MAIMON RESEARCH LLC

ARTIFICIAL INTELLIGENCE LARGE LANGUAGE MODEL INTERROGATION

REPRESENTATIONAL MEASUREMENT FAILURE IN HEALTH TECHNOLGY ASSESSMENT

ISPOR: NORMAL SCIENCE AND THE ABSENCE OF REPRESENTATIONAL MEASUREMENT IN GOOD PRACTICE - THE MAPPING DEBACLE

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

LOGIT WORKING PAPER No 3011 JULY 2026

www.maimonresearch.com

Tucson AZ

ASTRACT

ISPOR Good Practices Reports constitute an influential international methodological knowledge base for health economics and outcomes research (HEOR) and health technology assessment (HTA). Developed through expert Task Forces, these reports provide consensus recommendations covering economic evaluation, utilities, QALYs, modelling, uncertainty, patient-reported outcomes and healthcare decision making. This study asks whether this extensive methodological framework first satisfies the logically prior requirements of representational measurement.

The ISPOR Good Practices knowledge base was interrogated using a fixed 37-topic framework comprising 17 topics representing the requirements of representational measurement and 20 representing contemporary HEOR and HTA methodology. The measurement topics included measurement preceding arithmetic, scale properties, admissible transformations and arithmetic, dimensional homogeneity, ratio measurement, manifest and latent attributes, linear ratio measurement and Rasch measurement. The 20 contemporary topics included economic evaluation, utilities, QALYs, ICERs, decision analysis, simulation modelling, probabilistic sensitivity analysis, value-of-information analysis, budget impact and value assessment.

The result was categorical: 0/17 representational-measurement topics and 20/20 contemporary HEOR and HTA topics were explicitly represented. The Good Practices corpus therefore provides comprehensive recommendations for manipulating, modelling and interpreting numerical values without first establishing representational measurement as the gatekeeper determining whether those operations are admissible. Mapping provides a particularly revealing case study. Lawful transformation preserves an already established measurement structure: interval measures permit affine transformations and ratio measures similarity transformations, in each case involving the same attribute. Statistical mapping cannot promote ordinal or interval scores to ratio measurement, transfer ratio properties between scales, or transform one attribute into another. It creates predicted numbers, not measurement.

The ISPOR Good Practices knowledge base demonstrates a fundamental inversion of the scientific sequence required for quantitative claims: arithmetic and modelling precede establishment of lawful measurement. Consequently, mapping, utilities, QALYs and subsequent modelling cannot acquire scientific legitimacy through increasing statistical sophistication where their required measurement foundations have not been demonstrated. Mathematics permits numerical manipulation; representational measurement determines whether that manipulation has empirical meaning. The proposed alternative is a transition framework in which measurement precedes arithmetic: manifest attributes are addressed through linear ratio measures and latent attributes through Rasch measurement, followed by prospectively specified, evaluable and falsifiable therapy-impact claims. The Maimon Research nine-unit transformation program provides a structured pathway for implementing this transition in HTA teaching, research and practice.

INTRODUCTION: GOOD PRACTICE AND THE MEASUREMENT GATEKEEPER

The preceding interrogation of the ISPOR educational program-content knowledge base identified a categorical result: 0/17 topics representing the requirements of representational measurement compared with 20/20 topics representing contemporary health economics and outcomes research (HEOR) and health technology assessment (HTA) [1]. The significance of that finding rests upon the logically prior status of the 17 topics. They are not simply another group of methodological competencies to be considered alongside economic evaluation, utilities, QALYs, modelling and uncertainty analysis. They constitute the scientific gatekeeper. Measurement must be established before arithmetic is undertaken, and the properties of the resulting measure determine which mathematical operations are admissible. If that gatekeeper is not satisfied, there is no legitimate scientific basis for proceeding to the downstream quantitative framework.

ISPOR's educational program content, however, represents only one component of its influence upon contemporary HEOR. A second and potentially more consequential knowledge base is provided by its Good Practices Reports (now Good Practices). Since the development of the Task Force program, ISPOR has produced approximately 70 expert-consensus Good Practices Reports, with additional reports continuing to appear. ISPOR describes these reports as highly cited expert-consensus recommendations that establish international standards for HEOR and its application to healthcare decision making. Task Forces bring together international subject-matter experts, while wider review groups contribute comments and feedback before reports are finalized. The reports are published in *Value in Health* and made publicly available. Their intended audience extends across researchers, academics, students, technology assessors, manufacturers, regulators, payers and healthcare decision makers.

The purpose of the Good Practices library is therefore substantially different from that of an ordinary collection of methodological publications. These reports are intended to identify and codify what competent professional practice should be. ISPOR's own description of the Task Force process emphasizes methodological relevance, broad applicability, durability and an evidence-based approach. The reports have addressed economic evaluation, decision-analytic modelling, patient preferences, clinical outcomes assessment, observational research and numerous other areas central to contemporary HEOR. Their stated objective is to improve the quality of HEOR and its contribution to healthcare decision making. The Good Practices library can consequently be regarded as a formal methodological knowledge base through which ISPOR has helped define acceptable professional practice.

This creates an immediate scientific question. Do the Good Practices themselves pass the measurement gatekeeper? If ISPOR recommends how economic evaluations should be performed, how outcomes should be assessed, how utilities should be estimated, how models should be constructed and how quantitative evidence should inform decisions, the prior requirement remains unchanged: the quantities entering those procedures must possess the measurement properties required for the arithmetic performed. Professional consensus cannot waive this requirement. Neither can methodological sophistication, extensive peer review, international participation or longevity of use. Measurement must still precede arithmetic.

The present study subjects the ISPOR Good Practices knowledge base to the same fixed program-content interrogation applied to ISPOR education. Seventeen topics specify the requirements of representational measurement, including measurement preceding arithmetic, scale properties, ratio measurement, admissible transformations and arithmetic, dimensional homogeneity, the distinction between manifest and latent attributes, linear ratio measurement and Rasch measurement. Twenty topics represent the established methodological framework of contemporary HEOR and HTA, including economic evaluation, utilities, QALYs, ICERs, patient-reported outcomes, decision analysis, simulation modelling, probabilistic sensitivity analysis, value-of-information analysis, budget impact and value assessment. The question is deliberately narrow: before recommending what HEOR practitioners should do with numbers, does the Good Practices corpus establish the measurement conditions determining whether those operations are scientifically admissible?

The ordering is critical. The 17 and 20 topics are not competing scores to be balanced against one another. The 17 constitute the gate through which the 20 must pass. If representational measurement is established, quantitative analysis may proceed only with measures and operations satisfying those requirements [2,3]. For therapy-impact assessment this restricts measurement to two pathways consistent with the standards applied here: linear ratio measurement for manifest attributes and Rasch measurement for latent attributes. If the gatekeeper is failed, downstream manipulation to establishing lawful quantitative claims. Decision trees, Markov models, microsimulation, probabilistic sensitivity analysis and value-of-information calculations cannot retrospectively confer measurement properties upon quantities that lacked them at entry.

To illustrate the consequences of ignoring the gatekeeper requirements of lawful measurement, the paper presents mapping as a salutary case study. Mapping was selected because of the substantial attention it has received within ISPOR's Good Practices activities and because it provides an unusually transparent example of the distinction between numerical manipulation and measurement [4]. At its core, mapping uses a statistical relationship to generate one set of numerical values from another. However, elaborate the regression model, estimation procedure, validation exercise or treatment of uncertainty may become, the fundamental operation remains unchanged: numbers are used to predict other numbers.

Nothing in that process establishes lawful measurement. A mapping equation does not demonstrate that the source variable represents a unidimensional attribute with the scale properties required for the arithmetic performed. More importantly, even if the source variable were independently established as a lawful ratio measure, those measurement properties could not be transferred through a regression equation to the target variable. The target must establish its own measurement properties independently. Statistical association between two sets of numbers cannot create unidimensionality, equal units, invariance or a non-arbitrary zero.

Mapping therefore provides a particularly revealing demonstration of why the 0/17 finding matters. In the absence of representational measurement as the gatekeeper, statistical prediction can be mistaken for measurement and increasingly elaborate numerical manipulation can be presented as quantitative methodology. The contrast with Rasch measurement is instructive. For latent attributes, Rasch modelling begins with subjective observations and asks whether they satisfy the requirements necessary to construct an invariant measurement continuum. Mapping asks only whether one set of numbers can predict another. One is concerned with establishing measurement; the other with numerical prediction. ISPOR's substantial attention to mapping, alongside the absence of Rasch measurement as a foundational requirement under the fixed interrogation, therefore provides a striking case study of the consequences of program-content inversion.

THE ISPOR GOOD PRACTICES KNOWLEDGE BASE

The ISPOR Good Practices knowledge base is centered upon the approximately 70 Good Practices Reports developed through ISPOR Task Forces over more than two decades. These reports constitute the primary corpus for the present assessment. They are particularly important because they are not simply methodological articles reflecting the views of individual investigators. They are developed under ISPOR auspices specifically to identify, codify and disseminate recommended professional practice across health economics and outcomes research (HEOR). Collectively, they provide a formal statement of how ISPOR considers major components of HEOR and health technology assessment (HTA) should be undertaken.

The scope of the reports is extensive. Good Practices have addressed economic evaluation, decision-analytic modelling, budget impact analysis, health-state utilities, patient-reported outcomes, preference research, observational studies, real-world evidence, uncertainty, value assessment and related areas of quantitative analysis. The reports therefore span much of the methodological pathway through which evidence is generated and interpreted for therapy assessment and healthcare decision making. They provide an appropriate corpus for asking whether the scientific requirements governing measurement are established before recommendations are made concerning subsequent arithmetic, statistical analysis and modelling.

The Good Practices knowledge base, however, extends institutionally beyond the published reports. Reports are generated through Task Forces comprising recognized subject-matter experts and typically involve wider review and consultation. Their recommendations are subsequently reinforced through ISPOR conferences, presentations, educational programs, short courses, methodological resources and professional discussion. They also exist alongside ISPOR competency frameworks and educational activities defining the knowledge and skills expected of HEOR professionals. The published reports can therefore be regarded as the formal core of a wider professional knowledge environment through which methodological standards are developed, communicated and reproduced.

For purposes of the present interrogation, this distinction provides an important boundary. The primary unit of analysis is the formal Good Practices corpus. The wider ISPOR environment is not treated as an unlimited additional source from which isolated references can be collected to change the classification of individual topics. Rather, it establishes the institutional significance of the corpus and may provide supporting evidence concerning how particular Good Practice recommendations are interpreted and disseminated. This preserves a reproducible assessment while recognizing that the influence of Good Practices extends considerably beyond the reports themselves.

The assessment applies the same fixed37-topic framework used in the preceding ISPOR program-content interrogation. Seventeen topics represent the requirements of representational measurement; 20 represent the methods and constructs of contemporary HEOR and HTA. The question is not whether a term associated with measurement can be found somewhere within the Good Practices literature. The test is considerably more demanding: does the Good Practices knowledge base explicitly recognize representational measurement as the gatekeeper for quantitative analysis?

This distinction is fundamental. A report may discuss "measurement," psychometrics, reliability, validity, health-status assessment or utility measurement without addressing representational measurement. Similarly, discussion of scale construction does not establish recognition that arithmetic is constrained by scale properties. To satisfy the gatekeeper criterion, the knowledge base must establish that measurement precedes arithmetic; that admissible mathematical operations are determined by measurement-scale properties; that multiplication and division require ratio measurement; that dimensional homogeneity must be maintained; and that manifest and latent attributes require different measurement approaches. For therapy-impact claims, this leads to linear ratio measurement for manifest attributes and Rasch measurement for latent attributes.

The 17 representational-measurement topics are therefore not being counted alongside the 20 contemporary HEOR topics as though the assessment involved two equivalent bodies of program content. They are logically prior. The 17 constitute the gatekeeper through which the 20 must pass**.** If the Good Practices knowledge base fails to establish these measurement requirements, there is no legitimate scientific basis for proceeding to downstream arithmetic simply because that arithmetic is subsequently described in considerable methodological detail. Failure at the gatekeeper cannot be repaired by increasingly sophisticated modelling, statistical analysis, simulation or consensus recommendations.

The Good Practices interrogation consequently asks a more consequential question than whether ISPOR has occasionally discussed measurement. It asks whether, across the corpus specifically intended to define good professional practice, ISPOR has required practitioners to establish lawful measurement before undertaking the quantitative operations it recommends. This provides the basis for the two case studies examined subsequently. The QALY case asks whether ratio measurement is established before health-state values are multiplied by time and subsequently used in cost-effectiveness calculations. The mapping case asks whether the measurement properties of the source and target scores are established before sophisticated statistical procedures are employed to map one numerical system onto another.

The ISPOR Good Practices knowledge base is therefore defined for this study as a bounded formal corpus embedded within a much wider institutional system. The approximately 70 Good Practices Reports provide the evidence subjected to the fixed 17/20 interrogation; the Task Force structure, expert participation, educational activities, conferences, journals and competency framework establish why the findings matter. If representational measurement is absent from the formal corpus, the result cannot readily be dismissed as an omission from a particular course or the oversight of an individual author. It would indicate that the measurement gatekeeper is absent from an extensive expert-consensus system specifically created to define and disseminate good practice in HEOR.

That is the central issue addressed by the present assessment. ISPOR has devoted more than two decades to specifying how HEOR should be undertaken. The prior question is whether, before telling practitioners how to perform the analysis, its Good Practices knowledge base establishes whether the quantities involved are measures and whether the proposed arithmetic is scientifically admissible. If that gatekeeper is absent, the problem is not simply incomplete guidance. It is program-content inversion within the very corpus intended to define Good Practice.

PROGRAM CONTENT REVIEW

Organization
ISPOR — The Professional Society for Health Economics and Outcomes Research

Knowledge Base
ISPOR Good Practices for Outcomes Research Knowledge Base

Purpose

This interrogation evaluates the corpus of ISPOR Good Practices and predecessor Good Research Practices reports against the identical fixed17 + 20 topic framework used in the program-content assessments. The question is whether ISPOR's formal good-practice standards explicitly establish representational measurement as the foundation for quantitative claims before recommending the methods of contemporary HEOR and HTA.

This is a particularly demanding test because ISPOR does not describe these reports simply as publications expressing individual opinions. It describes them as highly cited expert-consensus guidance that establishes international good-practice standards and shapes the use of HEOR in healthcare decision making.

The corpus spans economic evaluation, budget impact, utilities, patient-centered research, clinical outcomes, modelling and simulation, value of information, HTA, reimbursement and health-policy applications.

TABLE 1: REPRESENTATIONAL MEASUREMENT

Topic	Explicitly recognized	ISPOR Good Practices evidence
Representational measurement	No evidence identified	—
Scales of measurement as constraints on arithmetic	No evidence identified	—

Topic	Explicitly recognized	ISPOR Good Practices evidence
Measurement precedes arithmetic	No evidence identified	—
Axioms of representational measurement	No evidence identified	—
Ratio measurement as prerequisite for multiplication and division	No evidence identified	—
True zero and equal units	No evidence identified	—
Admissible transformations	No evidence identified	—
Admissible arithmetic by scale type	No evidence identified	—
Dimensional homogeneity	No evidence identified	—
Manifest and latent attributes as distinct measurement classes	No evidence identified	—
Linear ratio measurement for manifest attributes	No evidence identified	—
Possession of latent attributes	No evidence identified	—
Rasch measurement as requirement for latent attributes	No evidence identified	—
Person–item conjoint measurement	No evidence identified	—
Specific objectivity and invariance	No evidence identified	—
Rasch logit ratio measurement for latent attributes	No evidence identified	—
Two admissible ratio measures for therapy impact	No evidence identified	—

Table 1 Result: 0/17 representational-measurement topics explicitly identified as foundations of the ISPOR Good Practices knowledge base.

This classification does **not** mean ISPOR's reports ignore measurement in the ordinary psychometric sense. For example, its Clinical Outcome Assessments report discusses measuring instruments, scores and measurement properties, while later Good Practices Reports address content validity and comparability of PRO measures. What was not identified is the specific content tested by the fixed instrument: representational measurement, arithmetic constrained by scale type, ratio measurement as prerequisite for multiplication and division, Rasch as the required latent-measurement framework, person–item conjoint measurement or specific objectivity.

It is equally important that Rasch material appearing elsewhere on ISPOR's website was excluded from this classification. Conference abstracts exist describing Rasch analysis, person-item maps and Rasch Measurement Theory. Those demonstrate that Rasch is known within the broader ISPOR community, but they are contributed conference research, not ISPOR Task Force standards.

TABLE 2: CONTEMPORARY HEALTH ECONOMICS AND HTA FRAMEWORK

Topic	Explicitly represented	ISPOR Good Practices evidence
Health economics	Yes	Economic-evaluation, health-economic and decision-making reports
Health technology assessment	Yes	HTA guideline development; HTA and policy reports
Economic evaluation	Yes	Dedicated Economic Evaluation corpus
Cost-effectiveness analysis	Yes	CEA Task Force and quality-improvement reports
Cost-utility analysis	Yes	Utility/QALY economic-evaluation reports
Pharmacoeconomics	Yes	Pharmacoeconomic-information and drug-cost Task Force reports
Utilities and preference-based outcomes	Yes	Health-state utility and mapping reports
Quality-adjusted life-years (QALYs)	Yes	Utility reports; vaccination economic evaluation
Incremental cost-effectiveness ratios (ICERs)	Yes	Incremental CEA/cost-per-QALY framework
Measurement and valuation of health outcomes / patient-reported outcomes	Yes	COA, PRO and health-state utility reports
Decision analysis	Yes	Decision-analytic modelling standards
Decision-tree modelling	Yes	Decision-model framework
Markov modelling	Yes	State-transition modelling Good Research Practices
Simulation and microsimulation modelling	Yes	Markov cohort, individual simulation, dynamic simulation
Probabilistic sensitivity and uncertainty analysis	Yes	Model Parameter Estimation and Uncertainty Analysis
Value-of-information analysis	Yes	Two dedicated VOI Good Practices reports

Topic	Explicitly represented	ISPOR Good Practices evidence
Budget impact analysis	Yes	Two Principles of Good Practice BIA reports
Health care priority setting and resource allocation	Yes	Economic-evaluation, opportunity-cost and allocation guidance
Value-for-money / value assessment	Yes	Value assessment framework and economic-value reports
Policy, reimbursement and decision use of economic evaluation	Yes	Pharmacoeconomics use, transferability and reimbursement guidance

Table 2 Result: 20/20 contemporary health-economics and HTA topics explicitly represented in the ISPOR Good Practices knowledge base.

The evidence on this side is not merely incidental. ISPOR has dedicated Good Practices series for budget impact, cost-effectiveness analysis alongside trials, health-state utilities, economic modelling, state-transition modelling, uncertainty analysis and value of information.

The QALY framework is particularly explicit. The 2019 Health State Utilities Task Force states that health-state utilities are required to calculate QALYs in cost-effectiveness models and provides recommendations for identifying, synthesizing and applying those utilities. An earlier mapping report gives good-practice recommendations specifically for generating preference-based outcomes for QALY-based economic evaluation.

Similarly, the modelling corpus is extensive. ISPOR's modelling Good Research Practices series includes conceptualization, state-transition modelling, discrete-event simulation, dynamic transmission modelling, parameter estimation, uncertainty analysis, transparency and validation. The state-transition report expressly includes Markov cohort simulation and individual-level Monte Carlo microsimulation.

ISPOR also formally connects these methods to healthcare decisions. Its pharmacoeconomics Task Force addressed how decision makers should interpret and use PE/HE studies, while its transferability report explicitly concerns economic evidence submitted for pricing and reimbursement decisions.

The interrogation therefore yields a categorical result:

Representational measurement: 0/17
Contemporary health economics and HTA: 20/20

ASSESSMENT

This result has a different character from the university program-content interrogations. In California, Australia or New Zealand, the question was what universities and research programs teach, research and transmit. Here the knowledge base comprises documents that ISPOR itself characterizes as good-practice standards. ISPOR states that its Task Force reports provide expert-consensus recommendations and set international standards for HEOR and its use in healthcare decision making. The result can therefore reasonably be characterized as good-practice standards inversion under the fixed instrument.

The 20/20 side is comprehensive. The reports do not merely acknowledge contemporary HTA methods; they provide recommendations for conducting them well. CEA has dedicated good-research-practice guidance. Budget impact has dedicated principles of good practice. Health-state utilities have dedicated recommendations for their use in cost-effectiveness models. State-transition and simulation models have consensus best-practice standards. Parameter uncertainty and VOI receive formal methodological guidance.

This makes the 0/17 result especially notable. ISPOR has repeatedly addressed questions of methodological rigor, validity, transparency, credibility, uncertainty and reporting. Its 2009 Quality Improvement of Cost-Effectiveness Research Task Force explicitly called for rigorous methodological standards and continuing improvement in economic evaluation. Yet within the corpus examined, no explicit good-practice foundation was identified requiring the prior establishment of representational measurement, ratio properties, admissible transformations or dimensional homogeneity before arithmetic is undertaken.

Nor can this absence be explained simply by assuming that ISPOR has no interest in measurement. Clinical outcome assessment is a substantial part of the corpus. The 2015 COA conceptual-foundation report explicitly discusses instruments, ratings or scores and adequate measurement properties. PRO Task Forces address content validity, instrument modification, translation, mode comparability and other measurement issues. The distinction is that these reports operate within clinical-outcomes and psychometric traditions rather than establishing the fixed representational-measurement requirements as constraints upon HEOR arithmetic.

The contrast is especially clear in relation to utilities and QALYs. The Good Practices corpus does not merely encounter these constructs; it provides formal guidance for their generation, selection and application. ISPOR's health-state utility report begins from the role of utilities in calculating QALYs and proceeds to recommend how they should be identified and used in cost-effectiveness models. Its 2026 child-health utility Good Practices Report similarly provides international guidance for deriving utilities for economic evaluation and notes their effect on cost-utility analysis and resource-allocation decisions.

The same structure appears in modelling. The Good Practices corpus asks how models should be conceptualized, parameterized, validated and subjected to uncertainty analysis. It provides standards for Markov models, individual simulation and other modelling approaches. What the 0/17 result indicates is that these downstream methodological standards are not preceded, within the tested framework, by explicit standards establishing whether the quantities entering those models possess the measurement properties necessary for the arithmetic performed.

The categorical nature of the result is therefore important. This is not a finding that representational measurement receives less attention than modelling. Under the fixed instrument, it is 0/17 versus

20/20. The corpus comprehensively specifies good practice for the established HEOR framework while none of the 17 representational-measurement foundations was identified as an explicit prerequisite governing that framework.

There is an additional methodological point worth emphasizing. ISPOR's broader website contains contributions that demonstrate awareness of precisely some of the issues absent from the Good Practices corpus. A 2003 ISPOR conference abstract, for example, argued that QALY arithmetic is compromised because life-years and utility values have different scale properties. Rasch analyses also appear repeatedly in conference research. Those contributions were deliberately **not** counted in Table A because the interrogation concerns what ISPOR has elevated to good-practice guidance, not everything its members have ever presented. Their existence actually sharpens the distinction between ideas present somewhere within the professional community and principles formally incorporated into ISPOR standards.

On the fixed criterion, the conclusion is therefore clear. The ISPOR Good Practices knowledge base demonstrates complete standards inversion**:** 0/17 versus 20/20. The organization has developed an extensive and sophisticated corpus specifying how contemporary HEOR and HTA methods should be conducted, modelled, validated, reported and used in decision making. What was not identified is an antecedent representational-measurement framework establishing whether the numerical quantities involved possess the properties required for the arithmetic those good-practice methods recommend.

CASE STUDY: THE MAPPING DEBACLE

In health technology assessment, mapping is the statistical prediction of utility scores from observations generated by another instrument, typically a disease-specific or generic patient-reported outcome measure. It is used when a clinical study has not collected the preferred multiattribute utility instrument, such as the EQ-5D, but an economic evaluation nevertheless requires utility values to calculate QALYs. A regression or related statistical model is constructed from a dataset in which both instruments have been administered, and the estimated relationship is then used to predict utility scores for patients for whom only the non-preference-based instrument is available.

Mapping was adopted primarily as a pragmatic response to missing utility data. Reference-case economic evaluation requires utilities and QALYs, yet many trials were not designed to collect the required preference-based scores. Rather than question whether the required construct possessed legitimate measurement properties, HTA developed methods for manufacturing the missing values indirectly. Increasingly elaborate regression techniques, response mapping, mixture models and other predictive approaches followed.

The deeper problem is that prediction does not establish measurement. A statistical relationship between two sets of scores cannot confer measurement properties on either. If the target utility scale lacks the properties required for the arithmetic subsequently performed upon it, predicting values on that scale does nothing to correct the deficiency. Mapping simply reproduces the target scale's measurement limitations with an additional layer of statistical uncertainty.

The mapping enterprise is therefore indicative of a broader measurement inversion within HTA. The objective became obtaining the numbers required by the economic model rather than first establishing whether those numbers represented a legitimate measure. Instead of measurement determining what arithmetic is permissible, the demands of the reference case determined what numbers had to be produced. Mapping is a particularly clear example of methodology being developed to sustain an established analytical framework rather than to satisfy the prior requirements of measurement.

The significance of the mapping problem extends well beyond ISPOR. It is particularly striking in the United Kingdom, where mapping currently occupies an explicit role in the NICE reference-case framework[5] [6]. Where EQ-5D-5L data have been collected, NICE's current approach requires those responses to be mapped onto the EQ-5D-3L value set for reference-case economic evaluation [7]. The exercise therefore does not merely concern predicting EQ-5D values from an unrelated patient-reported outcome. It involves statistically transforming one version of the EQ-5D numerical system into another so that the resulting EQ-5D-3L utility values can enter QALY-based economic models.

From the standpoint of representational measurement, this is an extraordinary procedure. Neither the EQ-5D-5L nor the EQ-5D-3L utility value has thereby been demonstrated to possess ratio-scale properties. Mapping cannot remedy that deficiency. It simply takes numerical information generated under one scoring system and uses a statistical relationship to generate numbers corresponding to another. The resulting EQ-5D-3L values do not acquire unidimensionality, invariant equal units or a non-arbitrary zero merely because they have been predicted from EQ-5D-5L responses.

Indeed, NICE's own mapping literature recognizes that the 3L and 5L systems cannot simply be regarded as equivalent. The response has been to develop statistical procedures linking them. But statistical linkage is not measurement. A mapping function can estimate what numerical value on one system is likely to correspond to observations on another; it cannot establish the measurement properties of either system.

The subsequent use of the mapped values makes the problem still more consequential. Predicted EQ-5D-3L utility values are treated as quantities suitable for discounting time to construct QALYs and ultimately for the calculation of incremental cost-per-QALY claims. Yet mapping has established none of the ratio properties required for that arithmetic. One set of numbers has been transformed into another set of numbers and then subjected to further arithmetic as though lawful measurement had been established somewhere in the process. It has not.

The NICE 5L-to-3L mapping requirement therefore provides an unusually clear example of measurement inversion. The methodological debate concerns which statistical procedure most successfully links the two numerical systems when the logically prior question is whether either provides the lawful measurement structure required for QALY arithmetic. Without that foundation, mapping from 5L to 3L does not solve a measurement problem. It merely reformulates a numerical one.

MAPPING AS LAWFUL TRANSFORMATION

Mapping should not be confused with lawful transformation between measurement scales. Statistical modelling can always generate new numbers, but creating numbers is not creating measures. Lawful transformation begins with an attribute for which measurement properties have already been established and preserves those properties when its numerical representation is changed.

For an ordinal scale, admissible transformations are monotonic transformations that preserve order. Ordinal scores establish only greater than, less than or equal ordering. Differences between scores have no invariant quantitative interpretation. An ordinal scale therefore provides no equal-unit structure that can be carried forward through arithmetic mapping to create interval or ratio measurement.

For an interval scale, equal units have been established but the zero is arbitrary. The admissible transformation is affine that allows both a change in the zero point and a change in the size of the unit**:**

$y = a + bx$, where $b > 0$.

Celsius and Fahrenheit provide the familiar example. Both represent the same attribute, temperature, with equal units within each system but different arbitrary zeros and unit sizes. The transformation changes the numerical representation while preserving the interval structure. It does not create measurement; measurement existed before the transformation.

For a ratio scale, equal units and a non-arbitrary zero have been established. The admissible similarity transformation preserves that zero:

$y = bx$, where $b > 0$.

Meters and feet or kilograms and pounds are examples. Again, the transformation concerns alternative representations of the same attribute. The ratio properties existed before transformation and are preserved by it.

These restrictions expose the problem with conventional HEOR mapping. An interval measure can be transformed lawfully into another interval representation of the same attribute through an affine transformation. A ratio measure can similarly be transformed into another ratio representation of that same attribute through a similarity transformation. What cannot be done is to use statistical modelling to promote an ordinal scale to an interval or ratio scale, transform an interval scale into a ratio scale, or transform one attribute into another and claim that measurement has been preserved.

This is particularly important when disease-specific scores are mapped onto so-called utility values. Unless the source and target possess the required measurement structures and represent the same attribute, regression produces predicted numerical values rather than a lawful transformation of measurement.

Mapping may therefore create equations, predictions and new numbers. It cannot manufacture a higher level of measurement from a lower one. Ordinal remains ordinal; interval remains interval; ratio remains ratio. Lawful transformation preserves measurement structure; it does not create it.

NUMERICAL STORYTELLING IS NOT MEASUREMENT LEGITIMACY

There is, of course, nothing mathematically preventing mapping. If numbers are assigned to observations, statistical procedures can be applied to them. Regression models can be estimated, coefficients calculated, predictions generated and measures of predictive performance reported. If the working assumption is that numerical values may be combined and transformed without regard to their measurement properties, mapping is entirely acceptable as a mathematical exercise.

The problem arises when the resulting numbers are interpreted as quantitative measures. Representational measurement does not ask whether arithmetic can be performed; arithmetic can always be performed on numbers. It asks whether the empirical relationships represented by those numbers justify the particular arithmetic operations undertaken and the interpretation subsequently placed upon their results.

This distinction is fundamental. Mathematics operates upon numerical symbols according to formal rules. It does not determine whether those symbols constitute measurements of an empirical attribute. That requires an independently established correspondence between the empirical and numerical relational systems. Scale type then determines which transformations and arithmetic operations preserve that correspondence.

Mapping that ignores these requirements may therefore be mathematically impeccable while being meaningless in representational-measurement terms. An ordinal score can be entered into a regression equation. An interval score can be used in multiplication. A multidimensional composite can be statistically related to a utility value. A computer will perform every calculation without objection. None of this demonstrates that the resulting numbers possess the measurement properties required for their proposed interpretation.

The issue is therefore not whether mapping can be done. It plainly can. The issue is whether the mapped values can legitimately be interpreted as measures and subjected to subsequent arithmetic. If the axioms of representational measurement are set aside, almost any numerical transformation becomes possible. If those axioms are accepted as the foundation of quantitative science, the permissible transformations are constrained by the measurement structure established before arithmetic begins.

Mathematics permits the calculation. Measurement theory determines whether the calculation means anything.

CONCLUSION

The ISPOR Good Practices library is founded on a preoccupation with numerical manipulation rather than scientific measurement. Its reports devote extraordinary attention to statistical estimation, mapping, utility weighting, probabilistic sensitivity analysis, simulation modelling, value-of-information analysis and increasingly sophisticated computational techniques. Yet all of these methods share a common characteristic: they manipulate numbers without first establishing that those numbers are lawful measures of empirical attributes. The result is an elaborate architecture of arithmetic unsupported by representational measurement.

Lawful quantitative transformation is scale-specific. Interval measures permit affine transformations, while ratio measures permit similarity transformations. In either case, the transformation preserves the measurement properties of the same unidimensional attribute. It does not create measurement, change the scale type or transform one attribute into another. The required measurement properties must exist before transformation is undertaken.

Mapping cannot promote an ordinal or interval score to ratio measurement, nor can the ratio properties of a target be transferred to a source through regression. If the source is ordinal, it remains ordinal regardless of how accurately it predicts numerical values associated with a ratio-scale outcome. If the source is interval, an admissible transformation preserves its interval status; it does not manufacture the non-arbitrary zero required for ratio measurement.

Nor can mapping transform one attribute into another. Physical function cannot become pain, fatigue cannot become need fulfilment, and a disease-specific latent attribute cannot become a multidimensional health preference merely because their numerical values are correlated. Regression may predict one from another, but prediction between attributes is not measurement transformation.

Legitimate transformations therefore preserve both the identity of the attribute and the measurement properties of the scale. Length may be expressed in meters or feet and mass in kilograms or pounds because each pair represents the same underlying attribute and each already possesses the required ratio properties. Likewise, an interval measure may be represented by another interval scale through an admissible affine transformation. In neither case does the transformation create measurement. It preserves measurement already established.

This failure denies the Good Practices library scientific status as a framework for quantitative claims concerning therapy impact. Scientific measurement is not created by regression, simulation or statistical sophistication. It must exist before arithmetic is undertaken. Once this principle is accepted, the numerical storytelling that characterizes much of the Good Practices framework becomes unnecessary because the quantities required to support many of its operations have not first been established. Mapping, composite utility scores, QALYs and probabilistic sensitivity analyses built upon such values do not merely raise methodological questions; they lack the measurement foundation required to support the quantitative interpretations placed upon them.

The solution is remarkably straightforward. Had the Good Practices program begun with the axioms of representational measurement rather than statistical methodology, it would have recognized that arithmetic follows measurement and can never substitute for it. There would have been no need to construct increasingly elaborate numerical procedures while leaving the measurement status of their inputs unresolved.

If the purpose of health technology assessment is to evaluate therapy impact, the measurement requirement comes first. Manifest attributes require lawful linear ratio measures. For latent attributes the solution is Rasch logit ratio measurement Mapping has nothing to do with these measures; thy are dictated by the axioms of representational measurement. Only after lawful measurement has been established can arithmetic support evaluable and falsifiable claims. Without that transition foundation, increasing mathematical sophistication does not produce better measurement. It produces increasingly sophisticated numerical storytelling.

The alternative is not another layer of statistical refinement but a return to measurement. Transition must begin by providing faculty and practitioners with the knowledge and practical tools required to reconsider the scientific foundations of what they teach and apply. It is precisely this gap that the Maimon Research nine-unit transformation program is intended to address [8]. Designed specifically for HTA, the program provides a structured transition from the legacy framework represented by the ISPOR Good Practices recommendations to one grounded in representational measurement. Beginning with the principle that measurement precedes arithmetic, the nine units progressively address the definition of attributes, scales of measurement, admissible arithmetic, dimensional homogeneity, manifest and latent attributes, linear ratio measurement, Rasch measurement, attribute possession and the construction of empirically evaluable therapy-impact claims. The purpose is not to append a measurement module to contemporary HTA while leaving its existing analytical structure intact. It is to demonstrate how the entire discipline changes once representational measurement becomes its organizing scientific principle. Utilities, mapping, QALYs and reference-case simulation are no longer taken as starting points; the starting point becomes the attribute for which a therapy-impact claim is required and the measurement system necessary to support that claim.

The nine-unit program then carries measurement into practice. Manifest therapy-impact attributes are addressed through lawful linear ratio measures, while latent attributes are addressed through Rasch measurement with requirements for unidimensionality, invariance, specific objectivity and possession of the defined attribute. These measures provide the foundation for prospectively specified claims and protocols defining the target population, comparator, attribute, measurement instrument, timeframe, success criterion and falsification rule. Therapy impact can then be evaluated against empirical observations'

The Program's objective is not simply to demonstrate why the ISPOR Good Practices framework fails to satisfy the requirements of lawful measurement, but to provide a coherent alternative for teaching, research and HTA practice. After forty years of arithmetic preceding measurement, the task is not to refine the inherited framework but to replace it with what quantitative science requires.

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] ISPOR: Normal science and the absence of representational measurement in HTA Education. Logit Working Paper No 3010. https://maimonresearch.com/logit-working-paper-no-3010-july-2026/

[2] Stevens S. On the Theory of Scales of Measurement. *Science*. 1946;103(2684):677-80

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

[4] Wailoo AJ, Hernandez-Alava M, Manca A, Mejia A, Ray J, Crawford B, Botteman M, Busschbach J. Mapping to Estimate Health-State Utility from Non–Preference-Based Outcome Measures: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value in Health*. 2017;20(1):18–27.

[5] van Hout B, Janssen MF, Feng Y-S, Kohlmann T, Busschbach J, Golicki D, Lloyd A, Scalone L, Kind P, Pickard AS. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value in Health*. 2012;15(5):708–715

[6] Longworth L, Rowen D. NICE DSU Technical Support Document 10: The Use of Mapping Methods to Estimate Health State Utility Values, NICE 2011

[7] National Institute for Health and Care Excellence (NICE). *Position statement on use of the EQ-5D-5L value set for England.* Published 2 November 2017; updated October 2019. NICE Corporate Document ECD16.

[8] Maimon Research. HTA Reconstruction. https://maimonresearch.com/hta-reconstruction-program-and-fees/