or analytical techniques insufficiently sophisticated. The problem is more fundamental. Measurement inversion detached arithmetic from lawful measurement and, in doing so, removed the conditions necessary for scientific correction from within the framework itself. Larger simulations, revised utilities and increasingly elaborate assumptions merely reproduce the same structural errors and arithmetic chaos.

The choice is therefore increasingly clear. Continue to defend a closed framework organized around numerical storytelling or transition toward measurable attributes, lawful ratio measures, explicit protocols and falsifiable claims. The issue is no longer whether measurement inversion exists. That question has been answered. The issue is whether Australia chooses to remain an exemplar of the *status quo* or become an exemplar of transition.

The conclusion is unavoidable: the reference-case structure of HTA is beyond redemption. Measurement inversion guaranteed this outcome decades ago. Once arithmetic routinely preceded measurement, closure became inevitable. Subsequent methodological sophistication merely concealed the problem while reproducing it. The future therefore does not lie in reforming the reference case. It lies in abandoning it.

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

Bond T, Zi Yan, Heene M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences (4th Ed.) New York Routledge, 2026