

Medicaid Formulary Decisions and the Institute for Clinical and Economic Review: Abandoning Pseudoscience in Imaginary Pharmaceutical Pricing Claims

Paul C Langley, PhD Adjunct Professor, College of Pharmacy, University of Minnesota

Abstract

Medicaid formulary committees and other gatekeepers face a difficult task. On the one hand they can utilize technical expertise in evaluating the real world evidence for clinical, quality of life and resource utilization claims for competing products while on the other hand they may be asked to assess claims built by simulation models for pricing and product access. A common option has been to take modeled claims from third parties such as the Institute for Clinical and Economic Review (ICER) at face value without challenging the model structure, its assumptions and its incremental cost-per-QALY claims set against competing products or the existing standard of care. Unfortunately, from the available evidence, it seems clear that many formulary assessment groups, last but not least those for whom the ICER modeling claims are targeted, have little if any appreciation of the limitations of ICER modeling. There are two substantive issues: (i) a failure to appreciate the limitations imposed by the standards of normal science for credible, empirically evaluable and replicable product claims and (ii) an understanding of limitations imposed by the axioms of fundamental measurement. In the latter case, a failure to recognize that the quality adjusted life year (QALY) is an impossible mathematical construct (hence the I-QALY). To these limitations should be added the potential for constructing competing imaginary claims. Surprisingly, ICER has provided the ideal opportunity to construct competing claims with the launch in late 2020 of the ICER Analytics cloud platform. Formulary committees and other health decision makers should be aware that claims based on the ICER Analytics platform together with competing lifetime modelled claims all fail the standards of normal science. Factoring these into formulary decisions is not only misguided but may have unintended consequences for pricing and access that may disadvantage significantly patients and caregivers. We have spent too much time debating the merits or otherwise of the I-QALY for targeted patient groups with the parties failing to recognize that the focus on simulated cost-per-I-QALY value assessments is a mathematical folly; I-QALY claims are a chimera. The I-QALY, at long last, should be abandoned together with modelled lifetime simulations. Medicaid formulary decision makers should rethink the required evidence base for formulary decisions and negotiations. Care should be taken to revisit previous negotiations where ICER recommendations have been utilized to support pricing and access.

INTRODUCTION

If we are concerned with the integrity of evidence based medicine and feedback from patients and caregivers in real world treating situations, then the tools at our disposal must be consistent with the requirements of fundamental measurement^{1 2 3 4}. This is widely recognized in the physical sciences where instruments are designed to capture single attributes (e.g., temperature, mass). Unfortunately, when decisions are based on modeled imaginary worlds, as in health technology assessment we find that an appreciation of the limitations imposed by fundamental measurement are either ignored or were never appreciated in the first place. Claims are made and embraced by those who should know better, in defiance of the standards of normal science. This is unacceptable. When formulary decisions have the ability to harm patients and caregivers, decisions should not be based on imaginary model simulations which are mathematically nonsensical, but on real world evidence.

If we accept the role of the standards of normal science, to include meeting those of fundamental measurement, then there is clearly a significant disconnect between the assessment standards typically applied in health technology assessment and those of normal science. Recognizing this disconnect is not new. Previous commentaries in *Innovations in Pharmacy* have made clear the manifest failing of health technology assessment with the widespread acceptance of modeled imaginary lifetime simulations to create (i.e., invent by assumption) claims for 'fair' pricing and access to pharmaceuticals. In the US, the Institute for Clinical and Economic Review (ICER), as the self-appointed technology assessment arbiter, plays a key role in recommendations for pricing and access based on imaginary assumption driven constructs; a position which is untenable.

The purpose of this commentary is to make the case that ICER's role in formulary decisions, not only for state Medicaid groups but also for agencies such as the Veterans Administration should be put to one side. While ICER views itself as providing an independent assessment of comparative product claims, 'fair' prices and access to products, this contribution is only acceptable if the assessment meets the standards of normal science. ICER's contributions do not meet these standards; they are imaginary pseudoscientific inventions⁵. A balanced review of the benefits and harms of new products, including ongoing disease area and therapeutic class reviews should not

Corresponding author: Paul C Langley, PhD
Adjunct Professor, College of Pharmacy
University of Minnesota, Minneapolis, MN
Director, Maimon Research LLC; Tucson, AZ
Email: langley@maimonresearch.com
Website: www.maimonresearch.net

accommodate imaginary claims; claims that fail the demarcation test between science and pseudoscience, sharing the Dover courtroom with intelligent design ⁶.

STANDARDS OF NORMAL SCIENCE

It has been recognized for the past 30 years that in health technology assessment hypothesis testing has been rejected in favor of creating (or inventing) approximate and impossible information to support formulary decisions ⁷. The reasons for this have been detailed in a recent commentary, but suffice to say it is a response to limited information on product performance following FDA approval and market entry ⁸. Rather than propose a research program to meet evidence gaps, leaders in the field opted for evidence creation through assumption-driven simulated lifetime modelled claims linked to limited phase 3 randomized trial data for product efficacy. Unfortunately these modelled claims lack credibility; they are not empirically evaluable let alone replicable across real world treatment settings. As the claims are driven by lifetime models that track hypothetical patient cohorts over their lifetime we have no idea if they are 'right' or if they are wrong, we will never know and we were never intended to know. The claims are not credible, empirically evaluable or replicable. For ICER and manufacturers who support this approach, this is a win-win situation. Claims can only be challenged by challenging assumptions; an ultimately futile exercise creating a multitude of competing models.

The failure to meet the standards of normal science is not just the rejection of hypothesis testing in favor of approximate information, but a more egregious failure: a rejection of the axioms of fundamental measurement. Science, if it is to advance through a qualified process of conjecture and refutation, of hypotheses testing and the discovery of new yet provisional facts, must rely on accurate measurement not on assumption driven non-evaluable claims ⁹. This is exemplified in the process of pharmaceutical product development from pre-phase 1 evaluations, through to pivotal phase 3 trials. Unless development and audit standards are met with agreed calibration of therapy response, then FDA review and marketing approval is impossible. Yet when we come to claims for marketing and cost-effectiveness claims these standards are, all too often, deliberately put to one side in favor of constructing approximate information to support a client's case.

In the physical sciences, and the more rigorous, and aware, social sciences such as education, psychology and mainstream economics, an understanding of the axioms of fundamental measures is recognized and is considered essential in measurement ¹⁰. Following the formalization by Stevens and others in the 1930s and 1940s, the axioms of fundamental measurement are well understood ¹¹. The measurement scales are nominal, ordinal, interval and ratio. Each scale of measurement has one or more of the following properties: (i) identity where each value has a unique meaning; (ii) magnitude

where values on the scale have an ordered relationship with each other but the distance between is unknown; (iii) invariance of comparison where scale units are equal to each other in an ordered relationship and known distance; and (iv) a true zero where no value on the scale can take negative scores. The implications for the ability to utilize a scale to support arithmetic operations (and parametric statistical analysis) are clear cut. A nominal scale is just a set of unique meanings but nothing else (e.g., gender). An ordinal scale has identity and magnitude in an ordered relationship but we do not know the distance between the values (i.e., it cannot support arithmetic operations, only non-parametric statistical evaluations, modes and medians). An interval scale has known differences but no true zero and can support only addition and subtraction (i.e., it can change the point on an integer line but only relative to other points). A ratio scale can support the additional operations of multiplication and division because it has a true zero (i.e., it can change the point on an interval line relative to zero).

In the early 1960s a new approach to fundamental measurement was introduced: probabilistic conjoint simultaneous measurement ¹². This new measure subsumed the existing fundamental measurement categories, providing a framework for identifying measurement structures in non-physical attributes. This provided the basis of going from ordinal to interval scales, becoming known as the Rasch model or Rasch Measurement Theory (RMT) where two attributes such as difficulty of a question and the ability of the respondent can be jointly evaluated to determine whether or not an interval scale measure might exist to capture a latent trait or attribute such as needs fulfillment quality of life as a measure of therapy response. Latent traits are not directly observed; only their outcomes which provides a basis for inferring the presence and amount of the latent trait. To achieve appropriate measures requires a deliberative process, not just allocating numbers to events. RMT does not replace statistical analysis, it precedes it.

A further key point is that if these axioms are to be applied, then any instrument must be dimensionally homogeneous; it must be unidimensional in applying to a single attribute whether this relates physical or non-physical (or latent) attributes. Construct validity demands that an instrument is focused on capturing a single attribute (e.g., temperature, weight). This requirement is invalidated in the overwhelming majority of patient reported outcome (PRO) measures. In the case of ICER where a key input to the imaginary model is the EQ-5D-3L utility score, the instrument, apart from being an ordinal scale, lacks construct validity as it combines five symptoms or health status attributes (e.g., mobility, self-care, usual activity, pain/discomfort, anxiety/depression) as elements in the utility scoring algorithm. This is a common and fatal feature of all preference and multiattribute utility systems (i.e., standard gamble, time trade-off, EQ-5D-5L, HUI Mk2 and Mk3, SF-36, SF-12, SF-6D and QWB). These are all ordinal measures. They may give the

impression of having interval properties but this is because the scores are typically presented on an integer or number line with equal intervals; a common mistake. The score could equally well be placed and ranked on a number line with unequal intervals. In either case the distance between values is unknown.

Most importantly, we cannot assume a given scale (e.g., utilities) is an ordinal, interval or ratio scale. Unless a scale is designed to have interval or ratio properties it is, by default, an ordinal scale; a ranking of raw scores. Understanding this points to the importance of Rasch measurement theory (RMT) which is quite clear in that it is only possible to create an interval scale for latent variables such as quality of life if there are techniques for translating raw scores or ordinal values to an interval scale¹². RMT has demonstrated for the last 60 years that these techniques are available; a measurement scale has to be designed to have the properties that are required for the latent attribute of interest.

ICER AND THE ICER ANALYTICS WINDFALL

Central to the ICER imaginary simulation is the quality adjusted life year (QALY); or, more accurately, the impossible or I-QALY. This is created by multiplying time spent in a disease stage (generated by the imaginary simulation) by a utility score (assumed to be a ratio scale with a range of 1 = perfect health to 0 = death). The product yields imaginary years of perfect health. The ICER simulation generates estimates of future years of perfect health for competing products and the associated assumed direct medical costs. These two imaginary quantities yield, in turn, lifetime imaginary cost-per-I-QALY estimates. These are compared to cost-per-I-QALY thresholds (imaginary societal willingness-to-pay thresholds) and recommendations made for possible pricing discounts to ensure the imaginary product price adjusted cost-per-I-QALY is less than a notional threshold. There are no empirically evaluable claims (by design).

Obviously, as noted above, there is a potential multiverse of simulations for individual products; models each yielding, by accident or design, different and possibly contradictory claims for pricing adjustments, budget impact and access recommendations, each driven by its own set of assumptions, yet with the QALY as lynchpin. The possibility of manufacturers developing competing simulation models to support claims for 'cost-effectiveness' by a judicious choice of assumption has long been recognized. A typical response to bring order to chaos with a referee, has been for single payer health systems to mandate standards for models developed and submitted by manufacturers, to review the model and then make a pricing and access determination. The classic example is the role of the National Institute for Health and Care Excellence (NICE) in the UK and its engagement with academic centers with long experience in judging the merits of imaginary model simulations and imaginary pricing¹³. These academic centers are, apparently, unaware of the standards of normal science.

ICER has attempted to emulate this approach. ICER contracts, again to academic imaginary modelling centers, the construction of a model. Following release of a draft report and public comments (often from manufacturers who recommend changing assumptions), ICER releases the final imaginary model with pricing and access recommendations. Recently, unlike the situation in the UK and other single payer health systems, ICER has decided to open up its product specific models; to turn the world upside down. ICER has released a 'do-it-yourself' platform, ICER Analytics, where there is access to an ICER 'backbone model' for all evidence reports^{14 15}. This allows the model assumptions to be changed and revised claims for pricing and access, as well as budget impact, 'imagined'. Why ICER has supported this is a puzzle as it devalues their 'base-case' reference model, making it perfectly obvious that it is nothing more than one of many options. Formulary committees and other decision makers now face the unenviable prospect of a model deluge with each claiming to be based on the ICER backbone model for that disease area, to include both new product claims and 'revised' claims from previous ICER evidence reports.

Given there are many preference or multiattribute instruments yielding utilities, each with its own set of symptom dimensions and responses within these dimensions, there is no 'universal' utility to create a 'universal' I-QALY. The ability to 'invent' utilities (e.g., guesses by key opinion leaders) and apply the ICER Analytics platform adds a pivotal option for model builders. Fundamental measurement standards will, no doubt, be ignored. ICER, for example, has been aware for some time of the ordinal nature of utility scales, yet chooses to ignore the science. Attempts to ask ICER to demonstrate that the EQ-5D-3L utility scale, for example has ratio properties has been doggedly resisted^{16 17}. ICER's position has been contradicted by one of the leading technology assessment textbooks where it is recognized that utility scales are not ratio scales¹⁸. Unfortunately, the case to defend the I-QALY then becomes more confused where the argument is then presented that because the EQ-5D-3L has interval properties it can support I-QALYs. This is clearly incorrect, apart from the fact that the EQ-5D-3L does not have interval properties.

The result is a disaster for ICER: the QALY is an impossible construct. We cannot multiply time spent in a disease state by an ordinal score. It is mathematically impossible. This means that any model constructed around IALYs and manipulation of those I-QALYs is nonsensical. The basic building block is a mathematical chimera. ICER and others appear to have the impression that because utility scores are typically presented on an integer number line with equal intervals that the raw score have interval properties. Having a scale with interval properties also does not mean that it has ratio properties.

The ICER Analytics cloud platform also offers the manufacturer to demonstrate why in ICER models the utility gains between

new and comparator products are often minimal; a lifetime I-QALY difference of only a few months or years. The answer is straightforward: the choice of a generic multiattribute instrument where the symptoms are limited and the ordinal response levels also limited. The EQ-5D-3L, for example has five symptoms and 3 response levels for each. As the symptoms covered in the EQ-5D-3L instrument are limited then it is likely that a number will not be relevant to the target disease state. This means that the utilities, given the EQ-5D-3L algorithm for creating a utility score, will cluster toward the 'perfect health' axis of the scale and that new products which may be targeted to an unmet medical need in that population will score poorly on any attempt to argue for their improved efficacy defined by five symptoms. Hence, in the imaginary ICER model the difference in lifetime I-QALYs will be minimal. Price and lifetime cost differences will drive incremental cost-per-I-QALY claims with the inevitable match against thresholds yielding substantial price discounts. The advent of ICER Analytics gives manufacturers the opportunity to challenge this with 'imaginary' utilities that better reflect patient and caregiver concerns in a competing imaginary simulation.

THE QALY DISQUIET

There have been numerous papers pointing to what are perceived as I-QALY limitations yet none of them raise the issue of fundamental measurement. A common critique is that I-QALY calculations inherently privilege treatments that extend the lives of those who can be restored to perfect health, and disadvantage the many who seek life-extending treatments despite having a disability or chronic condition that is not curable; reflecting earlier concerns over Oregon Medicaid cost-saving proposals with the I-QALY as a rationing tool^{19 20}. Indeed, one of the most often cited critiques of the I-QALY is remarkable for either being unaware or ignoring the limitations of fundamental measurement²¹. Other claims have examined the implications of adopting the imaginary ICER pricing thresholds. As an example a 5 state Medicaid-based evaluation found that if ICER recommendations were followed to reduce expenditures a significant number of patients would lose access to physician determined therapies in multiple sclerosis, rheumatoid arthritis, non-small cell lung cancer, multiple myeloma and psoriasis²². Again, the limitations due to fundamental measurement standards were not raised. This is, in fact, common. The California Health Care Foundation (CHCF) in a review of attempts to address the problem, as it saw it, of escalating prescription drug costs, again endorsed the I-QALY as a legitimate construct, although in presumptive ignorance of the violation of the axioms of fundamental measurement²³.

In May 2020 Oklahoma, in response to criticisms that the I-QALY was discriminatory, enacted the *Nondiscrimination in Health Care Coverage Act (HB2587)* which required state agencies to confer with disability groups prior to making decisions on coverage, reimbursement and utilization management. The key provisions are:

An agency shall be prohibited from developing or employing a dollars-per-quality adjusted life year, or similar measure that discounts the value of a life because of an individual's disability, including age or chronic illness, as a threshold to establish what type of health care is cost effective or recommended.

An agency shall be prohibited from utilizing such adjusted life year, or similar measure, as a threshold to determine coverage, reimbursement, incentive programs or utilization management decisions, whether it comes from within the agency or from any third party.

If the Oklahoma legislature members had had any understanding of the axioms of fundamental measurement, it should have been obvious that the I-QALY was automatically disqualified in any submission for product pricing and access.

At the same time the Massachusetts Health Policy Commission (HPC) which finalized its Drug Pricing Review in early 2020 to establish regulations to support evaluations of drug prices, proposes not only to review international reference pricing for selected drugs but also to ask manufacturers to provide market analyses, economic models, examination of similar drugs, cost-effectiveness analyses and comparative effectiveness analyses. These would include assessments, including reference prices, contracted to groups such as ICER and PORTAL (Research Program on Regulation, Therapeutics and Law), the latter affiliated to the Harvard Medical School. Again, it is not clear at this stage whether the HPC is aware of the absurd nature of the I-QALY modeling and the mathematical impossibility of any claims for pricing and product access.

It is worth noting that, primarily for disability concerns, the 2010 Patient Protection and Affordable Care Act (ACA) in establishing a Patient-Centered Outcomes Research Institute (PCORI) to conduct comparative effectiveness research explicitly forbids the Institute from developing or using cost-per-QALY as a 'threshold' in developing a Medicare formulary. It seems clear from the ACA that there was no awareness of the mathematically impossible nature of the I-QALY.

This does not mean that Federal agencies are not advocates of I-QALYs. The Centers for Disease Control and Prevention (CDC) for example, has a website devoted to their Diabetes State Burden Toolkit with its nonsensical estimates of annual I-QALYs lost to diabetes²⁴. This lack of awareness of the I-QALY as an impossible mathematical construct extends to the use of the US National Health Interview Survey (NHIS) and the Behavioral Risk Factor Surveillance System (BRFSS)²⁵.

What is remarkable is that, without exception, these criticisms of I-QALYs take the I-QALY construct at face value. No thought appears to have been given to the fact, not just that the I-QALY fails the axioms of fundamental measurement as a

mathematically impossible construct but that, in respect of the approximate information modelling, none apparently have an understanding of the standards of normal science: the proposition that claims must be credible, empirically evaluable and replicable is absent. It's a sad commentary on the forensic standards that state Medicaid agencies and others employ to judge the value of competing therapies. Certainly discrimination is an issue; but behind that is the fact the I-QALY is an impossible construct. That alone should be sufficient for its abandonment.

QALYS, ICER AND THE VETERANS ADMINISTRATION

The VA represents an interesting case study of the involvement of ICER in public sector formulary decision making. As background, we can note that VA along with the Department of Defense, the Public Health Service and the Coast Guard benefit from a separate program that caps the price of branded drugs. The VA is of particular interest as it is allowed to negotiate prices better than the lowest Medicaid best price. This is achieved through confidential negotiations with manufacturers by the VA to list drugs on its nationally preferred formulary by the VA Pharmacy Benefits Management Services (PBMS). For the past 3 years, however, the VA PBMS has contracted with ICER to provide clinical evaluations and imaginary modeled claims for cost-effectiveness. These are seen, in the latter case, as services to provide supplemental evidence. As a recent blog in *Health Affairs* informs us: *....missing from these information resources (pharmaceuticals) was a reliable source of non-biased cost-effectiveness analysis The opportunity to contract with the Institute for Clinical and Economic Review (ICER) was seen ... As a means to better contextualize value for decision making on behalf of veteran patients and US taxpayers* ²⁶.

While 'better contextualize' is an odd turn of phrase, it can be interpreted as ICER creating claims for cost-effectiveness, evidence by assumption, based on a lifetime incremental cost-per-I-QALY framework providing imaginary threshold 'fair' prices and access recommendations to support VA pricing negotiations and possible denial of access to new therapies. It is now three years from the start of the partnership, yet in that time no one has apparently seen fit to consider the weaknesses, indeed the fatal flaws, in the ICER value assessment case. This does not reflect well on the VA PBMS's forensic skills in evaluating technology assessment claims. The VA PBMS have apparently, taken ICER recommendations at face value without a more extensive assessment of the merits of this modeling framework, the denial of the standards of fundamental measurement together with ignorance of potential benefits and harms for VA patients.

ICER AND NEW YORK STATE MEDICAID

While the phrase 'ICER led formularies' is possibly an exaggeration, there are a number of state Medicaid formulary committees that apparently utilize imaginary ICER models to inform their decision making. Consider, as a case study, the

New York State Medicaid battle with Vertex Pharmaceuticals for pricing (or at least claiming a substantial kick-back) with Orkambi for cystic fibrosis. The saga has a passing interest, not least for the lack of awareness by the parties of the constraints imposed by the axioms of fundamental evidence; a debate over I-QALYS.

The Vertex imaginary pricing saga begins with the 2018 ICER report on cystic fibrosis ²⁷. The imaginary modelled claim presented by ICER and developed by the University of Minnesota School of Public Health, makes the assumption that the utility scores for the EQ-5D-3L have ratio properties; no proof is presented. The first step, in violation of measurement axioms, is to apply a linear function to create utilities corresponding to ppFEV₁ (percent predicted forced expiratory volume in 1 second) ranges. The second step, again in violation of measurement axioms, is to create the cystic fibrosis I-QALYS for time in a disease state. The third step is to add these I-QALYS over the lifetime of the hypothetical patient; this is not a violation of axioms, merely impossible. The I-QALYS are then discounted (together with discounted lifetime assumed direct medical costs) to create the mathematically impossible cost-per-incremental I-QALY claims. Anyone with an understanding of the axioms of fundamental measurement should, at this point, pause and reject the modeling.

Yet the model builders at the University of Minnesota proceed. To illustrate the absurdity of their creation consider the modeled imaginary and non-evaluable cost and I-QALY calculations (they should not be described as estimates; just imaginary numbers) for Orkambi for individuals with the homozygous F508del mutation. Hypothetical lifetime total costs (discounted) for best supportive care are \$2,108,199* (no drug costs) compared to Orkambi with best supportive care at \$6,983,336. While the degree of precision with these estimates may convince the less credulous that there is some merit to the calculation, they should not be deceived; after all we are modelling decades into the future. Corresponding to these imaginary costs are the equally imaginary I-QALYS: 14.74 I-QALYS in the case of best supportive care and 20.21 I-QALYS for Orkambi. Cost per I-QALY gained is \$890,739. This entirely imaginary and mathematically impossible calculation is then compared to the equally impossible cost-per-I-QALY ICER thresholds. Probabilistic sensitivity analysis is then applied to demonstrate that at no threshold within the range \$50,000 to \$500,000 can Orkambi be judged to be cost-effective. This is an entirely misleading exercise which raises once gain questions of the extent to which the audience is aware of the meaning and assumptions behind probabilistic sensitivity analysis, a technique that was developed solely to support imaginary claims for the likelihood of a product being cost-effective with lifetime simulation models. Forensic skills in evaluating health technology assessment claims seem conspicuously absent.

The analysis then proceeds to the question of pricing. This is presented, again in imaginary terms, to make the case for

pricing discounts for Orkambi. In annual price terms, the ICER estimated annual net price is \$264,000. The annual net price required to meet the various ICER imaginary thresholds range from \$55,652 for a \$50,000 threshold to \$165,824 for the \$500,000 threshold. While these results are clearly nonsensical, they provide media space for ICER to argue for substantial price discounting for Orkambi in the range proposed by the University of Minnesota model of 71% to 75%.

These modeled imaginary results were taken on-board by the New York Medicaid Drug Utilization Review Board. Drug utilization review boards are federally mandated for the states. In 2017 New York supplemented its DUR activities by establishing a drug spending cap in the Medicaid program. This gave officials the power systematically to limit the price of high-cost prescription drugs with the Board to determine a value-based price for the drug. Following this, negotiations with the manufacturer take place to set supplemental rebates. This is somewhat surreal as the parties are negotiating over imaginary claims where the notion of value is meaningless.

At the DUR Board's meeting on 26 April 2018, it was announced that in the case of Orkambi it was recommended that the unit price to achieve \$150,000 per I-QALY threshold the supplemental rebate target amount was the value resulting in a unit price equal to \$56.94 (net of all rebates). The value assessment was provided by ICER together with an imaginary threshold price analysis based on a range of QALYs modeled for the drug. Without wishing to impugn the skills and decisions of the Board the fact remains that the Board and its support staff were clearly unaware of the lack of scientific merit in ICER-type modelling. A decision involving millions of dollars was based entirely on a value assessment fantasy construct. They might care to revisit this decision with the ICER Analytics platform and come to a number of different imaginary conclusions. This could involve a pricing and rebate renegotiation.

It is worth noting that while the I-QALY is a mathematical fantasy, it is not a measure of the patient's perception of their quality of life) but rather the community's 'valuation' of the patient's HRQoL given the five symptoms defined, for example, by the EQ-5D-3L multiattribute instrument and the 3-level responses allowed by the instrument. The symptoms may have no apparent relation to the patients (and caregiver) needs nor do the five symptoms necessarily have any relevance to the experience of, say, cystic fibrosis. The typical I-QALY defense against patient-centric arguments is that the utility score is a generic measure needed to make comparisons between disease states. While the notion of the I-QALY as a Soviet-style central planning tool for resource allocation, to maximize a societal 'benefit' from a fixed health care budget, may have an appeal to those with a left-of-center persuasion, the discussion is fruitless as the I-QALY lacks mathematical credibility.

If they had known, Vertex could have stepped back and argued that a fair price benchmark built on an I-QALY simulated

lifetime model was an exercise in futility as the model not only defied the standards of normal science but that its cost-per-I-QALY 'fair price' is a construct that is mathematically impossible; pointing out that the notion of a 'fair value-based drug price' is a chimera.

BEYOND ICER PSEUDOSCIENCE

ICER's lack of awareness of the standards of normal science, including fundamental measurement should come as no surprise; it is shared by the majority of those in health technology assessment. This is seen in the belief in the ratio properties of ordinal utility scales; few are aware of the axioms of fundamental measurement.

ICER subcontracts the reference case modeling to university based modelling groups. These groups show the required expertise in constructing cost per I-QALY simulation models but a singular lack of appreciation of the standards of normal science, notably in respect of their failure to understand the axioms of fundamental measurement. The fact that the I-QALY is an impossible mathematical construct is a totally foreign notion; they cling tenaciously to the belief that ordinal utility measures are actually ratio measures in disguise. The implications of producing claims that lack credibility, fail to be empirically evaluable and replicable in real world treating environments, are not recognized.

At the same time there must be concern over the lack of technical expertise, forensic skills, in the audience for ICER's pricing and access recommendations. There are a number of issues to be resolved. First, the general lack of awareness of the standards of normal science; second, as part of this, the lack of awareness of the axioms of fundamental measurement; and third a lack of understanding of the credibility limitations implicit in the construction of lifetime simulation models. Absent an appreciation of any of these issues means that ICER's audience is prepared to take ICER's word for it. Certainly, there will be expertise in trial design, clinical outcomes and indirect comparisons as the audience, at least the membership the various CEPACs will include physicians, but even then there will be few physicians who will have been made aware of instrument development and measurement standards. This is exemplified in the plethora of physician authored PRO instruments, the overwhelming majority of which fail the standards for of fundamental measurement in assessing response to therapy²⁸. But these physicians are not alone, CVS in its adoption of ICER thresholds to determine formulary access also shows a singular lack of awareness of the standards of normal science^{29 30}.

While it might appear presumptuous to state the obvious, a significant part of ICER's audience's willingness to take the modeled claims at face value is because of a lack of understanding of the standards of normal science. ICER succeeds through ignorance; or at least the unwillingness of those in health technology assessment with the appropriate

skills to challenge ICER. The meme that ICER subscribes to is well entrenched. It is unusual, for example, to find in listening to the ICER review of the final evidence report any attempt to question the reference case modelling exercise. This should come as no surprise. After 30 years, with the high transmission fidelity of the I-QALY lifetime simulation meme, there should be no illusions as to the resilience, possibly for a number of years, of the I-QALY meme. Too many have too much to lose. After all, ICER's business case rests on this imaginary framework. For subscribers to this meme, truth is consensus.

Put simply: ICER has failed (or succeeded in its own fantasy terms) because of a decision to reject the standards of normal science and accept the prevailing I-QALY lifetime imaginary simulation meme. ICER is perfectly aware of this, but for years has decided to gloss over it. ICER is not alone; single payer health systems with formulary gatekeepers have fallen into the same trap. These include ICER's 'mentor' the National Institute for Health and Clinical Excellence (NICE) in the UK. ICER's failure to provide credible claims for pharmaceutical product pricing and access was guaranteed from the beginning. ICER chose to create approximate information to support its claims rather than the application of the standards of normal science to meet evidence gaps and provide formulary committees with robust and testable claims for product performance³¹. The result was inevitable. None of the evidence reports for product pricing and access published over the past decade by ICER have any claim to credibility. They are mathematically impossible and entirely imaginary, although in a contest of imaginary constructs there is the opportunity to be challenged through the ICER Analytics cloud platform. Similarly ICER recommendations for product pricing and formulary placement to state Medicaid groups and agencies such as the VA also fail the required standards. The notion of an ICER 'led' formulary has no credible basis. Formulary decisions should not be based on one-off imaginary claims. If claims are presented they should be seen in an evidence framework that responds to changing treating environments and supports ongoing disease area and therapeutic class reviews.

IMPLICATIONS FOR STATE MEDICAID

ICERs modelling and recommendations have no place in state Medicaid formulary deliberations. Relying on invented evidence is not acceptable and easily challenged. This standard would not be accepted for clinical claims. There are other options; ones that meet the standards of normal science and ensure a robust and coherent evidence base for product evaluations. It is up to the individual state Medicaid assessors to put in place formulary guidelines that abandon the easy way out of accepting I-QALY simulations while committing to meeting the standards of normal science. As a starting place, consideration could be given to the value assessment framework given in Version 3.0 of the Minnesota formulary submission guidelines³². These guidelines are built on three premises: (i) that all claims should conform to the standards of normal science with credible claims for single product

attributes; (ii) that all claims should accompanied by a protocol detailing how they are to be assessed and reported; and (iii) that all pricing is provisional, subject to value or outcomes contracting. Clearly, none of this is new; it simply brings together value assessment from real world evidence that meets the standards of normal science, including fundamental measurement.

At a more general level ICER (or other I-QALY model builders) may try to defend their role as reflective of accepted 'standards' but this does not insure against a reasoned critique. Patient advocacy groups are well equipped to mount such a challenge. Indeed the challenge may encompass not just a critique of ICER's reference case methodology but a description of an alternative real world evidence paradigm to meet evidence gaps and support ongoing disease area and therapeutic class reviews. If state Medicaid is to engage with manufacturers then the first step must be to develop formulary submission guidelines that are consistent with the standards of normal science. At this juncture it might be appropriate to recall the motto of the Royal Society of London (founded 1660): *nullius in verba* (take nobody's word for it)³³. Unfortunately, we take ICER's word for it. The lack of critical or forensic skills among recipients is undeniable.

CONCLUSIONS

Despite ample warnings, the belief in the I-QALY has endured. It is ironic that this belief rests on a mystical faith in the 'modified' properties of ratio scales. Reaching its apotheosis in the US with ICER's claim that all is well: we know that the utility scales, while may appear to only have ordinal measurement properties, are actually ratio scales in disguise. No doubt this alternative fact will be accepted by many analysts; after all it is a comfort zone which puts to one side hard questions regarding neglect of the standards of normal science in measurement theory for the past 30 years. Even so, ICER may believe in ratio utility scales but in defending this belief, apart from saying ICER 'understands' that the EQ-5D-3L utilities are on a ratio scale, ICER's attempts to put this belief in comprehensible terms involves faith in: (i) ordinal scales that are actually interval scales; (ii) an assumed interval scale (the EQ-5D-3L) that has the ratio properties of multiplication and division; (iii) ratio properties that are not necessary for the estimation of utility for use in producing I-QALY estimates; (iv) that to create I-QALYs all you need is an interval scale without a true zero; (v) that the EQ-5D-3L needs only to have interval properties to produce I-QALYs without any consideration of ratio scale properties; and (vi) belief in dimensionally heterogeneous multiattribute utility scores that lack construct validity and interval properties that actually have dimensional homogeneity^{34 35}. This is an amazingly complex and contradictory belief system. If true, it represents the most significant advance in measurement theory in the past 60 years; a ratio scale without a true zero.

The irony becomes even more apparent that with the recognition that if the I-QALY is an impossible mathematical construct this would have effectively killed the IQALY some 30 years ago. Rather than having to convince Federal and state authorities to ban the application of I-QALYs in certain circumstances, it would have been far easier to have pointed out its failure to meet the axioms of fundamental measurement. But, unfortunately, few were aware of these axioms and the limitations of ordinal utility scales. The result is an ongoing debate between I-QALY advocates such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and ICER and groups which have advocated alternative value assessment patient-centric frameworks such as the Partnership to Improve Patient Care. Despite a continuing opposition to the I-QALY the fundamental point has been overlooked: the irrelevant I-QALY construct³⁶. Even so, these alternative frameworks are only relevant if they conform to the standards of normal science, to include fundamental measurement. If not, their advocacy is a waste of time.

As the literature of the past 30 years has demonstrated, there have been well articulated concerns with the I-QALY. To date, the I-QALY and its supporters have largely withstood these criticisms. In retrospect, this has been a wasted effort. An understanding of the axioms of fundamental measurement is critical, with the basic premise that unless a measure or an instrument is designed to have certain properties then we

cannot assume, *ex post facto*, that it has those properties. As the physical sciences have demonstrated any attribute measure must be designed to have either interval or ratio properties. This is where utility scores fail. There is no evidence to suggest that the limitations of fundamental measurement entered their developmental considerations, including the standard for dimensional homogeneity.

Recognition and acceptance that utility scores are nothing more than ordinal measures effectively resolves any question of retaining the I-QALY as part of a modeling paradigm. Rather than debating the 'merits' of the I-QALY in particular situations, we can simply dispose of the I-QALY as a mathematically legitimate construct; a delusional (*ignis fatuus*) or will-o'-the-wisp fantasy. The I-QALY has no legitimacy. Debates over its merits and applications are easily disposed of; it is an illegitimate construct that fails the standards of normal science and should be discarded from any health technology lexicon. We can cut the Gordian I-QALY knot in favor of value assessment processes that meet the standards of normal science.

Conflict of Interest: PCL is an Advisory Board member and consultant to the Patient Access and Affordability Project, a program of Patients Rising

References

- ¹ Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. *Arch Phys Med Rehabil*. 1989. 70:308-312
- ² Grimby G, Tennant A, Testo L. The use of raw scores from ordinal scales: Time to end malpractice (Editorial) *J Rehab Med*. 2012;44:97-8
- ³ McKenna S, Heaney A, Wilburn J et al. Measurement of patient-reported outcomes. 1: The search for the holy grail. *J Med Econ*. 2019;22(6): 516-22
- ⁴ McKenna SP, Heaney A, Wilburn J. Measurement of patient reported outcomes. 2: Are current measures failing us? *J Med Econ*. 2019;22(6):523-30
- ⁵ <https://icer-review.org/blog/how-independent-assessment-of-drug-value-can-help-states/>
- ⁶ Piglucci M. Nonsense on Stilts: How to tell science from bunk. Chicago: University of Chicago Press, 2010
- ⁷ Neumann PJ, Willke R, Garrison LP. A Health Economics Approach to US Value Assessment Frameworks – Introduction: An ISPOR Special Task Force Report. *Value Health*. 2018;21:119-123
- ⁸ 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>
- ⁹ Wootton D. The Invention of Science: A New History of the Scientific Revolution. New York: Harper Collins, 2015

- ¹⁰ Langley PC, McKenna SP. Measurement, modeling and QALYs [version 1; peer reviewed] *F1000Research* . 2020; 9:1048 [\https://doi.org/10.12688/f1000research.25039.1](https://doi.org/10.12688/f1000research.25039.1)
- ¹¹ Stevens S. On the theory of scales of measurement. *Science*. 1946;103:677-680
- ¹² Bond T, Fox C. Applying the Rasch Model. New York: Routledge, 2015
- ¹³ Langley PC. Sunlit uplands: The genius of the NICE reference case. *Inov Pharm*. 2016;7(2): No.12. <https://pubs.lib.umn.edu/index.php/innovations/article/view/435/430>
- ¹⁴ Institute for Clinical and Economic Review. ICER Analytics Launched to Accelerate the Real World Application of ICER Reports, Helping Payers, the Life Science Industry and Health Care Policy Makers move from Insight to Action. Media Release. 1 December 2020. <https://icer-review.org/announcements/icer-analytics-launch>
- ¹⁵ Langley PC. Let a Thousand Models Bloom: ICER Analytics Opens the Floodgates to Cloud Pseudoscience. *InovPharm*. 2021; 12(1): No.5 <https://pubs.lib.umn.edu/index.php/innovations/article/view/3606/2668>
- ¹⁶ Langley PC. The Impossible QALY and the Denial of Fundamental Measurement: Rejecting the University of Washington value assessment of targeted immune modulators (TIMS) in ulcerative colitis for the Institute for Clinical and Economic Review (ICER). *InovPharm*. 2020;11(3). No. 17
- ¹⁷ Langley PC. To Dream the Impossible Dream: The commitment by the Institute for Clinical and Economic Review to rewrite the axioms of fundamental measurement for Hemophilia A and Bladder Cancer value claims. *InovPharm*. 2020; 11(4); No. 22 <https://pubs.lib.umn.edu/index.php/innovations/article/view/3542/2613>
- ¹⁸ 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
- ¹⁹ Smith WS. Key Questions for Legislators about the Institute for Clinical and Economic Review. Pioneer Institute. White Paper No. 189. January 2019
- ²⁰ McKenna J, Smith W. The Legality of QALY under the ADA. Pioneer Legal Public Comment. February 2020
- ²¹ Pettit D, Raza S, Naughton B et al. The Limitations of QALY: A literature review. *J Stem Cell Res Ther*.2016;6:4
- ²² Xcenda. State-based analysis: ICER-based formularies could upend access to current therapies in Medicaid. October 2019 <https://www.xcenda.com/insights/impact-to-patients-of-icer-formulary-implementation-in-medicaid>
- ²³ California Health Care Foundation. When the Price is not Right. ; State options on prescription drug pricing. 2017 <https://www.chcf.org/wp-content/uploads/2017/12/PDF-WhenStateRxPricing.pdf>
- ²⁴ Centers for Disease Control and Prevention. Diabetes State Burden Toolkit. <https://nccd.cdc.gov/Toolkit/DiabetesBurden/YLL/QALY>
- ²⁵ Lubetkin E, Jia H. Health-related quality of life, quality adjusted life years and quality adjusted life expectancy in New York City from 1995 to 2006. *J Urban Health: Bulletin New York Academy Medicine*, 2009;86(4):551-61
- ²⁶ Health Affairs Blog. VA and ICER at three years: Critics' concerns answered. June 15, 2020
- ²⁷ Institute for Clinical and Economic Review. Modulator Treatments for Cystic Fibrosis: Effectiveness and Value. Evidence Report and Meeting Summary. June 7, 2018. https://icer-review.org/wp-content/uploads/2017/10/CF_Final_Evidence_Report_06082018.pdf
- ²⁸ McKenna S, Heaney A. Composite outcome measurement in clinical research: the triumph of illusion over reality. *J Med Econ*. 2020; 23(10):1196-1204

²⁹ CVS Health. Current and new approaches to making drugs more affordable. August 2018.

³⁰ Langley PC. CVS Health and the imaginary worlds of the Institute for Clinical and Economic Review (ICER). *Inov Pharm.* 2018; 9(4):No. 4 <https://pubs.lib.umn.edu/index.php/innovations/article/view/1461/1344>

³¹ Langley PC. 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>

³² Langley P. Value Assessment, Real World Evidence and Fundamental Measurement: Version 3.0 of the Minnesota Formulary Submission Guidelines. *InovPharm.* 2020;11(4): No 12 <https://pubs.lib.umn.edu/index.php/innovations/article/view/3542/2613>

³³ Royal Society of London <https://royalsociety.org/about-us/history/>

³⁴ Institute for Clinical and Economic Review. Valoctocogene Roxaparvovec and Emicizumab for Hemophilia A without Inhibitors: Effectiveness and Value Response to Public Comments on Draft Evidence Report October 16, 2020 https://icer-review.org/wp-content/uploads/2019/12/ICER_Hemophilia-A_Public-Comment-Responses_101620.pdf

³⁵ Institute for Clinical and Economic Review. Nadofaragene Firadenovec and Oportuzumab Monatox for BCG-Unresponsive, Non-Muscle Invasive Bladder Cancer: Effectiveness and Value Response to Public Comments on Draft Evidence Report November 6, 2020 https://icer-review.org/wp-content/uploads/2020/02/Bladder-Cancer_Public-Comment-Response_110620.pdf

³⁶ Partnership to Improve Patient Care. Measuring Value in Medicine: Uses and Misuses of the QALY. Washington DC. 2017. <http://www.pipcpatients.org/resources/white-paper-uses-and-misuses-of-the-qaly-ethical-issues-and-alternative-measures-of-value>