MAIMON WORKING PAPERS No. 5 FEBRUARY 2022

COVID-19: PSEUDOSCIENCE, COST EFFECTIVENESS AND THE INSTITUTE FOR CLINICAL AND ECONOMIC REVIEW

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

Abstract

The experience with COVID-19 over the past two years has demonstrated the ability of individuals to accept misinformation regarding the nature of the virus, the emergence of mutations and the absolute denial of too many to accept the advice of medical professionals in the benefits, both personal and social, of vaccination. Unfortunately, the willingness to accept misinformation and the denial of the standards of normal science extends to claims for the cost-effectiveness of COVID-19 interventions. The recent report of the Institute for Clinical and Economic Review (ICER) repeats all of their mistakes of the past; it applies the standard assumption driven lifetime simulation model which has been roundly criticized. ICER's response has been to ignore criticism or offer poor excuses for its continued application. The COVID-19 simulation model is an analytical dead end. But ICER is not alone, the modeling framework is one that is endorsed by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), leaders in the field of economic evaluation in health care as demonstrated by the CHEERS 22 guidance released last month and the full support of editors of the leading health technology journals. All believe in the role of pseudoscience and the creation of imaginary value claims; rejecting the standards of normal science. The purpose of this brief commentary is to demonstrate that what ICER has produced is, from the standards of normal science, a complete waste of time. We cannot build non-evaluable claims on a series of selected 'reasonable' assumptions about an unknown future. ICER has been aware of the many criticisms of their reference case imaginary modeling for many years, yet refuses to accept any criticism of these modeled cost-effectiveness claims, where the notion of cost-effectiveness itself is meaningless. Why ICER takes this position is unclear although the ICER model for pricing and product access is its core business model.

INTRODUCTION

The Institute for Clinical and Economic Review (ICER) has never been put off by a lack of data in creating modeled cost-effectiveness value claims for pharmaceutical products. The analysis of the COVID-19 products is no different [1]. It presents an assumption driven lifetime simulation of value claims which are non-evaluable [2]. The model is, apparently, provisional and may change when input assumptions change with new data. What is not fully appreciated, notably among those who are naïve enough to take ICER economic evaluations at face value, is that any set of assumptions about an unknown future are no better (or worse) than any other set. ICER cannot fall back on the logical absurdity of one basket of assumptions gleaned from prior studies (or just guesses) being more realistic than another. This is emblematic of the ICER approach to value claims where the standards of normal science are put to one side and, in the case of assumptions, simple logic.

The purpose of this commentary is not to take issue at this time with ICER's interpretation of the limited clinical data for the four COVID-19 drugs considered, but to detail why the economic modeling to produce the impossible QALY claims is non-scientific or, to put it more bluntly, pseudoscience. It is not the intention here necessarily to single ICER out as the principal malefactor. The ICER modeling framework is the standard for health technology assessment. This meme or belief system that over the past 30 years has supported thousands of peer reviewed imaginary model claims; tracing its genesis to the early decision by leaders in the field to abandon hypothesis testing for approximate invented evidence where the claims were, by design, non-evaluable [3]. The meme is maintained not only by leaders in the area of health technology assessment who decided to reject hypothesis testing in favor of inventing approximate information to support formulary decisions at product launch when data are typically limited, but by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), as well as the leading journals in health technology assessment, evidenced by the wholehearted support for the recently published CHEERS 22 guidance for imaginary economic evaluations, led by *Value in Health* and the *BMJ* [4] .

IMAGINARY OUTCOMES FOR COVID-19

The COVID-19 ICER assumption driven lifetime simulation considers four COVID-drug interventions: (i) sotrovimab, a recombinbat human monoclonal antibody; (ii) molnupiravir a ribonucleoside analog; (iii) nirmatrelvir/ritonavir a combination treatment; and (iv) nirmatrelvir a protease inhibitor. The purpose of the technology assessment applies was to assess the modeled cost-effectiveness of these products, under a variety of assumptions, with a lifetime time horizon, capturing imaginary future costs and outcomes discounted by 3% per year, with a health sector perspective. The comparator in each case was usual care based on pivotal trial data. The model was a intent to treat analysis with a hypothetical cohort of patients, both vaccinated and non-vaccinated, with mild-0to-moderate COVID-19 treated in an outpatient setting entering the model with characteristics indicating a high risk of progression to severe disease or hospitalization.

Application of the model involves a cohort of patients transitioning between health states during cycles of one month over a lifetime horizon, modeling patients from treatment initiation until death. The model consists of an acute phase decision tree followed by a lifetime Markov model. The acute phase decision tree represents the COVID-19 infected period and assumed over time the highest setting of care received (e.g., outpatient management; emergency department visit; or inpatient hospitalization, with stratifications for level of respiratory support received). The lifetime Markov model consists of health states for alive and dead. Individuals in the alive health state who are assumed not to have experienced any long-term sequelae of COVID-19, have assumed costs and consequences characteristic of the general population. Hypothetical individuals who experience long-term sequelae of COVID-19 had additional utility decrements, costs, and mortality. Patients remain in the model until a transition to hypothetical death.

Cost effectiveness was estimated using QALYs, with incremental analyses comparing each intervention to usual care. Health outcomes and costs were assumed to be dependent on the highest setting of care received, respiratory support received if hospitalized. Model outcomes included costs, life years, quality-adjusted life years (QALYs), equal-value life years (evLYs), and inpatient hospitalizations.

None of these outcomes is empirically evaluable. This is by design for this class of lifetime model, which is the ICER standard for economic evaluations and claims for cost-effectiveness. As such, the model fails the standards of normal science, but is also subject to a number of deficiencies, chief amongst these is the role played by the QALY which, as will be pointed out below and has been in previous commentaries, is a mathematically impossible construct [5] [6] .

The imaginary base case results and incremental cost-effectiveness ratios are presented in Table 4.3 and 4.4 respectively. There is little to choose between any of the four products and usual care. The imaginary lifetime QALYs, including usual care range between 15.92 (usual care) and 15.96; the imaginary costs range from $297,000 (usual care) and $300,700 and life years ranging from 19.45 years (usual care) and 10.51 years; and the expected value of life years in the range 15.92 (usual care) to 15.97. The imaginary incremental analysis showed slightly more inter-product variation ranging from cost-per-QALY gained $6,000 to $69,000; cost per life year gained $5,000 to $58,000 and cost per evLY gained of $6,000 to $66,000. Cost per hospitalization averted ranged from $7,000 to $91,000. Finally, at a threshold of an imaginary $150,000 per QALY the treatment course price for the intervention alone ranged from $2,200 to $6,300. Note that while the criticism is directed to QALYs, the cost estimates are also imaginary, failing by design any attempt at evaluation.

None of these model results are to be taken seriously given the manifest deficiencies in the model: the denial of the standards of normal science and the insistence that the ordinal preference scores are ratio measures in disguise. If we consider the time and resources devoted to this exercise it can only be considered a monumental waste of time. A conclusion that applies to all ICER models.

DENYING THE STANDARDS OF NORMAL SCIENCE

Accepting the principle of demarcation between science and non-science (or pseudoscience) means recognizing that the scientific method rests on value claims that are credible, evaluable and replicable [7]. The ICER modelled value claims fail this requirement [8]. Rather than focusing on recommendations for meeting evidence gaps and proposing a framework for a process of the discovery, of new, yet provisional facts in the treatment of COVID-19, the ICER model is an analytical dead end; it is barren. At best, all ICER can offer is the possibility of other assumption driven non-evaluable claims when new data emerge which would be equally irrelevant in pricing and contracting with health systems. Interestingly, ICER provides the opportunity for a myriad of assumption driven competing models for specific products in disease areas which ICER has already modeled with the recently released *ICERAnalytics* package[9].

ASSUMPTIONS

The appear to be a quaint belief among those model builders committed to the invention of approximate information that it is possible to claim that a choice of assumptions regarding an unknown future can be justified by their realism; presumably based on past studies that appeal to the analysts involved. In other words, implicit in the lengthy list of assumptions that accompany ICER models (and including assumptions regarding the structure of the model) is the belief that these are the most 'reasonable' with a rationale for the assumption capturing key findings from existing studies. The same argument applies to the relative risks that are assumed for the model. This belief was abandoned over 250 years ago by David Hume in his proposed 'problem of induction' [10]. Often put in terms of the statement 'all swans are white', the problem of induction makes clear that even the so-called universal laws cannot be proved; there is always the possibility of a black swan. Or as Magee eloquently puts it: in logic, and with an unknown (by definition) future, the fact that while past futures have resembled past pasts it does not follow that all future futures will resemble future pasts [11]. Accepting the solution to the problem of induction, as Karl Popper pointed out creates an asymmetry: while we can disprove a statement or value claim, we cannot prove the claim. This eliminates logical positivism and belief in verification; Popper claimed he was solely responsible for its demise, with the publication in 1934 of *Logik der Forschung* [12].

If you make assumptions about an unknown future then this is just personal choice; there is no justification other that your psychology finds them to be the most 'reasonable' based on past observations. One assumption is no different from another; you may prefer one but this is not an objective choice. An assumption is an assumption which means, of course, that a multitude of COVID-19 models for these same products and the same timeframe could be constructed, each producing its own set of imaginary outcomes and pricing recommendations with no basis for choosing one model and its cost-per-QALY claims, and threshold prices any better (or worse) over another.

ORDINAL PREFERENCES

The preference utility scores that support the cost-per-QALY claims in the ICER model are from two recent papers reporting values for the EQ-5D-3L/5L [13] [14]. Unfortunately, neither paper considered the issue of fundamental evidence. To create a QALY you need a bounded ratio measure; that is, a measure with interval properties (or invariance of comparison) and with a true zero, capped at unity.
As a worthy complement to ICERAnalytics, is the commonly used Tufts University Medical Center's Cost Effectiveness Analysis (CEA) database of over 36,000 ordinal preference scores (health state weights) and cost-per-QALY claims from over 8,000 cost-per-QALY studies and assembled over the past 46 years [15] . The Tufts CEA, as an example, includes negative values as part of the database of preference scores (e.g., opioid abuse -0.064; tuberculosis-0.0.55) without realizing the implications of this for any claim that the scores are anything but ordinal with the absence of a true zero Again, in common with ICER, the Tufts CEA assumes that the ordinal preference have ratio measurement properties; in retrospect this belief, like the ICER model claims, appears a waste of time.

Indeed, there is now abundant evidence that denies that the EQ-5D 3L and EQ-5D-5L, and other direct and indirect multiattribute instruments, meet this bounded ratio requirement. There is no true zero as the various country specific scoring algorithms can produce negative health state values across numerous disease stares; there is no evidence for interval properties and the only reason the utility score is capped at unity is that utilities are defined as decrements for health state characteristics from unity. As no one thought that the issue of fundamental measurement was relevant in designing these instruments, they invariable overshot (despite econometric tweaking) the death = zero marker, producing states worse than death. No one thought that instead of fitting their model for preference scores to the data they might follow Rasch Measurement Theory and let the model determine the required data elements to fit the model [16] In fact, a bounded ratio scale is difficult to construct and, as far as quality of life is concerned, the only example is the Rasch measurement based need fulfillment or N-QOL bounded ratio scale which has been recently published [17].

The fact that the preference scores are ordinal rather than ratio is, without being too melodramatic, the death blow for ICER modeling. Ordinal scores cannot support arithmetic operations, including multiplication, only non-parametric statistics. This means that the QALY, multiplying time spent (a ratio measure) by an ordinal preference score is mathematically impossible. This brings down the entire ICER edifice of lifetime QALYs, lifetime cost-per-QALY claims, cost-per-QALT Y thresholds and value based pricing. OI course, ICER denies this with the somewhat odd claim that health economists have confidence that the preference scores have mystical ratio properties. A weak defense of an indefensible belief, but one that continues to take center stage in health technology assessment.

This complete lack of understanding of the limitations imposed by ordinal scores is demonstrated in the application of Covis-19 related disabilities (Table E9). The first step, mathematically disallowed, is to create an age adjusted utility (0.87) by discounting the unit utility of perfect health (an ICER adjustment). The second step, also disallowed, is to consider four disutilities ranging from emergency department visits (-0.30) to hospitalization with mechanical ventilation (-0.60). In this last case the presumed, yet mathematically impossible utility is 0.87 – 0.60 to give a utility score of 0.27. This entire exercise is absurd because the ordinal scale lacks invariance of comparisons; the EQ-5D-3L/5L algorithms, which give quite different scores for the same health state, were not designed to create scores with interval, let alone ratio properties.

It is worth noting that these disutilities do not match the utility weights presented in the website of the Tufts CEA database where all COVID-19 health state weights are negative (i.e., health state worse than death) which is not the case for the ICER report where the COVID-19 health states are all positive. According to the Tufts database the health state weights presented on the website (which capture direct and indirect multiattribute preference scores) a preference score of 0.27 (the worst outcome in the ICER model) is equivalent to a preoperative total hip or knee arthroplasty with COVID-19 weights ranging from -0.19 to -0.6. Needless to say, the Tufts database which is now 46 years old, has not apparently considered the implications of negative preference weights in terms of the axioms of fundamental evidence and the impossibility of applying any preference score to create QALYs.

VERIFICATION

One feature that stands out in the ICER modeling is their need to justify the model through a process of verification. As the claims are imaginary, the issue of falsification does not arise; instead we have a self-referential process where the imaginary modeled claims are verified by, first, providing the proposed model structures to manufacturers for feedback; second, parameter values were varied to establish face validity of results; third, the draft report was circulated again to manufacturers; and finally the model was compared to other published models (there were none apart from two unpublished papers). This hardly a ringing endorsement although there is no evidence that manufacturers challenged the model in terms of the standards of normal science. In any event, comparing the ICER model to similar models seems a pointless exercise apart from trying to identify which assumptions in the model are the culprits; illustrating that obvious point that one model cannot be judged 'superior' to another unless Hume's problem is violated .

FUTURE UNCERTAINTY

If the model attempts to parse an unknown future, a common feature of the ICER models and others is the treatment of uncertainty. In one respect this can be considered a fudge to give the model and its imaginary claims street credibility. One way and multiway sensitivity analyses for both the base case model and for the add-on scenarios that are always proposed, culminates in probabilistic sensitivity analysis (PSA). Indeed, the leading textbook in technology assessment is nothing more than a primer for the creation of imaginary claims, sensitivity analysis and PSA [18]. No recognition is made of the fact that preference scores are ordinal and the QALY is mathematically impossible; it joins ISPOR and the CHEERS 22 guidance in failing to appreciate that preference scores for QALYs must have bounded ratio properties, with no concern regarding the role of created evidence and imaginary claims.

A HOLLOW CLAIM

ICER's claim that the result of the economic modelling, the imaginary recommended prices, are cost-effectiveness is best ignored. The modelling suffers from manifest deficiencies that any claim fir 'reasonable pricing' and cost-effectiveness has no substance. The claim for cost-effectiveness can be rejected on two grounds: first, it rests upon a patently indefensible analytical framework and, secondly, it defies the standards of fundamental measurement to define effectiveness, driven by a QALY that bundles attributes into a single metric. The EQ-5D-3L/5L preference scores are by design multiattribute, with 5 symptoms or attributes and with three and five response levels respectively. The response levels are ordinal. As a result we have a preference score which is not only ordinal but one which lacks dimensional homogeneity and thus construct validity. Following the standards set by the axioms of fundamental evidence any score should relate to a single attribute with ratio or interval measurement properties. This is ignored entirely in the ICER modeling and, by extension, the CHEERS 22 guidance. A claim for cost-effectiveness, which is impossible with cost-per-QALY modelling, is made redundant by its ordinal base. Advocates of the cost-per-QALY belief system who envisage the QALY

and claims for cost-effectiveness to guide resource allocation in health systems should revisit their assumptions.

CONCLUSIONS

Health technology assessment, specifically in the case of economic evaluations, occupies a unique position; alone among the physical and mature social science where advocates such as ICER reject the standards of normal science in favor of pseudoscience. The only discipline where decisions, in this case for the pricing and access to pharmaceuticals, are built entirely on imaginary and invented value claims for cost-effectiveness. It is as though the scientific revolution of the 17th century had never occurred with the founding of institutions such as the Royal Society (1660) with its motto: *nullius in verba* (take nobody's word for it). In the case of ICER we are being asked to take their word for it. This is a retreat to Aristotle where 'science' or 'belief' is about rhetoric, persuasion and authority; not about abandoning consensus views when are at odds with the evidence [19].

REFERENCES

[1] : Yeung K, Whittington M, Beinfeld M et al. Special Assessment of Outpatient Treatments for COVID-19; Draft Evidence Report. Institute for Clinical and Economic Review, February 3, 2023. https://icer.org/assessment/COVID-19-2022/.

[2] Langley P. Nonsense on Stilts – Part 1: The ICER 2020-2023 value assessment framework for constructing imaginary worlds. *Inov Pharm*. 2020;11(1):No. 12

[3] 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

[4] Husereau D, Drummond M, Augustovski F et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 22) Statement: Updated reporting guidance for health economic evaluations. *ValueHealth*. 2022;25(1):3-9

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

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

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

[8]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

[9] Langley P. Let a Thousand Models Bloom: ICER Analytics Opens the Floodgates to Cloud Pseudoscience. *InovPharm*. 2021;12(1):No. 5

[10] Hume D. An Enquiry Concerning Human Understanding, 1748

[11] Magee B. Popper. London: Fontana Books, 1974

[12] Popper K. Logik der Forschung first published in Vienna in 1934 (dated 1935) published in English as The Logic of Scientific Discovery by Hutchinson in 1969

[13] Poteet S, Craig BM. QALYs for COVID-19: A Comparison of US EQ-5D-5L Value Sets. Patient. 2021;14(3):339-345.

[14] Sullivan PW, Ghushchyan V. Preference-Based EQ-5D index scores for chronic conditions in the United States. Medical decision making : An international journal of the Society for Medical Decision Making. 2006;26(4):410-420. 117.

[15] Tufts University Medical Center. Center for the Evaluation of Value and Risk in Health. Cost Effectiveness Analysis (CEA) database. https://cevr.tuftsmedicalcenter.org/databases/cea-registry

[16] Bond Y, Fox C. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. New York: Routledge, 2015

[17] Langley P. McKenna S. Fundamental Measurement: The Need Fulfilment Quality of Life (N-QOL) Measure. *InovPharm*.2021;12(2):No. 6

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

[19] Wootton D. The Invention of Science: A new history of the scientific revolution. New York: Harper Collins, 2015