MAIMON WORKING PAPER No. 4 MARCH 2024

FALSE PREMISES AND IMAGINARY CLAIMS: THE PSEUDOSCIENCE OF THE PHARMACEUTICAL BENEFITS ADVISORY COMMITTEE GUIDELINES

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

Abstract

Despite continuing criticism, the belief system in health technology assessment (HTA) continues to be unique in putting to one side the standards of normal science and those for fundamental measurement. This is seen in the commitment to creating assumption driven modeled simulations, such as the reference case models required by NICE in the UK and the PBAC in Australia. Perhaps the most egregious failure is the continued belief in quality adjusted life years (QALYs) as the prime focus for pricing and access decisions for pharmaceutical products and devices. The QALY is an impossible mathematical construct as it involves multiplying utility or preferences, ordinal number scales by the estimated time a modelled hypothetical population spends in disease stages. This is disallowed by fundamental measurement requirements. Illustrative of this lack of understanding of the QALY construct, the current efforts in the US to disallow QALYs in Federal programs focuses on its implications for those with disabilities rather than on the straightforward case that it is just a mathematically impossible construct. This is not the only false premise and associated imaginary claims for cost-effectiveness. The purpose of this paper is to consider seven false premise that support HTA. These are: (i) the pre-eminence of imaginary claims; (ii) the irrelevance of demarcation; (iii) the apotheosis of cost-effectiveness; (iv) the rejection of induction; (v) the rejection of fundamental measurement; (vi) the acceptance of composite ordinal scores as single attribute ratio measures in disguise; and (vii) the acceptance of the cost-per-QALY thresholds to drive resource allocation

INTRODUCTION

The failures of the health technology assessment (HTA) belief system or meme have been extensively documented [1]. They rest upon the failure to apply the standards of normal science to value claims for pharmaceutical products and devices together with a complete failure to appreciate the unique role of the standards of fundamental or Rasch measurement to support valid and replicable value claims [2] [3] [4]. Over the past 30 years the most egregious fault lies with the concept of the quality adjusted life year (QALY) where a single score (I will not call it a measure) is taken as a universal metric to capture quality of life as a value claim for competing therapy interventions and thresholds [5]. Putting on one side the most obvious point of criticism that there is no universally agreed multiattribute QALY given the application of different clinical composites and their descriptions to support algorithms driving preference or utility scores, the fundamental error is to ignore the importance of fundamental or Rasch measurement. Composite or multiattribute scores fail the simple and well-established standard for any value claim: it must have the demonstrated properties that it is unidimensional, linear, interval and invariant . If the term

measure is to be applied then the instrument must be designed to have these properties. Absent these properties the instrument, such as the EQ-5D-3L/5L yields only a composite ordinal score. Put simply. It is impossible to create a QALY with a composite ordinal score as this cannot support any of the standard arithmetical operations; the creation is mathematically nonviable.

Once this fairly obvious objection is recognized, at least in modern measurement theory with the Rasch model of the 1960s, the contribution of modelled cost-per-QALY claims, claims for cost-effectiveness and cost-per-QALY thresholds collapses. It is not the question just of modelled imaginary claims which fail standards for empirical evaluation, but the patent falsity of the premises to support these claims. This represents 30 years of wasted effort in HTA. The thousands of published peer-reviews cost-per-QALY claims, resting on composite metrics, are simply irrelevant. It is not clear how many of the thousands of modeled cost-per-QALY claims are nothing more than marketing devices where the sponsor of the model has indicated that it should defend the product at a price, deemed cost-effective, that is consistent with the manufacturers intended market launch price. All we know is that there is a strong presumption that manufacturer supported models are all too often consistent with known cost-per-QALY thresholds [6].

The question which has to addressed is why this belief has dominated decision making in HTA? Social psychology provides a possible explanation in terms of suggestibility, a willingness to yield to a social pressure, a high degree of social interaction and imitation such that the group behaves like a crowd. This is reinforced by the fact that people respond to narratives than to facts and data, where the more compelling the narrative the less the application of critical thinking. An audience of cognitive misers who avoid rigorous analysis in favor of heuristics. If the intent is to convince an audience, the more compelling narratives possible will be devised; analytical ability is corroded by compelling narrative irrespective of the 'truth elements' in that narrative; false premises are accepted as part of a persuasive narrative [7].

Without getting into an extended semantic debate, it is of interest to consider what the term 'falsity' means in the premises accepted by HTA. If a false claim is promoted, does this imply that those involved are aware that it is a lie because they have put the truth to one side? In HTA this may not be the case as those involved in assumption driven modelling may be unaware of the standards of normal science and fundamental measurement. Or they may consider these standards to be irrelevant because they do not consider that HTA should be focused on what Popper describes as the evolution of objective knowledge [8]. To an independent observer, however, it is difficult to conceive of a discipline that is preoccupied with creating imaginary claims and not with the process of discovery of new facts for therapeutic impacts as exemplified by the CHEERS 2022 guidance for the creation and submission of imaginary HTA modelled claims to unsuspecting journal editors and reviewers [9]. After all, the essence of the 17th century scientific revolution is summed up in the motto of the Royal Society 'nullius in verba' (take no one's word for it) yet with assumption driven non-evaluable claims that are promoted as critical decision variables in therapy pricing and access there is no further step; we are asked to take the model builders word for it although it is an analytical dead end.

It is difficult to believe that leaders in the HTA discipline are not unaware of the standards of normal science and fundamental measurement. This appears, unfortunately to be the case with the

leading text on imaginary claims attempting to introduce a notion of fundamental measurement but failing to understand what it involves in value claims; at best it is confused [10]

The purpose of this paper is to consider the narrative of HTA in terms of the false premises that support the HTA narrative; to assess the appropriateness of the charge that the HTA belief system in respect of assumption driven modelled non-evaluable claims is pseudoscience. The case that the HTA narrative rests on what are best described as false premises. This is particularly apposite given current Australian Health Technology Assessment Policy and Methods Review [11] The case to be made is that the HTA belief system narrative lacks credibility, as detailed in a submission to the review committee [12]. But, like so many of what may be described as delusional beliefs, an accounting of false premises may fail to shake belief. If this is the case then the downside of the narrative will continue with a continued preoccupation with imaginary claims as the chief barrier to progress in HTA and with the PBAC riding out the storm.

THE MEANING OF PSEUDOSCIENCE

The Oxford English Dictionary (OED) defines pseudoscience as: *A pretended or spurious science; a collection of related beliefs about the world mistakenly regarded as being based on scientific method or as having the status that scientific truths now have.* The emphasis by a number of writers, as noted by Hansson, is that pseudoscience is non-science posing as science; accepted beliefs masquerade as genuinely scientific ones [13]. In other words, pseudoscience is seen to involve a sustained effort to promote standpoints different from those that have scientific legitimacy. Pseudoscience deviates from the quality criteria of science: reliability, fruitfulness and practical usefulness.

When the standards of normal science are summarized, it is to consider credible claims, empirical evaluation and replication (and reproduction). The first question is, therefore, the status of a credible claim: what distinguishes a credible claim from pseudoscience? Hansson proposes two criteria to classify a claim as pseudoscience, where the second criteria can take a narrow (ii) or wider form (iii):

(i) *It is at variance with the most reliable knowledge about its subject matter that is currently available; and*

(ii) *it is part of a non-scientific doctrine whose major proponents try to create the impression that it is scientific or*

(iii) *it is part of a doctrine whose major proponents try to create the impression that it represents the most reliable knowledge on its subject matter.*

Criteria (i) is a revised version of an earlier criteria which simply stated that the activity is not scientific. The revised version follows from consideration that demarcation recognizes the quality of science; the function of science as a process of discovery or fact-finding that has the objective of providing the most reliable current information. The process of discovery follows rules; pseudoscience while often attempting to mimic science, fails at his basic level. This does not, it

should be emphasized, imply, following Popper a rigid application of the test for falsifiability. We could opt for a more sophisticated falsification as detailed by Lakatos or adopt a multi-criterion as opposed to a mono-criterion approach, a check list, to label a theory or analytical framework as pseudoscience practice. Elements of such a list could include a belief in the authority of leaders in a field (particularly apt for HTA), non-testable claims, the process by which claims are created, nonreplicable or nonreproducible claims and the rejection of refutation. A major problem with the multi-criterion approach is which criteria do we choose, in particular in the assessment of clinical claims where it is suggested that the majority of such claims are false.

In terms of criteria (ii) and (iii) above, the question is the extent we wish to case a narrow or wider net for a definition of activities that we characterize as pseudoscience. The narrower net is to consider activities characteristic of individuated sciences or belief systems that are seen as specific branches of knowledge. The wider net (iii) sees science as activities which share a common focus with the individual sciences as merely examples.

As Hansson makes clear, there is a critical distinction between science and pseudoscience [12]. In the case of science there is agreement on the essential unity of the objectives of scientific enquiry and the process of creating and evaluating credible or falsifiable claims; ultimately, the determination of beliefs that are epistemologically warranted: *arguably, the crucial issue is not whether something is called 'science' but whether it is claimed to have the function of science, namely to provide the most reliable information about its subject-matter* [14] . On this criterion, as will be discussed in detail later, it can hardly be claimed that the HTA assumption driven simulation provides the most reliable information. In fact, we might call it essentially information free. There is no focus on discovery or the reporting of discovery of new facts. At best, there is a summary of the latest clinical trial data, which has been reported elsewhere. Assumption driven claims are not information where the assumption choice within the same or different model structures can produce any number of non-evaluable cost-effectiveness claims for the same product.

.Focusing on the HTA belief in assumption driven simulations the question first and foremost is why was the reference case instruction to populate Markov models developed? It is made quite clear that the focus was on creating approximate information [15]. To plug an evidence-gap at product launch is to quickly create imaginary evidence that decision makers could be convinced to act upon. The decision was made to reject normal science, to reject hypothesis testing and the possibility of falsification and to come down on the nonscience side of demarcation; truly intelligent design. In respect of criteria (i) above the HTA belief system summarily rejected the possibility of developing, through a planned research strategy, new facts creating reliable knowledge consistent with the standards of normal science. Instead of a commitment to mind-independent objective knowledge of provisional facts to support therapy decisions, the intent was to construct claims based on mind-dependent approximate information; not to recognize hypothesis testing but to subvert it.

The term "bullshit" was introduced into philosophy by Harry Frankfurt in in a 1986 essay and followed by a book in 2005 [16]. The term is used to describe a type of falsehood that does not amount to lying. A person who lies deliberately chooses not to tell the truth; they are inescapably concerned with truth values. An effective lie is one that designed under the guidance of truth.

Whereas a person who utters bullshit (a.k.a humbug) is not interested in whether a statement is true or false, only in its suitability for his or her purpose. Bullshit is seen as a claim that is phony, often promoted as a program of fake claims with deliberate misrepresentation. Bullshit is indifferent to the truth or the pursuit of the truth. Both the bullshitter and the liar are perceived as attempting to communicate the truth; their success depends on how well they deceive. But while the liar leads us away from a correct apprehension of realty, the bullshitter has no interest in the truth value of statements or claims. The bullshitter is concerned only with getting away with what they can with a credulous audience. A culpable lack of concern with the truth where there is no intention of trying to describe reality, the elements are just picked out to suit the purpose. No attention is paid to the authority of truth. This is why, for Frankfurt, bullshit is the greater enemy of truth than lying.

Whether HTA assumption driven modelling to produce imaginary claims falls into this category raises a number of questions: the role of intent is uppermost. The intent of modelling is presumably not to misrepresent as there is no reference point for misrepresentation, although perpetuating a required outcome through manipulation of model structure and assumptions is presumably fraud and an ever-present incentive by the less scrupulous. It is also difficult to judge the intent of the modeler: is there a lack of interest in whether the model or its assumption choice is true or false. Is it seen as a profitable consulting opportunity? All we can say is that, at best, the model builder presumably chooses a model structure and assumption data points because they seem 'reasonable'. As the model extends to capture the lifespan of a hypothetical population there is no basis for assigning degrees of belief in data elements (the induction problem) nor is there an independent reference point for approximation that, as will be noted, meets the standards for normal science and fundamental measurement. There can presumably, not be a reference point for approximate claims based on false premises that can extend decades into an unknown future to create an imaginary cost-effectiveness claim.

Against this is the question, again, of whether or not the model builder cares whether or not the things that are said describe reality correctly; are they just picked out, in the case of model structure and assumptions, to fit a purpose? Against this is a key point which comes back to the view that the model building is phony: we can have no faith whatsoever is the truth-value of one model over another because none have any truth-value. There is no future reality that can be captured or even approximated. The notion of approximate information is nonsensical. However, the concept or notion of the validity of approximate information (whatever that means as the data elements must, by definition, lack reference points for the future) is the driver for modeled imaginary claims.

Can we categorize PBAC-type modelled claims as falling under the bullshit umbrella? It is certainly pseudoscience. A common characterization of pseudoscience is that it is not only false and unfalsifiable but that its practitioners are indifferent or take no care to assure the truth of their claims. In more formal terms are we in a position to argue that the supporters and practitioners in assumption driven model simulations lack epistemic conscientiousness? Do supporters of the HTA assumption driven model framework actually subscribe to the meme or are they inhabiting parallel worlds? A superposition where they recognize the importance, call it truth, of the standards of normal science and, possibly to a lesser extent, standards of fundamental measurement yet when

asked to create a case for cost-effectiveness fall back on reference case modelling and pseudoscience. Given the tenacity with which the HTA belief system is held by heath economists, pharmacists and others in Australia, and other single payer health systems, the ability to hold one's nose seems unlikely. Understandably, few seem to want to rock the boat. It is difficult to believe that those involved in health technology assessment are unaware of the standards of normal science even though possibly unaware of the standards of fundamental measurement. If this is the case then it is difficult to judge the motivation for supporting a belief in the creation of non-evaluable imaginary claims; and where these claims are seen as a key decision criterion.

In a recent paper, Moberger has proposed that pseudoscience should be seen as a special case of bullshit, understood as "a culpable lack of epistemic conscientiousness" ; an indifference to getting to the truth or a total lack of concern with the truth. This is the overriding characteristic of assumption driven modelled simulations where the truth is not even out there. But this is not an excuse for a lack of concern. One might counter this with a reason for actively pursuing the truth. No effort is made, for example to proposing a research program to discover new facts. On this lack of culpable concern and indifference towards the truth alone we could argue for modelled assumption driven non-evaluable claims as an analytical and uninformative dead end.

The argument presented by Moberger leads to his belief that pseudoscience is a special case of bullshit. [17]. As the model builder can be culpably indifferent towards the truth of statements (e.g., a product is cost-effective) or to be culpably unconscientious with respect to their truth; *an ineptitude that neither presupposes nor rules out indifference towards the truth* [17] . Moberger proposes that pseudoscience is bullshit with scientific pretensions. The HTA model makes the claim that it is asserting propositions on a scientific issue, the modeling gives the appearance of being scientifically rigorous while creating claims that a patently non-falsifiable. An outcome that is built into the modeling for approximate information. The model creates claims that are not only non-falsifiable but they were intended to be non-falsifiable. If we accept Moberger's position that pseudoscience is a special case of bullshit, then this is a special case of pseudoscience where the question of falsifiability can never arise. The PBAC guidelines have enforced a program of bullshit; a program which has the explicit approval of the health economics community in Australia. The bullshitter aims to get away with what he says; describing reality is immaterial. The elements of the case are picked out to support the present purpose and convincing others that, in the guise of a scientific approach, the model is endeavoring to communicate the assumption driven simulation version of the truth; a truth that can extend decades into the future.

The HTA imaginary modelled claim narrative is compelling; appearance in bullshit is a key characteristic. It provides a global solution that captures all elements of the decision, prices, costs and outcomes, into a single global metric with the resolution of a fundamental question: is a new product cost-effective! Embraced by single payer health systems by agencies such as NICE in the UK, CADTH in Canada, PBAC in Australia, PHARMAC in New Zealand and ICER in the US, the aspiring practitioner in HTA, particularly where there appear to be no critiques offered in HTA course (many of dubious veracity) the narrative is accepted without question. Whether this is to be labelled delusional behavior on a stand with dispensational premillennialism, is an open question;

yet the hallmarks are there. Indeed, whether the current 5-year review of PBAC methods and models will achieve a breakthrough to a new start in HTA is a moot point; belief can be tenacious.

FALSE PREMISES SUPPORTING THE HTA BELIEF SYSTEM

In the case of PBAC modelling a further issue arises: the truth of the premises that support the assumption driven simulation model. This is not to define truth as the presence of a substantive claim for statistical regularity but to question whether the premises and intent of the analysis are true or false. The position taken here is to consider truth versus falsity from two perspectives: the standards of normal science where conclusions fail the required evidentiary standards and the case where there is a failure to meet standards of Rasch or fundamental measurement where the only acceptable metric is a measure that is unidimensional, linear, interval and invariant.

Importantly in HTA, the term premise also includes patently false declarative statements to justify the truth value of a further proposition or conclusion. The truth value, or proof of a conclusion, depends on both the truth of the premises and the validity of the argument. In the case of the PBAC, the conclusion is expressed as a claim for cost-effectiveness based on creation of assumption driven modelled simulations. In this context, a false premise is an incorrect proposition that forms a basis for the false conclusion of cost-effectiveness. Even so, it should be noted that that the logical validity of an argument, the proof of a conclusion, is a function of its internal consistency, not the truth value of its premises. In the case of PBAC modelled claims, the question is not one of internal consistency but the fact that the analytical framework encompasses both false premises and imaginary conclusions or non-evaluable claims for cost-effectiveness for which no truth value can be assigned. If a premise is false, but not by a judgement of its frequentist or evidentiary base which may be a compelling argument for its rejection, but by its failure to meet standards for Rasch or fundamental measurement, then it should also be rejected.

The emphasis on false premises extends to the appearance or structure of the simulation model. While this may be justified in terms of simplifying the stages experienced in a disease state with associated transition probabilities expressed, for example, as a Markov process, the choice and application of a particular populated Markov process yields false claims based on false premises (e.g., incremental cost-per-QALY claims). While it is possible to justify the model structure as a reasonable approximation or simplification, it remains an incorrect proposition. The claim may be made that it is valid, but it remains the basis for false claims for cost-effectiveness. This leads to an important feature of the appeal of model simulations: appearance. The argument is straightforward: appearance is an essential part of fallacies where they must have the appearance, however quickly seen through, of being valid [18]. An argument that appears singularly apposite when applied to the HTA meme where the assumption driven modelled simulation to produce imaginary claims continues to occupy center stage as it has done for the past 30 years where the leaders in the field have assiduously promoted these false claims [19]. False claims, it must be noted, that are virtually impossible to separate from the truly false claims for cost-effectiveness produced by paper mills and less scrupulous players, including academics and their associated groups [20].

This commitment to a false belief system is unnerving. There are clearly fundamental flaws to have rendered this meme unsustainable. Yet it continues to be an article of faith by those whose

training should have rejected it. This has not happened with the result that HTA has the unique and unenviable status as the only discipline (if that is the right word) where the standards of normal science and fundamental measurement are absent. There is a culpable lack of concern with the truth of one's statements; modelled assumption driven claims for therapy impact are accepted at face value; *the truth is out there, but so are lies*.

SEVEN FALSE PREMISES

Seven false premises that support the current HTA belief system or meme are identified. They range from the widely held premise or belief that imaginary claims have a pre-eminent position in formulary decisions (False Premise 1) to an acceptance of composite or multiattribute ordinal preferences as interval measures in disguise (False Premise 5) to a belief in cost-per QALY thresholds (False Premise 7). These premises are the foundation of the HTA belief system; a foundation that supports the pseudoscience of modelled deliberate non-falsifiable claims for invented comparative cost-effectiveness, delivered as approximate information.

FALSE PREMISE 1: THE PRE-EMINENCE OF IMAGINARY CLAIMS

The HTA meme is predicated on the creation of assumption driven modelled simulations that rely on assumptions and produce non-evaluable cost-effectiveness claims that lack credibility in terms of both the standards of normal science and fundamental measurement. Whether we view the belief in the imaginary modelled claim as an understandable and convenient shortcut, given the limited evidence at product launch, to establish an imaginary claim for cost-effectiveness or as a deliberate and continuing narrative that both HTA practitioners and the recipients of imaginary modelled claims know is false is a moot point. The recent publication of the updated CHEERS 2022 guidelines for submitting imaginary claims would suggest that the advocates for the simulated imaginary claims models are not deliberately avoiding the required standards of normal science and fundamental measure but have no idea of their relevance for value claims [1 14] .

This belief systems puts to one side the commitment to the discovery of new facts, a process which Popper described as the search for objective knowledge [8]. Rather than attempt the serious activity of tracking the impact of a therapy and the assessment of value claims, HTA falls back on formulating non-evaluable imaginary claims for cost-effectiveness. The result has been over the past 30 years an embarrassment of imaginary modelled cost-effectiveness claims numbering in the thousands, many littering the peer reviewed literature with no purpose other than to serve as marketing devices. With CHEERS 2022 and its attempted reinforcement of the HTA meme, creating approximate information, all we can look forward to is a continuing commitment to imaginary modelled claims where neither authors nor decision makers are seeming aware of the absurdity of the analytical framework . It is just an easy way out but where no claim can be taken seriously. The model builder is in the envious position that absent falsification, they can never be held accountable.

At the risk of repetition, once point must be made abundantly clear: the belief in assumption driven imaginary simulations by leaders and professional groups such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and their satellites such as the PBAC make clear that so-called pharmacoeconomics or HTA is a belief system where the standards of normal

science and fundamental measurement have no role. It is as though the scientific revolution of the 17th century had never occurred and where over a century of debate and agreement on fundamental measurement is totally irrelevant. This is the foundation false premise: a position unique in the physical and social sciences.

FALSE PREMISE 2: THE IRRELEVANCE OF DEMARCATION

Judged by the standards of demarcation between science and non-science, the HTA belief system for modelled claims falls squarely in the latter camp [21]. It shares this space with delusionary cults as well the many the millions of believers in intelligent design, creationism's successor. The demarcation standard is that all value claims must be empirically evaluable [22]. For those believing in the HTA meme, this standard is, apparently, irrelevant. This puts HTA at variance with the standards that emerged in the scientific revolution of the 17th century: the appeal to superior evidence [23] and a commitment, in Popper's terms, to the potential for falsification of value claims. This is of significance in therapy value claims; they must be capable of replication and reproduction in target patient populations. It is impossible, of course, to apply these requirements to imaginary claims that are designed to be empirically non-evaluable. This also means that the level of confidence in therapy-based pivotal trial endpoints in translation to treatment practice is seriously qualified.

If the concept of demarcation is rejected then what are the consequences for HTA? If there is no boundary, the essence of demarcation, between science and non-science then we have to accept that claims which are empirically non-evaluable have an equal epistemic status with those that are empirically evaluable. Perhaps we have to fall back on a subjective assessment of epistemic reliability to differentiate 'true' claims from 'false' claims and what Hansson describes as 'epistemic warrant'. Can we, outside of demarcation, make any claims for cognitive reliability? Should we focus on the science/pseudoscience distinction where 'an epistemic field is a group of people and their theories and practices, aimed at gaining knowledge of some sort'? But does this knowledge have to be factual or true? Are imaginary constructs knowledge?

There is no doubt that there are borderline areas of science (e.g., string theory, inflationary cosmology, field studies) but these are more unresolved activities than rest on 'nature' or access to empirical data. The HTA imaginary claims do not occupy a borderline area; there is no possibility that any of the claims for cost-effectiveness can even be empirically evaluated. It is not that one set of assumptions to populate one set of imaginary cost-effectiveness claims is (to an unknown extent) superior to another but the fact that whichever claim is flavor of the PBAC month it includes assumptions that are patently false with claims that are not empirically evaluable (even in a distant galaxy sharing the same wave function). Appeals to a supernatural designer will hopefully fall on deaf ears despite pseudoscientific stratagems such as sowing seductively deceptive false doubts.

Abandoning any commitment to demarcation ensures that HTA imaginary cost-effectiveness pseudoscientific claims are a performance without an audience; or at least an audience that is less that credulous. Unfortunately, the audience is receptive either from a lack of training to instill essential beliefs or a lack of ability to challenge beliefs. As Dawkins points out in his case for

viruses of the mind: lack of evidence can be a positive virtue [24]. Of particular relevance to HTA is the belief that mystery is a positive virtue and quoting Tertullian *Certum est quia impossible est* (It is certain because it is impossible). Perhaps a suitable epitaph for the QALY.

FALSE PREMISE 3: THE APOTHEOSIS OF COST-EFFECTIVENESS

The claim that a product is cost-effective at a price subject to threshold cost-per-QALY barriers, while widely accepted, is meaningless. While one can appreciate its brevity and appeal, those making such claims are caught up in a framework which is an analytical dead end. Nevertheless, among those who either reject or are unaware of the standards of normal science and fundamental measurement, where modelled simulations have the appearance of being the quintessence of good practice to inform formulary committees, they are accepted without question. Certainly, there may be questions and quibbles raised as to modelling requirements (e.g., the cost-benefit discount rate) but these are irrelevant once the analytical framework is rejected and appearance is seen as nothing more than a masquerade.

Australia is not unique in what is essentially an unquestioning belief and acceptance of imaginary cost-effectiveness claims. As Dawkins points out, in comparing mind viruses to computer viruses, the mind virus will be hard for their victims to detect; a deep inner conviction that a belief is 'true or right, or virtuous' which appears to owe nothing to evidence or reason [24]. A relativist belief system that is continually reinforced through education, training, conferences, messages from the leaders in the field and the close association with others who share, uncritically, the same views. A situation that is enhanced by the HTA media reporting a cost-effectiveness recommendation when they have no idea of the false analytical status of the claim.

The foundation for blanket model driven cost-effectiveness claims is fatally flawed. There is no possibility of an 'adjustment'; we require a new paradigm which, unlike the notional of evolution following Popper, owes nothing to the past of imaginary HTA claims. This systemic failure is evidenced by guidelines for reference case models published by single payer system gatekeepers and their emulators. They provide an uncritical and enduring basis for false premises and imaginary claims [25]. The possibility that inputs to imaginary models may be false is, of course, immaterial. Any assumption, false or otherwise, is accepted as long as it passes muster with the PBAC and its imaginary model evaluators or academe assumption police in their acceptance critiques of modelled imaginary claims.

Attempting to create non-empirically evaluable modelled claims for cost-effectiveness, where a judicious choice of assumption for the model can create a situation where most prices can be claimed to be cost-effective is hardly a serious form of analysis. It is, as noted above, bullshit with scientific pretensions. The model builder has no interest in the truth-value or otherwise of the modelled claims. This is not necessarily a question of lying but of laziness; a lack of willingness to challenge the basis for cost-effectiveness claims. The PBAC has encouraged, if not mandated, a program of creating bullshit with scientific pretensions with its reference case standards. This has lasted for some 30 years. The result is a stable of consultants and those attached to academic positions, that are well versed in PBAC requirements and are ideally suited to apply and manipulate

these to achieve the best price and access case demanded by sponsors, where none of parties has any interest in truth-value.

Unfortunately, given the experience in the sciences, there is a presumption that the ease of creating dubious or false claims for cost-effectiveness merely encourages the mass production of paper mill studies [26]. This is more than the odd claim but a commitment to the contribution of paper mills. Where the center piece in HTA is modeled imaginary claims that could account for hundreds if not thousands of these studies that have appeared in the literature. This is not, presumably, a problem that is exclusive to the promotion of new products, but may be nothing more than CV padding. Even so, the saving grace, the acceptance that these cost-effectiveness claims are entirely imaginary and should not be taken seriously, may disincentivize the attraction.

FALSE PREMISE 4: THE REJECTION OF INDUCTION

Assumptions, of course, play a key part in the creation of imaginary claims, both in the choice of model structure and the population of data points within that structure. These assumptions may be based on pivotal clinical trial data (which may be false or incapable of replication) or on a literature search What is all too apparent is the limited base for all many if not all assumptions in HTA imaginary models, where only one or two references may be cited to support that choice. Again, all too many assumptions, such as the choice of preference or utility scores, may not only be limited in their applications but demonstrably false in their neglect of fundamental measurement.

Claims that some assumptions are more appropriate or more 'realistic' than others to support and justify long term and even lifetime modelling of a hypothetical population by stages of disease run into Hume's problem of induction (first formulated in 1748) [27]. If we are to apply the principle of induction then it lacks proof or disproof by experience; we have to assume it holds even in probabilistic terms. The problem of induction as Russell so eloquently put it in his 1912 essay is that in terms of past associations between cause and outcome:

> *We know that all these rather crude expectations of uniformity are liable to be misleading. The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken* [11].

Put more cogently: the fact that past futures have resembled past pasts does not mean that future futures will resemble future pasts. Even in probabilistic terms, we might have greater predictive faith in an association or outcome, but this presupposes assumptions about literature or sense-based completeness of observations. This is the Achilles heel of assumption driven imaginary claims: it rests not only on the 'credibility' of assumptions but on an assumption regarding the acceptance of the principle of induction.

It is important to distinguish two forms of argument: deductive arguments and inductive arguments [28]. The former refers to arguments where the truth of the premises guarantees the truth of the conclusion; they are deductively valid and necessary for truth preservation (e.g. mathematical proofs), The latter refers to arguments based on statistical frequencies. The conclusion will hold in a significant proportion of possible worlds where the premises hold. However, *to argue that past*

instances and regularities lead to conclusions about future instances and general principles where the truth of premises provides some degree of support for the truth of the conclusion is open to substantive objections. The focus on statistical regularity, typically described as the Uncertainty Principle, runs into the objections that, first, the argument cannot be deductive and second that the claim cannot be probable as that presupposes the Uncertainty Principle (the argument is circular). Only if the Uncertainty Principle can be conclusively verified, can inductive inferences be truly justified; a position that is untenable.

When this is raised the true believer in HTA will presumably follow in the response laid down before Galileo with traditionalists refusing to look through the telescope. Unfortunately, the problem of induction cannot be put to one side as it effectively undercuts the basis for imaginary simulated claims. This does not deny the need to base claims on assumptions, but to insist that the claims can be empirically evaluated and falsified. This gives a basis for a reassessment of these assumptions to decide which are acceptable. A singularly fruitless undertaking as the status of one more non-evaluable imaginary claim is another cost-effectiveness claim.

FALSE PREMISE 5: THE REJECTION OF FUNDAMENTAL MEASUREMENT

Once the standards for fundamental or Rasch measurement are applied to the various generic multiattribute and disease specific patient reported outcome (PRO) instruments in HTA, the inescapable conclusion is that all composite instruments and 95% of disease specific instruments are failures. The reason for this wasted effort over the past 30 to 40 years is clearcut: if the required standard is, as it must be, for value claims to be based on PROs that the instrument has to be for a single attribute: a unidimensional measure that has linear, interval and invariance properties.

The Rasch model, where data are selected to meet the mathematical standards of the model, is both necessary and sufficient to yield a translation from observation to measurement. Rasch is unique; there are no exceptions. This has held for over 60 years where the Rasch contribution provides the framework for assessing one human attribute at a time (e.g., need fulfillment as a measure of quality of life). The Rasch principle is quite straightforward:

> *A person having a greater ability should have the greater probability of solving any type of item in question, and, similarly, one item being more difficult than another means that for any person the probability of solving the second item is the greater one* [2]

If the attribute of interest is quality of life, then for a target patient group, this can be proposed and hopefully measured by Rasch standards in terms of the needs of that target population. Needs are expressed as item statements of increasing difficulty, where for a given distribution of abilities (typically normal) a therapy intervention my facilitate item needs of greater difficulty being resolved by an increasing proportion of the tare population in terms of expected response probabilities. Linear and interval measure allows the standards tools of statistical analysis to be applied; they cannot be applied in any non-Rash score or putative measure.

Apart from multiattribute composite ordinal generic scores, the failure to meet Rasch standards is all too evidence in the many hundreds of disease specific PRO instruments. The most widely used

scoring framework is to propose a series of item statements where the responses are in Likert or integer format. Numbers are applied to the integer response category and the responses aggregated to create an overall score. Certainly, Rasch modelling can be applied to polytomous as well as dichotomous data with application of the Rasch Rating Scale Model or the Rasch Partial Credit Model developed over 40 years ago [2 3]. The mistake made by those proposing value claims based on integer summation is to assume *a priori* (but not consciously) that all items in the instrument are of equal difficulty and that the thresholds between the steps are of equal distance or value. The Rasch model makes no such assumptions: items are not of equal difficulty and the integer progression for each item is only an order.

The Bond et al summary of the unique status of the Rasch model is worth noting (p. 274) [2]. The key difference between the Rasch model and Item Response Theory (IRT), classical test theory (TST) and other traditional models is that the Rasch model fits the data to the model requirements while with IRT and TST the model is fitted to the data. In other words, for Rasch the model has primacy while with the latter the data have primacy. IRT and TST models are exploratory and descriptive of the data while the Rasch model is confirmatory and predictive. The IRT and TST models try to take account of all the data; if necessary, by adding more explanatory variables (tweaking). The Rasch model requires the data items to fit the model; typically selected from patient interviews in a disease area. This fit focuses on the size and structure of residuals applying probabilistic conjoint measurement. If the required fit standards are achieved, we can claim that the results can be used as a measurement scale with linear. interval and invariant measurement properties. Such a claim is impossible with IRT and TST models.

Attempting to fit a model to the data accounts for the fact that the various multiattribute models, while anchored at unity, cannot ensure a true zero and hence the possibility of health states worse than death. Apart from the disallowed composite nature of various health algorithm scores, it also ensures that these scores cannot meet the standards of unidimensional, linear and interval

These errors are compounded once we consider mapping; the transformation of one set of therapy responses into another (e.g., transforming a disease specific ordinal score to a multiattribute (e.g., EQ-5D-3L) composite ordinal score from instruments jointly administered to a target patient population. Mapping exercises in peer reviewed publications continue, where neither the author nor the reviewer/editor have any idea of the constraints of fundamental measurement. Neither party appreciates that to report a crosswalking or mapping exercise from disease specific measures to a generic measure where the former is an ordinal scale and the claim is to create an ordinal composite score equivalent, is denied by fundamental measurement. Again, a completely meaningless exercise.

It is worth noting that there is no accord given to the standards of Rasch measurement for value claims in the leading textbooks and journals (notably *Value in Health* and *Pharmacoeconomics*). While *Value in Health*, the house journal of ISPOR has published a number of papers that apply Rasch criteria to established PRO instruments, there is no credence given to Rasch practice standards and the unique contribution of fundamental measurement for PROs. ISPOR is still wedded to the QALY and the credibility of imaginary yet non-evaluable cost-effectiveness claims.

FALSE PREMISE 6: THE ACCEPTANCE OF COMPOSITE ORDINAL SCORES AS SINGLE ATTRIBUTE RATIO MEASURES IN DISGUISE

For more than 30 years HTA has been obsessed with the QALY; it is the centerpiece of assumption driven simulations and an object of veneration. Understandably, for those who accept the QALY uncritically, including those who have a vested and financial interest, apostasy is to be firmly put down. Any hint that the QALY is a measurement will o'the wisp is treated with scorn. HTA seems to be trapped in its belief in the QALY, an unchallenged mystery. CHEERS 2022 presents the QALY as an unassailable building block; it could, of course, not do otherwise. To this should be added the role of the EuroQol organization, which has a vested interest in QALYs, in its continuing promotion of the various ordinal preference algorithms: first the EQ-5D-3l, then the EQ-5D-5L and now the EQ-HWB [29]. All fail the standards of fundamental measurement.

Despite this embrace of the QALY the fact remains, and has been true for 60 years, that the preference scores that support the QALY fail the standards of fundamental measurement. This is dictated by the requirements of Rasch mathematics: the necessary and sufficient framework for transforming observations to measurement. Observations reported by patients and caregivers are not measurement. They need to be transformed to create a unidimensional, linear, interval and invariant metric. This is the only basis for measurement irrespective of whether the instrument supporting the observations comprise dichotomous or polytomous responses.

While it may come as no surprise that HTA programs typically fail to report on Rasch measurement, it is Rasch measurement that devalues the QALY to irrelevance. At the same time, this ignorance of Rasch measurement ensures that the QALY is accepted at face value as a 'measure' by those less acquainted with fundamental or Rasch measurement or who propose to ignore it.

The QALY failure can be explained in simple terms: if you want to create a QALY, then the utility or preference score must be unidimensional, linear, interval and invariant with a true zero: a ratio scale. While this standard has been recognized for over 60 years, the HTA application of algorithms that support utility and preference scores made no provision for this requirement. Instead, they took without question a multiattribute approach, typically with health state preferences scored by the time trade off (TTO) technique which produces ordinal scores (with positive and negative outcomes) and with these TTO scores the basis for ordinal utility or preference weights. These are then combined in an algorithm to create an ordinal scale in order to fit the health status data. What was not foreseen, is that fitting a model to data to create an algorithm to create a utility or preference score for health states in the range 0 – 1 is unlikely to succeed. This was the case across the board with different algorithms, with 1 = perfect health and other scores for 'less perfect' health states determined as decrements from the unity reference.

Health states are composite bundles of health states and dimensions and as such, as we have known since the 1950s, disallowed by measurement theory. But this constraint, the focus on a single unidimensional attribute that might capture quality of life was ignored; or most probably never considered with the commitment to defining health states. One of the more intractable results was the overshooting of zero (supposedly death) to give negative scores or states worse than death and,

inevitably, negative QALYs. Despite repeated tweaking to squeeze the numbers into the desired range, negative values or states worse than death were commonplace. In the latest US valuation of health states, the preference range for the EQ-5D-5L is 1 to -0.573, with 20% of health states taking negative scores. Even if this tweaking were, by happenstance successful for all possible health states, this would not solve the problem as the scale would still be ordinal.

One of the more bizarre claims in HTA has been made by ICER that, when challenged, regarding the measurement properties of preference or utility scales, the following statement was posted:

> *As we have expressed before we (and most health economists) are confident that changes in the EQ-5D (and other multiattribute utility instruments) do have ratio properties. The EQ-5D value sets are based on time trade-off assessments (which are interval level), with preference weights assigned to different attributes. We fail to see why this should be considered an ordinal (ranked) scale. The dead state represents a natural zero point on a health-related quality of life. Negative utility values on the EQ-5D scale represent states worse than dead. We do not find this lacks face validity* [30].

This is complete nonsense: a ratio measure is defined as one that is unidimensional, linear interval, invariant and with a true zero. The Rasch model, applied to PRO observations, is unique as the necessary and sufficient condition to yield an interval scale with the required properties and an approximation to a true zero (no negative measures) . The dead state is not a natural or true zero point; as some 20% or more responses to value sets take negative values. The notion of states worse than death is a necessary afterthought in order to defend the indefensible. It is not clear how those who are beyond death actually feel about it as the value score for a stare worse than death is not based on responses from patients but from public preferences. Hence negative QALYs (ignoring ordinal properties of health state values), it has been argued, should be abandoned in favor of only positive response values [31]. This could be achieved just cutting of values that are negative and assuming they are zero but this overlooks the problem of dead respondents.

.Also note that time trade-off is not an interval measure, the TTO preferences are only ordinal scores. There was no thought given when the EQ-5D and other generic multiattribute instruments were being developed to conform to Rasch or fundamental measurement standards. This would have simply illustrated that composite bundles of health states cannot yield a measure that has the required Rasch properties. The health economist may 'have confidence' but this is just a fudge because the assumption, which is unsupported, that is needed to holds this edifice of approximate imaginary information driven by assumption driven simulations together. It lacks any validity; it is simply gibberish as evidenced by the belief ('confidence without proof'') that a ratio measure with a true zero can take negative values.

It is perhaps not surprising that the many criticisms of QALYs, for example bias against those with disabilities, fail to make the point that the QALY is mathematically impossible [32]. This is the position taken with respect to efforts currently underway in the US with H.R. 485 *Protecting Health Care for all Patients Act of 2023* to Amend Title XI of the *Social Security Act* to prohibit the use of the QALY or similar measures in coverage and payment determinations in Federal

programs [33]. While this has passed the House the likelihood of passing the Senate seems problematic (House vote 211 to 208). There is a certain grim humor to see an effort to restrict the application of the QALY, without realizing that the QALY should be discarded (and never even adopted) as an impossible mathematical construct; focusing on disabilities misses the point.

FALSE PREMISE 7: THE ACCEPTANCE OF THE COST-PER-QALY THRESHOLDS TO DRIVE RESOURCE ALLOCATION

If it was thought that in health care systems with fixed budgets, the QALY could drive resource allocation, then the supporters must be disappointed. Proposals that resources could be allocated between therapies and therapy areas on the basis of cost-per-QALY estimates and cost-per-QALY thresholds is simply absurd. If for no other reason that the QALY rests on composite utility or preference scores which have only ordinal properties; unless a universal QALY equivalent single attribute, linear and invariant ratio (not interval) score is accepted, then any attempts should be abandoned. Ratio properties are an essential requirement where there is a true zero (or acceptable approximation) for the measure, avoiding the inevitable negative utilities and preferences that emerge from the various multiattribute, utility decrement efforts to report clinically-based quality of life.

The inherent ability to falsify and create imaginary cost-effectiveness claims in assumption driven simulations is matched when the gatekeepers establish cost-per-QALY thresholds. In the UK, the NICE threshold is a modeled claim supporting a price that is less than £20,000 per QALY; in Australia there are no fixed thresholds although a review of PBAC decisions would suggest a range of $45,000 to $60,000. In common with other cost-per-QALY thresholds, these are only ordinal scores not ratio measures. . There is a growing literature on the possible applications of cost-effectiveness thresholds, although the various contributions fail to consider the application of standards of fundamental measurement, let alone the creation of assumption driven simulated imaginary cost-utility claims to support QALYs and claimed costs-per-QALY for products and comparators [34 35].

Irrespective of the basis on which the numerator as a presumed value for money criterion, the opportunity cost of new expenditure, and value for money criterion the exercise, as a resource allocation or budget tool, it is simply a waste of time. Not only is there little if any agreement on the estimate of opportunity cost as a central planning tool, but the QALY denominator, to emphasize the point, is a mathematical impossibility involving multiplying a modelled estimate of time spent by a composite ordinal score. Even if disbelief was suspended over the actual QALY calculation, there is no agreement on either the choice of multiattribute generic instrument preference or utility ordinal scores from those presently available or the benefits to be gained from making the instrument more comprehensive to capture other aspects of wellbeing as in the case of the proposed EQ-Health and Wellbeing (EQ-HWB) instrument which, despite the title of the overview is not a fundamental measure. The EQ-HWB is, once again, a composite score intended to complement the composite scoring EQ-5D-3L/5L. A cynic might of course make the point that if you are in a hole, don't dig any deeper. An activity that all too many health economists seem willing to pursue given a false belief in measurement properties.

Attempting to assess the impact varying cost-per-QALY thresholds, for those who believe in the QALY having fundamental measurement is, as noted, a singularly useless activity. The classic example of a futile and wasted effort is the debate in the UK over cancer therapies following the decision by NICE to reject submissions because the cost-per-QALY was too high. While the entire episode is an example of what might charitably be called idiotic is amply documented by an exercise of raising cost effectiveness thresholds for end-of-life drugs in the UK (with unfortunate eugenic implications) [36].

But the effects could be more insidious. In Australia the hurdles placed by the PBAC on meeting submission standards for new products may just turn off manufacturers who see the Australian market as too small to warrant any submissions: a feature that may play a key and adverse role in rare diseases. Even if the manufacturer is aware that the PBAC guidelines are meaningless, that may still be not worth the effort if they point it out. What they may find equally galling is the failure of the PBAC to acknowledge this.

CONCLUSIONS

Those who are indifferent toward the truth of their statements are, of course, lacking in epistemic conscientiousness. If there is concern for truth value it is in the interest of both patients , health systems and manufacturers to refrain from engaging in assumption driven simulations to justify imaginary claims for cost-effectiveness. Acceptance, or at least a lack of critical appraisal, of these false premises ensures that the current belief system in HTA is not only a major barrier to progress in our understanding of the impact of therapy options on target patient populations but is an analytical dead end on its own terms. In common with Aristotelian thought with its focus on systematizing and observation, there is no concept or commitment in HTA to the process of the discovery and assessment of new facts: supporting a research program that should be proposed on the market entry of new products focusing on value claims for clinical impact, PRO response and resource allocation. Creating imaginary modelled claims as the ultimate decision criteria is a dead end. There is no need for blanket and non-evaluable QALY-based cost-effectiveness claims; these activities are a waste of time. For Frankfurt, bullshitting involves an indifference to the truth. It's an interesting question as to how this is to be interpreted in the PBAC modelling: is it an indifference to the facts as we know them or an indifference towards the discovery of new yet provisional facts?

For those supporters of the PBAC guidelines for imaginary claims, a rebuttal is in order, not just to defend the fact that it lacks truth value, as the imaginary cost-effectiveness proposition cannot be true or false but an unknown value, but the detailed criticisms levelled here to support the conclusion that we are dealing with a framework of analysis, if that is the right word, which is best characterized, following Frankfurt and Moberger, as bullshit with scientific pretensions.

While this may be pushing the envelope too far, there is broader question that after 30 years and the enthusiastic if mistaken uptake of the HTA meme: what have been the social costs versus social benefits for Australian patients, caregivers and families? Over the past 30 years the PBAC have imposed false criteria for the acceptance and rejection of therapy submissions,

including impacting decision by manufacturers to either withdraw from or not enter the Australian market with new and innovative therapies. What impact has this had? Has there been any change in the health status of the Australian population? Or have we been chasing, amid academic enthusiasm, nothing more than a modelled cost-effectiveness will o'the wisp.

If there is a decision from the ongoing PBAC review to make minor adjustments and continue with the assumption driven model simulations then we are, for the next 30 years, in the somewhat bizarre situation where, in logic, there are by design only unknown values; cost-effectiveness claims, where the claims are neither true nor false. If this is interpreted as an unconcern with the truth, refusing to establishing testable propositions to meet evidence gaps as part of a therapy impact research program, then we might reasonably label the exercise as, once again, bullshit with scientific pretensions.

Against this commitment to only cosmetic changes to the activities of the PBAC is the compelling case for a new start in HTA in Australia which recognizes the requirements of normal science and fundamental measurement in the appraisal of new therapies and tracking these over the product lifetime. The framework for a new start has been detailed in a recently released certificate program by the College of Pharmacy, University of Wyoming: *A New Start in Health Technology Assessment* [37]. This program, in marked contradistinction to the established HTA belief system rests on only three premises. These are:

- All value claims must refer to single attributes for defined patient populations that meet the demarcation standards for normal science: they must be credible, evaluable and replicable
- All value claims, notably for patient or caregiver reported outcomes. must be consistent with the limitations imposed by the standards of fundamental measurement: they must be unidimensional with linear, interval and invariance properties
- All value claims must be supported by an agreed protocol detailing how they are to be assessed in a meaningful timeframe

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

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

[4] 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

[5] 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)

[6] Xie F, Zhou T. Industry sponsored bias in cost-effectiveness analysis: registry-based analysis. *BMJ.* 2022:377 https://www.bmj.com/content/bmj/377/bmj-2021-069573.full.pdf

[7] Bernstein W. The Delusions of Crowds. New York: Atlantic Monthly Press, 2021

[8] Popper K. Objective Knowledge: An Evolutionary Approach (Rev Ed.). London: Clarendon Press, 1979

[9] 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 https://www.valueinhealthjournal.com/article/S1098-3015(21)03146-6/pdf

[10] 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

[11] Australia, Health Technology Assessment Policy and Methods Review https://www.health.gov.au/our-work/health-technology-assessment-policy-and-methods-review

[12] Langley P. Submission to the Australian Health Technology Assessment Policy and Methods Review [February 2024]: Rejecting the PBAC Guidelines for Imaginary Cost-effectiveness Value Claims: A Proposed New Start for Health Technology Assessment in Australia. Maimon Working Papers No. 2 February 2024 https://maimonresearch.com/wp-content/uploads/2024/02/Maimon-Working-Papers-February-2024-No.-2-REV.pdf

[13] Hansson S. *Defining Pseudoscience and Science* in Pigliucci M, Boudry M. Philosophy of Pseudoscience: Reconsidering the Demarcation Problem. Chicago: University of Chicago Press, 2013

[14] Hansson O. Science and Pseudo-Science. *The Stanford Encyclopedia of Philosophy* (Fall 2021) Ed. Zalta E. https://plato.stanford.edu/archives/fall2021/entries/pseudo-science/

[15] 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.

[16] Frankfurt H. *On Bullshit*. The Importance of what we care about: Philosophical Essays. Cambridge: Cambridge University Press, originally published *Raritan Quarterly Review*. 1986; 6(2):81-100 and reprinted as a book *On Bullshit*. Princeton University Press, 2005

[17] Moberger V. Bullshit, Pseudoscience and Pseudophilosophy. *Theoria*. 2020;86:595-611

[18] Hansen H. Fallacies, *The Stanford Encyclopedia of Philosophy* (Spring 2023 Edition), Zalta E & Nodelman U (eds.), https://plato.stanford.edu/archives/spr2023/entries/fallacies/

[19] 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 https://www.valueinhealthjournal.com/article/S1098-3015(21)03146-6/pdf

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

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

[22] Pigliucci M, Boudry M. Philosophy of Pseudoscience. Chicago: University of Chicago Press, 2013

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

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

[25] 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/wp-content/uploads/2024/03/Maimon-WP-No-16-Sept-2023-V2.pdf

[26] Noorden R. How big is sciences fake-paper problem? *Nature*. 6 November 2023

[27] Hume D. An Enquiry Concerning Human Understanding, 1748

[28] Dutilh Novaes C. Argument and Argumentation. *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), Zalta E & Nodelman U (Eds.) https://plato.stanford.edu/archives/fall2022/entries/argument/

[29] Brazier J, Peasgood T, Mukuria C et al. The EQ-HWB: Overview of the Development of a Measure of Health and Wellbeing and Key Results. *Value Health*. 2022;25(4):482-91

[30] Institute for Clinical and Economic Review. Lupus Nephritis. ICER-lupus nephritis-comments-folio_031221.pdf

[31] Schneider P. The QALY is ableist: on the unethical implications of health states worse than death. *Qual Life Res*. 2022:31;1545-1552

[32] Browne J, Coyer D, Stevens W. Is the QALY fit for purpose. *Am J Accountable Care*., 2021; 9(2)

[33] Langley P. Will we finally abandon the mathematically Impossible QALY with H.R. 485 passing in the Senate. *Maimon Working Paper No. 3*. March 2023 https://maimonresearch.com/wp-content/uploads/2024/03/Maimon-WP-No.-3-March-2024-V5.pdf

[34] Sampson C, Zamora B, Watson S et al. Supply-Side Cost-Effectiveness Thresholds: Questions for Evidence-Based Policy. *App Health Econ Health Policy*. 2022;20(5):651-667

[35] Sampson C, Leech A, Garci9a-Lorenzo B. Editorial: Opportunity costs in health care: Cost-effectiveness thresholds and beyond. *Front Health Serv*. 2023: 3.1293592

[36] Collins M, Latimer N. NICE's end of life criteria: who gains, who loses. *BMJ* : 27 April 2013

[37] [37] University of Wyoming. School of Pharmacy. Certificate Program: *A New Start in Health Technology Assessment.* https://www.uwyo.edu/pharmacy/resources/certificate-program-a-new-start-in-health-technology-assessment.html