MAIMON WORKING PAPERS No. 12 AUGUST 2021

WHERE IGNORANCE IS BLISS: THE INSTITUTE FOR CLINICAL AND ECONOMIC REVIEW (ICER), THE UNIVERSITY OF WASHINGTON CHOICE INSTITUTE AND ATOPIC DERMATITIS SIMULATION PRICING AND ACCESS RECOMMENDATIONS

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

Abstract

It has been demonstrated conclusively that value and utility preference scores have only ordinal properties. This means, as has been pointed out on numerous occasions that the quality adjusted life year (QALY) is a mathematically impossible construct. The implications are profound: some 30 years of health technology assessment is rendered worthless due to a failure to recognize the well documented limitations imposed by the axioms of fundamental measure. Yet these are ignored in favor of persevering with the QALY as though it has mystical measurement properties, a ratio scale in disguise that allows QALYs to be imagined. This denial of the relevance of these axioms by the Institute for Clinical and Economic Review (ICER) has been the long-standing position and in the case of the just published ICER final evidence report on JAK Inhibitors and monoclonal antibodies (released on 17 August, 2021), shared by its contracted model builders at the College of Pharmacy, CHOICE Institute at the University of Washington, Seattle. The fact that an academic center should support ICER's denial of fundamental measurement and the construction of evidence should come as no surprise as professional groups such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) is equally complicit in this denial. Interestingly, this denial appears to extend to ignoring peer reviewed studies that have considered alternative frameworks for evaluating quality of life in atopic dermatitis. This applies to the Quality of Life Index of Atopic Dermatitis (QoLIAD) score and its correlation with efficacy outcomes in atopic dermatitis. The purpose of this commentary is to address once again ICER's (and the CHOICE Institute) failure to understand the difference between science and pseudoscience, irrespective of numerous critiques of the denial and misuse of preference scales. The creation of mathematically impossible simulations continues; driven by assumption without credible and evaluable claims for therapy response. ICER is operating in an analytical dead end; a basis for creating pricing and access recommendations that should be rejected out of hand. Yet, ICER perseveres. Truly, ignorance is bliss.

INTRODUCTION

The Institute for Clinical and Economic Review's (ICER) denial of the standards of normal science and embrace of pseudoscience is well documented [1] . The latest ICER report on atopic dermatitis is firmly in this long-held tradition of inventing imaginary evidence to support product claims [2]. Resting on the imaginary shoulders of the QALY, ICER lives to create, typically through academic centers that should know better, lifetime assumption driven simulation models to support its self-appointed mission to invent social pricing and denial of care for products entering the US market [3]. This is a futile

endeavor, a complete waste of time, for reasons that are well established [4]. Its genesis can be traced back to the early 1990s when in order to save time and make claims for cost effectiveness for formulary submissions, pharmacists and others in health technology assessment decided that if evidence was not available at product launch it should be invented [5]. Hypothesis testing was rejected (too time consuming) in favor of inventing evidence (the euphemistic phrase was 'approximate information). Although it seems odd that you can create lifetime modeled approximate information when there is no reference point to judge the worth of 'approximate'; approximate to what? Based on assumptions about the future all one had to do was change the assumptions, even reverse engineer, and create a competing set of claims for social pricing and access. Nevertheless this was embraced by the health technology assessment profession as evidenced by the premier textbook for inventing imaginary cost-effectiveness claims [6] .

The attractiveness of the approximate imaginary information belief system (or meme) is undeniable. It is not often that claims are made that can never be empirically evaluated or replicated. Indeed, in the UK, where the National Institute for Health and Care Excellence (NICE) is the ICER lodestar, there are academic institutions where forensic skills developed over years are employed to assess the validity of the imaginary cost-per-QALY models presented by manufacturers, proposing imaginary amendments or even an alternative imaginary reality. While ICER does not expose its models to this level of imaginative inquiry, the bottom line is that ICER is in a win-win situation: recommendations are made that are incapable of empirical evaluation. We don't know if ICER is right, or wrong; we will never know and we were never intended to know. All we can offer, although a waste of time, is to change assumptions including the model structure and parameter values and come up with competing cost-per-QALY claims; none of which, in turn, will be empirically evaluable. This is facilitated through the *ICERAnalytics* choose your assumptions platform for selected product models [7].

IMAGINARY ICER RECOMMENDATIONS

The ICER report considers six atopic dermatitis therapies: abrocitinib (Pfizer); tralokinumab (LEO Pharma); baricitinib (Olumiant®, Eli Lilly and Incyte); upadacitinib (Rinvoq®, AbbVie); ruxolitinib (Incyte); and dupilumab (Dupixent®, Regeneron and Sanofi); the objective to propose pricing recommendation and claims for budget impact. Both sets of claims are imaginary, driven by an assumption fueled five year simulation model. The centerpiece to this imaginary presentation is the ICER devised imaginary health-benefit price benchmark (HBPB). This, according to ICER is the highest imaginary price a manufacturer should charge for a treatment. The eugenic implications for access to and denial of care are clear: [8] this highest price is *based on the amount of improvement in overall health (defined by the preference score attributes) patients receive from that treatment, when a higher price would cause disproportionately greater losses in health among other patients in the health system due to rising overall costs of health care and health insurance. In short, it is the top price range at which a health system can reward innovation and better health for patients without doing more harm than good.* The fatal flaw is that the entire exercise is based on a failure to recognize the standards of normal science, notably the axioms of fundamental measurement, and a naive belief that the imaginary QALY can support health care allocation decisions. Health care resource allocation cannot be based on imaginary constructs. The HBPB is meaningless, implying as it does that some health states, from a community preference for health attributes perspective, are more 'worthy' of support than others. Just as eugenic criteria were pseudoscience, so the ICER HBPB criteria are equally pseudoscience.

While not to be taken seriously, ICER's recommended HBPB ranges as follows abrocitinib, $30,600-$41,800 per year; tralokinumab, $25,700-$35,000 per year; baricitinib, $24,400-$33,300 per year (which would require a 0-16% discount with current US list price of $29,000), upadacitinib, $30,400-$41,500 per year (which would require a 35-53% discount off the treatment's current US list price of $64,300) and dupilumab, $29,000-$39,500 per year, which would require a 6-31% discount off the treatment's current US list price of $41,800).

The attraction of applying generic preference scores to support pricing and access recommendations are that the instrument is less than sensitive to therapy impact due to the limited range of symptoms covered, which may be of little relevance to the disease area and target patient. This seen clearly in the ICER report where imaginary QALY equivalents over the five year time horizon range from 2.98 QALYs for standard of care (topicals) to 3.59 for Abrocitinib (the range for all comparators is 3.23 to 3.59 QALYs). Incremental QALY gains over the standard of care (topicals) range from 0.26 in the case of Baricitinib to 0.61 for Abrocitinib, with even smaller increments for Baricitinib and Upadacitinib compared to Dupilumab at 0.12 and 0.03 QALYs respectively. Translating these incremental QALYs into time gains in a five year time horizon, comparing Abrocitinib to standard of care gives incremental 0.61 QALYs or 7.32 months while if Dupilumab is the comparator the gain is 0.12 QALYs or 1.44 months (42.3 days) over five years. Incremental cost-per-QALY claims and the application are driven almost entirely by hypothetical costs. ICER's case is unsustainable; the QALY (or imaginary I-QALY) is a mathematically impossible construct [9]. None of these claims are empirically evaluable and, based on ordinal scores, entirely imaginary. The distance between scores such as preference and QALY estimates are unknown. The rule of thumb is: if you want to minimize imaginary therapeutic gains expressed as QALYs then use an ordinal generic preference score. Needless to say the choice of competing preference score with alternative manipulations will produce different results. The ICER results, as detailed below, say nothing about whether the need of patients is addressed; the model is driven by community preferences for a bundle of clinical symptoms and response levels defining a generic health state which may have little to do with health states relevant to atopic dermatitis populations. The notion of perfect health (preference score equal unity) is entirely contrived.

Claims for the future based on claims from the past are logical nonsense (otherwise known as the problem of induction). But this is a mainstay of ICER modelling. The pervading assumption is the ability to use past claims to support future assumptions, forgetting that claims from the past cannot support claims on the future. Unfortunately, all too many decision makers take ICER's assumptions and recommendations as if they were holy writ; whether this is just a negotiating tactic or a more concerning failure to appreciate the standards of normal science (which underpin drug development) is an open question. The downside, of course, is that patients and caregivers can be denied therapy. A recent commentary has described this as eugenics by the back door: if preferences are based, as they are, on the views of a community sample on the value of health states then we face the issue of 'worth' in the allocation of health care. We can use these preferences and ICER's modeling of QALY increments and costs to refuse care to the 'less worthy', restricting it to the more 'worthy'. More pernicious, is the presence of negative preference scores, or 'state worse than dearth'. The eugenic association is obvious, but more to the point is that if there are negative scores bundled with positive ordinal scores to generate an average preference score (which is mathematically impossible) then for these health state cost-per-QALY estimates will be inflated with smaller average preference scores but the same costs, and claims for price discounting and access more disadvantageous for that target patient population.

IMPOSSIBLE REPLICATION

Any attempt to replicate the ICER model simulation is virtually impossible given the lack of supporting information. Consider the preferences (utilities) employed in the model. Clarification on your use of preference scores required more information than that provided in the draft evidence report. Unfortunately, we have no idea as to what these scores actually are for mild, moderate and severe stages of AD. They are blanked out. All we have is the Delphic utterance from the internationally respected CHOICE expert group that in constructing their imaginary assumption driven claims for the pricing and recommendations for atopic dermatitis therapies were '*weighted by a single set of health state utility values from pooled manufacturer data to derive quality-adjusted life-years (QALYs)'.* Seeking further clarification on these utility scores the process is described by the University of Washington expert group as follows:

> ***We derived health state utilities for the non-responder and responder states by pooling utility estimates from manufacturer submitted data. We estimated weighted average utility values for each health state, combining estimates from all treatments with data available by health state. We considered therapy-specific health state utility values to capture benefit beyond EASI score, however the available evidence did not support differential utility scores by treatment (p. 42).***

No further details are given; all that is provided is a list of atopic dermatitis trials. This is unfortunate because if the protocols for the various AD trials are reviewed (Clinicaltrials.gov: ECZTRA 1&2; MEASURE UP 1 & 2; AD UP; and SOLO 1&2) there is no evidence from the list of secondary outcomes for each of these of any health related quality of life or just quality of life instrument that is designed to generate either direct or indirect preference scores. At best, we have the ordinal Dermatology Life Quality Index (DLQI) in two trials (ECZTRA 1 & 2 and SOLO 1 &2) which simply provides an aggregate of 10 4-level Likert scales (scores 0 – 30). Other than that we have no idea how the University of Washington CHOICE Group of experts then proceeded to create utility values for a ratio scale with a true zero and a range of 0 = death to 1 = perfect health. ICER was asked to clarify but no reply was received. We must presume, as these are all secondary endpoints for the various protocols that they were all powered to create a 'composite' utility scale.

It might also be pointed out that if these various inputs from manufacturers are patient reported outcomes with ordinal properties, then the calculations vaguely described by the University of Washington CHOICE group are mathematically impossible (with a further concern that they lumped together utilities from different instruments). Ordinal scales can only support non-parametric assessments. It seems clear that the Washington group has little or no understanding of the limitations on statistical and econometric analysis imposed by the axioms of fundamental measurement (let alone building imaginary simulation models); if so, this is a major concern that ICER and the University of Washington should address. As a renowned university research group one would have thought their training would have included measurement theory and some elementary logic regarding assumptions to include Hume's Problem of Induction, the rationally unfounded premise that the future will resemble the past [David Hume, 1711 - 1776) [10].

Given this, it might be pointed out that in the previous ICER review of atopic dermatitis and subsequent imaginary modelling for Dupilumab in moderate to severe atopic dermatitis ICER provided EQ-5D-3L utility values (source Sanofi data on file) [11].For patients with moderate disease (IGA), the utilities ranged from 0.684 (baseline) to EASI 50 0.892, EASI 75 0.895 and EASI 90) 0.907 while for

severe disease (IGA4) the baseline was 0.536 to EASI 75 0.535, EASI 75 0.090 and EASI 90 0.911. Needless to say, the results presented failed to note that the EQ05D-3L has only ordinal properties which nullifies the analysis.

What is puzzling is that there are a range of preference scores for AD available from the literature. Although as might be expected, there appears to have been no attempt to undertake a systematic review of the QoL (HRQoL) literature in atopic dermatitis. Perhaps the expert group did not think a systematic review worthwhile? It is surprising because there are a number of reviews in the last 10 years that could have been updated. Even

HEALTH STATES WORSE THAN DEATH

The question of the distribution of health states defined by ordinal preference scores is not addressed by ICER and the University of Washington experts: the prevalence and distribution of health states worse than death is ignored. Again, this brings in the eugenics overtones which ICER apparently subscribes to in is belief in allocating resources by heath state and QALY. The fact that all generic preference instruments support negative preference scores is well established; they all subscribe to eugenic criteria. While often glossed over, the presence of negative scores has important implications. The most obvious is that if a preference algorithm can produce negative preference health state values then it cannot claim to have a true zero. It cannot support the standard arithmetic operations (e.g., multiplication) and cannot support QALYs; QALYs are mathematically impossible. A more important issue is how are we to interpret a negative preference score based on community sampling? Is health care to be withdrawn or denied?

As a first step, any application of ordinal preference scores should include a ranking of patients to indicate the proportion of patients with negative scores; attempting to claim QALY increments when a significant proportion of respondents report health states worse than death would give a misleading impression of benefits. In addition, there is the possibility that if an 'average preference score is presented to support QALY estimates, the presence of negative preferences could, as noted, 'deflate' the preferences and hence QALY values (although with ordinal scales averages cannot be computed; it's mathematically impossible). Are we to imply that the ICER preference averages include an allowance for withdrawal or denial of health care to these 'negative' souls?

Unfortunately, ICER and the modelling experts at the University of Washington are in no position to resolve this issue as they have no access to the distribution of preference scores that support the utility data points they abstract from the literature; authors typically will not report them. It is not as though negative health states are unlikely to occur. In the case of the latest US valuation of EQ-5D-5L health states (5 symptoms, 5 response levels) we find that of the 3125 possible health state values, 624 (20%) take negative scores (range -0.573 – 1.0) [12]. Indeed, one study using the EQ-5D-5L has reported negative preference scores in topic dermatitis ranging from -0.003 to -0.53 [13].

Should the sub-group of negative preference patients be separately identified in ICER's imaginary modelling? Are they considered less 'worthy'? Should ICER avoid deflating the preference scores due to the presence of 'states worse than death'? One approach would be to use, assuming ICER has preference distributions, non-parametric ranking comparisons as one criterion for evaluating benefits before and after therapy. This is most unlikely to occur. It would actually be a waste of time as the ordinal scores, based on a bundle of health states, lack dimensional homogeneity, unidimensionality and construct validity [14] , The preference scores themselves, although there are many different ones

to choose from, all lack the standards required to assess response to therapy: for one simple reason, they try of capture multiple attributes at one time rather than following the standards of the physical science and main line social science in focusing on one attribute at time. Responses can then be evaluated across a spectrum of required attributes.

MEASUREMENT BLINKERS

Without exception, studies claiming to evaluate the QoL or HRQoL in atopic dermatitis fail appreciate that all the disease specific instruments such as the Dermatology Life Quality index (DLQI) and the Children Dermatology Life Quality Index (CDLQI) that claim to capture aspects of quality of life in atopic dermatitis together with the generic preference instruments, including comparative studies with generic preference measures including the SF-6D, the EQ-5D-5L all fail to meet the required standards of fundamental measurement [15] [16] [17]. It is worth noting that although the question of fundamental measurement was not addressed, a review of classical measurement properties concluded that only the QoLIAD and DLQI merited further evaluation in atopic dermatitis [18]. The result is that, despite considerable attention given to QoL (and HRQoL) in the last 20 or more years, there are no acceptable measures of QoL in atopic dermatitis. This is not unusual in chronic disease states.

QoLIAD: AN INCONVENIENT TRUTH

One aspect of the current ICER report on atopic dermatitis is the apparent willingness of ICER (and CHOICE) to ignore published and peer reviewed studies that point to a response assessment that meets the standards of normal science. The respective material was cited in an evidence commentary to ICER, but failed to materialize in the final report. We have, in fact, a QoL measure for atopic dermatitis that meets the required measurement standards. This is the Quality of Life Index for Atopic Dermatitis (QoLIAD) [19]. It has been revised and used to create interval scores in atopic dermatitis trials, most recently Dupilumab in moderate to severe atopic dermatitis [20].

Given the focus on measurement theory, it should be noted that the QoLIAD instrument applies Rasch Measurement Theory to create interval response scores consistent with these requirements. The study found that compared to mean QoLIAD scores at baseline, Dupilumab significantly improved the QoLIAD score at 12 weeks of treatment against placebo. These scores were significantly correlated with changes in efficacy outcomes including EASI, 5-dimensionsl pruritis, pruritis NRS, total SCORAD and SCORAD VAS scores for sleep.

It is worth noting that a (weak) case for generic preference and the imaginary QALY is that comparisons can be made across disease states for resource allocation, to support supplementing and denying therapy in the eugenics tradition. In the case of the QoLIAD it has been shown to co-calibrate, using Rasch analysis, with other disease specific instruments across disease states. The specific example is the Psoriasis Quality of Life Scale (PSORIQoL) where common need fulfillment items provide the anchor [21] [22] [23]. Both the QoLIAD and PSORIQoL meet the required measurement standards.

CONCLUSIONS

Reviewing ICER responses to questions by third parties, not only in atopic dermatitis assessments, one is left with the feeling that if a truth or fact is inconvenient, ICER (and their consultants) just ignore it.

After all, only a handful of reviewer will actually read the report and assess the arguments against the ICER models. The standard ICER response to criticism is that the case for the I-QALY in imaginary assumption driven simulations to create imaginary recommendations for pricing and access is that everyone else does it, a shared mystical experience. There may be other mystical experience enjoyed in human populations, but there is no reason to adopt them. The belief is made quite clear that, for ICER (and presumably the experts at the University of Washington) the ordinal generic preference score, even with well documented negative values and lack of dimensional homogeneity, is truly a ratio measure in disguise; a mystery to which health technology assessment practitioners subscribe. Supported by the leaders in the field, the approximate information meme holds fast. The imaginary QALY stands firm, it provides a notionally robust yet imaginary measure, and the assumption driven simulation needs this acceptance of a memetic mystery.

ICER simulation modeling framework is not just an analytical dead end, but a framework for inventing non-evaluable evidence that should never have been attempted in the first place. Emulating agencies such as NICE is not a defense for producing pseudoscience. It is not a defense of claims for pricing and access to pharmaceuticals that are clearly nonsense. That measurement theory is a key standard for normal science is uncontroversial; it has been recognized for centuries. If ICER begs to differ in its embrace of preference scores with mystical ratio properties it should make clear that this is just a convenient belief to maintain its networking and its business model; a retreat to wishful thinking and belief in a mystery ratio scale with negative values. Belief in the credibility of ICER claims by all too many decision makers in health care is a belief, sad to say, that reflects a pervasive ignorance of standards for hypothesis testing and the discovery of new facts. ICER is a barrier to this endeavor to the detriment of patients, caregivers and medical practitioners. It is eugenics by the back door. Issues of the quality of life to support pricing and access, let alone investment in new products in atopic dermatitis deserve more than this uninspired and insignificant modelling travesty.

REFERENCES

[1] Langley P. 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/2444/2348

[2] Atlas SJ, Brouwer E, Fox G, Carlson JJ, Campbell JD, Agboola F, Hansen RN, Pearson SD, Rind DM. JAK Inhibitors and Monoclonal Antibodies for the Treatment of Atopic Dermatitis: Effectiveness and Value; Evidence Report. Institute for Clinical and Economic Review, July 9, 2021. https://icer.org/assessment/atopic-dermatitis-2021/#timeline

[3] 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. 2
https://pubs.lib.umn.edu/index.php/innovations/article/view/3992/2855

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

[5] 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

[6] 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

[7] Langley P. 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

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

[9] Langley P. The Great I-QALY Disaster. *InovPharm*. 2020; 11(3): No 7 https://pubs.lib.umn.edu/index.php/innovations/article/view/3359/2517

[10] Hume D. An Essay Concerning Human Understanding. 1748

[11] Zimmermann M, Rind D, Chapman R et al. Economic evaluation of Dupilumab for moderate-to-severe atopic dermatitis: A cost-utility analysis. *J Drugs Derm*. 2018;17(7):750-6

[12]Pickard A, Law E, Jiang R et al. United States valuation of Eq05D-5L health states using an international protocol. *ValueHealth*, 2019;22(8):931-41

[13] Andersen L, Nyeland, Nyberg F. Higher self-reported severity of atopic dermatitis in adults is associated with poorer self-reported health-related quality of life in France, Germany, the UK and the USA. *Br J Dermatology. 2020;182:1176-83*

[14] McKenna S, Heaney A, Langley P. Fundamental Outcome Measurement: Selecting patient reported outcome Instruments and Interpreting the data they produce *InovPharm*. 2021;12(2):No. 17 https://pubs.lib.umn.edu/index.php/innovations/article/view/3911/2764

[15] Chernyshov P, Tomas-Aragones L, Manolache LK et al. Quality of life measurement in atopic dermatitis. Position paper of the European Academy of Dermatology and Venereology (EADV) Task Force on quality of life. *J Eur Adad Dermatology Venereology*. 2017;31:575-593.

[16] Chiesa Fuxench Z, Block J, Boguniewicz M et al. Atopic dermatitis in America Study: A cross-sectional study examining the prevalence and disease burden of atopic dermatitis in the US adult population. *J Invest Dermatology*. 2018;139:

[17] Silverberg J, Gelfand J, Margolis D et al. Health utility scores of atopic dermatitis in US adults. *J Allergy Clin Immunol Pract*. 2018;7(4)

[18] Heinl D, Prinsen A, Deckert S et al. Measurement properties of adult quality-of-life measurement instruments for eczema a systematic review. *Eur J Allergy Clin Immunol*. 2016;71:358-70

[19] Whalley D, McKenna SP, Dewar AL et al. A new instrument for assessing quality of life in atopic dermatitis: international development of the Quality of Life Index for Atopic Dermatitis (QoLIAD). *Br J Dermatol*. 2004; 24(1):105-13

[20] Tsianakas A, Luger TA, Radin A. Dupilumab treatment improves quality of life in adult patients with moderate-t-severe atopic dermatitis: results from a randomized, placebo controlled clinical trial. *Br J Dermatol*. 2018;178(2):406-14

[21]McKenna S, Cook S, Whalley D et al. Development of the PSORIQoL, a psoriasis-specific measure of quality of life for use in clinical practice and trials. *Br J Dermatology*. 2003;149(2):323-31

[22] Twiss J, McKenna S. Comparing the impact of psoriasis and atopic dermatitis on quality of life: co-calibration the PSORIQoL and QoLIAD. *Qual Life Res*. 2015;24(1):105-13

[23] Langley P. McKenna S. Fundamental Measurement: The Need Fulfilment Quality of Life (N-QOL) Measure. *InovPharm*.2021;12(2):No. 6 https://pubs.lib.umn.edu/index.php/innovations/article/view/3798

.