# MAIMON RESEARCH LLC

# ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

# REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

# NEW ZEALAND: : 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 1812 JUNE 2026**

[www.maimonresearch.com](www.maimonresearch.com)

**Tucson AZ**

## ABSTRACT

*New Zealand has long been recognized for its pragmatic approach to healthcare decision making. Through PHARMAC and a small number of influential university research groups, the country has developed a reputation for disciplined resource allocation and evidence-based policy. Yet the foundations of health technology assessment (HTA) in New Zealand have rarely been examined from the perspective of measurement theory. This paper evaluates the extent to which PHARMAC and the Universities of Auckland and Otago recognize and apply the principles of representational measurement that are required to support quantitative claims regarding therapy impact.*

*Interrogations of HTA knowledge bases reveal a consistent pattern of measurement inversion. Across all three institutions, propositions that are false within measurement theory receive strong endorsement, including claims that ratio measures can have negative values, that the QALY is a ratio measure, that subjective instrument scores can be transformed into ratio measures through aggregation, and that reference-case simulations generate falsifiable claims. At the same time, propositions that are fundamental to quantitative science receive little support, including the requirements of unidimensionality, dimensional homogeneity, ratio measurement and the principle that measurement must precede arithmetic.*

*The results are reinforced by a second interrogation focused on Rasch measurement. Endorsement probabilities for core Rasch propositions are consistently at floor levels, indicating an almost complete absence of recognition of the only framework capable of transforming observations of latent attributes into lawful measures. The findings suggest that New Zealand HTA is characterized not only by endorsement of inadmissible arithmetic but also by the absence of the measurement framework necessary to replace it.*

*The paper argues that the resulting framework generates arithmetic chaos. Health-state descriptions are transformed into utilities, utilities into QALYs and QALYs into simulation-based claims for cost-effectiveness through a succession of arithmetic operations whose measurement foundations are never established. Against this background, a reconstruction of HTA is proposed based upon a simple principle: there are only two lawful forms of measurement for therapy assessment. Manifest attributes require linear ratio measures, while latent attributes require Rasch logit ratio measures. Together these provide the only basis for empirically evaluable, replicable and falsifiable claims regarding therapy impact. The future of HTA in New Zealand lies not in refining utilities, QALYs and simulation models but in restoring measurement to its central role in scientific evaluation.*

## INTRODUCTION

For more than four decades health technology assessment (HTA) has operated under a fundamental misconception. The prevailing assumption has been that the principal challenge in evaluating therapies is to generate numbers. Once numerical values are available, whether utilities, QALYs, cost-effectiveness ratios or simulation outputs, they are assumed to provide a quantitative basis for decision making. The result has been an extraordinary expansion in analytical complexity. Health-state descriptions have been transformed into utilities, utilities into QALYs and QALYs into cost-effectiveness claims supported by increasingly sophisticated simulation models. Yet throughout this process a more fundamental question has largely been ignored: what is being measured?

New Zealand provides a particularly interesting case study. Unlike larger jurisdictions with extensive HTA bureaucracies, New Zealand has long been recognized for its pragmatic approach to healthcare decision making. Through PHARMAC and a small number of influential university research groups, the country developed a reputation for disciplined resource allocation and evidence-based policy. Yet in HTA, New Zealand followed the same path as many larger health systems. Utilities, QALYs and simulation models became accepted as the language of value assessment, while the measurement foundations necessary to support these numerical constructions remained largely unexplored.

The irony is that New Zealand was well positioned to avoid this outcome. The distinction between manifest and latent attributes provides a complete framework for therapy assessment. Manifest attributes require linear ratio measures. Latent attributes require Rasch logit ratio measures. Together these provide the only lawful basis for quantitative claims regarding therapy impact. There was no need to construct utilities from health-state descriptions, no need to create QALYs and no need to populate simulation models with entities whose measurement status was unknown. Nevertheless, New Zealand adopted the reference-case paradigm and became increasingly committed to a framework in which arithmetic displaced measurement.

Recent interrogations of PHARMAC and two New Zealand academic HTA knowledge bases reveal the consequences [1] [2] [3] . Utilities, QALYs and simulation modelling receive strong endorsement, while representational measurement, dimensional homogeneity and Rasch measurement receive little recognition. These findings point to measurement inversion, where arithmetic precedes measurement, and to arithmetic chaos, where one inadmissible numerical operation builds upon another. The result is a framework that generates increasingly sophisticated numerical outputs while lacking the measurement foundations required to support them.

The purpose of this paper is therefore straightforward. It is to demonstrate that, for the assessment of therapy impact, there are only two lawful measures. Recognizing this fact not only explains the emergence of measurement inversion and arithmetic chaos within New Zealand HTA, but also provides the foundation for its reconstruction.

The irony is that this path was entirely unnecessary. Once therapy outcomes are recognized as either manifest or latent attributes, the requirements for lawful measurement become straightforward. Manifest attributes require linear ratio measures. Latent attributes require Rasch logit ratio measures. Together these provide a complete framework for the assessment of therapy impact. There was no need to construct utilities from health-state descriptions, no need to create QALYs and no need to populate simulation models with entities whose measurement status was unknown

The neglect of this question has had profound consequences. Across the physical sciences, engineering and the natural sciences, arithmetic follows measurement. Quantitative operations are undertaken only after the measurement properties of the entities involved have been established.

In HTA, however, the reverse has become commonplace. Numerical operations are routinely applied to constructs whose measurement status is either unknown or assumed. This phenomenon, described throughout this series of papers as measurement inversion, has created the conditions for arithmetic chaos. Numbers are multiplied, divided, aggregated and incorporated into simulation models without first demonstrating that the underlying attributes possess the properties necessary to support those operations.

The source of much of this confusion lies in a failure to distinguish between two fundamentally different classes of attributes. Some attributes are manifest. They are directly observable and can be measured through straightforward counting or observation. Hospital admissions, hospital days, emergency department visits, physician encounters, treatment persistence and resource utilization are examples. These attributes support linear ratio measurement with meaningful zeros and interpretable magnitudes. Other attributes are latent. Pain, fatigue, depression, physical functioning, symptom burden, treatment satisfaction and need fulfilment cannot be observed directly. Their existence must be inferred from observations. These attributes require a completely different measurement framework.

Once this distinction is recognized, the path forward becomes remarkably clear. There are only two lawful forms of measurement available for the assessment of therapy impact. Manifest attributes require linear ratio measures. Latent attributes require Rasch logit ratio measures. Together these provide the only basis for constructing empirically evaluable, replicable and falsifiable claims regarding treatment outcomes. No other measurement framework is required. More importantly, no other measurement framework is capable of supporting the arithmetic operations necessary for quantitative science.

The significance of this conclusion extends far beyond a technical debate over measurement theory. It challenges the foundations of the contemporary reference-case paradigm. Utilities, QALYs and simulation models are all dependent upon assumptions concerning measurement that are rarely examined and seldom justified. By contrast, a framework built upon linear ratio measures and Rasch logit ratio measures restores the principle that measurement must precede arithmetic. It provides a direct route from observable evidence to evaluable claims without the need for utility construction, imaginary future worlds or non-falsifiable simulations.

The purpose of this paper is therefore straightforward. It is to demonstrate that, for the assessment of therapy impact, there are only two lawful measures. Recognizing this fact not only resolves the confusion that has characterized modern HTA but also provides the foundation for its reconstruction. Measurement inversion and arithmetic chaos are not inevitable. They are the consequence of abandoning the discipline of measurement. The future of HTA depends upon restoring it.

## 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.

## 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.

HTA ignores this requirement. Utilities, preference scores and questionnaire responses are routinely manipulated as though they possess ratio properties. Yet the measurement status of these constructs is rarely established. Numbers are generated and arithmetic operations applied without first demonstrating the scale to which the numbers belong. The result is a framework in which numerical manipulation substitutes for measurement.

The consequences become particularly important when considering utility construction. Both standard gamble and time trade-off techniques generate numerical values. The critical question is not whether numbers are produced but what scale of measurement those numbers represent. If they are ordinal, then 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.

The significance of this oversight cannot be overstated. Utilities, QALYs and cost-effectiveness ratios all depend upon arithmetic operations that require ratio measurement. If the underlying values fail to satisfy ratio requirements, then the resulting constructions lack quantitative meaning. The arithmetic may be technically correct, but it is not admissible. The distinction between computation and measurement is therefore critical. Numbers can always be manipulated. The question is whether the resulting outputs possess any scientific significance.

The scales of measurement thus provide the first diagnostic test for closure in HTA. A discipline that cannot distinguish between nominal, ordinal, interval and ratio scales has lost sight of the conditions necessary for quantitative inquiry. Once that distinction disappears, arithmetic becomes detached from measurement and numerical outputs become ends in themselves. The result is measurement inversion: arithmetic preceding measurement.

## 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.

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 implications for 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. 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. Once that failure occurs, closure follows. The entire sequence of utilities, QALYs and reference-case models rests upon a foundation that never satisfied the conditions necessary for quantitative measurement.

## NEW ZEALAND: MEASUREMENT INVERSION IN PHARMAC AND THE UNIVERSITIES OF AUCKLAND AND OTAGO

The results presented in Table 1 provide a striking picture of the intellectual foundations of health technology assessment in New Zealand. Across PHARMAC, the University of Auckland and the University of Otago there is a consistent pattern of endorsement for propositions that are false within the framework of representational measurement, coupled with a consistent rejection of propositions that are true. The significance of this finding extends beyond disagreement over technical issues. The results point to a systematic failure to recognize the conditions necessary for quantitative measurement and, consequently, to the presence of measurement inversion throughout New Zealand HTA.

**TABLE 1**

**MEASUREMENT INVERSION: PHARMAC, AUCKLAND, OTAGO**

| CANONICAL STATEMENT | CATEGORICAL PROBABILITY | | |
|---|---|---|---|
| | PHARMAC | AUCKLAND | OTAGO |
| FALSE STATEMENTS | | | |
| 1, Ratio measures can have negative values | 0.95 | 0.85 | 0.85 |
| 2.The QALY is a ratio measure | 0.95 | 0.90 | 0.95 |
| 3, Summations of subjective instrument responses are ratio measures | 0.95 | 0.85 | 0.90 |
| 4. The QALY is a dimensionally homogeneous measure | 0.96 | 0.90 | 0.35 |
| 5. Reference case simulations generate falsifiable claims | 0.95 | 0.85 | 0.90 |
| TRUE STATEMENTS., | | | |
| 6. Measures must be unidimensional | 0.15 | 0.20 | 0.15 |
| 7. Multiplication requires a ratio measure | 0.15 | 0.10 | 0.10 |
| 8. There are only two classes of measurement: linear ratio and Rasch logit ratio | 0.05 | 0.05 | 0.05 |
| 9.Measurement precedes arithmetic | 0.15 | 0.15 | 0.10 |
| 10. The outcome of interest for latent traits is the possession of that trait | 0.05 | 0.10 | 0.10 |

The first observation is the remarkable consistency of the results. The proposition that ratio measures can have negative values receives endorsement probabilities between 0.85 and 0.95 across all three institutions. Similarly, the proposition that the QALY is a ratio measure receives endorsement probabilities of 0.90 to 0.95. The claim that summations of subjective instrument responses create ratio measures is endorsed at probabilities between 0.85 and 0.95. These are not marginal findings. They indicate strong support for propositions that directly contradict the requirements of measurement theory.

The same pattern is evident in attitudes towards the QALY and the reference case. PHARMAC assigns a probability of 0.96 to the proposition that the QALY is dimensionally homogeneous, while Auckland assigns 0.90. Although Otago is less committed to this proposition, with a probability of 0.35, it nevertheless endorses the QALY as a ratio measure at 0.95. Likewise, the proposition that reference-case simulations generate falsifiable claims receives endorsement probabilities of 0.85 to 0.95. These findings suggest that simulation modelling and QALY construction are accepted as legitimate quantitative activities without regard to the measurement requirements necessary to support them.

Equally revealing is the treatment of propositions that are true. The requirement that measures be unidimensional receives endorsement probabilities of only 0.15 to 0.20. The proposition that multiplication requires a ratio measure receives endorsement probabilities between 0.10 and 0.15. The proposition that measurement precedes arithmetic receives probabilities between 0.10 and 0.15. Perhaps most striking of all, the proposition 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—receives a probability of only 0.05 in all three institutions.

These findings are not isolated observations. Taken together they reveal a coherent pattern. The requirements for measurement receive almost no support, while the numerical constructs built upon those requirements continue to be endorsed. This is the defining characteristic of measurement inversion. In every quantitative science, measurement is established before arithmetic is undertaken. In New Zealand HTA, the evidence suggests the opposite sequence. Arithmetic is accepted as legitimate while the measurement foundations necessary to justify it are either ignored or rejected.

The implications are substantial. Consider the endorsement of the proposition that multiplication does not require ratio measures. This single belief undermines the entire logic of utility construction and QALY estimation. If multiplication can be applied to ordinal preference structures without establishing ratio properties, then any numerical output can be treated as though it were a measure. The distinction between lawful and unlawful arithmetic disappears. The result is a framework in which numerical manipulation becomes self-justifying.

This helps explain the strong endorsement of the QALY. The QALY depends upon the multiplication of time by a utility value. Yet if the utility lacks ratio properties, the operation is inadmissible. The resulting QALY cannot acquire measurement properties simply because multiplication has occurred. Nevertheless, PHARMAC, Auckland and Otago all strongly endorse the proposition that the QALY is a ratio measure. The implication is that the arithmetic outcome is accepted while the measurement question remains unasked.

The same problem is evident in attitudes towards latent attributes. The proposition that the outcome of interest for latent traits is possession of that trait receives endorsement probabilities of only 0.05 to 0.10. This indicates little awareness of the central objective of Rasch measurement. If latent attributes such as pain, fatigue, functioning or quality of life are not understood as attributes whose possession must be measured, then it is hardly surprising that preference scores, utility algorithms and questionnaire summations are accepted as substitutes for measurement.

Viewed collectively, the results point to a knowledge base that has become detached from the foundations of measurement science. The issue is not a lack of statistical sophistication. New Zealand HTA is fully capable of constructing utilities, QALYs and simulation models. The issue is that these numerical activities are undertaken without recognition of the measurement conditions necessary to justify them. Arithmetic has replaced measurement as the organizing principle of evaluation.

The consequence is arithmetic chaos. Health-state descriptions are transformed into utility values. Utility values become QALYs. QALYs become inputs to simulation models and cost-effectiveness claims. At each stage one arithmetic operation is built upon another, despite the absence of evidence that the underlying entities possess lawful measurement properties. The resulting outputs may appear precise, but precision is not measurement.

The results for PHARMAC, Auckland and Otago therefore point to a common conclusion. New Zealand has institutionalized a framework characterized by measurement inversion and sustained by arithmetic chaos. The problem is not unique to New Zealand; similar patterns have been observed internationally. Nevertheless, the consistency of the findings suggests that the issue is deeply embedded within New Zealand HTA education, research and policy. The challenge is not to refine utilities, QALYs or simulation models. The challenge is to reconstruct HTA around lawful measurement. Until that occurs, the quantitative claims generated by the current framework will remain vulnerable to the charge that they are numerical constructions without demonstrated measurement foundations.

Given the importance of the Rasch model for providing the necessary and sufficient conditions for transforming ordinal observations to ratio measurement and the Rasch logit ratio scale for latent attribute possession, it is of interest to see the extent to which Rasch resonates for HTA in New Zealand.

## LINEAR RATIO MEASURES AND MANIFEST ATTRIBUTES

The first lawful form of measurement relevant to therapy assessment is the linear ratio measure. Although the concept is straightforward, it has been largely overshadowed within contemporary HTA by an emphasis on utilities, QALYs and simulation models. Yet it is linear ratio measurement that provides the foundation for quantitative science. Without ratio measurement there can be no meaningful multiplication, division, comparison of magnitudes or quantitative inference. The assessment of therapy impact therefore begins not with utilities or health-state descriptions but with the identification of attributes that can be measured directly.

Manifest attributes are directly observable characteristics. Their existence does not have to be inferred from responses to questionnaires, preference exercises or health-state classifications. They can be counted, observed or recorded. Examples include hospital admissions, hospital days, emergency department visits, physician encounters, ICU hours, treatment persistence, medication possession, treatment switching and resource utilization. These are not hypothetical constructs. They are observable events that occur within the real world of patient care.

The defining feature of a manifest attribute is that it supports a meaningful zero. A patient with zero hospital admissions has experienced no admissions. A patient with zero emergency department visits has not attended an emergency department. Similarly, a patient with ten hospital days has experienced twice as many hospital days as a patient with five hospital days. The magnitudes are interpretable because the attribute possesses ratio properties. Arithmetic operations involving addition, subtraction, multiplication and division are therefore admissible.

This may appear self-evident, but its implications for HTA are profound. Once an attribute is measured on a linear ratio scale, claims regarding therapy impact become immediately evaluable. A manufacturer can claim that a therapy reduces hospital admissions by a specified amount over a defined period. The claim can be tested empirically. Data can be collected, analyzed and replicated. Independent investigators can attempt to confirm or refute the result. The claim is therefore falsifiable. This is precisely how quantitative science is intended to operate.

The contrast with the reference-case framework is striking. Rather than beginning with directly observable attributes, the reference case focuses on health-state descriptions, utility construction and simulation models. The result is a framework in which arithmetic operations are applied to entities whose measurement status remains uncertain. Linear ratio measures avoid this problem entirely. The measurement properties of the attribute are established before arithmetic is undertaken. Measurement precedes arithmetic rather than the reverse.

Manifest attributes also provide a direct link between clinical outcomes and health system objectives. A reduction in hospital admissions, emergency department visits or treatment discontinuations has immediate practical significance. The outcome is understandable to clinicians, patients, payers and policy makers. Unlike utility scores, the measure does not require interpretation through preference algorithms or hypothetical valuations. The attribute itself is the outcome of interest.

Importantly, the use of linear ratio measures does not imply that all therapy impact can be reduced to observable events. Many important treatment outcomes involve latent attributes such as pain, fatigue, functioning and quality of life. These cannot be measured directly and require a different measurement framework. The existence of latent attributes, however, does not diminish the importance of manifest attributes. Rather, it reinforces the need to distinguish clearly between the two.

The reconstruction of HTA therefore begins with a recognition that manifest attributes require linear ratio measurement. This is not a matter of methodological preference but a consequence of the requirements of representational measurement. When the attribute is directly observable, the objective is to construct a measure that preserves meaningful magnitudes, supports admissible arithmetic operations and allows empirical evaluation. Linear ratio scales satisfy these requirements. They provide a lawful basis for quantitative claims regarding therapy impact and establish the first of the two measurement frameworks necessary for a scientifically credible HTA.

The importance of this conclusion cannot be overstated. Once manifest attributes are recognized as requiring linear ratio measurement, much of the complexity that characterizes contemporary HTA becomes unnecessary. Claims can be framed directly in terms of observable outcomes, evaluated through explicit protocols and subjected to replication and falsification. The result is a return to the principles that govern quantitative inquiry in every other scientific discipline. For manifest attributes, the path is clear: linear ratio measurement provides the only lawful foundation for assessing therapy impact.

## RASCH LOGIT RATIO MEASURES AND LATENT ATTRIBUTES

While manifest attributes can be measured directly through observation and counting, many of the outcomes of greatest interest in therapy assessment are latent attributes. These include pain, fatigue, depression, anxiety, physical functioning, symptom burden, treatment satisfaction and need fulfilment. Such attributes cannot be observed directly. A patient cannot be examined and assigned a direct measure of pain in the same way that a hospital admission can be counted or a length of stay recorded. The existence of the attribute must instead be inferred from observable indicators. This distinction is crucial because latent attributes require a fundamentally different measurement framework.

The challenge of latent measurement has long been misunderstood within HTA. The conventional approach has been to administer questionnaires, assign numerical scores to responses and aggregate those scores to create a summary measure. The resulting totals are often treated as though they were quantitative measures capable of supporting arithmetic operations and comparisons of therapy impact. Yet the assignment of numbers to questionnaire responses does not create measurement. The resulting scores remain ordinal structures. They provide rankings but do not establish equal intervals, meaningful units or ratio properties.

This is where Rasch measurement assumes central importance. Developed by Georg Rasch and subsequently formalized within the framework of representational measurement, the Rasch model provides the only established method for transforming observations of latent attributes into lawful measures [8]. The objective is not to generate scores but to construct a measurement system in which persons and items are located simultaneously on a common latent continuum. The resulting measure is expressed in logits, the natural logarithm of the odds ratio.

The importance of the Rasch framework lies in its strict measurement requirements. The model demands unidimensionality, local independence and invariance. Items must collectively represent a single underlying attribute. Responses must be independent once the latent trait is taken into account. Most importantly, the relationship between persons and items must be invariant across populations and contexts. These requirements are not optional. They are the conditions necessary for measurement.

The outcome of interest in Rasch measurement is not a questionnaire score. It is the possession of the latent trait itself. A patient's position on the latent continuum represents the extent to which the attribute is possessed relative to other persons and items within the measurement system. Therapy impact is therefore assessed through changes in possession of the latent trait rather than changes in arbitrary ordinal scores. This distinction separates Rasch measurement from conventional psychometric approaches, item response theory models and preference-based utility instruments.

The resulting Rasch logit scale possesses ratio properties through the odds structure underlying the model. Changes in logits can be interpreted quantitatively and transformed into odds ratios. The measurement framework therefore supports meaningful comparisons and provides a lawful basis for assessing therapy impact in latent attributes. Unlike summed ordinal scores, the Rasch measure is not an artifact of arbitrary scoring conventions. It is the outcome of a measurement model specifically designed to satisfy the requirements of representational measurement.

The implications for HTA are substantial. Many outcomes currently represented through utilities, quality-of-life instruments and patient-reported outcome measures are latent attributes. Yet the conventional approach treats questionnaire responses as though arithmetic operations alone can create measurement. Rasch measurement demonstrates that this is not the case. Measurement requires a formal measurement model, not a scoring algorithm. The objective is not to produce numbers but to construct lawful measures.

The reconstruction of HTA therefore requires recognition of a simple principle. Manifest attributes demand linear ratio measures. Latent attributes demand Rasch logit ratio measures. There is no third category. Once this distinction is accepted, much of the confusion surrounding patient-reported outcomes, quality-of-life instruments and utility construction disappears. The question is no longer how to manipulate ordinal scores through increasingly sophisticated algorithms. The question is whether the latent attribute has been measured according to the requirements of Rasch measurement.

For latent attributes, the conclusion is clear. Rasch logit ratio measurement provides the only lawful basis for assessing therapy impact. It therefore represents the second and final measurement framework required for a scientifically credible system of health technology assessment. Once combined with linear ratio measures for manifest attributes, the foundation exists for a reconstructed HTA grounded in lawful measurement, empirical evaluation and falsification rather than measurement inversion and arithmetic chaos.

## NEW ZEALAND: THE ABSENCE OF RASCH MEASUREMENT

If the findings for measurement inversion reveal the foundations of contemporary HTA in New Zealand, the Rasch interrogation results reveal an equally important absence: the virtual non-existence of Rasch measurement within the HTA knowledge bases of PHARMAC, the University of Auckland and the University of Otago (Table 2).

## TABLE 2

## ENDORSEMENT OF RASCH MEASUREMENT: PBAC, AUCKLAND AND OTAGO

| CANONICAL RASCH STATEMENT | CATEGORICAL PROBABILITY | | |
|---|---|---|---|
| | | | |
| 1. There are only two classes of measurement: linear ratio and Rasch logit ratio | 0.05 | 0.05 | 0.05 |

| | | | |
|---|---|---|---|
| 2. Transforming subjective responses to interval measurement is only possible with Rasch rules | 0.05 | 0.05 | 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 |
| 4. The outcome of interest for latent traits is the possession of that trait | 0.05 | 0.10 | 0.10 |
| 5. The Rasch rules for measurement are identical to the axioms of representational measurement | 0.05 | 0.05 | 0.05 |

The results are striking. Four of the five Rasch statements receive endorsement probabilities of only 0.05 across all three institutions. The remaining statement, that the outcome of interest for latent traits is the possession of that trait, receives probabilities of only 0.05 for PHARMAC and 0.10 for Auckland and Otago. These are effectively floor-level responses. They indicate not merely disagreement with Rasch measurement principles but an almost complete absence of recognition of their role in measurement theory and therapy assessment.

The importance of these findings should not be underestimated. Rasch measurement is not simply another psychometric technique competing with alternative approaches. Rather, it represents the only established framework capable of transforming observations of latent attributes into lawful measures while satisfying the requirements of representational measurement. Attributes such as pain, fatigue, symptom burden, physical functioning, treatment satisfaction and quality of life cannot be observed directly. They are latent traits. The central challenge is therefore not statistical analysis but measurement. Rasch provides the solution by creating a logit ratio scale that supports quantitative assessment of the possession of the latent trait.

Yet none of these principles appear to be recognized within the New Zealand HTA knowledge base. The proposition that there are only two lawful forms of measurement—linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes—receives an endorsement probability of only 0.05. The proposition that subjective responses can be transformed into interval and ratio measurement only through Rasch rules receives the same response. Most significantly, the proposition that the Rasch logit ratio scale is the only basis for assessing therapy impact in latent traits also receives a probability of 0.05 across all institutions.

The implications are profound. Without Rasch measurement, latent attributes are reduced to ordinal scores, preference valuations and utility algorithms. Questionnaire responses may be summed, transformed and manipulated mathematically, but the resulting outputs remain numerical constructs rather than measures. The absence of Rasch awareness therefore helps explain the simultaneous endorsement of utilities, QALYs and reference-case simulations observed in the broader interrogation results. If Rasch measurement is unknown or ignored, then there is no obvious alternative to the utility-based paradigm. Arithmetic fills the vacuum left by measurement.

Particularly revealing is the response to the statement that the outcome of interest for latent traits is the possession of that trait. This proposition lies at the heart of Rasch measurement. The objective is not the generation of scores but the measurement of the extent to which a person possesses the latent attribute. Yet endorsement probabilities remain close to zero. This suggests that latent attributes are viewed primarily through the lens of questionnaire scoring and utility construction rather than as entities requiring formal measurement.

The final statement is perhaps the most important. The proposition that Rasch rules are identical to the axioms of representational measurement receives an endorsement probability of only 0.05 across all three institutions. This finding points directly to the source of the problem. The issue is not merely unfamiliarity with Rasch methodology. It is unfamiliarity with the foundations of measurement itself. The same principles that govern quantitative measurement in the physical sciences govern the measurement of latent attributes. Rasch measurement is significant precisely because it operationalizes those principles.

Taken together, these findings reveal a remarkable paradox. New Zealand HTA displays strong confidence in utilities, QALYs and simulation modelling while showing almost no recognition of the only framework capable of providing lawful measurement for latent traits. The result is a system in which arithmetic dominates assessment while measurement remains absent. The interrogation therefore complements the broader findings on measurement inversion. It demonstrates that the problem is not simply endorsement of inadmissible arithmetic. It is also the absence of the one measurement framework capable of replacing it.

For New Zealand, the challenge is therefore not the refinement of utility algorithms or simulation models. It is the introduction of measurement into a field that has become dominated by arithmetic. Until Rasch measurement and the principles of representational measurement are recognized, HTA will remain vulnerable to the criticism that its quantitative claims rest not on measurement but on numerical construction. The interrogation results suggest that this challenge remains largely unrecognized within PHARMAC and the country's leading academic institutions.

## WHY THERE ARE NO OTHER MEASURES

The argument advanced in this paper is intentionally simple. For the assessment of therapy impact there are only two lawful forms of measurement: linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes. This conclusion raises an obvious question. What about the numerous measurement systems, utility instruments, patient-reported outcome measures and composite indices that currently dominate HTA? If these are not lawful measures, what are they?

The answer lies in a failure to distinguish between numbers and measures. The assignment of numbers to observations does not, in itself, create measurement. Numbers may be used as labels, rankings, classifications or scores without satisfying the requirements of representational measurement. The critical issue is not whether a numerical value exists but whether the numerical value preserves the properties of the attribute being measured and supports admissible arithmetic operations.

This distinction explains why so many contemporary HTA constructs fail as measures. Utility instruments such as the EQ-5D and the Health Utilities Index begin with health-state descriptions and preference valuations. The resulting utility scores are generated through scoring algorithms that combine attribute levels and preference weights. Yet the existence of an algorithm does not establish measurement. Health-state classifications are not measures; preference structures are ordinal and arithmetic operations applied to ordinal structures cannot create ratio measures. The final utility score is therefore a numerical construct rather than a lawful measure.

The same criticism applies to summed questionnaire scores. Instruments such as quality-of-life scales, symptom inventories and treatment satisfaction questionnaires frequently assign numbers to response categories and aggregate the results into a total score. This practice remains widespread despite a simple limitation: ordinal scores support ranking, not arithmetic. Summing ordinal responses does not transform them into interval or ratio measures. The resulting total score remains an ordinal construct irrespective of the complexity of the scoring procedure.

The QALY provides perhaps the most influential example. The QALY is often presented as a universal measure of health benefit capable of supporting comparisons across diseases and interventions. Yet the QALY depends entirely upon the utility values from which it is constructed. If the utility lacks ratio properties, multiplication by time cannot create a lawful measure. The resulting QALY is therefore not a ratio measure but the numerical consequence of inadmissible arithmetic. The same conclusion applies to the cost-effectiveness ratios and simulation outputs that subsequently depend upon QALYs as model inputs.

This is why the search for alternative measurement frameworks within HTA has been so unproductive. The problem is not that existing instruments require refinement. The problem is that they attempt to achieve measurement through arithmetic rather than through measurement itself. Utility algorithms, summed scores, preference weights and simulation models all share a common assumption: that numerical manipulation can compensate for the absence of lawful measurement. It cannot.

Once the distinction between manifest and latent attributes is recognized, the situation becomes much clearer. Manifest attributes require direct observation and linear ratio measurement. Latent attributes require Rasch measurement and logit ratio scales. Every outcome relevant to therapy assessment falls into one of these two categories. There is no intermediate class of attribute requiring a separate measurement framework. There is no need for utility construction, preference algorithms or composite indices. The apparent diversity of measurement approaches in HTA is largely an illusion created by the proliferation of scoring systems that generate numbers without generating measures.

This conclusion explains the emergence of measurement inversion and arithmetic chaos throughout contemporary HTA. The discipline has attempted to create quantitative claims without first establishing lawful measurement. The result has been a proliferation of numerical constructs that are routinely treated as measures despite lacking the properties required to support that interpretation. The solution is not further methodological complexity. The solution is to return to the two measurement frameworks that satisfy the requirements of representational measurement. Linear ratio measures and Rasch logit ratio measures are not simply preferable alternatives. They are the only lawful measures available for the assessment of therapy impact.

## THE END OF ARITHMETIC CHAOS

Arithmetic chaos is not an inevitable feature of health technology assessment. It is the consequence of abandoning a simple scientific principle: measurement must precede arithmetic. Once this principle is ignored, any numerical construct can be manipulated as though it were a measure. Health-state descriptions become utilities, utilities become QALYs, QALYs become simulation outputs and simulation outputs become cost-effectiveness claims. At every stage, arithmetic operations are undertaken without establishing whether the entities involved possess the measurement properties necessary to support those operations. The result is a framework in which numerical sophistication substitutes for measurement.

The consequences of this inversion have been profound. For more than four decades HTA has devoted enormous intellectual effort to the refinement of utilities, preference algorithms, simulation models and cost-effectiveness frameworks. Yet each of these developments rests upon the assumption that lawful measurement has already been achieved. The evidence suggests otherwise. The widespread acceptance of utilities and QALYs has obscured a more fundamental problem: the absence of measures capable of supporting the arithmetic operations imposed upon them.

Recognition that there are only two lawful measures immediately changes the discussion. Once therapy outcomes are classified as either manifest or latent attributes, the measurement requirements become clear. Manifest attributes require linear ratio measures. Latent attributes require Rasch logit ratio measures. The distinction eliminates the need for utility construction, preference weighting systems and composite indices. Therapy impact can be assessed directly through attributes that possess demonstrable measurement properties.

The implications for HTA are transformative. Formulary submissions no longer require hypothetical future worlds populated by utility scores and simulated QALYs. Instead, manufacturers present claims based upon measurable attributes supported by explicit protocols. These claims can be evaluated empirically, replicated by independent investigators and subjected to falsification. The focus shifts from the generation of imaginary estimates to the assessment of real-world evidence.

The same transformation applies to disease-area reviews and therapeutic-class evaluations. Rather than comparing therapies through cost-per-QALY estimates derived from simulation models, comparisons can be based upon measurable outcomes relevant to patients, clinicians and health systems. Hospital admissions, treatment persistence, resource utilization, symptom burden, physical functioning and need fulfilment can all be assessed through lawful measurement frameworks. The resulting evidence is transparent, evaluable and directly relevant to decision making.

Most importantly, the recognition of only two lawful measures restores the distinction between numbers and measures. Numerical outputs are no longer accepted simply because they are produced by sophisticated algorithms. Every claim must demonstrate the measurement properties of the attribute involved. Arithmetic becomes subordinate to measurement rather than a substitute for it.

This is why reconstruction is both possible and necessary. The problem confronting HTA is not a lack of analytical sophistication. The problem is that analytical sophistication has been allowed to replace measurement. Once the discipline returns to linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes, the foundations of arithmetic chaos disappear. Measurement inversion is reversed, lawful measurement is restored and HTA can once again function as a quantitative science. The path forward is therefore neither complex nor mysterious. It begins with a simple recognition: there are only two lawful measures, and everything else is arithmetic.

## IMPLICATIONS FOR THE RECONSTRUCTION OF HTA

Recognition that there are only two lawful measures has profound implications for the future of health technology assessment. The issue is not merely one of measurement theory. It affects every stage of therapy evaluation, from clinical development and formulary submissions to disease-area reviews, therapeutic-class assessments and professional education. Once the distinction between manifest and latent attributes is restored, reconstruction becomes a practical process rather than an abstract methodological debate.

The first implication concerns formulary submissions. The reference-case paradigm assumes that the central objective of a submission is the construction of a cost-effectiveness model supported by utilities, QALYs and simulation outputs. Reconstruction reverses this emphasis. The purpose of a submission becomes the presentation of empirically evaluable claims regarding therapy impact. These claims must be expressed either as linear ratio measures for manifest attributes or Rasch logit ratio measures for latent attributes. Each claim must be supported by an explicit protocol describing the target population, comparator, attribute definition, measurement framework, evaluation period, success criteria and replication strategy. The focus shifts from hypothetical future projections to measurable outcomes.

The second implication concerns clinical development. Manufacturers can no longer regard formulary submissions as a separate activity undertaken after regulatory approval has been secured. The development of attribute claims must begin during clinical development. Manifest and latent outcomes should be identified during Phase II and Phase III studies, with measurement strategies embedded in trial design. There can be no discontinuity between the evidence generated for regulatory approval and the evidence required for therapy assessment. Both should be part of a single measurement strategy directed toward evaluable claims.

The third implication concerns disease-area and therapeutic-class reviews. Under the reference-case framework, these reviews are often dominated by simulation outputs, utility estimates and cost-per-QALY comparisons. Reconstruction replaces these constructs with measurable attributes. Therapies are compared according to their demonstrated impact on manifest and latent outcomes. The result is a framework that is transparent, reproducible and directly linked to patient experience and health-system objectives. Claims can be tested, challenged and revised as new evidence emerges.

The fourth implication concerns professional education. For more than four decades, HTA education has focused on utilities, QALYs, simulation modelling and economic evaluation.

Reconstruction requires a different set of skills. Researchers, policy analysts, clinicians and pharmacy faculty must become familiar with representational measurement, dimensional homogeneity, Rasch measurement, protocol design, empirical evaluation and falsification. The emphasis moves from model construction to measurement construction. Understanding how to create lawful measures becomes more important than understanding how to manipulate numerical outputs.

Finally, reconstruction changes the purpose of HTA itself. The objective is no longer to generate estimates of value through increasingly elaborate simulations. The objective is to support credible claims regarding therapy impact that can be evaluated in real-world populations. Measurement regains its primary role. Arithmetic becomes subordinate to measurement rather than a substitute for it. The consequence is a framework grounded in observation, replication and falsification rather than hypothetical projections and non-evaluable claims.

Once it is recognized that there are only two lawful measures, the path forward becomes surprisingly clear. Linear ratio measures provide the framework for manifest attributes. Rasch logit ratio measures provide the framework for latent attributes. Together they offer the foundation for a reconstructed HTA capable of supporting the standards of quantitative science. The challenge is not conceptual but institutional. It requires a willingness to move beyond the reference case and to embrace a measurement-based paradigm for the assessment of therapy impact.

## CONCLUSION

The central argument of this paper is straightforward. For the assessment of therapy impact there are only two lawful forms of measurement: linear ratio measures for manifest attributes and Rasch logit ratio measures for latent attributes. This conclusion follows directly from the requirements of representational measurement and provides a clear foundation for quantitative claims regarding health outcomes. Once these two measurement frameworks are recognized, much of the complexity that has come to characterize contemporary HTA is revealed as unnecessary.

The history of the reference-case paradigm can be interpreted as a gradual movement away from measurement and toward arithmetic. Utilities, QALYs, cost-effectiveness ratios and simulation models were developed in the belief that increasingly sophisticated numerical methods would provide a stronger basis for decision making. Yet sophistication cannot compensate for the absence of lawful measurement. Arithmetic operations applied to entities whose measurement properties have not been established cannot create quantitative meaning. The result is measurement inversion, where arithmetic precedes measurement, and arithmetic chaos, where numerical constructs are treated as though they were measures despite the absence of supporting evidence.

The consequences have been substantial. For more than four decades HTA has relied upon utilities derived from health-state descriptions, preference valuations and scoring algorithms. These utilities have been transformed into QALYs, incorporated into simulation models and used to support reimbursement decisions affecting millions of patients. Yet the legitimacy of the entire framework depends upon assumptions concerning measurement that remain unresolved. The problem is not the quality of the arithmetic. The problem is that arithmetic has been allowed to replace measurement.

Recognition that there are only two lawful measures provides a way forward. Manifest attributes can be assessed through linear ratio measures that support direct observation, replication and falsification. Latent attributes can be assessed through Rasch logit ratio measures that satisfy the requirements of measurement for unobservable traits. Together these frameworks provide the basis for empirically evaluable claims regarding therapy impact without recourse to utilities, QALYs or imaginary future worlds generated by simulation models.

The challenge facing HTA in New Zealand is therefore not one of refinement but reconstruction. The objective is not to improve the reference case but to replace it with a framework grounded in lawful measurement. Once measurement is restored to its proper place, arithmetic becomes a tool rather than a substitute for scientific inquiry. The future of HTA depends upon a return to first principles: measurement precedes arithmetic, claims must be evaluable and falsifiable, and quantitative science begins not with numbers but with measures. There are only two lawful measures. Everything else is arithmetic.

## 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] New Zealand: PHARMAC endorses false measurement for health technology assessment. Logit Working Paper No 1261. https://maimonresearch.com/logit-working-paper-no-1261-april-2026/

[2] New Zealand: The absence of representational measurement in health technology assessment at the University of Auckland. Logit Working Paper No 779. https://maimonresearch.com/logit-working-paper-no-779-may-2026/

[3] New Zealand: The absence of representational measurement in health technology assessment at the University of Otago. Logit Working Paper No 778. https://maimonresearch.com/logit-working-paper-no-778-may-2026/

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