12-11-2017

Transparency, Imaginary Worlds and ICER Value Assessments

Paul C. Langley

University of Minnesota, langley@maimonresearch.com

Follow this and additional works at: https://pubs.lib.umn.edu/innovations

Recommended Citation

https://pubs.lib.umn.edu/innovations/vol8/iss4/11

INNOVATIONS in pharmacy is published by the University of Minnesota Libraries Publishing.
Transparency, Imaginary Worlds and ICER Value Assessments

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

Abstract
The Institute for Clinical and Economic Review (ICER) is seen as offering a credible platform for evaluating the pricing policies for pharmaceutical products and devices. Over the past few years ICER has presented a stream of reports, many of which have recommended substantial price discounts where the results of a lifetime cost-per-QALY modeling suggests they are out of line with notional willingness to pay thresholds and arbitrary budget constraints. At the same time, there have been growing concerns over the lack of transparency in the ICER value assessment process, focusing in particular on the refusal by ICER to allow access to its value assessment modeling framework. The purpose of this brief commentary is to point out that the position taken by ICER over model access is not defensible; the arguments given are specious. This ongoing refusal undercuts the ICER claim to be independent and the credibility of ICER recommendations for price discounting. The solution is for ICER to commit to a transparent process of value assessment, allowing in particular access to its models and for the ICER model to be subject to an independent assessment. At the same time, manufacturers and other stakeholders should have access to the model with the opportunity to challenge the model through developing model frameworks which they feel better represent product value. This advocacy, it should be noted, does not reflect acceptance of the ICER lifetime cost-per-QALY value assessment framework. Health care decision makers would be better served by a value assessment framework that provided short-term credible, evaluable and replicable claims, facilitating meaningful feedback to decision makers, and not on the construction of simulated imaginary worlds.

Keywords: ICER, VMAT2 inhibitors, PARP inhibitors, value assessments, pseudoscience, credible claims, imaginary worlds, simulations

Conflicts of Interest: I declare no conflicts of interest or financial interests that I or members of my immediate family has in any product or service discussed in the manuscript, including grants (pending or received), employment, gifts, stock holdings or options, honoraria, consultancies, expert testimony, patents and royalties.

Introduction
The purpose of this commentary is to call into question ICER’s apparent reluctance to engage with manufacturers and other stakeholders in a transparent and, hopefully, jointly rewarding value assessment process. Unlike agencies such as the National Institute for Health and Care Excellence (NICE) in the UK and the Pharmaceutical Benefits Advisory Committee (PBAC) in Australia, transparency for ICER is very much limited in respect of the modeled claims for cost-effectiveness. Manufacturers and other interested third parties are denied access to the actual cost-effectiveness model and are asked to take at face value the description of the model, its structure, parameter inputs and claims for discounted quality adjusted life years (QALYs), years lived (YTD) and direct medical costs. Set against notional willingness-to-pay thresholds, and an arbitrarily defined budget limit, the value assessment then makes claims for necessary downward pricing adjustments for a manufacturer’s product; pricing adjustments which can be substantial.

The lack of transparency in the ICER modeling and review process, notably in respect of the cost-effectiveness model has been recognized for some time. Concerns have been expressed by pharmaceutical manufacturers and industry organizations such as the National Pharmaceutical Council. The importance of raising the issue of transparency has to be seen in the attention given to ICER pricing recommendations. This attention to ICER, according to a recent critique of the revised ICER value assessment framework, stems from ‘…the fact that it has surfaced as a credible response to demand among the US private payers for health technology assessment (HTA) to counter high drug (and device) prices’ 1. This critique, while addressing a number of lingering concerns with the updated value assessment framework, does not address the question of transparency and access to the underlying cost-effectiveness model. Also, it should be noted that in accepting the health technology assessment (HTA) methodological framework in its focus on the construction of lifetime cost-per-QALY modeled claims, this critique puts to one side the more fundamental issues of claims credibility, evaluation, replication and feedback.

Whether the status accorded ICER value assessment claims is warranted or not is a moot point. In previous commentaries in INNOVATIONS in Pharmacy, the ICER HTA methodology has been criticized with the principal argument being that, in..."
following current HTA standards, it fails the standards of normal science. The claims made for competing products lack credibility because they are neither evaluable nor replicable. As part of this ongoing critique four ICER reports have been reviewed. Most recently, a comprehensive review has been presented detailing what are seen as the major methodological flaws in the construction of ICER lifetime cost-per-QALY models, with recommendations for abandoning the ICER modeling approach in favor of short-term models that generate evaluable claims.

This latest review of the ICER methodology should be seen in the context of some 30 commentaries published in *INNOVATIONS in Pharmacy* since July 2016. The focus of these commentaries has been to point to the failure of HTAs to meet the standards of normal science. This is not an exceptional standard as it has been in place since the scientific revolution of the 17th century. Failure to generate claims that meet these standards puts technology assessment at risk of being labeled as pseudoscience, sharing a platform with intelligent design (in the construction of imaginary worlds) rather than natural selection.

**Assumptions and Claims**
The purpose of the present commentary is not to address just the question of the scientific status of the ICER HTA methodology, but to consider ICER’s attitude to transparency in presenting and evaluating value assessment recommendations. The argument presented is that the lack of transparency in developing and presenting modeled claims, together with the absence of a comprehensive review process, must cast doubt on the veracity of such claims and, in the wider context, the acceptance of recommendations for product price discounting.

To give two recent examples: (i) the final evidence report for PARP inhibitors in ovarian cancer and (ii) the final evidence report for VMAT2 inhibitors in tardive dyskinesia. Both reports rely upon long-term or lifetime cost-per-QALY models (including cost-per-outcome variants). There is no attempt in the modeling to generate evaluable short-term claims. As such, if the current HTA standards are accepted, the modeled claims and the consequent recommendations for downward pricing adjustments have to be taken at face value. In the case of the PARP inhibitors, for example, the use of olaparib for maintenance therapy resulted in annual costs of approximately $247,600. At estimated net prices, the cost-effectiveness of olaparib versus placebo was estimated to be approximately $324,000 per QALY and approximately $289,000 per life-year gained. Discounts from WAC to reach willingness to pay thresholds ranged from 8% to 87% ($50,000 to $150,000 per QALY). In the case of the VMAT2 inhibitor products, modeling, generated an ICER versus placebo for valbenazine of $752,080 and an ICER versus placebo for deutetetabenazine of $1,100,773. Against an annual WAC for valbenazine of $75,789 the value assessment concluded that the price should be $11,260; the corresponding figures for deutetetabenazine were $90,071 and $9,158 respectively.

Indeed, in looking back to the 1990s in the dawn and halcyon days of modeling HTAs, a perennial complaint of health systems was the marketing of ‘black box’ claims by manufacturers. A model would be presented, typically in a simple Excel format, and health systems were invited to punch in the relevant parameters to see how the cost-effective the product would be. It was intriguing as to how high the probability was that the product was cost-effective. The only difference is that now we have a more complex black box.

A further issue concerns the absence of a comprehensive, published, independent assessment of the ICER model. In contrast, when a manufacturer submits a modeled claim following the NICE or PBAC guidelines, it is sent to a contracted external assessment agency. Electronic versions of the model also have to be submitted. In the case of NICE, the External Review Group (ERG) provides a full report that may accept the model, suggest minor or major changes, or propose an entirely new model. The ERG review is then assessed by NICE and the results published. It is this degree of transparency and public domain access that is essential if ICER is to maintain its status.

**Challenging Transparency**
An obvious question is why should ICER abandon its present value assessment process? There are three reasons: first, if there is to be any confidence in the merits of the ICER value assessment, then there needs to be a commitment by ICER to full transparency; second, there is always the concern that the model may be defective; and third, that even if not judged defective in its structure and internal operations, it is simply one of a range of models that could have been developed or even reverse engineered to generate countervailing results. Models which, for example, could be tailored to a particular segment of the US health market. If the ICER model is just one among many, with no particular claim to being in pole position, then health systems and other decision makers need to be aware of the strengths and shortcomings of the ICER model.

Given these concerns, it seems pointless to delay any model access to, say, 12 or 18 months after the release of a final report. If mistakes have been made or pertinent advice on clinical appraisals and modeling parameters have been ignored or overlooked, it is incumbent upon ICER to remedy these issues before final pricing recommendations are made. After all, if a recommendation for a price discount against a notional willingness-to-pay threshold is subsequently found to be in error then the respective manufacturer or manufacturers are, presumably, entitled to feel aggrieved given the potential revenue stream losses.
ICER puts forward two arguments for not releasing copies of its cost-effectiveness model. Both are specious. The first argument is that ‘...top flight academic health economists and their academic institutions require that they retain the intellectual property to the executable model and have the ability to use it for future academic purposes’. The second argument is the ‘very real practical barrier that it is not possible to simply hand over a model and expect someone, even someone very skilled to know how to dissect or run the model without extensive help from the model builder’ with the added defense that ‘We have explored this issue with our academic modeling network and received consistent guidance that it is not feasible to them to assist all stakeholders in this effort during the development of the model’. However, there is a faint ray of hope: ‘...we will continue to discuss with our modelling collaborators the possibility of release of additional model information after a suitable “embargo” period to allow for academic publication’. By which time possibly, from the perspective of the manufacturer launching a new product, any damage will have been done.

Without wishing to be unduly cynical regarding whether ICER actually wants reviewers to have access to the cost-effectiveness model, a reasonable response is that, in the present case of the PARP inhibitors with the Modelling Group at the School of Pharmacy, University of Colorado contracted to develop the PARP model, it should be made quite clear that they have no rights to restrict access to the model by stakeholders and external reviewers on the assumption that at some time in the future they may wish to submit the model to peer review and possible publication. The same arguments apply in the case of the College of Pharmacy Modeling Group at the University of Illinois at Chicago where they apparently have presumptive rights to the VMAT2 model. If these groups wish to meet the standards of normal science, then the model access should be a condition of contract as ‘academics for hire’. If they wish to publish results, all well and good, but this should not be a reason to limit transparency. They are no different from other consultants. There is not even the defense, in the absence of information supplied by a manufacturer, that the model contains commercial information.

The second reason is just as ingenious. We, the potential model assessors, don’t apparently have the skills to dissect and fully appreciate the complexities of the model, and even if we had those skills the developers don’t have time to assist us; the we know best defense. Apart from pointing out that the access required is to the completed model and not while the model is being developed, it is simply nonsense to say that it would not be possible for stakeholders such as pharmaceutical manufacturers (who have been globally supporting similar models for some 20 years in making submissions to NICE and other agencies) either to apply their internal skills or to contract for them. There are academic review groups in the UK (e.g., ScHARR at the University of Sheffield) and in Australia (e.g., Adelaide Health Technology Assessment, University of Adelaide) who certainly have the skills to assess the model, point to deficiencies, suggest improvements or even propose an alternative model framework. There are also a number of consulting groups in the private sector with the required skills. Given the stakes, there is no doubt that manufacturers would find the resources to engage reviewers and place these reviews in the public domain.

Exercising the Imagination

There is no gold standard for constructing a modeled imaginary world or simulation. The fact that models may be constrained by standards mandated by the assessment agency, the NICE reference case is an example, may restrict the available options, but still gives consultants and others considerable latitude in model design. Also, the fact that modeled claims typically take a natural or lifetime course of disease, also gives additional latitude in the choice of assumptions. If claims are not evaluable, as would be the case in a lifetime cost-per-QALY models, then this eliminates an important constraint on the exercise of the modeler’s imagination.

In the absence of a gold standard, particularly for a one-size-fits-all imaginary world, means that model builders have few constraints. Certainly, they are constrained by evidence for the course of the disease and target patient characteristics for the index intervention. Beyond that they have a range of options for the structure of the model, application of some variant of the ubiquitous Markov framework, the choice of disease progression staging, selecting transition probabilities, the choice of inputs and the values or parameters attached to those inputs. Indeed, it is entirely possible, and understandable, that those underwriting a model may insist that the model put the product in question in the most favorable light. After all, if formulary approval, a favorable tier placement and a premium price are the objectives, then there is little doubt that such a result will be forthcoming. The model is fulfilling its role as an advocate for the product. In the same vein, product competitors may underwrite models that put their product in the best light against competitor models.

Whether or not the primary purpose of modeling cost-effectiveness claims is to advocate a favorable outcome for a product, there is a role for independent assessors. This is recognized by NICE and the PBAC. This does not mean, however, that groups such as ICER who put themselves forward as independent arbiters should not meet the same standards. This is a particular concern in the US market where target populations for PARP inhibitors or more recently VMAT2 inhibitors are dispersed across a fragmented health care delivery system. ICER in this context is no better than a consulting group commissioned by a manufacturer to prepare a dossier for submission following, for example, the Academy.
of Managed Care Pharmacy (AMCP) Format for formulary submissions. At least the AMCP has the good grace to suggest that an electronic version of the model should be presented. Indeed, in the case of the US, it could be argued that a ‘one size fits all’ model, even if it allows manipulation of key parameters or develops a range of treatment and outcome scenarios, with the inevitable defense of using sensitivity analysis and probabilistic parameters to capture ‘uncertainty’, has to give way to models targeted to patients in a range of health delivery systems. Access to an electronic model would be a necessary first step to ensuring that, to some extent at least, non-evaluable model claims were consistent with target patient characteristics and the decision-making environment.

In the last resort, however, there is no ‘gold standard model’. As noted, in the absence of evaluable credible claims, even the most ‘realistic’ model fails the standards of normal science. This is apart from the obvious point that it is impossible to grade models by their degree of ‘realism’, although this appears to be the intent of model builders. Looking forward 20 – 30 years in modeled claims certainly stretches credulity. There is no feedback to decision makers on the outcomes achieved in treatment practice. There is no possibility of replicating claims assessments across diverse health care delivery systems. We may be able to exercise our imaginations in developing models but, in the absence of platforms for evaluating credible claims, the exercise is, in the last resort, essentially a waste of time and resources.

Nonetheless, constructing lifetime-cost-per-QALY simulations is a major preoccupation in HTA. This stems, in large part, from the standards for formulary submission either recommended or mandated by professional associations or by health system decision makers. The AMCP formulary submission format and the good research practices recommended by the International Society for Pharmacoeconomics and Outcomes Research would be in the former category, the reference case guidelines is to acknowledge: Economic evaluations are designed to inform decisions. As such they are distinct from conventional research activities, which are designed to test hypotheses. Whether decision makers wish to be informed by the construction of imaginary worlds is a moot point.

Care also has to be taken in not being overly-imaginative in constructing lifetime cost-per-QALY models. All too often the evidence base is simply too weak to support even the attempt at modeling. Unfortunately, the natural impulse is for authors to persevere, defending their decision by a brief *mea culpa* statement of data limitations as if that absolved them from any responsibility that their conclusions might actually be acted on. While, as noted, a model should not be judged on the realism of its assumptions, the critical issue is one of experimentation and observation. Even so, it is always worth noting the assumptions supporting the model. In the case of the VMAT2 inhibitors value assessment there is an extensive list of assumptions made to support the base case model. These include assumptions regarding long-term response to treatment, assumptions regarding discontinuation of therapy, assumptions regarding prior treatment before entering the model and assumptions regarding primary care visits to for responders/non-responders in the long term (p.ES16). Equally importantly, it is always worth noting the limitations acknowledged by the model builder. Again, in the case of the VMAT2 inhibitors these include: (i) effectiveness data that were based on limited intermediate measures from clinical trials; (ii) tardive dyskinesia severity measures that do not accurately reflect disease burden on overall quality of life; (iii) lack of data on discontinuation of target modeled medication due to adverse events beyond the first year; (iv) lack of robust data on non-drug costs of tardive dyskinesia; and (v) inability to include sub-populations who may differ from the average tardive dyskinesia patient due to a lack of data (p. ES23).

The response to the challenge of constructing imaginary worlds has been overwhelming with literally thousands of models developed, published and presented to journals and assessment agencies over the past 30 years. Apart from the fact that few of these models have generated credible and evaluable claims, they continue to be a bread and butter product for academic centers and consulting groups. Constructing lifetime imaginary worlds has the obvious appeal that the claims are immune to failure with the result that the only challenge can come from a review of the model itself in the choice of model structure, the choice and measurement of input parameters and the further choice of model scenarios. These standards are well entrenched and, as noted in a previous review of the latest revision to the Canadian guidelines, it is recognized by assessment agencies that they are not intended to meet the standards of normal science. Rather than supporting a process of what Popper would describe as conjecture and refutation in improving our understanding of cost-outcome models to generate credible, evaluable and replicable cost-effectiveness claims, the response as stated in the latest version of the Canadian guidelines is to acknowledge: Economic evaluations are designed to inform decisions. As such they are distinct from conventional research activities, which are designed to test hypotheses. Whether decision makers wish to be informed by the construction of imaginary worlds is a moot point.
Quis custodiet ipsos custodes 23
If we accept, for present purposes, that non-evaluable, assumption-based, imaginative lifetime cost-per-QALY claims have a positive contribution to make as ‘information’ to health system decision makers, then we should, presumably, endeavor to seek assurance that the model itself meets some standard for assessing the construct.

Unlike the review process mandated by NICE in the UK and the PBAC in Australia, ICER has no required formal process whereby its clinical evaluation and cost-effectiveness model is subject to independent professional scrutiny. The NICE Single Technology Appraisal Process (STA) is a good example of the level of transparency and engagement that should be required 24. If ICER was to commit to a similar process of external review then there are a number of elements in the process of independent assessment that should be met.

First, before any engagement between stakeholders and ICER is initiated, in particular the manufacturers whose products are under scrutiny, ICER should agree the nominated agency that is to undertake an independent assessment of the modeled case. The academic group contracted by ICER to build the model should be informed of the choice made and should declare they have no conflict of interest in respect of the assessment group. The contractors for the value assessment report will be advised that their value assessment, to include their clinical assessment of the products or devices under review, their proposed decision framework, modeling assumptions and inputs and the various pricing scenarios will all be subject to review. They will also be advised that all value assessment reviews, unless there is commercially confidential information, will be posted to an ICER public access website.

Second, it is important that the modeled value assessment should be put in the context of previous modeled assessments that have been undertaken by other agencies (e.g., NICE, PBAC) or presented as peer reviewed publications. noting which previous modeled cases have been supported by interested manufacturers. A detailed assessment of these previous modeled claims should be presented, in particular models claims that have generated potentially evaluable and replicable claims (and whether these have been subject to empirical evaluation). While the details of these models will obviously differ (e.g., unit prices of the target pharmaceutical products or devices), the substance of the clinical assessment and the intent of the modeling (e.g., lifetime QALY estimates) will often be similar. This material needs to be reviewed with details on any differences in clinical claims with model frameworks described and compared to the model framework developed for the ICER assessment.

Third, the agency selected for the independent review, will have been advised that their assessment including recommendations will be posted to the public domain. At the same time, by request, any stakeholder to the ICER value assessment process should have access rights to the model and all supporting materials (e.g., copies of referenced papers).

Fourth, there should be agreement between ICER and stakeholders on a set of ‘key questions’ that the independent assessment agency should be asked to address. This ‘assessment template’ should, first and foremost, address the issue of the extent to which the modeled claims meet the standards of normal science: are the modeled claims presented credible, evaluable and replicable? Does the model contracted by ICER provide for feedback to formulary committees so that they can judge the relative clinical and cost-outcomes claims for the competing products? If not, then the assessor should be asked to categorize the model as one that is ‘for information only’ and not one that meets the standards of normal science. As such, it would sit alongside other imaginary modeled worlds with the assessor pointing to the differences between these various ‘worlds’ in choice of model framework and the assumptions built into the model. Particular note might be taken of the utility estimates and whether these are considered as robust.

It should be noted that there are modeling review templates that have been proposed. These include the CHEERS (Consolidated Health Economic Evaluation Reporting Standards) and the more recent PROST (Protocol Standards) checklist 25 26. The CHEERS checklist was designed to optimize the reporting of conventional health economic evaluations, to include both the reporting of evaluations alongside clinical trials as well a non-evaluable lifetime cost-per-QALY models. The PROST checklist is focused on evaluating and replicating evaluable claims through prospective studies where patients are tracked from existing data sources as well as prospective observational studies. It is an integral part of the proposed Minnesota Guidelines for Formulary Evaluation 27. The PROST checklist is not intended to be applied to an assessment of modeled imaginary worlds where there is no intention of presenting claims that can be assessed and reported to formulary committees. The latest version of PROST includes checklist items for evaluating next generation sequencing (NGS) claims 28. It should also be pointed out that the PROST template requires manufacturers to present a protocol for assessing their claims in treatment practice. Obviously, in the case of ICER, it is impossible to put forward a protocol to test claims that are generated by lifetime cost-per-QALY models.

Fifth, as stakeholders with a major interest in modeled claims for efficacy and recommendations for pricing, manufacturers should be given the opportunity for presenting their own modeled claims. These models may range from those intended to generate credible, evaluable and replicable claims for short-term assessment and feedback, to models which adopt a long-term or lifetime cost-per-QALY framework, intended as
‘information only’ offerings. These models may be constructed by consultant or other groups under contract with the manufacturer. Again, these should be posted to the public domain. This should be relatively straightforward given their likely prior experience in submissions to health technology agencies. The intent here, of course, is to point out to health care decision makers that once the standards of normal science are put to one side, then, from the perspective of constructing any number of competing claims from competing imaginary worlds, the only constraint is the lack of imagination of the model builders and their willingness to embrace ‘realistic’ assumptions.

Sixth, there needs to be ample time for responses to the ICER model. There should not be a rush to present claims. Rather there needs to be time (as noted below) for a considered response. Three to four months would be reasonable.

Finally, given ICER is the initiator of this review process and its intent in formulating pricing recommendations, ICER should present responses to the results of the review process. These should be posted on-line, with further responses or rebuttals by stakeholders also placed in the public domain. ICER should then be in a position to complete a final value assessment report.

If ICER is not prepared to engage in this process to support a more comprehensive and transparent process of value assessment, then it may be incumbent upon an outside group or, indeed, individual manufacturers, to establish a website where assessments of the respective ICER reports are posted. These could include:

- an assessment of the model with independent reviews of the model in a standard template format (based on requested access to the electronic version of the model)
- outcomes from alternative model structures (to include a systematic review)
- protocols for short-term claims assessment for evaluation and feedback to formulary committees
- assessments by interested manufacturers and interested third parties (e.g., health system formularies)
- responses to questions directed to ICER and outside groups responsible for the modeling

**Conclusions**

ICER’s apparent unwillingness to engage in a comprehensive and meaningful assessment of the structure and assumptions of their cost-effectiveness model by third parties is unacceptable. Put bluntly, unless ICER is prepared to meet the standards for independent review set by agencies such as NICE and the PBAC, then any conclusions drawn from the model regarding pricing adjustments and budget impact must be put to one side. While this may be seen as an unnecessarily harsh conclusion, the fact that ICER has placed itself as an independent arbiter of value assessments, free of any obligations to manufacturers and health system decision makers, requires an independent and transparent review process. This is not intended to cast aspersions on the academic groups contracted with the ICER to construct models. Rather, experience with model reviews in the UK and Australia points to the leeway possible in structuring models and choice of assumptions in the construction of these imaginary worlds.

Insisting on an independent assessment of ICER models, the cost-per-QALY imaginary world paradigm, is not to endorse the present methodology. As noted in previous commentaries, these constructs fail to meet the standards of normal science. This alone would be sufficient grounds to reject the methodology and consequent value assessments. Even, however, if ICER was prepared to abandon their commitment to what has been described in previous reviews of ICER as pseudoscience and commit to modeling frameworks designed to generate credible and evaluable claims, these should also be subject to independent review. The difference here, however, is that that the review would extend to the evaluation of protocols proposed by ICER to evaluate claims and to track and report the results of the application of these protocols in target patient populations. ICER would be accepting the role of hypothesis testing.
References


10 Pigliucci M. Nonsense on Stilts: How to tell science from bunk. Chicago: Chicago University Press, 2015


14 School of Health and Related Research (SchARR), University of Sheffield, Sheffield UK. https://www.sheffield.ac.uk/scharr


21 Canadian Agency for Drugs and Technologies in Health (CADTH). Guidelines for the economic evaluation of health technologies: Canada. Ottawa: CADTH, 2017


23 Juvenal. Satires (Satires VI) lines 347-8: literally ‘who will guard the guards themselves’.


