THE STANDARDS OF NORMAL SCIENCE: PSEUDOSCIENCE IN HEALTH TECHNOLOGY ASSESSMENT

Recent Current Issue Reports have made the case that the methods espoused by groups such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), the Academy of Managed Care Pharmacy (AMCP), and the Institute for Clinical and Economic Review (ICER) to support modeled claims for cost-effectiveness fail the standards of normal science. This follows from their focus on the construction of imaginary lifetime simulations where value claims are driven by imaginary, incremental cost-per-QALY calculations set against arbitrary willingness-to-pay thresholds. This approach to technology assessment should be rejected. It adds nothing to our discovery of new facts because it fails the demarcation test between science and pseudoscience. Here's why.

The Pseudoscience Position

ISPOR reaffirmed its commitment to the construction of imaginary worlds in a 2018 Special Task Force Report.[1] This task force was convened to "review relevant perspectives and appropriate approaches and methods to support the definition and use of high-quality value frameworks"; presumably "high value" imaginary constructs, although no criteria are suggested that may distinguish "high value" from "low value" imaginary worlds, or how a metric might be applied to inform those who are willing to utilize imaginary worlds in formulary decisions. [2]

Expressing a concern that in attempting to simplify the problem of value, competing value frameworks could end up making *ad hoc* assumptions and simplifications not supported by theory or evidence, the Report takes the position that frameworks that do not capture the full costs and benefits of treatment may distort decisions, although "full" is not defined outside of broad reference case guidelines. ISPOR's approach is "to define value on the basis of microeconomic principles, recognizing that value is best defined as what individuals (or others acting on their behalf) would be willing to pay to acquire more health care or other goods and services." This is best achieved, the Report maintains, by informing resource allocation decisions by "approximating" the value of interventions in incremental cost per quality adjusted life year (QALY) term, capturing length and quality of life in an imaginary reference case construct.

This approximate value of interventions is achieved through the construction of lifetime imaginary worlds. A hypothetical population is tracked, on the basis of a series of assumptions detailing the structure of the simulation, the choice of a therapy and a comparator, the assumed transition of the target patient group through the various stages of disease over their lifetimes as determined by selected clinical response parameters, and the time spent in each disease state. Each stage of the disease is

assigned a utility score extracted from the literature, and the result is the creation of average imaginary lifetime QALYs for the drug and its comparator. Incremental lifetime costs per QALY are estimated and matched to baseline "willingness to pay" thresholds. Thus is the approximate – yet imaginary – "value" of a new product assessed compared, for example, to the standard of care. However, ICER takes these comparisons and uses them to make recommendations for price discounting and access. The imaginary and arbitrary construct is not intended to provide credible, evaluable, and replicable claims – the construct is incapable of providing such information. Yet ICER claims it is the "state of the art" in health technology assessment. Judged against the standards of normal science, the ICER construct qualifies as nothing more than pseudoscience.

The Standards of Normal Science

The requirement for testable hypotheses in the evaluation and provisional acceptance of claims made for products and devices is so fundamental as to be unexceptional. For hundreds of years, scientists and other thinkers have accepted as a baseline requirement that if a research agenda is to advance – if there is to be an accumulation of knowledge – there must be a process of discovering new facts.

Indeed, as early as the 16th century, Leonardo da Vinci (1452 – 1519), in notes that appeared posthumously for his *Treatise on Painting* (published in 1641), clearly anticipated the standards for the scientific method, which were widely embraced a century later in rejecting thought experiments that fail the test of experience.

By the 1660s, the scientific method, following the seminal contributions of Bacon, Galileo, Huygens, and Boyle, had been clearly articulated by associations such as the Academia del Cimento in Florence (1657) and the Royal Society in England (founded 1660; Royal Charter 1662) with their respective mottos *Provando e Riprovando* – "prove and again prove," and *nullius in verba* – "take no man's word for it."[3]

By the early 20th century, standards for empirical assessment were put on a sound methodological basis by Sir Karl Popper (1902-1994) in his advocacy of a process of 'conjecture and refutation.[4] [5] Hypotheses or claims must be capable of falsification; indeed, they should be framed in such a way that makes falsification likely. Falsification of our hypotheses forces us to reconsider our models and the assumptions built into those models. This leads to the obvious point that claims or models should not be judged on the realism or reasonableness of assumptions, or on whether the model represents – for a public advocacy research group such as ICER – their belief in lifetime comparative cost-per-QALY outcomes in a future hypothetical reality.

Although Popper's view on what demarcates science from pseudoscience is now seen as an oversimplification involving more than just the criteria of falsification, the demarcation problem remains.[6] Certainly, there are different ways of doing science, but what all scientific inquiry has in common is the "construction of empirically verifiable theories and hypotheses." **Empirical testability is the one major characteristic distinguishing science from pseudoscience; theories must be tested against data.** The development of pharmaceutical products and the evidence standards required by the Food and Drug Administration (FDA) for product evaluation and marketing approval is driven by stringent adherence to the scientific method. However, paradoxically, once a product is launched, to support its claims about cost-effectiveness, ICER's modeled pricing and access recommendations abandon the scientific method.

The rejection of a research program that meets the standards of normal science by groups such as ICER is best exemplified by the latest version

of the Canadian health technology guidelines, which state, in part: "Economic evaluations are designed to inform decisions. As such, they are distinct from conventional research activities, which are designed to test hypotheses." [7] While this position puts modeled health technology assessment in the category of pseudoscience, it is also what may be described as a relativist position.

Rather than subscribing to the position that the standards of normal science are the only standards to apply in health care decisions and value claims, the relativist believes that all perspectives are equally valid. Health care decisions are to be understood sociologically. No one body of evidence is superior to another. Results of a lifetime modeled simulation are on an equal basis with those of a pivotal Phase 3 randomized clinical trial. For the relativist, the success of a scientific research program – in this case one built on hypothetical models and simulations – rests not on its ability to generate new knowledge but on its ability to mobilize the support of the community. Basing decisions on models and simulations underpins the consensus view that evidence is **constructed**, never **discovered**. This camp would have you believe that instead of discovery, science is about rhetoric, persuasion and authority.[11] Truth is consensus.

The Health Technology Assessment Meme

If truth is consensus, how is this consensus, resting upon the construction of imaginary worlds, maintained? In this case, the "consensus" consists of over 30 years' worth of cost-effectiveness modeled claims. The ISPOR consensus, embraced by ICER, on health technology assessment can be usefully characterized as a meme: a unit of cultural transmission or a replicator, or as Dawkins' suggests, a mind virus.[8] Applying this term to the imaginary worlds of health technology assessment is deliberate. It underpins the interpretation of ISPOR and ICER's continued, unqualified acceptance of the reference case as its core business model, as a sociological phenomenon. The ICER reference case that constructs evidence to support pricing and affordability decrees can be characterized as an analog of gene pool propagation "by leaping from body to body via sperm or eggs."[8]

Human beings are good at imitation. The reference case meme appears to be adept in its infectivity, supported by an organizational infrastructure to defend it against competing views, ensuring survival through supporting propagation, longevity, acceptability, and copying fidelity. The widespread adoption and propagation of this meme in seen with literally thousands of imaginary world technology assessments published. ISPOR's continuing enthusiastic efforts at transmission and infection ensure a high copying fidelity with few willing to reject the "faith and mysteries" of their parents' belief system.[9]

Driven by Assumption

Knowledge is provisional and permanently so; we cannot, at any stage, prove that what we "know" is true. Attempting to believe or justify our belief in a theory is logically impossible. However, by empirical assessment, we can try and demonstrate our preference for one theory over another and apply it to the best of our knowledge.

Constructing imaginary worlds that were never intended to generate potentially falsifiable claims cannot, therefore, be defended by an appeal to the "truth" of their assumptions. If a health technology assessment claim is built upon a series of assumptions, it's reasonable to ask: what is the status of the various assumptions? Are they to be viewed as "reasonable" or "realistic" metrics for an unknown future reality? Have they been selected from the literature because they seem appropriate? Are they the "best available" from limited data?

More to the point, there is a belief that the fact that the selected assumptions are based, where feasible, on an empirical study validates the choice of assumption. For example, if the model is intended to incorporate utilities that have been reported in one or two studies for progression and time spent in the stages of a disease over a hypothetical future lifetime, then there is an immediate methodological issue. To claim that an assumption is valid is to revisit David Hume's (1711-1776) induction problem: an appeal to facts to support a scientific statement. Unfortunately, as Hume pointed out, no number of singular observations can logically entail an unrestricted general statement. Certainly, there may be comfort in reporting that the claim that all swans are white has not been contradicted yet (until that Qantas vacation in Western Australia), so one fully expects the next swan to be white. But as Hume pointed out, this is a fact of psychology and does not entail any general statement. From a utility perspective, the fact that one hundred papers have agreed (within limited bounds) on generic utilities from the same instrument for a target population in a disease state stage is immaterial. We cannot secure this assumption: it cannot be "established by logical argument, since from the fact that all past futures have resembled past pasts, it does not follow that all future futures will resemble future pasts." [10]

Despite ICER's continued embrace, logical positivism is dead. It died some 80 years ago. All knowledge is provisional. Popper's contribution was to make clear that Hume's problem with induction can be resolved. We cannot prove the truth of a theory, or justify our belief in a theory or attendant assumptions, since this is to attempt the logically impossible. **We can only justify our preference for a theory by continued evaluation and replication of claims.** Constructing imaginary worlds, even if the justification is that they are "for information" is, to use that memorable phrase of Jeremy Bentham's (1748-1832), "nonsense on stilts." If there is a belief, as the one subscribed to by ICER, in the sure and certain hope of the relevance of approximate information created by imaginary worlds, a belief to drive formulary and pricing decisions, then it needs to be made clear that this is a belief that lacks scientific merit.

Certainly assumptions can be a critical element of models; the difference is these models should support testable hypotheses. This is echoed by Isaac Newton (1642-1727), with René Descartes (1596-1650) as his target in saying, "I do not feign hypotheses." Descartes, in Newton's view, had "produced fantastic and untestable ideas, then assumed them to be true and used them as building blocks of his philosophy." [11]

Neglecting Measurement

A final point is the neglect in the reference imaginary guidelines of the need for precision in measurement. Surely, in the social sciences, as it is true of economics, the intent should be that the measurement standards should match those of the mainstream sciences. This is not the case in health technology assessment. As detailed in a previous IPAA Current Issues Report, the focus on constructing QALYs imposes the need to utilize ordinal utility scores to support lifetime cost-per-QALY estimates. Although these are only imaginary constructs, the manipulation of raw scores from ordinal scales is invalid. The claims have no substance.

The Threshold Confusion

Media attention on the ICER recommendations is typically focused on the pricing and access recommendations driven by value claims based on willingness-to-pay thresholds. Acceptance of these claims reflects a willingness to accept these claims at face value rather than a more nuanced assessment, which recognizes that: (i) the recommendations are unique to the model and its assumptions, meaning a different utility score would generate different recommendations, and (ii) that there is a potential multiverse of competing imaginary worlds that all subscribe to the reference case. Again, we face the belief in

"approximate information" when the term itself is meaningless. Again, in a world driven by assumption, any series of recommendations could be produced, adding to the confusion by decision makers over which "approximate information" source to believe (if belief should ever enter into it).

A Failure on All Counts

It is unusual to find a willing acceptance of an analytical method, in what purports to be a relevant social science, that fails the standards of normal science. Perhaps the answer is that, if cost-effectiveness claims were to be made to support formulary decisions, then the evidence base was limited for the target patient populations. There was no option but to build imaginary worlds, carefully circumscribed by the mists of sensitivity analysis, to convince the recipients that the "approximate information" had an undefined value in decision making.

If so, then we should be considering whether to point out the emperor has no clothes – admit that formulary decisions for new therapies are inevitably based on limited information. Rather than make up the evidence, we might consider a more conservative approach. One that includes a patient-focused needs assessment and quality of life measurement, and a research program to track response to therapy. We should discover new information, continue evaluating it, and leave room to adjust our claims accordingly.

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

---

[1] Neumann P, Willke R, Garrison L. A health economics approach to US value assessment frameworks – Introduction: An ISPOR Special Task Force Report [1] *Value Health*. 2018;21:119-123
[2] Langley PC. ICER, ISPOR and QALYs: Tales of Imaginary Worlds. 2019;10(4);No. 10 https://pubs.lib.umn.edu/index.php/innovations/article/view/2266
[3] Wootton D. The Invention of Science: A new history of the scientific revolution. New York: Harper Collins, 2015.
[4] Popper KR., The logic of scientific discovery .New York: Harper, 1959.
[5] Lakatos I, Musgrave A (eds.). Criticism and the growth of knowledge. Cambridge: University Press, 1970.
[6] Piglucci M. Nonsense on Stilts: How to tell science from bunk. Chicago: University of Chicago Press, 2010)
[7] Canadian Agency for Drugs and Technologies in Health (CADTH). Guidelines for the economic evaluation of health technologies: Canada. Ottawa: CADTH, 2017
[8] Dawkins R. The Selfish Gene (30th Anniversary Ed). Oxford: University Press, 2006
[9] Dawkins R. A Devil's Chaplain. New York Houghton Miflin, 2004
[10] Magee B. Popper. London; Fontana, 1973
[11] Briggs R. The Scientific Revolution of the seventeenth century. Longman, 1971.