MAIMON WORKING PAPERS No. 17 July 2022

ANOTHER OWN GOAL: ICER's IMAGINARY CLAIMS FOR THE COST-EFFECTIVENESS AND PRICING OF WEIGHT MANAGEMENT THERAPIES

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

The publication of ICER's draft evidence report on the effectiveness of weight management therapies illustrates, once again, ICER's determination to put to one side the standards of normal science and fundamental evidence in favor of assumption driven simulations that produce invented imaginary claims [1]. ICER is perfectly aware of the criticisms but, as it apparently has no other option, it perseveres, together with its academic consultants, to present assessments which are meaningless [2]. After all, if data, as in the case of weight loss therapies are limited, as ICER acknowledges, then the ICER solution is to invent approximate information to fill the gaps rather than the more demanding task of suggesting a research program to meet evidence gaps.

The fundamental flaw in the ICER approach is to construct a simulation model built on assumptions. In the obesity case it is a Markov state transition model informed by assumptions based on the results of clinical trials, prior economic models, literature reviews and stakeholders. The model takes a lifetime horizon. This means that, by design, none of the claims made for the competing products has any hope of being judged empirically. The claims are entirely imaginary; there are any number of possible competing models each resting on another sit of assumption and each producing imaginary claims.

The first problem is this question of the role of assumptions. If this is a lifetime assumption driven construct based on prior observations to support assumptions then there is no reason to believe that these assumptions are valid for the future, even if it is a model. In logic, when we look to future events, the fact that past futures have resembled past pasts does not mean that future futures will resemble future pasts. If ICER is to be taken seriously then claims for the relative merits of competing products for obesity it must be in terms that are empirically evaluable. We have no idea if ICERs claims are right or wrong; we will never know and we were no intended to know. ICER cannot claim, in justification for its choice of dozens of assumptions, that these, for an unknown future, are realistic. We might as well advocate the choice of one video game over another on the grounds that the characters and scenarios are more realistic.

This is quite different from the standards of normal science, including mainstream economics and applied health economics, where the analytical frameworks are consistent with the standards of normal science and fundamental measurement given the focus on regression modeling for hypothesis testing. Hypotheses as to value claims are foreign to the ICER analytical framework.

The critical difference is that models and economics, and not pharmacoeconomics, create evaluable value claims. ICER ignores these considerations. Normal science standards for credibility, empirical evaluation and replication of value claims are of no interest; nor are measurement standards for a ratio scale to support claims for response to therapy.

Second, having assumed that patients move through various disease stages over their modeled hypothetical lifetimes (where it is assumed each stage lasts for exactly 12 months), health gains from therapy options are mainly derived from increased utility, with utility gain applied for each unit of weight loss due to treatment for each health state. This application of utility gains to health states in the model is a fundamental error which, apart from other failings, means that claims for QALYs and costs are meaningless. Rather than a focus on quality of life that is disease specific, ICER chooses to apply multiattribute generic measures that focus on clinical parameters not on the needs of patients who are obese.

Third, to create claims for quality adjusted life years (QALY) you require (i) a unidimensional utility measure which has ratio measurement properties, so that is can support multiplication and (ii) a measure that is bounded by 1 and 0. That is, it must have a true zero; under no circumstances can the measure produce negative utility values. The utility measures used in the obesity model fail on both these criteria; apart from the fact that the analysis fails to point out that a range of different utility measures are used together (the EQ-5D-3L, the time trade off (TTO) and the SF-36), which is invalid as they are different constructs. At the same time, the utility measures are composite ordinal scores; not ratio scores. Rather than separately valuing different aspects of health-related quality of life (HRQoL) such as pain and depression, the instruments rely on health state descriptions which bundle together different clinical attributes. This means they are dimensionally heterogeneous and lack construct validity. They are also ordinal measures; the algorithms used to construct these measures combine ordinal measure of symptom response levels. The scores lack interval properties, invariance of comparisons, because no one thought when they were developed that they should meet fundamental measurement standards.

While these might appear technically abstruse arguments, the bottom line is that as an ordinal score cannot capture response to therapy (as we don't know the difference between the scores) the ordinal utilities cannot support QALYs [3] [4]. When an ordinal score is used inappropriately to discount time spent in a health stat, as in the obesity model, it cannot create a QALY; it is mathematically impossible. ICER is aware of this criticism but chooses to ignore it claiming that utilities have secret ratio properties [5].

The fact that the QALY is mathematically impossible invalidates the entire ICER modelling exercise for weight management therapies. Add to this the assumptions for lifetime costs, specified for disease states, which are also entirely imaginary and we have value claims that must be rejected out of hand. The base case results (Table 4.4) are entirely fanciful with QALYs ranging from 16.95

to 17.85 and total costs ranging from $178,000 to $385,000 lack any meaningful interpretation. The incremental cost-effectiveness ratios (Table 4.6) share a similar fate with cost per QALY gained of $9,000 for Phentermine/Topiramate to $1,262,000 for Liraglutide. These are not believable and must be disregarded. Indeed, a feature which is common to many ICER models is the trivial difference between modelled QALY outcomes; in this case a difference of only 0.90 QALYs. This means that the burden of comparative claims falls entirely on the cost assumptions, an interesting result if the intent is to consider possible price discounting to meet imaginary threshold cost-per-QALY values.

The threshold cost-per-QALY values are, of course, meaningless as the QALY is a mathematically impossible construct (Table 4.7) even if presented in technically respectable (for the model analytical framework) in probabilistic sensitivity analysis terms. These are, for those seeking price discounting and pricing caps as a policy imperative a useful if meaningless result with only two weight management therapies proposed as being cost effective at the $100,000 threshold: Phentermine/Topiramate (85.3% likelihood) and Bupropion/Naltrexone (53.9% likelihood). This analysis leads, inevitably, to ICER's recommendations, which will be seized upon by the media, for price discounting (Table 4.11). The annualized imaginary price to achieve a $100,000 per QALY gained threshold would have to be reduced to $7,587 (from $13,618) for Semaglutide, to $3,607 for Liraglutide (from $11,760); to $1,465 for Phentermine/Topiramate (from $$2,990) and $2094 for Bupropion/Naltrexone (from $2,142). None of these recommendations should be taken seriously.

Mention should also be made of ICER's budget impact analysis, with detailed results presented in the case of semaglutide at the estimated net price of $13,618 per annum. The ICER budget impact threshold is of interest because it allows ICER to make claims for the number of eligible patients who could be treated within 5 years without crossing the threshold of $777 million per year for the US. The estimate for the next 5 years is 3.99%, with the figure increasing to 28.51% if the cost were $5,300 and a threshold of $50,000 per QALY. These estimates, given the application of impossible QALYs and the budget impact assumptions, provide additional imaginary ammunition to defend price discounting and/or restrictions on access to therapy. Once again, ammunition for price caps based on imaginary claims.

Finally, ICER claims that the obesity model has been validated (largely by comparing to equally imaginary other obesity simulations). This is a misuse of the term as we cannot claim that any model, imaginary or otherwise has been 'validated or confirmed' without reference to the empirical status of the value claims. Science is a process of discovery; in weight management therapies we are concerned, as always, with the strengths and limitations of existing therapies judged against the possibility of disproof of the claims made. This is central to progress in science; we must consider the asymmetry, recognized for almost a century in the philosophy of science (and in the philosophy of economics), that we can never prove a proposition; to validate or confirm

a claim is impossible. We can only fail to disprove it. All ICER has to offer is a one-off assumption driven simulation creating imaginary QALYs and costs, claiming they have been validated, with no chance whatsoever of being either being proved or disproved.

These observations are not lightly undertaken. It is widely recognized that, in subscribing to the modeling standards common in pharmacoeconomic that ICER is promoting in its model claims an analytical dead end. Unlike mainstream economics and applied health economics, both of which subscribe to recognized standards, ICER's position is to reject science in favor of non-science, to create modelled claims that have no relevance to pricing and formulary decision making whatsoever. The manifest deficiencies in the ICER model are well known: the argument is based on assumptions that attempt to create and unknown future, the application of QALYs which is just plain wrong with claims made that are impossible, under any circumstance, to assess empirically. We are asked to take ICER's word for it when the approach can support any number of alternative assumption driven models with the same result; there is no basis for claiming the superiority of ICER's imaginary construct and claims over any other. The essence of science is inquiry and the discovery of new facts for therapy impact; science appeals to the evidence. ICER is not interested; the discovery new facts is less appealing than inventing them. ICER promotes a will o'the wisp ersatz modeling exercise in an attempt mistakenly to inform decisions by assumption which should be rejected out of hand. We surely can do better!

REFERENCES

[1] Atlas SJ, Kim K, Beinfeld M et al. Medications for Obesity Management: Effectiveness and Value; Draft Evidence Report. Institute for Clinical and Economic Review, July 13, 2022. https://icer.org/assessment/obesitymanagement-2022/.

[2] Langley P. Nothing to Cheer About: Endorsing Imaginary Economic Evaluations and Value Claims with CHEERS 22 [version 1; peer review: 2 approved]. *F1000Research* 2022, **11**:248 (https://doi.org/10.12688/f1000research.109389.1)

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

[4] Langley PC and McKenna SP. Measurement, modeling and QALYs [version 1; peer review: 2 approved] *F1000Research* 2020, 9:1048 https://doi.org/10.12688/f1000research.25039.1

[5]Langley P. Peter Rabbit is a Badger in Disguise: Deconstructing the Belief System of the Institute for Clinical and Economic Review in Health Technology Assessment. *InovPharm*. 2021; 12(2): No.20