MAIMON WORKING PAPER No. 22 NOVEMBER 2023

THE UNIVERSITY OF WYOMING CERTIFICATE PROGRAM: A NEW START IN HEALTH TECHNOLOGY ASSESSMENT

Paul C Langley Ph.D., Adjunct Professor, College of Pharmacy, University of Minnesota, Minneapolis, MN and Lecturer, School of Pharmacy, University of Wyoming, Laramie WY

BACKGROUND

Health technology assessment occupies a unique place in the social and physical sciences: costeffectiveness value claims to support formulary decisions are based on assumption driven modelled simulations which are both false and imaginary. Under the commitment to approximate information the notion of demarcation and falsification are put to one side [1]. For the first time since the scientific revolution of the 17th century we are asked to reject hypothesis testing in favor of non-evaluable claims [2]. The standards of normal science are irrelevant as is any commitment to the standards of fundamental measurement, notably for patient reported outcomes. As a discipline, if this is the correct term, what has been described as the technology assessment meme, is an analytical dead end [3]. There is no commitment to the discovery of new facts in therapy response; indeed, it is actively discouraged by the thought leaders in the subject area and not even considered by the thousands of those who subscribe to this belief system.

For those who subscribe to the standards of normal science and fundamental measures, to a belief that science is concerned with progress and the challenging of existing claims, the demise of the health technology assessment (HTA) belief system is long overdue. A reasoned critique is required to point to the manifest deficiencies of this belief system and the need for a new start in health technology assessment. This is the purpose of the Wyoming Certificate program.

THE CHALLENGE OF THE NEW START

It is not clear why the current belief system in HTA technology assessment has endured for so long when it clearly fails to meet the standards for scientific enquiry that are accepted by the physical sciences and other social sciences. The most charitable explanation is that it reflects a lack of commitment in professional training including understanding measurement theory and the standards of normal science. A less charitable interpretation is that it is an easy way out. Why engage in a long-term research program to assess therapy impact, building on the results of pivotal clinical trials, which address issues of replication and reproduction of clinical claims, in subjective instrument response, when an assumption driven simulation with imaginary non-evaluable claims can be sold to an audience with little ability or interest in challenging the analytical framework? If this is the case, does it, at a more fundamental level, reflect a lack of concern or indifference towards the truth of statements or even a willful blindness, an avoidance of the known truth, in the advocacy of approximate information? [4] [5]

Good science is hard, but the answer is not to avoid the challenge, taking refuge in assumption driven non-evaluable simulations to create approximate information that is not only false but risible when proposed as a viable basis for formulary decisions. It is not a question of making the case, which is self-evident, that Rasch measurement is more instructive in decision making than

assumption driven simulated claims and patient reported outcomes (PRO) instruments that produce only ordinal, composite scores, but the fact that to attempt any such comparison is clearly nonsensical [6]. Rasch measurement is unique in providing the necessary and sufficient framework for PRO response to therapy claims [7]. This is a major theme in the Certificate Program. At the same time, as the Certificate Program makes clear, the notion of a quality adjusted life year (QALY) where preferences are composite and ordinal, is a mathematically impossible construct [8]. This provides a needed counterpoint to belief in and modelled applications of QALYs and the literally thousands of false QALY-based claims that have been published.

The challenge for those who support and believe in the current health technology assessment meme is to understand why it must be rejected. The purpose of the Certificate Program is to provide the answer, making clear that this rejection is long overdue. Rather than building on false claims and assumptions plucked from the literature, thus ignoring or demonstrating a lack of awareness of the problem of induction, the Certificate Program offers a framework of analysis that focuses on demarcation and falsification of claims; claims that are expressed in terms that meet Rasch measurement standards for single attributes whether the claim is clinical, in terms of PROs or resource allocation and drug utilization, supported in each case by a claims assessment protocol. This provides a meaningful basis for a research strategy to support the discovery of new facts for therapy impact, including meeting perceived evidence gaps, over the life cycle of the product.

A major concern with the current belief in supporting decisions with imaginary claims for costeffectiveness is that it is an open door to false claims; the equivalent to a paper mill that generates Markov models and trusts that they are never challenged even though the claims are deliberately non-evaluable yet support the sponsors product [9] [10]. Any set of assumptions can be justified as 'realistic' as a picture of an unknown future [11].

THE NEW START BASIC PREMISES

The new start is committed to the application and endorsement of the standards of normal science and fundamental measurement. Health care decisions cannot be based on imaginary, assumption driven claims for cost-effectiveness. We have to do better than rely on the multiattribute QALY as a gold standard in creating approximate information; unless the QALY can be demonstrated to have linear interval measurement properties, capturing a single unidimensional attribute, health technology assessment has no claim to relevance. This is impossible. Rather there is a concern that the current standards in health technology assessment encourage a belief in the importance of consciously rejecting the standards of normal science and fundamental measurement [12]. By focusing on disease specific value claims, and rejecting multiattribute generic preferences and quality adjusted life years (QALYs), there is a pressing need to understand the impact of modern or Rasch measurement theory to construct patient reported outcome (PRO) instruments that support meaningful claims for response. Rasch measurement is not new; it was proposed and accepted in the 1950s in education and psychology, but ignored in HTA from the 1980s with the commitment to multiattribute generic instruments and patient reported outcomes that produce nothing but ordinal scores from observations. We have to backtrack; to admit that the commitment to observations rather than measurement has effectively crippled health technology assessment [6]. What was

overlooked, and continues to be overlooked, is that meaningful measurement is based on the properties of interval scales. If this lesson is rejected, then health technology assessment has nothing to say in capturing patient response to therapy. We have to do better.

This new start in HTA rests on three premises:

- All value claims for therapy impact, whether for clinical endpoints, PRO, drug and resource utilization must meet the standards of normal science for credibility, empirical evaluation and replication;

- All value claims must be for instruments supporting single attributes that meet Rasch measurement standards or rules as interval or ratio scores in order to capture response to therapy; and

- All value claims must be supported by a protocol detailing how the claim is to be assessed and reported.

Accepting these premises means a new paradigm in health technology assessment. The current meme has to be rejected. This is not an easy challenge as there are entrenched vested interests in academia, industry and government who have a lot to lose if the current belief system is overturned. After 30 years of belief, thousands of false QALY claims and associated cost-effectiveness claims that litter the published literature will, by default, have to be, in effect, excised. Perhaps the most wide-ranging repercussion will be for patient reported outcome (PRO) instruments as the overwhelming majority (over, probably, 90%) fail Rasch measurement standards. If we are to measure response to an intervention, then a PRO must be for a single attribute with linear, interval and invariant properties. This is only achieved, as the Certificate Program makes clear, with Rasch measurement; standards which have been recognized for over 60 years. Yet these are standards that are ignored in HTA.

PROGRAM MODULAR STRUCTURE

The Certificate Program comprises 14 modules, each addressing a key topic to define the new HTA paradigm. Each module comprises extensive notes (with 85,000 words for the notes overall), an audiovisual presentation and a short true-false and multiple-choice assessment for each module. Successful completion of the Certificate Program requires attaining minimum scores in each module; entry to a module requires successfully completing the previous module.

Modules are grouped into three parts. These are:

PART I:

The four modules in Part I have two objectives. First, to detail the required evidentiary standards for any value claim for product performance in terms of (i) the standards of

normal science and (ii) the failure of assumption driven multiattribute modeled simulations to produce value claims that meet the required standards. This is achieved by deconstructing the recently released CHEERS 2022 Guidance for creating and submitting imaginary costeffectiveness claims to journals (and endorsed by a number of editors) [13]. These first modules represent the theme that underpins the role for a new start in HTA assessment: understanding the importance of demarcating science from non-science, the critical role of Rasch or modern measurement theory to transform observations to measurement and the need to reject assumption driven modelled simulations for imaginary cost-effectiveness claims.

Module 1: Science versus non-science: Understanding the importance of demarcation in the acceptance of value claims

Module 2: Ratio a and interval measures: Appreciating the importance of interval and ratio measures to support value claims

Module 3:Assumptions and Hume's problem of induction; Understanding that assumptions cannot be used to validate modeled value claims

Module 4: CHEERS 2022 - Tenacity of false belief systems in pharmacoeconomics:

PART II

The five modules that comprise Part II of the program focus on the failure of assumption driven modeled simulations in health technology assessment. In the quest for approximate information, to pass the demarcation test: they fail to meet standards for credibility of claims, the ability to be empirically evaluated and replicated in other target patient populations within a disease area. Rasch measurement is a major focus with a proposed new format for reporting Rasch model results to capture the extent to which target patient populations possess a latent construct and the impact of a therapy intervention on the degree of possession [14].

The modules are:

Module 5: Truth is not consensus: Consider whether there is any justification for lifetime modeled claims in formulary decisions

Module 6: Failure of multiattribute generic preference measures: Understand the case for rejecting multiattribute preference measures in value claims for therapies such as the EQ-5D3L/5L

Module 7: The impossible QALY: Understand, despite its acceptance, why the QALY based on ordinal scores must be rejected

Module 8:Impossible value claims: Consider the case for single attribute ratio value claims in formulary submissions

Module 9:Abandoning models in value claims: Consider the circumstances under which modeled value claims are acceptable

PART III

The five modules in Part III of the program set out the standards for establishing and evaluation value claims for therapies in health technology assessment that ensure that they are a firm basis for formulary submissions. Not only must all value claims be presented as single attributes whether for clinical claims, patient reported outcome claims, drug utilization and resource utilization, but they must be supported by an evaluation protocol and, if required, support outcomes-based contracting and ongoing disease area and therapeutic class reviews.

The modules are:

Module 10: Guidelines for value claims in formulary submissions: Introducing a proposed format for therapy value claims that meet required evidentiary standards

Module 11: The patient voice: Introducing the needs-fulfillment quality of life measure for patients and caregivers

Module 12: Selecting PRO claims: Introducing criteria for evaluating measurement standards for disease specific PRO claims

Module 13:Formulary submission guidelines: Proposal for a formulary submission package for value claims and protocols

Module 14:Questions a formulary committee should ask: Questions to address to ensure value claims meet standards of normal science and fundamental measurement

THE MEASUREMENT IMPERATIVE

Central to establishing meaningful value claims for pharmaceutical products and devices is measurement. HTA has been blighted for over 30 years in subscribing to composite measures, both generic and disease specific that are not measures as understood in the physical and other social sciences., Generic measures such as the EQ-5D-3L/5L and algorithms derived from these measures to create so-called preference scores, fail modern or Rasch measurement. As detailed in the Certificate Program these instruments produce only ordinal scores; there was no stated intent to create an interval or ratio measure. The same error occurs in the more recent EQ-Health and Wellbeing (EQ-HWB) instrument [15]. It is a composite ordinal score with no stated intent to meet the required Rasch measurement standards. Despite the apparent years of effort to create thus ‘extension’ to the EQ-5D-3L/5L instruments, it was designed to fail from the absence of any commitment to a unidimensional interval with a transformation to a ratio scale.

As detailed in the Certificate Program, the generic HTA instruments are measurement failures; the implications are interesting because, as detailed also, assumption driven modelled simulations must also fail, not just because they are imaginary assumption driven simulations, but because they include these measurement failures. By extension, the application of incremental cost-per-QALY claims to generate claims for cost effectiveness and the application of sensitivity test probability sensitivity analysis are also measurement failures. This leads to a key question: does a claim for comparative cost-effectiveness have any meaning? The Certificate Program details why this is the case in current use and spells out the conditions required to make empirically evaluable costeffectiveness claims consistent with Rasch requirements.

The failure of the generic measures extends to disease specific instruments. The EORTC stable of cancer-specific measures and the attempt to create a generic cancer instrument all fail Rasch standards [16] They are simple integer summations from Likert responses and are only ordinal measures; they cannot support viable claims for cancer therapy response. As integer score summations are all to common in disease specific instruments, all fail Rasch standards. All we are left with are a handful of instruments, as detailed in the Certificate Program that meet the required Rasch standards. This is why the Certificate Program details how Rasch standards instruments can be developed; a process supported by a range of software packages available for over 30 years.

Finally, a common mistake is to equate item response theory (IRT) with Rasch measurement. The fact is that they are conceptually different with IRT failing to meet the required standards for fundamental measurement [17]. The reason is straightforward: IRT follows classical measurement by fitting the IRT model to the data while Rasch measurement fits the data to the model by establishing the standards required to create unidimensional, linear, interval and invariant measurement. IRT, judged by the unique contribution of Rasch to fundamental or modern measurement, fails to provide a satisfactory framework for evaluating therapy response.

Interestingly, this failure to appreciate the unique nature of Rasch measurement is made clear in a recently released FDA technical specification document for outcomes assessment data submissions [18]. The FDA endorses IRT without any consideration of Rasch measurement. It is, in short, misinformation showing a lack of appreciation of the standards for interval-based measures of response to therapy for clinical trial claims.

MEETING THE REQUIRED STANDARDS IN HTA

It is not the intent of the Certificate Program just to point to the failures in the current HTA belief system but to set the stage for a new HTA paradigm. Part 3 of the program provides a framework for developing formulary submissions that both meet the required normal science and the standards of fundamental measurement. The focus is on formulary submissions where a manufacturer makes evaluable claims for their product. Certainly, data at launch may be limited but this is not an excuse for filling in data gaps with assumptions and imaginary non-evaluable claims. Rather, the new start paradigm focus is on evaluable claims supported by protocols so that these claims, involving for example replication of pivotal trial claims, can be tracked. Other claims may be for PROs as well as compliance and resource utilization. In all cases the protocol drives the assessment. There is no need for imaginary simulations extending over decades which are designed to be supportive of a product, which leads to deliberately manufactured claims, but not for empirical evaluation.

NEXT STEPS

The Certificate Program provides a new start framework for a paradigm in HTA that meets the standards of normal science and fundamental measurement. The structure and content of the Certificate Program have been detailed in this Working Paper. For those wishing to register and take the program, the link is:

https://www.uwyo.edu/pharmacy/resources/certificate-program-a-new-start-in-health-technology-assessment.html

If further information is required on the content of the Certificate Program, Dr Langley can be reached through his Wyoming email plangley@uwyo.edu. For question on the administration of the program and registration issues please send enquiries to Ms. Jen Paintin jpaintin@uwyo.edu.

REFERENCES

[1] Neumann P, Willke R, Garrison L: A Health Economics Approach to US Value Assessment Frameworks – Introduction: An ISPOR Special Task Force Report. ValueHealth. 2018; 21: 119–123

[2] Wootton D. The Invention of Science: A New History of the Scientific Revolution. New York: Harper Collins 2015

[3] Langley P. Nothing to Cheer About: Endorsing Imaginary Economic Evaluations and Value Claims with CHEERS 22 [version 1; peer review: 2 approved]. F1000Research 2022, 11:248 (https://doi.org/10.12688/f1000research.109389.1)

[4] Langley P. Willful Blindness and Value Claims in the Health Technology Assessment Meme: Some considerations. Maimon Working Papers. No. 16 September 2023 https://maimonresearch.com/wpcontent/uploads/2023/09/Maimon-Working-Paper-No-16-Sept-2023-V2.pdf

[5] Langley P. Bullshit with Scientific Pretensions: Assumption driven simulated claims in health technology assessment. Maimon Working Papers No. 20 October 2023 https://maimonresearch.com/wpcontent/uploads/2023/10/Maimon-WP-No-20-October-V3.pdf

[6] Wright B, Linacre J. Observations are always ordinal; measurements, however, must be interval. Arch Phys Med Rehabil. 1989; 70(12):857-60 https://www.researchgate.net/publication/20338407_Observations_are_always_ordinal_measurements_however_must_be_interval/link/5563b02408ae9963a11ef326/download

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

[8] Langley PC and McKenna SP. Measurement, modeling and QALYs [version 1; peer review: 2 approved]. F1000Research 2020, 9:1048 https://doi.org/10.12688/f1000research.25039.1

---

[9] Langley P. The Challenge for Health Technology Assessment: Paper Mills, False Claims and the Endorsement of Imaginary Claims. Maimon Working Papers No. 14 August 2023 https://maimonresearch.com/wp-content/uploads/2023/08/Maimon-Working-Paper-No-14-August-2023.pdf

[10] Ritchie S. Science Fictions: How fraud, bias, negligence and hype undermine the search for truth. New York: Henry Holt, 2020

[11] Langley P. Facilitating bias in cost-effectiveness analysis: CHEERS 2022 and the creation of assumption-driven imaginary value claims in health technology assessment [version 1; peer review: 3 approved]. F1000Research 2022, 11:993 (https://doi.org/10.12688/f1000research.123709.1)

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

[13] Husereau D, Drummond M, Augustovski F, et al.: Consolidated health economic evaluation reporting standards 2022 (CHEERS 2022) explanation and elaboration: a report of the ISPOR CHEERS II good practices task force. Value Health. 2022; 25: 10–31

[14] Langley P. Enhancing the Rasch Response Model for Value Claims: Latent Trait Possession and Formulary Evaluations [rev]. Maimon Working Paper No. 21. October 2023 https://maimonresearch.com/wp-content/uploads/2023/11/Maimon-Working-Paper-No.-21.pdf

[15] Langley P. After the QALY: Measurement and the road not taken (Part 1: The EQ-HWB). Maimon Working Paper No. 8, June 2023 https://maimonresearch.com/wp-content/uploads/2023/11/MaimonWorking-Paper-No.-8-Part-1-1.pdf

[16] Langley P. After the QALY: Measurement and the road not taken (Part 2: The QLU-C10D instrument. Maimon Working Paper No. 9. June 2023 https://maimonresearch.com/wpcontent/uploads/2023/11/Maimon-Working-Paper-No-9-Part-2.pdf

[17] Stemler S, Naples A. Rasch Measurement v. Item Response Theory: Knowing when to cross the line. Practical Assessment, Research and Evaluation. 2021;26(26):Article 11 https://scholarworks.umass.edu/pare/vol26/iss1/11/

[18] US Department of Health and Human services. Food and Drug Administration. Center for Drug Evaluation and Research (CEDER). Submitting Clinical Trial Datasets and Documentation for Clinical Outcome Assessment Using Item Response Theory. November 2023. https://www.fda.gov/media/173587/download