MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

CANADA: THE ABSENCE OF RATIO MEASUREMENT AND THE CLOSURE OF HEALTH TECHNOLOGY ASSESSMENT

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

LOGIT WORKING PAPER No 1816 JUNE 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

This paper argues that contemporary health technology assessment (HTA) in Canada has reached a state of closure. Closure occurs when a discipline can no longer distinguish between numbers and measures and continues methodological development within an accepted framework while neglecting the foundational requirements of measurement. The central claim is that Canadian HTA has institutionalized measurement inversion, where arithmetic precedes measurement rather than measurement preceding arithmetic.

The analysis focuses on the measurement foundations of utility-based assessment, including time trade-off (TTO), standard gamble (SG), the Health Utilities Index (HUI), quality-adjusted life years (QALYs) and reference-case simulation models. Although these approaches differ in methodology, they share a common assumption: that numerical values derived from health-state descriptions can be treated as lawful quantitative measures. This paper argues that no evidence is provided that utility scores satisfy the requirements of ratio measurement. In particular, there is no demonstration of unidimensionality, meaningful zero, invariance, dimensional homogeneity, or compliance with the axioms of representational measurement. Consequently, utilities remain numerical constructions rather than demonstrated measures.

The paper examines the HUI as a particularly important Canadian example of measurement inversion. The HUI combines multiple health dimensions through a multiplicative utility algorithm and applies adjustment factors to generate a utility score. Yet neither the attribute-level coefficients nor the resulting utility score is shown to possess ratio-scale properties. More fundamentally, the HUI combines heterogeneous attributes without demonstrating the existence of a single underlying quantitative attribute, thereby violating the requirements of dimensional homogeneity.

Evidence for closure is drawn from large language model interrogations of Canada's Drug Agency (CDA) and five leading Canadian HTA research centers. The results reveal strong endorsement of propositions incompatible with representational measurement together with weak endorsement of propositions defining lawful measurement. A second set of interrogations demonstrates an almost complete absence of awareness of Rasch measurement, despite Rasch methodology providing the only accepted framework for transforming observations on latent attributes into lawful measures.

The paper concludes that Canadian HTA lacks a coherent measurement strategy for both manifest and latent attributes. Utilities, QALYs and reference-case simulation models are therefore symptoms rather than causes of closure. The solution is reconstruction rather than reform. HTA must be rebuilt around the only two lawful measures relevant to therapy assessment: linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes. Claims for therapy impact must be empirically evaluable, replicable and falsifiable if HTA is to recover scientific credibility.

INTRODUCTION

Closure occurs when a discipline can no longer distinguish between numbers and measures. It is reached when methodological debate continues within an accepted framework while the foundational assumptions that sustain that framework are no longer questioned. This paper argues that closure is now an unavoidable characteristic of health technology assessment (HTA) in Canada. The reason is straightforward. For more than four decades, utility-based HTA has proceeded as though the problem of measurement had already been solved. It has not.

The development of utility-based assessment has followed two principal routes. The first relies upon time trade-off (TTO) techniques to generate utility values from responses to health-state descriptions. The second relies upon standard gamble (SG) techniques to derive utility values from choices involving risk and uncertainty. In Canada, the most prominent example of the latter approach is the Health Utilities Index (HUI), where health-state descriptions are valued through SG-based procedures and transformed into utility scores [1]. Although these approaches differ in methodology, they share a common objective. Responses to health-state descriptions are converted through a scoring algorithm into numerical values, typically anchored on a scale where 0 represents death and 1 represents full health. These values are then treated as utilities.

Once created, the origin of the utility score becomes largely irrelevant. The reference-case framework can accommodate utility values generated by TTO, SG, HUI, EQ-5D, AQoL, SF-6D or any other preference-based system. Utility scores may be reported as measures of health-related quality of life, multiplied with time to generate quality-adjusted life years (QALYs), or inserted into simulation models to support claims regarding cost-effectiveness and value for money. An extensive literature has developed comparing TTO and SG valuations, examining alternative algorithms, assessing additive and multiplicative scoring systems, and proposing refinements to improve agreement between estimates. Yet from the perspective of measurement theory, these debates are largely beside the point.

The critical question is not whether TTO is superior to SG, whether additive algorithms are preferable to multiplicative algorithms, or whether one utility instrument produces higher scores than another. The critical question is whether any of these approaches create a lawful ratio measure. Unless that question is answered, the resulting numerical values remain numbers rather than measures. The apparent sophistication of the scoring algorithm cannot alter this fact.

The evaluation of therapy impacts requires ratio measurement. Claims regarding improvement, deterioration, comparative effectiveness, value for money and cost-effectiveness all depend upon the existence of quantities that satisfy the conditions required for lawful arithmetic. The existence of a ratio scale cannot be assumed. It must be demonstrated. Yet neither TTO nor SG establishes that the resulting utility scores satisfy the axioms of representational measurement. Neither demonstrates unidimensionality, invariance, dimensional homogeneity or the existence of a true zero. The resulting utility values therefore possess unknown measurement properties. They are numerical constructions rather than demonstrated quantitative measures.

This failure occurs before the first arithmetic operation is undertaken. Utilities cannot acquire ratio properties through subsequent manipulation. QALYs cannot become lawful measures through multiplication of utility scores by time. Simulation models cannot generate scientifically credible quantitative claims merely because increasingly sophisticated statistical and computational techniques are applied. Once the requirement for ratio measurement has been ignored, every subsequent stage rests upon an unsupported assumption. The result is measurement inversion: arithmetic precedes measurement rather than measurement preceding arithmetic.

The significance of this argument extends beyond criticism of particular utility instruments or modelling frameworks. The problem is structural. If utility scores generated through either TTO or SG lack demonstrated ratio properties, then every framework that relies upon those scores inherits the same limitation. The consequences are unavoidable. Utilities become assumed measures. QALYs become assumed measures. Simulation model outputs become assumed measures. Debates then focus upon model structure, discount rates, time horizons, utility algorithms and sensitivity analyses while the prior question of measurement remains unaddressed.

This is why closure becomes inevitable. Once a discipline accepts utility scores as measures without demonstrating the properties necessary for measurement, all subsequent methodological development occurs within the boundaries established by that original assumption. Researchers may disagree over valuation techniques, health-state descriptions, model specifications or statistical methods, yet they continue to share the same foundational premise. The possibility that utility scores may not constitute lawful measures is excluded from consideration. Closure therefore does not arise because the discipline reached the correct answer. Closure arises because the discipline ceased to ask the foundational question: measurement precedes arithmetic. A question which is absent in the textbooks, reports and publications of HTA [2]

The Canadian experience provides a particularly revealing example of this process. Canada witnessed the development of both SG-based utility construction through the HUI and the widespread adoption of utility-based economic evaluation through cost-utility analysis and reference-case modelling. Although these traditions evolved through different methodological pathways, both converged upon the same assumption: that preference scores derived from health-state descriptions could be treated as quantitative measures. The result is that closure emerges irrespective of the route chosen. Whether utility scores are generated through TTO or SG procedures, the same measurement problem remains unresolved.

If this interpretation is correct, then evidence of closure should be visible throughout the Canadian HTA community. It should be reflected not only in utility instruments and modelling practices but also in the knowledge structures that support HTA teaching, research and policy assessment. Institutions operating within a framework of measurement inversion would be expected to endorse propositions that assume the legitimacy of utility scores, QALYs and simulation model outputs while showing little recognition of the requirements for ratio measurement itself.

This expectation can be examined directly through interrogation of HTA knowledge bases. If closure is present, interrogations of Canada's Drug Agency, formerly CADTH, and leading Canadian HTA research centers should reveal a consistent pattern. There should be widespread endorsement of propositions that assume utilities and QALYs are valid quantitative measures together with limited recognition of representational measurement, dimensional homogeneity, admissible transformations, unidimensionality, ratio scales and the distinction between manifest and latent attributes. Most importantly, the same pattern should emerge regardless of whether the institution's intellectual heritage is rooted primarily in TTO-based utility construction, SG-based utility construction, or a combination of both. Closure predicts convergence.

This is, in fact, the case. A comprehensive interrogation of the HTA knowledge for major research centers points to a significant level of measurement inversion together with a virtual complete absence of awareness of Rasch measurement [3]. This is a critical gap as it means that, following the standards for scales of measurement and the axioms of representational measurement there is no concept of the Rasch logit ratio scale. Although Rash was developed over 60 years ago, the fact that the Rasch rules are the necessary and sufficient condition for transforming observations on latent attributes to ratio measurement is not on the agenda.

The purpose of this paper is to examine the relationship between ratio measurement, measurement inversion and closure in Canadian HTA. The central argument is that closure is not a consequence of the HUI, utilities, QALYs or simulation models. These are symptoms rather than causes. Closure occurs at the point where numerical preference scores are accepted as measures without demonstrating the properties necessary for lawful measurement. Once that step is taken, the subsequent development of utility instruments, QALYs and simulation models becomes merely the elaboration of an initial error. The interrogations of Canadian HTA agencies and research centers provide evidence that this process has become institutionalized, creating a self-reinforcing framework in which the distinction between numbers and measures has effectively disappeared.

SCIENCE DEMANDS MEASUREMENT

The scientific revolution of the 17th century established a principle that remains fundamental to all quantitative inquiry: before claims can be made about the world, the attributes of interest must be measured [4] This requirement was not formalized through modern measurement theory until much later, but it was implicit in the work of the earliest experimental scientists. Galileo did not begin with arithmetic. He began with the identification of attributes that could be observed, compared and measured. Distance, time, velocity and acceleration were treated as properties of the physical world whose measurement was a prerequisite for explanation. Arithmetic followed measurement, not the reverse.

This principle became embedded within the development of science. The history of scientific progress is inseparable from the history of instrument development. Telescopes, microscopes, clocks, balances, thermometers and spectrometers were not created to generate numbers. They were created to measure specific attributes. The objective was always the same: to establish a correspondence between an empirical attribute and a numerical representation that preserved the relevant relationships among observations. The success of science depended upon the reliability of this correspondence. Without measurement there could be no quantitative claims.

Equally important was the recognition that measurement refers to attributes. A measure is meaningful only if it refers to a specific property or characteristic. Length, mass, temperature and time are all examples of attributes. They possess identities independent of the numbers used to represent them. The same principle applies outside the physical sciences. If an attribute is to be measured, it must first be identified. Measurement is therefore inseparable from the concept of an attribute. Without an attribute there can be no measure.

Representational measurement theory formalized these insights [5]. Measurement came to be understood as the assignment of numbers to attributes according to rules that preserve empirical relationships. The emphasis remained on the attribute rather than the number. Numbers do not create measurement. They merely represent measured attributes. Consequently, arithmetic operations are only meaningful when the measurement properties of the attribute have first been established.

The importance of this principle is evident in the philosophy of science associated with Popper [6]. Scientific claims must be capable of falsification. A claim that cannot be subjected to empirical testing cannot be regarded as scientific. Yet falsification presupposes measurement. To test a claim regarding an attribute, there must first be a means of measuring that attribute. The claim itself must refer to a measurable property of the world. Measurement therefore precedes both arithmetic and falsification.

This relationship is often overlooked. Falsification is frequently discussed as though it were an abstract methodological requirement independent of measurement. In reality, the two are inseparable. Claims are evaluated through observations of attributes. If the attribute has not been measured, the claim cannot be tested. The requirement for measurement is therefore embedded within the very concept of scientific inquiry.

The implications for health technology assessment are immediate. If HTA is to function as a scientific discipline, it must begin by identifying the attributes that are relevant to therapy assessment and establishing lawful measures for those attributes. Only then can arithmetic operations be justified and only then can claims be subjected to empirical evaluation. When this sequence is reversed, measurement gives way to numerical construction and science gives way to speculation. The lesson of the scientific revolution remains unchanged: quantitative inquiry begins with measurement.

This principle has profound implications for the framework endorsed by Canada's Drug Agency (CDA), formerly CADTH, and by INESSS in Quebec. Current guidance assumes that responses to health-state descriptions can be transformed into utility scores and that these utility scores can subsequently be employed as measures of health-related quality of life, combined with time to generate quality-adjusted life years, and incorporated into simulation models that support claims regarding cost-effectiveness and value for money. Yet each of these applications presupposes the existence of lawful measurement. Without measurement, the resulting arithmetic operations have no scientific foundation.

The critical issue is not the method used to generate utility scores. Over the past four decades extensive attention has been devoted to comparing time trade-off and standard gamble techniques, evaluating alternative scoring systems, and refining utility algorithms. In Canada, the Health Utilities Index represents one of the most influential outcomes of this process. Yet these debates are secondary. Whether utility values are generated through TTO, SG, HUI, EQ-5D or any other preference-based instrument is ultimately irrelevant if the resulting scores fail to satisfy the requirements of ratio measurement. The source of the number does not alter its measurement status.

The central question is therefore not whether one utility instrument is superior to another. The central question is whether any utility-generating procedure endorsed by Canadian HTA guidance creates a lawful ratio measure. If it does not, then utility scores remain numerical constructions rather than measures. The subsequent multiplication by time of those scores transformed to proportions to generate QALYs becomes inadmissible. Likewise, the incorporation of utilities and QALYs into reference-case simulation models cannot transform those numerical constructions into scientifically meaningful quantities. Arithmetic cannot create measurement. Measurement must come first.

The challenge facing Canadian HTA is therefore institutional rather than methodological. The question is not how utilities should be constructed but why the Canadian HTA framework continues to rely upon utility-based claims despite the absence of demonstrated ratio measurement. This question lies at the heart of closure. Once the requirement for lawful measurement is ignored, debate shifts toward the refinement of utility instruments, valuation techniques, model structures and sensitivity analyses. The foundational issue disappears from view. Numerical constructions are accepted as measures, arithmetic proceeds as though measurement had been established, and the distinction between numbers and measures is lost.

The purpose of the following sections is to demonstrate that this loss has become embedded within Canadian HTA. The issue is not the merits of one utility instrument over another, nor is it the relative performance of alternative valuation techniques. The issue is that the Canadian HTA framework presupposes the existence of ratio measurement without ever demonstrating that ratio measurement exists. The consequence is measurement inversion. Arithmetic precedes measurement, and closure becomes inevitable.

THE SCALES OF MEASUREMENT

The requirement that science begins with measurement immediately raises a second question: what constitutes a measure? The answer lies in the scales of measurement. Since the work of Stevens in the mid-twentieth century, it has been recognized that not all numerical assignments possess the same measurement properties [7]. Numbers may be used as labels, rankings, intervals or ratios. The distinction is crucial because the arithmetic operations that can be performed depend entirely upon the properties of the scale.

The most elementary scale is the nominal scale. Nominal classifications simply identify categories. Numbers assigned to categories serve as labels and nothing more. The numerical values have no quantitative meaning. Ordinal scales represent the next level of complexity. Ordinal measures support ranking. One observation can be greater than, less than or equal to another. Yet the distances between ranks are unknown. An ordinal scale provides order but not magnitude. Arithmetic operations involving addition, subtraction, multiplication or division are therefore inadmissible.

Interval scales introduce an additional property. Equal numerical differences correspond to equal differences in the attribute being measured. Temperature measured in degrees Celsius provides the classic example. While interval scales permit addition and subtraction, they lack a true zero.

Consequently, multiplication and division are not meaningful. A temperature of 20 degrees Celsius is not twice as hot as a temperature of 10 degrees Celsius because the zero point is arbitrary.

Ratio scales represent the highest level of measurement. Ratio scales possess all the properties of interval scales together with a meaningful, non-arbitrary zero. Length, mass and time are familiar examples. Because the zero point represents the absence of the attribute, multiplication and division become admissible. A distance of ten meters is twice a distance of five meters. A duration of twelve months is twice a duration of six months. Ratio measurement therefore provides the foundation for quantitative science because it supports the full range of arithmetic operations.

These distinctions are not matters of convention or preference. They determine what can and cannot be done with numerical observations. Arithmetic is constrained by the scale of measurement. The scale must therefore be established before arithmetic is undertaken. This principle is fundamental. It is not possible to begin with arithmetic and subsequently infer the existence of measurement properties.

Modern representational measurement theory strengthened this conclusion. The assignment of numbers to an attribute is only meaningful if the numerical representation preserves the relevant empirical relationships. Measurement is therefore a property of the attribute and its representation, not a consequence of numerical manipulation. Numbers do not become measures because they are used in calculations. The measurement properties must exist before the calculations are undertaken.

The implications for Canadian health technology assessment are immediate. Utility scores derived from instruments such as the HUI, EQ-5D and related preference-based systems are routinely treated as though they possess ratio properties. These scores are subsequently employed as measures of health-related quality of life, multiplied to generate QALYs and incorporated into the reference-case modelling frameworks endorsed by Canada's Drug Agency and INESSS. Yet the measurement status of the underlying utility values is rarely established. Numbers are generated and arithmetic operations applied without first demonstrating the scale to which those numbers belong.

The consequences become particularly important when considering utility construction. Both standard gamble and time trade-off techniques generate numerical values from responses to health-state descriptions. The critical question is not whether numbers are produced but what scale of measurement those numbers represent. If they are ordinal, multiplication is inadmissible. If they are interval, multiplication remains inadmissible because interval scales lack a true zero. Only if ratio properties can be demonstrated does multiplication become permissible. Yet this demonstration is never attempted. Instead, ratio properties are assumed rather than established.

This point exposes the weakness of the extensive literature comparing standard gamble and time trade-off valuations. Researchers have devoted considerable effort to comparing alternative valuation methods, refining scoring algorithms, evaluating additive and multiplicative formulations, and examining sources of variation in utility estimates. Yet these debates presuppose that the measurement problem has already been solved. The central issue is not whether one valuation technique generates different scores from another. The central issue is whether either technique creates a lawful ratio measure. If neither does so, then the differences between valuation approaches are of little scientific significance.

The significance of this omission extends far beyond utility construction. Canadian HTA guidance requires utility scores to function simultaneously as measures of health-related quality of life, discount factors for the construction of QALYs, and inputs to simulation models intended to support claims regarding cost-effectiveness and value for money. Each application presupposes ratio measurement. If ratio properties are absent, then the resulting QALYs and model outputs are numerical constructions rather than lawful quantitative measures. The arithmetic may be technically correct, but it is not scientifically admissible.

The distinction between computation and measurement is therefore critical. Numbers can always be manipulated. Computers can generate increasingly sophisticated outputs from increasingly sophisticated algorithms. But computation does not create measurement. Scientific significance depends not on the complexity of the calculation but on the measurement properties of the quantities being manipulated. If those properties are absent, the resulting outputs remain numerical constructions irrespective of their apparent precision.

The scales of measurement therefore provide the first diagnostic test for closure in Canadian HTA. A discipline that cannot distinguish between nominal, ordinal, interval and ratio scales cannot distinguish between numbers and measures. Once that distinction disappears, arithmetic becomes detached from measurement, utility construction becomes detached from scale theory, and numerical outputs become substitutes for evidence. The result is measurement inversion: arithmetic preceding measurement rather than measurement preceding arithmetic. Closure follows because the discipline no longer possesses the conceptual framework necessary to recognize its own foundational error.

THE PROPERTIES OF RATIO MEASURES

The distinction between the scales of measurement immediately raises a more fundamental question: what properties must a measure possess if it is to support arithmetic operations? The answer is provided by representational measurement theory. Representational measurement is concerned with the conditions under which numerical assignments faithfully represent the empirical properties of attributes. It is not concerned with numbers themselves but with the relationship between an attribute in the real world and its numerical representation. The central question is simple: do the numbers preserve the structure of the attribute being measured?

This requirement led to the development of the axioms of representational measurement. These axioms define the conditions that must be satisfied before an attribute can be regarded as quantitative and represented by a lawful numerical scale. Although the formal mathematical treatment is complex, the underlying principles are straightforward. The attribute must be capable of ordering. Differences among observations must be meaningful. The relationships among observations must be preserved when represented numerically. Most importantly, the resulting numerical structure must support the arithmetic operations that researchers wish to perform.

The Canadian experience provides a particularly revealing illustration of what occurs when these requirements are ignored. The HUI is widely presented as a preference-based measure of health-related quality of life and has played an influential role in Canadian and international health technology assessment. The HUI begins with descriptions of health states across multiple dimensions of health. These health-state descriptions are subsequently valued through standard gamble procedures and transformed through a multiattribute scoring algorithm into utility values. The resulting scores are then treated as measures and employed in economic evaluation, burden-of-disease assessments, QALY construction and policy analysis.

The key feature is the HUI3 multiplicative scoring function. The HUI literature describes the overall utility score as being generated by a multiplicative multi-attribute utility function in which attribute-level coefficients are multiplied together and transformed by a scaling constant.

The commonly cited HUI3 formula is:

$$U = 1.371 \times (Vision \times Hearing \times Speech \times Ambulation \times Dexterity \times Emotion \times Cognition \times Pain) - 0.371$$

where each attribute level is assigned a coefficient and the coefficients are multiplied to generate the overall utility score. The critique of the HUI multiplicative algorithm rests on two separate but related concerns. The first concerns the measurement properties of the attribute-level coefficients themselves. The second concerns the admissibility of combining coefficients drawn from different attributes into a single utility score.

The first issue is straightforward. The HUI algorithm assumes that the coefficients attached to levels of vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain possess the properties required for multiplication. Yet there is no demonstration that these coefficients satisfy the requirements for ratio measurement.

Before any quantity can be regarded as a ratio measure, a number of conditions must be satisfied. The first requirement is that there must be a clearly defined attribute that is the object of measurement. Measurement is always measurement of something. Length, mass, temperature, pain and mobility are examples of attributes. Without a clearly specified attribute there can be no meaningful discussion of measurement.

The second requirement is that the attribute must be unidimensional. The measure must refer to a single quantitative property rather than a collection of unrelated characteristics. Unidimensionality is essential because arithmetic operations are meaningful only when they are applied to observations of the same attribute. Combining different attributes does not create a measure of a new attribute unless the existence of that attribute can be independently demonstrated.

The third requirement is that the scale must possess a meaningful and non-arbitrary zero. A ratio scale differs from an interval scale because zero represents the absence of the attribute being measured. This permits meaningful ratio comparisons. A value of 10 units can legitimately be described as twice a value of 5 units because both observations are referenced to a true zero.

The fourth requirement is that ratio comparisons must be meaningful throughout the scale. It must be possible to say that one observation possesses twice, three times or half the magnitude of another observation. If such statements cannot be justified, the scale cannot be regarded as a ratio measure regardless of the numerical values assigned.

The fifth requirement is that the scale must satisfy the axioms of representational measurement. The numerical representation must preserve the empirical structure of the attribute. The relationships observed among empirical observations must be reflected in the numerical assignments. Measurement is therefore not simply the assignment of numbers but the establishment of a lawful correspondence between an attribute and its numerical representation.

The final requirement is invariance. The measure must retain its meaning across populations, settings and applications. A ratio measure cannot change its interpretation because a different group of respondents is examined or because the measure is applied in a different context. The attribute being measured and the numerical representation of that attribute must remain stable.

Only when these conditions are satisfied can a measure legitimately claim ratio status and support the full range of arithmetic operations. The burden of proof rests with any proposed measure to demonstrate compliance with these requirements before arithmetic operations are undertaken. This applies both to the individual scale values of the algorithm as well as to the final utility score with the adjustment factors.

The HUI literature focuses on the estimation of preference weights but provides little evidence that the resulting coefficients possess the properties necessary to justify their treatment as ratio measures. The burden of proof rests with the HUI framework, yet that proof has never been provided [8] .

Even if this difficulty could somehow be overcome, a more fundamental problem remains. Ratio measurement applies to defined attributes. The legitimacy of arithmetic operations depends not only upon the scale properties of the measures but also upon dimensional homogeneity. It is entirely legitimate to combine ratio measures when the resulting quantity corresponds to a well-defined attribute. In the physical sciences, length multiplied by width yields area, and area multiplied by length yields volume. The resulting entities have clear empirical interpretations and satisfy the requirements of dimensional analysis.

The HUI algorithm is fundamentally different. It combines coefficients attached to a collection of distinct attributes, including vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain. These are not manifestations of a single attribute but separate dimensions of health. Multiplying coefficients across these dimensions undoubtedly produces a number, but the existence of a number does not establish the existence of a measure. The resulting quantity is labeled a utility score, yet no evidence is provided that it corresponds to a single underlying attribute possessing quantitative structure. The arithmetic therefore creates a numerical construct rather than a measure.

An oft overlooked feature of the HUI utility algorithm is the use of adjustment factors or scaling constants. In the HUI3 system, the multiplicative product of the attribute coefficients is multiplied by a constant and then adjusted by subtracting a second constant. These numerical adjustments are introduced to ensure that the final utility score conforms to a predefined range in which perfect health equals 1.0, death equals 0.0 and some health states may be assigned negative values.

The critical question is: what exactly are these adjustment factors? They are not attributes. They are not measures. They are dimensionless numerical constants introduced to transform one number into another number. Their role is purely arithmetic. They do not contribute any empirical content and they do not establish the measurement properties of the resulting utility score.

This distinction is important because adjustment factors cannot create measurement where measurement has not already been demonstrated. If the attribute-level coefficients are not established as ratio measures, multiplying them by a scaling constant does not confer ratio properties on the result. Similarly, subtracting a constant from the product cannot create a meaningful zero or establish compliance with the axioms of representational measurement.

Indeed, the presence of additive adjustment factors raises an additional question. Ratio scales are defined by a meaningful, non-arbitrary zero and are preserved only under similarity transformations. Introducing constants to shift the numerical range may create a convenient scoring convention, but it does not demonstrate that the resulting utility score is a lawful ratio measure. The adjustment factors therefore illustrate a broader problem within the HUI framework: arithmetic manipulation is used to create numerical outputs while the measurement status of those outputs remains unproven.

This is the central weakness of the HUI. The problem is not simply that the ratio status of the coefficients remains unproven. The more serious difficulty is that the multiplicative algorithm combines heterogeneous attributes without demonstrating dimensional homogeneity or identifying a single attribute being measured. The result is a utility score that may have numerical precision but lacks the measurement foundations required by the axioms of representational measurement. The HUI therefore illustrates measurement inversion in its purest form: arithmetic is undertaken first and measurement is assumed to emerge from the calculation.

This distinction is fundamental. The HUI combines responses across multiple health dimensions and transforms them into a utility score. Yet there is no demonstration that the resulting score represents a unidimensional attribute. There is no demonstration that the score possesses a true zero. There is no demonstration of additivity, cancellation, invariance or any of the other conditions required by representational measurement theory. Most importantly, there is no demonstration that the resulting values support meaningful statements regarding magnitude. The assignment of numerical values to health states does not establish that one state contains twice, half or three times the quantity of an attribute represented by another state.

The significance of this omission extends beyond the HUI itself. The HUI is not unique because it fails to establish ratio measurement. It is important because it makes the omission visible. The instrument demonstrates how utility construction can proceed without any attempt to establish the measurement properties required to support subsequent arithmetic operations. Utility values are generated, interpreted as measures and then incorporated into QALY calculations and economic models. Arithmetic follows automatically from the assumption that measurement has already been achieved.

Viewed in this way, the HUI provides one of the clearest examples of measurement inversion in Canadian HTA. Measurement theory requires that the quantitative properties of an attribute be established before arithmetic operations are undertaken. The HUI reverses this sequence. Numerical scores are generated and arithmetic applications follow, while the prior question of measurement remains unresolved. In short, unless it can be demonstrated otherwise against the criteria for scales of measurement, the utility score is just a number. This effectively closes the HUI instrument group.

The importance of the HUI therefore lies not in the particular values it generates but in what it reveals about the intellectual foundations of utility-based assessment. It demonstrates a complete failure to recognize the indispensable role of ratio measurement in supporting arithmetic applications. The issue is not whether the scoring algorithm is correct. The issue is whether the resulting score constitutes a lawful measure. Without that demonstration, the entire analytical framework rests upon unverified assumptions.

This is why representational measurement occupies such an important place in the present argument. It provides the standard against which all quantitative claims must be judged. The issue is not whether a calculation can be performed. Modern computing allows virtually any calculation imaginable. The issue is whether the calculation is meaningful. Meaningful arithmetic requires meaningful measurement. The axioms of representational measurement exist precisely to establish that distinction.

The implications for Canadian HTA are unavoidable. If the outputs of standard gamble and time trade-off exercises fail to satisfy the axioms of representational measurement, then they cannot be regarded as ratio measures. If they are not ratio measures, utility construction becomes inadmissible. If utility construction is inadmissible, then the QALY loses its measurement foundation. If the QALY loses its measurement foundation, then the reference-case models endorsed by Canada's Drug Agency and INESSS lose their claim to quantitative validity. The problem therefore begins long before cost-effectiveness analysis or simulation modelling. It begins with the failure to demonstrate that the numbers generated by preference exercises possess the properties required of ratio measures.

MANIFEST AND LATENT ATTRIBUTES

The discussion of representational measurement points to a distinction that is largely absent from Canadian health technology assessment but is fundamental to lawful measurement: the distinction between manifest and latent attributes. Every claim regarding therapy impact must ultimately refer to one or the other. Failure to recognize this distinction has contributed significantly to the emergence of measurement inversion and the subsequent development of utilities, QALYs and reference-case simulation models.

A manifest attribute is directly observable. Its existence does not have to be inferred because it can be counted, recorded or measured directly. Examples include hospital admissions, emergency department visits, hospital days, physician encounters, treatment persistence, medication possession and survival time. These attributes are visible in the world and their measurement presents no conceptual difficulty. The challenge is simply to establish an appropriate ratio scale. Once this has been achieved, arithmetic operations become admissible because the measurement properties of the attribute are known.

Latent attributes are fundamentally different. They cannot be observed directly but must be inferred from observable indicators [9]. Pain, fatigue, depression, anxiety, symptom burden, physical functioning and need fulfilment are examples. Their existence is not directly observed but inferred from responses to carefully constructed items. The central challenge is therefore not observation but measurement.

This distinction has profound implications. Manifest attributes and latent attributes require different measurement frameworks. The mistake made by contemporary HTA is to assume that numerical scores generated from questionnaires, preference exercises and health-state descriptions somehow solve the measurement problem. They do not. Assigning numbers to observations creates data, not measures. The question remains: what measurement framework is capable of transforming those observations into lawful measures?

For manifest attributes the answer is straightforward. Linear ratio measurement provides the required framework. The attribute possesses a meaningful zero, supports comparisons of magnitude and satisfies the requirements for arithmetic operations. Hospital days, for example, can be counted directly. Ten hospital days represent twice the duration of five hospital days. The arithmetic is meaningful because the measurement properties of the attribute have been established.

Latent attributes require a different solution. Because they cannot be observed directly, they cannot simply be counted. Their measurement requires a model capable of demonstrating that observations reflect positions on an underlying quantitative continuum. The objective is not the creation of scores but the measurement of the extent to which an individual possesses the latent trait. This requirement leads directly to Rasch measurement. Rasch measurement provides the necessary and sufficient framework for transforming observations into measures through the construction of invariant latent variable scales. The resulting Rasch logit ratio measure provides a lawful basis for quantitative claims regarding latent attributes.

The Canadian HTA experience demonstrates what happens when this distinction is ignored. Consider the Health Utilities Index. The HUI begins with health-state descriptions covering multiple domains including vision, hearing, speech, ambulation, dexterity, cognition, emotion and pain. Some of these domains involve observable performance. Others involve latent psychological states. Yet no attempt is made to determine whether the attributes involved are manifest or latent. No attempt is made to establish the measurement requirements appropriate to each attribute. Instead, responses are combined through a utility algorithm to generate a single utility score.

The consequence is remarkable. Before asking whether a lawful measure has been created, the HUI assumes that the measurement problem has already been solved. Questions concerning attribute identification, dimensionality, latent trait measurement, invariance and measurement structure disappear from view. The result is a utility score whose measurement status remains unknown. The issue is not whether the score is ordinal, interval or ratio. The issue is that the prior question concerning the nature of the attribute has never been addressed.

Equally significant is the complete absence of Rasch measurement from this tradition. Neither the HUI framework nor the utility-based assessment frameworks endorsed by Canada's Drug Agency and INESSS recognize the role of Rasch measurement in the evaluation of latent attributes. The distinction between manifest and latent measurement is effectively absent from the Canadian HTA knowledge base. As a consequence, latent attributes are treated as though they can be represented by preference scores, utility values or questionnaire totals without demonstrating the measurement properties required for lawful inference.

The importance of distinguishing between manifest and latent attributes cannot be overstated. Every therapy claim must refer to an attribute. That attribute must either be directly observable or inferred. There is no third category. Once this distinction is recognized, the apparent complexity of measurement largely disappears. Manifest attributes require linear ratio measures. Latent attributes require Rasch logit ratio measures. Together these two frameworks provide the complete basis for therapy assessment.

This observation has major consequences for Canadian HTA. Utilities, QALYs and preference scores are often presented as though they constitute a separate class of measurement. They do not. They are neither manifest attributes nor latent attributes. Rather, they are numerical constructions derived from responses to health-state descriptions and preference exercises. The central question therefore remains unanswered: what attribute is being measured? Until that question is addressed, the resulting numerical outputs cannot claim the status of lawful measures.

The distinction between manifest and latent attributes therefore provides another diagnostic test for closure. A discipline that cannot identify the attribute of interest, determine whether it is manifest or latent, and select the appropriate measurement framework has lost sight of the foundations of quantitative inquiry. The evidence from Canadian HTA suggests that this distinction has been absent for more than four decades. Closure did not emerge because the wrong utility instrument was selected or because the wrong scoring algorithm was adopted. Closure emerged because the discipline never asked the prior question: what attribute is being measured and how should that attribute be measured? The Rasch concept of the possession of the latent attribute has never been raised yet is the reason for the role of the Rasch logit ratio scale.

THE RATIO MEASURE AND THE INVERSION OF MEASUREMENT

The AI large language model interrogations of HTA knowledge bases provide compelling evidence of measurement failure in Canadian HTA. Across agencies, research centers, journals and educational programs there is strong endorsement of propositions that are inconsistent with ratio measurement. Ratio measures are assumed to permit negative values. QALYs are assumed to be ratio measures. Summations of subjective responses are assumed to create ratio measures.

QALYs are assumed to be dimensionally homogeneous. At the same time, propositions that define ratio measurement receive little support. Measures need not be unidimensional. Multiplication does not require ratio measures. Measurement does not precede arithmetic. These findings are not isolated misconceptions. Together they point to a systematic decades long misunderstanding of the properties that distinguish ratio measures from all other numerical constructions.

To illustrate this measurement inversion for Canada a summary of endorsement for 5 false and 5 true statements is presented in Table 1. This provides a comprehensive picture of the intellectual foundations of contemporary health technology assessment (HTA) in Canada. Six leading Canadian institutions were examined: the Toronto Health Economics and Technology Assessment Collaborative (THETA), the Health Technology and Policy Unit (HTPU) at the University of Alberta, the Institute of Health Economics (IHE), the Institute for Clinical Evaluative Sciences (ICES), Canada's Drug Agency (CDA) and the Centre for Clinical Epidemiology and Evaluation (C2E2) at the University of British Columbia. Together these organizations represent the core of the Canadian HTA enterprise, encompassing academic research, policy analysis, technology assessment, outcomes research and reimbursement decision-making.

TABLE 1

MEASUREMENT INVERSION: CANADA

CANONICAL STATEMENT	CATEGORICAL PROBABILITY ENDORSEMENT					
	THETA	HTPU	IHE	ICES	CDA	C2E2
FALSE STATEMENTS						
1, Ratio measures can have negative values	0.90	0.90	0.90	0.90	0.85	0.90
2.The QALY is a ratio measure	0.90	0.85	0.95	0.85	0.85	0.90
3, Summations of subjective instrument responses are ratio measures	0.90	0.80	0.90	0.80	0.85	0.80
4. The QALY is a dimensionally homogeneous measure	0.95	0.85	0.95	0.85	0.80	0.85
5. Reference case simulations generate falsifiable claims	0.90	0.80	0.90	0.80	0.85	0.85
TRUE STATEMENTS.,						
6. Measures must be unidimensional	0.10	0.10	0.10	0.10	0.15	0.10
7. Multiplication requires a ratio measure	0.10	0.10	0.10	0.10	0.15	0.10
8. There are only two classes of measurement: linear ratio and Rasch logit ratio	0.05	0.05	0.05	0.05	0.10	0.05

9.Measurement precedes arithmetic	0.10	0.10	0.10	0.10	0.15	0.20
10. The outcome of interest for latent traits is the possession of that trait	0.05	0.05	0.05	0.05	0.10	0.05

The findings are remarkable not because of their diversity but because of their uniformity. Across all six institutions there is strong endorsement of propositions that are false within the framework of representational measurement and correspondingly weak endorsements of propositions that are true. The pattern is so consistent that it is difficult to interpret the results as anything other than evidence of a deeply embedded conceptual framework that has become detached from the requirements of measurement theory.

The significance of this conclusion extends beyond Canada itself. Canada occupies a distinctive place in the history of HTA. Through the development of the Health Utilities Index (HUI), the widespread adoption of utility-based assessment and the influence of Canadian researchers on international HTA practice, Canada has played a major role in shaping contemporary approaches to economic evaluation. If the Canadian HTA knowledge base demonstrates systematic measurement inversion, then the implications extend far beyond national borders. They call into question one of the intellectual foundations of modern HTA. With over 230 interrogations reported on the Maimon Research website www.maimonresearch.com the findings are conclusive: each interrogation produce measurement inversion. Canada in not alone

The most striking feature of Table 1 is the overwhelming endorsement of propositions that are false. The statement that ratio measures can have negative values receives probabilities ranging from 0.85 to 0.90. The proposition that the QALY is a ratio measure receives probabilities ranging from 0.85 to 0.95. The proposition that summations of subjective instrument responses create ratio measures receives endorsement probabilities of 0.80 to 0.90. Similarly, the proposition that the QALY is dimensionally homogeneous receives probabilities of 0.80 to 0.95.

These are not modest levels of support. They indicate strong confidence in propositions that directly contradict the requirements of representational measurement. Ratio measures require a meaningful non-arbitrary zero. They do not permit negative values. Dimensional homogeneity requires compatibility of dimensions under arithmetic operations. Ordinal preference structures cannot be transformed into ratio measures simply through numerical manipulation. Yet these propositions receive widespread endorsement across the Canadian HTA landscape.

The same pattern is evident in attitudes towards reference-case simulation modelling. The proposition that reference-case simulations generate falsifiable claims receives endorsement probabilities ranging from 0.80 to 0.90. This finding is particularly revealing because falsifiability occupies a central place in scientific inquiry. A claim that cannot in principle be subjected to empirical refutation cannot be regarded as scientific. Yet reference-case models project hypothetical outcomes into imaginary future worlds whose assumptions can be continuously modified and recalibrated. Their outputs are not empirical claims but conditional numeric projections. Nevertheless, they are treated as though they satisfy the standards of normal scientific investigation.

The implications of these findings become even clearer when attention shifts to the true statements. The proposition that measures must be unidimensional receives endorsement probabilities of only 0.10 to 0.15. The proposition that multiplication requires a ratio measure receives virtually identical probabilities. The proposition that measurement precedes arithmetic receives endorsement probabilities of only 0.10 to 0.20. These findings suggest that the foundational principles governing quantitative measurement receive little recognition within the Canadian HTA knowledge base.

Perhaps the most revealing result concerns the statement that there are only two classes of measurement relevant to therapy assessment: linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes. This proposition receives endorsement probabilities of only 0.05 across five of the six institutions and only 0.10 for Canada's Drug Agency. These are effectively floor-level responses. They indicate not merely disagreement but near-complete absence of recognition.

The significance of this result cannot be overstated. The distinction between manifest and latent attributes lies at the heart of measurement. Manifest attributes are directly observable. Hospital admissions, emergency department visits, physician encounters, hospital days and treatment persistence can be counted and represented through linear ratio scales. Latent attributes such as pain, fatigue, symptom burden and physical functioning cannot be observed directly. Their measurement requires Rasch methodology and logit ratio scales. Together these two measurement frameworks provide a complete basis for therapy assessment.

Yet the Canadian HTA knowledge base appears largely unaware of this distinction. Instead, utilities, QALYs and simulation models dominate the evaluative framework. The consequence is measurement inversion. Rather than establishing the measurement properties of attributes before undertaking arithmetic operations, arithmetic becomes the starting point. Numerical manipulation substitutes for measurement.

This inversion is clearly visible in the treatment of the QALY. The QALY occupies a central position within Canadian HTA. It serves as a common metric for comparing therapies across disease areas and forms the basis of cost-utility analysis. Yet the QALY depends entirely upon the assumption that utility values possess ratio properties. If utilities are not ratio measures, then multiplication with time is inadmissible. The resulting QALY cannot become a ratio measure merely because arithmetic has occurred.

The Canadian interrogation results suggest that this problem is not recognized. The QALY is strongly endorsed as a ratio measure, while the proposition that multiplication requires a ratio measure receives virtually no support. This is measurement inversion in its purest form. The arithmetic outcome is accepted while the measurement conditions necessary to justify that outcome are ignored.

The same conclusion applies to subjective instruments. Across all six institutions there is strong endorsement of the proposition that summations of subjective responses create ratio measures. This belief underlies much of contemporary quality-of-life assessment. Questionnaire responses are assigned numerical values, summed and transformed through scoring algorithms. The resulting outputs are then treated as quantitative measures. Yet arithmetic cannot create measurement. Summing ordinal responses does not transform them into ratio scales. The resulting scores remain numerical constructs whose measurement is typically ordinal scores.

INITIAL AND SUBSEQUENT INVERSIONS

The significance of this failure becomes apparent when examining the origins of utility construction. Both time trade-off and standard gamble techniques generate numerical scores. Yet the critical question is never addressed: do these scores possess the properties required of ratio measures? The answer is assumed rather than demonstrated. No evidence is offered that the outputs satisfy the axioms of representational measurement. No evidence is offered that they possess a meaningful zero, support ratio comparisons or maintain invariance across contexts. Nevertheless, they are immediately treated as though they were lawful measures.

This is the first inversion. Numbers are accepted as measures without establishing their measurement status. The distinction between numerical assignment and measurement disappears. Once that distinction is lost, the second inversion follows naturally. Arithmetic operations are applied to entities whose measurement properties remain unknown. Additive and multiplicative algorithms are employed to create utility values. Utility values are multiplied by time to create QALYs. QALYs are incorporated into simulation models that generate claims regarding cost-effectiveness and value for money. At each stage arithmetic assumes the existence of ratio measurement while simultaneously avoiding the requirement to demonstrate it.

The result is arithmetic chaos. One inadmissible operation is layered upon another until the final outputs appear as precise quantitative estimates. Yet the precision is illusory. Arithmetic cannot create ratio measurement. Multiplication cannot transform an ordinal structure into a ratio scale. Aggregation cannot create measurement properties that were absent from the original observations. The entire sequence remains dependent upon the initial assumption that the outputs of time trade-off and standard gamble exercises are lawful measures. Once that assumption is challenged, the subsequent constructions collapse.

The central weakness of the reference case model in health technology assessment is that it creates the appearance of a quantitative claim without satisfying the conditions required for a scientific claim. The outputs of the model are not discoveries about the world. They are numerical projections generated from a structure of assumptions chosen by the modeler. Every component of the analysis, the disease pathway, treatment effects, transition probabilities, utility weights, discount rates, time horizon, persistence assumptions, mortality risks, and resource utilization estimates, is selected and assembled within an analytical framework that predetermines the range of possible outcomes.

The resulting incremental cost-per-QALY estimate is therefore not an empirical observation. It is a model artifact. Change the assumptions, alter the structure, select a different utility instrument, modify the transition probabilities, or extend the time horizon, and a different result emerges. The model does not reveal a property of the world; it reveals the consequences of the assumptions embedded within the model.

More importantly, the reference case claim cannot be falsified. Scientific claims must be presented in a form that allows empirical evaluation. A claim must specify what evidence would demonstrate that it is wrong. This is the central lesson of Popperian science. Yet a lifetime simulation projecting costs and outcomes over decades cannot be subjected to empirical refutation within any practical decision-making timeframe. There is no protocol, no target population, no prospective observation period, and no clearly defined endpoint against which the claim can be tested. The claim is therefore insulated from failure.

Sensitivity analyses do not solve this problem. They merely demonstrate how the model behaves when assumptions are altered. They are exercises in internal consistency rather than empirical validation. A model may prove robust to changes in assumptions while remaining entirely disconnected from observable reality.

The consequence is that the reference case does not function as a scientific instrument. It functions as a rhetorical device. It produces precise numerical outputs that convey an impression of objectivity while avoiding the fundamental requirement that claims be exposed to the possibility of failure. The apparent precision of the model disguises the fact that no evaluable claim has been made.

The irony is that the stated purpose of health technology assessment is to support evidence-based decision making. Yet the reference case abandons the evidentiary standards that define science. Rather than generating hypotheses capable of replication, refutation, and revision, it creates imaginary future worlds populated by assumptions and then reports the consequences of those assumptions as if they were evidence. The result is not scientific assessment but numerical storytelling. The role of HTA should be to generate credible, testable, and reproducible claims regarding therapy impact. The reference case achieves the opposite. It substitutes simulation for evidence, assumption for observation, and consensus modeling for scientific inquiry.

This explains why closure is inseparable from measurement inversion. Closure does not occur because researchers make mistakes. Mistakes are corrected within functioning scientific disciplines. Closure occurs because the concepts required to identify the mistake have never been appreciated. Ratio measurement has never been treated as the foundation of quantitative inquiry. Instead, numerical manipulation becomes an end in itself. Utilities are refined, QALYs recalculated and simulation models expanded without ever returning to the original question: what are the measurement properties of the entities entering these calculations?

The consequence is that what exists as contemporary HTA inhabits an analytical framework in which arithmetic has replaced ratio measurement as the basis for quantitative claims. This is not merely a methodological weakness. It is a rejection of the standards that have governed measurement since the scientific revolution of the 17th century and that were later formalized through representational measurement theory. The result is a discipline that generates increasingly sophisticated numbers while remaining unable to demonstrate that those numbers represent lawful measures. Closure follows not because HTA has exhausted its analytical possibilities but because it abandoned the measurement foundations upon which quantitative science depends decades ago; it has abandoned lawful science.

THE ABSENCE OF RASCH MEASUREMENT

A paradox in Canadian HTA is that despite an implicit recognition in HUI health state descriptions of latent attributes, no mention is made in the HTA knowledge bases of the contribution of Rasch measurement. This is a common omission in HTA; the implications are significant. Once the distinction between manifest and latent attributes is made then the logical conclusion is to focus on two measures: linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes.

To illustrate this omission and the lack of awareness, consider Table 2 which summarizes the endorsement of 5 Rasch related statements for the 5 research centers and the CDA. The results presented in Table 2 provide perhaps the clearest evidence of the measurement foundations of contemporary Canadian HTA. Across all six organizations—THETA, HTPU, IHE, ICES, CDA and C2E2—endorsement probabilities for the five Rasch statements are effectively at floor level. Four organizations return values of 0.05 for every statement, while Canada's Drug Agency reaches only 0.10. These are not merely low levels of endorsement; they indicate an almost complete absence of recognition of Rasch measurement as a framework for quantitative assessment.

This finding is important because Rasch measurement occupies a unique position within measurement theory. Unlike conventional scoring systems, psychometric scales and item response models, Rasch measurement is explicitly concerned with the transformation of observations into measures. The objective is not the production of scores but the construction of a measurement instrument capable of locating both persons and items on a common quantitative continuum. The outcome is the Rasch logit ratio scale, the only framework capable of transforming observations of latent attributes into lawful measures.

The first statement, that there are only two classes of measurement relevant to therapy assessment, linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes, received virtually no endorsement. This is particularly revealing because it indicates that the distinction between manifest and latent attributes is itself largely absent from Canadian HTA. Without that distinction there can be no coherent measurement strategy.

TABLE 2

ENDORSEMENT OF RASCH MEASUREMENT: CANADA

	ENDORSEMENT CATEGORICAL PROBABILITIES					
STATEMENT	THETA	HTPU	IHE	ICES	CDA	C2E2
1. There are only two classes of measurement: linear ratio and Rasch logit ratio	0.05	0.05	0.05	0.05	0.10	0.05
2. Transforming subjective responses to interval measurement is only possible with Rasch rules	0.05	0.05	0.05	0.05	0.10	0.05
3. The Rasch logit ratio scale is the only basis for assessing therapy impact for latent traits	0.05	0.05	0.05	0.05	0.10	0.05
4. The outcome of interest for latent traits is the possession of that trait	0.05	0.05	0.05	0.05	0.10	0.05
5. The Rasch rules for measurement are identical to the axioms of representational measurement	0.05	0.05	0.05	0.06	0.10	0.05

Similarly, the proposition that subjective responses can only be transformed into measurement through Rasch rules receives no meaningful support. This suggests that the dominant assumption remains that questionnaire scores, preference scores and composite indices can somehow acquire measurement properties through aggregation, weighting or statistical manipulation. Rasch measurement rejects this position entirely. Measurement cannot be created through arithmetic. It must be demonstrated through adherence to the Rasch model.

The same pattern is evident in the responses to statements concerning possession of a latent trait and the relationship between Rasch measurement and the axioms of representational measurement. Both receive near-zero endorsement. Yet these are among the defining characteristics of Rasch measurement. The objective is to estimate the extent to which an individual possesses a latent attribute, while simultaneously satisfying the same measurement requirements that govern quantitative inquiry elsewhere in science.

Taken together, these findings indicate that Canadian HTA possesses little if any awareness of Rasch measurement as the only lawful framework for transforming observations of latent attributes into measures. This is not a minor omission. It means that while utilities, QALYs and simulation models are widely endorsed, the only recognized measurement framework for latent attributes remains effectively invisible. The consequence is that Canadian HTA has no coherent measurement strategy for latent outcomes and remains committed to numerical constructions that lack demonstrated measurement properties.

CONCLUSION: THE FUTURE FOR HTA IN CANADA

Closure occurs when a discipline can no longer distinguish between numbers and measures.

Once it is accepted that Canadian HTA fails to recognize the central role of ratio measurement in supporting quantitative, evaluable and falsifiable claims, the conclusion is unavoidable: the present framework is closed and has been for some 40 years. This is not simply a matter of methodological weakness or the need for incremental refinement. It is a recognition that the foundations of contemporary HTA were established without satisfying the requirements of measurement. The consequence is that utilities, QALYs and reference-case simulation models cannot be regarded as scientific instruments for evaluating therapy impact.

The Canadian interrogations reveal two related failures. The first is the widespread endorsement of propositions incompatible with ratio measurement. The second, and perhaps more significant, is the near-total absence of awareness of Rasch measurement. Across Canada's Drug Agency, THETA, HTPU, IHE, ICES and C2E2, endorsement probabilities for core Rasch principles were effectively at floor level. There was little recognition that only two forms of measurement are relevant to therapy assessment: linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes. There was similarly little recognition that Rasch measurement provides the only lawful framework for transforming observations of latent attributes into measures.

These findings are particularly important because Canadian HTA occupies a distinctive position in the international development of the discipline. Through the Health Utilities Index (HUI), Canada became one of the principal architects of utility-based assessment. Through Canada's Drug Agency and provincial assessment programs, it subsequently embraced the reference-case framework built around utilities, QALYs and simulation models. The result is a uniquely revealing combination. Canada helped create one of the world's most influential utility instruments while simultaneously demonstrating little awareness of the only accepted framework for measuring latent attributes.

The HUI illustrates the problem. Numerical values derived from health-state descriptions were accepted as measures without demonstrating compliance with ratio measurement. These values became utilities, utilities became QALYs and QALYs became inputs to simulation models. Yet the Canadian interrogations suggest little recognition of the measurement requirements necessary to justify any stage of this progression. At the same time, the near absence of Rasch measurement indicates that Canadian HTA lacks a coherent alternative framework for evaluating latent attributes such as pain, fatigue, functioning and quality of life.

The implications are therefore uncompromising. Utilities must be abandoned. QALYs must be abandoned. Reference-case simulation models must be abandoned. These constructs are not measures and cannot be transformed into measures through increasingly sophisticated arithmetic, statistical manipulation or modelling exercises. Their continued use can only perpetuate arithmetic chaos.

The task facing Canada is therefore not reform but reconstruction. Four decades of methodological accumulation have produced an extensive literature devoted to utility construction, QALY estimation and simulation modelling, yet have left unresolved the prior question of measurement itself. The challenge is to rebuild HTA around the only two lawful measures: linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes. Claims must be capable of replication, reproduction and falsification in real populations and over defined time horizons.

The responsibility extends beyond Canada's Drug Agency and provincial assessment bodies. Canadian universities, HTA research centres and professional training programs must reconstruct their curricula and research agendas around the principles of measurement science. Unless this occurs, each new generation of researchers will continue to reproduce the assumptions that created closure in the first place.

The choice facing Canada is therefore clear. Continue to defend utilities, QALYs and simulation models despite the absence of lawful measurement, or reconstruct HTA around the standards that govern every other quantitative science. The future of Canadian HTA depends upon that choice.

ACKNOWLEDGEMENT

I acknowledge that I have used OpenAI technologies, including the large language model, to assist in the development of this work. All final decisions, interpretations, and responsibilities for the content rest solely with me.

REFERENCES

[1] Furlong W, Feeny D, Torrance G et al. Multiplicative Multi-Attribute Utility Function for the Health Utilities Index Mark 3 (HUI3) System. CHEPA Working Paper 98-11, McMaster University, 1998

[2] Drummond M, Sculpher M, Claxton K et al. Methods for the Economic Evaluation of Health Care Programmes.(4th Ed). New York: Oxford University Press, 2015

[3] Maimon Research, Canada Interrogations https://maimonresearch.com/ai-llm-rest-of-world/#canada

[4] Wootton D. The Invention of Science: A New History of the Scientific Revolution. New York: HarperCollins, 2015

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

[6] Popper K**.** *The Logic of Scientific Discovery*. London: Hutchinson; 1959

[7] Stevens S. On the Theory of Scales of Measurement. *Science*. 1946; 103 (2684), 677–680

[8] Furlong W., Feeny D, Torrance W. . *Health Utilities Index (HUI) Procedures Manual: Algorithm for Determining HUI Mark 2 (HUI2) / Mark 3 (HUI3) Health Status Classification Levels, Health States, Single-Attribute Utility Scores and Overall Health-Related Quality of Life Utility Scores from HUI Questionnaires*. Dundas, Ontario: Health Utilities Inc.2012

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