

Peter Rabbit is a Badger in Disguise: Deconstructing the Belief System of the Institute for Clinical and Economic Review in Health Technology Assessment

Paul C Langley, PhD

College of Pharmacy, University of Minnesota – Twin Cities MN; Maimon Research, Tucson AZ

Abstract

The Institute for Clinical and Economic Review (ICER), a Boston-based consulting group, has seen itself as the lead organization in the US for evaluating pharmaceuticals and, at product launch, making recommendations for pricing and access. Previous commentaries in Innovations in Pharmacy have made the case that the ICER analytical framework is nonsensical. It abandons the standards of normal science in favor of inventing evidence through unsupported assertions regarding measurement properties and lifetime assumption driven simulations. It has been labeled pseudoscience. Yet ICER persists in its belief that all preference scales have ratio properties. ICER believes it can disregard these standards, notably in respect of the axioms of fundamental evidence, and continue its technology assessment activities. Challenging a belief system is not undertaken lightly, although in the case of ICER the belief system is built on such shaky foundations that the effort seems almost superfluous. This deeply held belief, shared apparently by the majority of health economists according to ICER, that all preference scores have ratio properties with a true zero, is easily overturned: if it has ratio properties how is it that preferences scores have been known for over 30 years to recognize health states worse than death? In other words, they can have negative preferences. Recognizing this manifest contradiction is important because it brings into relief the wider belief system of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) who share the same mythical certainties. A key issue is one of cultural relativity: can we accept with equanimity the parallel existence of two belief systems in health technology assessment when one is clearly nonsense? The answer proposed here is clearly no; although unfortunately the blowback by ICER and ISPOR will ensure the survival at least in the near term of their unfortunate meme.

INTRODUCTION

Challenging strongly held belief systems with their mythical uncertainties is not a task lightly undertaken. The penalties for disagreeing can be severe, as Michael Servetus (1511-1553) found to his cost when he stayed overnight in Geneva in 1553, coming inadvertently into the clutches of a vicious and unforgiving John Calvin (1509-1564)¹. A belief system or meme is an intractable position. It does not have to make practical or even logical sense. It can be replete with mysteries; the more they are obviously nonsensical only makes the belief even stronger. An attempt to leave this citadel of belief has significant downsides: ostracism, rejection of published papers, failed marriages and denial of promotion and tenure. Not quite matching the threat of the inquisition and subsequent stage managed *auto da fé* events with green wood to extend the agony, yet the penalties can be severe.

For our purpose, a key question is why this dominant meme in health technology assessment, characterized as the rejection of hypothesis testing in favor of the invention of approximate information, came into being. While not quite a Damascene conversion, the epiphany, if that is the right word, came from a concern with the limited evidence base to support product

cost-effectiveness claims at product launch in the early 1990s by thought leaders in the nascent subject area of health technology assessment. They decided that hypothesis testing would be put to one side in favor of inventing approximate information to support formulary decision making through lifetime simulation modelling². It also was a lot quicker.

But that was only the start. To ensure survival you need to develop a creed, badger beliefs, to support this nascent meme; to ensure transmission fidelity and attract converts. This was accomplished through the development of techniques, with a supporting language; to ensue adherents did not stray too far from the core activities. At the same time, care was taken to minimize apostasy. Certain questions were off limits: notably discussions of the applications of the scientific method, including the question of limitations imposed by the axioms of fundamental measurement. Quality adjusted life years (QALYs) became the holy grail of modelling. Measurement theory was never discussed.

For the Institute for Clinical and Economic Review (ICER) the badger belief assertion that ordinal scales have hidden ratio properties is a non-starter for any disputation. The ingrained belief that the ordinal preference scale has hidden ratio property simply rejects the question out of hand. Peter Rabbit is, without question, a badger in disguise. It admits of no empirical evaluation; it is asserted. Generations of children might object and in tears disagree, but ICER will stand firm in its badger belief. This would apply to the author, Beatrix Potter (1866-1943), who might challenge this rewriting, as well as her

Corresponding author: Paul C Langley, PhD
Adjunct Professor, College of Pharmacy
University of Minnesota, Minneapolis, MN
Director, Maimon Research LLC; Tucson, AZ
Email: langley@maimonresearch.com
Website: www.maimonresearch.net

publishers. Perhaps, as explored here, Peter Rabbit might point out to the badgers of ICER than preference scales can generate negative utilities or values; a fact that has been evident for over 30 years. This means, as detailed here, that there cannot be, by definition, a ratio scale without a true zero or an ordinal scale with hidden ratio properties unless the ICER belief system is so ingrained that this belief is possible: a veritable faith based mystery.

We might even rephrase this strange belief system of scales with negative properties having magical ratio properties in terms of the Bertrand Russell (1872-1970) and Ludwig Wittgenstein (1889-1951) Cambridge encounter in 1911 where instead of a rhinoceros or an elephant in the lecture room we have Peter Rabbit as a badger, where Russell might raise the question “there is not a badger disguised as Peter Rabbit in the lecture room”. For interest, it might be noted that the *Tale of Mr Tod* was published within a few months of this encounter³.

THE ICER/ISPOR TRUTH IS CONSENSUS

Judged by a number of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) good practice guidelines, the invention of evidence through simulation modelling was seen as the only practical approach; evidence for cost-effectiveness claims and an analytical framework to deliver the imaginary goods had to be invented. At no stage over the past 20 years has ISPOR even produced: (i) a practice guideline that focused on the evidentiary standards of normal science for credible, empirically and replicable claims; or (ii) a practice guideline that focused on the evaluation of patient reported outcome (PRO) claims in terms of the axioms of fundamental measurement. This latter point is important because of the neglect of PRO instruments that allow a meaningful response to therapy⁴. That is, the neglect of PROs that meet the required standards of fundamental measurement and emulate the measurement protocols of the physical sciences⁵. Absent these standards, PRO instruments should be put to one side.

The ISPOR truth by consensus meme puts to one side the standards of normal science and any consideration of the limitations of fundamental measurement; a statement or assertion is true simply because people generally agree with them. An ISPOR team will decide on the merits of any debate in these imaginary scenarios. We have the commitment to imaginary simulated claims which cannot and never were intended to be empirically evaluated (let alone expressed in credible empirically evaluable terms). Imaginary claims, with the security blanket of sensitivity analysis, are in badger belief, both necessary and sufficient for formulary decisions.

BARRIERS OF LANGUAGE

The language of a belief system or a meme can act as a barrier to criticism; even mutual incomprehension. This can raise a can of worms when we consider language and its application in the 20th century. The chief protagonist here is Wittgenstein in his

belief that certain, otherwise simple empirical questions, can be considered out of bounds: the question has no meaning.

If we accept Wittgenstein’s proposition that ‘the limits of my language is the limits of my world’ then language, notably the introduction of technical terms to support a meme and its application become important. Although often sharing a common language, the relativists and realists in health technology assessment have a number of barriers in place to repeal unwanted ideas and concepts. To the extent that the language determined beliefs and the limits of enquiry, it can also reinforce belief. We will consider the Karl Popper (1902-1994) and the Wittgenstein Cambridge encounter of 1947.

THE QALY AS HOLY GRAIL

It is difficult not to understate the hold that the quality adjusted life year (QALY) has on the ISPOR and ICER meme for technology assessment. Incremental cost-per-QALY imaginary simulations are the lodestone of technology assessment. This is easy to understand: take away the unconstrained badger belief in the QALY and the edifice of health technology modelled assessments collapses. Samson can demolish the Temple. It is sad to think that generations of eager health technology enthusiasts have embraced the QALY without recognizing it is a mathematically impossible construct – unless you believe that ordinal preference scales have hidden ratio properties. This represents the singular failure in the ISPOR/ICER meme; a failure to face simple logic and the impact of levels of measurement. Belief trumps logic.

Without being overly tedious, the argument is simple: preferences, whether utilities or values, are on an ordinal scale; they are ordinal scores⁶ This applies to both direct and indirect preference scales.. Nothing has been put forward to dispute this position. As such, you cannot multiply a ratio measure (time) by an ordinal score. It is mathematically impossible. Of course you can as these are just numbers, but it is meaningless. If you want to combine a measure with another measure, then both have to have ratio properties. We have known this from Thurstone’s contributions of the early 19th century and Stevens’ contributions of the 1940s^{7 8}.

One interpretation of this disaster is simple: leadership and ignorance. Those developing multiattribute instruments and proselytizing in favor the time trade off (TTO) and standard gamble (SG) techniques, had no idea that they should be aiming for a measure with interval and, hopefully, ratio properties; a measure in the latter case that could support the QALY holy grail where death is not an arbitrary zero but a universal constant. Otherwise they could just sell tickets for an ordinal score raffle, which could present negative values and a cooked chicken with their choice of zero.

THE ICER MEASUREMENT CREED

To protect its use of the QALY, ICER has a badger belief creed; a set of unsupported assertions that guide, indeed dictate, ICER's aims. As stated:

*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*⁹.

This creed has a confused deconstructive message:

- The creed is in terms of belief based on a failure to understand the standards of normal science, including fundamental measurement;
- The creed acknowledges an internal contradiction where a ratio scale can have negative preferences;
- The creed believes that the dead state represents a natural zero point although is not clear if this is viewed as a universal constant (a true zero on a ratio scale) or just an arbitrary zero which can be applied to an interval scale;
- The EQ-5D-3L/5L and other multiattribute instruments are believed to have hidden ratio properties although they are ordinal scales (a badger in disguise);
- The time trade off (TTO) scores are believed without proof to have interval properties (another badger);
- These unproven interval properties mysteriously transmute to a ratio scale (mystical badger);
- Ordinal scales are believed without proof to have ratio properties (a badger mystery);
- It is believed to be perfectly acceptable to add up scores for different attributes to create a single score;
- Preference scales such as the EQ-5D-3L/5L can generate negative values or states worse than death;
- The 0 = death point on an ordinal scale ensures it has ratio properties consistent with negative preferences;
- Negative preferences on a ratio scale have face validity;
- The EQ-5D-3L/5L multiattribute scales were designed to have face validity while allowing negative preferences or states worse than death consistent with their ratio properties.

As detailed below, these are nothing more than a series of assertions with no attempt to consider them as empirical propositions. This is a weak defense of an indefensible position.

CULTURAL RELATIVITY

One perspective on why ICER (and ISPOR) continue to deny the relevance of hypothesis testing with their advocacy of imaginary simulations, together with the denial of the axioms of fundamental evidence, is to consider belief in terms of cultural relativity or, in the vernacular, anything goes. Under what is described as the 'strong program'; as Wootton describes it, the position taken is that the values and aspiration of scientists answer to a sociological explanation¹⁰. This rests on the notion of symmetry between beliefs. If beliefs differ then they should be evaluated in their own terms, not by reference to standards held by other belief systems. Badger belief systems should not be evaluated by the standards of non-badger belief systems. Put simply: rationality is always culturally relative. The strong program, goes further, *denying that scientific claims are ever adopted because they fit the evidence better than the alternatives* (p. 580)¹⁰. This postmodernist view, in common with most postmodernist views, is nonsense. If the view that all beliefs are equivalent is accepted then we must accept ICER's position that the ratio scale can accommodate a non-true zero and negative preferences. This may violate, even in simple logic, the axioms of fundamental measurement but in this parallel badger belief system it makes perfect sense. But, as Wittgenstein is reported to have pointed out, the logic of mathematics may not be relevant; in this lecture room $2 + 2 = 4$; in the next lecture room the belief is that $2 + 2 = 5$. As Wootton continues: *To insist that the issue of validity must be separated from the issue of credibility is to insist that well-founded beliefs be treated as if they are unfounded beliefs* (p. 582)¹⁰. The belief, held by generations of children that Peter Rabbit is not a badger in disguise (the issue was never raised by Beatrix Potter) should be treated as on a par with the assertion that Peter Rabbit is a badger in disguise (absent any evidence to support this proposition). The beliefs are equivalent and entirely rational.

If ICER supporters are to be labelled 'relativists', in the badger camp, the question then arises: is this a relativism of convenience to support a misguided analytical framework while supporting a successful business model with attendant accolades or are they all true badger-believers? Should ICER/ISPOR be viewed as a cete of badgers? A cete of true believers that judge their analytical approach (the term is used loosely) only within their own belief system. Questions that are raised by reference to a non-badger belief system (e.g., claims with empirical credibility) are not only irrelevant but make no sense in their terms. It is an article of faith, interpreted in sociological terms, that the creed holds: there is a ratio scale with a true zero (death = 0) yet, perhaps in contradiction, also preference states worse than death. This is not uncommon in belief systems (or memes) to hold mutually contradictory beliefs; indeed such mysteries, rife in religions, may engender even stronger belief.

What ICER holds as a central tenet of their analytical framework is clearly relativist: evidence is never discovered, only

constructed within a particular social community that joins in subscribing to a common belief system. This wider community is, of course, ISPOR and supplicant community of those who believe in creating approximate information as opposed to hypothesis testing. No one body of evidence is superior to another; approximate information is equivalent to hypothesis testing if we are to 'inform' decision makers. Success, in these terms is judged on the ability to garner support. The EQ-5D-3L/5L may be failed constructs, but belief in these scales remains, supported by an effective outreach program of believers or missionaries, in some cases by those with a financial interest (and those refusing, for one reason or another, to abandon their belief). Within this community 'truth is consensus'; a position held by ICER in its assertions regarding measurement theory and the confidence of badger believers. For the relativist a belief system cannot be overturned by new evidence; the badger belief system is nothing more than rhetoric, persuasion and authority. The search for new knowledge is irrelevant; laurels can be rested on the creation of assumption driven imaginary simulations.

To the non-badger belief system, with its commitment to evidence, hypothesis testing and the discovery of new facts, stretching back to the 17th century, the badger cete is logically and linguistically indefensible; indeed, an object of scorn. Care has to be taken; one might object to Wittgenstein's view of language, but yet agree with him on the importance of evidence and the scientific method. As detailed in the *Tractatus* and in *On Certainty*, Wittgenstein was no relativist (e.g., his rejection of logical positivism) and, even though taken as an inspiration by badger believers, he would not only have rejected relativism out of hand but also claims built on imaginary or existential simulations^{11 12}.

We should not, as Wootton notes, confuse the symmetry principle where good and bad science can be explained in the same way with impartiality, where failed science should be studied as carefully as successful science¹⁰. This is a major theme of this commentary. Focusing on a sociological defense of a failed science, and the ICER framework is clearly an analytical dead end, but should not excuse us for not assessing the reasons for failure as impartially as we can. Perhaps only an impartial resolution will keep the genie in the bottle (or the badgers in their cete). After all, for the majority of ICER's audience, the analytical framework is still a black box.

HIPPOPOTAMUS IN THE LECTURE ROOM

The issue of linguistic convention still remains. This is illustrated, in the context of evidence, in the confrontation in 1911 in Cambridge between Bertrand Russell (1872 - 1970) and Wittgenstein as to whether there was a rhinoceros (or elephant) in the room, with Wittgenstein refusing to admit there was not. It hinges on the nature of evidence and whether belief or a belief system can ever be overturned by an appeal to evidence. To paraphrase Russell's recollection:

Wittgenstein maintained that all existential propositions are meaningless. As this was a lecture room I invited him to consider the proposition that 'There is no hippopotamus in the room at present'. When he refused to believe this, I looked under all the desks without finding one, but he remained unconvinced'¹⁰.

The encounter, was commented on by Russell "My German engineer, I think, is a fool. He thinks nothing empirical is knowable. I asked to admit there was not a elephant in the room, but he wouldn't"¹³ [note: it is unclear from Russell's recollection whether there was not a elephant or not a hippopotamus]. From the ICER badger perspective, would Wittgenstein have refused a response if Russell has said 'There is not a badger in the room'? While we can agree that the badger is an existential entity, the reasons for Wittgenstein's response (or lack of) is not clear. Badgers, like hippos are rarely found in Cambridge lecture rooms.

One explanation is that Wittgenstein thought the statement made no sense and hence saw no reason to answer. Consider the distinction between a proposition and an assertion. A proposition is a statement that it makes sense to ask whether it is true or false; an assertion is a statement that one claims is true (hence the ICER creed). Assertions are more restrictive than propositions (example: "It's snowing" is a proposition not an assertion) but if I open the door and say 'it's snowing' I am asserting the proposition is true; hence it is an assertion. Everything that is a proposition can also be an assertion while all assertions are not necessarily propositions; i.e., a claim that may be true or false. While the distinction is contextual, it is clear that the ICER belief system in measurement is a series of assertions with no intention of opening them up to refutation. For Wittgenstein there are meaningful and non-meaningful assertions, which leads to the question of a background to the proposition. What Russell thought was a proposition was not, for Wittgenstein, in fact a proposition or assertion. Hence the term used by Wittgenstein 'asserted proposition', as an everyday statement that can be true or false as compared to one that is nonsense.

AN UNCOMMON GROUND

Whether one agrees with Wittgenstein regarding the concept of a form of representation, for our purposes it is the distinction between a proposition and an assertion that is of interest. If a belief system rests on assertions, by assertion truth is consensus, then there is little common ground for arguing for propositions. Importantly, it is not just a belief system built on assertions that is a barrier, but the structure of the belief system and its language. To believers in the non-badger belief system an analytical framework that rests on core assumptions that are clearly false and develops techniques such as probabilistic sensitivity analysis (PSA), is nonsensical; techniques that do nothing other than act as a cover for assertions and assumptions that deny elementary logic.

Amongst a global community that believes in the scientific method and the role of measurement in instrumentation to support hypothesis testing, we have a nested global community of non-believers. Relativists in health technology assessment who have accepted an entirely different reference framework; one that predates the scientific revolution in constructing imaginary worlds with a complete disregard for empirically evaluable claims or propositions in favor of assertions. Whether understood by the leaders in health technology assessment, their decision to abandon hypothesis testing in favor of inventing approximate imaginary information to support formulary decisions is a denial of the scientific method; a major reversal of any commitment to the scientific method. Indeed, an implicit acceptance of verification, as a central analytic of logical positivism, with the consequence that all mathematical proofs and logical inferences are tautologies. As such, they give us no information about the real world, they are devoid of substance and are only about the internal relationships of the proof. This seems apt when deconstructing the ICER badger belief system, taken with the standard paragraph in ICER evidence reports that verification is either in terms of the internal structure of their model or its comparison to other models in the same disease area. Claims are made, but none meet the required standards in normal science of credibility, empirical evaluation and replication. The thought that verification has been long been rejected in terms of falsification and probabilistic confirmationism seems not to have occurred to them.

To a realist, the badger technology assessment belief system is truly Orwellian; O'Brien would clearly approve. You might hold up 3 fingers but if you are told there are really four you will assert (and actually see) there are four. There is no evidence from an external reality that would shake that belief; it is disallowed¹⁴. More to the point, the relativist position is an embarrassment. For the first time in 400 years since the invention of science in the 17th century, we have a commitment to imaginary claims in decision making by analysts that share a common language, endorsed by leading academic groups and applied by government agencies such as the National Institute for Health and Care Excellence (NICE) in the UK. This makes the intractability of the meme, the embrace of facile ideals, psychologically fascinating. This is not just the belief that direct and indirect preference scales have, not just hidden interval, but hidden ratio properties, but that ordinal preferences are an integral part of the creation of imaginary claims for cost-effectiveness. This, again, is clearly in the tradition of denying the relevance of external evidence; the relevance of accepting propositions. We are now dealing with a bloat of assertive hippos (hippopotami) [or a crash of rhinoceroses] in the belief, commitment and publication of imaginary claims.

To argue for cultural relativity ignores how the language for what we now call science has evolved¹⁰. The notion of science, in the 17th century, was a process of a growing vocabulary; science was invented. Observations were made for which there

was no nomenclature; cloud classification, planets and moons and so forth. So the language evolved to accommodate these and establish a common vocabulary (at first in Latin), and continues to do so. The downside is that a language and vocabulary can also emerge that supports a relativist rather than a realist belief system. If a language does not evolve then events such as the scientific revolution or the rapid acceptance of Einstein's theories will not occur¹⁵. Certainly, a static language cannot support a commitment to a dynamic social science; the need for an evolving vocabulary to capture new observations and techniques. Communication and rational thought are only possible between people when there is an agreed and comprehensible common ground.

A POKER EVENING

If we fast forward some 36 years to the evening of 25 October 1946 and a meeting of the Moral Sciences Club in Cambridge, we come to the famous (or infamous) poker confrontation where, in a heated argument with Popper, Wittgenstein is believed to have waved a poker at Popper, shouting 'you are wrong'; Russell took the poker and Wittgenstein stormed out of the room slamming the door.

Wittgenstein had invited Popper, possibly at the suggestion of Russell, to discuss 'philosophical problems'. The details of the actual debate are elusive but have been characterized by one reviewer as a row over the role of philosophy in addressing real problems (Popper) as opposed to linguistic puzzles (Wittgenstein)¹⁶. The debate over the debate continues with no end, at least from the Wittgenstein perspective, where advocates maintain that Wittgenstein is too complex a philosopher to categorize neatly (or understand) the complexities of language in communication. In contrast, the overwhelming majority of practical scientists and those in the philosophy of science mainstream, see Popper's focus on real problems, the growth of knowledge, and their evaluation as winning.

From our perspective in health technology assessment, the case should be closed. The question is whether in rejecting Popper and taking a relativist position on the invention of simulated evidence, ISPOR/ICER have inadvertently rekindled this debate. If the issue is about belief, language and grammar, a relativistic position that holds truth is consensus and that one belief system, or language including internal contradictions, is as good as another, then we have just descended into solipsism where ICER presents an assertive linguistic defense of the indefensible. This may appeal to those who want to support the ISPOR/ICER imaginary belief system (whatever their motives), it certainly does not appeal to those who take science seriously, agreeing with Popper, and point to the failure of ISPOR/ICER to advance beyond pseudoscience¹⁷.

As Lakatos (Imre Lakatos 1922-1974) points out:

*For centuries knowledge meant proven knowledge – proven either by the power of the intellect or by the evidence of the senses. Wisdom and intellectual integrity demanded that one must desist from unproven utterances and minimize, even in thought, the gap between speculation and established knowledge*¹⁸.

What actually transpired at the Cambridge meeting is, as noted, hazy and subject to mixed recollections, including the accusation that Popper ‘embellished’ his recollections, even to the extent of lying. The evening started with Popper presenting a paper ‘Are there philosophical problems?’ or, as a subtext, are their problems as opposed to what he saw as Wittgenstein’s position that there were only linguistic puzzles. According to the account by Edmonds and Eidinow, Wittgenstein’s preoccupation was with how we use words, how language functions, how to give words meaning to avoid the systematic misuse of language; are there statements which defy the logic of our language¹⁹? Other so-called philosophical problems could be dismissed as questions of logic and mathematics, not philosophy. One question raised by Popper was how we obtained our knowledge, not only about the external world but about ourselves, about the nature of reality? This was surely not by induction? Logical positivism and verification, the so-called Vienna School, had long been abandoned, with Popper and Wittgenstein prime architects. Apparently, the discussion came to a head when Wittgenstein asked for an example of a moral problem that philosophers could address; Popper apparently replied ‘Not to threaten visiting lecturers with a poker’; hence Wittgenstein’s response. However, while Wittgenstein is believed to be an important impetus in the relativist emergence in the 1960s, statements from the *Tractatus* and *On Certainty* would not only suggest otherwise but also point to the convergence between Wittgenstein’s and Popper’s views on induction and hypothesis testing as early as the 1920s.

Popper was correct in raising the question of the reach of philosophical problems. Certainly the use and abuse of language is critical, contexts where meanings are poorly articulated, as in the badger belief ICER response that we are ‘confident’ in our assertions, and we do not find this ‘lacks face validity’ and, most concerning, the view that ‘most health economists’ believe ordinal scales have ratio properties. These can be dismissed out hand as assertive nonsense, pointing to a cavalier misuse of language, to the extent of being deliberately misleading.

Given this, Popper is also correct in pointing to a range of philosophical problems. In particular, and central to our thesis, the question of progress and the growth of knowledge. We cannot address this issue in terms of the misuse of language, falling back to a counter-productive relativistic philosophy, but

through a commitment to the discovery of new facts; not old assertions recycled by ICER through imaginary constructs which fail Popper’s demarcation test.

So, one way of viewing what transpired is agreeing that science and philosophy can tell us new things about the world, but only by putting to one side relativism, pseudoscience, mythical mysteries and linguistic games. Perhaps the most important contribution is to focus on demarcation: what criteria distinguish pseudoscience (bunk) from science?¹⁸ . The distinction may not be hard and fast, but it certainly relegates the ISPOR/ICER activities and beliefs to the Dover Courtroom to join and defend intelligent design.

Wittgenstein in the history of science is important because while he would not have described himself as relativist (if that term had any meaning for him), the influence of his assumed relativism is taken for granted; this gives relativism an uncalled for credibility¹⁰. Wittgenstein clearly believed that we can separate good science from bad science, where one of the defining criteria is the support for progress and, in Kuhnian terms (Thomas Kuhn, 1922-1992), evidence driven paradigm shifts²⁰. The unfortunate aspect is of course that Wittgenstein’s arguments are often puzzling, contradictory and often allow more than one interpretation; hence the interpretations that surround the hippopotamus/rhinoceros event. The key point is that paradigm shifts are impossible within the badger belief system.

PROTECTING SIMULATED ASSERTIONS

The invention of evidence through assertions and assumption driven lifetime simulations, supported by the varied applications of modeling techniques is designed, not to produce evaluable claims which admit of an appeal to external evidence, but to smother them in a blanket of plausibility²¹. This is made clear by ensuring that the claims are impossible to ‘validate’ in ICER terminology, other than within the model itself. Discounted lifetime incremental cost-per-QALY claims are magical constructs which defy any form of external assessment, unless in some bizarre universe one is prepared to create a 30 to 40 year data set to challenge them. Certainly, model assumptions may change with ‘new facts’ from clinical trials and new publications, but these are just subsumed within a selective analytical framework that uses these to create more imaginary non-evaluable claims on the future.

Applying assumptions still defies elementary logic. It is one thing to deny the role of an appeal to evidence to support propositions; it is another to ensure that such an appeal can never take place. In other words; we make sure there is no hippopotamus (or disguised badger) in the lecture room to start with. We are left with a discipline focused on creating non-evaluable assertions with a probabilistic tag; assertions which can never be considered propositions.

The fact that alone amongst the social sciences, health technology assessment embraces an assertion driven relativistic belief system is, as noted, an embarrassment. For 30 years the leaders of this relativistic and quixotic world have denied the role of hypothesis testing in favor of creating non-evaluative assertions, approximate imaginary information, the parallels with quasi-religious belief systems are too close which is why the term meme is more than apt. With organizations such as ISPOR the support for this belief system is well entrenched as evidenced by the publication of practice guidelines for the young model builder to replicate imaginary claims. Transmission fidelity is high, supported by the existence of mysteries such as the impossible or I-QALY which merely serves to reinforce belief. As noted, there is no ISPOR publication on the standards of normal science or the limitations imposed by the axioms of fundamental measurement; perhaps that cat is best left in the bag. It is only recently that the denial of fundamental measurement appeared as a letter to the Editor in the ISPOR house journal *Value in Health* ²². The message was quite clear: 30 wasted years in technology assessment and a willingness to reject the axioms of fundamental measurement. The letter asked, possibly over optimistically, that *Value in Health* might inform its readership.

The relativist position is anti-science yet the ICER/ ISPOR belief persists; it fails to recognize (or refuses to recognize; or doesn't want to recognize) that the purpose of science, the progress of science, is to seek provisional new knowledge through, as Popper in his early work would describe it, a process of conjecture and refutation ²³. A dynamism that admits of paradigm shifts when belief systems are challenged and a crisis sets the stage for a new consensus. Dynamism, the possible emergence of a new consensus, is absent from the static and unchangeable analytical framework of the relativistic meme; a virus of the mind ^{13,14}. Consider the use by ICER in their creed of the word 'confident'. In the belief systems prior to William Gilbert (1544? - 1603) we find confidence in beliefs aplenty, but no widespread commitment to experimentation or even the advance of ideas (beliefs) in a form suitable for experimentation. Indeed, if we consider the language of science and its evolution through the 17th century, the concept of experimentation was resisted. As Wootton makes clear. for mediaeval philosophers such as William Grossteste (c.1175 – 1253):

Experience and experimentation were thus invoked only to fill gaps in a fundamentally deductive system of knowledge, never to question the reliability of deductive logic itself; and these gaps were always of limited significance within a curriculum centered upon Aristotle's texts ¹⁰.

For ICER/ISPOR shared mythical certainties resist intellectual speculation.

DECONSTRUCTING THE ICER CREED

In order to illustrate the lack of any redeeming features in the ICER assertive creed, consider the following five fundamental errors in logic and measurement:

- assumptions to support lifetime modelled simulations
- ignorance of attributes and the axioms of fundamental measurement
- failing to recognize that time trade off can only create ordinal scales
- failing to recognize that the QALY is an impossible mathematical construct
- failing to consider the implications of states worse than death

ASSUMPTIONS: WHAT WHITE SWANS?

It is a fact of elementary logic that what has happened in the past cannot lay claim to what might happen in the future. Presumably to those in ISPOR/ICER, the issue of induction (Hume's problem; David Hume 1711-1776) was not an objection standing in the way of developing simulations of imaginary therapeutic futures, driven by past assumption, populated by I-QALY claims, to produce pricing and access assertions that had no empirically evaluable content. Put simply: the ISPOR and ICER models are derailed by their ignorance of induction and the demise of logical positivism by the mid-20th century. It is not as though this is a recent proposition; Hume proposed in the 18th century and Wittgenstein went to lengths to demolish it in his *Tractatus*, followed by Popper and his focus on conjecture and refutation. These are, apparently, concepts that are entirely foreign to ISPOR/ICER.

While we might consider modifications to Popper and the rise of confirmationism, where scientific theories are defined by the accumulating probabilities of experimental success ²⁴, the fact remains that the commitment to assertions and an assumption driven future imaginary world is logically indefensible and an analytical dead end. We cannot confirm non-evaluative assertions. There can be no experimentation. There is no progress. If you assert, on past observations in the UK, that all swans are white, consider a vacation in Western Australia where they are black; unless of course you retreat and insist on defining a swan as a bird with white feathers. It is assumed the black swans are unconcerned. The badger belief system should collapse on these terms alone.

ATTRIBUTES AND THE AXIOMS OF FUNDAMENTAL MEASUREMENT

In the absence of ISPOR practice guidelines on the application and misapplication of the axioms of fundamental measurement, a primer is in order. As ICER in its assertions appears confused on the agreed levels of measurement, as recognized in the physical sciences and the more mature social sciences, the confusion should be addressed even at this late

stage. Following the formalization by Stevens and others in the 1930s and 1940s, scales used in statistical analyses are classified as nominal, ordinal, interval or ratio⁷. Each scale has one or more of the following properties: (i) identity where each value has a unique meaning (nominal scale); (ii) magnitude where ordered values on the scale have an ordered relationship with each other but the distance between each is unknown (ordinal scale); (iii) invariance of comparison where scale units are equal in an ordered relationship with an arbitrary zero (interval scale) and (iv) a true zero (or a universal constant) where no value on the scale can take negative scores (ratio scale). The implications for the ability to utilize a scale to support use of arithmetic operations (and parametric statistical analysis) are clear. Nominal and ordinal scales do not support any mathematical operations; only nonparametric statistics. Interval scales can support addition and subtraction while ratio scales support the additional operations of multiplication and division as they have a true zero. This zero point characteristic means it is meaningful to say the one object is twice as long as another. To measure any object on a ratio scale it has to be demonstrated that all criteria for an interval scale have been met with a true zero. These conditions are, it appears, totally foreign to the badger discipline. Indeed, the suspicion is that there is no perception of a pressing need to open this measurement window. Defenestration is long overdue.

Importantly, in fundamental measurement ordinal scales together with interval and ratio measures refer to single attributes; not to scales which attempt, as in the case of multiattribute instruments such as the EQ-5D-3L, to create an aggregate score that combines individual scores for each attribute. What was ignored (or overlooked) in the instrument development was that each attribute measure (e.g., pain, mobility) must have dimensional homogeneity, unidimensionality and construct validity. This has to be demonstrated in the development of true measurement scales with intended interval or ratio properties. Assertion (or unfounded belief) is not proof; it may be an article of faith among some (but certainly not the majority of) health economists, let alone mainstream economists, but that leads us nowhere (apart from wars of religion). If an instrument is to have a given property this has to be built into its design and development. In this respect it is important to distinguish the term 'score' from 'measure': a measure refers to a scale that meets interval or ratio properties while score refers to a scale with nominal or ordinal properties. In both instances the interval or ratio scale must be for a single attribute, physical or latent, to meet the standards of fundamental measurement. The attribute must have coherence, relating to key concepts (e.g., need-fulfillment quality of life²⁵) and an assessment of measurement options (e.g., Rasch measurement theory²⁶).

These properties have been accepted for decades yet have been ignored, by design or otherwise in the development of direct and indirect preference scales for health technology assessment. The first step promised disaster: attempt to take a

health state description comprising a cluster of attributes and assuming, incorrectly, that this could be valued either as a single description (TTO) or by a fixed set of attributes (EQ-5D-3L). The ready acceptance of these preference scales, without recognizing their failure to meet fundamental measurement standards has had, in retrospect, a disastrous impact on health technology assessment. A situation where an organization such as ICER, following faithfully as a disciple of ISPOR, is promoting a relativistic solution for formulary pricing and access decisions. If there is a major failing, among many others, it is the mistaken impression that ordinal scales are magically transformed to interval scales. If you place the scores on a number line with equal intervals, you can then claim a metamorphosis into a ratio scale. You could equally well put the scores on a number line with unequal intervals. A ratio scale, in the absence of a natural true zero, is a more complex undertaking, in particular with latent constructs. This negates decades of experience in the physical, medical device and more mature social sciences. ISPOR appears prepared to endorse this charade with ICER lapping it up.

Once the criteria of fundamental measurement are applied, not only to multiattribute generic instruments but to the menagerie of published PRO instruments specific to disease areas or target patient populations, it becomes readily apparent that the overwhelming majority of PRO instruments fail to meet required measurement standards. The reason is obvious: no one thought that fundamental measurement was relevant. All that had to be done was to add up scores from the various component scales (usually Likert) that compose the instrument and present an aggregate score to assess response to therapy. No one thought of emulating the instrument development standards of the physical or more mature social sciences where the focus is on measuring specific attributes.

ORDINAL TIME TRADE OFF SCALES

Time trade-off (TTO) creates ordinal scales. While some would like to think it has interval and even ratio properties it fails on two counts: (i) the TTO can produce states worse than death so that it fails to have a true zero or any pretense to a ratio scale; and (ii) as the TTO is applied to a health state description, this description is multiattribute and not a single attribute. This means it is impossible to capture the description as a 'single value' the multiattribute description is dimensionally heterogeneous, lacking unidimensionality and construct validity. Even minimal descriptive changes can result in different scores. To this we should add that respondents find the TTO valuation exercise cognitively challenging with responses clumping towards the presumed end points and the middle of the hypothetical range. We cannot assume that it produces an interval scale; a conclusion that is reinforced by the lack of concordance by patients with the behavioral assumptions for respondents that underpin the scale. The current view is that any claim that the TTO has interval properties is by assertion and unsupported belief²⁷.

THE IMPOSSIBLE QALY

The centerpiece of the badger belief system is the I-QALY; a construct, as noted, that is impossible (hence I-QALY) as it requires multiplying a ratio measure (time) by an ordinal score (preference); typically a multiattribute score which lacks dimensional homogeneity, unidimensionality and construct validity⁵. Even if it were possible to claim that the multiattribute preference scale had interval properties, the impossible or I-QALY is still the outcome because to support the QALY construct through multiplication you need a preference ratio scale; one that was designed to have ratio properties. An example of this is the proposed need fulfillment or N-QOL measure, a bounded ratio scale that meets the required measurement properties and can generate estimates of needs adjusted life years (NALYs)²⁸.

While the impossible or I-QALY construct, given an appreciation of measurement theory, is obvious any attempt to recognize it would destroy the centerpiece of badger belief: the incremental cost-per-QALY and the application of cost-per-QALY thresholds. Simulation models are designed, in their commitment to generic scales, to generate non-evaluable QALY claims. If the QALY is cast aside, in a truly biblical sense, then a methods framework such as that carefully detailed in the leading textbook, with its assertions and techniques, is redundant²³. A major element of the relativistic belief system does not collapse through external evidence (which it ignores), but from its internal contradictions. Marx and post-modernist disciples such as Foucault, were they alive, would be pleased.

STATES WORSE THAN DEATH

We have known for many years that all direct and indirect preference instruments create negative utilities or states worse than death²⁹. In the case of indirect multiattribute instruments this is the result of attempting to fit equations to the data rather than selecting items to fit the model. This is not just a 'rare' occurrence; in the case of the EQ-5D-5L a recent study found that for five countries the percentage of negative valuations of health states ranged from 9% to 33%³⁰. There is also evidence that for negative values with the EQ-5D-5L, these values show little discrimination regarding the 'severity' of the state worse than death, random noise and other factors dominating the valuation. If the belief is that the preference scores are on number line with interval properties in the range 0 to 1, the presence of states worse than death must disabuse badgers of this belief. These cannot be ratio scales because there is no true zero, a natural minimum quantity, below which a valuation cannot occur.

Let's be quite clear about the implications of states worse than death. If we subscribe to the belief that the QALY is a meaningful this must rest on a belief that any preference scale, for either direct or indirect values or utilities, must in turn have a true zero. If this condition is not met, under any circumstance, then the preference scale is, at best, an interval scale although this has to be proved. As a default then, we must assume that

the preference scale is ordinal. This holds irrespective of the health state decision, whether defined in terms of a descriptive bundle of attributes (SG and TTO), or a more 'refined' and formal multiattribute instrument (e.g., EQ-5D-5L, AQoL). In the latter case preferences are created from an arbitrary starting point of unity (perfect health defined by no problem with any symptom) utility decrements are driven by rules and preference weights (described as TTO tariffs) with preference applications approaching zero (death). Unfortunately, in an effort to hit the zero point, all multiattribute scales 'overshoot' (undershooting is not allowed). They yield negative preferences. This does not mean that certain health state descriptions will necessarily yield a negative value or utility, but that at least one respondent will attach a negative preference to one health state. This decision does not reflect just the description but can include attitudes to risk, cognitive understanding, possible interdependence between symptoms and other personal and environmental factors. If negative preferences exist, then we have the interesting theoretical proposition that in the limit the lower bound of the preference scale in minus infinity. Thus dividing any preference score by the scale denominator (minus infinity to 1 or just minus infinity) yields zero. Of course, the instrument rule makers may try to set negative lower bounds and even resort to continual tweaking of regression models to get a better 'fit' to their data and hopefully eliminate the likelihood of the pesky negative score. Unfortunately, this is a wasted effort: what needs to be proved is that under no circumstance can any respondent to the instrument return a negative preference score. Such a proof is impossible. However remote, the likelihood exists; there is no universal constant defining the true zero. There can be no true zero and hence a ratio scale argument for either direct or indirect preference elicitation is untenable.

To this would be added the mistake of creating a scoring algorithm in the first place. As noted, measurement refers only to single attributes. If a combination of attributes is proposed then each attribute must be on a ratio scale. This is not the case with the multiattribute instruments as witnessed by the EQ-5D-3L where the five symptoms or attributes are all on 3-response level ordinal scales and reflect merely clinician preferences for summary descriptions of general health. Moving to a 5-level has proved to be disastrous as the scores are different with failed attempts to map between them^{31 32}. An activity that is doomed as the scales are ordinal. You can only 'map' between ratio scales and, in certain circumstances, between interval and ratio scales. A point that ISPOR, in presenting practice guideless, has overlooked³³.

THE ROAD NOT TAKEN

A reasonable question is why the leaders in health technology assessment rejected the scientific method in favor of inventing evidence, convincing thousands that this was the way forward. The excuse provided by the leaders was that we needed to construct imaginary simulations and invent evidence because, at product launch good quality evidence is insufficient²³. We

can even attempt to resolve uncertainty with value of information analysis applied to constructed imaginary scenarios. A more practical approach is to balance uncertain early benefits against longer term more real world evidence; not to create imaginary cost-effectiveness claims, but for formulary committees to require manufacturers to underwrite protocols submitted for prospective claims assessment. This was proposed almost 20 years ago and has been maintained in the latest version 3 of the Minnesota proposed formulary guidelines³⁴. Pricing and access requirements would be provisional, subject to claims assessments being reported to the committee. This would establish required measurement standards in the process of discovery with, if required, providing support for value contracting.

Unfortunately, convincing agencies such as NICE to reject reference case invention of approximate imaginary information has proved counter-productive. It is all too easy to construct a model; engineer the model to support the manufacturer's preferred price and convince a committee that this is the last word in technology assessment. This stifles any further need to discover new facts; non-evaluable assertions will suffice. The product is on formulary and the manufacturer only has to consider its ongoing marketing and sales position *vis á vis* competing products and new entrants. This may involve further modelling, but the incentive is not necessarily there; ICER does not revisit regularly its previous modelled claims to accommodate new evidence from the literature and revised assumptions.

A case can, of course, be made that it is not in ICER's interest even with its new cloud based platform, *ICERAnalytics* where the ICER model structure and its assumption can be changed to create entirely new assertions regarding pricing and access. For those interested to do so and with time on their hands on a wet Sunday afternoon, the opportunity is there to create a multitude of competing models and non-evaluable claims; an assertive paradise of competing non-evaluable claims³⁵.

PROBABILISTIC SENSITIVITY ANALYSIS

To those in the badger belief system, the ultimate model outcome is PSA. This technique, proposed in the late 1990s, is to generate simulated likelihood assertions for comparative cost-effectiveness from invented evidence. A comparison of products, with the application of cost-per-QALY thresholds, produced a series of assertions, associated with contrived probabilities that various products at different price levels were cost-effective. Formulary committees were asked to embrace these assertions in their formulary assessment deliberations. This, from a realist perspective, is complete nonsense. There is no basis and never would be for these likelihood claims (and attendant probabilities) ever to be empirically evaluated. Different assumptions and assertions could produce (and in some cases were designed to produce) alternative probabilistic claims. A multitude of models could produce a multitude of non-evaluable PSA claims. As more data to support changing

assumptions became accessible a new deluge of likelihood claims could be envisaged. These would be for decades into the future, a veritable Disneyland of claims or, to be precise, assumption driven assertions for those prepared to put the standards of normal science to one side.

Not surprisingly, within the badger system, belief in the imaginary PSA information content as a driver of formulary decisions is widely held; truth, again, is consensus. ICER is a devoted advocate. The logical position that the simulations driving PSA are built on assumptions and assertions projecting into the future that, for a realist, lack any credibility, is not part of the discussion. This apparently is approximate imaginary information (or assertions) that are equivalent to that created by randomized clinical trials where the protocol is designed to capture comparative cost-effectiveness claims.

Formulary committees are in a dilemma: do they accept non-evaluable probabilistic or likelihood assertions of cost-effectiveness at different pricing levels as valid decision inputs or should they be thrown out. To the realist the answer is obvious. The gap between speculation and established knowledge, in Lakatos' terms, is wide. Admittedly, the PSA diagrams are pretty as candidate for a realist wallpaper collection; the problem is that they are meaningless however many Monte Carlo simulations are presented to populate a cloud diagram.

REPUTATIONAL CHALLENGES

There is always satisfaction for the non-believer in deconstructing a belief system and relegating it to the nearest metaphysical dumpster. In technology assessment this is straightforward and long overdue. Indeed, there is the feeling that the need to protect the belief system against charges that were not even considered in its infancy is not a reasoned defense but an *ex post facto* contrived defense; a defense to protect reputations and the inordinate amount of money spent over the last 30 years to construct imaginary worlds. The importance of the network of believers should not be underestimated: these include academic and research centers, PSA missionaries as well as the multitude of believers presently employed by pharmaceutical manufacturers. All have a lot to lose. After all, admitting to generations of graduate students that the emperor has no clothes is an unanticipated recantation, with obvious religious analogies with worthless indulgencies and relics.

If the true believer, an unreformed relativist, holds firm then the debate ends there. No logic or evidence will ever convince the believer that he/she may be wrong. They will drink the cool aid. Their belief system, even if an observer thinks it's weird, is immune to any criticism. The only salvation, if that is the correct word, is for third parties, such as agencies evaluating pharmaceuticals, to reject submissions that treat ordinal preferences as though they were ratio scales, rejecting

constructs such as the impossible or I-QALY, thus allowing the badger belief system to wither away to an empty cete.

Over the last 30 years, millions of dollars have been spent, by both public and private sectors, on constructing imaginary simulations to support marketing claims for products. Tens of thousands of papers have been published in leading journals (*Value in Health, Pharmacoeconomics, Journal of Medical Economics*). ISPOR and ICER have been in the forefront of these endeavors. The media and all too many patient associations have given their support. By and large, these investments in imaginary claims have paid off. Who would not invest in a product that can never be shown to be wrong in the claims made? To the cynic: we can always prove a product is cost-effective (without having to reduce price!). Paradoxically, manufacturers may continue to support the meme on these grounds. The outcomes from a contribution to a church collection are far less certain; in this badger case the outcomes are predetermined. John Calvin would no doubt approve.

CONCLUSIONS: THE MADNESS OF CROWDS

Can a belief system have a use by date? Even the anticipation of the Parousia? The feature that stands out in terms of both ISPOR and ICER is their denial (or refusal to even consider) the limitations imposed by the axioms of fundamental measurement. Certainly, the belief system has proved to be inordinately popular and incredibly lucrative, but the end is uncertain. This is not from the many reminders of the limits of measurement, which have been a constant over 30 years, but the willingness of third parties to support the ISPOR and ICER packaged responses. Put simply, if the I-QALY fails so does the belief system. This is no small matter; it is more than an academic ripple in a decaf coffee cup or a Senior Common Room spat. It is a failure that impacts one of the major sectors of the US economy.

To an important degree, the continued popularity of the relativistic imaginary simulation has been the result of national agencies such as NICE buying into this relativistic belief system. The NICE reference case has been emulated globally, in many cases with a NICE police force equivalent of academics, like an inquisition, giving manufacturer's submissions the good housekeeping seal of imaginary approval. If it makes life easy, although nonsensical, then manufacturers will fall into line. Hiring a consultant to create an imaginary simulation to reference case standards is an easy and low cost option. A *danse macabre*: where both parties want a quick and low cost

resolution in pricing and access recommendations. Neither party is willing to expose the underbelly of the badger and the failure to meet measurement standards; both seek a comfortable relationship that hits the checkboxes and gains formulary approval. The claims are never revisited; a rite of passage has been observed and pricing negotiations, including rebates, can begin. Welcome to the real world.

It is not the intention here to force believers to abandon a commitment to imaginary simulations and the rejection of fundamental measurement. That would be the moral equivalent of convincing a class of 5-year olds in mid-December that Father Christmas is a myth. Why shatter such innocent belief? Yet, while innocent it may be of its rejection of normal science and fundamental measurement, the fact is that it is fatally flawed. If those promoting approximate imaginary information to support formulary decisions are to maintain a steadfast belief, then (as with the Salvation Army) they need to be prepared to stand behind those beliefs; tambourines and those irritating Christmas bells. Just asserting belief, a deeply held faith in the badger technology assessment meme, taking refuge in relativism, is unlikely to convince realist critics.

In an important sense, ISPOR and ICER are between a rock and a hard place. On the one hand, they embrace relativism, maintaining that truth is consensus, and defend in the ISPOR practice monographs, the construction of approximate imaginary information and the denial of the standards of normal science. On the other hand, their ongoing neglect (or refusal to consider) the limitations imposed by the axioms of fundamental measurement undercuts their belief system; all preference scores are ordinal and the I-QALY is an impossible construct; let alone their embrace of logical positivism. As noted, at no time has ISPOR considered a practice guideline for measurement. It lurks in the background; no effort to suppress these criticisms will succeed. This implies that the end of shelf life will be at least entertaining. Will rhetoric, persuasion and authority trump the realism of the scientific method? Will common sense prevail? Will third parties reject ICER and the QALY? Or will we have a post-Reformation parallel belief world? We may have two groups of advocates of health technology assessment: dogmatic followers of the relativist badger ISPOR/ICER meme and more practical realists looking to confront claims with real world evidence and the discovery of provisional new facts for therapy impact. Hopefully, poker will not be involved.

CONFLICTS OF INTEREST

PCL is an Advisory Board member and consultant to the Patient Access and Affordability Project, a program of Patients Rising

REFERENCES

- ¹ Goldstone L&N. *Out of the Flames*. London: Century, 2003
- ² Neumann PJ, Willke R, Garrison LP. A Health Economics Approach to US Value Assessment Frameworks – Introduction: An ISPOR Special Task Force Report. *Value Health*. 2018;21:119-123
- ³ Potter B. *The Tale of Mr Tod*. London: Frederick Warne and Co, 1912
- ⁴ McKenna S, Heaney A. Composite outcome measurement in clinical research: the triumph of illusion over reality. *J Med Econ*. 2020; 23(10):1196-1204
- ⁵ McKenna, Heaney A, Langley P. Fundamental Outcome Measurement: Selecting Patient Reported Outcome Instruments and Interpreting the Data they Produce. *InovPharm*. 2021;12(2):No.17
<https://pubs.lib.umn.edu/index.php/innovations/article/view/3911/2764>
- ⁶ Langley P. The Great I-QALY Disaster. *InovPharm*. 2020; 11(3): No 7
<https://pubs.lib.umn.edu/index.php/innovations/article/view/3359/2517>
- ⁷ Thurstone L. A law of comparative judgment. *Psychological Review*, 1927;34(4): 273–286
- ⁸ Stevens S. On the theory of scales of measurement. *Science*. 1946;103:677-680
- ⁹ Institute for Clinical and Economic Review. *Lupus Nephritis*. ICER-lupus nephritis-comments-folio_031221.pdf
- ¹⁰ Wootton D. *The Invention of Science: A new history of the scientific revolution*. New York: Harper Collins, 2015
- ¹¹ Wittgenstein L. *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul, 1922
- ¹² Wittgenstein L. *On Certainty*. London: Basil Blackwell, 1969
- ¹³ MacDonald JF. Russell, Wittgenstein and the problem of the rhinoceros. *Southern J Philosophy*. 1993;31(4):409-24
- ¹⁴ Orwell G. 1984. London: Secker & Warburg, 1949
- ¹⁵ Briggs R. *The Scientific Revolution of the Seventeenth Century*. London: Fontana, 1969
- ¹⁶ Gopnik A. Finding our inner scientist. *Daedalus*. Winter 2004; 21-28
- ¹⁷ Pigliucci M. *Nonsense on Stilts: How to tell science from bunk*. Chicago: University of Chicago Press, 2010
- ¹⁸ Lakatos I. Falsification and the methodology of scientific research programmes. In *Criticism and the Growth of Knowledge*. I Lakatos, A Musgrave. Eds. London: Cambridge University Press, 1970
- ¹⁹ Edmonds D, Eidinow J. *Wittgenstein's Poker: The Story of a Ten-Minute Argument between Two Great Philosophers*. New York: Harper Collins, 2001
- ²⁰ Kuhn T. *The Structure of Scientific Revolutions*. 2nd Ed. Chicago: University of Chicago Press, 1970

- ²¹ 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
- ²² Langley P, McKenna S. Fundamental Measurement and Quality Adjusted Life Years. *Value Health*. 2021;24(3):461
- ²³ Magee B. Popper. London: Fontana, 1974
- ²⁴ Rodych V. Popper versus Wittgenstein on Truth, Necessity and Scientific Hypotheses. *J General Philosophy Science*. 2003;34:323-336
- ²⁵ McKenna S, Wilburn J. Patient value: its nature, measurement, and role in real world evidence studies and outcomes-based reimbursement. *J Med Econ*. 2018;21(5): 474-80
- ²⁶ Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. New York: Routledge, 2015
- ²⁷ Lugné AK, Krabbe PF. An overview of the time trade-off method: concept, foundation, and the evaluation of distorting factors in putting a value on health. *Exp Rev Pharmacoecon Outcomes Res*, 2020;20(4): 331-42
- ²⁸ Langley P, McKenna S. Fundamental Measurement: The Need Fulfillment Quality of Life (N-QOL) measure. *InovPharm*. 2021;12(2): No. 6 <https://pubs.lib.umn.edu/index.php/innovations/article/view/3798/2697>
- ²⁹ Patrick D, Starks H, Cain K et al. Measuring preferences for health states worse than death. *Med Decis Making*. 1994;14:9-18
- ³⁰ Gandhi M, Rand K, Luo N. Valuation of health states considered to be worse than death – An analysis of composite time trade-off data from 5 EQ-5D-5L valuation studies. *Value Health*. 2019;22:37-76
- ³¹ Devlin N, Brazier J, Pickard A et al. 3L, 5L. What the L? A NICE Conundrum. *Pharmacoeconomics*. 2018;36:637-40
- ³² Wailoo A, Hernández M, Pudney S et al. An international comparison of EQ-5D-3L and EQ-5D-5L, for use in cost-effectiveness analysis. *Value Health*. 2021;24(4):568-74
- ³³ Wailoo AJ, Hernandez-Alava M, Manca A, et al. Mapping to estimate health-state utility from non-preference-based outcome measures: an ISPOR Good Practices for Outcomes Research Task Force Report. *Value Health*. 2017; 20(1):18-27.
- ³⁴ Langley P. Value Assessment, Real World Evidence and Fundamental Measurement: Version 3.0 of the Minnesota Formulary Submission Guidelines. *InovPharm*. 2020;11(4):No. 12 <https://pubs.lib.umn.edu/index.php/innovations/article/view/3542/2613>
- ³⁵ Langley P. Let a Thousand Models Bloom: ICER Analytics Opens the Floodgates to Cloud Pseudoscience. *InovPharm*. 2021;12(1): No. 5 <https://pubs.lib.umn.edu/index.php/innovations/article/view/3606/2668>