MAIMON WORKING PAPERS No. 18 SEPTEMBER 2023

HEALTH TECHNOLOGY ASSESSMENT: THE FALSE CLAIMS PSEUDOSCIENCE

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

Abstract

The intriguing feature of health technology assessment (HTA) is that for over 35 years it has promoted in textbooks, practice guidelines and teaching a commitment to a belief system or meme that is based on creating false claims to support pricing and formulary acceptance of pharmaceutical products and devices. This has been achieved by a focus on a single claim objective in HTA: cost-effectiveness. To achieve this HTA has promoted the creation of false claims from assumption driven simulations with the quality adjusted life year (QALY) as a key parameter. The problem is the neglect, or lack of knowledge, of fundamental measurement and the standards for value claims where each claim must be for a single attribute with unidimensional, linear, interval and invariant properties. This commitment to creating false claims stands in contrast to the question of false claims in science and social science where false claims have been described as the replication and reproduction crisis. Research is faked, data sets are made up and claims made which are entirely imaginary. While science rests on replication, the provisional confirmation of previous results, the presence of false claims effective derails this process; even to the extent of reproducing claims from the original data set. The purpose of this note is to examine the commitment in HTA to false claims where these are not the provenance of a few, isolated individuals, supported by the activities of paper mills and model builders for marketing, but a deep-seated belief in the merit of false claims where evidence for cost-effectiveness is created.

INTRODUCTION

Deliberate false claims in science and some social sciences, notably psychology, have attracted the opprobrium they well deserve. Set alongside the focus on objective knowledge, of progress and discovery, deliberate falsification must be seen to be the antithesis of the commitment to the provisional acceptance of empirically evaluable claims that adhere to the standards for demarcation; to separate sense from the nonsense of pseudoscientific truths [1]. When we turn to health technology assessment (HTA) we face a more important distinction between the practice by a minority of deliberately fomenting false claims within a discipline which recognizes the standards for normal science and fundamental measurement and a discipline that is essentially founded on the rejection of normal science and fundamental measurement, endorsing the fomenting of false claims. HTA is clearly in the latter camp, occupying a unique position where practitioners reject the standards of normal science and fundamental measurement. Given this, it seems more reasonable to describe HTA, not as a discipline but a meme or belief system that, from a relativistic perspective, denies there is a unique scientific method [2] [3]. A meme that denies the possibility of false claims. HTA exists from the belief that truth is consensus and that evidence is never discovered but constructed within the HTA community [4]. The success of HTA, as with intelligent design and the various religions, rests on its ability to mobilize the belief of a community by rhetoric, persuasion and authority. There can never be an appeal to superior evidence.

The purpose of this brief commentary is to point to the long-standing commitment, not to false claims, in the usual context of manufactured science fictions, or the activities of participants to support bias in the promotion or suppression of false or countervailing claims, but to the promotion of assumption driven simulations as the gold standard for formulary evaluations. A gold standard which puts false claims front and center for pricing and patient access. A commitment to a meme that denies normal science and fundamental measurement, where evidence is created rather than discovered.

THE REPLICATION CRISIS IN HTA

It is widely accepted that across a number of disciplines there is a failure to replicate study findings; disciplines ranging from psychology to neuroscience imaging, economics, evolutionary biology and organic chemistry [5]. This is probably the tip of the iceberg as, when set against the sheer volume of published research, particularly with human subjects, few replications are attempted. The implications are profound; to what extent can we trust published studies even though they have passed peer review and the ministration of publication bias by both reviewers and journal editors with the known bias to favor positive results. Average replication rates across journals have been proposed as a maximum of 1%.

In this context it is important to distinguish between replication and reproduction. The former refers to a different data set where the same questions are asked while the latter refers to reproducing the same results from the same data set. This raises an important question: should reproduction precede replication. After all, if there is doubt that the original claims can be reproduced from the original data, where in all too many cases the original data set cannot be resurrected, then why proceed? Reasons for this vary including poor record keeping, poor protocols and even reluctance on the part of authors to support a re-evaluation. If reproduction is impossible, and even where it is possible there are only a few reported efforts that provide confidence, then reproduction seems a waste of time. The study claims should be ignored.

There is no replication or reproduction crisis in HTA. The reason is that the discipline supports and endorses false claims. This is not a question or research practice and the incentives to produce false claims but the failure to apply critical standards in the creation of claims; claims for incremental cost per quality adjusted life year (QALY) and cost-effectiveness which are false. This follows from a lack of interest, amounting to a rejection, of the standards of normal science in proposing, evaluating and replicating or reproducing value claims together with a further lack of interest or misunderstanding of measurement. The charge comes down to a simple statement: all value claims for outcomes in HTA must be for single attributes which are unidimensional, linear, interval and invariant [6].

Instead, we have a commitment to assumption driven simulations to produce non-evaluable and false cost-effectiveness claims (including the impossible QALY) and, in patient reported outcomes, a focus on integer scores as the basis for claimed responses to therapy interventions; integer scores that have only ordinal characteristics [7]. There is a replication and a reproduction crisis, not by the activities of individual participants in their quest to create false outcomes, but a systemic failure in the embrace of methodologies that inevitably produce false outcomes.

A question that is easily resolved is whether assumption driven simulations creating imaginary cost-effectiveness claims should be seen as no different from those invented claims based on constructed patient responses to non-administered questions. The answer is that they are both creating false (or imaginary) data to support false value claims. The only difference is that the false claims of simulations are non-evaluable by construct while the data created from false patient records are designed to support a testable proposition but one that has been designed not to be falsified, again by construct. As there is no real distinction between these two approaches to inventing data to support value claims then these are best characterized, as Ritchie describes it, as science fictions [5] .

QUALITY ADJUSTED LIFE YEARS (QALYs)

The systemic failure to recognize the critical role of credible interval measures to capture health status and response to therapy is shown in the espousal of multiattribute preference and utility scores from instruments such as the EQ-5D-3l/5L. These epitomize the failure to create interval measures that meet modern or Rasch measurement standards. First developed in the mid-1970s, the belief that it is possible to combine various clinical attributes into a single metric fails because a credible measure must relate to a single attribute [8]. Certainly, it may be possible to combine single attributes into a composite measure, but the precondition is that each attribute has been transformed from counts or observations into an interval or ratio measures manifesting a latent construct [9]. This was not done; instead, health state descriptions involving combination of response levels to a handful of clinical dimensions were evaluated to produce a community score capped at unity applying standard gamble or time trade off techniques to a sample of a bundle of symptoms. Unfortunately, applying these techniques yielded only an ordinal preference or utility score which gave both positive and negative values; the symptom responses for each clinical dimension were also on an ordinal scale. This made it impossible to put any meaning on the difference between these scores and when the weights for each basket of responses were patched into an algorithm to create an overall score, which was manipulated to try and fit to the data, this in turn only created a composite ordinal score. Capped at unity to define perfect health (for 5 health dimensions), worsening health was captured by decrements from unity. While there may have been a naïve belief that the decrements would remain in the range zero to unity, with zero intertied as death defined by the set of symptoms, the algorithm actually yielded negative values or states worse than death defined as negative ordinal scores.

The lack of success in multiattribute scores is compounded by their application to create QALYs. Again, ordinal scores cannot support the standard arithmetical operations; multiplication to get a discounted time spent in a disease state; the task is impossible [10] [11]. This is unfortunate because the assumption driven simulation is founded on QALYs with cost-effectiveness defined in terms of QALYs. As these are impossible, the cost-effectiveness claim, which is multiattribute with an absence of linear, interval and invariant properties, is a false metric. It also fails the standards for normal science. Certainly, the results could be reproduced, but this involves the trivial task of re-running a software package. Replication is a wasted effort because there is nothing to replicate; it is a one-off software exercise.

FALSE CLAIMS AND THE PATIENT

The principal reason for simulation models is to put to one side the tedious process of developing and testing interval claims, hypothesis testing, in favor of assumption driven simulations creating approximate information, driven by claims for realistic assumptions, to support formulary decisions. In other words, a conscious decision, particularly when robust data are limited at product launch, to create data to short cut the evidence development process. This is not approximate information as there is no reference point for an approximate assessment; false claims are the currency. A situation identical to the creation of false claims by evidence creation from the products of paper mills and invented data bases from the less scrupulous academic and research teams.

The problem for HTA is no different from that in other areas where validated false claims are presented; we have no basis on which to judge the merits of a value claim. Certainly, there are efforts to assess the extent of bias in individual study results as well as study selection for systematic reviews, but this can only go so far. Unless there is robust evidence, first for reproduction and second for replication we are forced to a default position of disbelief. As the HTA meme, with its deliberate avoidance of the standards of normal science and fundamental measurement, must be judged to produce only false, not approximate,claims, there is little scope to judge competing products as they are proposed for formulary acceptance following marketing approval. Approval which itself is based on minimal pivotal trial data with no requirements for replication let alone reproduction.

CONCLUSIONS: THE COMMITMENT TO FALSE CLAIMS

If HTA practitioners see themselves, with their simulated assumption driven claims, as the saviors of formulary committees, then they a sadly mistaken. They are only promoting the advocacy of false claims. The potential implications for patens are significant. Despite favorable clinical claims from pivotal trials, which have been judged acceptable, an assumption driven simulation can still result in false recommendations for pricing, access and budget impact. The problem for patients and their advocates in target populations is that there is no awareness that what they might be protesting is simply a catalogue of false claims. The primary reason why HTA has been able, for 35 or more years, to ignore fundamental measurement is that few QALY critics are even aware of Rasch standards, let alone the deficiencies of Markov modelling. The fact that any valid claim must rest on a credible interval measure is a foreign concept. The unique importance of Rasch modelling as the necessary and sufficient means to transform ordinal scores to interval and ratio measures is not so much as disregarded but a requirement for interval measurement to which they have never been exposed [12] [13] .

To recognize the unique role of Rasch modelling and interval or ratio measurement as integral to HTA is, to use an overblown metaphor, opening Pandora's Box (a.k.a 'opening a can of worms'). It is the first step to a realization that HTA is characterized by deliberately promoted techniques to create false claims which fail standards for falsification. . Rasch is not mentioned in the leading textbook; there is no mention of the need to create single attribute, linear, interval and invariant measures to support evaluable therapy claims [14]. Rather, the focus is on creating assumption driven simulations to prove the case for false claims. The notion of discovery and progress is completely

absent; there is no objective reality to challenge. All that is needed is to populate a Markov framework with assumptions, notably the impossible QALY, to make false cost-effectiveness recommendations. This message continues with the CHEERS 2022 guidance for submitting assumption driven modelled claims for cost-effectiveness [15] .

ISPOR is no different with its Good Practices Reports which are intended to provide expert consensus guidance recommendations to set international standards for HTA and its application in decision making. The effect is somewhat different with the absence of any reference to the standards of normal science and fundamental measurement [16]. The importance of interval and ratio scales, let alone the Rasch model for transforming ordinal to interval or ratio single attribute measures is never mentioned . There is no appreciation that classical test theory requires a Rasch standard interval or ratio measure; ordinal scales are treated as if they had interval or ratio properties.

The HTA meme rests on false and unsubstantiated foundations. Unlike other disciplines, it is a meme designed to create false claims The belief is so well entrenched that is difficult to see if the meme can change to recognize standards for normal science and measurement. Advocates will cling to the meme; false claims will be canonized. There will be no progress and discovery of new yet provisional facts in HTA. Hopefully, with continued criticism pointing to the absurd endorsement of false facts in HTA, eventually, the 35-year commitment to pseudoscience and unthinking trust in assumption driven simulations, will evaporate.

REFERENCES

[1] Pigliucci M. Nonsense on Stilts: How to tell science from bunk. Chicago: University of Chicago Press, 2010

[2] Dawkins R. A Devil's Chaplain. New York: Houghton Mifflin, 2004

[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

[4] Wootton D. The Invention of Science: A new history of the scientific revolution. New York: HarperCollins, 2015

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

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

[7] Langley P. Medication adherence, Rasch measurement and the Morisky MMAS-8 instrument. *Maimon Working Papers* No 15 August 2023 https://maimonresearch.com/maimon-working-paper-no-15-august-2023

[8] McKenna S, Heaney A. Composite Outcome measurement in clinical research: The triumph of illusion over reality? *J Med Econ*. 2020;23(10): 1196-1204

[9] Andrich D, Marais I. A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences. Singapore: Springer, 2019

[10] Langley P. The Great I-QALY Disaster. *InovPharm*. 2020; 11(3): No 7

[11] Langley P. McKenna S. Fundamental Measurement: The Need Fulfilment Quality of Life (N-QOL) Measure. *InovPharm*.2021;12(2):No. 6

[12] Merbitz C, Morris J, Grip J. Ordinal scales and the foundations of misinference. *Arch Phys Med Rehabil*. 1989;70:308-32

[13] Wright B, Linacre J. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil.* 1989; 70(12):857-60

[14] 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

[15] Husereau D, Drummond M, Augustovski F et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 22) Statement: Updated reporting guidance for health economic evaluations. *ValueHealth*. 2022;25(1):3-9

[16] Caro JJ, Briggs A, Siebert U, et al. Modeling good research practices –Overview: a report of the ISPOR-SMDM Modeling good practices task force– 1. *ValueHealth* 2012;15:796-80315.