MAIMON WORKING PAPER NO. 5 APRIL 2021

A COMMITMENT TO PSEUDOSCIENCE: DO WE NEED PROBABILISTIC SENSITIVITY ANALYSIS?

ABSTRACT

Probabilistic sensitivity analysis is statement of our level of ignorance in the commitment to approximate imaginary information in health technology assessment. In response to limited information to support cost-effectiveness claims at product launch, the leaders or the illuminati in health technology assessment agreed to reject hypothesis testing in favor creating approximate imaginary information. This rejection of the standards of normal science created a situation where claims for competing pharmaceutical products and devices were couched in terms of incremental cost-per-QALY models with the assumptions driving the imaginary simulation protected by sensitivity scenarios to defend the choice of assumption. Probabilistic sensitivity analysis is a formalization of this to create imaginary claims for the likelihood of product pricing being cost-effective. This framework gave, to those unfamiliar with the standards of normal science and the pre-eminent role of hypothesis testing to discover new, yet provisional facts, an aura of scientific respectability. Looking forward some 30 odd years or the timeframe of the imaginary simulation model driving the probabilistic claims, proponents could claim superficial rigor in their analysis, convincing the less knowledgeable in their audience, and make claims for the role of invented evidence. This is, of course, just pseudoscience; sharing the Dover courtroom platform with intelligent design. It rejects 400 years of science and the fundamental point: what demarcates science from gibberish is the ability to formulate credible claims, empirically evaluate and replicate them in different treatment settings. No, we don't need probabilistic sensitivity analysis. The claim for resolving uncertainty is no excuse to abandon the standards of normal science.

INTRODUCTION

It has been noted in previous Maimon Working Papers that health technology assessment has the dubious distinction, alone of the social sciences, of promoting the invention of approximate imaginary information to support claims for competing pharmaceutical products and devices [1]. In the early 1990s, the leaders or illuminati in the field of health technology assessment rejected claims assessment through hypothesis testing in favor of claims creation through the invention of evidence. As previously noted, this flies in the face of some 400 years of advances in the physical sciences where the dominant paradigm has been one of the discovery of provisional new facts. For those who subscribe to the standards of normal science, the demarcation between science and pseudoscience, or more properly, gibberish, are claims that are credible, empirically evaluable and replicable [2]. Indeed, there is always uncertainty as to product performance, possibly increased with limited information, at market entry but

that is no reason to invent evidence on assumptions that are patently false and an analytical framework that reject the standards of normal science [3].

Health technology assessment, judged by the standards of normal science, is an abject failure. The meme or belief system, which is well entrenched after some 30 years, accepts without question the invention of competing claims for cost-effectiveness. This relies in part on the ignorance of the standards of normal science, most notably measurement theory, and the needs to accelerate the process of claiming cost-effectiveness. The most egregious mistake is probably the impossible quality adjusted life year or I-QALY where the construct relies on multiplying a ratio scale (time) with an ordinal scale (utility) [4] . Rather than, as those subscribing to the standards of normal science would advocate, arguing for an evidence base to evaluate claims though hypothesis testing, it is easier (and more lucrative from a consulting perspective) to invent evidence to support claims [5]. Initially, concerns about the merits of hypothesis testing were a concern. These were, apparently quite easily pushed aside, in favor of a relativist position that inventing model simulations was as valid as hypothesis testing. We now have a belief system, shared by thousands, that truth is consensus and that modeled claims, although pseudoscience, are the way forward.

SENSITIVITY ANALYSIS

It occurred early on in the advocacy and development of simulated lifetime cost-effectiveness claims that one needed a 'protective belt' to protect modeled assumptions: assumptions which came from clinical trials, the so-called literature and, all too frequently, pure guesswork. This last point could be seen in the process of 'simplifying' assumptions for model structures and the rules driving the progression of the hypothetical target population through natural disease stages. No thought was ever given to formulating empirically evaluable claims. When these imaginary model constructs were 'validated' this was typically against other equally imaginary models. This, of course, led to an obvious criticism: why choose one model out of a potential multitude of competing models [6].

Sensitivity analysis involves, quite simply in both deterministic and probabilistic forms, attaching parameter boundaries to selected assumptions and the re-running the simulation to see how much claims were impacted. As these sensitivity claims could never be empirically evaluated, it seems a waste of time but the audience bough into this exercise as demonstrating the appropriateness of the chosen meme or belief system, with its attendant mysteries.

PROBABILISTIC SENSITIVITY ANALYSIS

Any number of studies presented one and two way sensitivity assessments of modeled claims in the early 1990's. None were that convincing apart from demonstrating the subjective 'flexibility' of claims where a model to support a particular claim could easily be constructed, supported by a judicious selection of parameters to 'sensitize'. But the icing on the cake was the promotion of probabilistic sensitivity analysis to characterize parameter uncertainty. A technique with an apparent technical plausibility that obscured the dubious assumptions and neglect of the standards of normal science on

which the modeling for cost-effectiveness claims was based. One unfortunate consequence was the decision by several agencies proposing guidelines to adopt this framework.

The analysis proceeds from the choice of decision framework or model structure, and assigns distributions to the various model parameters to capture, or at least characterize uncertainty, from evidence available. These distributions are sampled, using Monte Carlo simulations, with each sample set producing estimates of expected costs and benefits. This continues (say 10,000 times) to produce a 'likely' (by assumption) range of values of the parameters and an estimate, which can never be empirically evaluated, which is considered the 'correct' estimate of imaginary costs, outputs and net benefits. One obvious problem is how to justify choice of a parameter distribution if, as in the case of utility scores, the scale is ordinal. The nature of ordinal scales is that a parameter distribution is impossible as the distance between 'numbers' is unknown with only non-parametric techniques allowed. For example, in the case of utilities, ignoring negative values, the estimation is based on means and standard errors with the gamma or lognormal distribution prime candidates. This is incorrect; the utility scale cannot support these. Or, to make the more general point, all parameter inputs must be from a ratio scale or products of ratio scales.

Putting aside minor issues of fundamental measurement, the end-product is a presentation to represent decision uncertainty or how uncertain is the claim being made. With estimates of lifetime (usually) expected cost, expected befit and expected net benefit, given choice of particular cost-effectiveness thresholds, a decision can be proposed. If benefits are expressed as QALYS we immediately hit a brick wall as the QALY is a mathematically impossible construct; a distribution of QALYs is just mathematical nonsense. Which leaves us with the imaginary cost-effectiveness acceptability curve analysis where the probability of being cost-effectiveness is matched to various cost-effectiveness thresholds for each product. The probability that each is cost-effective is the proportion of times it has the highest net benefit. Again, to emphasize, if the QALY is the measure of effectiveness, then we have just nonsense.

NONSENSE ON STILTS

Previous evaluations of modeled claims, notably, in the US, those created by the Institute for Clinical and Economic Review (ICER) have pointed to the manifest deficiencies of the ICER invention of evidence to support cost-effectiveness claims. ICER is not alone; they join government supported agencies such as the National Institute for Health and Care Excellence (NICE) in the UK, as well as the practice guidelines for creating imaginary simulations issued by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) in their promotion of created or invented evidence to support formulary decisions. ICER cannot, in fact, defend its reference case methodology. It has a naïve belief that utility scales have hidden ratio properties, one of the great technology assessment mysteries, and that the QALY is a mathematically possible construct. This is nonsense; a level of nonsense that joins the equally naïve belief that evidence for formulary decisions can be created from reference case simulation modeling. These various probabilities are, of course, non-evaluable (apart from the fact that they extend over the lifetime of hypothetical patients) plus the unfortunate fact that as effectiveness is typically simulated from QALYs, he entire exercise is a manifest waste of time.

While there is no doubting the popularity of the cost-effective acceptability framework, the fact is that it adds nothing to supporting imaginary claims for cost-effectiveness. Certainly, uncertainty is present in evaluating the competing benefits and costs of medications, but this can be more easily and believably accommodated through adhering to the standards of normal science. There is no doubt, of course, that belief will linger as the impossibility of the 'mystery' of the impossible QALY and the visual appeal of probabilistic sensitivity analysis with its scatter diagrams and net benefit threshold diagrams will ensure the mystery endures. In the relativistic world of technology assessment truth is consensus [7].

EXHIBIT NO. 1: WELCOME TO FAIRYLAND

To illustrate the resilience of the technology assessment meme and probabilistic sensitivity analysis, we need go no further than the well regarded, and now in its 4th edition, Drummond et all textbook for constructing imaginary claims that fail the standards of normal science [8]. This is a deficiency primer. Exhibit 1 is characterized by its failure to address a number of issues:

- The absence of any statement as to the standards of normal science, credible claims, empirical evaluation and replication, and why these have been rejected in favor of creating approximate imaginary information with lifetime assumption driven simulation models
- A misunderstanding of the axioms of fundamental measurement which are quite clear on the requirement that if a proposed metric is to be mathematically acceptable then the elements comprising that metric (e.g., the QALY) must each have ratio properties
- Presenting direct and indirect utility instruments but with no consideration of the implications of these instruments creating negative scores (i.e., states worse than death) together with the lack of construct validity, unidimensionality and dimensional homogeneity by adding together symptoms or attributes to a single score
- Failing to acknowledge that he scores produced by the direct and indirect utility instruments are ordinal scores and the consequent mathematical impossibility of creating ratio scales but proceeding to assume they have ratio properties both in supporting (the direct instruments) the multiattribute indirect utility algorithms and the mathematically impossible QALY
- Insisting that the only way forward in technology assessment is to invent claims for competing products through lifetime cost-per-I-QALY simulation following the reference case methodology of agencies such as NICE when other agencies and formulary committees might object to making decisions from imaginary claims
- Failing to acknowledge that probabilistic sensitivity analysis fails the demarcation test between science and non-science
- The absence of any consideration of options that recognizing the importance of the standards of normal science (e.g., a product evidence base to support claims evaluation)
- The possibility of provisional pricing and value contracting to support claims assessment to support access to new technologies on a provisional basis with recourse to inventing evidence to drive early approval

The defense offered by Drummond et al for the construction of imaginary claims to support early approval is that in practical terms restricting the accumulation of new evidence to research is the only feasible option because the possibility of otherwise accumulating new data will no longer be possible as a consequence of approval. The trade-off is, therefore, between inventing evidence through modeling over hypothetical lifetimes versus establishing a framework for future claims assessment. The former is favored because the practicality of the latter, given lack of interest once on formulary, is uncertain. A fairy story trumps real world evidence; unless, of course, formulary committees demand real world evidence and are prepared only to give provisional approval. The fact that NICE is committed to invented claims, with a legion of academic advisors to give a good housekeeping seal of approval to fairytale modeled claims, does not mean that anyone else has to follow this lead.

CONCLUSIONS

In this uncertain world it is difficult to gage how much longer the charade of invented claims and 'believe me' probabilistic sensitivity analysis will continue. It is riddled with errors and assumptions that are patently wrong; notably from an apparent disregard of measurement theory. It is possibly one thing to invent claims but another to invent claims that are easily shot down. In the US, at least, the concern is one of litigation. Challenges from manufacturers, physicians and patient groups that formulary decisions are based on nonsense. Indeed this is already happening with agencies refusing to accept QALYs in formulary decisions. While this rejection is on the grounds of discrimination and not the more abstruse consideration of measurement theory, there is now a willing audience to take groups such as ICER to task for peddling claims that have no basis in reality, only approximate imaginary information.

REFERENCES

[1] Langley P. The herd and the meme: Why ordinal scales have to be ratio measures in disguise. Maimon Working Paper No. 3 April 2021 www.maimonresearch.net

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

[3] Langley PC and McKenna SP. Measurement, modeling and QALYs.*F1000Research.* 2020; 9: 1048 https://doi.org/10.12688/f1000research.25039.1

[4]Langley P. Nonsense on Stilts - Part 1: The ICER 2020-2023 Value Assessment Framework for Constructing Imaginary Worlds. *InovPharm*.2020;11(1): No.12 https://pubs.lib.umn.edu/index.php/innovations/article/view/2444

[5]Langley P. The Great I-QALY Disaster. *Inov Pharm*. 2020;11(3): No 7 https://pubs.lib.umn.edu/index.php/innovations/article/view/3359/2517

[6] Langley P. Let a Thousand Models Bloom: ICER Analytics Opens the Floodgates to Cloud Pseudoscience. InovPharm.2021;2(1): No. 5 https://pubs.lib.umn.edu/index.php/innovations/article/view/3606/2668

[7] Wootton D. The Invention of Science. New York: Harper Collins, 2015

[8] 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