MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

CANADA: KNOWLEDGE BASE CONSENSUS ON MEASUREMENT INVERSION IN HEALTH TECHNOLOGY ASSESSMENT FOR SIX CANADIAN RESEARCH CENTERS

Paul C Langley Ph.D Adjunct Professor, College of Pharmacy, University of Minnesota, Minneapolis, MN

LOGIT WORKING PAPER No 850 APRIL 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

This study presents a canonical logit interrogation of six Canadian academic health technology assessment (HTA) research centers, examining the extent to which their knowledge bases recognize and apply the axioms of representational measurement. Using a standardized 24-item diagnostic instrument, each knowledge base is assessed in terms of its endorsement of propositions that are either consistent with, or violate, the requirements for valid measurement. Responses are expressed as categorical probabilities and transformed into normalized logits, allowing comparison across institutions.

The results are consistent, stable, and highly structured. Across all six centers, statements that are true under the axioms of representational measurement—such as the requirement for unidimensionality, the necessity of ratio scales for multiplication, and the precedence of measurement over arithmetic—are weakly endorsed, with probabilities concentrated in the lower range and corresponding negative logits. In contrast, statements that are false but central to HTA practice—such as the classification of QALYs as ratio measures, the assumption of dimensional homogeneity, and the aggregation of utility-based constructs—are strongly endorsed, with probabilities in the upper range and positive logits.

The narrow range of probability–logit assignments across centers indicates a high degree of conceptual uniformity. There are no outliers or countervailing positions that suggest internal recognition of measurement constraints. The pattern observed is not one of isolated inconsistency, but of systematic inversion.

These findings support the conclusion that the Canadian HTA academic knowledge base operates within a framework in which numerical constructs are treated as measures in the absence of the conditions required for measurement. The implications are immediate: quantitative claims derived from such constructs cannot be considered evaluable, replicable, or falsifiable in the scientific sense. At the same time, the results point to a clear alternative based on manifest ratio measures and Rasch logit ratio scales, consistent with the axioms of representational measurement.

INTRODUCTION

Over the past several decades, health technology assessment (HTA) has evolved into a central component of healthcare decision-making in publicly funded systems. Academic research centers, policy-linked institutes, and data-driven evaluation units have collectively contributed to a substantial body of analytical work designed to inform questions of resource allocation, comparative effectiveness, and value for money. In Canada, this activity is distributed across a network of research domains that include formal HTA centers, health economics groups, and evidence-generating organizations. These institutions differ in structure, mandate, and methodological emphasis, ranging from decision-analytic modeling and cost-effectiveness analysis to real-world evidence and outcomes research. Despite this diversity, they share a common objective: to produce quantitative claims that support healthcare decision-making. The six centers are:

- Toronto Health Economics and Technology Assessment Collaborative (THETA)
- Institute for Clinical Evaluative Sciences (ICES)
- Institute of Health Economics (IHE)
- University of Alberta Health Technology and Policy Unit (HTPU)
- Department of Community Health, Dalhousie University
- Centre for Clinical Epidemiology and Evaluation (C2E2), University of British Columbia

At the core of HTA is the use of numerical constructs, costs, outcomes, utilities, and composite measures such as quality-adjusted life years (QALYs) which are combined, aggregated, and compared through arithmetic operations. These operations presuppose that the quantities involved meet the requirements of measurement as defined in representational measurement theory. Specifically, valid arithmetic requires that attributes be measured on scales with appropriate properties, including unidimensionality, dimensional homogeneity, and, where multiplication or ratio construction is involved, the presence of a true zero. These conditions are not optional; they are necessary for the interpretation of numerical results as measures rather than as ordered labels or constructed indices.

The present study builds on a series of interrogations of Canadian HTA-related knowledge bases using a standardized 24-item canonical diagnostic grounded in these measurement axioms. Each knowledge base is treated as a defined domain, encompassing its published outputs, methodological frameworks, and underlying assumptions. The interrogation process applies a fixed set of statements where each of which is either true or false under the axioms of representational measurement and evaluates the extent to which these statements are endorsed within the knowledge base. Endorsements are expressed as categorical probabilities and transformed into normalized logits, allowing for comparison both within and across domains.

The purpose of this paper is to move beyond individual interrogations and to examine whether a consistent pattern emerges across multiple Canadian research centers. Six distinct knowledge bases, representing academic, policy-linked, and evidence-generating domains, are included. These domains differ in institutional form, geographical location, and methodological orientation, providing an appropriate test of whether recognition of measurement principles varies across the HTA landscape. If these knowledge bases exhibit divergent profiles, this would suggest that measurement awareness is context-dependent. If, however, the results converge, this would indicate the presence of a shared structure of false measurement.

The findings presented in this paper demonstrate that such convergence exists. Across all six knowledge bases, there is a consistent pattern in which statements grounded in the axioms of measurement are weakly endorsed, while statements that conflict with these axioms but are embedded within HTA practice are strongly supported. This pattern is not confined to a particular type of institution or analytical approach. It is observed across academic centers, policy-linked units, and data-driven research domains. The implication is that the Canadian HTA research community exhibits not a diversity of measurement perspectives, but a consensus. The nature of that consensus, and its implications for the validity of HTA claims, are the focus of the analysis that follows.

FROM INTERROGATION TO CONSENSUS

This paper marks a transition from a series of individual interrogations of Canadian health technology assessment (HTA) research centers to a system-level evaluation of their collective structure. Previous analyses have applied a standardized 24-item canonical diagnostic to a range of Canadian HTA-related knowledge bases, each treated as a defined domain encompassing its published outputs, methodological frameworks, and implicit assumptions. These interrogations were conducted independently, with no prior assumption that results would converge. The present paper brings these results together to address a more fundamental question: do Canadian HTA research centers exhibit variation in their recognition of measurement principles, or do they demonstrate a consistent and reproducible pattern?

The importance of this question lies in the role of measurement within HTA. The construction and interpretation of claims regarding costs, outcomes, and value depend on the properties of the quantities involved. Arithmetic operations including multiplication, aggregation, and ratio construction are only valid if the underlying attributes are measured on scales that satisfy the axioms of representational measurement. These include unidimensionality, dimensional homogeneity, and, for ratio operations, the existence of a true zero. Without these conditions, numerical outputs cannot be interpreted as measures, regardless of their apparent precision or methodological sophistication.

Each interrogation in this series has employed an identical framework. A fixed set of 24 canonical statements, each of which is either true or false under the axioms of measurement, is applied to the knowledge base. The knowledge base is instructed to evaluate each statement, and its response is expressed as a categorical probability reflecting the strength of endorsement. These probabilities are then transformed into normalized logits within a fixed range (±2.50), providing a continuous metric that supports comparison across statements and across domains. The use of fixed probability categories and a consistent transformation ensures that differences in results reflect differences in the knowledge base, not variation in method.

The move from individual interrogation to synthesis introduces a different objective. The purpose is no longer to characterize a single knowledge base, but to determine whether the results across multiple domains converge to a common profile. Six Canadian HTA-related knowledge bases are included, representing a range of institutional forms: academic research centers, policy-linked institutes, and data-driven evidence units. These domains differ in geography, function, and methodological emphasis, providing an appropriate test of whether measurement recognition is context-dependent or structurally determined.

The central concept introduced in this paper is that of consensus. This is not consensus in the conventional sense of agreement among researchers or institutions. Rather, it refers to the convergence of knowledge base structure when interrogated under a common framework. If each domain, despite its differences, produces a similar pattern of probabilities and logits across the canonical statements, then the result is a consensus profile. Such a profile would indicate that the underlying assumptions governing HTA practice are shared across institutions, independent of their specific roles or analytical approaches.

The results presented in the following sections demonstrate that this is indeed the case. Across all six Canadian knowledge bases, there is a consistent pattern in which statements that are true within the axioms of measurement are weakly endorsed, while statements that are false but embedded within HTA practice are strongly supported. The convergence of these results, with minimal variation across domains, indicates that the Canadian HTA research landscape is characterized by a stable and reproducible structure. The nature of this structure, and its implications for the interpretation of HTA claims, form the basis of the analysis that follows.

THE CANADIAN HTA RESEARCH LANDSCAPE

The Canadian health technology assessment (HTA) research landscape is characterized by a set of institutional domains that differ in structure, mandate, and methodological orientation, yet collectively contribute to the production of evidence used in healthcare decision-making. These domains include academic HTA centers, policy-linked research institutes, and data-driven evidence units. While there is no single centralized academic HTA authority in Canada, a relatively small number of research centers account for the majority of HTA-related output, methodological development, and policy engagement.

The six knowledge bases included in this analysis reflect this diversity of institutional forms. They comprise academic research centers such as the University of Toronto Health Economics and Technology Assessment Collaborative and the Health Technology and Policy Unit at the University of Alberta; policy-oriented and independent research organizations such as the Institute of Health Economics; data and evidence platforms such as the Institute for Clinical Evaluative Sciences and the Centre for Clinical Epidemiology and Evaluation; and regionally embedded academic domains such as the Dalhousie University Department of Community Health and Epidemiology. Together, these institutions span the principal modes through which HTA is conducted in Canada, including economic evaluation, decision modeling, real-world evidence generation, and health services research.

Each of these knowledge bases has been interrogated independently using the standardized 24-item canonical diagnostic. The full results of these interrogations, including detailed probability and logit tables for each center, are available on the Maimon Research website. The present paper does not reproduce these detailed outputs. Instead, it draws on them to construct a synthesis profile that captures the common structure across domains. This approach allows the analysis to move from institutional description to structural evaluation.

The inclusion of multiple and diverse knowledge bases serves a specific purpose. If recognition of measurement principles were dependent on institutional role, methodological focus, or geographical location, then variation would be expected across these domains. Academic centers engaged in economic modeling might differ from data-driven evidence units, and policy-linked organizations might exhibit distinct patterns compared to university-based research groups. By selecting knowledge bases that differ along these dimensions, the analysis provides an appropriate test of whether such variation exists.

The results, however, point in the opposite direction. Despite differences in institutional form and analytical emphasis, the six knowledge bases exhibit a high degree of similarity in their response profiles. This lack of variation is central to the argument that follows. It indicates that the underlying assumptions governing HTA are not localized to particular institutions, but are shared across the Canadian research landscape. The diversity of institutional settings therefore strengthens, rather than weakens, the case for a consensus structure.

In this context, the Canadian HTA research landscape can be understood not as a collection of independent domains with distinct methodological perspectives, but as a set of interconnected knowledge bases that operate within a common analytical framework. The consistency observed across these domains provides the empirical basis for the synthesis presented in the following sections.

INTERROGATION FRAMEWORK AND ANALYTICAL METHOD

The analysis presented in this paper is based on a standardized interrogation framework designed to evaluate whether a knowledge base recognizes the conditions required for valid measurement. The framework is anchored in the axioms of representational measurement and applies a fixed diagnostic instrument comprising 24 canonical statements. Each statement is constructed so that it is either unequivocally true or false under these axioms. The purpose is not to elicit opinion, but to determine whether the knowledge base, as represented in its corpus of outputs and methodological conventions, reinforces or rejects propositions that are fundamental to measurement.

Each knowledge base is treated as a defined domain. This domain includes its published research, methodological practices, teaching materials, and the assumptions that can be inferred from these sources. The interrogation is conducted by presenting the canonical statements sequentially, with the knowledge base instructed to evaluate each statement in terms of whether it is supported or contradicted. Importantly, the interrogation is constrained: the knowledge base is not invited to reinterpret the statements, but to assess them within the context of its own established practices and conventions. This ensures that responses reflect the structure of the knowledge base rather than the flexibility of interpretation.

Responses to each statement are expressed as categorical probabilities drawn from a fixed set. These categories are deliberately coarse 0.05, 0.10, 0.20, 0.30, 0.75, 0.80, 0.85, 0.90, 0.95 to avoid spurious precision while allowing meaningful differentiation in the strength of endorsement. The assigned probability represents the extent to which the knowledge base reinforces the proposition embodied in the statement. Low probabilities indicate weak endorsement or rejection, while high probabilities indicate strong endorsement.

These probabilities are subsequently transformed into normalized logits using the standard log-odds transformation. The transformation maps bounded probability values onto a continuous scale, which is then normalized to a fixed range (±2.50) using established rounding conventions. The resulting logit values provide a metric that allows for systematic comparison both within and across interrogations. Equal differences in logits correspond to equal differences in the strength of endorsement, enabling the identification of patterns that are not apparent from ordinal probability categories alone.

Within a given interrogation, the set of logits across the 24 statements constitutes a profile of the knowledge base. Differences in logit values indicate the relative strength with which individual propositions are supported or rejected, allowing the identification of internal structure. Across interrogations, the use of a fixed probability set and consistent logit transformation ensures comparability. Because each knowledge base is evaluated using the same instrument and scaling, differences in profiles can be attributed to differences in the knowledge base rather than variation in method.

The objective of this framework is not to produce a summary score or ranking. Instead, it is to identify structural alignment with, or divergence from, the axioms of measurement. The pattern of probabilities and logits across the canonical statements provides a reproducible representation of how a knowledge base treats the conditions required for valid measurement. When applied across multiple domains, the framework allows for the identification of common structures, including the presence of consistent patterns that transcend institutional differences.

RESULTS – A CONSENSUS PATTERN OF MEASUREMENT INVERSION

The results of the six Canadian HTA knowledge base interrogations are summarized in Table 1. For each of the 24 canonical statements, the table identifies whether the statement is true or false under the axioms of representational measurement, together with the observed range of categorical probabilities and the corresponding range of normalized logits across the six domains. The purpose of this presentation is to consolidate the individual interrogations into a single profile and to make explicit both the level and the consistency of endorsement for each statement. The use of ranges allows variation across knowledge bases to be assessed directly.

TABLE 1: SUMMARY OF STATEMENT RESPONSES: SIX CANADIAN RESEARCH CENTRE KNOWLEDGE BASE DOMAINS

STATEMENT	RESPONSE 1=TRUE 0=FALSE	ENDORSEMENT OF RESPONSE CATEGORICAL PROBABILITY RANGE	NORMALIZED LOGIT (IN RANGE +/- 2.50) RANGE
INTERVAL MEASURES LACK A TRUE ZERO	1	0.15 to 0.20	-1.75 to -1.40
MEASURES MUST BE UNIDIMENSIONAL	1	0.10	-2.20
MULTIPLICATION REQUIRES A RATIO MEASURE	1	0.10	-2.20
TIME TRADE-OFF PREFERENCES ARE UNIDIMENSIONAL	0	0.80 to 0.90	+1.40 to +2.20
RATIO MEASURES CAN HAVE NEGATIVE VALUES	0	0.90	+2.20

EQ-5D-3L PREFERENCE ALGORITHMS CREATE INTERVAL MEASURES	0	0.80 to 0.90	+1.40 to +2.20
THE QALY IS A RATIO MEASURE	0	0.85 to 0.95	+1.75 to +2.50
TIME IS A RATIO MEASURE	1	0.90	+2.20
MEASUREMENT PRECEDES ARITHMETIC	1	0.10	-2.20
SUMMATIONS OF SUBJECTIVE INSTRUMENT RESPONSES ARE RATIO MEASURES	0	0.80 to 0.90	+1.40 to +2.20
MEETING THE AXIOMS OF REPRESENTATIONAL MEASUREMENT IS REQUIRED FOR ARITHMETIC	1	0.10	-2.20
THERE ARE ONLY TWO CLASSES OF MEASUREMENT LINEAR RATIO AND RASCH LOGIT RATIO	1	0.05	-2.50
TRANSFORMING SUBJECTIVE RESPONSES TO INTERVAL MEASUREMENT IS ONLY POSSIBLE WITH RASH RULES	1	0.05	-2.50
SUMMATION OF LIKERT QUESTION SCORES CREATES A RATIO MEASURE	0	0.80 to 0.90	+1.40 to +2.20
THE QALY IS A DIMENSIONALLY HOMOGENEOUS MEASURE	0	0.85 to 0.95	+1.75 to +2.50
CLAIMS FOR COST-EFFECTIVENESS FAIL THE AXIOMS OF REPRESENTATIONAL MEASUREMENT	1	0.10 to 0.15	-2-20 to -1.75
QALYS CAN BE AGGREGATED	0	0.85 to 0.95	+1.75 to +2.50
NON-FALSIFIABLE CLAIMS SHOULD BE REJECTED	1	0.15 to 0.30	-1.75 to -0.85
REFERENCE CASE SIMULATIONS GENERATE FALSIFIABLE CLAIMS	0	0.80 to 0.90	+1.40 to +2.50
THE LOGIT IS THE NATURAL LOGARITHM OF THE ODDS-RATIO	1	0.30 to 0.35	-0.85 to -0.65
THE RASCH LOGIT RATIO SCALE IS THE ONLY BASIS FOR ASSESSING THERAPY IMPACT FOR LATENT TRAITS	1	0.05	-2.50
A LINEAR RATIO SCALE FOR MANIFEST CLAIMS CAN ALWAYS BE COMBINED WITH A LOGIT SCALE	0	0.75 to 0.80	+1.15 to +1.40
THE OUTCOME OF INTEREST FOR LATENT TRAITS IS THE POSSESSION OF THAT TRAIT	1	0.05	-2.50
THE RASCH RULES FOR MEASUREMENT ARE IDENTICAL	1	0.05	-2.50

TO THE AXIOMS OF REPRESENTATIONAL MEASUREMENT			

Note: Where no range is reported all six knowledge bases are allocated to the same categorical endorsement probability.

The synthesis reveals a striking and highly structured pattern. Across all statements, the ranges of both probabilities and logits are narrow, with minimal variation between knowledge bases. This consistency is observed despite differences in institutional form, methodological emphasis, and geographical location. The results do not display dispersion; they display convergence. The profile that emerges is not one of heterogeneity, but of consensus.

The first and most important observation concerns statements that are true under the axioms of representational measurement. These include propositions such as the requirement for unidimensionality, the necessity of dimensional homogeneity, the restriction of multiplication to ratio scales, and the principle that measurement must precede arithmetic. Across all six knowledge bases, these statements are associated with low endorsement probabilities, typically in the range of 0.05 to 0.15, with occasional values of 0.20. When transformed into normalized logits, this corresponds to a tightly bounded range from approximately -2.50 to -1.75. For several core statements, the values are fixed at the lower bound of the scale (-2.50), indicating complete non-possession of the principle within the knowledge base.

There is no evidence of variation in this respect. No knowledge base strongly endorses any foundational measurement axiom. Academic HTA centers, data-driven evidence platforms, and policy-linked research units all exhibit the same response structure. The implication is that the axioms representational measurement required for valid measurement are not recognized as part of the analytical framework in any of the interrogated domains.

In direct contrast, statements that are false under the axioms of measurement but are embedded within HTA practice are strongly endorsed. These include propositions such as the treatment of QALYs as ratio measures, the aggregation of QALYs, the interpretation of preference-based utility scores as interval or ratio quantities, and the use of composite measures derived from ordinal inputs. Across all six knowledge bases, these statements are associated with high endorsement probabilities, typically in the range of 0.80 to 0.95. The corresponding normalized logits fall between approximately +1.40 and +2.50, with many statements clustering at the upper end of this range.

Again, there is no divergence. No knowledge base rejects these propositions, and none exhibits weak endorsement. While the exact probability assigned to a statement may vary slightly, the direction and magnitude of endorsement are consistent. These constructs are not marginal; they are central and reinforced across all domains.

The relationship between these two sets of results defines the central finding of this analysis. The pattern is not one of isolated error or inconsistency. It is symmetrical. Statements that are true are systematically weakly endorsed, while statements that are false are systematically strongly endorsed. When expressed in logit terms, this produces a clear separation: true statements occupy a negative range (-2.50 to -1.75), while false statements occupy a positive range (+1.40 to +2.50). There is no overlap between these ranges. The resulting profile is a structured inversion of measurement principles.

The strength of this conclusion is reinforced by the absence of outliers. Across six independent interrogations, there is no instance of a knowledge base deviating from the overall pattern. For several key statements, identical probability and logit values are observed across all domains. For example, statements concerning the necessity of Rasch-based measurement, the classification of measurement scales, and the possession of latent traits are uniformly assigned probabilities of 0.05, corresponding to logits of -2.50. Conversely, core HTA constructs such as the QALY as a ratio measure and the aggregation of QALYs consistently receive probabilities of 0.95, with logits of +2.50. These are repeated values, not approximations.

A small number of statements show slightly greater dispersion. The statement concerning the rejection of non-falsifiable claims exhibits probabilities ranging from 0.15 to 0.30, with corresponding logits from -1.75 to -0.85. The definition of the logit as the natural logarithm of the odds ratio shows a similar range. However, these variations do not alter the overall pattern. In each case, the logits remain within the same directional range, and no statement crosses from negative to positive or vice versa.

Taken together, these results provide clear evidence of a consensus profile across Canadian HTA research centers. This consensus is not explicitly articulated within the field; it is revealed through the structure of responses to the canonical diagnostics with the large language model for each. The consistency of probability and logit ranges across all domains demonstrates that the same set of assumptions governs the treatment of measurement in HTA, independent of institutional context.

The implications are straightforward. The Canadian HTA knowledge bases operate as if measurement exists where it does not, and ignores the conditions required for measurement where they are essential. This is not a technical issue. It is a reversal of the basic relationship between measurement and arithmetic.

To make this explicit: multiplication requires a ratio scale. Across all six knowledge bases, the statement that multiplication requires a ratio scale is assigned a probability of 0.10 and a logit of approximately -2.20. At the same time, the multiplication of utility values by time, central to the construction of QALYs is strongly endorsed. The requirement is rejected; the operation is accepted. The contradiction is not resolved; it is institutionalized.

The same pattern holds for dimensional homogeneity. The statement that the QALY is a dimensionally homogeneous measure is strongly endorsed, with probabilities up to 0.95. Yet the conditions required to establish dimensional homogeneity are not recognized. Again, the conclusion is not derived from measurement; it is assumed.

What makes this finding decisive is not any single statement, but the consistency across all statements and all knowledge bases. There are no counterexamples. There is no domain in which measurement axioms are strongly endorsed. There is no domain in which core HTA constructs are rejected. The ranges do not vary in direction. They vary only slightly in magnitude, and always within the same side of the scale.

This removes any plausible alternative explanation. The pattern cannot be attributed to differences in institutional focus, methodological preference, or data sources. Academic centers, policy-linked institutes, and data-driven research platforms all produce the same profile. The conclusion is that the structure is shared; all six concur with the place of measurement inversion in HTA claims.

The consequence is that HTA, as currently practiced, does not operate within a framework of measurement. It operates within a framework of numerical construction. Numbers are generated, combined, and compared, but the properties that would make those numbers measures are not established. The appearance of quantitative analysis is preserved, but its foundation is absent. This is best described as numerical storytelling.

For a non-technical reader, the conclusion can be stated simply. If you ask whether these knowledge bases recognize the rules required to measure something, the answer is no. If you ask whether they nevertheless use numbers as if those rules had been satisfied, the answer is yes. The data in Table 1 show both results clearly.

CONCLUSION – WHERE DO WE GO FROM HERE?

The results presented in this paper leave little room for qualification. Across six Canadian HTA knowledge bases, representing the principal institutional forms through which health technology assessment is conducted, there is a consistent and reproducible pattern: the axioms required for valid measurement are not recognized, while numerical practices that reject these axioms are strongly endorsed. This is not a marginal defect. It is a foundational failure. The conclusion is unavoidable: HTA, as currently constructed by these six centers rests on a framework of false measurement. These results are identical to a similar assessment for research centers in Australia and the United States.

If this is accepted, then the implications are not incremental. They are decisive. It is not sufficient to refine existing models, adjust parameter inputs, or improve reporting standards. The problem lies deeper. The constructs that define HTA such as time trade-off preferences, multiattribute instruments, utilities, QALYs, the reference case, and cost-effectiveness ratios are not salvageable through modification. They fail because they violate the conditions required for measurement. They must be abandoned.

Start with time trade-off preferences and multiattribute instruments. These are presented as methods for generating quantitative representations of health-related quality of life. In practice, they are procedures for assigning numerical labels to ordinal judgments. There is no evidence that the resulting values satisfy the requirements of interval or ratio measurement. They lack a true zero, they are not unidimensional in any defensible sense, and they cannot support multiplication or meaningful aggregation. Yet these values are treated as if they were measures. This is not a technical oversight. It is a category error that has been institutionalized.

Utilities, derived from these instruments, inherit the same defects. They are treated as quantities that can be multiplied by time, summed across individuals, and compared across interventions. None of these operations is justified. The numerical outputs retain their form but lose their meaning. They are not measures of anything that satisfies the axioms of representational measurement. They are constructed indices that have been granted a status they do not possess.

The QALY, built from these utilities, is the central false artifact of this system. It is presented as a ratio measure of health gain, capable of supporting arithmetic operations and policy decisions. The interrogation results show that this claim is strongly endorsed across all knowledge bases. At the same time, the conditions required to support it are absent. The QALY is not dimensionally homogeneous, it does not rest on a valid scale, and it cannot be interpreted as a measure in any scientific sense. It is, at best, a numerical convenience with no known measurement properties. At worst, it is a vehicle for the systematic misrepresentation of value.

The reference case, which codifies these practices, compounds the problem. It standardizes the use of QALYs, utilities, and cost-effectiveness ratios, embedding false measurement within formal guidelines. What is presented as methodological rigor is, in fact, the enforcement of invalid constructs. The reference case does not resolve the problem of measurement; it institutionalizes its absence.

Cost-effectiveness claims, derived from these constructs, are the final step. Ratios of costs to QALYs are presented as meaningful quantities that can guide decisions about resource allocation. Yet both the numerator and the denominator fail to meet the requirements of ratio measurement. Costs, when expressed as aggregates of heterogeneous resource inputs, violate dimensional homogeneity. QALYs, as already established, are not measures. The resulting ratio is therefore not a measure of anything. It is a number produced by arithmetic operations applied to quantities that do not support those operations.

The cumulative effect is a system of analysis that produces numerical outputs without measurement. Over the past forty years, this system has generated an extensive literature, trained generations of analysts, and informed policy decisions across multiple jurisdictions. It has done so without establishing the conditions required for the claims it makes. This is not a case of slow progress or partial success. It is a legacy of false measurement.

The absence of measurement has consequences beyond technical validity. It represents a failure of duty of care. Health technology assessment is not an abstract exercise. Its outputs are used to make decisions that affect patient access to therapies, the allocation of resources, and the prioritization of interventions. When these decisions are justified on the basis of numerical claims that do not meet the standards of measurement, the responsibility for those decisions is compromised. The appearance of quantitative rigor cannot substitute for the existence of measurable evidence.

Equally, there has been no evolution of objective knowledge. In a scientific framework, claims are subject to falsification, refinement, and replacement. Measurement provides the basis for this process, allowing hypotheses to be tested against observed reality. In the absence of measurement, this process cannot occur. The constructs of HTA are not tested; they are repeated. Models are elaborated, not challenged. The framework reproduces itself without engaging with the conditions that would allow it to evolve.

Where, then, do we go from here? The answer is both simple and demanding. HTA must be reconstructed on a foundation of valid measurement. This requires a clear distinction between two classes of attributes. For manifest attributes, those that can be directly observed, measurement must be expressed in linear ratio terms. Resource use should be counted in natural units: hospital days, clinic visits, doses administered, hours of care. These are quantities that support arithmetic operations.

For latent constructs, those that cannot be directly observed, measurement must be established through models that satisfy the axioms of representational measurement. The Rasch model provides such a framework. It allows the transformation of ordinal responses into interval measures with defined properties, supporting meaningful comparison and change. Without the unique Rasch transformation, latent constructs remain unmeasured and cannot support quantitative claims.

This reconstruction requires the abandonment of existing constructs that fail these standards. Time trade-off preferences, multiattribute instruments, utilities, QALYs, the reference case, and cost-effectiveness ratios must be set aside. They cannot be repaired. Their continued use perpetuates a system in which numerical outputs are mistaken for measurement.

The alternative is not the absence of evaluation, but its redefinition. HTA can and should produce claims that are credible, evaluable, and replicable. These claims must be grounded in measures that satisfy the axioms of representational measurement and must be specified within protocols that allow for empirical testing. Only then can HTA move from numerical storytelling to scientific inquiry. The results presented here require that the past forty years of HTA be understood in a specific and limited sense. This is not a cumulative scientific tradition that can be refined or extended. It is a body of work constructed on numerical operations that were never grounded in valid measurement. Over this period, a set of practices including time trade-off preferences, multiattribute utility instruments, QALYs, and cost-effectiveness ratios that became institutionalized as if they satisfied the requirements of representational measurement. They do not. The interrogation results demonstrate that the conditions required for measurement were neither recognized nor applied.

This does not imply that the work lacks internal coherence or policy influence. On the contrary, the framework has been highly successful in establishing itself as the dominant paradigm for evaluating healthcare interventions. Its influence is precisely the problem. Numerical outputs have been interpreted as measures, and decisions have been justified on that basis, despite the absence of measurable attributes. The appearance of quantitative rigor has substituted for the existence of measurement.

The appropriate response is not to reinterpret or rehabilitate these constructs, but to recognize their limits. The legacy of HTA in Canada is not a foundation to be repaired, but a boundary to be acknowledged. Its central claims cannot be carried forward as evidence. The obligation is not to explain the past, but to prevent its continuation by replacing it with a framework that meets the requirements of measurement and supports the evolution of objective knowledge.

The findings of this paper make clear that such a transition has not yet occurred. The consensus identified across Canadian research centers is not a consensus of scientific validity, but of shared departure from the requirements of measurement. Recognizing this is the first step. Acting on it is the task that follows.

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.

WEBSITE REFERENCES

CANADA: THE ABSENCE OF REPRESENTATIONAL MEASUREMENT AND THE TORONTO HEALTH ECONOMICS AND TECHNOLOGY ASSESSMENT COLLABORATIVE (THETA)

CANADA: THE ABSENCE OF REPRESENTATIONAL MEASUREMENT AND THE INSTITUTE FOR CLINICAL EVALUATIVE SCIENCES (ICES)

CANADA: THE ABSENCE OF REPRESENTATIONAL MEASUREMENT AND THE INSTITUTE OF HEALTH ECONOMICS (IHE)

CANADA: THE ABSENCE OF REPRESENTATIONAL MEASUREMENT AND THE UNIVERSITY OF ALBERTA HEALTH TECHNOLOGY AND POLICY UNIT (HTPU)

CANADA: THE ABSENCE OF REPRESENTATIONAL MEASUREMENT AND THE DALHOUSIE UNIVERSITY DEPARTMENT OF COMMUNITY HEALTH

CANADA: THE ABSENCE OF REPRESENTATIONAL MEASUREMENT AND THE CENTRE FOR CLINICAL EPIDEMIOLOGY AND EVALUATION (C2E2) UNIVERSITY OF BRITISH COLUMBIA