MAIMON WORKING PAPERS No. 11 APRIL 2022

TITLE: FOXES AND HENHOUSES - THE ICER/ NORC WHITE PAPER ON RARE DISEASE DRUGS POLICY

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

Abstract

It is ironic that the Institute for Clinical and Economic Review (ICER) in collaboration with the National Opinion Research Center (NORC) proposes recommendations for the more rapid acceptance and introduction of rare disease therapies when ICER is, in fact, the principal barrier to achieving these goals in its commitment to creating imaginary claims for cost-effectiveness. The manifest deficiencies of the ICER assumption driven modelling approach to creating cost-effectiveness claims is well known; it is an analytical dead end due to its creation of evidence, typically at product launch, which generates non-evaluable claims which defy the standards of normal science and modern measurement theory. In the case of rare disease compounds, where the evidence base at product launch is often very limited, the incentive to create modelled claims instead of a commitment to a research program to discover new facts, progress in science, is put to one side in favor of a modelled easy fix. The attractions of this approach are clear; it provides a low-cost and plausible way of establishing price and access recommendations for rare compounds. Unfortunately, for those who subscribe to this belief in invented claims, it is meaningless. Fortunately, there is an alternative framework to support the entry of rare disease compounds: a return to the standards of normal science and fundamental measurement with a focus on establishing individual empirically assessable single attribute value claims to support provisional negotiations for pricing and access with health systems, subject to their ongoing evaluation. A commitment to a longer-term strategy to discover new facts for the benefits or otherwise of the rare disease compound.

INTRODUCTION

The purpose of the Institute for Clinical and Economic Review (ICER) and National Opinion Research Center (NORC) White Paper is to examine potential reforms to current policies and practices related to orphan drug development, pricing and coverage; exploring potential risks as well as advantages of reform options (Pg. 5) [1]. The purpose of this present commentary is to point to a fundamental failing in the White Paper: the failure to consider the application of the standards of natural science and the limitations imposed by fundamental measurement on valuing the contributions of new compounds, both for orphan drugs and, more narrowly, rare diseases. In other words, a failure by ICER/NORC to reject the current health technology belief system; to reject non-science in favor of science [2].

The failings of the current approach to health technology system are well known and extensively documented [3]. ICER as a committed follower. ICER is also aware of the manifest deficiencies of the belief in assumption driven imaginary simulation to create non-evaluable cost-effectiveness claims; resting in

large part on the mathematically impossible quality adjusted life year (QALY) [4]. There is no doubt as to ICER's commitment to inventing evidence and denying the standards of normal science and fundamental evidence; a commitment shared by hundreds of other analysts (including NORC apparently). For over 30 years, heath technology assessment has put aside any commitment to the discovery of new facts for new therapies to meet evidence gaps at product launch, in favor of assumption driven simulation models designed to invent evidence to fill those gaps and to present decision makers with non-evaluable claims for cost-effectiveness. While this is clearly an analytical dead end, and a path that should never have been followed in the first place, with support from ICER and others such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the recently released, and endorsed by ISPOR, guidance for creating imaging product claims, CHEERS 22, the belief system thrives: as shown in this White Paper. While analysts who subscribe to the approximate information belief system for rare diseases, application of CHEERS 22 is mathematically impossible [5].

DECONSTRUCTING MR FOX

There have been numerous critiques of the approximate information belief system, many directed towards ICER and its imaginary claims, yet ICER takes refuge in beliefs that are clearly mathematically impossible, notably its deeply held belief in the Holy Grail metric of quality adjusted life years (QALYs) to support its models and cost-effectiveness claims. Indeed, to claim a product is cost-effective at an ICER determined price, is mathematically absurd, let alone failing to give any guidance as to how this claim can ever be evaluated in the real world; which is impossible anyway. Without going into the details and arguments to demonstrate why ICER, now joined by expert opinion at NORC, believes in impossible created claims, the following flaws (fatal in almost every case) in the ICER business model for creating claims are self-evident; at least for those who subscribe to the standards of normal science.

This QALY based modelling is easily demolished. The QALY is based on multiplying model time spent in a disease state by a preference score (or utility). This is assumed to have a range from zero (=death) to 1 (= perfect health). However, to create a QALY you need to multiply time (a ratio measure) by another ratio measure; unfortunately, the preference score is only an ordinal measure. This means that the QALY is an impossible mathematical construct; it only exists in the mind of the model builder. There are three reasons: first, when the various preference systems and algorithms were created in the late 1980s, no one thought about the measurement requirements of a preference score which must have bounded ratio properties with a true zero; second, the instrument must have interval properties (implicit in a ratio measure) where there is invariance of comparisons; again, this was neglected; and, third, the instrument must be dimensionally homogeneous or have unidimensional properties. The various multiattribute preference scores (e.g., EQ-5D-3L/5L) fail to meet these standards. There is no true zero; the preference algorithms can create negative values or states worse than death, which could resonate, of course, in rare disease populations. A true zero means that under no circumstances can preference algorithm ever create negative values for a health state (the EQ-5D-3L range is from I to -0.58). The lack of invariance means that preference scores, with possible negative QALYs, cannot capture response to therapy. At best we can apply non-parametric statistics and create, for the ranked

preference scores, medians and modes. We cannot even compare the 'distance' between successive medians and label this a response; all it means is that one median number is 'higher' than another. Finally, multiattribute preference scores reflect the bundling together in a single equation of selected symptoms (or attributes) and response levels. This is disallowed because the various response levels are ranked ordinal responses. We can only combine attributes is they each have ratio properties. Preferences, therefore, are dimensionally heterogeneous and lack construct validity; combinations are disallowed. This means, of course that before we even think of creating a QALY, the various preference algorithms which either combine additively or multiplicatively symptoms and responses, with *ad hoc* 'adjustments' to try an avoid negative values, are themselves disallowed.

The failures of the QALY and attempts to add QALYs together are not the only fatal flaw (presumably we can have more than one fatal flaw) in modelling simulated claims. What is also overlooked is the issue of applying modern measurement theory (Rasch Measurement Theory) to creating interval scores for single attribute or latent constructs; and in some instances, a transformation to a single attribute bounded ratio scale [6]. The choice of symptom and response levels in multiattribute measures are essentially clinician determined (with *ex post facto* minimum inputs from patients). They are also generic, which means the symptoms and response levels may have nothing to do with patient and caregiver experience and need in rare diseases. The patient voice is missing. If quality of life claims are a possible value claim then they should reflect the extent to which patient or caregiver needs are fulfilled following the introduction of a new therapy. But we are not focusing on just clinician inputs; clinical success, a dictated by a clinician, may have little relevance to meeting the needs of patients and caregivers. We need to consider a single attribute, latent construct, based on extensive patient or caregiver input to develop, if it proves possible, a measure of the need fulfillment. This measure, which relies on RMT to fit data items to the attribute measure, has been applied since the early 1990s with some 30 disease states presently covered. All too few, unfortunately are rare diseases but the possibility of developing such as instrument should be considered as part of the product development process for new compounds as a value claim which meets required fundamental measurement standards. The key to RMT is the simultaneous assessment of the needs of the patient or caregiver and the ability of the patient or caregiver to meet those needs, as they become more difficult, with a new therapy. This creates n interval scale for quality of life.

A further flaw in the approximate modeled information belief system is the treatment of assumptions. There is a belief that assumptions to support modeling for an unknown future can be justified by the realism of assumptions chosen for the model. These would support the choice of model structure (usually a Markov or semi-Markov process) the data elements that populate it; including in particular the choice of preference (or utilities) to create imaginary QALY claims within the model. The problem is one of simple logic; claims from the past, observations, cannot justify their application in the future. This is Hume's problem of induction, first proposed in 1748. Put simply: the fact that past futures have resembled past pasts does not mean the future futures will resemble future pasts. The choice of assumptions to populate a modeled simulation of an unknown future cannot rest on past confirmation, irrespective of the number of observations or the scope of a systematic review. Any choice is in the mind of the analysis; it reflects his/her psychology. We must abandon all justification for our

expectations about the future. The implication is interesting: if we cannot claim superiority of one model and its claims over any other on the 'realism' of its assumptions then one model's claims have the same status (or lack of status) as any other. In rare disease, an ICER model may recommend substantial price discounting, but that is one recommendation from one of possible multitude of models each with different assumption permutations. The claims should not be taken seriously; in fact they should never be made in the first place.

But there is a more fundamental issue which takes us to the essence of evaluable value claims: the question of falsification. The promotion by groups such as ICER of imaginary claims for QALY-driven cost-effectiveness models effectively excludes any opportunity to test hypotheses and discover new facts in rare disease therapies; it also excludes, by design, any attempt to falsify claims. The formulary committee is asked to accept as face value the claim for imaginary cost-effectiveness. Recipients may be at ease with this, particularly if they have been trained in the approximate information belief system to accept without questions the belief in imaginary modelling; where truth is consensus supported by rhetoric and authority [7] . Despite this denial, the process of drug discovery in rare disease and across the board, must meet the standards of normal science in protocol driven claims and their assessment. A failure to meet recognized standards for acceptance leads to rejection and abandoning many compounds. ICER and the approximate information belief system is quite clear that this is rejected. Instead, it embraces the non-science or pseudoscience of invested non-evaluable claims.

A FAILURE TO COMUNICATE

Addressing the question of value assessments (pg. 22) by health technology assessment groups, the report commits a fundamental error: it fails to distinguish assumption driven invention of data from a commitment to a structured research program, the discovery of new, yet provisional facts. Assumption driven lifetime simulation modeling in rare disease is a non-starter; and has been since the invention of science in the 17th century and the focus on empirical evaluations. Assumption driven simulations are, of course, ICER's business model. The report argues that 'Less robust data complicate the process of designing cost-effectiveness models to evaluate the long-term value for money of orphan drugs …. (where) clinical experts and patients and families often have to supply model inputs in lieu of good data from clinical trials and broader epidemiological studies". Whether data are robust or not is irrelevant; realistic assumptions are a fallacy to support simulation models; the proposal to 'better' define cost-effectiveness models is a non-starter. The result, apparently from constructing these imaginary simulations, is that "…the higher prices of orphan drugs often drive unfavorable cost-effectiveness results". This, again apparently, is not mitigated by the "Application of higher cost-effectiveness thresholds …". Even if required data elements were accessible, they should not be plugged into a simulation claims for an unknown future, stretching decades ahead. ICER overlooks a simple point of logic: the problem of induction [8]. The fact that past futures have resembled past pasts does not mean that future futures will resemble future pasts; there can never be a claim, in logic, that one assumption to support a modeled future is more 'realistic' than another, even if supplied by clinical experts, patients and their families at ICER's request [9]. An assumption collection box is irrelevant; claims from the past,

which are the basis for 'realistic' assumptions, cannot support claims on the future. All swans are not, apparently, white; we cannot prove or disprove induction by an appeal to experience.

ICER is wedded to the creation of evidence through imaginary assumption driven simulations, presumably bolstered by claims for the greater realism of assumptions when 'better' data become available [10]. This embrace of non-science, the deliberate creation of non-empirically evaluable cost effectiveness claims, is a fundamental weakness which should have been noticed by reviewers of this reports and the many participants involved in system development [11]. So-called claims for "long-term value for money" fail the standards of normal science and fundamental measurement. Value assessments are not built on analytical dead ends, although many analysts pursue these will o'the wisps and even try to improve on their imaginary frameworks to light the way. Although unlikely, given ICER's interests, this absurd approach to value claims should have been recognized and abandoned; at least for a readership who might be unfamiliar with these standards.

Certainly, there is a need for a stronger evidence base to support value claims, their protocols and their evaluation, and we should applaud efforts to meet more stringent evidence standards; but this does not mean we invent evidence, as ICER has done for its assumption driven simulated models going back for 10 years or more. Registries are one option but before jumping to such conclusions we should ask a more pertinent question: what are evidence requirements and expected measurement standards for value claims? Not only does the report ignore completely the standards of normal science where value claims must be credible, evaluable and replicable but also required measurement standards where these single attribute value claims have ratio or interval properties. Unless we are clear about the nature of value claims for technology assessment in rare disease, we might as well forget the entire exercise. The unwelcome truth, at least for cost-effectiveness imaginary models, is that there is no universal metric to capture cost-effectiveness; the concept is irrelevant (and mathematically impossible). We have to focus on a research strategy, including the design of RCTs that capture single attributes. This may not be as exciting as a single quality adjusted life year (QALY)-based blanket lifetime benefit claim with imaginary pricing points and recommendations, but it is a fact of life and mathematically defensible. As a final point, given the previous reference to QALYs, note that nowhere in the report is there any mention of the QALY although cost-per-QALY and threshold claims which are based on QALYs are critical in ICER modeling. The QALY is, for this report, the ghost in the room. Perhaps it is not a metric that should be raised in discussions of orphan and rare drugs; or perhaps the answer is more prosaic: the QALY is an impossible mathematical construct, as ICER is well aware [12].

A NEW START FOR RARE DISEASES

If we abandon the approximate information assumption driven lifetime models, and there is no alternative in health system decisions, although journal editors may accept fantasy models, then we have to accept the standards of normal science and fundamental measurement. These will not support lifetime modeling, although there is the possibility of models based on clinical trials and for short term extrapolations as long as the required standards are met to support, for example, composite claims as extensions of single attribute ratio claims. Of course, retrospective data can support econometric

modelling (e.g., determinants of adherence/compliance behavior) and these may support further value claim hypotheses for empirical evaluation. In all cases, however, the modeling must be supported by data input measures that meet ratio or interval properties; blanket claims for lifetime modeled cost-effectiveness claims are mathematically impossible, and hence irrelevant in health care decision making and policy prescriptions. It should be made clear to all parties in rare disease therapy options that no credence should be given to assumption driven lifetime simulations to support some national reference notion of a social price or value-based price for an orphan drug or rare disease compound or, as the White Paper suggests, a potential (undefined) value-based national benefit (Pg. 37). The last thing we should aim for is to bring QALYs or some similar gold standard metric that defies the standards of normal science and fundamental measurement, to sustain what is, in the ICER stable, a commitment to pseudoscience.

CONCLUSIONS: SAVING THE HENS

No one doubts the challenges of value assessment, pricing and access for rare disease drugs. A situation made more intractable, unfortunately, by a continued insistence and emphasis by ICER/NORC on the positive contribution of assumption driven invented claims from lifetime simulation models; the ICER creation. We can certainly do better if we decide to reject non-science in favor of science, establishing evidence standards for value claims and value assessment.

If value claims and their supporting protocols, and possible contracting are the only option, then it is up to the manufacturer and health system to decide how to implement value claims that meet the required standards. One size does not fit all; irrespective of ICER's attempts in the past to extend its assumption driven simulation modelling to orphan drugs and rare disease, this is a futile and unnecessary endeavor. Decisions on pricing and access must be decided by negotiation, subject to requirements a health system may put in place for invited formulary submissions. Attempts to balance both innovation and affordability through impossible modeled simulations must be rejected. Certainly, we must consider eliminating barriers to investment in compounds for rare disease to meet a significant unmet medical, but this does not require imaginary cost-effectiveness models and pricing recommendations as a meaningless barrier. It is up to the parties involved to establish the basis for pricing and access, subject to the considerations of the standards for normal science and fundamental measurement. We don't need Mr Fox to set criteria for acceptance or rejection.

REFERENCES

[1] Pearson C, Schapiro L, Pearson S. The Next Generation of Rare Disease Policy: Ensuring both Innovation and Affordability. Institute for Clinical and Economic Review, April 2022

[2] 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;21:119-25

[3] 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)

[4] 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

[5] 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

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

[7] Dawkins R. A Devil's Chaplain. New York: Houghton Miflin, 2004

[8] Russell B. The Problems of Philosophy. 1912

[9] Magee B. Popper. London: Fontana, 1974

[10]Langley P. Supping with the Devil: Belief and the Imaginary World of Multiple Myeloma Therapies Invented by the Institute for Clinical and Economic Review. *Inov Pharm*. 2021; 12(3): No. 6.

[11] Pigliucci M. Nonsense on Stilts: How to Tell Science from Bunk. Chicago: University of Chicago Press, 2010

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