MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

CANADA: THE HEALTH UTILITIES INDEX: FROM MEASUREMENT INVERSION TO ARITHMETIC CHAOS

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

LOGIT WORKING PAPER No 1708 JUNE 2026

www.maimonresearch.com

Tucson AZ

ABSTRACT

The Health Utilities Index (HUI) occupies a unique position in the history of health technology assessment (HTA). Developed in Canada and subsequently adopted internationally, the HUI became one of the principal sources of utility values used in quality-adjusted life years (QALYs), health-adjusted life expectancy, population health reporting, economic evaluations and reference-case simulation models. Through its widespread use by researchers, policy makers, Statistics Canada and HTA agencies, the HUI helped establish utility-based assessment as a central component of modern healthcare decision making. This paper examines a more fundamental question: does the HUI generate lawful measures capable of supporting the arithmetic operations upon which these applications depend?

Drawing upon a recent interrogation of the HUI knowledge base, the paper argues that the answer is no. Foundational propositions concerning representational measurement, unidimensionality, ratio measurement, dimensional homogeneity and Rasch measurement receive little or no endorsement, while utility construction and its associated arithmetic remain largely unquestioned. These findings point to a pervasive pattern of measurement inversion, where arithmetic precedes measurement and numerical manipulation substitutes for the establishment of lawful measures.

The analysis follows the pathway from health-state descriptions to preference valuations, from preference valuations to utility scores and from utility scores to QALYs and simulation models. Particular attention is given to the ordinal nature of preference structures and to the multiplicative scoring algorithm at the heart of the HUI. The paper argues that ordinal preference rankings cannot be transformed into ratio measures through arithmetic operations alone and that the multiplicative algorithm assumes precisely those measurement properties it is required to demonstrate. The resulting utility score is therefore not a measure but the numerical output of an inadmissible arithmetic procedure.

The implications extend far beyond the HUI itself. Utility values generated by the HUI continue to inform population health reporting, health-adjusted life expectancy calculations, cost-utility analyses and reimbursement decisions. Any failure to establish lawful measurement at the utility stage necessarily propagates throughout these frameworks. The result is arithmetic chaos: a sequence of numerical constructions in which one inadmissible arithmetic operation builds upon another.

The paper concludes that the HUI represents more than a utility instrument. It provides a case study in the institutionalization of measurement inversion within Canadian HTA. The broader lesson is that numerical outputs should not be confused with measures. The future of HTA lies not in refining HUI utility algorithms but in restoring lawful measurement through linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes.

INTRODUCTION

The Health Utilities Index (HUI) occupies a unique position in the history of health technology assessment (HTA) [i]. Developed in Canada and subsequently adopted throughout the world, the HUI became one of the most influential systems for generating utility values to support quality-adjusted life years (QALYs), health-adjusted life expectancy, population health assessments and reference-case simulation models. Through its widespread use in research, reimbursement submissions and public policy, the HUI helped establish utility-based assessment as a central component of modern HTA. Its influence extends well beyond Canada, shaping methodological standards adopted by government agencies, academic research centers and professional organizations internationally.

The importance of the HUI lies not merely in its popularity but in its role within a much larger analytical framework. HUI utility scores are routinely employed as inputs to QALY construction, cost-utility analyses and simulation models designed to estimate the long-term value of competing therapies. As a consequence, the measurement properties of HUI utilities are of fundamental importance. If the utility values generated by the HUI possess lawful measurement properties, then the arithmetic operations that follow may be justified. If they do not, then every subsequent calculation inherits that deficiency.

A recent interrogation of the HUI knowledge base presents a troubling picture [ii]. Foundational propositions concerning representational measurement, dimensional homogeneity, ratio measurement and Rasch measurement receive little support, while the assumptions required to sustain utility construction remain largely unquestioned. The results point to a pervasive pattern of measurement inversion. Measurement inversion occurs when arithmetic precedes measurement; when numerical manipulation is undertaken before demonstrating that the entities entering the calculation possess the properties required for quantitative analysis. The significance of this finding extends far beyond the HUI itself. Measurement inversion is not merely a methodological curiosity. It is a diagnostic signal.

The purpose of this paper is to examine the implications of that signal. The concern is not whether the HUI is statistically sophisticated, widely used or historically influential. Nor is the concern whether particular valuation studies or scoring algorithms might be improved. The central question is more fundamental: does the HUI generate lawful measures? To answer this question, the paper follows the pathway from health-state descriptions to preference valuations, from preference valuations to utility scores and from utility scores to QALYs and simulation models. The argument advanced is that the HUI provides a clear illustration of how measurement inversion gives rise to arithmetic chaos. Once arithmetic becomes detached from lawful measurement, numerical outputs may become increasingly sophisticated, but they cannot be assumed to possess quantitative meaning. The consequences extend from the HUI itself to the broader framework of utility-based assessment that has dominated HTA for more than four decades.

THE HUI AND THE PROMISE OF UTILITY MEASUREMENT

The Health Utilities Index emerged from a simple but ambitious objective: to create a single numerical indicator capable of representing an individual's overall health status. The challenge was immediately apparent. Health is not a single observable characteristic. It encompasses multiple dimensions including vision, hearing, speech, ambulation, dexterity, cognition, emotion

and pain. The HUI addressed this challenge by constructing a classification system in which each attribute was represented by a series of levels describing differing degrees of impairment or functioning. A person's health status could then be represented by a combination of attribute levels across these dimensions.

At first sight, this appears a reasonable undertaking. Health-state descriptions can provide a structured framework for characterizing disease burden and treatment impact. The difficulty is that classification is not measurement. A health-state description may distinguish one health profile from another, but it does not establish that the resulting categories possess quantitative properties. The HUI nevertheless sought to move beyond description by attaching values to health states through preference elicitation techniques, principally the standard gamble procedure.

The standard gamble asks respondents to express preferences among alternative health states under conditions of uncertainty. The resulting valuations are intended to reflect the relative desirability of different health outcomes. These preference values then become inputs to a multiplicative scoring algorithm that generates a single utility score. It is this utility score that subsequently becomes the foundation for QALY construction and a wide range of HTA applications.

The attraction of the HUI is obvious. By transforming complex health experiences into a single number, it appears possible to compare therapies across diseases, populations and interventions. A utility score can be averaged across patients, compared across groups and incorporated into economic evaluations. The promise of utility measurement is therefore simplicity. Diverse experiences of health and illness are reduced to a common numerical metric.

This promise explains the extraordinary influence of the HUI. Utility values generated by the instrument have been employed in cost-utility analyses, reimbursement submissions, population health studies and health-adjusted life expectancy calculations. Statistics Canada has used HUI scores in national health surveys and population reporting. Academic researchers have used HUI utilities to compare disease burdens, evaluate interventions and estimate treatment benefits. Within HTA, HUI utilities became one of the principal building blocks for QALY construction and simulation modelling.

Yet beneath this apparent success lies a question that has received remarkably little attention. What exactly is the measurement status of the utility score produced by the HUI? The construction process begins with multidimensional health-state descriptions and proceeds through preference valuations to a single numerical value. At what point does measurement occur? Is the utility score a lawful measure, or is it merely the numerical output of a scoring algorithm? These questions are fundamental because the legitimacy of every subsequent arithmetic operation depends upon the answer.

The interrogation of the HUI knowledge base, discussed in the previous paper, provides an important clue. The interrogation demonstrated almost no endorsement for propositions concerning unidimensionality, ratio measurement, representational measurement and Rasch measurement. Equally important, there was little support for the proposition that measurement must precede arithmetic. These findings suggest that the HUI was never developed within a

framework primarily concerned with measurement. Rather, the emphasis was placed upon classification, preference elicitation and utility construction.

This distinction is critical. The objective of the HUI was not merely to describe health states but to generate utility values capable of supporting arithmetic operations. The question is therefore not whether the HUI produces numbers. Clearly it does. The question is whether those numbers possess the measurement properties necessary to justify their subsequent use in QALYs, simulation models and policy evaluations. The answer to that question requires a closer examination of the pathway from health-state descriptions to utility scores. It is here that the signal of measurement inversion first appears and where the foundations of arithmetic chaos are laid.

UTILITIES AS ORDINAL SCORES

A striking feature of both the HUI framework and the interrogation results is the absence of any recognition of the scales of measurement first described by Stevens in 1946 [iii] . Stevens distinguished among nominal, ordinal, interval and ratio scales, emphasizing that the arithmetic operations that can be applied depend entirely upon the measurement properties of the scale involved. Ordinal scales provide rank order. They indicate that one state is preferred to another, but they do not establish equal intervals, constant units or a true zero. Consequently, ordinal scores cannot support the arithmetic operations required for quantitative measurement.

This omission is critical because the HUI utility framework originates in preference valuations. Respondents' express judgments regarding the relative desirability of alternative health states. These valuations provide an ordering of preferences, but there is no demonstration that they possess interval or ratio properties. Yet once incorporated into the HUI scoring algorithm, these ordinal structures are magically transformed into utility scores that are subsequently treated as though they were quantitative measures. The resulting utilities are averaged, compared, multiplied and incorporated into QALYs and simulation models.

Equally revealing is the interrogation itself. The responses demonstrate little awareness of the implications of Stevens' classification. The distinction between ordinal and ratio measurement is largely absent in textbook presentations [iv] This is important because the entire utility framework depends upon treating preference-based ordinal structures as though they support arithmetic operations. The result is a fundamental confusion between ranking and measurement. Numbers are generated, but the existence of numbers does not establish that measurement has occurred.

The failure to recognize the limitations of ordinal scores has profound implications for the HUI framework. Ordinal structures can support ranking, but they cannot support arithmetic operations such as multiplication, division or the construction of weighted scoring algorithms intended to create quantitative measures. Attaching weights to preference responses and combining them through a multiplicative formula does nothing to alter the underlying measurement status of the data. The preference structure remains ordinal. Arithmetic cannot transform an ordinal ranking into a ratio measure simply by generating a new number. Indeed, the application of such arithmetic is itself inadmissible because the properties required to justify the operation have never been established. The HUI scoring algorithm therefore illustrates the central problem of measurement inversion: arithmetic is used as a substitute for measurement.

The final utility score, expressed on a conventional scale between 0 and 1 and sometimes extending below zero, is often interpreted as though it were a quantitative measure of health-related quality of life. Yet the scoring process provides no evidence that the resulting value possesses interval or ratio properties. The utility score is therefore not a measure but the numerical output of an algorithm applied to ordinal preference structures. Its apparent precision conceals the fact that it has no demonstrated measurement properties. The resulting utility is simply a number, and the subsequent use of that number in QALYs, health-adjusted life expectancy and simulation models merely extends the chain of inadmissible arithmetic.

THE INADMISSIBLE MULTIPLICATIVE ALGORITHM

At the heart of the Health Utilities Index (HUI) lies a multiplicative scoring algorithm intended to transform health-state descriptions and preference valuations into a single utility value. This utility is subsequently treated as a quantitative measure of health-related quality of life and employed in the construction of QALYs, health-adjusted life expectancy, simulation models and cost-effectiveness claims. The legitimacy of this entire framework rests upon a simple question: are the arithmetic operations involved admissible? The answer depends not on the sophistication of the algorithm but on the measurement properties of the entities entering the calculation.

A multiplicative algorithm imposes demanding requirements on its inputs. Multiplication is not an arithmetic operation that can be applied indiscriminately to any set of numbers. For multiplication to be meaningful, the variables involved must possess lawful measurement properties. In particular, they must be measured on ratio scales with meaningful non-arbitrary zeros and support the comparison of magnitudes. Multiplication presupposes measurement. It cannot create measurement where measurement does not already exist.

This principle is widely recognized in the physical sciences. Distance may be calculated as speed multiplied by time because both speed and time are ratio measures. Similar examples are found throughout physics, chemistry and engineering. The legitimacy of the arithmetic operation derives entirely from the measurement properties of the quantities involved. If those properties are absent, the multiplication becomes inadmissible.

The HUI algorithm does not satisfy this requirement; the axioms of representational measure are absent [v] . The scoring system begins with health-state descriptions comprising multiple attributes and levels. These descriptions are not measures. They are classifications that distinguish among health profiles. Preference exercises, typically based on standard gamble techniques, are then used to assign values to these health states. The resulting preference scores provide rankings and valuations, but there is no demonstration that they possess ratio properties. They indicate relative desirability rather than measurable quantities.

The crucial step occurs when these preference-derived values are entered into a multiplicative scoring algorithm. The algorithm combines attribute weights through a formula designed to generate a utility score, often expressed on a conventional scale ranging from 0 to 1, with some health states assigned negative values. The existence of the algorithm creates the appearance of quantitative rigor. Yet the appearance is deceptive. The arithmetic operation assumes precisely

what it is required to demonstrate. It assumes that the preference weights possess the measurement properties necessary to support multiplication.

This assumption is never justified. Ordinal preference structures do not become ratio measures because they are inserted into a multiplicative formula. Arithmetic cannot transform rankings into quantities. The output of the algorithm cannot possess stronger measurement properties than the inputs from which it is derived. If the inputs are ordinal preference structures, the resulting utility cannot become a ratio measure simply because multiplication has occurred.

The problem is therefore not the complexity of the formula but its admissibility. Multiplication requires ratio measures. The HUI framework provides no evidence that the preference weights entering the calculation satisfy this requirement. Consequently, the resulting utility value is not a demonstrated measure. It is the numerical output of an inadmissible arithmetic operation.

The allowance of negative utility values further reinforces this conclusion. Ratio measures require a meaningful non-arbitrary zero and do not permit negative values. Yet HUI utilities may fall below zero for health states judged worse than death. This feature alone is inconsistent with the interpretation of utilities as ratio measures. A scale that permits negative values cannot simultaneously satisfy the requirements of ratio measurement.

The consequences extend well beyond the utility score itself. HUI utilities are routinely employed as inputs to QALY construction. Utility values are combined with time to estimate quality-adjusted life years. Yet if the utility lacks ratio properties, the multiplication of utility by time is equally inadmissible. The resulting QALY inherits the deficiencies of the utility structure from which it is derived. The process then continues through simulation models, health-adjusted life expectancy calculations and cost-effectiveness analyses. At each stage the numerical outputs inherit the limitations of the original utility construction.

This is the essence of arithmetic chaos. An inadmissible multiplicative algorithm generates utility values. Those utility values become inputs to further arithmetic operations. The resulting outputs are treated as though they were lawful measures despite the absence of any demonstration that measurement has occurred. Numerical sophistication increases while measurement remains absent.

The judgment is therefore straightforward. The HUI multiplicative algorithm lacks measurement legitimacy because the entities entering the calculation do not possess demonstrated ratio-scale properties. The arithmetic operation is inadmissible. The resulting utility is not a utility measure but a numerical construct generated by applying multiplication to ordinal preference structures. Its apparent precision conceals the absence of lawful measurement. The utility is simply a number, and the subsequent use of that number in QALYs, simulation models and policy analyses merely extends the chain of inadmissible arithmetic.

The HUI thus provides a clear illustration of the sequence that characterizes much of contemporary HTA. Measurement inversion is the signal. The inadmissible multiplicative algorithm is the mechanism. Arithmetic chaos is the consequence. Closure is the outcome.

MEASUREMENT INVERSION AS A DIAGNOSTIC SIGNAL

The interrogation of the HUI knowledge base provides an important insight into the intellectual foundations of utility construction in Canada. The results do not merely identify disagreement with particular propositions concerning measurement. Rather, they reveal a coherent pattern in which the requirements of representational measurement receive almost no support while the numerical structures built upon those requirements continue to be accepted and applied. This pattern is best described as measurement inversion.

Measurement inversion occurs when arithmetic precedes measurement. In every quantitative science, the first requirement is to establish that the attribute under consideration possesses the properties necessary for measurement. Only then can arithmetic operations be undertaken. The interrogation results suggest that the HUI framework reverses this sequence. Numerical operations are accepted as legitimate while the measurement foundations that would justify those operations are largely ignored.

Several findings are particularly revealing. The proposition that measures must be unidimensional receives an endorsement probability of only 0.05. The proposition that multiplication requires a ratio measure also receives an endorsement probability of 0.05. Equally low endorsement is observed for the proposition that meeting the axioms of representational measurement is required before arithmetic can be undertaken. These are not minor methodological issues. They are foundational requirements for quantitative inference. Their rejection indicates that the measurement question occupies little place within the HUI framework.

The results become even more striking when attention turns to latent attributes. The HUI is explicitly concerned with characteristics such as health status, functioning and quality of life. These are not directly observable attributes. They are latent constructs inferred from observations and classifications. Yet the propositions that there are only two relevant forms of measurement, linear ratio measurement for manifest attributes and Rasch logit ratio measurement for latent attributes, that subjective responses can only be transformed into interval measurement through Rasch rules, and that Rasch measurement provides the basis for assessing therapy impact in latent traits all receive endorsement probabilities of only 0.05. In effect, the only established framework for transforming subjective observations into lawful measures is absent from the HUI knowledge base.

This finding is important because it reveals what the interrogation is signaling. The issue is not whether the HUI produces numerical outputs. The issue is whether those outputs have been shown to possess measurement properties. The interrogation suggests that this question is never seriously addressed. The existence of a scoring algorithm is treated as evidence of measurement. The generation of a numerical value is assumed to confer quantitative meaning. Yet arithmetic cannot create measurement. Numbers may be produced without establishing that measurement has occurred.

The proposition that measurement precedes arithmetic receives an endorsement probability of only 0.10. This single result may be the most revealing finding in the entire interrogation. It suggests that the HUI framework does not regard measurement as a prerequisite for numerical

manipulation. Instead, arithmetic operations appear to be accepted as self-justifying. Once a numerical value is generated, it is treated as though it possesses the properties necessary for further arithmetic operations.

This is why measurement inversion should be regarded as a diagnostic signal. The signal does not in itself demonstrate arithmetic chaos. Rather, it identifies the conditions under which arithmetic chaos becomes possible. If measurement is no longer required before arithmetic, there is no principled basis for determining whether multiplication, division, aggregation or comparison are admissible. The distinction between lawful and unlawful arithmetic operations disappears. Numerical outputs can be manipulated without regard to their measurement status.

The significance of this signal extends beyond the HUI. The same pattern has been observed in interrogations of HTA agencies, academic research centers, professional organizations and journals in multiple countries including Canada. The HUI therefore provides not merely an example of measurement inversion but one of its most influential expressions. Because HUI utilities subsequently become inputs to QALYs, simulation models and population health assessments, any failure to establish lawful measurement at the utility stage necessarily propagates throughout the entire analytical framework.

The interrogation results therefore point to a simple but profound conclusion. The HUI does not merely exhibit measurement inversion. It institutionalizes it. The numerical outputs generated by the utility algorithm are accepted as though their measurement status had been demonstrated, when the interrogation suggests precisely the opposite. Measurement inversion is therefore the signal. The next question is what follows when arithmetic is built upon foundations whose measurement properties have never been established. The answer is arithmetic chaos.

FROM UTILITY CONSTRUCTION TO ARITHMETIC CHAOS

The significance of measurement inversion becomes fully apparent when the construction of the HUI utility score is examined in detail. The interrogation results indicate that the requirements of representational measurement are largely absent from the HUI knowledge base. The question is therefore straightforward: what happens when arithmetic proceeds without first establishing measurement? The answer is arithmetic chaos.

The process begins with health-state descriptions. The HUI classifies health according to multiple attributes, including vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain. Each attribute is represented by a number of levels intended to describe varying degrees of impairment or functioning. The result is a multidimensional health-state classification system capable of describing a vast number of possible health profiles.

At this stage there is no measurement. The health-state descriptions are classifications. They provide a means of distinguishing one health profile from another, but they do not establish quantitative properties. The categories are descriptive rather than metrical. They identify differences but not magnitudes. Nevertheless, the HUI framework proceeds to the next stage.

Health states are valued through preference elicitation exercises, principally the standard gamble technique. Respondents are asked to express preferences among alternative health states under conditions of uncertainty. The resulting values are intended to represent the relative desirability of different health outcomes. Importantly, these values reflect preferences rather than quantities. They indicate rankings and relative judgments. There is no demonstration that they possess constant units, ratio properties or any of the characteristics required for quantitative measurement.

At this point the process should stop. Preferences are not measures. A ranking of health states cannot, by itself, support arithmetic operations. Yet the HUI framework proceeds to transform these preference structures into utility values through the application of a scoring algorithm.

This is the critical transition. Preference scores are combined through a multiplicative function to generate a utility value. The utility is then reported on a conventional scale anchored at 1.00 for perfect health and 0.00 for death, with negative values permitted for states considered worse than death. The resulting number is presented as though it were a measure of health-related quality of life.

The difficulty is obvious. Arithmetic cannot create measurement. A multiplicative algorithm applied to entities that lack demonstrated measurement properties cannot generate a lawful measure. The scoring formula may be mathematically sophisticated, but sophistication is not measurement. The utility score is simply the numerical output of the algorithm. Its existence does not demonstrate that the resulting value possesses ratio properties.

The use of multiplication is particularly problematic. Multiplication is not a neutral operation. It requires that the entities entering the calculation possess appropriate measurement properties. This requirement is recognized throughout the physical sciences. Multiplication of quantities is admissible only when the underlying scales support the operation. The interrogation results demonstrate almost no endorsement for this proposition. The HUI framework therefore applies multiplication without first establishing that the preference values being multiplied possess ratio characteristics.

The consequences are immediate. The utility value inherits the limitations of the underlying preference structure. If the inputs are ordinal, the output cannot become a ratio measure merely because a multiplicative function has been applied. Arithmetic operations do not transform rankings into quantities. They simply produce new numbers. The utility score may appear precise, but precision does not establish measurement.

The allowance of negative utility values introduces an additional difficulty. Ratio measures require a meaningful non-arbitrary zero and do not permit negative values. Yet HUI utilities can fall below zero for health states deemed worse than death. This feature alone creates a profound inconsistency with the interpretation of utilities as ratio measures. A scale that permits negative values cannot simultaneously satisfy the requirements of ratio measurement.

The creation of the utility score is therefore the first manifestation of arithmetic chaos. Numerical operations are undertaken without demonstrating that the entities involved possess lawful

measurement properties. The resulting utility is then treated as though it were a quantitative measure of health status.

The process does not end there. The utility becomes an input to QALY construction. Utility values are combined with time to estimate quality-adjusted life years. This operation is often presented as self-evidently legitimate. Yet the legitimacy of the QALY depends entirely upon the measurement status of the utility. If the utility is not a ratio measure, multiplication by time cannot create a lawful quantitative outcome. The QALY simply inherits the deficiencies of the utility structure from which it is derived.

The utility then enters simulation models. Costs, utilities, transition probabilities and assumptions are combined to generate estimates of long-term therapy impact. Each layer of modelling adds further numerical complexity. Yet none of these operations addresses the original measurement problem. The simulation model merely extends the chain of arithmetic. Numerical sophistication increases while measurement remains absent.

This is the defining characteristic of arithmetic chaos. The framework becomes increasingly elaborate while the measurement status of the entities involved remains unresolved. Health-state descriptions become preference scores. Preference scores become utilities. Utilities become QALYs. QALYs become simulation outputs. Simulation outputs become cost-effectiveness claims. At every stage arithmetic proceeds, as though measurement had already been established.

The remarkable feature of the HUI is not that it produces utility scores. It is that those utility scores became one of the most influential inputs to modern HTA. Through the HUI, arithmetic chaos was not confined to a single instrument. It became embedded in QALY construction, health-adjusted life expectancy, population health assessment and reference-case simulation modelling. The interrogation results signal the source of the problem. Examination of the utility construction process reveals its consequences. Measurement inversion is the signal; arithmetic chaos is the result.

STATISTICS CANADA, QALYS AND THE ILLUSION OF MEASUREMENT

The influence of the HUI extends far beyond the construction of utility values. Its significance lies in the fact that these utilities have been incorporated into some of the most widely used measures of population health and therapy assessment in Canada. Through Statistics Canada, health-adjusted life expectancy calculations, economic evaluations and reference-case simulation models, HUI utilities have become embedded in the infrastructure of health policy. The consequence is that any deficiency in the measurement properties of the utility score propagates throughout a much larger analytical framework.

Statistics Canada provides an important illustration. HUI scores are routinely reported as indicators of population health status. Average scores are calculated for demographic groups, disease categories and population subgroups. Comparisons are made between age groups, income groups and regions. Changes in average HUI scores are interpreted as evidence of improvements or

declines in health status. The implicit assumption is that the utility score possesses quantitative meaning and that differences between scores can be interpreted as differences in health.

Yet the interrogation results and the analysis presented in this paper suggest a very different interpretation. The HUI utility is not a demonstrated measure but the numerical output of a preference-based scoring algorithm. The crucial question therefore remains unanswered: what exactly does a difference between two HUI scores represent? If one population reports an average HUI score of 0.95 and another reports a score of 0.89, what is the unit of difference? What property of health is being measured? How can a multidimensional health-state classification system be transformed into a single quantitative measure through arithmetic operations alone?

These questions become even more important when HUI utilities are used to calculate health-adjusted life expectancy. Life expectancy is a lawful ratio measure. Time possesses a true zero and supports the arithmetic operations required for quantitative analysis. The adjustment applied through the HUI utility is intended to convert life expectancy into a measure that reflects both quantity and quality of life. Yet if the utility lacks ratio properties, the resulting calculation becomes problematic. The operation may generate a numerical output, but the measurement status of that output remains unresolved.

The same issue arises in QALY construction. HUI utilities are frequently employed as quality-adjustment weights in the estimation of quality-adjusted life years. The resulting QALYs are then used to compare therapies, support reimbursement decisions and populate simulation models. The arithmetic appears straightforward. Utility is combined with time to generate a common metric. The difficulty is that the legitimacy of the operation depends entirely upon the measurement properties of the utility. If the utility is not a ratio measure, the resulting QALY cannot acquire ratio properties simply because multiplication has occurred.

The implications extend further still. QALYs generated from HUI utilities become inputs to cost-effectiveness models and reference-case simulations. These models may project outcomes over decades or even lifetimes. They may generate estimates of incremental cost-effectiveness ratios, health benefits and resource implications. Yet every output produced by the model inherits the properties of its inputs. If the utility lacks lawful measurement properties, no amount of modelling sophistication can resolve the problem. The model simply propagates the deficiency.

This is the broader significance of arithmetic chaos. The issue is not confined to the HUI itself. It extends through every framework that depends upon HUI utilities. Statistics Canada, health-adjusted life expectancy calculations, QALY estimation and simulation modelling all assume that the utility score possesses measurement properties sufficient to justify arithmetic manipulation. The interrogation results suggest otherwise. Measurement inversion at the level of utility construction becomes arithmetic chaos at the level of policy analysis.

The remarkable success of the HUI therefore presents a paradox. The instrument became one of the most influential utility systems in the world. Its outputs entered population health reporting, economic evaluation and HTA decision making. Yet the interrogation results suggest that the measurement foundations required to support these applications were never established. The result

is an illusion of measurement. Numerical outputs are routinely interpreted as though they were lawful measures when the evidence indicates that they are not.

The HUI is therefore more than a utility instrument. It is a case study in how false measurement inversion can propagate through an entire analytical framework. The consequences extend from utility construction to population health assessment, from QALYs to simulation models and ultimately to reimbursement decisions. The signal identified by the interrogation thus becomes visible in practice. Arithmetic has displaced measurement, and the resulting numerical structures have acquired an authority that their measurement foundations cannot support.

CONCLUSION

The Health Utilities Index occupies a distinguished place in the history of health technology assessment. For more than four decades it has been employed to generate utility values for QALYs, health-adjusted life expectancy, economic evaluations and simulation models. Through its adoption by researchers, policy makers and government agencies, the HUI became one of the principal foundations of utility-based assessment. Its influence on the development of modern HTA is difficult to overstate.

Yet the interrogation results and the analysis presented in this paper point to a troubling conclusion. The issue is not the sophistication of the HUI, the quality of its valuation studies or the complexity of its scoring algorithm. The issue is that the measurement question was never resolved. Foundational propositions concerning unidimensionality, ratio measurement, representational measurement and Rasch measurement receive virtually no endorsement within the HUI knowledge base. At the same time, the framework continues to generate utility values that are subsequently treated as though they possess quantitative properties sufficient to support arithmetic operations.

This is the essence of measurement inversion. Measurement no longer precedes arithmetic. Instead, arithmetic becomes the mechanism through which measurement is assumed to emerge. Health-state descriptions are transformed into preference scores, preference scores into utilities, utilities into QALYs and QALYs into simulation outputs. At no stage is there a demonstration that lawful measurement has occurred. The result is arithmetic chaos: a sequence of increasingly sophisticated numerical operations undertaken without establishing the measurement status of the entities involved.

The consequences extend far beyond the HUI itself. Utility values generated by the HUI continue to inform population health reporting, health-adjusted life expectancy calculations, cost-utility analyses and reimbursement decisions. Through these applications, the assumptions embedded in utility construction have become embedded in the wider HTA framework. The HUI therefore serves as a powerful illustration of how measurement inversion can propagate throughout an entire knowledge system. More importantly, does the lack of understanding, even awareness, of scales of measure and representational measurement represent a deep seated and more distributed measurement malaise in Canadian HTA? The interrogation results from other targeted Canadian HTA knowledge bases suggests that it does.

The significance of the HUI is not that it failed after implementation. Rather, the evidence suggests that the framework was never capable of achieving its stated objective. A utility score derived from multidimensional health-state descriptions and preference valuations cannot acquire lawful measurement properties through arithmetic manipulation alone. The problem was present from the outset. In this sense, the HUI was dead at liftoff.

The broader lesson is clear. Numerical outputs should not be confused with measures. The future of HTA does not lie in refining utility algorithms, recalibrating preference weights or constructing ever more sophisticated simulation models. It lies in restoring the primacy of measurement. For manifest attributes, this requires empirically evaluable linear ratio measures. For latent attributes, it requires Rasch logit ratio measures. Only through lawful measurement can HTA support credible, evaluable and falsifiable claims regarding therapy impact. Measurement inversion is the signal. Arithmetic chaos is the consequence. Closure follows. The challenge now is reconstruction.

Canada occupies a distinctive place in the history of health technology assessment. Through the development of the Health Utilities Index (HUI), the growth of major academic research centres and the influence of national organizations such as CADTH and INESSS, Canada has played a significant role in shaping the international HTA landscape. Canadian researchers helped establish utility-based assessment as a central component of economic evaluation and contributed to the widespread adoption of QALYs, health-adjusted life expectancy and reference-case simulation models. As a consequence, Canada provides an important case study for examining a fundamental question: does contemporary HTA in Canada satisfy the requirements of lawful measurement?

The evidence presented throughout this section suggests that the answer is no. Interrogations of Canadian HTA agencies, academic research centers and the HUI knowledge base reveal a consistent pattern of measurement inversion. Foundational propositions concerning representational measurement, ratio measurement, dimensional homogeneity and Rasch measurement receive little support, while utilities, QALYs and simulation models continue to be strongly endorsed. The result is a framework in which arithmetic routinely precedes measurement. Numerical constructs are accepted as though they were lawful measures without first demonstrating that the attributes involved satisfy the conditions necessary for quantitative inference.

The implications are particularly important because Canada was not merely a consumer of the reference-case paradigm; it helped create some of its most influential components. The HUI, in particular, became one of the principal sources of utility values for QALYs and population health assessment. Yet examination of the HUI scoring framework suggests that health-state descriptions, preference valuations and multiplicative utility algorithms are combined through a sequence of arithmetic operations whose measurement legitimacy is never established. The result is arithmetic chaos. One inadmissible operation builds upon another until the final outputs emerge as utilities, QALYs, health-adjusted life expectancy estimates and cost-effectiveness claims. These outputs are presented as measures of therapy impact despite the absence of lawful measurement foundations.

The challenge facing Canada is the same challenge confronting HTA internationally. The issue is not whether Canadian institutions will continue to function or whether utility-based analyses will

continue to be published. The issue is closure. Closure occurs when a framework can no longer justify its claims according to the standards of quantitative science. The purpose of this section is to examine the Canadian evidence in detail and to consider the implications for reconstruction. The future of HTA in Canada lies not in further refinement of utilities, QALYs and simulation models, but in a return to lawful measurement, empirical evaluation and falsification as the basis for claims regarding therapy impact.

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

[i] Furlong W, Feeny D, Torrance G, Barr R. The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies*, Annals of Medicine*, 2001;33:375-84

[ii] Canada: Deconstructing the Epistemic Knowledge Base of the Health Utilities Index of Instruments. Logit Working Paper No 60. https://maimonresearch.com/logit-working-paper-no-60/

[iii] Stevens S. On the theory of the scales of measurement. *Science*, 1946; 103(2684), 677–680.

[iv] 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

[v] Krantz D, Luce R, Suppes P, Tversky A. *Foundations of Measurement, Volume I: Additive and Polynomial Representations*. New York, NY: Academic Press, 1971