INNOVATIONS in pharmacy

Volume 8 | Number 4 Article 12

12-11-2017

Another Imaginary World: The ICER Claims for the Long-Term Cost-Effectiveness and Pricing of Vesicular Monoamine Transporter 2 (VMAT2) Inhibitors in Tardive Dyskinesia

Paul C. Langley
University of Minnesota, langley@maimonresearch.com

Another Imaginary World: The ICER Claims for the Long-Term Cost-Effectiveness and Pricing of Vesicular Monoamine Transporter 2 (VMAT2) Inhibitors in Tardive Dyskinesia

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

Abstract
The recently released value assessment of vesicular monoamine transporter 2 (VMAT2) inhibitors in tardive dyskinesia by the Institute for Clinical and Economic Review (ICER) relies upon a long-term modeling exercise to support recommendations for what the ICER sees as the appropriate pricing for these products if prices are to be judged 'cost-effective'. In this case, the recommendations are for a substantive price reduction of some 90% over WAC. Needless to say, this recommendation is unlikely to be welcomed with open arms by the respective manufacturers of valbenazine and deutetrabenazine. Unfortunately, as has been argued in a number of commentaries published over the past 18 months in INNOVATIONS in Pharmacy, the ICER endorsed health technology assessment methodology that underpins this exercise in building a modeled imaginary world to justify product pricing recommendations is fatally flawed: it does not meet the standards of normal science. Rather than addressing the issue of claims validation for VMAT2 products, the question of generating modeled evaluable claims, among others, for clinical, quality of life and resource utilization outcomes, the analysis focuses on claims that are neither credible nor evaluable and, of course, non-replicable. A more positive and useful approach would be for ICER to focus on a framework where claims could be assessed in the short term to provide feedback to health system decision makers, physicians and patients. Instead, we are asked to believe that we can model 20 or 30 years into the future to establish non-evaluable claims for pricing and, ultimately, access.

Keywords: ICER, VMAT2 inhibitors, value assessments, pseudoscience, limited evidence, credible claims, Imaginary Worlds

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

Irrespective of the disease state that is under consideration, commentaries published in the past 18 months in *INNOVATIONS in Pharmacy* have pointed to the lack of scientific merit that can be accorded current standards in health technology assessment (HTA) [1]. This is seen in the modeling of lifetime cost-per-quality adjusted life year (QALY) models to establish the notional cost-effectiveness of pharmaceutical products and devices. These lifetime models or simulations of therapy interventions, where a target patient disease group with assumed characteristics is tracked by assumption over its remaining assumed life expectancy, following an assumed staging of disease progression, are intended to generate a robust 'value' framework for 'informing' health care decision makers. While the weight, if any, that decision makers place upon recommendations from constructing such imaginary worlds is unclear, the key point is that this is the methodology that ICER puts center stage in its value assessments.

Corresponding author: Paul C Langley, PhD
Adjunct Professor
College of Pharmacy University of Minnesota
Director, Maimon Research LLC
5061 North Apache Hills Trail, Tucson, AZ 85750
Email: langley@maimonresearch.com

Unfortunately, if this methodology is judged against the standards of normal science it fails. The modeled claims are not credible, evaluable or replicable. This has been pointed out in commentaries on ICER value assessments of heart failure (Entresto), multiple sclerosis, rheumatoid arthritis and cholesterol reduction (PCSK9 inhibitors)[2 3 4 5]. At the same time, a more recent commentary has stressed once again the limitations of the ICER methodology in a review of the recently updated value assessment framework[6].

Establishing a credible basis for pricing recommendations is of critical importance in marketing medications and devices. Unfortunately, modeled ICER recommendations are often taken at face value without health care decision makers necessarily understanding the mechanics and limitations of the modeling and value assessment methodology. This has the potential for limiting the commercial success of a product or device with possible adverse consequences for the target treatment population in limiting access to therapy. In the most recent commentary in *INNOVATIONS in Pharmacy*, attention was drawn to the refusal of ICER to allow access to the actual value assessment model[7]. While access to the model and independent assessments of the ICER model structure, including the options for developing alternative models that yield quite different recommendations, does not imply an acceptance of the lifetime cost-per-QALY methodology, access

does set the stage for challenging the model . A challenge which could make clear, as it is proposed to demonstrate here, that recommendations for pricing rest on an assumption driven construct given the lifetime or long-term perspective that is adopted. This does not mean, of course, that ICER should not continue to base value assessments on the construction of imaginary worlds. If ICER believes this makes a substantive and necessary input to formulary decisions then they should continue. Whether these models are of interest to health system decision makers is another question.

Credible Claims

The case to be put forward in this commentary is that the standards for cost-effectiveness and the subsequent recommendations for price adjustments in the ICER value assessment of vesicular monoamine transporter 2 (VMAT2) inhibitors in tardive dyskinesia are not , from the standpoint of normal science, acceptable. In presenting this case (and at the risk of going over ground that has been covered in previous commentaries in *INNOVATIONS in Pharmacy*), it needs to be reemphasized that if the standards of normal science are relevant in formulary decision making and pricing, then claims made for products should be in terms that are credible, evaluable and replicable. The standards of normal science, as they should apply in health technology assessment, are captured in the motto of the Royal Society (founded 1660; Royal Charter 1662): *Nullius in verba* (take no man's word for it). If product claims, including those for cost-effectiveness, are to meet these standards then they have to: (i) involve the construction of an empirically evaluable, coherent model and (ii) facilitate the testing of hypotheses through experimentation or observation. A model is not to be judged by the realism of its assumptions [8] . To argue that a lifetime cost-per-QALY based claim is to be accepted as an input to formulary decision making on the grounds that the simulation is, by assumption, 'realistic' is unacceptable. Not only is it impossible to claim that a simulation projecting forward 20 or 30 years is realistic but, as detailed below, health technology assessment standards in respect of pricing and compliance, ensure that it is not intended to be 'realistic'. This, as noted in previous commentaries is succinctly put in the latest edition of the Canadian health care technology guidelines: *Economic evaluations are designed to inform decisions. As such they are distinct from conventional research activities, which are designed to test hypotheses* [9] [10].

The standards for hypothesis testing through experimentation and observation have been in place since the 17th century[11]. They demarcate science from pseudoscience or, as more strongly stated by Pigliucci, they demarcate natural selection from intelligent design[12]. Claims that meet the standards of normal science can apply to clinical outcomes, comparative effectiveness, quality of life, resource utilization, discontinuation in treatment practice and other measures relevant to formulary decisions. The point is that is possible to develop evaluable claims, to ensure feedback to decision makers where, for example, patients are tracked through observational studies to assess the impact of target therapies in treatment practice.

ICER and VMAT2 Inhibitors

Valbenazine (Ingreeza) and deutetrabenazine (Austedo) are two VMAT2 inhibitors approved by the Food and Drug Administration (FDA) in 2017; both indicated for tardive dyskinesia in adults. Deutetrabenazine is also indicated for chorea associated with Huntington's disease. ICER issued a draft evidence report for VMAT 2 inhibitors at the beginning of October 2017 and a response to comments from stakeholders together with the final evidence report on the 21st November 2017 [13] [14] [15]. The ICER value assessment also included tetrabenazine (Xenazine). As this product is used off-label in tardive dyskinesia and the data for it are limited, it is not included in this review.

According to the ICER value assessment, the base-case modeled results for both products generated increased costs and increased QALYs compared to placebo over an assumed lifetime time horizon (Table ES9). The discounted lifetime costs, where no adjustment was made for possible price increases over the patent lifetime of the respective product, were $185,000 and $6,900 for valbenazine and its placebo comparator respectively, with corresponding discounted lifetime costs of $220,000 and $6,600 for deutetrabenazine. Lifetime modeled discounted QALYs were 15.35 for valbenazine and 15.37 for deutetrabenazine with their corresponding placebo discounted lifetime QALYs of 15.12 and 15.18 respectively. The differences in these imaginary QALYs are 0.23 for valbenazine and 0.19 for deutetrabenazine (equivalent to a modeled projection of 84 days and 69 days respectively). The apparent precision that attaches to these QALY claims should not obscure the fact that they are entirely imaginary and, by construct, unevaluable.

Expressed in discounted incremental cost effectiveness (ICER) terms in comparison with the placebo arm of the model g, lifetime ratios were $752,000 per QALY for valbenazine and $1.101 million per QALY for deutetrabenazine (Table ES10). As these are clearly outside a notional willingness-to-pay threshold of $150,00 per discounted QALY, the annual price required to achieve this for valbenazine (80 mg day) would have to be $11,260 (versus current annual WAC $75,789) and $9,158 (versus current annual WAC of $90,071) for deutetrabenazine (4 mg tablets/day).

ICER also projected that costs per symptom-reduced year for tardive dyskinesia were approximately $71,000 for valbenazine and $105,000 for deutetrabenazine versus placebo. ICER notes: *It is difficult to judge the importance of these results, however, as there are no clear benchmarks of cost per tardive symptom-reduced year for comparison*. These

may, however, be a more relevant metric for health systems, particularly (as noted in previous commentaries) QALY measures are of little interest in US health care and are, under the Affordable Care Act (ACA) explicitly put to one side[16]. Once again, it is unclear how these imaginary claims would be evaluated, if indeed they were ever intended to be evaluated.

Understandably, taken at face value, the results of this modeled outcome for price discounting to achieve a $150,000 willingness-to-pay threshold will be of concern to the respective manufacturers. Against this modeled result, the case made in this commentary is, if the standards of normal science are applied, rather than accepting the construction of an imaginary world, these results should be put to one side. Any concern by manufacturers is misplaced.

Irrespective of whether one believes or not that the standards of normal science apply in health technology assessment, the bottom line is that there is no universal 'gold standard' imaginary world in lifetime cost-per-QALY model building. All models rest upon decisions as to which model structures and assumptions appeal to the model builder. The inputs to the model rest on the vagaries of the available evidence, the assumptions driving how that evidence is manipulated and extrapolated, how those assumptions are 'reasonably' justified and the criteria proposed for determining whether the constructed incremental cost-per-QALY value meets notional willingness to pay thresholds. Pricing recommendations, for what they are worth, necessarily follow.

Obviously, different models will yield different outcomes. Modifying assumptions within the model will yield different outcomes. In the absence of credible, evaluable and replicable claims to support hypothesis testing, one model is as 'good' (or as 'bad') as another. They all, however, share the same methodological flaw. Indeed, it would be a somewhat confusing situation if manufacturers and other stakeholders developed their own models. How would one model claim be judged from another? Would ICER claim the pole position? We could envisage a multiverse of models, each resting on modifications expressed as sensitivity scenarios from their own base-case lifetime scenarios. They all, however, share the same fatal methodological flaw. A situation which, it has been noted in a previous commentary, is found in diabetes modeling[17].

Consider, as an example, evidence from a 'revised' ICER or competitor model that suggests a remodeled estimate of QALY gains. This is not an unreasonable outcome given the often dubious evidence base that supports utility assumptions in lifetime models. If for example, the discounted lifetime QALY difference vs. placebo was 2.4 and not 0.24 for valbenazine, the ICER at $74,287 would be below the annual WAC. Would this lead to a recommendation for a WAC increase? Applying the same argument to the imaginary discounted QALY estimate for deutetrabenazine of 1.9 versus 0.19 would yield an ICER vs. placebo of $112,523. This would imply a significantly smaller price discount than recommended by ICER.

While this example of varying estimated incremental QALYs is trivial, it points to the importance of challenging, within the model framework, key assumptions. In this case, the measurement of discounted QALY gains. Assuming one subscribes to the information role of creating imaginary worlds then a challenge could be made in the case of the VMAT2 inhibitors. Unfortunately, in the model presented by ICER this is not the case. Comments received by ICER as part of the public comment process point to the absence of acceptable utility estimates. ICER in its response to the criticism that it understates the disutility of tardive dyskinesia argued that while there are 'limited data on the impact of tardive dyskinesia on patient's quality of life' and that if data do emerge then ICER '*may* develop an evidence update on an *ad hoc* (*emphasis added*) basis', the overriding consideration for ICER is to focus 'its evaluations to inform policy decisions at or near the time of regulatory approval'. A consideration which, presumably, has to be balanced against the evidence base and gaps in that evidence base. Or, put somewhat differently, the evidence for making the utility assumptions may be shaky, but we need to include it so that we can generate pricing recommendations in an 'immediate' timeframe.

Whether ICER would ever return to a reassessment of utility scores or other assumptions is an open question. It would be hoped that ICER would both make a commitment and undertake to support studies that allowed them to revisit the utility scores. Unfortunately, as noted (see above) in a recent commentary in *INNOVATIONS in Pharmacy*, the refusal by ICER to give manufacturers and others access to the actual cost-effectiveness model means that it is impossible to challenge the assumptions ICER deems appropriate [7]. It is also impossible to judge the claim by ICER that the application of one-way sensitivity analyses and probabilistic sensitivity modeling overcome limitations in data inputs and even the absence of data in constructing imaginary worlds.

While none of the manufacturers' comments in respect of the VMAT2 inhibitors challenge the relevance of constructing a lifetime cost-per-QALY imaginary world, their comments and the ICER responses point to the flexibility accorded model builders in accepting or rejecting proposals to 'modify' assumptions. The impression given in the responses by ICER is that while there are a number of unresolved issues and that these are unlikely to be resolved in the near future, to include QALY measures that may better align with patient and caregiver experience of tardive dyskinesia, ICER cannot wait until these data become available. ICER sees its mission as one of recognizing the paramount importance in informing decision makers of ICER's assessment of the value of products

in clinical practice; the model may be imperfect, assumptions have to be made and they ICER hopes to capture some uncertainty or the lack of evidence in their sensitivity analyses. This is the best of all possible imaginary worlds.

The Clinical Evidence Base

ICER admits that the evidence base for the clinical benefits and harms for valbenazine and deutetrabenazine is limited. In the case of valbenazine three publications and five abstracts for the KINECT 2 and KINECT 3 placebo-controlled randomized clinical trials (RCTs) together with two open-label extensions supported the analysis, with two publications and seven abstracts for two placebo controlled RCTs (ARM-TD and AIM-TD) and one open-label extension supporting the deutetrabenazine model. The studies were rated good or fair quality. The limited number of studies together with protocol heterogeneity did not allow a formal indirect comparison of the two VMAT2 inhibitors.

Trial duration was 6 weeks and 12 weeks for valbenazine and deutetrabenazine respectively. The primary endpoint in the valbenazine RCTs was the change in abnormal involuntary movement scale (AIMS) score at 6 weeks. Secondary outcomes were: investigator rated clinical global impression of change (CGIC) and patent reported clinical global impression of change (PCIC). Both trials showed a significant improvement in AIMS scores (a 50% or more reduction) at six weeks while evidence for an improvement assessed by the CGIC was mixed with the KINECT 3 study showing no significant difference on the CGIC score and CGIC responders compared to placebo. Evidence was also mixed on improvements in PCIC with one trial reporting a statistically significant improvement but no difference in another.

At 12 weeks, the primary efficacy endpoint was change from baseline in the AIMS score (> 50% improvement) for the deutetrabenazine trials. Both trials showed a statistically significant improvement over placebo. Evidence was mixed for the CGIC measure with differences observed in one of three arms of one trial and no difference observed in the other. There were no significant differences observed in the PGIC scale and in the modified craniocervical dystonia questionnaire (mCDQ-24).

The limited number of trials and the mixed results, even with subsequent post-hoc, subgroup and pooled analyses of the data, point to the importance of basing assumptions for modeled efficacy and effectiveness claims on a comprehensive evidence base, to include replication of pivotal phase 3 trial results. While the results reported supported FDA recommendations, the limitations inherent in the trial designs, the duration of the trials and the reported primary (powered) and secondary (underpowered) endpoints point to the care that has to be taken in modeling. This is of particular import when these clinical outcomes are extrapolated for the lifetime of the respected modeled imaginary patient cohorts, even when therapy discontinuation assumptions are built into the model.

Utilities and QALYs

Given the role played by utilities in the VMAT2 inhibitor model and pricing recommendations based upon projected ICERs, it should be noted that none of the trials attempted to capture quality of life as a primary or even secondary endpoint. Apart from the obvious question of which quality of life instrument is appropriate in tardive dyskinesia (generic? condition specific?), the gains in utility scores, which are central to the ICER case for pricing adjustments were modeled from the literature. It is worth quoting in full the process by which the utility scores were derived. This is critical as the construction of the score is by assumption and not actually reported by patients experiencing moderate to severe tardive dyskinesia compared to controls. The value assessment states:

> *The mean utility for the modeled population with improved TD was 0.82, which in part reflects utility scores associated with the underlying condition. [18] [19] To this utility, we applied utility decrement of 0.095 to those patients with moderate to severe TD. The utility gained from improvement in TD was assumed to be independent of any other underlying conditions. We chose this utility decrement for TD because it was the only available utility estimate that directly assessed the impact of TD. The estimate was based on standard gamble utilities from subjects rating TD relative to perfect health and was directly elicited from healthy individuals upon viewing the symptoms of moderate to severe TD, independent of any underlying conditions. Limitations of soliciting utilities from healthy volunteers include that the final utility depends on the accuracy of the scenarios presented[20]. However, patients who have the condition in question often adapt to their conditions and frequently report utilities that are higher (or decrements that are smaller) than those reported from healthy volunteers.* [Note: original references retained]

Building assumption on assumption, with implicit assumptions regarding the application of standard gamble methodology drawn from a single study is questionable. In view of the importance of utility scores in the modeling, it might have been more appropriate for ICER to have undertaken their own empirical valuation of utility scores.

The appropriateness of the standard gamble-based score was raised by Neurocrine Biosciences (valbenazine) in their public comment on the draft ICER model. They believed that the 'TD disutility of 0.095 is too small to correctly capture the impact of TD'. They continue: 'Our preliminary results show that the

mean EQ-5D index score (note: not clear if it is EQ-5D-3L or EQ-5D-5L) for respondents with TD is 0.625 and the mean utility for propensity score matched respondents without TD is 0.750 (a difference of 0.125)'. They suggest that a 'utility decrement of at least double the value used is more appropriate'.

ICER, however, was not to be budged. The response was that they had chosen the best available evidence for the impact of tardive dyskinesia on health utilities, giving the full benefits of completing eliminating tardive dyskinesia symptoms to all patients whose AIMS scores improved by 50% in clinical trials, together with sensitivity assessments. However, they do on to state that 'As new evidence emerges, we *may* (emphasis added again) develop an evidence update on an ad hoc basis'. Given the short review timeframe, it is unlikely that the model outcomes reported in the ICER draft report could have been revisited.

In a previous evaluation of an ICER value assessment, it was cautioned that a 'rush to judgement' might lead to modeled imaginary claims that might subsequently have to be 'modified' (or even rejected). This caution, of course applies across the board to assumptions, including parameter distributions supporting sensitivity analyses. Utility has been singled out because of its central role but also, it is worth noting, because there are a range of quality of life instruments which yield, from underlying differences in their constructs, scoring algorithms and measurement properties, different scores for the same disease state and staging of severity in that disease state.

Other Model Assumptions

The authors of the cost-effectiveness model set out their key assumptions and the rationale for these in Table 5.3. Without going into the detail of this table it is worth noting how limited the evidence base actually was for model building, particularly where the focus was presumably on constructing a 'believable' lifetime cost-per-QALY imaginary world. Consider four assumptions and their rationale:

i. *Response*: The assumption was that therapy response remained constant for all responders; patients did not improve or decline beyond their initial response to therapy while remaining in the model 'improved tardive dyskinesia' state. *Rationale*: There was limited information on the individual change in response to therapy over time; furthermore, there was no information available on the impact of tardive dystonia severity on quality of life.
ii. *Discontinuation*: Long-term discontinuation rates were modeled from open-label studies with less than one year of observation. Following the first model cycle, discontinuation rates were modeled as being 50% of that observed in the first cycle. *Rationale*: There was no information regarding discontinuation rates of therapies beyond the clinical trial extensions. The discontinuation rate was lowered by 50% following the first year, it was argued, because for most therapies, patients typically discontinue their medications at a higher rate in the first year of treatment.
iii. *Resources*: The assumption was that patients who responded to treatment were assumed to have no added primary care and neurologist visit costs related to tardive dyskinesia; those not to treatment were assumed to have two additional primary care and two additional neurological visits per year. *Rationale*: There were no data on the costs associated with treating tardive dyskinesia; it was likely that patients whose tardive dyskinesia has improved will incur future office visits.
iv. *Mortality*: The assumption was that tardive dyskinesia treatments have no effect on mortality. *Rationale*: no studies were identified demonstrating an impact of VMAT2 inhibitors on mortality in patients with tardive dyskinesia.

From the perspective of the standards of normal science, in the absence of a modeling framework that is designed to generate credible, evaluable and replicable claims, a debate over the relative merits of competing assumptions seems pointless. Perhaps the only justification is that, as in the case of the two VMAT2 inhibitors, challenging ICER to justify assumptions for utility scores, quality of life claims, therapy adherence and persistence, will underscore the often tenuous empirical base on which the unevaluable modeled claims rest. Perhaps there should be standards for judging whether or not a model should be attempted if the information vacuum is 'sufficiently empty'.

Of course, we could only hope that when evidence gaps are noted in the construction of 'believable' imaginary worlds, the model builders or ICER could have suggested a research program to capture those data. These evidence gaps may be just as important to building credible and evaluable modeled claims as they are to future 'ad hoc' enhancements to the ICER model that have been promised in the response to public comments.

Acknowledged Limitations

Finally, it should be noted that, apart from the choice of assumptions to drive the model, the authors list what they consider to be limitations. The key limitations include (this list is not presumably exhaustive):

- effectiveness data that were based on limited intermediate measures from clinical trials
- current tardive dyskinesia severity measures that do not accurately reflect disease burden on overall quality of life

- a lack of data on discontinuation of tardive dyskinesia medication due to adverse events beyond the first year
- lack of robust data on non-drug costs of tardive dyskinesia
- inability to include sub-populations that may differ from the average tardive dyskinesia patient, due to a lack of sub-population data

Acknowledging the limitations of the model must cast doubt on whether or not it is intended to be taken seriously. Or, more to the point, whether recommendations for imaginary ICERs and price reductions should be taken seriously. As it stands, the model presented in the ICER report is only one of many that could be constructed. Of course, if one wishes to maintain the integrity of constructing imaginary lifetime cost-per-QALY claims, then the prospect is that at some time, hopefully in the not too distant future, ICER will present a 'revised' model that recalibrates the 'base-case' to accommodate additional observations.

Conclusion

Previous commentaries on the ICER value assessment model have cautioned against taking recommendations for pricing adjustments at face value. For precisely the same arguments that have been put forward in this review of the VMAT2 inhibitor models, previous recommendations have questioned the willingness to base recommendations on the construction of imaginary worlds. Further questions have to be raised when ICER then assess pricing decisions on the basis of budget constraints and the application of willingness-to-pay thresholds.

The objections to this commitment to constructing imaginary worlds to support pricing adjustments are twofold: (i) there is any number of imaginary worlds that could be constructed and (ii) the creation of claims that are not credible, evaluable or replicable, failing to provide feedback to formulary committees, is unacceptable. Certainly, ICER can persevere in generating value assessments that rest upon imaginary constructs, defending data limitations through sensitivity analyses and the creation of targeted imaginary scenarios. Whether the construction of imaginary worlds can be said to 'inform' decision makers is another question. The fundamental objection is that ICER value assessment models, however 'realistic' the construct is intended to be, fail the standards of normal science. We have no idea if they are right or if they are wrong – and we will never know.

What are possible next steps? This has been made clear in previous commentaries. Put to one side the obsession with creating imaginary worlds; the fact that these are a cornerstone of health technology assessment is irrelevant. The focus should be on the evidence base. Can we develop evaluable short-term modeled claims, extrapolating from clinical trials, as opposed to claims stretching 20 or 30 years ahead? This would allow health systems to evaluate claims in treatment practice. If so, then we should put hasty and indefensible efforts at value assessment to one side, focusing instead on evaluating credible, evaluable and replicable claims for products such as valbenazine and deutetrabenazine.

References

[1] *INNOVATIONS in Pharmacy*. Formulary Evaluations. https://pubs.lib.umn.edu/innovations/formulary-evaluations/

[2] Langley PC. Cost-Effectiveness and Formulary Evaluation: Imaginary Worlds and Entresto Claims in Heart Failure. *Inov Pharm.* 2016;7(3): Article 6. https://pubs.lib.umn.edu/innovations/vol7/iss3/6

[3] 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): Article 12. https://pubs.lib.umn.edu/innovations/vol8/iss1/12

[4] 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): Article 10. https://pubs.lib.umn.edu/innovations/vol8/iss2/10

[5] Langley PC. Rush to Judgement: Imaginary Worlds and Cost-Outcomes Claims for PCSK9 Inhibitors. *Inov Pharm.* 2017;8(2): Article 11. https://pubs.lib.umn.edu/innovations/vol8/iss2/11

[6]Langley PC. Resolving Lingering Problems or Continued Support for Pseudoscience? The ICER Value Assessment Update. *Inov Pharm.* 2017;8(4): Article 7. https://pubs.lib.umn.edu/innovations/vol8/iss4/7

[7] Langley PC. Transparency, Imaginary Worlds and ICER Value Assessments. Inov Pharm. 2017;8(4): Article 11. https://pubs.lib.umn.edu/innovations/vol8/iss4/11

[8] Popper KR. Objective Knowledge. Rev Ed. Oxford: Clarendon Press 1986

[9] Canadian Agency for Drugs and Technologies in Health (CADTH). Guidelines for the Economic Evaluation of Health Technologies: Canada (4th Ed). Ottawa: CADTH, 2017i

[10] Langley PC. True North: Building Imaginary Worlds with the Revised Canadian (CADTH) Guidelines for Health Technology Assessment. *Inov Pharm.* 2017;8(2): Article 9. https://pubs.lib.umn.edu/innovations/vol8/iss2/9

[11] Wootton D. The Invention of Science. New York: Harper, 2015

[12] Pigliucci M. Nonsense on Stilts: How to tell science from bunk. Chicago: Chgicago University Press, 2010

[13] Institute for Clinical and Economic Review (ICER). Vesicular monoamine transporter 2 inhibitors for tardive dyskinesia: Effectiveness and value. Draft evidence report. October 2, 2017. https://icer-review.org/wp-content/uploads/2017/04/NECEPAC_TARDIVE_DRAFT_EVIDENCE_REPORT_10022017.pdf

[14] Institute for Clinical and Economic Review (ICER). Vesicular monoamine transporter 2 inhibitors for tardive dyskinesia: Effectiveness and value. Response to public comment. November 21, 2017. https://icer-review.org/wp-content/uploads/2017/11/NECEPAC_TD_Response_to_Comments_112117.pdf

[15] Institute for Clinical and Economic Review (ICER). Vesicular monoamine transporter 2 inhibitors for tardive dyskinesia: Effectiveness and value. Draft evidence report. October 2, 2017. https://icer-review.org/wp-content/uploads/2017/04/NECEPAC_TD_Evidence_Report_11212017.pdf

[16] Langley PC. Great Expectations: Cost-Utility Models as Decision Criteria. *Inov Pharm.* 2016;7(2): Article 14. https://pubs.lib.umn.edu/innovations/vol7/iss2/14

[17] Langley PC, Rhee TG. Imaginary Worlds: The Status of Simulation Modeling in Claims for Cost-Effectiveness In Diabetes Mellitus. *Inov Pharm.* 2016;7(2): Article 17. https://pubs.lib.umn.edu/innovations/vol7/iss2/17

[18] Calvert NW, Burch SP, Fu AZ, Reeves P, Thompson TR. The cost-effectiveness of lamotrigine in the maintenance treatment of adults with bipolar I disorder. *J Man Care Pharm : JMCP.* 2006;12(4):322-330.

[19] Wang PS, Ganz DA, Benner JS, Glynn RJ, Avorn J. Should clozapine continue to be restricted to third-line status for schizophrenia? : a decision-analytic model. *J Mental Health Policy Economics*. 2004;7(2):77-85.

[20] Lenert LA, Ziegler J, Lee T, Sommi R, Mahmoud R. Differences in health values among patients, family members, and providers for outcomes in schizophrenia. *Med Care*. 2000;38(10):1011-1021.