MAIMON WORKING PAPER No 19 SEPTEMBER 2025

THE FAILURE OF ICHOM: PART 1 THE ABSENCE OF MEASUREMENT

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

ABSTRACT

The International Consortium for Health Outcomes Measurement (ICHOM) was created to provide global "standard sets" of outcomes for benchmarking and value-based care. Its appeal lies in the promise of comparability: if providers use the same tools, international comparisons and policy learning seem possible. Yet beneath this attractive vision lies a profound flaw. The instruments endorsed by ICHOM are not measures. They are disease-specific, multiattribute questionnaires selected by consensus, usually on the grounds of familiarity or convenience, with no filter for measurement legitimacy. The result is not a global framework of measurement but the globalization of pseudo-measurement.

Representational Measurement Theory (RMT) has, for decades, defined the axioms required for numbers to qualify as measures: unidimensionality, additivity, solvability, cancellation, and invariance. By the 1970s these standards were codified in Foundations of Measurement, while Rasch modeling, made explicit by Wright in 1977, provided the practical solution for latent constructs. Rasch alone ensures that ordinal responses can be transformed into interval scales, enforcing unidimensionality and invariance across populations. ICHOM has ignored this settled science, institutionalizing summed scores and composites that fail the most basic requirements of measurement.

The appeal of ICHOM's approach is sociological, not scientific. Policymakers crave simple metrics, hospitals want dashboards, and funders demand comparability. ICHOM delivers this by producing numbers that look like measures, but are not. This mirrors the trajectory of the QALY and the Tufts cost-effectiveness registry, which appear rigorous but rest on utilities that fail RMT axioms. The ignorance is not accidental; it is convenient. Confronting measurement theory would collapse the edifice of pseudo-standards.

The consequences are serious. Policy is distorted, hospitals are misled by spurious benchmarks, patients are harmed by false claims of benefit, and research is corrupted by non-measures as trial endpoints. Globally, the credibility of value-based healthcare is eroded.

ICHOM's influence is undeniable, but influence is not legitimacy. Its standard sets enshrine the absence of measurement. The only remedy is reconstruction: ratio scales for manifest attributes and Rasch-based logit scales for latent constructs. Unless ICHOM pivots to these foundations, it will remain what it is today: a global façade of science, harmonizing error rather than measurement.

INTRODUCTION: STANDARDS WITHOUT MEASUREMENT

The International Consortium for Health Outcomes Measurement (ICHOM) was founded with a bold and seemingly unassailable mission: to create global "standard sets" of outcomes for different diseases, enabling providers, health systems, and payers to compare results, benchmark performance, and improve value in healthcare. At first glance the appeal is obvious. If everyone uses the same outcomes in the same disease area, then international comparisons become possible, variation can be identified, and health systems can learn from each other. The idea is elegant, persuasive, and politically attractive.

Yet beneath the rhetoric lies a fatal flaw. The outcomes selected for ICHOM's standard sets are not measures at all. They are multiattribute and disease-specific instruments chosen by consensus panels, often on the grounds of familiarity or face validity, with no filter for measurement legitimacy. The result is not a global framework of measurement but a global framework of pseudo-measurement, a harmonization of instruments that do not meet the most elementary requirements of measurement science. What ICHOM institutionalizes is not comparability but confusion: a global language of non-measures parading as measures.

It is worth emphasizing the ICHOM is not alone in putting to one side the importance of measurement for disease specific instruments. This neglect is a defining feature of health technology assessment. For 80 years the standards for measurement have been ignored; instead, the focus has been on ignoring the axioms of representational measurement theory (RMT) in favor of valuing health state descriptions [1]. These fail the first axiom of RMT: unidimensionality.

The argument of this paper is straightforward. Representational Measurement Theory (RMT) has long defined the conditions under which numbers qualify as measures. Since the 1970s, with the publication of *Foundations of Measurement* and the uptake of Rasch modeling, the standards have been explicit, rigorous, and available. To claim to set "standards" in outcomes while ignoring these foundations is to abandon science. ICHOM exemplifies this abandonment. Its standard sets are the very opposite of standards: they enshrine the absence of measurement.

The answer is simple, and damning: ICHOM, its staff, and the academic support groups it relies on have no idea of the theory of measurement. This is not a trivial gap but a categorical failure in education and experience of other disciplines. The axioms of representational measurement were agreed and codified more than fifty years ago: unidimensionality, additivity, solvability, cancellation, and invariance. These are not optional refinements. They are the necessary conditions for any numerical assignment to qualify as measurement. Without them, numbers are at best labels, at worst distractions masquerading as science.

By 1977, the role of Rasch measurement, proposed in 1960, as the practical solution for latent constructs had been made explicit by Wright [2 3]. Rasch's probabilistic model was shown to provide the indispensable bridge between the axioms of representational measurement and applied outcomes research. Unlike classical test theory or general item response theory, Rasch enforces the axioms: items and persons must conform to a unidimensional continuum, responses are

modeled as a function of the difference between person ability (or possession) and item difficulty, and invariance is tested across populations. If the data fit the model, then and only then do we have an interval scale suitable for arithmetic and comparison [4]. Wright's 1977 insight made it clear that Rasch was not just another statistical model, but the only model consistent with measurement theory.

WHERE IGNORANCE IS BLISS

The success of ICHOM's "standard sets" rests less on scientific rigor than on the convenience of ignorance by experts and stakeholders. By gathering disease-specific instruments and presenting them as benchmarks of best practice, ICHOM has offered health systems and clinicians a semblance of order. The promise is simple: here is a catalogue of tools, endorsed by experts, that can be used across registries, clinical trials, and comparative studies. The appeal is obvious. In a world where outcomes measurement is fragmented and contested, ICHOM appears to deliver a common language. But appearances are not science. The instruments that fill these standard sets are not measures. They are almost always summed ordinal scores spanning multiple constructs, or composites that collapse disparate attributes into a single number. Their acceptance signals not scientific progress but collective amnesia for the axioms of representational measurement theory. In this amnesia, ignorance has become bliss.

RMT, codified between the 1940s and the 1970s, laid out with ruthless clarity the conditions under which numbers can claim the status of measures. Stevens, in his seminal 1946 paper, drew the boundary between ordinal and interval assignments, making clear that arithmetic without regard to admissible transformations produces nonsense [5]. Suppes in the 1950s set down the axioms of extensive measurement, showing that additivity could be justified only when empirical concatenations preserved structure [6]. Luce and Tukey in the 1960s codified the axioms of additive conjoint measurement, single and double cancellation, solvability, the Archimedean property, and demonstrated how these support the representation and uniqueness theorems [7]. And in 1971, Foundations of Measurement, Volume I, gave the definitive synthesis [1]. Rasch, working independently in the 1960s, had already provided the probabilistic model that allows ordinal responses to be transformed into interval scales when the data fit. By 1977, Wright had shown that Rasch was not just another statistical device but the only one aligned with the axioms of measurement. By the time ICHOM was launched decades later, there was no excuse for ignorance. The standards were not obscure or contested; they were settled science.

Yet ICHOM pressed ahead as though none of this existed. Instead of insisting that instruments meet the requirements of unidimensionality, additivity, and invariance, it endorsed what was available, familiar, and convenient. A questionnaire that bundles fatigue, pain, and mood into a single total score was not rejected but celebrated as efficient. A global quality-of-life instrument built on ordinal Likert items was not interrogated for its failure to support arithmetic but promoted as a standard. The bliss lies in not asking whether these numbers are measures, because to ask is to discover that they are not. Once that discovery is made, the entire edifice of comparability collapses. Better, then, not to ask.

The parallel with the Tufts Cost-Effectiveness Analysis Registry is striking [8]. For more than three decades, Tufts has collected published cost-per-QALY studies into a searchable database that now

includes thousands of entries . Policymakers, analysts, and journal editors use it as an authoritative resource. It seems to provide a panoramic overview of the "evidence" on cost-effectiveness, distilled into clean ratios of dollars per QALY. But like ICHOM's standard sets, this resource is a monument to ignorance. Every QALY it contains is constructed from utilities that are not measures, derived from preference tariffs built on ordinal data. These pseudo-measures are then multiplied by time, a legitimate ratio variable, to create a product that violates dimensional homogeneity. The resulting QALYs are undefined, their ratios mathematically meaningless. Yet the database flourishes, not because the entries are scientifically valid, but because users choose not to confront the axioms of measurement. Here, too, ignorance is bliss as the term cost-effectiveness in RMT axiom terms is meaningless.

The shared pathology is the refusal to engage with RMT. For manifest attributes the axioms of measurement are satisfied by direct inspection if they admit true zeros, equal intervals, and proportional comparisons. Claims framed in these terms are legitimate: they can be tested, replicated, and falsified. For latent constructs measurement requires Rasch. Only when items conform to the Rasch model can responses be transformed into interval scales, yielding logits that preserve constant relative differences, enforce unidimensionality, and test invariance. Without Rasch, summed scores remain ordinal and cannot sustain arithmetic. This is not an esoteric demand but the minimal condition for science. Yet both ICHOM and Tufts proceed as if these standards were irrelevant.

The result is not measurement but numerology. When ICHOM endorses a heart failure questionnaire that yields a total score by adding symptoms, function, and social participation, it does not matter how widely that score is used: it cannot sustain subtraction or division. When Tufts reports that a therapy costs $75,000 per QALY, it does not matter how many studies agree: the denominator is not a measure, and the ratio is meaningless. Both enterprises flourish only because their users are willing to suspend disbelief. In the absence of a measurement filter, consensus or convention substitutes for science.

The deeper irony is that this ignorance is voluntary. It is not that the standards of measurement are unknown. They have been published, codified, and taught for decades. They are ignored because they are inconvenient. To recognize that most legacy patient-reported outcome instruments are non-measures is to admit that ICHOM's catalogue collapses. To recognize that QALYs are undefined is to admit that the Tufts database is an archive of nonsense. The institutions survive by pretending that these issues do not exist, and by reassuring themselves that policy utility is justification enough. That is the bliss: to avoid the discomfort of acknowledging that one's life's work rests on pseudo-measurement.

But ignorance cannot be a permanent refuge. Claims that cannot be falsified are not scientific claims. Numbers that cannot sustain arithmetic are not measures. Consensus cannot transmute ordinal sums into interval scales. The longer ICHOM persists in this illusion, the deeper the credibility crisis becomes. For eventually, the recognition will come.

COLLECTIVE MEASUREMENT AMNESIA IN HTA

Health technology assessment (HTA) presents itself as the scientific guardian of healthcare resource allocation, but its education and practice are marked by a striking and persistent amnesia. HTA training programs and textbooks have ignored RMT for half a century. This amnesia is not an innocent oversight but a product of historical path dependence and institutional convenience [9]. In the 1970s and 1980s, as health systems looked for methods to ration scarce resources, economists and policy makers embraced utilities, QALYs, and reference case models. These tools gave the appearance of rigor while sidestepping the problem that their numbers were not measures at all. HTA courses, responding to the demand from agencies and funders, trained students in how to operate this machinery rather than asking whether its foundations were valid. Simulation modeling, regression analysis, and probabilistic sensitivity analysis became the standard curriculum, while RMT and Rasch were never mentioned.

The omission also reflects a lack of psychometric literacy. Health economists are typically trained in statistics and modeling but not in the axioms that determine whether numbers qualify as measures. As a result, educators pass on what they know: the rituals of cost-effectiveness analysis rather than the science of measurement. The collective memory of measurement theory—well understood in other sciences—was effectively erased from HTA.

This silence has been reinforced by institutional self-protection. If HTA education acknowledged RMT, it would expose the incoherence of its core instruments. EQ-5D utilities, QALYs, and reference case models all fail the axioms. To admit this would undermine decades of guidelines, publications, and careers. Far easier to maintain the illusion of rigor by never raising the measurement question at all.

The result is collective amnesia: a field that behaves as though measurement theory does not exist, despite its clarity and availability for over fifty years. Until HTA restores memory—ratio scales for manifest claims, Rasch-calibrated logit scales for latent constructs—it will remain trapped in numerical storytelling, insulated from the very science it claims to embody.

UNDERSTANDING ICHOMS'S NEGLECT: CONSENSUS IS NOT SCIENCE

Against this settled background, ICHOM's neglect is inexcusable. Launched in the second decade of the twenty-first century, it could not plead ignorance. The science of measurement had been clarified, debated, and resolved decades earlier. ICHOM entered the field with this entire intellectual scaffolding in place and, like the leaders in HTA, chose to ignore it.

Instead, ICHOM pursued a program of consensus. Disease-specific working groups were convened to select outcomes, almost always defaulting to legacy questionnaires. These instruments, built on summed ordinal scores across multiple attributes, fail every relevant axiom of RMT. They are not unidimensional, they do not permit additivity, and they do not sustain invariance. Yet ICHOM endorsed them as "standard sets," presenting pseudo-measures as if they were interval or ratio scales, and exporting this illusion globally.

This is why ICHOM fails. It has confused consensus with science, endorsement with measurement. It has treated numbers as measures when the conditions of measurement were never satisfied. The indictment is severe: ICHOM with HTA has institutionalized ignorance of measurement theory, perpetuating false claims for fifty years after the axioms were settled and the Rasch solution was codified. Unless it confronts this failure, it cannot survive as a scientific enterprise. It will remain only as a clearinghouse of non-measures, reproducing the very error it was supposed to correct.

AXIOMS AND THEOREMS OF REPRESENTATIONAL MEASUREMENT IN MANIFEST AND LATENT CLAIMS

At the heart of RMT is the idea of a mapping between an empirical relational system and a numerical relational system. To qualify as measurement, this mapping must be structure-preserving. If objects stand in some empirical relation, say, “longer than” the numbers assigned must preserve that order relation under transformations. The four classical scale types described by Stevens (nominal, ordinal, interval, ratio) are distinguished by the transformations under which structure is preserved. Nominal assignments are preserved under one-to-one relabeling. Ordinal assignments are preserved under monotonic increasing transformations. Interval assignments are preserved under positive linear transformations. Ratio assignments are preserved under similarity transformations, meaning multiplication by a positive constant. These distinctions are not semantic niceties but mathematical constraints: they determine which operations are legitimate. Subtraction, for instance, is meaningful only on interval and ratio scales; division is meaningful only on ratio scales.

The axioms of additive conjoint measurement provide the formal conditions under which interval scales can be constructed. These include axioms such as single cancellation, double cancellation, solvability, and the Archimedean property. Single cancellation asserts that if one ordered pair of attributes exceeds another in a particular dimension, this ordering should hold consistently. Double cancellation provides the stronger condition needed to sustain additivity across dimensions. Solvability guarantees that for any combination of attribute values, there exists a corresponding numerical representation. The Archimedean axiom ensures that repeated concatenations of an attribute do not lead to paradoxes of scale. Together, these axioms support the representation theorem, which states that if the axioms hold, then there exists a numerical scale preserving the empirical structure. The accompanying uniqueness theorem specifies which transformations preserve that representation, thereby defining the legitimate arithmetic.

For manifest claims in HTA, such as “therapy reduces hospital days by ten,” these axioms can be directly applied. Days are a ratio scale: they admit a true zero (zero days), have equal intervals (each day is the same length), and permit proportional comparisons (ten days is twice five). Similarly, weight in kilograms, time in minutes, or drug doses in milligrams meet ratio standards because they satisfy the axioms of extensive measurement, as described by Suppes in the 1950s. Concatenating two one-kilogram masses yields the same as a two-kilogram mass, demonstrating additivity. The uniqueness theorem tells us that multiplying all values by a positive constant (converting kilograms to pounds) does not affect the structure. These attributes therefore provide a secure foundation for evaluable claims: differences and ratios have defined meanings, and claims can be falsified against data.

Latent constructs present a more challenging case. Fatigue, pain interference, or dyspnea severity are not directly observable. We infer them from responses to items with ordered categories. Without a measurement model, those responses are at best ordinal: "always" is more than "sometimes," but we do not know by how much. Summing such responses produces ordinal totals, which cannot sustain the axioms of additivity or invariance. This is where Rasch measurement provides the essential bridge. Rasch proposed in 1960 a probabilistic model in which the probability of a given response depends solely on the difference between person ability (or possession of a latent trait) and item difficulty, both expressed on the same logit scale. If data fit this model, then the conditions of unidimensionality, additivity, and invariance are satisfied.

The Rasch model embodies the conjoint measurement axioms in probabilistic form. Ordered thresholds ensure that categories function monotonically. Fit statistics test whether responses adhere to the model's assumptions. Differential item functioning (DIF) analysis ensures invariance across groups: if items behave differently for men and women, the invariance axiom is violated and the item must be removed or adjusted. When the data fit, the representation theorem holds: we have an interval scale of logits mapping the empirical relations of item responses to numerical distances. The uniqueness theorem specifies that only linear transformations of the logits are admissible, preserving interval properties. Unlike ordinal sums, Rasch-calibrated logits support subtraction (differences in severity) and legitimate effect size calculations.

Wright's contribution in 1977 was to make explicit that Rasch measurement is not just another statistical option but the only framework aligned with representational measurement theory. Other item response theory (IRT) models are flexible, allowing parameters to vary to fit data, but they do not guarantee the preservation of axioms. Rasch insists that data must fit the model, not the reverse, and only then are interval properties secured. This makes Rasch uniquely suited for latent constructs in HTA: it transforms ordinal item responses into interval scales, enabling claims about therapy impact to be evaluable and falsifiable.

In practice, this means that for each latent construct trait of interest, such as fatigue, dyspnea, or depression, items must be written to reflect only that trait, tested for fit to the Rasch model, and calibrated on a logit scale. Each patient then has a possession score indicating their position on the continuum. Therapy impact can be assessed as a change in possession, expressed in logits, which can be compared across groups and evaluated statistically. The claim that a therapy reduces fatigue severity by 0.5 logits is meaningful because it rests on an interval scale anchored in the axioms of measurement.

The lesson is clear: the axioms of RMT apply universally. For manifest attributes, ratio scales can be confirmed by direct empirical checks. For latent attributes, Rasch ensures conformity to the axioms. In both cases, the result is the same: numbers that are legitimate measures, supporting claims that are credible, evaluable, replicable, and falsifiable. Any departure from these standards—whether summed scores, composite indices, or preference utilities, produces numbers that look like measures but are not. They cannot sustain arithmetic, cannot generate falsifiable claims, and therefore cannot serve as the foundation of science.

Central to RMT is the necessity of unidimensionality. A measure must vary along a single continuum. Length is measurable because all rods vary in one dimension of extension. Time is

measurable because all events can be ordered along a single line. Pain, if it is to be measured, must vary along a single latent trait of intensity. Multiattribute composites that collapse different domains, pain, mobility, anxiety, social functioning, are not measures, because no unidimensional continuum underlies them. They are indices, and unless proven otherwise, they remain ordinal at best.

This is the first lesson ICHOM ignores. By treating disease-specific indices as if they were measures, ICHOM collapses attributes without unidimensionality. It mistakes consensus for science.

ICHOM'S STANDARD SETS: CONSENSUS WITHOUT MEASUREMENT

ICHOM presents its work as the culmination of expert consensus. For each disease area, panels of clinicians, researchers, and sometimes patients are convened to decide what outcomes matter most. From a list of possible instruments, those judged most relevant, feasible, and familiar are chosen. The process is consultative, inclusive, and authoritative. But it has no scientific warrant.

Consider what is missing. No disease-specific set is filtered for measurement legitimacy. No Rasch analyses are conducted to test unidimensionality. No evaluation is made of invariance across populations. Instead, instruments are selected for face validity or precedent. The EQ-5D is frequently recommended as a generic anchor, despite the fact that it is a multiattribute index failing every axiom of RMT. Disease-specific quality-of-life questionnaires are adopted despite being simple sums of ordinal items, with no transformation, no calibration, and no interval property. The result is a patchwork of indices, ordinal scores, and non-measures paraded as standards.

The irony is striking. ICHOM claims to be setting global "standards." Yet its process ignores the only standards that matter: the axioms of measurement. A standard meter is defined by a physical property; a standard set of outcomes, if it were scientific, would be defined by adherence to measurement theory. ICHOM's standards are defined instead by consensus panels. This is not standardization but institutionalization of non-science.

The consequences are severe. Data collected under ICHOM protocols cannot sustain scientific claims. Scores are not comparable across populations. Changes cannot be interpreted as real differences on a continuum. International benchmarking becomes a comparison of non-measures, giving the illusion of comparability where none exists.

WHY ICHOM SUCCEEDS WITHOUT MEASUREMENT

Why has ICHOM been so successful despite its lack of measurement rigor? The explanation is sociological rather than scientific. Policymakers crave simple metrics. Hospitals want benchmarks they can display on dashboards. Funders demand comparability across institutions and countries. ICHOM delivers all of this, not by producing measures, but by producing numbers that mimic the appearance of measures.

The strategy is rhetorical. Outputs are neatly quantified, easily tabulated, and readily slotted into performance reports. They convey the look and feel of scientific precision, even though they rest

on ordinal composites that fail the axioms of measurement. This appearance alone is enough to persuade policymakers and funders that they are dealing with science, when in reality they are dealing with consensus-based indices.

The appeal mirrors the rise of the QALY in HTA: a deceptively simple number that can be multiplied, averaged, and compared across diseases. Its very portability makes it powerful, even as its scientific legitimacy collapses under scrutiny. Whether it qualifies as a measure is irrelevant to those who deploy it. What matters is its rhetorical utility; its ability to generate apparent comparability and support decision-making. This is why ICHOM has thrived: not because it produces truth, but because it produces numbers that are useful. Its success is therefore sociological, not scientific; a triumph of appearance over substance.

CONSEQUENCES OF THE ABSENCE OF MEASUREMENT

The absence of measurement in ICHOM's approach is not a harmless technicality. When decisions are built on ordinal tallies masquerading as interval metrics, the cascade of harms is predictable, immediate, and large.

First, policy becomes theater. Governments and health systems that base reimbursement, procurement, and strategic priorities on ICHOM-derived "benchmarks" are effectively signing off on decisions that lack empirical warrant. Resources are diverted not to interventions proven to change measured states, but to those that merely nudge unreliable indices. That is not neutral inefficiency; it is misplaced spending that denies other effective care. In constrained health systems, every misallocated dollar is a patient turned away, a delayed surgery, a drug not funded. This is moral harm, not abstract error.

Second, hospitals are tied to by their own dashboards. Administrators who believe they are comparing performance across units, regions, or countries are comparing ordinal sums that do not support subtraction, averaging, or meaningful change. A hospital "improvement" of three points may be statistically meaningless, an artifact of measurement noise, or a shift in one trivial item rather than a real change in patient health. Management choices made on these spurious signals, staffing shifts, penalty regimes, public reporting can punish competent clinicians and reward smoke-and-mirror improvements. Careers, reputations, and livelihoods hang on numbers that have no scientific teeth.

Third, patients are misled and harmed. When clinicians report "better outcomes" because an endorsed questionnaire's summed score rose, patients are fed a false narrative of progress. They may be exposed to continued or intensified interventions on the assumption of benefit; conversely, truly helpful therapies may be discontinued because the noisy index failed to show a change. Informed consent collapses into ritual when the metrics used to justify treatment choices do not measure what they claim to measure.

Fourth, research and innovation are warped. Trials that use non-measures as endpoints produce results that cannot be aggregated, compared, or meaningfully interpreted. Meta-analyses become exercises in counting incompatible things. Promising interventions may be abandoned because they fail to move an index that was never a valid target; worthless interventions may be adopted

because they produce nominal index gains through gaming, framing, or cultural bias. Scientific learning stalls; false positives proliferate; opportunity costs compound.

Fifth, the global credibility of value-based healthcare is at stake. International comparisons that look precise in tables and reports are statistical illusions. Policymakers in low- and middle-income countries who import these "standards" risk institutionalizing cosmetic metrics rather than building measurement capacity. The rhetoric of comparability becomes an instrument of epistemic colonialism: exporting convenient, non-evaluable tools that lock in dependence on external validation rather than local scientific development.

Finally, there is legal, ethical, and reputational risk. Institutions that present pseudo-measures as evidence expose themselves to lawsuits, audits, and public scandal when the fog lifts. Funders and citizens will rightly ask: on what basis were these life-and-death allocations made? The answer — "on consensus indices that were never measures" is a confession that will not comfort those harmed.

CONCLUSION: REBUILDING STANDARDS OF MEASUREMENT

ICHOM has secured global influence. Its "standard sets" are cited across continents, embedded in registries, and promoted as the benchmark for value-based care. But influence is not legitimacy. What ICHOM has spread is not measurement but its absence. By bypassing representational measurement theory and ignoring the Rasch model, it has institutionalized pseudo-measurement on a global scale. The result is not science but the codification of error.

The solution cannot be tinkering at the margins. The only remedy is reconstruction on the foundations of measurement. For manifest constructs, this means ratio scales: days alive, units of therapy, meters walked, grams of HbA1c reduced. For latent constructs, this means Rasch-based logit ratio scales, built under strict unidimensionality, tested for invariance, and calibrated to interval standards. These are the only admissible forms of evidence. They exhaust the legitimate options. Everything else fails the axioms and collapses under scrutiny.

ICHOM must face this truth. Its standard sets are not standards at all; they are conventions of convenience, harmonizing noise rather than measurement. They offer comparability only in the illusionary sense of aligning errors. This is not global progress; it is the globalization of pseudo-science. Unless ICHOM pivots to ratio and Rasch, it will remain what it is today: a clearinghouse of non-measures, the institutionalization of absence, a façade of science built on consensus rather than truth.

The path forward is demanding but obvious. Rebuild on the only foundations that can sustain evaluable, replicable, falsifiable claims. Anything less is not reform but surrender. The alternatives for ICHOM, survival or collapse, are the subject of Part II.

ACKNOWLEDGEMENT

Portions of this paper were drafted and edited with assistance from ChatGPT (version 5; OpenAI). The author reviewed, verified, and refined all AI-assisted text and assumes full responsibility for the accuracy, integrity, and originality of the final content.

REFERENCES

[1] [1]Krantz D, Luce R, Suppes P, Tversky A**.** *Foundations of Measurement*, Volumes I–III. New York: Academic Press, 1971 (Vol. I); 1989 (Vol. II); 1990 (Vol. III).

[2] Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research, 1960. (Expanded ed., Chicago: University of Chicago Press, 1980)

[3] Wright B. "Solving Measurement Problems with the Rasch Model." *Journal of Educational Measurement* 14, no. 2 (1977): 97–116

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

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

[6] Suppes P. A Set of Independent Axioms for Extensive Quantities. *Portugaliae Mathematica* 1951; 10: 163–172

[7] Luce R, Tukey J. Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement. J *Math Psychol. 1964;* 1(1): 1–27

[8] Center for the Evaluation of Value and Risk in Health (CEVR). *The Cost-Effectiveness Analysis (CEA) Registry*. Tufts Medical Center

[9] MacKillop E, Sheard S. Quantifying Life: Understanding the History of Quality-Adjusted Life-Years (QALYs). *Social Science Medicine.* 2018; 211: 359-366