

Yet another Ersatz World: The ICER Final Evidence Report for Additive Cardiovascular Therapies

Paul C Langley, PhD

Adjunct Professor, College of Pharmacy, University of Minnesota

Abstract

Previous commentaries in the Formulary Evaluation section of INNOVATIONS in Pharmacy have pointed to the lack of credibility in modeled claims for cost-effectiveness and associated recommendations for pricing by the Institute for Clinical and Economic Review (ICER). The principal objection to ICER reports has been that their modeled claims fail the standards of normal science: they are best seen as pseudoscience. The purpose of this latest commentary is to consider the recently released ICER report for Additive Cardiovascular Disease therapies. This report should not be taken seriously in its claims for cost-effectiveness and pricing in cardiovascular disease (CVD). The analytical framework applied by ICER fails to meet the standards of normal science in demarcating science from pseudoscience. Irrespective of the value judgements and recommendations of an ICER report, these lack credibility. They were never intended to be evaluable and replicable across treatment settings. The claims made are constructed, driven by assumption, and should be put to one side by health system decision makers. In this review the focus is on to the ICER modeled estimates of utility scores in CVD, the insistence on utilizing a generic utility algorithm (the EQ-5D-3L) and the consequent quality adjusted life year (QALY) estimates. Two issues are raised that will be the subject of future commentaries: the lack of appreciation of fundamental measurement and (ii) the importance of the patient voice in benefit claims. Given the importance in the ICER methodology of QALYS, the ad hoc nature of the ordinal utilities introduced to the cardiovascular model must raise concerns over the role the ICER evidence report may play in health care decision-making. These concerns extend to the claim by ICER that, on ICER's own affordability threshold for individual new molecular entities, the anticipated uptake of these therapies may raise questions of overall affordability. Again, we are dealing with an arbitrary construct that may adversely impact patient access.

Keywords: Additive cardiovascular therapies, ICER, pseudoscience, generic QALYs, ordinal utility values, cost-per-QALY claims

Introduction

The construction of assumption driven model worlds to support incremental cost-per-QALY claims for pricing and access recommendations is the hallmark of the ICER business model. The purpose of this commentary is to review ICER's latest report on additive treatments for cardiovascular disease (ATCD), released on 17 October 2019¹. Following previous commentaries on ICER evidence reports the focus will be on the scientific status of the ICER reference case methodology, with particular reference to the claims made for health related quality of life (HRQoL).

The evidence report considers separately two additive therapies for cardiovascular events for those with stable cardiovascular disease: rivaroxaban (Xarelto, Janssen Pharmaceuticals) and icosapent ethyl (Vascepa®, Amarin Corporation). The report concluded that there was a high certainty that rivaroxaban plus aspirin or acetylsalicylic acid

(ASA) significantly reduced the risk of cardiovascular death, stroke, or MI in patients with stable CVD, with limited evidence of any health benefit compared to dual antiplatelet therapy (DAPT). In respect of those with established CVD or a high risk of cardiovascular events treated with statins, there is a high certainty that icosapent ethyl provides a small to substantial net health benefit. At the same time, while the products were deemed cost effective they were then declared not to meet the second ICER hurdle of affordability. This led to ICER issuing an Access and Affordability Alert to signal that the health care costs associated with the potential uptake of the therapies may be difficult for health systems to absorb.

Central to the ICER case for declaring products to be cost-effective is the construction of a lifetime simulated reference case model. The purpose of the ICER ersatz model is to construct, by assumption, incremental cost-per-quality adjusted life year (QALY) claims for the ICER hypothetical target patient population in target disease states. In the case of CVD the model, to represent by assumption an unknown clinical future, was constructed from a lifetime Markov cohort framework. The model compared the addition of rivaroxaban to ASA therapy to ASA alone and the addition of icosapent ethyl to optimal medical management. The two products were modeled separately but within a similar model structure for a common hypothetical population. The modeled evaluation concluded that, in the base case incremental cost-per-quality adjusted life year framework, rivaroxaban versus optimal medical management yielded \$36,000 per QALY gained while

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

icosapent ethyl versus optimal medical management yielded \$18,000 per QALY gained. Both are below the lowest ICER willingness-to-pay threshold of \$50,000 per lifetime QALY.

If a model is constructed by assumptions based upon the translation of clinical data, in this case pivotal randomized controlled trials, together with assumptions based upon data points in the literature or assumptions to be taken at face value with little if any literature based justification, then the model should be seen as just one of many that can be constructed. As the choice of assumptions and the various scenarios presented by ICER as adjuncts to its base case model within the reference case Markov framework demonstrate, a range of alternative cost-per-QALY outcomes are feasible. Add to this the options for alternative hypothetical model frameworks (e.g., rule driven microsimulation models) and the door opens to a multiverse of possible cost-per-QALY hypothetical outcomes. The evidence report model can be easily challenged by other imaginary constructs given the options open to change assumptions and the construction of competing models within the same reference case paradigm^{2 3 4}. A situation where, it might be noted, we can then throw in a further option for the modeling multiverse: the choice of utility metric to drive the QALY estimate. As demonstrated in the health technology assessment literature over the past 30 years, ICER-type modeled cost-per-QALY cost-effectiveness claims for therapy options are entirely discretionary. Indeed, in all too many cases we find the products of sponsors of modeled cost-effectiveness outcomes confirmed as cost-effective by the model.

The purpose of this commentary is to point out that the ICER modeled claims for rivaroxaban and icosapent ethyl, together with ICER claims for affordability should not be taken seriously. They both fail to meet the standards of normal science. As previous commentaries have pointed out, if an ersatz incremental cost per QALY model is constructed, then any number of similar models can be constructed,^{5 6 7 8 9}. None of the claims made for clinical and comparative cost-effectiveness are credible, evaluable and replicable. As such, formulary committees have no idea whether ICER recommendations are right or even if they are wrong, they will never know and they were never intended to know.

Embracing the reference case paradigm means that the ICER value pronouncements for pricing and access are not, in fact, ever meant to meet the standards of normal science. They are immune to failure and do nothing to advance our knowledge for therapy impacts of the therapies in CVD target patient populations^{10 11 4}.

In order to demonstrate the poverty of the ICER business model, a useful distinction is between *value judgements* and *value claims*. A value judgement, an assessment in terms of one's standards or priorities, characterizes the ICER reference case and the construction, by assumption, of an evidence framework. ICER is the sole arbiter of the value metric and of

the criteria that support its claims for pricing and product access. A value claim takes a totally different perspective: it represents the outcomes that patients, caregivers and providers consider appropriate for a target patient population. It recognizes that patient reported outcomes instruments, including those that cast themselves as capturing quality of life (QoL) (or, more narrowly, health related quality of life - HRQoL) must be empirically based and subject to evaluation by third parties. A value claim must be credible, evaluable and replicable. Value claims support an ongoing dialogue that captures the contribution and limitations to our present knowledge of therapeutic options. Rather than arbitrarily imposed criteria for value judgements, where claims are seen all too often through the blinkered lens of a selected generic metric driven by a simulated model world, value claims recognize that knowledge is provisional, not constructed, and that the accumulation of knowledge must be evidence based.

ICER and NICE

It is important to note that ICER has no legislative or regulatory mandate for health technology assessment in the US. While it has taken upon itself the mandate of sole arbiter for rigorous and independent value assessments, its position is significantly different from that of the National Institute for Health and Care Excellence (NICE) in the UK. While ICER may be seen as a simulacrum of NICE (NICE-lite), the facts are that (i) it is not operating in a single payer health care system and (ii) it has no legislative role to provide guidance on the acceptance of technologies as NICE does within the English National Health Service (NHS). ICER's perceived and self-appointed role as an arbiter of value judgements for the US health market, supported by proclaimed processes of stakeholder involvement, clinical benefit assessment, model building and, ultimately, voting by an ICER appointed expert panel on the merits or otherwise of target therapies, should not obscure the fact that the end-result are value judgements that rest on constructed evidence. Its clinical assessments and modeled claims have no more weight (if any) in decision making than other assessment and models that populate the CVD technology assessment literature.

NICE takes a reference case approach to establish model parameters. In this case, however, rather than an in-house model developed by its staff, manufacturers are asked to submit their own reference case model. This is typically a lifetime incremental cost-per-QALY model with the EQ-5D-3L generic HRQoL instrument as the standard utility measure. The model is then submitted, if considered appropriate, to an independent third party (usually an academic center) for appraisal. The appraisers, with experience in the review of ersatz lifetime model simulations, can (i) accept the model; (ii) modify the model or (iii) create an alternative model. This allows transparency in the review process with the final decision by NICE.

The key point is that with NICE and other countries such as Australia, New Zealand and Ireland who have followed NICE's lead in mandating the construction of simulated or modeled worlds to support formulary submissions, the requirement has legislative and regulatory backing. Value judgements driven by the various country models are mandated as key elements in formulary decision making. While modeling reference case lifetime value judgements might be objected to on grounds of scientific merit, there is an acceptance of the approach. While this might seem odd, as noted in a recent commentary: *The playing field is level and all parties know the rules of the 'game'. There are even imaginary world referees, typically in academic institutions, who will adjudicate the manufacturer's imaginary submission. They can pronounce whether it is acceptable, modifiable or should be replaced by the referees own proposal for an imaginary world. NICE, as senior referee, is the judge*¹².

There is no reason why ICER, in its NICE-lite incarnation, should assume that value judgements based on constructed evidence from simulated worlds should have relevance to health care decision making in the US. The US is not a single payer health system. There is no legislated or regulatory across-the-board requirement for imaginary reference case modeling to support value judgements. Certainly, ICER might believe in the sure and certain hope that incremental cost-per-QALY lifetime simulation models are the 'state of the art' in health technology assessment, a position taken by professional groups such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR); this does not mean that the ICER business model standard is appropriate. Indeed, under the Affordable Care and Patient Protection Act (2010) it is made clear that the Patient Centered Outcomes Research Group (PCORI) must exclude discounted cost-per QALY or similar measures as threshold values for priority setting in health by the Centers for Medicare & Medicaid Services (CMS)¹³. While this exclusion gives pause to those advocating the use of QALYs in pricing and access, the debate overlooks a more substantive concern: the ICER business model lacks scientific merit. It is best seen, as detailed below, as pseudoscience. As the ICER reference case methodology does not meet the standards of normal science then any value judgement claims for products should be rejected.

Meeting the Standards of Normal Science

The requirement for testable hypotheses in the evaluation and provisional acceptance of claims made for products and devices is unexceptional. Since the 17th century it has been accepted that if a research agenda is to advance, if there is to be an accretion of knowledge, there has to be a process of discovering new facts. Indeed, as early as the 16th century Leonardo da Vinci (1452 – 1519) in notes that appeared posthumously in 1540 for his *Treatise on Painting* (published in 1641) clearly anticipated the standards for the scientific method which were widely embraced a century later in rejecting thought experiments that fail the test of experience. By the 1660s, the scientific method, following the seminal contributions of Bacon, Galileo, Huygens and Boyle, had

been clearly articulated by associations such as the Academia del Cimento in Florence (1657) and the Royal Society in England (founded 1660; Royal Charter 1662) with their respective mottos *Provando e Riprovando* (prove and again prove) and *nullius in verba* (take no man's word for it)¹⁴.

By the early 20th century standards for empirical assessment were put on a sound methodological basis by Popper (Sir Karl Popper 1902-1994) in his advocacy of a process of 'conjecture and refutation'^{15 16}. Hypotheses or claims must be capable of falsification; indeed, they should be framed in such a way that makes falsification likely. Life becomes more interesting if claims are falsified because this forces us to reconsider our models and the assumptions built into those models. This leads, then, to the obvious point that claims or models should not be judged on the realism or reasonableness of assumptions or on whether the model 'represents' for a public advocacy research group such as ICER their perception of a future, yet unknown, reality.

Although Popper's view on what demarcates science (e.g., natural selection) from pseudoscience (e.g., intelligent design) is now seen an oversimplification involving more than just the criteria of falsification, the demarcation problem remains¹⁷. Certainly, there are different ways of doing science but what all scientific inquiry has in common is the 'construction of empirically verifiable theories and hypotheses'. Empirical testability is 'one major characteristic distinguishing science from pseudoscience'; theories must be tested against data. Indeed, paradoxically, while the development of pharmaceutical products and the evidence standards required by the Food and Drug Administration (FDA) for product evaluation and marketing approval is driven by adherence to the scientific method, once a product is launched and claims made for cost-effectiveness and, in the case of ICER, pricing and access recommendations, the scientific method is put to one side. Pseudoscience succeeds science.

The rejection of a research program that meets the standards of normal science by groups such as ICER is best exemplified by the latest version of the Canadian health technology guidelines where it is stated: *Economic evaluations are designed to inform decisions. As such they are distinct from conventional research activities, which are designed to test hypotheses*¹⁸. While this position puts modeled health technology assessment in the category of pseudoscience, it is also what may be described as a relativist position. Rather than subscribing to the position that the standards of normal science are the only standards to apply in health care decisions and value claims, the relativist believes that all perspectives are equally valid. Health care decisions are to be understood sociologically. No one body of evidence is superior to another. Results of a lifetime modeled simulation are on an equal basis with those of a pivotal Phase 3 randomized clinical trial. For the relativist, the success of a scientific research program, in this case one built on hypothetical models and simulations, rests not on its ability to generate new knowledge but on its ability to mobilize the support of the community.

Basing decisions on models and simulations underpins the consensus view that evidence is constructed, never discovered. Instead of coming to grips with reality, science is about rhetoric, persuasion and authority¹¹. Truth is consensus.

Models and Assumptions

It is accepted that knowledge is provisional and permanently so. This stems from the obvious point that we can at no stage prove that what we 'know' is true. Attempting to believe or justify our belief in a theory is logically impossible. What we can do, by empirical assessment, is to try and demonstrate our preference for one theory over another (and apply it to the best of our knowledge).

ICER's response to public comments makes it clear that not only are the modeled value judgments driven by assumption but that there is considerable scope in the assumptions selected to 'drive' alternative reference case lifetime models. The ICER model assumptions can be considered from three perspectives: (i) assumptions which derive from the clinical literature specific to the target therapies and (ii) general assumptions from the literature; and (iii) convenience assumptions that represent the ICER modeling perspective. In the first group we can point to trial end-points. Amarin in its stakeholder submission for the final draft evidence report notes that given *the strong scientific evidence from REDUCE-IT, it is critical and appropriate to incorporate all statistically significant and clinically meaningful endpoint data for icosapent ethyl into the base-case Markov model to rigorously assess its economic value and budget impact*¹⁹. ICER respondent by saying that, in its view it would focus on three endpoints in its base case MACE analysis and noting that this was an early decision made prior to the model analysis plan and that *In many instances, the FDA's decision around a trial's primary endpoint may not be well aligned with the strongest evidence that informs lifetime costs and QALYs*. It is not clear what ICER means by 'strongest evidence' if this is to support lifetime (i.e., unknown but by assumption) costs and measures of HRQoL. A further issue is the extent to which the hypothetical ICER target CVD population is consistent with the target populations defined by RCT protocols. Janssen makes the point, in the budget impact analysis, that prevalence estimates used by ICER for its product are grossly overestimated. ICER responds that: *We are estimating patients under the approved label, not based on the eligibility criteria for COMPASS*.

In the second group, assumptions from the literature, Amarin again notes that ICER relies upon somewhat dated CVD cost data to drive its assumptions. ICER responds: *We agree that there may be higher quality evidence sources for certain model inputs or that could be used to relax some of the model assumptions* (emphasis added). *However, the scope of this model exercise does not include detailed patient-level evidence generation in the eight-month review timeline. The exercise involves the team identifying best-available evidence sources. Therefore, to provide actionable critiques of this review process would involve suggesting alternative available evidence sources*

that are considered to be of the same or higher level of quality. In other words, we use what is available, even though the evidence source to justify an assumption may not be of high quality.

In the third group are assumptions about future costs and prices. As ICER notes in its response to a further Janssen comment on ICER's assumption that the price of their product will remain steady (but not with an annual 3% price increase) that: *Following standard health economic practice we assumed that the net price of rivaroxaban would remain the same over time, as we have no way to predict price increases or decreases in the future* (emphasis added). This caveat does not, apparently, apply to other modeled lifetime assumptions.

Constructing ersatz worlds which were never intended to generate potentially falsifiable outcomes cannot be defended by an appeal to the 'truth' ('quality of the evidence source') of their assumptions. If a health technology assessment claim is built upon a series of assumptions, a reasonable question is to ask what is the status of the various assumptions? Are they to be viewed as 'reasonable' or 'realistic' metrics for an unknown future reality? Have they been selected from the literature because they seem appropriate? Are they the 'best available' from limited data? Unless there are agreed criteria for assessing the 'quality' of an assumption, we face potentially competing claims for the 'quality' of assumptions in competing models.

More to the point, there is a belief that the fact that the selected assumptions are based, where feasible, on an empirical study validates the choice of assumption. For example, if the model is intended to incorporate utilities that have been reported in one or two studies (usually as few as that) for progression and time spent in the stages of a disease, then there is an immediate methodological issue. To claim that an assumption is valid is to revisit Hume's induction problem (David Hume 1711-1776): an appeal to facts to support a scientific statement. Unfortunately, as Hume pointed out, no number of singular observations can logically entail an unrestricted general statement. Certainly, there may be comfort in reporting that 'so far' the claim that all swans are white has not been contradicted (until that Qantas vacation in Western Australia) so that one fully expects the next swan to be white. But as Hume pointed out, this is a fact of psychology and does not entail any general statement. From a utility perspective, the fact that one hundred papers have agreed (within limited bounds) generic utilities from the same instrument for a target population in a disease state stage is immaterial. We cannot secure this assumption: it cannot be 'established by logical argument, since from the fact that all past futures have resembled past pasts, it does not follow that all future futures will resemble future pasts'²⁰. Claims, for the relevance of a constructed imaginary world built on the assumption that the model elements have been validated by observation is simply nonsensical.

Despite ICER's continued embrace, logical positivism is dead. It died some 80 years ago. All knowledge is provisional. Poppers contribution was to make clear that Hume's problem with induction can be resolved. We cannot prove the truth of a theory, or justify our belief in a theory or attendant assumptions, since this is to attempt the logically impossible. We can only justify our preference for a theory by continued evaluation and replication of claims. Constructing imaginary worlds, even if the justification is that they are 'for information' is, to use Bentham's (Jeremy Bentham (1748-1832) memorable phrase 'nonsense on stilts'. If there is a belief, as subscribed to by ICER, in the relevance of constructing simulated worlds to drive formulary and pricing decisions, then it needs to be made clear that this is a belief that lacks scientific merit.

Cardiovascular Disease: Which Generic QALY?

Previous commentaries in this series have raised the concern that an unqualified use of the term QALY may give decision makers the impression that there is a common QALY standard that has been agreed to in health technology assessment ²¹. This is far from the truth. ICER uses the term, almost indiscriminately, without qualifying its claims that the utility metric driving the QALY estimate is based on an often arbitrary choice of measure. If the intent is to mandate a specific generic utility metric as is the case in the NICE reference case, then for US preference measures there are a number of options: EQ-5D-3L, EQ-5D-5L, SF-36, SF-12, SF-6D. Confusion can arise when, as in the ATCD evidence report, ICER refers to the EQ-5D, without qualifying whether it is the 3-level or 5-level successor variant (introduced in 2009). Reviewing source documents referenced by ICER points to the 3-level variant. The EQ-5D-5L, introduced in 2009, is preferred given the floor and ceiling effects of the EQ-5D-3L and its lack of sensitivity. NICE in the UK is still struggling with the use of the 3-level as opposed to the 5-level. The current position is that the preferred measure is the EQ-5D-3L. If data are collected using the EQ-5D-5L system, utility values in reference case analyses should be calculated by mapping the descriptive systems data to the 3L value ²².

The point to note, however, is not just the limited number of responses open to patients within the health dimensions captured in the generic measure, but the limited ambit of those measures. Are these measures appropriate if the intent is to represent health related quality of life (HRQoL) in CVD? The EQ-5D-3L, for example, is based on five broad health dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Patients respond: no problems, some problems and major problems. The SF-6D, by contrast, has six health dimensions (with levels of response in parentheses): physical functioning (6), role limitations (4), social functioning (5), pain (6), mental health (5) and vitality (4).

Without going into details of each of the various preference-based multi-attribute health status systems it should be emphasized that the decision as to which generic measure to use either in a modeling exercise, a clinical trial or observational

study does matter as the systems are far from identical. They differ in their coverage of health dimensions, in the defined levels, the description of these levels, the severity of the most severe level, the populations surveyed, the instruments used to determine the preference scoring and the theoretical approach for modeling the preference data into a scoring formula ²³. The same patients can have quite different scores depending on choice of instrument. This is seen even with the two versions of the EQ-5D. The EQ-5D-5L yields, at least in the case of rheumatoid arthritis where it is possible with data from the US National Data Bank for Rheumatic Diseases (NDB) to contrast it with the EQ-5D-3L for the same respondents ²⁴. The EQ-5D-5L yielded higher utility scores and a tighter distribution of those scores, with consequent different claims for cost-effectiveness. This is attributable to the five rather than three response levels of the EQ-5D-5L: no problems, slight problems, moderate problems, severe problems and extreme problems.

Similar findings occur with the EQ-5D-5L and EQ-5D-3L when econometric modelling is used to apply mapping algorithms to transition either between the two versions of the EQ-5D as well as mapping to health state utilities from non-preference bases outcomes measures (e.g., HAQ-DI scores in RA). Although there are now good practice guidelines developed by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) for those undertaking mapping studies, the fact remains that different mapping algorithms can produce different utility values with the analyst hopefully defending their choice of method ²⁵. Although NICE, as noted, has been attempting to transition from the EQ-5D-3L to the EQ-5D-5L over the past few years it faces significant challenges, not only in developing a value score, but because the switch has been demonstrated to lead to substantially different estimates of incremental QALYs and claims for cost-effectiveness. This has been shown using both the NDB and the EuroQoL Group data sets ²⁶.

ICER has used both utility values gleaned from selected studies where the study has either presented utility values captured directly from target patient populations or utility values captured indirectly from clinical markers. What does not seem to have been appreciated is the implications for ICER value judgements of options open in: (i) choice of generic utility instrument; (ii) choice of mapping; and (iii) the application of willingness to pay thresholds to support ICER recommendations for possible price discounting and/or access to care under its affordability alerts.

In respect of point (iii) above, it should be made clear that any value judgement (e.g., for price discounting) based upon a willingness to pay threshold (e.g., \$50,000 per incremental QALY) is only meaningful if the technique(s) used to generate utility values as inputs to QALY estimates, *ceteris paribus*, are specified. As utility values are discretionary, defined by the generic instrument and the techniques applied to map to that instrument, a cost per QALY claim will not only be specific to the

characteristics of the target patient population in a disease state for the products selected, but to the construction or choice of utility metric. A mandatory or fixed willingness to pay threshold for incremental QALYs will yield different recommendations by ICER for possible price discounting depending on the value metric as the incremental QALY count will differ. Not only are the ICER models imaginary constructs, yielding different recommendations depending upon the choice of assumption, but the utility value algorithm is discretionary. There is no standard QALY or standard mapping algorithm to support across the board application of fixed willingness to pay thresholds. ICER should create different willingness to pay thresholds for the various utility metrics if its claims for price discounting are to be consistent across the various therapies reviewed. As an example, consider an incremental cost-per-QALY where the incremental costs are fixed but the incremental QALY count is different. One generic instrument may yield 5 lifetime incremental QALYs while another yields 10 incremental QALYs. If incremental costs are \$500,000 (over the hypothetical average patient lifetime), the first model will yield an incremental cost per QALY of \$50,000, the latter \$100,000. If the willingness to pay threshold is \$50,000, the second model will lead to an ICER recommendation for price discounting while the first model results in a no change recommendation. If ICER is to be consistent in its recommendations then the \$50,000 willingness to pay threshold cannot be applied to both.

A failure to appreciate the impact of alternative utility metrics on value judgements is seen in the ACC/AHA statement on cost and value methodology²⁷. As part of the value assessment for the quality of evidence in health economic studies they recommend integrating a 'Level of Value' into clinical guideline recommendations. The levels are defined in terms of cost per QALY gained (e.g., high value an ICER < \$50,000 per QALY gained). No account is taken of the choice of utility metric and its potential impact on value level.

Similar concerns are raised in respect of the application of the CVD PREDICT microsimulation model²⁸. This model, which was developed to the comparative and cost-effectiveness of CVD policies has been applied in an assessment of financial incentives and disincentives for food purchases under the US Supplemental Nutrition Assistance Program (SNAP)²⁹. The evaluation included estimates of QALYs gained and costs per QALY under various SNAP policy scenarios for timelines of 5, 10, 20 and the lifetime of participants. Unfortunately, the authors failed to indicate which utility metric was used to generate QALY estimates. A more recent microsimulation of the economic impact of incentivizing diet with CVD PREDICT in Medicare and Medicaid recipients raised the same issues³⁰.

It is absurd for ICER to claim, as it does in its recent draft evidence report for Oral Semaglutide for Type 2 Diabetes, where it cobbles together utility values from the HUI3 with mapped EQ-5D-3L metrics that: *The utility values for events*

modeled from the risk equations were drawn from two sources due to a lack of a single comprehensive source of health-related quality of life inputs. It is also important to point out that the two sources used different preference-weighted measures (EQ-5D and HUI3), and these two instruments are known to produce slightly different utility estimates (emphasis added; pg. 73)³¹. It is unclear what is meant by 'slightly different' as no references are given for a direct comparison of the two utility value scores for identical patients in the Type 2 diabetes target patient group. We might as well say, as ICER would no doubt endorse, that it is immaterial as far as ICER is concerned whether it uses for its in-house modeling the EQ-5D-3L or EQ-5D-5L utility values in its modeling for the same evidence report as the two yield only 'slightly different' estimates even though the evidence to date would suggest this is patently false. If ICER is to appoint itself as the sole arbiter in point position as the go-to health technology assessment agency then it needs to set standards for its imaginary reference case modeling that have at least minimum credibility. As it stands a company such as Novo Nordisk, the manufacturer of Oral Semaglutide should make it quite clear that any ICER modeled claims should be rejected out of hand.

It is worth also pointing out that utility values captured in the in the ACTD ICER model are taken from publicly available literature with modifications introduced by the model builders. The scores were primarily from a chronic disease study that provided nationally representative EQ-5D-3L scores for a range of chronic conditions defined by ICD9 codes³². Disutility scores were applied to modeled MI, stroke, severe atrial fibrillation, major bleeding and acute non-fatal major adverse limb events (Table 4.7). In the case of stroke, as the severity can vary, a weighted average stroke utility was computed (Table E5) from utility estimates stratified by dependency level from the Virtual International Stroke Trials Archive (VISTA)³³.

Unfortunately, ICER does not explain, in the ACTD model, its reasons for the choice of utility metric, or the criteria it applies in that selection. Apart from the possibility that it was all they could find, why did they choose the EQ-5D-3L instrument? In this respect it is of interest to note ICER's response to an observation by the National Forum for Heart Disease And Stroke Prevention that: *We continue to view QALY as an imperfect metric because it has potential for discrimination against those with baseline disabilities comorbidities and advanced age, all of which are common in CAD patients.* The response from ICER is somewhat disingenuous: *The QALY is the gold standard for measuring how medical treatment improves and lengthens patients' lives, and therefore has served as a fundamental component of cost-effectiveness analysis in the US and around the world for more than 30 years. Because the QALY records the degree to which a treatment improves patients' lives, treatments for people with serious disability or illness have the greatest opportunity to demonstrate more QALYs gained and justify higher prices.* Two questions may be raised: (i) why has a particular QALY system been selected and (ii) the

possibility that utility metrics for functional status and symptom change may have nothing to do with how patients, rather than treating physicians, view the benefits from therapy in meeting their needs.

Ordinal versus Cardinal Measures

Discussions over the application of utility metrics may seem surreal (if not irrelevant) once we consider the measurement properties of the EQ-5D-3L and other preference-based multi-attribute systems. While future commentaries will address the question of measurement in health technology assessment in more detail, it is sufficient for our purposes to point out that these instruments generate raw scores, captured as ordinal scales, rather than interval or cardinal calibrations^{34 35}. In the case of ordinal scales while intervals (e.g., based on utility raw scores and increments within a scale of 0 – 1) are assumed to be equal, the intervals are, by definition, unknown. If attempts are made to manipulate the scores mathematically (e.g., estimating means and standard deviations, estimating change scores or effect sizes) the results are not logically valid³⁶.

The fact that utility metrics are typically expressed through ordinal scales rather than attempting through Rasch modeling to translate the ordinal to a cardinal or interval scale means that conclusions based on these metrics should be put to one side. The ICER QALY is not defensible on purely measurement grounds. The Rasch model, if its standards are met, means that the cardinal translation complies with the two axioms of measurement theory: invariance of comparison within the scale and sufficiency in the total score³⁶. Although Rasch Measurement Theory (RMT) has been widely used in the last 60 years in education and psychology to achieve coherent unidimensional scores to capture latent constructs, it has yet to gain a firm acceptance in health technology assessment. In consequence, groups such as ICER, following the notional ISPOR standard for ‘state of the art’ modeling for value claims, continue to produce evidence reports where recommendations for price discounting and affordability fail to meet required fundamental measurement axioms³⁷. The application of threshold willingness to pay criteria founders because the reference case incremental cost-per-QALY estimates have no logical basis.

ICER Disclaimers

It would have been more informative for ICER to have pointed out:

- That the use of the EQ-5D (or other multi-attribute measure) has ordinal measurement rather than RMT cardinal calibration
- Mathematical manipulations (e.g., means, standard deviations) to create utility stage of disease lifetime profiles and QALYS are not logically valid

and, for those who are committed to ordinal measurement:

- There is no agreed QALY gold standard or universally accepted single utility scoring algorithm, rather there are both generic and disease specific instruments that claim to capture the relevant dimensions of quality of life (or, more accurately, health related quality of life) both across disease states and for CVD as a target disease state
- Although not mentioned in the evidence report (you would have to check the references) it should have been pointed out that the utility metric used by ICER in this report (it differs by ICER report) is the EQ-5D-3L measure, which is only one of a number of generic utility metrics and that can yield different values within disease states for defined target populations. Hence, when combined with modeled time estimated as being spent in a disease stage, which will depend on the model and its assumptions, different utility algorithms will give different QALY estimates
- Certainly, QALYs have been a fundamental component of cost-effectiveness studies although the majority are constructed models which can yield, even in the same disease state, different claims for cost-effectiveness given different model structures, assumptions and choice of utility metric (which also raise the question whether the term ‘cost-effective’ is meaningful)
- To be totally transparent, ICER should have reviewed the various utility metrics and HRQoL outcomes for CVD with both generic and disease specific instruments and defended their choice of the EQ-5D-3L option
- ICER should also have detailed the health dimensions and levels captured by the EQ-5D-3L and made the case that this instrument, in contrast to others, (i) captures the most important and relevant health dimensions for CVD experience in the target patient population and (ii) is appropriate to the impact of the two therapies considered for change over time as CVD progresses and adverse events occur
- ICER should have demonstrated, from its literature review, the QALY change (or just the utility metric change) for the application of the EQ-5D-3L is considered clinically meaningful in the target CVD population
- ICER might also have noted that there is a ‘successor’ utility metric, the EQ-5D-5L and pointed out that the metrics reported (particularly if they have been captured for the target CVD population) may differ between these two versions of an instrument with the same health dimensions (but different response levels)
- ICER might have pointed out that in a lifetime reference case model, claims for utility metrics are by assumption, with possible interpolations from ‘expert opinion’ or their own modelers, and that there is no way the claimed metrics and overall QALY estimates can be ever be evaluated as empirical value claims

- If the concern is with evidence based value claims, rather than simulated value judgements, ICER might point out that if the protocols for the pivotal CVD trials had included the EQ-5D-3L as a primary/secondary endpoint then the outcomes reported may have been different from those assumed by ICER in its modeling from third party sources
- Last but not least, the impression given that any concerns with QALYs have long been settled is false: there is an ongoing debate over the theoretical basis for preference-based utilities, the application of weighting and scoring algorithms, the application of generic utility instruments in older populations and rare diseases, the role of caregivers and the relevance of generic utility measures in target disease states such as CVD^{38 39}.

It is understandable that ICER's views on QALYs should be positive (and defensive), after all lifetime reference case based value judgments are an integral, if not the central, element of their business model. Unfortunately, adherence to this meme puts ICER in a straightjacket. If the reference case is abandoned or if ICER 'relents' and allows non-generic utility measures as part of its imaginary modeling, then it opens the door to even more competing imaginary worlds.

Even so, the question remains unanswered: what are the characteristics of the EQ-5D-3L instrument that make it appropriate for assessing health status in CVD? It would be a mistake to describe the ICER model in 'capturing' a generic utility metric with resultant QALY claims, as indicative of the quality of life of patients in the target patient group. Far from it; at best, unless ICER can demonstrate otherwise, the QALY claim reflects the responses of patients to the health dimensions captured in the generic measure together with the choice of levels reported within each dimension. ICER is assuming that these health dimensions are the most relevant in CVD. ICER fails to point out that the last decade has seen a robust and extended debate over the relevance of the narrow generic measures of value in economic evaluations⁴⁰.

As a general representation of health status, instruments such as the EQ-5D may be appropriate in national health surveys to give a 'broad brush' picture of current health status, it is another issue entirely to consider them as appropriate in target disease states.

Of course, if the ICER reference case means that ICER is committed to the use of generic HRQoL measures as the metric for QALY claims then it should be made clear that ICER has no intention of modifying its modeling to accommodate disease specific measure. Irrespective of whether or not these might be preference based and meet required psychometric and FDA audit standards, let alone required measurement properties, ICER will not include them. Patient advocates might point out that these generic measures are not designed to capture HRQoL

or QoL specific to CVD or other target populations in disease state, but their protests will be ignored.

Disease Specific Instruments

If ICER is concerned about justifying its application, from third party sources, of a generic instrument such as the EQ-5D-3L in CVD, it might take as reference points two widely used HRQoL measures: the MacNew and HeartQoL instruments. Without endorsing the relevance of lifetime simulated models to support non-evaluable value judgement claims, it would be instructive if ICER compared the health dimensions of the EQ-5D-3L with both of these instruments.

The MacNew instrument was designed to evaluate how daily activities and physical, emotional and social functioning are impacted by persons with coronary heart disease⁴¹. It consists of 27 items in three domains: physical limitations (13-items), emotional functioning (14 items) and social functioning (13 items). In addition, there are 5 items that capture symptoms: angina/chest pain, shortness of breath, fatigue and dizziness. The questions refer to the previous 2 weeks, with a low respondent burden (10 mins). It is now in 25 language versions.

The HeartQoL questionnaire is a hybrid developed from the MacNew and two condition specific questionnaires measuring HRQoL in patients with ischemic heart disease. The instrument comprises 14 items, 10 capturing physical status and 4 items for emotional well-being, together with summary scores. A comparison test-retest of the two instruments pointed to their comparability with the focus on the HRQoL as a 'core' instrument⁴². The HeartQoL, together with the MacNew, has been validated in patients with angina, myocardial infarction, and heart failure. The HeartQoL was developed to be a single, reliable and valid core HRQoL instrument for patients with chronic heart disease for between diagnosis comparisons and changes in HRQoL following interventions^{43 44 45}.

The Patient Voice

A further issue that should be noted is whether or not HRQoL instruments, such as the HeartQoL actually capture a latent QoL construct for patients with CVD. Since the mid-1990s there has been a debate in health technology assessment on the concept of patient value and whether HRQoL instruments, which focus on clinical functioning and symptoms, reflect the patient voice: the extent to which competing therapies meet patient needs. Value is hypothesized, within the needs model, *to be dependent on the extent to which human needs are fulfilled*⁴⁶. In this value framework, the presence of disease and its treatment are considered major influences on needs fulfillment. Certainly, HRQoL disease specific measures may capture clinical manifestations of disease. We may infer, indirectly, that this may go some way towards needs fulfillment. This is not sufficient. We require an instrument that is patient centric (and not just an outcomes measure) and one that provides a cardinal measure, a unidimensional scale, of needs fulfilment⁴⁶.

There are, in fact, over 30 patient centric, needs fulfillment measures, that have been developed and applied⁴⁷. The focus of these instruments is to provide a rating scale that is assumed to measure directly a common QoL latent construct. If it can be demonstrated that they meet the standards for Rasch Measurement Theory (RMT), the successor to Classical Test Theory (CTT), then the instrument can be assumed to measure the underlying construct with the final selection of response items generating a technically valid unidimensional scale with a single score. This is not a profile instrument; rather it is an index to support claims that needs are potentially fulfilled⁴⁷. As disease-specific needs measures, with items generated directly from qualitative interviews with target patient populations, are developed to the same RMT standard, they can support direct comparisons between diseases. As McKenna et al conclude: *(i) most available patient reported outcomes measures fail to employ a meaningful construct theory to guide their development and, consequently cannot be validated and (ii) employ CTT rather than a response model ... (so that) it is not possible to relate scores on the scales to a construct theory*³⁵.

Exeunt ICER QALYs?

If our concern is with developing claims that are credible, evaluable and replicable, rather continuing to subscribe to pseudoscience in the construction of ersatz worlds, then the continuing role of QALYs in health system decision making is problematic. While we might argue for the information role of ersatz claims, should we continue with those claims if we know (but ignore) the failure of ordinal utility scores to meet interval scaling standards? A more cynical position might be to argue that QALYs are only seen as relevant in decision making because lifetime ersatz worlds rely on generic ordinal utilities, applied by stage of disease, to create simulated lifetime QALY counts. This is not only a necessary measure of value but the only one applicable if value propositions are driven by matching incremental cost per QALY estimates to willingness to pay thresholds. Take away utilities on measurement grounds, then the QALY edifice and claims for incremental cost per QALY differences collapses, together with the notion that ersatz worlds have a meaningful role. To complete the demolition, we might then bring in the need for measures of value that represent the patient voice not the views of physicians on therapy benefits.

The earlier QALY commentary, by the present author, following a review of the practical impact of modeled cost-per-QALY claims and the unlikely event that they would ever be followed up as credible and evaluable hypotheses, concluded that: *In retrospect, it is doubtful, that the great expectations for QALYs could ever be realized outside of reference case imaginary worlds, or the willingness of decision makers to suspend belief in the standards of normal science, and accept lifetime cost-per-QALY claims as decision criteria. Unless, therefore, a case can be made for short-term and evaluable QALY claims, there seems little scope for QALYs, and associated cost-per-QALY claims, as inputs to formulary decision making. Perhaps, as Pip says to*

Estella, it has been 'a vain hope and an idle pursuit'⁴⁸. After over 30 years perhaps we can put QALYs to one side and return to clinically and quality specific endpoints in comparative claims for pharmaceutical products in disease and therapeutic areas⁷. The qualification to add here is to recognize the importance of fundamental measurement and the needs of the patient if we move to more relevant and evaluable benefit claims.

Product Affordability and Budget Impact Thresholds

Although a therapy may meet ICER's arbitrary willingness to pay thresholds for cost-effectiveness as determined by the imaginary modeled construct and the choice of utility metric matched to the appropriate willingness to pay threshold, this first hurdle may be surmounted only to be halted at the second hurdle: ICER's potential budget impact threshold.

In May 2019 ICER determined that the annual budget impact threshold for each individual new molecular entity would be \$819 million. If projected annual US spending on a specific drug exceeds this threshold then ICER will determine the maximum number of eligible patients who would be able to receive the therapy, at multiple possible pricing points (lower than the price deemed cost effective in the first hurdle analysis) without exceeding the threshold. In effect, the ICER proposal is for a central planning rationing regime with recommendations for patient access, presumably through some form of prior authorization, irrespective of the benefits that excluded patients might receive.

How is this molecular ceiling created? ICER calculates an estimated annual threshold for all net health care cost growth for all drugs and divides this by an estimate of FDA new molecular entity approvals. This yields an average annual threshold for average cost growth per individual new molecular entity (current estimate \$409.6 million). This is then doubled to give the \$819 million threshold for individual new molecular entities that ICER has arbitrarily decided separate the affordable from the non-affordable.

Whether anyone should take this back-of-the-envelope rationing alert seriously is a moot point. In the case of both rivaroxaban and icosapent ethyl, ICER issued what it describes as an Access and Affordability Alert. Apparently, the ICER convened group of experts at the final meeting on the draft evidence, stated that they would consider using rivaroxaban in 30% or more of eligible patients while, according to ICER's affordability projections, only 6% of patients in the US could be treated in a given year before crossing the threshold of \$819 million. Similarly, while ICER's clinical experts at the meeting stated they believed the majority of eligible patients would want to be on icosapent ethyl, only 4% of eligible patients could be treated before breaching the budget threshold.

To recommend a ceiling for patient access to meet a notional budget threshold is to put to one side assessed clinical benefits for the individual patient, and whether this merits additional

funds being allocated, as well as potentially creating waiting lists for access. It is all well and good to recommend prior authorization but without recommended criteria for approval/refusal, it is a hollow recommendation. After all, it would be presumably possible to translate the aggregate budget limit into QALYs and estimate the allocation of QALYs to each molecular entity and estimate the number of patients allowed to utilize the therapy! Unfortunately, this would raise the question again of why generic ordinal QALYs are used when the focus is presumably (again) of the benefits and harms to patients. ICER would also have to argue that, if the object from a societal perspective is to maximize health benefits then it would be reasonable for ICER to nominate other products in specific disease areas that could either be dropped from formulary or have price reductions so that resources could be shifted to 'high value' therapies. This, of course, is unlikely but the first step would be to agree a utility metric that is standardized across disease states to populate imaginary ICER cost-per-QALY worlds and which met fundamental measurement standards.

Conclusions: The Patient Perspective

If we are to understand the contribution of additive therapies in cardiovascular disease in real world treating environments, through for example protocol driven observational studies, then the focus should be on those attributes of a target disease state, defined for a target population, that attempt to build upon claims from RCTs and reflect the patient voice. Typically, RCTs are of short duration and while it is possible to track patients adherence patterns with some clinical markers from laboratory tests, the ability to track patient centric outcomes in a needs-fulfillment measure would seem a key contribution. Of course, if ICER announced that it would only develop models where the EQ-5D-5L was the mandated value metric then it would not only have to revisit all previous reports and provide revised recommendations for pricing and access, but admit that this would do nothing to alleviate concerns with the application of ordinal measures to support recommendations for pricing and access. It would be of interest to speculate on the legal implications of this for pricing and access decisions imposed on

manufacturers by decisions of health systems based on ICER models.

If contributing to our knowledge of therapy response in CVD is an objective, then the ICER reference case simulation with constructed value judgements adds nothing. Indeed, it may have a negative impact. A similar conclusion holds in respect of the CVD PREDICT microsimulation package should ICER decide to assess its outcomes vis à vis regression modeling in a Markov framework. Indeed, ICER value judgements may have the effect of limiting access to therapy where formulary committees take the ICER value judgements at face value without appreciating the pseudoscientific nature of the simulation that is driving those claims both for cost-effectiveness and affordability.

Next Steps

In CVD disease we should be looking, not to ersatz pseudoscientific constructs, but to a program focused on patient response and patient needs that goes beyond RCTs to provide input for therapy decisions and the tracking of patients. This perspective is not, of course new. We have ample experience with building patient registries and of building on existing patient registries to capture the potential of novel therapies. We also have experience in extrapolating from RCTs and building short term models that can be evaluated and the results reported back to formulary committees in a meaningful time frame. At the same time, there are a range of data bases to support 'data mining'. Rather than constructing imaginary worlds to support hypothetical value judgements for pricing and access, ICER would be better employed developing models that actually add to our knowledge of the impact of additive cardiovascular therapies to support formulary decisions and treatment guidelines, notably measures of patient benefit that meet RMT standards.

Conflicts of Interest PCL is an Advisory Board Member and Consultant to the Institute for Patient Access and Affordability, a program of Patients Rising.

References

- ¹ ICER. Additive Therapies for Cardiovascular Disease: Effectiveness and Value. Final Evidence Report. 17 October 2019. <https://icer-review.org/material/cvd-final-evidence-report/>
- ² Langley PC. Resolving Lingering Problems or Continued Support for Pseudoscience? The ICER Value Assessment Update. *Inov Pharm.* 2017;8(4): No 7 <https://pubs.lib.umn.edu/index.php/innovations/article/view/933>
- ³ Langley PC. Transparency, Imaginary Worlds and ICER Value Assessments. *Inov Pharm.* 2017;8(4): No 11 <https://pubs.lib.umn.edu/index.php/innovations/article/view/926>

- ⁴ Langley PC. Alternative Facts and the ICER Proposed Policy on Access to Imaginary Pharmacoeconomic Worlds. *Inov Pharm.* 2018;9(2): No. 10 <https://pubs.lib.umn.edu/index.php/innovations/article/view/1300>
- ⁵ Langley PC. Cost-Effectiveness and Formulary Evaluation: Imaginary Worlds and Entresto Claims in Heart Failure. *Inov Pharm.* 2016;7(3): No. 6 <https://pubs.lib.umn.edu/index.php/innovations/article/view/449>
- ⁶ Langley PC. Multiple Sclerosis and the Comparative Value Disease Modifying Therapy Report of the Institute for Clinical and Economic Review (ICER). *Inov Pharm.* 2017;8(1): No. 12 <https://pubs.lib.umn.edu/index.php/innovations/article/view/492>
- ⁷ Langley PC. Imaginary Worlds and the Institute for Clinical and Economic Review (ICER) Evidence Report: Targeted Immune Modulators for Rheumatoid Arthritis. *Inov Pharm.* 2017;8(2): No. 10. <https://pubs.lib.umn.edu/index.php/innovations/article/view/515>
- ⁸ Langley PC. Rush to Judgement: Imaginary Worlds and Cost-Outcomes Claims for PCSK9 Inhibitors. *Inov Pharm.* 2017;8(2): No. 11 <https://pubs.lib.umn.edu/index.php/innovations/article/view/516>
- ⁹ Langley PC. Another Imaginary World: The ICER Claims for the Long-Term Cost-Effectiveness and Pricing of Vesicular Monoamine Transporter 2 (VMAT2) Inhibitors in Tardive Dyskinesia. *Inov Pharm.* 2017;8(4): No 12 <https://pubs.lib.umn.edu/index.php/innovations/article/view/927>
- ¹⁰ Langley PC. Resolving Lingering Problems or Continued Support for Pseudoscience? The ICER Value Assessment Update. *Inov Pharm.* 2017;8(4): No 7 <https://pubs.lib.umn.edu/index.php/innovations/article/view/933>
- ¹¹ Langley PC. Transparency, Imaginary Worlds and ICER Value Assessments. *Inov Pharm.* 2017;8(4): No 11 <https://pubs.lib.umn.edu/index.php/innovations/article/view/926>
- ¹² Langley PC. ICER, ISPOR and QALYs: Tales of Imaginary Worlds. *InovPharm.* 2019;10(4): No 10 <https://pubs.lib.umn.edu/index.php/innovations/article/view/2266>
- ¹³ Persad G. Priority setting, cost-effectiveness and The Affordable Care Act. *Am J Law Med.* 2015;41:119-166
- ¹⁴ Wootton D. *The Invention of Science: A new history of the scientific revolution.* New York: Harper Collins, 2015.
- ¹⁵ Popper KR., *The logic of scientific discovery.* New York: Harper, 1959.
- ¹⁶ Lakatos I, Musgrave A (eds.). *Criticism and the growth of knowledge.* Cambridge: University Press, 1970.
- ¹⁷ Piglucci M. *Nonsense on Stilts: How to tell science from bunk.* Chicago: University of Chicago Press, 2010)
- ¹⁸ Canadian Agency for Drugs and Technologies in Health (CADTH). *Guidelines for the economic evaluation of health technologies: Canada.* Ottawa: CADTH, 2017
- ¹⁹ ICER. *Additive Therapies for Cardiovascular Disease: Effectiveness and Value – Response to public comments on Draft Evidence Report.* 12 September 2019 <https://icer-review.org/material/cvd-icers-response-to-comments/>
- ²⁰ Magee B. Popper. London; Fontana, 1973
- ²¹ Langley PC. Great Expectations: Cost-utility models as decision criteria. *Inov Pharm.* 2016;7(2): No 14 <https://pubs.lib.umn.edu/index.php/innovations/article/view/437>
- ²² NICE. *Position statement on the use of the EQ-5D-5L value set for England (updated October 2019)*
- ²³ Drummond M, Sculpher M, Torrance et al. *Methods for the Economic Evaluation of Health Care Programmes 3rd Ed.* Oxford University Press, 2005.

- ²⁴ Hernández-Alava M, Pudney S. Econometric modeling of multiple self-reports of health states: The switch from EQ-5D-3L to EQ-5D-5L in evaluating drug therapies for rheumatoid arthritis. *J Health Econ.* 2017;55:139-51
- ²⁵ Wailoo A, Hernández-Alava M, Manca A et al. Mapping to estimate health-state utility from non-preference-based outcome measures. An ISPOR Good Practices for Outcomes Research Task Force Report. *Value Health.* 2017;20:18-27
- ²⁶ Hernández-Alava M, Wailoo A, Grimm S et al. EQ-5D-5L versus EQ-5D-3: The impact on cost-effectiveness in the United Kingdom. *Value Health.* 2018;21(1):49-56
- ²⁷ Anderson J et al. ACC/AHA Statement on Cost/Value Methodology in Clinical Practice Guidelines and Performance Measures. *Circulation.* 2014;129:2329-2345
- ²⁸ Pandya A, Sy S, Cho S et al. Validation of a cardiovascular disease policy micro-simulation model using both survival and receiver operating characteristic curves. *Med Decis Making.* 2017;37(7):802-14
- ²⁹ Mozaffarian D, Liu J, Sy S et al. Cost-effectiveness of financial incentives and disincentives for improving food purchases and health through the US Supplemental Nutrition Assistance Program (SNAP): A microsimulation study. *PLOS Med.* 2018;15(10):e1002661
- ³⁰ Lee Y, Mozaffarian D, Sy S et al. Cost-effectiveness of financial incentives for improving diet and health through Medicare and Medicaid: A microsimulation study. *PLOS Med.* 2019;16(3):e1002761
- ³¹ ICER. Oral Semaglutide for Type 2 Diabetes: Effectiveness and Value. Draft Evidence Report (updated). 12 September 2019. https://icer-review.org/wp-content/uploads/2019/04/ICER_Diabetes_Draft-Evidence-Report_091219-2.pdf
- ³² Sullivan P, Ghushchyan V. Preference-based EQ-5D index scores for chronic conditions in the United States. *Med Decis Making.* 2006;26(4):410-20
- ³³ Ali M, MacIsaac R, Quinn T et al. Dependency and health utilities in stroke: Data to inform cost-effectiveness analyses. *Eur Stroke J.* 2016;2(1):70-76
- ³⁴ 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):516-522
- ³⁵ McKenna S, Heaney A, Wilburn J. et al. Measurement of patient-reported outcomes. 2: Are current measures failing us? *J Med Econ.* 2019;22(6):523-30
- ³⁶ Grimby G, Tennant A, Tesio L. (eds.). The use of raw scores from ordinal scales: Time to end malpractice. (Editorial). *J Rehab Med.* 2012;44:97-98
- ³⁷ 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;21:119-123
- ³⁸ Pettitt D, Raza S, Naughton B et al. The Limitations of QALY: A literature review. *J Stem Cell Res Ther.* 2016. 6:4
- ³⁹ De Smeldt, Clays E, De Bacquer D. Measuring health-related quality of life in cardiac patients. Editorial. *Eur Heart J – Quality of care and clinical outcomes.* 2016; 2:149-50
- ⁴⁰ Brazier J, Tsuchiya A. Improving cross-section comparisons: Going beyond the health-related QALY. *Appl Health Econ Health Policy.* 2015;13:557-565

- ⁴¹ Höfer S, Lim L, Guyatt G et al. The MacNew Heart Disease Health Related Quality of Life Instrument: A summary. *Health Qual Life Outcomes*. 2004;2(3)
- ⁴² Lee W, Chinna K, Bulqiba A et al. Test-retest reliability of the HeartQoL and its comparability to the MacNew heart disease health-related quality of life questionnaire. *Qual Life Res*. 2016;25(2):351-357
- ⁴³ Oldridge N, Höfer S, McGee H et al. The HeartQoL: Part 1. Development of a new core health-related quality of life questionnaire for patients with ischemic heart disease. *Eur J Prev Cardiol*. 2014;21(1):90-7
- ⁴⁴ Oldridge N, Höfer S, McGee H et al. The HeartQoL Part II. Validation of a new core health-related quality of life questionnaire for patients with ischemic heart disease. *Eur J Prev Cardiol*. 2014;21(1):98-106
- ⁴⁵ De Smedt D, Clays E, Höfer S et al. Validity and reliability of the HeartQoL questionnaire in a large sample of stable coronary patients: The EUROASPIRE IV Study of the European Society of Cardiology. *Eur J Prev Cardiol*. 2016;23(7):714-21
- ⁴⁶ McKenna S, Wilburn J. Patient value: its nature, measurement, and role in real world evidence studies and outcomes based reimbursement. *J Med Econ*. 2018;21(5):474-80
- ⁴⁷ Galen Research, Manchester, UK. www.galenresearch.com
- ⁴⁸ Dickens C. *Great Expectations*. London, 1861