MAIMON WORKING PAPERS No 23 NOVEMBER 2021

ANOTHER IMAGINARY ICER FUTURE: ASSUMPTION DRIVEN INVENTED CLAIMS FOR TIRZEPATIDE IN TYPE 2 DIABETES

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

Abstract

Despite sustained criticism that the technology assessment claims promoted by the Institute for Clinical and Economic Review (ICER) and its supporting consulting academic research groups fail the standards of normal science, they continue to be the subject of draft and final evidence reports. The latest contribution to these fantasy constructs, the draft evidence report for tirzepatide for Type 2 diabetes, makes clear that ICER and its consultants have failed to appreciate the demarcation test, the distinction between science and non-science. For comparative value claims to be acceptable they must be credible, empirically evaluable and replicable. The ICER claims fail on all counts. Any value claim must meet the appropriate measurement standards for single attributes, with a meaningful timeframe for reporting the claim to health system decision makers. A blanket claim for cost-effectiveness is unacceptable. Creating a simulation model built on an assumption driven simulation stretching decades into the future not only defies the elementary rule of logic, Hume's problem, but rejects the standards that have held for the past century; rejecting inductivism in favor of falsification and its current applications. Modeling by assumption ensures that-the ICER tirzepatide model is just one model from potentially a myriad of others that can produce any number of non-evaluable claims for any pharmaceutical product or device Unfortunately, ICER and its expert consulting group are not alone; for 30 years the invention of non-evaluable 'evidence' claims have been the mainstay of health technology assessment. The belief is firmly held that invention of evidence by assumption is sufficient to inform formulary committees in pricing and access decisions. Health technology assessment must be the only discipline where evidence to support critical decisions in population health is invented and where none of the claims have any chance of ever being empirically assessed. The draft evidence report for tirzepatide should be rejected out of hand as an exemplar of this belief. Manufacturers and others concerned with hypothesis testing and real world evidence to support discovery in should reject outright the ICER analytical dead end. They should recognize that ICER and its expert consulting group occupy a unique position that rejects science in favor of non-science [metaphysics and pseudoscience] which they share with intelligent design. The purpose of this commentary is to make clear the manifest failings of the ICER commitment to inventing evidence, notably the failure to understand the limitations of fundamental measurement in value claims for competing products.

Keywords: invented claims, tirzepatide, ICER imaginary world, failing measurement theory

INTRODUCTION

The manifest deficiencies of the modelling framework and recommendations for pricing and access put forward by the Institute for Clinical and Economic Review (ICER) and its consulting expert

academic groups, are well known [1]. ICER is well aware of these but has decided, possibly to support its business model, not to respond to detailed questions that ask ICER to demonstrate that its models conform to the standards of normal science, in particular the axioms of fundamental measurement. This requirement is completely absent in the tirzepatide (Eli Lilly) ICER draft evidence report [2]. Instead we have a modeling framework that is designed to put to one side any consideration of credibility and empirical evaluation. We have no idea if the value claims and health benefit price benchmark pricing (HBPB) are right or wrong, we will never know and we were never intended to know. The framework and the present ICER model represent an analytical dead end.

Not surprisingly, after 30 years of inventing evidence in health technology assessment, supported by the leading textbook as primer for inventing evidence, the academic expert group members, consultants to ICER, concur in believing, as ICER does, *that all health economists apparently are confident that changes in the EQ-5D (and other multiattribute utility instruments) do have ratio properties* [3] Confidence is not proof and despite multiple requests of ICER a proof has never been forthcoming to demonstrate that the various ordinal generic multiattribute preference scores are in fact ratio measures in disguise.

But the concern goes deeper. Despite being unable to demonstrate ratio properties, the preference scores from the various instruments are applied to create the lodestar of invented evidence, the quality adjusted life year (QALY). This is an impossible mathematical construct as the relevant preference scores have only ordinal properties, apart from lacking dimensional homogeneity and construct validity. The impossible or I-QALY then becomes the key driver, apart from invented and non-evaluable direct medical cost assumptions, in generating non-evaluable incremental cost-per-QALY claims and social pricing recommendations in a model that looks 20 or 30 years into the future, denying the simple logic of Hume's problem of induction, supported by a claim for validation that is risible.

THE TIRZEPATIDE MODEL

The imaginary tirzepatide model developed by the ICER expert model consulting group is a patient level, Monte Carlo microsimulation first used in the ICER 2019 report evaluating oral semaglutide[4] . Unlike the traditional Markov cohort model with imaginary deterministic results, the microsimulation base case results for each outcome are the average of multiple assumption driven simulations to achieve statistical convergence. The conclusions of this commentary apply equally to this earlier model for oral semaglutide and its claims. The objectives of this simulation are not to produce credible and evaluable simulated value claims, but the invention of microsimulated imaginary non-evaluable claims for future costs, quality of life, clinical events and mortality associated with Type 2 diabetes as the basis for pricing and access to therapy.

By design (and intent) these are not empirically evaluable claims; there is no intent to develop a simulation to generate claims that can be empirically evaluated with real world data or future clinical trials. The imaginary microsimulation proceeds in two stages: event microsimulation and calculation of mean results from the pool of imaginary simulated patients' lifetime outcomes. The modeled

interventions for the lifetime of the hypothetical patient cohort were assumed to be tirzepatide plus background therapy versus (i) background therapy alone; (ii) injectable semaglutide and (iii) empagliflzoin. The assumptions for the simulation were based on short term clinical data and literature reviews as there were, not surprisingly, no long-term cardiovascular outcomes trial data existing for tirzepatide; intermediate outcomes were utilized to support assumptions. In other words, the data to support the model had to be assumed or invented. The key assumptions are detailed in Section 4.2 (Table 4.1) of the report. Ordinal utilities (or more accurately preferences) for the modelling (Section E2, Table E3) were captured by the ordinal multiattribute Health Utilities Index (HUI Mk3) and taken from the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial.

The imaginary results of this microsimulation all fail the test of being credible and empirically evaluable. The lifetime costs (suitably discounted) are incapable of any matching to target real world data bases, detailing unit impact by NDC and CPT codes. As far as can be ascertained, the base drug costs carry through the entire lifetime simulation (aggregated over the patients in the model). Clinical or non-drug costs were, without explanation, adapted from the literature'. The imaginary simulated base case for tirzepatide (Table 4.4) provided microsimulated total lifetime imaginary costs of $284,000 (95% range $255,000 to $315,000) with mathematically impossible quality adjusted life years (QALYs) of 4.69 (range 4.48 to 4.90). Reference case therapy involved discounted imaginary costs of $263,000 (range $235,000 to $291,000) and impossible QALYs of 4.14 (range 3.95 to 4.32) with no chance of any of these cost or QALY hypotheticals of ever being empirically evaluated or, indeed, set alongside competing imaginary models (e.g., with different multiattribute utilities and assumptions for the future).

The imaginary simulation modelling concludes with what are described as cost-effectiveness ratios (Table 4.5) defined as costs per QALY gained. These range for tirzepatide with a mean of $38,000 from -$33,000 to $91,000. It is of interest to note that the analysis supports negative values although it is not clear how these are to be interpreted for pricing recommendations. Avoiding this conundrum, the analysis then proceeds to provide annual prices to achieve imaginary cost effectiveness thresholds from $50,000 per outcome to $200,000 per outcome, with respective prices from $4,800 to $6,100. These cost-effectiveness claims are, as detailed below, meaningless.

A MULTITUDE OF MODELS

There is a simple point of logic (Hume's problem) that is overlooked (or never considered) in building assumption driven simulations: assumptions based on past observations have no claim on assumed future observations [5]. Just because past futures have resembled past pasts does not mean that we can assume that future futures will resemble future pasts [6]. In short, we cannot claim that one assumption about the future can be considered more 'realistic' than another because it is based on past observations. An assumption is an assumption; we might believe in the 'realism' of assumptions about the future, stretching decades into that imagined future, but that just reflects the model builders' psychology regarding assumptions [7]. This means, obviously, that the tirzepatide model, with its apparent realism of assumptions, is just one possible set of assumptions among a potential myriad of other tirzepatide imaginary simulation models. None can claim superiority over another as there are

no independent output claims criteria for making such an assignment. The tirzepatide microsimulation model thus is only one of many alternative models, both microsimulation or otherwise, that could have been developed to invent imaginary claims for competing social prices and deny ICER center stage. ICER and their consultants reject the only basis for evaluating the merits of the simulation: an appeal to the evidence for outcome claims that are possible to evaluate empirically; in short, as detailed below, a rejection of predictive validation for simulation models. Once we admit the fact that there can be a multitude of competing models, the ICER business model collapses.

SCIENCE VERSUS NON-SCIENCE

Following the seminal contribution of Karl Popper in the 1930s to demarcating science from non-science, the rejection of inductivism or logical positivism, the focus is on falsification [8]. If a claim is to have merit then it must be capable of empirical evaluation. Producing claims that are neither credible nor capable of empirical evaluation is a waste of time; there is no notion of null versus alternative hypotheses. In theory development, background knowledge is clearly important, but in evaluating a theory that is generating claims, these claims must be evaluable. It is this demarcation between science and non-science (metaphysics and pseudoscience) that is critical. If decisions are based purely on inductivist reasoning by assumption that, by design, admits of no possibility of empirical evaluation, then we are in the fantasy realm of relativism. That is where, in the case of health technology assessment, claims for cost-effectiveness are created. This is a sociological perspective where the same sort of explanation can be given to all types of knowledge claim [9]. Thus to a relativist, a microsimulation model producing non-evaluable claims has the same merit as a research program producing empirically evaluable claims; they should be treated as such in decision making. It is illegitimate to say that the simulated non-evaluable claim is less valuable than the empirical claim even if it is argued that there is good evidence for it. As Wootton notes, this belief in the notion of symmetry in claims excludes from consideration the very feature which makes scientific arguments distinctive; their appeal to superior evidence [9] . On these grounds, the demarcation between science and non-science, metaphysics and pseudoscience is one that has been ignored in health technology assessment for over 30 years: the strongly held belief that invented and non-evaluable claims (approximate information) must be an essential part of formulary decision making [10]. So-called 'pharmacoeconomists' have rejected willingly one of the key tenets in science: falsification. This is the contribution of the ICER expert group to the imaginary formulary assessment of tirzepatide which is fully consistent with the approximate imaginary evidence philosophy of the International Society of Pharmacoeconomics and Outcomes Research (ISPOR) and its global audience; and equally irrelevant.

HUME'S PROBLEM

Although it is some 360 years since Hume's problem was first formulated, ICER and the expert consulting group, not to forget ISPOR, have argued for the ability to essentially rank assumptions, justifying their choice as past claims on the future. This falls foul of Hume's problem: an assumption is just an assumption. This does not deny the role of assumptions in modeling or simulations, but only if the microsimulated claims are presented as likely to fail; then a failure of the claim leads us to reconsider the choice of assumption as part of the model. But this does not provide an excuse for

basing choice of assumptions with a notion of their being futuristically realistic where the model (or simulation) fails by design to have empirically evaluable claims. This just opens the door to multiple model options for selected products in disease areas driven by choice of assumption, where models can even be reverse engineered to produce the required comparative claims for price and access; possibly as marketing exercises.

MEASUREMENT

A perennial complaint directed towards those charged with instrument development for patient reported outcomes (PROs) in health technology assessment, is their apparent lack of understanding of the limitations imposed by the axioms of fundamental measurement. The failure is to appreciate that the overwhelming number of instruments, both generic and disease specific, are ordinal scales that cannot support the basic arithmetic operations of addition, subtraction, multiplication and division. Those charged with developing the tirzepatide microsimulation model to support cost-per-QALY claims fall squarely into this category. The failure of the model extends, not to the production, by design, of non-evaluable claims, but in the key position played by the HUI Mk3 ordinal preferences and the resulting mathematically impossible QALY. The HUI Mk3 preferences are ordinal; there was no apparent consideration that this scale should have interval let alone ratio properties. It was just assumed (without recognizing the assumption) that the scale would have to be ratio; otherwise it would not be able to support time in a disease state multiplied by a corresponding preference score to create a QALY. You require a bounded ratio scale where there is a true zero – death - and an upper bound 1 = perfect health, defined in terms of a single attribute; the algorithm creating the reference scores must be bounded in the senses that it is impossible, under any circumstances, for the algorithm to produce negative values. The HUI Mk3 can produce negative scores, while at the same time being dimensionally heterogeneous and lacking construct validity as a multiattribute scale. The same factors that render all multiattribute generic scores ordinal scales apply to the HUI Mk3.

With the decision in the early 1990s by leaders in health technology assessment that the focus should be on creating approximate information rather than hypothesis testing, the stage was set for imaginary simulations driven by assumption as the basis for formulary submissions. This has been the standard for the last 30 years with the leading textbook in technology assessment, as noted above, is essentially a primer for creating imaginary simulations and imaginary comparative cost-effectiveness claims [11] . For many moving on will be difficult; we are asking for a return to normal science and the rejection of non-evaluable product claims.

THE GREAT I-QALY DISASTER

Central to the ICER and expert group simulation modelling is the commitment to the QALY (as the unfortunate universal gold standard single metric for effectiveness), incremental cost-per-QALY claims and the application of cost-per-QALY thresholds to support pricing and the notion of an ICER defined social price or health benefit price benchmark (HBPB). As noted above, it has been demonstrated on a number of occasions that the QALY is a mathematical impossibility [12]. Under the axioms of fundamental measurement an ordinal scale cannot support multiplication; this is precisely what the

generic multiattribute preference algorithms (e.g., EQ-5D-3L/5L) produce. Proposing that a QALY can be constructed by multiplying simulated time spent in a disease state by an ordinal preference score is simply disallowed. The problem is that no one developing these generic preferences gave any thought to measurement theory. While you might assume that preference scores are actually ratio measures in disguise this 'understanding' falls short on three counts: first, there is no true zero as the preference algorithms produce negative scores; second, the preference scores by their very definition lack dimensional homogeneity (unidimensionality) and construct validity; and, third, there has been no recognition that it is important first to develop a scale with interval measurement properties and, second, translate this to a bounded ratio scale.

All generic preference scores combine symptoms or attributes which are presented as ordinal scales. They cannot be combined. To combine attributes you require each to have a ratio score; just as you require a single attribute ratio score to deflate time spent in a disease state by some version of a latent quality of life construct. As detailed below, a claim for cost-effectiveness is impossible if it rests on a multiattribute definition of effectiveness that fails to meet the standards of fundamental measurement. Outcomes are bundled together with no recognition that they should be empirically evaluable as single attributes; let alone able to be empirically evaluated.

There is, of course no 'universal' QALY; this would require agreement on a 'universal' preference scores with mythical ratio properties, from the range of scores available. Unfortunately, as the Versailles Conference taught us, committees of experts can never come to an agreement where developer egos are at risk. Although attempted, there is no possibility of a universal preference algorithm being agreed. This would be, in any event, a fruitless endeavor, given that all commonly used preference algorithms have only ordinal scaling properties. Discussions that have occurred on a 'universal QALY' have failed to consider the required measurement properties of a universal preference scale, let alone agreement on the chosen instrument[13] .

DIRECT MEDICAL COSTS

The assumptions regarding direct medical costs in the tirzepatide simulation are designed such that that any attempt to evaluate the costs with an appeal to real world data sets is doomed to failure let alone their discounted lifetime basis which ensures this is the case. If claims for costs are not expressed as resource units, as CPT or NDC codes, then it is impossible to go any further. It does not apparently occur to ICER or the expert group to consider resource utilization as a first step to claiming aggregate direct medical costs. This, of course, is consistent with the claims for QALYs where the construct is mathematically impossible. This leaves us with meaningless cost-per QALY claims, let alone incremental cost per QALY claims and the cost per QALY thresholds, when assessed in direct medical cost assumptions terms.

TORNADO DIAGRAMS

The fact that the simulation model is driven entirely by assumption with assumptions regarding the impossible QALY and the impossible costs deemed to be appropriate for decades into the future, renders the creation of tornado diagrams irrelevant. None of the elements of the tornado diagram

are, by design, empirically evaluable. In the case of tirzepatide the tornado diagrams for incremental costs (Table 4.2) and incremental QALYs (Table 4.3) are meaningless. Any number of competing microsimulations could produce alternative tornado diagrams (complete with alternative discounting assumptions) and all would be equally meaningless as the microsimulation is driven by assumption.

PROBABILISTIC SENSITIVITY ANALYSIS

Developed in the 1990s as a meaningful imaginary extension to the creation of assumption driven imaginary simulations, the purpose of probabilistic sensitivity analysis (PSA) is to provide assumption driven estimates of the likelihood that imaginary product prices are cost effective. Although the draft evidence report for tirzepatide does not present a PSA it is worth noting as another technique that defies accepted standards. PSA suffers from two fatal weaknesses: first, as the modeling is entirely by assumption there can be any number of PSA claims from competing sets of assumptions and, second, that the concept of a single metric to capture cost-effectiveness is mathematically impossible.

BUDGET IMPACT

As well as proposing imaginary social prices, ICER and the expert group also present an imaginary budget impact analysis; the downside of which is the potential for this to support denial of care with unfortunate eugenics implications [14]. For ICER, potential budget impact is defined as the total differential imaginary cost of using tirzepatide rather than the relevant existing therapy for the treated population, calculated as differential health care imaginary costs (including intervention costs) minus any offsets in these costs from averted health care events. All costs are undiscounted and estimated over a five-year time horizon. Assuming a placeholder price for tirzepatide of $4,643, ICER predicts that only 61.6% of eligible patients could be treated within 5 years of market entry and a 20% uptake each year without crossing, at the US national level, the potential budget impact threshold of $734 million. At annual prices of $4,800, $5,200 and $5,700 the eligible patients treated drops to 51.9%, 31.7% and 22.8%. These imaginary projections should not be taken seriously as they are driven by the results of the cost-effectiveness simulation model.

VALIDATION

We have extensive experience of validation with predictive simulation models going back 30 years or more. The focus of predictive simulation is on software to create statistically accurate models representing the behavior of real life systems and processes. ICER, true to its commitment to inventing non-evaluable claims for pharmaceutical products, does not involve itself with predictive simulation; indeed it is never mentioned as a criterion for modelling assessment. Validation of ICER models does not involve predictive simulation validation; indeed it is impossible. This is implicit in the creation of claims build on assumption driven simulations stretching decades into the future. We will never know if ICER and its expert modelers are right or wrong, we will never know and were never intended to know. Instead we have one simulation model out of a potential 'infinity' of models, each addressing the same cost-effectiveness question, each coming to a non-validated conclusion. For ICER and the expert modeling group, validation is self-referential. It involves reviewing the internal

consistency of the model, agreeing that the assumptions are appropriate and comparing the model to two or three other similar models. To this extent it is believed it has achieved its purpose of creating imaginary claims that are a valid basis for formulary decisions and social pricing recommendations, with the added smokescreen of constructed scenarios and sensitivity analyses to convince the uninformed.

Ignoring predictive validation is, to be blunt, a misrepresentation of the concept of statistical validation, whether intended or otherwise. Model validation for ICER refers to the process of confirming that the imaginary model achieves its intended purpose of generating invented claims. Predictive validation involves, after agreement on outcome measures that can be empirically evaluated, a comparison of modeled outcomes compared to an independent external data set. Data used in the estimation of model parameters cannot be included in the external data set. Each data point must be a single attribute with ratio measurement properties. The simulated predicted data points are compared or plotted to observed external data points using standard statistical techniques. ICER not only ignores predicted validation but ensures that there is no possibility that the simulation model is capable of producing such data points in the first place as all outcome are imaginary and impossible to assess from real world data bases. The point is, of course, that simulated predictions must be consistent with the constraints of external data (including proposed new data sets) and with evaluation in a time frame that is meaningful for decision makers. This is of no interest to ICER or the expert modeling group.

This does not mean that the term simulated or predictive validation has no intrinsic meaning; it is logically possible to assess a simulated claim if it is presented as output that can be falsified [15] [16] . Following the contribution of Karl Popper in the 1930s and the case for falsification, where the problem of induction meant that verification was logically impossible, knowledge grows through a rigorous process of conjecture and refutation with evaluable hypotheses, not to an analytical dead end of assumption driven imaginary simulations capped by meaningless PSA claims. This applies to simulation modelling where the objective is to compare simulation model outputs with real world data. Putting this to one side renders the exercise irrelevant to decision making in health care.

COST EFFECTIVENESS

Applying the term cost-effectiveness to ratios that fail the axioms of fundamental measurement is clearly a waste of time. The term cost-effectiveness is widely used, typically in the context of assumption driven simulations. The term effectiveness is not a single metric representing a single attribute with ratio properties. Costs, also discounted, are equally impossible to evaluate empirically. Claims that tirzepatide can be considered cost-effective are a disallowed amalgam of different ordinal measures presented as ordinal generic preferences. If, as is typically the case, effectiveness is defined in terms of QALYs, then the metric proposed is mathematically impossible.

Claims for cost-effectiveness, cost-utility or cost-benefit are inadmissible unless the numerator and denominator have single attribute ratio properties and can be empirically evaluated. Any claim for the likelihood of being cost effective (or not) driven in the case of tirzepatide by a microsimulation model

where costs are impossible to evaluate and the effectiveness is defined in terms of QALYs are clearly inadmissible. Indeed, there is no possibility of a construct we may label 'cost-effective' or a price that may be through the application of PSA labelled cost effective in probabilistic terms. Both are inadmissible. Unless carefully designed to capture specific evaluable attributes the 'general' term cost-effectiveness is meaningless.

CONCLUSION

The microsimulation model presented in the ICER draft evidence report for tirzepatide is emblematic of the failure of the approximate invented information paradigm that has been the mainstay of health technology assessment for the past 30 years. We have little to show for it. This model, in common with others that have proposed to look 20 or 30 years into the future to invent claims for competing products is an analytical dead end. It should play no part in formulary decisions. It is constructed, in defiance of the standards of normal science, on constructs such as the QALY which are mathematically impossible or, more specifically, a lack of appreciation of the limitations imposed by the axioms of fundamental measurement. Most importantly, none of the claims made for lifetime QALYs, lifetime direct medical costs or incremental cost-per-QALY claims has any chance whatsoever of being empirically evaluated; indeed, they were never intended to be evaluable. Validation of claims in the ICER and expert group simulation modeling is never achieved. Whether this is deliberate is an open question; all that can be said is that these analytical follies have no place in health technology assessment and formulary decision making. However, this is not to deny a prospective role for simulation modelling in health technology assessment; only to deny a role in imaginary constructs that deny the contribution of predictive simulation, falsification and reassessment of the model against real world data.

REFERENCES

[1] Langley P. Peter Rabbit is a Badger in Disguise: Deconstructing the Belief System of the Institute for Clinical and Economic Review. *InovPharm*. 2021; 12(2): No 22
https://pubs.lib.umn.edu/index.php/innovations/article/view/3992/2855

[2] Lin G, Brouwer E, Nikitin D et al. Tirzepatide for Type 2 Diabetes; Draft Evidence Report. Institute for Clinical and Economic Review, November 9, 2021. https://icer.org/assessment/diabetes-type-2-2022/#timeline.

[3] Langley P. Suppling with the Devil: Belief and the Imaginary World of Multiple Myeloma Therapies Invented by the Institute for Clinical and Economic Review. *InovPharm*. 2021; 12(3): No 6
https://pubs.lib.umn.edu/index.php/innovations/article/view/4215/2917

[4] Rind D, Guzauskas G, Fazioli K et al. Oral Semaglutide for Type 2 Diabetes: Effectiveness and Value. Institute for Clinical and Economic Review, November 1, 2019. http://icer-review.org/material/diabetesevidence-report/

[5] Hume D. An Enquiry Concerning Human Understanding. 1748

[6] Magee B. Popper. London: Fontana, 1973

[7] Russell B. The Problems of Philosophy. London, 1912.

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

[9] Wootton D. The Invention of Science: A New History of the Scientific Revolution. New York; HarperCollins, 2015

[10] Neumann P, Willke R, Garrison L. A Health Economics Approach to US Value Assessment Frameworks – Introduction: An ISPOR Special Task Force Report. *Value Health*. 2018;21:119-123

[11] Drummond M, Sculpher M, Claxton K et al. Methods for the Economic Evaluation of Health Care Programmes. New York; Oxford University Press, 2015

[12] 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

[13] Drummond M, Brixner D, Gold M et al. Toward a consensus on the QALY. *ValueHealth*. 2009;12(Supp1): S31-5

[14] Langley P. Abandoning eugenics and the QALY. *InovPharm*.2021;12(3):No. 20
https://pubs.lib.umn.edu/index.php/innovations/article/view/4291/2939

[15] Herskovitz P. A theoretical framework for simulation validation: Popper's falsification. *Int J Modeling Simulation.* 1991;11(2):51-55

[16] Klein E, Herskovitz P. Philosophical foundations of computer simulation validation. *Simulation Gaming*. 2005;35(3):303-329