MAIMON WORKING PAPERS No. 16 JUNE 2022

FACILITATING BIAS IN COST-EFFECTIVENESS ANALYSIS: ASSUMPTION DRIVEN IMAGINARY VALUE CLAIMS IN HEALTH TECHNOLOGY ASSESSMENT

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

Abstract

The current standards for health technology cost-effectiveness assessment rest on the creation of lifetime assumption driven modeled simulations for imaginary pricing and consequent patient access recommendations. A recent paper in the BMJ reports on a detailed assessment of some 8,192 cost-effectiveness analyses and concludes that industry sponsored modeled claims were more likely to publish ICER's below a $50,000 threshold than non-industry sponsored studies, supporting the claim that the product was cost-effective. This is an entirely unsurprising result; indeed, the opposite can occur with a modeled claim deliberately resulting in ICER is excess of $50,000. This methodology is well entrenched with the recently published CHEERS 2022 guidance for creating imaginary cost-effectiveness modeled claim ensuring an open season for deliberately manipulated cost-effectiveness claims. Unfortunately, the BMJ analysis cannot stand because the limitations imposed by fundamental measurement are overlooked. Manipulation of ICERs and claims for cost-effectiveness are impossible. including cost-utility thresholds, because the preference or utilities supporting the creation of QALYs are ordinal sores; the QALY is mathematically impossible. Putting this caveat aside, on its own terms the study points to systematic bias in industry sponsored modeled claims. Acceptance of such claims is self-defeating. As this commentary argues, not only does the approximate information meme fail the required standards for normal science and fundamental measurement, but the meme is self-defeating because of the opportunity for sponsored systematic bias in formulary submission and other cost-effectiveness claims. This meme that facilitates bias has to be rejected in favor of a NEW START paradigm for health technology assessment that focuses on evaluable single attribute value claims, meeting the required standards for normal science and fundamental measurement.

INTRODUCTION

A recent commentary in F1000Research made the case for the rejection of the current standards, the belief system or meme, in health technology assessment in favor of a NEW START paradigm that meets both the standards for normal science and fundamental measurement [1]. Value claims in normal science are required to be credible, evaluable and replicable; while fundamental measurement requires value claims to have interval or ratio measurement properties. The NEW START paradigm for value claims in health technology assessment requires the claim to be for a single attribute, to be unidimensional and to allow empirical evaluation within a meaningful and short timeframe. The NEW START rejects claims for cost-effectiveness as these are composite measures, with the claim resting on assumption driven,

modeled lifetime simulations to produce outcomes that are entirely imaginary, putting to one side standards for normal science and measurement. The current health technology framework fails the demarcation test; it is pseudoscience [2].

Assumptions regarding model structure and parameter values are the core of the current health technology assessment (HTA) meme. The focus is on lifetime models or those that are intended to capture approximate information for therapy response over the natural course of a disease. The models are not designed to create evaluable outcomes, but to simply 'inform' health system decision makers with simulated claims [3]. But more crucial is the fact that as there is no basis in logic for believing claims from the past will hold in the future, there can be any number of competing non-evaluable model claims. It is not a question of model transparency and justifying choice of assumption; the model is just one of a possible multitude of models each based on assumptions drawn from the literature and completed pivotal clinical trials; there is no justification for one set of assumptions than any other.

If value claims for competing products are based on assumption, then there is the obvious incentive, in a competitive environment where formulary listing at a preferred price can have significant financial implications for industry, to propose modeled claims for cost-effectiveness which support the preferred price and formulary placement of the product. This brings in the question of bias and whether or not there is substantive evidence to support claims for manufacturers 'gaming' the system to create favorable cost-effectiveness claims, notably for drugs.

The purpose of this brief commentary is to make the case that by continuing to support the current health technology assessment meme, bias is inevitable. Irrespective of efforts made to police the system with requirements for greater transparency and training for formulary assessors, if a modeled value claim is required to support a specific claim for cost-effectiveness, it will be constructed. Questions of bias will be easily deflected in justification for model structure and data inputs. Proposing guidelines for submitting imaginary modeled claims for cost-effectiveness, exemplified by CHEERS 2022 with associated checklist will do no more than facilitate such endeavors [4]; justifying a continuing belief in model claims rather than rejecting a technology assessment meme which is manifestly deficient, fatally flawed and failing the demarcation test between science and pseudoscience.

THE McMASTER-TUFTS STUDY

A recent evaluation presented in the BMJ, makes the claim that in the universe of modeled cost-effectiveness claims, bias is pervasive in industry sponsored studies [5]. This is, of course, an unexceptional conclusion as there is a wealth of evidence to support claims for bias where industry funded cost-effectiveness claims are likely to report favorable results to the sponsor. The focus of the study is on the Tufts University Cost-Effectiveness Analysis Registry and the reporting of claims for studies published between 1976 and 2021; the majority in the last 10 years [6]. The study identified 8,192 studies, of which 46.5% were for drugs. A range of study categories were identified, but aggregated to industry sponsored as opposed to non-industry. Studies were categorized in terms of disease and methodological characteristics, with incremental cost-effectiveness ratios defined in terms of three cost per quality

adjusted life year (QALY) thresholds: $50,000, $100,000 and $150,000. A total of 8,192 CEAs were evaluated. With 2,437 (29.7%) sponsored by industry. Of these, 90.3% were model based (compared to 89.2% for non-industry studies) with 78.7% of industry sponsored CEAs having an ICER below $50,000 (65.4% for non-industry sponsored).

In the base case logistic regression with 8,192 CEAs, industry sponsored claims were more likely to conclude that the intervention was cost-effective than the comparator with a threshold of $50,000.; an adjusted odds ratio of 2.06 and a 95% CI 1.82 to 2.33. Similar results were reported for the other two thresholds. In terms of magnitude, industry sponsored CEAs ICER was 33% lower than non-industry sponsored (95% CI -40% to -26%). The analysis suggested that the industry sponsored bias was systemic, existing across a wide range of disease and study designs. A further subgroup analysis found that CEAs for drugs accounted for almost three-quarters of industry sponsored studies compared to just over a third for non-industry sponsored studies, with one of the largest sponsorship biases. Perhaps not surprisingly, the analysis found least bias among trial based as opposed to modeled studies.

ACCEPTING PSEUDOSCIENCE

To accept the argument for industry bias in modeling cost-effectiveness claims seems somewhat paradoxical when the basis on which the assessment has been undertaken accepts a meme that fails the standards for normal science and measurement. The Tufts University data base pays no attention, or is ignorant of, the standards for fundamental measurement. Multiattribute utility or preference scores (e.g., EQ-5D-3L/5L) are presented as a fee-based 'help-yourself' emporium but with no understanding that these are ordinal scores. As such they cannot support the universally favored yet mathematically impossible quality adjusted life year (QALY) and the various economic models [7]. Ordinal scores cannot support multiplication; incremental cost-per-QALY (ICER) claims are also mathematically impossible as are cost per-QALY thresholds. The criteria for assessing industry bias; the extent to which industry ICERs squeak under thresholds is an impossible measure. In this important sense, the entire analysis is redundant. Add to this is the failure to recognize that the entire modelling meme is unsustainable; non-empirically evaluable value claims are simply a chimera.

If protagonists are to argue for industry bias, then they must embrace the pseudoscientific basis of approximate information modelled outcomes, including the notion of ordinal cost-effectiveness as a believable metric. If the modeling is rejected, then cost-effectiveness claims disappear; together with the measurement of bias. The ready availability through the Tufts database for access to ordinal preferences or utilities makes, unfortunately, the construction of imaginary claims that much easier. The application of the Tufts data is illegitimate. Certainly, it is possible to apply a detailed regression analysis, but that presumes that all data elements in the model meet required measurement standards; the Tufts data certainly do not.

THE PRICE OF FAILURE

While it is one thing to point out that under the current approximate information, health technology assessment meme it is possible to have a range of competing assumption driven models, it is another to attempt to provide a *prima facie* case that for industry it is a question of how to game the system to generate assumption driven cost-effectiveness claims that are deliberately constructed to meet favorable cost-per-QALY thresholds. A process for modeling and choice of assumptions that is made more straightforward once cost-per-QALY thresholds are established and the objective is to create a cost-effectiveness case that deliberately yields a value just short of the threshold value.

In an important sense, the approximate information assumption driven simulation meme is hoist with its own petard. A standard for health technology assessment for 30 years; it is also a meme that is driven by non-evaluable biased value claims; biased or otherwise. The fact that the standards of normal science for credible, evaluable and replicable claims are rejected merely opens Pandora's Box. There are not the resources and skills, with a failure in training and education, to assess (and reject) these approximate information models. As long as these standards are accepted in education and by health system decision makers and journal editors, as witnessed by the endorsements of the CHEERS 2022 guidance, the opportunity and rewards from 'threshold' gaming will continue. This absence of resources and skills, it might be added, is an added incentive to create favorable CEA claims. It is all very well for agencies such as NICE in the UK and the PBAC in Australia to establish academic review centers, staffed by those who have devoted their professional lives to challenging and modifying assumptions in imaginary lifetime models; but this is only possible if the resources are available and if there is a willingness to engage with industry in what is best seen as a pointless activity.

Of course, even if a concerted effort was made in graduate programs, health systems and even industry to provide a more coherent basis for developing and evaluating model claims, the exercise would be a waste of effort. It is not, to emphasize, the problem of systemic bias, but of a more fundamental objection to the assessment of CEAs and ICERs in modeled imaginary claims. The assessment meme fails the standards of normal science and measurement and should not be subject to an assessment of bias in the first place. The findings are of interest, and indicative of the ease with which models potentially can be deliberately manipulated, but of no interest for the required standards for formulary submissions and the discovery of new facts for therapeutic benefit.

IMPLICATIONS FOR ICER MODELLED CLAIMS

Although the BMJ analysis does not identify industry versus non-industry CEA studies by country, the analysis has some major implications for the modeling by the Institute for Clinical and Economic Review (ICER) in the US. Previous commentaries have pointed out that the ICER embrace of assumption driven simulated model claims defies the standards of normal science and fundamental measurement [8 9]. Yet, with the continued acceptance of the approximate information meme, ICER through its stable of academic consultants, keeps producing imaginary modeled recommendations for pricing, with threshold cost-per-QALY cutoffs, and access to the selected drugs. The BMJ study points quite clearly to the futility of this

approach in that ICER's modeling produces only one or a selection of many possible imaginary value claims for cost-effectiveness and pricing to achieve threshold cutoffs. This is made clear by the commitment by many manufacturers to tailor their modeling to meet threshold constraints. This is not to justify, as in the case of the BMJ study results, a commitment to minimize the opportunities for bias through review or even the support for 'blinded' competing models; this is impossible and unnecessary given the deficiencies of approximate information modeling. What the BMJ study makes clear is that once an approximate information meme is the vehicle for cost-effectiveness claims, health system decision makers should reject those claims. Virtually any value claim, engineered or otherwise, must be rejected. The existence of systematic bias is a salutary reminder of how futile is this approach to value claims; it might be taken one step further to argue that bias is inevitable in any choice of assumptions. The problem with this argument is that the presence and extent of possible bias is unknown as there is no reference point for judging assumptions about an unknown future. Hopefully, the BMJ study will be seen in the US as further evidence for rejecting assumption driven cost-per-QALY simulations where one imaginary study is as good or as bad as another, with no value at all to health system decision makers. There should be no encouragement given to industry to compete with ICER in the imaginary cost-effectiveness claims stakes.

NEW START PARADIGM

As detailed in previous commentaries and proposed formulary submission guidelines, NEW START rejects completely the existing assumption driven approximate information meme [1]. Rather, it starts from the premise that if technology assessment in health care is to be meaningful it must meet the standards for normal science and measurement. Value claims must be single attribute and target patient population or specific disease states. All claims must be for unidimensional attributes with interval or ratio measurement properties irrespective of whether they are for clinical outcomes, patient reported outcomes (PROs) or drug and resource utilization. Composite or multiattribute generic claims are unacceptable, as will be disease specific PRO claims that fail to meet Rasch or modern measurement standards. This means that the bulk of PRO claims must be rejected as they only produce ordinal scores.

NEW START also minimizes the opportunities for gaming the system. This follows from the requirement that all claims must be supported by an evaluation protocol to detail how the claim is to be evaluated and reported in a meaningful time frame. If all value claims are required to be empirically evaluable there is a firm bass for rejection in an ongoing process of conjecture and refutation; a process that is alien to the approximate information meme The fact that NEW START value claims are only provisional means that there is no possibility of squeaking under the radar with imaginary, purpose built modeled value claims.

CONCLUSIONS

The uncomfortable truth is that for many participants there is no incentive to reject the current meme. Not only is it well entrenched, but industry as the prime mover with deep pockets has no incentive to change; it is an easy option to build a one-off modeled imaginary value claim for cost-effectiveness rather than engage in a long-term research program of therapy response for the target patient population or disease state. Over the past 30 years thousands of analysts have embraced assumption driven imaginary

claims; careers have been built on it and professional associations have unquestioningly adopted it. There are too many with too much to lose.

Against this is the growing realization that the meme for technology assessment is not only built on sand with a complete disregard for the standards of normal science and fundamental measurement, but the real and demonstrated ability of the meme to facilitate systematic bias across study designs and diseases in industry sponsored assumption driven modeled cost-effectiveness threshold value claims. Sponsored modeling will always support, if published, the sponsor's product. After all, that is what many believe consultants are for in health technology assessment. The meme, in this very real sense, despite its acceptance, is self-defeating. Embracing CHEERS 2022 is no guarantee that the pursuit of self-serving assumption driven modeled claims will not continue; indeed CHEERS 2022 encourages it in its support for the submission of modeled claims to journals; journals which have limited resources and skills seriously to challenge and reassess the specific assumptions driving the model. Indeed, it is not just a self-defeating analytical framework but one that from the very start has popularized the support for non-evaluable claims. A position unique among the physical and social sciences.

Formulary committees and other health decision players have no reason to believe, let alone support assumption driven modeled imaginary claims for cost-effectiveness. Instead, the option is there with the New Start formulary submission package which not only meets the required standards but in its emphasis on evaluable protocol driven, single attribute value claims limit the opportunities for systematic bias.

REFERENCES

[1] 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)

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

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

[4] 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

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

[6] Center for the Evaluation of Value and Risk in Health (CEVR), Tufts University Medical Center https://cevr.tuftsmedicalcenter.org/databases/cea-registry

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

[8] Langley P. Nonsense on Stilts – Part 1: The ICER 2020-2023 value assessment framework for constructing imaginary worlds. *Inov Pharm*. 2020;11(1):No. 12

[9] 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