Rejecting False Claims from Markov Simulations in Alzheimer's Disease

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

Dear Editor,

Innovations in Pharmacy have just published an article in the Formulary Evaluations section that applies an assumption driven Markov simulation model to produce false or imaginary cost-effectiveness claims for patients with differing levels of severity in Alzheimer's Disease [1].

According to the article's abstract:

The research aimed to explore the cost-effectiveness of a hypothetical gene therapy for patients with Alzheimer's disease at varying degrees of severity. A Markov model with a 20-year time horizon was constructed for simulated cohorts with mild cognitive impairment due to Alzheimer's disease, assigned to receive either standard of care or a one-time gene therapy administration. Varying costs of care due to disease severity and treatment efficacy were utilized to determine the effect of those variables at different willingness-to-pay thresholds.

Tucked away at the end of the paper in the discussion section is a reference to a paper which presented the case to abandon entirely the current belief system in health technology assessment (HTA) supporting modeled claims for cost-effectiveness [2]. The referenced paper proposed that the current HTA meme should be rejected in favor of a new HTA paradigm that recognized the role of the standards of normal science where all value claims should be credible, empirically evaluable and replicable with fundamental or Rasch measurement ensuring all value claims must be for single attributes or unidimensional with linear, interval and invariant properties

The statement of interest is:

Finally, there are concerns regarding the mathematical validity of QALY as a utility estimate and the use of an assumption-based modeling framework to provide economic evaluation of novel therapeutics. While our analysis conformed to the common methodology currently utilized in cost-effectiveness analyses, we acknowledge the importance of challenging the limitations of typical approaches to health technology assessment, including the identification and estimation of key parameters and future prediction solely based on past observations. Original and innovative approaches in health technology assessment are needed to address these shortcomings and they deserve increased attention from researchers and key decision makers.

This extract mischaracterizes the position placed for a new start in health technology assessment claims. It suggests that while there might be concerns with the current meme or belief in creating assumption-driven imaginary modelled claims, we should stay with the present methodology, while looking to 'original and innovative approaches'. The problem is that there is no place in science or social science for the current commitment to assumption driven simulations which ensure the creation of false claims for pharmaceutical products and devices. It is not a question of remediating shortcomings but of rejecting the practice of creating false claims to support formulary decisions. We know the required standards as they have been in place since the scientific revolution of the 17th century; it is up to practitioners in HTA to reject their commitment to false claims and join the mainstream in the physical and social sciences [3].

We have known for some 60 years that for a measure of response we need a calibrated interval or ratio scale [4 5]. This position was formalized by Rasch in the 1960s and has been used globally as the basis for measurement [6]. There have been ample red flags over the past 40 years for the application in PROs, warnings not to abandon interval scales. These have been ignored with the focus on quality adjusted life years (QALYs) driven by the construction of community health state ordinal composite utilities and preferences.

As such, composite ordinal scales fail as a measure, there is no justification for their support for non-evaluable claims for cost-effectiveness as the principal guide to resource allocation in health systems.[7]. The result, as in the present case for Alzheimer's Disease, is that ISPOR's claimed cost-per-QALY estimates and recommendations for pricing are meaningless. It is a puzzle why these models are willingly accepted by journal editors.

A question that is easily resolved is whether assumption driven simulation creating imaginary cost-effectiveness claims should be seen as no different from those fraudulent 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 unsubstantiated value claims. The

Correspondence: Paul Langley, PhD
University of Minnesota
College of Pharmacy
Email: langley@maimonresearch.com

DOI: https://doi.org/10.24926/iip.v14i3.5701

only difference is that the 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 [8].

The need to formally construct an interval scale is a critical point because unless there is an interval measure there is no basis for traditional or classical statistical analysis [9]. This is seen, as noted in the case of the QALY, where time spent in perfect health is discounted by an ordinal preference score. Ordinal scores can only support non-parametric statistics; not multiplication. As the QALY is impossible, then the entire modelling exercise collapses and the QALY and consequent incremental cost-per-QALY claims and thresholds are meaningless. It is not a question of a challenge to this situation but of pointing out, that it collapses from its own manifest deficiencies in failing to understand, or be aware of, the required standards for fundamental measurement.

The stand for measurement is clear: a credible measure must have single attribute, linear, interval and invariant properties. This means that when we start from counts or observations (ordinal scores) the Rasch model is the necessary and sufficient condition to transform these to interval and ratio measures. The QALY, typically based on scores (preference or utilities) created by multiattribute instruments creating a composite ordinal score, is a failure; we cannot create an interval score from composite multiattribute counts or observations. It is false to assume you can [10].

The authors also write that there should be a challenge to 'the identification and estimation of key parameters and future predictions solely based on past observations '. This is not an option for a challenge. There is a simple point of logic called the problem of induction, first raised by David Hume in 1748 [11] -- the fact that past futures have resembled past pasts does not mean that future futures will resemble future pasts [12]. This means that we have to reject 25 years of hypothetical claims based on a literature search for useful or realistic assumptions. It makes no sense. The only exception is where a value claim is empirically evaluable in a relatively short time frame where, following the demarcation standard, a failure to meet required assessment standards, leads to a rejection and reconsideration of a hypothesis which may involve reconsidering assumptions made as part of the prediction; this is commonplace in the physical and other social sciences.

The emphasis in HTA on assumption driven claims makes it open to false claims; including a judicious choice of assumptions to make a threshold cost-per-QALY claim favorable to the sponsor. Indeed, there is considerable evidence that a high proportion of published industry-sponsored Markov and similar models appear to be deliberately constructed to give results favorable to the sponsor's product [13 14 15 16].

The endorsement of false claims is made worse by the endorsement by leading journal editors to accept such modelled claims (e.g., *BMJ, Journal of Medical Economics, Value in Health, Pharmacoeconomics*) with the launch of the CHEERS 2022 guidance for submitting false modeled claims to journals [17]. A situation which is even further facilitated by editorial publication bias where it appears virtually unheard of for a negative simulated modelled false cost-effectiveness claim to be published. All it needs is to modify a few assumptions. Interestingly, in the UK, NICE employs academic groups who are specialists in choosing assumptions to review submitted modeled simulation proposals; a luxury which, presumably, *Innovations* cannot afford. Even so, it seems a pointless exercise as any model can be re-jigged to give the answer you want; with leaders of the meme or belief system fending off critics.

The bottom line is that these models are trivially easy to create with off the shelf Markov software packages. Their attractiveness, is that none of the claims can ever be empirically assessed. This, again, points to the unique nature of HTA: a win-win for all with no hope of a challenge except by changing assumptions with a claim that a new set of assumptions are more 'realistic' for an unknown future than an alternative collection. Bias is inevitable [18].

On a more positive note: there is a quality of life interval measure (needs fulfillment) that has been developed in Alzheimer's Disease for caregivers [19]. This measure meets Rasch standards and can be applied to evaluate the extent to which needs are fulfilled and how the needs met may change with a new therapy. The measure is interval and can be transformed to a bounded ratio scale with a range 0 – 1, reported in *Innovations* [20 21] . This means, unlike the multiattribute ordinal preference or utility measures which are anchored at unity with utility decrements, that there is no overshooting zero (death) to give negative scores for states worse than death.

As a final point, I certainly don't want to advocate retraction of this article which is the first false claim simulation model published in *Innovations*. It should stand as a warning to those who may be considering *Innovations* as a suitable publication. If published, this letter may accomplish this and bring submissions into accord with the preamble to the Formulary Evaluations section of the journal which asks for standards for normal science and measurement in submitted papers, not imaginary modelled simulations that put the discovery of provisional new facts, to one side. This is not new: if we go to the motto of the Royal Society "nullius in verba", awarded its charter in 1662 (take no man's word for it) it is clear that non-empirically

DOI: https://doi.org/10.24926/iip.v14i3.5701

evaluable imaginary claims would not have got any traction. The concern was to escape the shackles of an Aristotelian philosophy in its denial of progress and discovery; to focus on the advancement of science and natural knowledge. The position that Aristotelian philosophy 'is inept for new discoveries' resonates with the HTA meme [22]; a continued production of non-evaluable modelled claims is not the basis for progress and discovery. Truth is not consensus even if supported by rhetoric, persuasion and authority by leaders in the HTA meme [3].

Disclaimer: The statements and opinions are those of the author.

[1] Kieu C, Look K. Cost-effectiveness of a hypothetical gene therapy for Alzheimer's Disease: A Markov simulation analysis. *InnovPharm*. 2023;14(3): No. 3

[2] 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

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

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

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

[6] Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research, 1960

[7] Langley P. Peter Rabbit is a Badger in Disguise: Deconstructing the Belief System of the Institute for Clinical and Economic Review in Health Technology Assessment. *InovPharm*. 2021; 12(2): No.20

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

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

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

[11] Hume D. An Enquiry Concerning Human Understanding. 1748

[12] Russell B. The Problems of Philosophy. 1912

[13] Xie F, Zhou T. Industry sponsored bias in cost-effectiveness analysis: registry-based analysis. *BMJ.* 2022:377

[14] Langley PC, Rhee T. The Imaginary Worlds of ISPOR: Modeled Cost-Effectiveness Claims Published in Value in Health from January 2016 to December 2016. *Inov Pharm.* 2017;8(2): No. 14

[15] Langley PC, Rhee TG. More Imaginary Worlds: A Systematic Review of the Status of Modeled Cost-Effectiveness Claims Published in the Journal of Medical Economics from January 2016 to December 2016. *Inov Pharm.* 2017;8(3): No. 13

[16] Langley PC. Imaginary worlds: Modeled claims for cost-effectiveness published in PharmacoEconomics January 2015 to December 2015. *Inov Pharm* . 2016;7(2): No. 9.

[17] 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

[18] 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

[19] McKenna S, Bouse M, Heaney A et al. International Development of The Alzheimer's Patient Partners Life Impact Questionnaire (APPLIQue). *Am J Alzheimers Dis other Demen*. 2020;35:153331752095

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

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

[22] Glanville J. The Vanity of Dogmatizing. London: H Eversden, 1661

DOI: https://doi.org/10.24926/iip.v14i3.5701