INNOVATIONS in pharmacy

Volume 8 | Number 2 Article 9

5-4-2017

True North: Building Imaginary Worlds with the Revised Canadian (CADTH) Guidelines for Health Technology Assessment

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

True North: Building Imaginary Worlds with the Revised Canadian (CADTH) Guidelines for Health Technology Assessment

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

Abstract

In March 2017 the Canadian Agency for Drugs and Technologies in Health (CADTH) released the 4th edition of their Guidelines for the Economic Evaluation of Health Technologies: Canada. These guidelines, which were first published and revised for a 3rd edition in 2006 are intended to help decision makers, health systems leaders and policy makers make well-informed decisions. They are designed, apparently, to support best practice in conducting health technology assessments in Canada. The purpose of this commentary is to consider whether or not the evidence standards proposed and the consequent modeled claims for economic effectiveness meet the standards of normal science: are the CADTH standards capable of generating claims for competing products that are credible, evaluable and replicable? The review argues that the standards proposed by CADTH do not meet the standards expected in normal science. Technical sophistication in building reference case imaginary worlds is not a substitute for claims that are experimentally evaluable or capable of assessment through systematic observation. There is no way of judging whether imaginary claims are right or even if they are wrong. CADTH is not alone in setting standards that fail to meet the standards of normal science. Recent commentaries on formulary submission guidelines in a number of other countries, to include Ireland, the Netherlands, France, Australia, the UK and New Zealand conclude that they are subject to the same criticism. If the CADTH guidelines were never intended to support feedback to health system decision makers, then this should be made clear. If not, then consideration should be given to withdrawing the guidelines to ensure they conform to these standards. Hopefully, future versions of the CADTH guidelines will address this issue and focus on a rigorous research program of claims assessment and feedback and not the building of imaginary worlds.

Keywords: CADTH Guidelines, economic evaluations, imaginary worlds, pseudoscience, simulations

Introduction

Canada, in common with countries including the UK, Ireland, the Netherlands, New Zealand and Australia, has issued guidelines to support formulary submissions [1 2 3 4 5 6]. In the US recommendations for formulary submission standards have been proposed by the Academy of Managed Care Pharmacy (AMCP) together with modelling standards proposed by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) [7 8 9]. Guidelines have also been proposed as a standard for the European Union under the EUnetHTA umbrella [10]. A common feature of many of these guidelines, as in the Canadian case, is to mandate a cost-utility reference case for the modeling of disease interventions, with particular reference to chronic disease.

Previous publications and a number of formulary evaluation commentaries in this series have made clear that in putting to one side a commitment to the standards of normal science, in which modeled claims or hypotheses are credible, evaluable and replicable, decision makers in health care have a limited evidence base for formulary decisions [11 12 13 14 15]. A lifetime cost-per-QALY model is not designed to generate evaluable claims [16]. It is a construct that is defended by its sufficient correspondence to a perceived reality. Validation focuses on the core model and its assumptions. Whether or not the model can support evaluable claims and whether or not these claims could ever be evaluated is apparently irrelevant. Indeed, in the draft of the present guidelines issued in October 2016 it was made quite clear that, from CADTH's perspective: 'Economic evaluations are designed to inform decisions. As such, they are distinct from conventional research activities, which are designed to test hypotheses' [17] . This position continues to be held. The standards of normal science are not intended to apply to the modeled claims; hypothesis testing is explicitly rejected.

In rejecting the standards of normal science, advocates of models that are intended to 'inform' decision makers in health care systems (whatever that means) rather than establish a practical research program, put to one side a commitment to standards that have been in place since the seventeenth century in favor of what may be described as pseudoscience: intelligent design rather than natural selection [18]. In an effort to avoid this characterization, guidelines have been proposed by the Program in Social and Administrative Pharmacy at the University of Minnesota that reject imaginary constructs in favor of credible, evaluable and replicable claims; claims which apply equally well to clinical outcomes as well as those for comparative cost-effectiveness and budget impact [19].

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
Web: www.maimonresearch.net

Formulary submissions are to be supported by protocols to detail how the claims are to be evaluated and reported. This requirement is not new. It was put forward as a standard over ten years ago in formulary submission guidelines developed for the Wellpoint (now Anthem) health system in the US [20] [21]. In a recent overview of the 22 commentaries on health technology assessment published in INNOVATIONS in Pharmacy since mid-2016 the importance of distinguishing science from pseudoscience (a.k.a true bunk) was emphasized [22]. The concern is that technology assessment agencies persist in their commitment to the construction of imaginary worlds and ask decision makers to take them seriously.

The purpose of this commentary is to review the standards proposed for health technology assessment in the 4th edition of the CADTH guidelines. The commentary adopts the same format as that followed in the reviews of other health technology assessment guidelines: (i) a brief overview of the structure and standards proposed; (ii) a discussion; and (iii) suggestions for possible revisions to the guidelines. Questions that are addressed are: do the guidelines support claims for new medicines or medicinal products that are credible, evaluable and replicable? Are the guidelines capable of supporting a progressive research program of experimentation and systematic observation? Are the guidelines capable of supporting feedback on product performance to physicians, patients and health system decision makers?

Guideline Structure

The CADTH guidelines are intended to help decision makers, health systems leaders and policy makers make well informed decisions. They are designed to support best practice in conducting health technology assessments in Canada. The submission guidelines cover:

- Guideline statements: summary points for the main sections
- Section1: Decision Problem
- Section 2: Types of Evaluations
- Section 3: Target Population
- Section 4: Perspective
- Section 5: Time Horizon
- Section 6: Discounting
- Section 8: Modelling
- Section 9: Effectiveness
- Section 10: Measurement and valuation of health
- Section 11. Resource Uses and Costs
- Section 12: Analysis
- Section 13: Uncertainty
- Section 14: Equity
- Section 15: Reporting
- Appendix 1: Standard Reporting Format
- Appendix 2: Reference Case (defined in terms of Sections 1 through 14).

Section1: Decision Problem

For the CADTH guidelines, economic evaluations are intended to inform decisions; they are not intended to test hypotheses. The decision problem entails identifying the perspective of the problem to be addressed together with the relevant costs and outcomes. The assessment should be comparative reflecting the variety of interventions relative to the decision problem. The time horizon should be identified for the specified costs and outcomes. Where relevant, sub-groups for the intervention should be identified. *Reference Case: specify the interventions, setting, perspective, costs, outcomes, time horizon and target population.*

Section 2: Types of evaluations

The recommended evaluation is generic cost-utility analysis. This is intended to allow broad comparisons across different conditions and interventions, facilitating resources based on maximizing health gains. However, the guidelines do not recommend or mandate a specific utility instrument. The absence of a mandated specific instrument renders claims for cross disease and product areas difficult (if not impossible): different HRQoL instruments 'can produce very different utilities for the same health state'. The guidelines recognize this limitation, but argue it speaks to the need for further methodological advances. *Reference Case: Cost-utility analysis with outcomes captured as QALYs.*

Section 3: Target Population

The submission should specify the target populations(s), together with their description. This should include patient characteristics, disease severity and grading, and the distribution of comorbid states. Models should address questions of heterogeneity in the target population and its quantitative impact. To avoid post hoc data dredging, robust evidence should be provided to justify sub-group analysis for more homogeneous target populations. *Reference Case: Identify population(s) for intervention, stratified to distinct sub-groups if necessary*.

Section 4: Comparators

Comparators should be identified as part of the decision problem. Interventions currently used and potentially displaced should be identified together with interventions likely to be available in the near future. These should relate to the complete clinical pathway. If current technologies are of poor or uncertain value, best supportive care should be considered as inclusion as a comparator. Comparators could be management strategies intended to maximize impact of primary therapy. Comparators should be clearly identified to account for all relevant costs and outcomes. *Reference case: All relevant interventions including current care.*

Section 5: Perspective

The perspective of the reference case is the publicly funded health care payer. The costs should be those incurred by the Canadian public payer with all meaningful patient outcomes. These could include government programs and services beyond healthcare. Supplementary non-reference case analyses could accommodate wider and multiple perspectives. *Reference case: Publicly funded health care payer.*

Section 6: Time Horizon

The time horizon should be based on the condition and the likely impact of the intervention. In the reference case, the horizon should be long enough to capture all potential differences in costs and outcomes associated with the intervention. A longer term analysis allows for the exploration of uncertainty, but does not imply that primary data must be collected from patients or affected populations over this time period. In chronic conditions or when interventions have differential effects on mortality a lifetime horizon is most appropriate. If no long term differences, a shorter time line may be considered. *Reference case: Long enough to capture all relevant differences in the future costs and outcomes associated with the intervention*.

Section 7: Discounting

Costs and outcomes (utilities) over the timeframe of the model should be discounted at a rate of 1.5% per annum. *Reference case: Costs and outcomes at a rate of 1.5% per annum*.

8. Modelling

The model, where a probabilistic rather than a deterministic framework is preferred, should be consistent with the decision problem describing the clinical or care pathway. Choice of model is at the discretion of the manufacturer in representing the natural history of the target treating population. The model should be validated: face validity of the model structure, assumptions, data and results. Validation does not include empirical assessment (i.e., hypothesis testing) of modeled claims. *No Reference case equivalent*.

9. Effectiveness

A comprehensive search of data sources should be undertaken to inform estimates of effectiveness and harms: fitness for purpose, credibility and consistency. Justify surrogate end-points. Capture uncertainty through reference case probabilistic analysis. *No reference case equivalent.*

10. Measurement and Valuation of Health

QALYs are the 'gold standard' value of the effect of an intervention, using health preferences from an indirect (i.e., multi-attribute) generic classification (EQ-5D-3L or EQ-5D-5L, HUI, SF-6D). Health preferences should reflect those of a general Canadian population. No preference for one measure over another. *Reference case: Identify, measure and value all relevant health outcomes with Canadian population preferences from an indirect method of measurement*.

11. Resource Use and Costs

Should be based on Canadian sources with valuation from perspective of publicly funded health care payee. Data sources should be selected that reflect opportunity cost. If appropriate, broader societal impact should be captured in intervention impact on paid or unpaid work by patients and caregivers. *Reference case: Estimate resource units and costs using jurisdiction of interest from perspective of public health care payer.*

12. Analysis

Economic evaluation assessed based on incremental cost-effectiveness ratio (ICER). In reference case typical application will be Monte Carlo probabilistic simulation. *Reference case: Probabilistic sensitivity analysis with ICERs.*

13. Uncertainty

Capture uncertainty through probabilistic analysis. Present results as cost-effectiveness acceptability curves and cost-effectiveness acceptability frontiers. Structural uncertainty to be addressed through scenario analysis. *Reference case: Compare to non-reference case analysis, apply cost-effectiveness acceptability curves and acceptability frontiers.*

14. Equity

All outcomes weighted equally. Describe fully relevant target populations. *Reference case: Weight equally regardless of characteristics of those receiving or affected by intervention.*

15. Reporting

Critical requirements are (i) transparency and (ii) description of quality assurance processes. *No reference case equivalent.*

Discussion

There is no requirement in the CADTH guidelines for claims for competing products to be presented that are credible, evaluable and replicable. Credible claims, in this context referring to their ability to be empirically evaluated not as outcomes that are judged credible because the model is considered to be sufficiently representative of the 'real' world. There is no intention in the guidelines that any modeled claim should be evaluable. This conclusion applies to models irrespective of whether or not they take a short-term or a long-term or lifetime perspective. In these respects the CADTH guidelines are no different from those in Australia, or from those for the UK with the National Institute for Health and Care Excellence (NICE), Ireland, New Zealand and the Netherlands.

In common with the majority of other guidelines that have been reviewed, the CADTH guidelines mandate a reference case. This is similar to the reference case mandated by NICE but without the requirement for a preferred generic utility instrument (in the case of NICE the EQ-5D). The other key difference is the absence of any attempt to apply willingness-to-pay thresholds for cost-per-QALY claims. As noted below, this is probably wise given the absence of a preferred utility instrument.

Constructed Evidence

Unfortunately, as noted in the case of the PBAC guidelines and others, in the absence of claims that are evaluable and replicable, there can be no assurance that they are credible. To argue that they are acceptable and can play a role in 'informing' decision makers, in the absence of any empirical evidence to support the claims, is to adopt a relativist position. Hypothesis testing of competing clinical claims and the cost-effectiveness claims that rely upon them are, for CADTH, irrelevant to formulary decisions. Economic evaluations are not a research activity. Questions of the credibility of claims, including clinical claims, which one imagines include patient reported outcomes such as pain, satisfaction with care, side effects and even quality of life, can be put to one side even though these are essential to a composite claim for the effectiveness part of the cost-effectiveness model. There is no need, presumably, to consider the null hypothesis that there is no difference in patient reported outcomes between competing therapies?

For a relativist evidence is never discovered, only constructed within a particular social community [23]. The community of pharmacoeconomists who accept a reference case paradigm to support imaginary claims for competing pharmaceutical products through 'informing' decision makers would, as relativists reject any arguments that one body of evidence is superior to another. A research program is not seen as one that generates new knowledge through claims evaluation and replication but one that is judged on its ability to persuade and mobilize community support for invented facts. Such a research program puts to one side any notion of the progress of science, of the process through which new evidence overturns consensus views, in favor of rhetoric and authority. Reference case imaginary modeling, the rejection of experiment through hypothesis testing or systematic observation, is seen as the 'gold standard'. This acceptance of a gold standard to support the construction of imaginary worlds is in direct contrast to the motto of the Royal Society (founded 1660; Royal Charter 1662): *nullius in verba* (take no man's word for it') [24]. As stated on the Royal Society website, this motto 'is an expression of the determination of Fellows to withstand the domination of authority and to verify all statements by an appeal to facts determined by experiment'. This stricture applies equally well to the uncritical acceptance of clinical trials where there is little evidence for replication of results as well as to the constructed evidence and conclusions of long-term and lifetime economic models. Whether this is in the best interests of provincial authorities who may rely on CADTH assessments is a moot point.

Rejecting constructed evidence to support therapeutic claims raises questions as to the relevance of technical standards for constructing long-term or lifetime models. The detail in the standards required by CADTH for modeling the imaginary world of the reference case points to a potential substantial misallocation of time and resources to justify a cost-utility model that at the end of the day supports unevaluable claims.

Immunity to Failure

While simulations can fail, lifetime cost-per-QALY modeled claims are immune to failure. Presumably, that is their attraction to both manufacturers and assessment agencies. The only basis on which a non-evaluable modeled claim can be challenged is on a review of the structure of the core model (e.g., state transition models) and the assumptions of the model (e.g., state transition probabilities). In practical terms, it is possible not only to build models that produce competing cost-per-QALY claims as well as models that reverse engineer to generate competing results. In both the UK and Australia models are subject to independent review and appraisal. The assessors may recommend structural changes to the core model or challenge the basis on which assumptions have been derived (e.g., attaching quality of life weights to therapy states). Presumably, the criterion employed is whether or not the appraisers consider the model is a 'sufficient' representation of the reality captured by the natural course of the disease and the impact of competing interventions. If it is judged 'sufficient' then the non-evaluable results necessarily follow. Further assessment is unnecessary. The model and its claims can then be forwarded for peer review and publication, joining the thousands of other non-evaluable cost-effectiveness and cost-utility models that have been published over the past 30 years.

If these unevaluable claims are intended to inform decision makers in health systems them CADTH should make clear to agencies such as provincial health authorities that they are only intended to do this. They are not intended to generate evaluable claims. The provincial health authorities can then decide whether this is a useful input to decision making or whether they should be put to one side in favor of a more rigorous and credible assessment protocol.

Validation

Irrespective of the extent to which model builders claim that their imaginary worlds are a valid construct and have a pivotal role to play in informing decision makers, the validation standards proposed in the CADTH guidelines make no claim for validation that involves prospective, protocol-driven empirical evaluation in target treating populations. The validation

process, as reported in the submission is intended to cover internal validity, external validity and cross-validity. Where the question of external validity with actual data is addressed it is in terms of 'whether the model estimates are consistent with other reliable and preferably independent data sources'; hypothesis testing is not seen as essential and if 'the objective of the analysis is to make projections about the future, the validity of the model and the associated projections should be assessed as data become available'. Presumably, it is not a concern if these comparisons are impossible to make or if, in the future, these data are not available.

In practice, validation for CADTH means demonstrating that *patients track though the model structure appropriately, typically following treatment guideline algorithms and stage evaluations for the timeline of disease progression, together with empirical justification of the model assumptions and choice of parameters and then comparing the model with similar models to assess comparability. The bottom line is a judgement of 'sufficient' correspondence to the assessor's perception(s) of an unknown future reality.

Cost-utility Analysis

In the absence of a commitment to a protocol to evaluate claims expressed in cost-utility terms, the preference for utilities as an end-point (Section 3A.1) creates a further barrier to evaluation [25] . Utilities are not collected on a regular basis by health care systems as part, for example, of searchable electronic medical records to support systematic observation of outcomes for claims assessment. Given this, it seems odd that CADTH would focus on utilities as the preferred effectiveness measure as this would appear to guarantee that the outcome claims are (and always were intended to be) non-evaluable.

There is also the question of choice of utility measure or instrument and the possible need to crosswalk or map from one instrument to another. While the guidelines do not mandate a specific QALY measure, they do not recommend mapping from one instrument another. This means that, while generic QALYs are put forward as a gold standard for comparing outcomes across products and disease areas, thus 'facilitating resources based on maximizing health gains' modeled imaginary claims that rest on disparate QALY measures mean that such a comparisons is impossible. This same objection would apply if CADTH wished to apply willingness-to-pay cost-per-QALY thresholds.

Even if a single utility instrument is mandated, technology assessment agencies can still shoot themselves in the foot. Consider the NICE EQ-5D debacle. Concern with the sensitivity of the EQ-5D-3L to changes in the quality of life led to introducing five levels rather than three of severity: (i) no problems: (ii) slight problems; (ii) moderate problems; (iv) severe problems; and (v) extreme problems rather than (i) no problems; (ii) some problems; and (iii) extreme problems. Unfortunately, the application of the EQ-5D is now in doubt as the attempt to move to a more sensitive version of the instrument with five levels within each health domain (EQ-5D-5L) as opposed to the three levels (EQ-5D-3L) has led to a situation where these two versions produce substantially different estimates of cost-effectiveness [26]. This is because of the combined effect of differences in the way individuals respond to the changed descriptive system and the changed valuation system in the 5L compared to the 3L. The two versions are not consistent with each other. This problem is likely to plague NICE for years given the number of accepted modeled evaluations and claims for cost-per-QALY outcomes based on the EQ-5D-3L.

This result should not, in retrospect, have come as a surprise as there have always been difficulties in mapping between the various multi-attribute HRQoL instruments and comparing these to the more fundamental measures of standard gamble and time trade-off [27]. It also raises the intriguing possibility that manufacturers, in their submissions to CADTH, could select from the various utility instruments, selecting the one that gave the 'best' results to support cost-effectiveness claims and pricing.

Even if one accepts the case for non-evaluable 'for information only' imaginary claims, the CADTH reference cases stands or falls on the willingness to accept the manufacturers choice of a QALY measure. As CADTH is unwilling to nominate a preferred QALY (e.g., the HUI with Canadian preferences), the fact that within a disease area manufacturers with competing products can, by accident or design, opt for a particular multi-attribute instrument (e.g., some may choose the EQ-5D-3L; others the EQ-5D-5L) means that the cost-outcomes claims are not comparable. A situation which becomes more troublesome when, as NICE has found out, previously accepted submissions may have used the EQ-5D-3L. The picture becomes even more problematic, if not bizarre, when cost-per-QALY comparisons are between disease areas where 3 or 4 generic multi-attribute measures may be jostling for attention (e.g., the HUI vs. the nominated cancer generic measure (FACT-G). It would seem far easier to fall back on disease specific effectiveness measures. At least these would make sense to health system decision makers, physicians and patients.

Clinical Standards

There are two further issues that are not addressed in the CADTH guidelines that are important for moving from imaginary worlds to evaluable claims. These are: (i) evidence for the replication of clinical efficacy and safety claims for the pivotal phase 2 and phase 3 trials; and (ii) evidence of adherence to and persistence with therapy for the new product and the comparator. Clearly, attention needs to be given to how claims for efficacy and safety as well as

comparative adherence and persistence translate to a Canadian treating environment, but these are best seen as secondary considerations to the reproducibility of claims and compliance. As will be discussed below, these issues provide, along with claims for treatment effect, key elements in the case for protocol driven claims assessment and reporting.

Long-term Uncertainty

One defense of non-evaluable simulations is that concerns with a model or simulation as a 'sufficient' representation of the real world is taken account of by capturing structural and parameter uncertainty. This is not a tenable argument as it sidesteps the issue of credible claims and their evaluation. Any number of competing lifetime or long-term simulation could be constructed with account taken of structural and parameter uncertainty, accompanied by the usual tornado diagrams, cost-effectiveness acceptability curves and even value of information exercises to fill the gaps in the modeled claims. However much we try to embellish the modeled imaginary world the fact remains that we don't know, and we will never know, whether the announced claims for long-term or lifetime comparative therapeutic benefit are right or even if they are wrong.

Replication of Clinical Claims

The guidelines do not, as far as can be ascertained, address the issue of the replication of clinical claims. It is all well and good to ask for spreadsheet summaries of the relevant RCT data, supported, in the absence of head-to-head trials, with network meta-analysis and other assessments of comparative efficacy (but not effectiveness). Analyses that make a number of implicit assumptions regarding acceptance of the quality and replicability of the RCTs themselves. There is now an abundant literature on irreproducibility in scientific research, including Amgen's attempts to reproduce benchmark studies, Bayer's validation of new drug target claims and an increasing failure rate where phase 3 trials attempt to reproduce results [28] [29] . At the same time, pre-registration of National Heart, Lung and Blood Institute trials since 2000 has been associated with a decline in the reporting of positive results [30]. In a survey undertaken by the journal *Nature* over 70% of researchers have tried and failed to reproduce another scientist's experiments and over 50% have failed to reproduce their own experiments [31]. Given these concerns, it could be argued that they place a premium on the need to reproduce clinical claims in target Canadian patient populations.

Adherence and Persistence

A puzzling feature of lifetime reference model claims is the neglect of adherence and persistence behavior. The CADTH guideline is no exception. Nowhere in the instructions for constructing the imaginary reference case is there any guidance on including adherence and persistence behavior. This appears an odd oversight as the guideline is explicit in the requirement for a default lifetime model. If this implies that the hypothetical patient cohort are persistent with therapy over their lifetime, this flies in the face of decades of accumulated evidence which shows that by the end of two years from an index prescriptions, probably less than one third of patients are persistent with therapy with an even smaller proportion maintaining a adherence at a clinically meaningful level [32]. If this is the case, then to model competing therapies assuming full compliance over the lifetime of the patient cohort would seem pointless. A default timeline for models of two to three years would appear to be more reasonable if the intention is to inform decision makers. If more realistic disease specific imaginary worlds are mandated by CADTH, then there could presumably be a requirement for a systematic review of adherence and persistence patterns in that disease area for comparator therapies to set the appropriate model timelines.

Recently reported adherence and persistence patterns with the new oral anti-coagulants (NOACs) in non-valvular atrial fibrillation are an example of short term compliance [33] [34] [35] [36]. Overall, these estimates suggest that by the end of one year after the index prescription persistence with warfarin is in the range 60 to 70% with a corresponding NOAC rate of 70 to 80%. By the end of year 2, persistence is likely to be 15 to 20% lower. Beyond two years is sheer speculation, although it would not be unreasonable, given evidence for persistence in other chronic disease states to assume that the overwhelming majority of patients have discontinued within 3 to 4 years. Given the age at which treatment is usually initiated for atrial fibrillation, deaths to patients need to be factored in to persistence estimates. In the edoxaban pivotal trial, for example, 10-8% of patients died before the end of the trial. Other recent examples of short-term adherence and persistence behavior could be given. Examples would include multiple sclerosis where the majority of patients discontinue therapy within 2 years and statins in cardiovascular disease [37] [38].

This neglect of adherence and persistence on claims for clinical effect and cost-effectiveness, even in the construction of imaginary worlds, is puzzling given the attention to the impact of private insurance cover on prescription drug utilization in Canada. As outpatient prescriptions are outside the scope of the Canada Health Act, patients are often faced with significant out-of-pocket costs. Policies to meet these costs vary by province and the variability in insurance cover. These may include policies for 'catastrophic' drug costs as well as potential support for low income groups where drug-costs may consume a significant proportion of income. Law et al, in a study utilizing the 2007 Canada Community health survey, reported that 1 in 10 Canadians receiving a prescription reported cost-related non-adherence [39]. A more recent study by Kratzer et al utilizing the 2008 Canadian Community Health Survey found, for health cohorts in Ontario with asthma, high blood pressure or diabetes, that patients with private insurance were more likely to take prescribed drugs than those

without [40]. Most recently, with data from the Barriers to Care for People with Chronic Health Conditions, Hennessy et al reported on the association between the level of out-of-pocket spending and the likelihood of cost-related non-adherence [41]. The study found that those spending at least 5% of their income on prescriptions drugs were more likely to report cost-related prescription non-adherence than those spending less than 5%. Overall, from the various studies the proportion of individuals and families with low income and high cost non-adherence ranged from 5% to 10%. Although Canadian data are more limited in their ability to capture adherence and persistence behavior from an index prescription than in the US, the neglect of the impact of costs on adherence and persistence, together with the absence of any attempt to factor in likely patterns of adherence and persistence with drug classes and disease categories is a major oversight in the CADTH guidelines.

Lifetime Costs

A further puzzling feature is the assumption that, other than being discounted at a rate of 1.5% per annum, costs remain unchanged over the lifetime of the hypothetical patient cohort. This seems an odd requirement if the object is to present a model of long term or lifetime costs and benefits. In the US, for example, there is ample evidence for pricing strategies (price gouging?) by manufacturers for ongoing semi-annual and annual price increases both over the patent life of a drug and beyond. The price increases being typically accompanied by co-payment waivers, coupons and other discounts to maintain market share. In the case of disease modifying treatments (DMTs) in multiple sclerosis a recent study of the trend in annualized drug costs for nine DMTs from 1993 to 2014 found annualized increases ranging from 7.9%. to 35.7% [42]. Four of the DMTs had annualized cost increases greater than 20% and four in the range 13% to 16.8%. While these annualized changes were two to three time bigger than in other countries, the potential for annualized price increases together with possible price increases in direct medical costs should, presumably, be factored into reference case models. Lifetime cost increase assumptions may, of course be irrelevant if the majority of patients have discontinued therapy or report low rates of medication possession within 2 to 3 years from product listing.

Pipeline Competitors

It is unlikely, over the lifetime of a patient cohort, that there will be no therapies entering the market place to compete with and replace existing medications. In the reference case model this is not the case. Patients are assumed to remain with the indicated drug over their lifetime. Again, this flies in the face of evidence for drug turnover in target populations where patients are switched to new therapies. This switching may reflect a lack of response to the index drug in the treatment arm or may involve moving to a combination therapy. Again, this appears an odd assumption but one that is, unfortunately, driven by the focus on constructing evidence to establish credibility rather than on a more practical perspective of evaluating claims for feedback to physicians and formulary committees.

Next Generation Sequencing

The likelihood of competitor therapies and therapy combinations is also made more likely by the introduction of next generation sequencing (NGS) where assay platforms will recommend linking sub-groups of patients defined by mutation clusters to monotherapy or combination therapies [43 44]. An obvious application is in late stage cancer but there will be applications earlier in the treatment pathway. In these scenarios there will be a premium placed on umbrella RCTs and tracking evaluable claims and reporting in real time to clinicians and health system decision makers on clinical outcomes and resource utilization. If a disease area or target tumor group is characterized by a distribution of patients by mutation cluster then a 'simplistic' reference case model is hardly a viable basis for therapy choices when multiple pathways are involved and patients are individually selected for an assay driven intervention.

Redrafting the CADTH Guidelines

Although it is unlikely that the CADTH guidelines will be redrafted given their recent launch and their commitment to the reference case imaginary model paradigm as the basis for submissions by manufacturers, the present reference case format can be relatively easily modified. As detailed in the review of the PBAC guidelines, the modification can be accomplished by establishing a cut-off for the modeling of clinical and cost-effectiveness claims. The key issue is one of supporting feedback to formulary committees. In the Canadian case the principal audience would be provincial health department, and potentially insurance companies to ensure coverage of the product. The proposed cut-off could be two years. The modeled claims, whether they are extrapolated from clinical trials or represent a de novo model, would establish primary clinical and resource utilization endpoints, together with the proposed unit cost of the drug. The manufacturer's submission would propose, as part of the claims submission, a protocol to evaluate these claims within a maximum 2-year timeframe to report back to CADTH. The results would be posted to the CADTH website.

As detailed in previous commentaries in this series, together with the proposed Minnesota guidelines, the manufacturer is responsible for (i) establishing comparative claims for a new product that are credible, evaluable and replicable and (ii) underwriting a protocol that supports claims evaluation and reporting of results. It should be entirely at the manufacturer's discretion as to the choice of clinical and effectiveness endpoints. There would be no requirement for a reference case or for highly technical, lifetime probabilistic simulations. This does not mean that quality of life is necessarily put to one

side; rather it has to be considered and justified as a meaningful endpoint within a two year timeframe. If there are claims for long-term clinical outcomes, then these should be assessed via surrogate markers which can be assessed within 2 years. Focusing on a 2-year timeframe will reinforce claims for clinical impact. As the majority of RCTs and attempts to make indirect comparisons are typically short term, (with the typical RCT not extending beyond 6 months) extrapolating clinical benefits from a short timeframe will avoid trying to justify longer term, non-evaluable and more uncertain benefits.

As recommended in the review of the PBAC guidelines, extrapolated or modeled claims should be required to include anticipated adherence and persistence behavior. This behavior should be required to be reported as part of the study protocol with summary measures, for example, to include median time to discontinuation of therapy and proportion of patients maintaining a medication possession ratio (MPR) or days covered by therapy (DCT) > 0.8. An interesting corollary here is that if the benefits of a potential new therapy are qualified significantly by anticipated adherence and persistence behavior, then the manufacturer may propose an intervention strategy (funded by the manufacturer) to improve such behavior. This could be justified in cost-effectiveness terms where the benefits from improved adherence and compliance are set against the costs of implementing and monitoring the intervention.

CADTH, in redrafting the guidelines to accommodate a 2-year timeframe will have to judge the extent to which it makes sense to require submissions to incorporate the technical modeling standards that characterize the construction and reporting of claims from lifetime imaginary worlds. Rejecting attempts to model the long-term course of a disease does not mean throwing the bathwater out with the baby, but it may mean reconsidering the relevance not of simple decision frameworks but those frameworks such as Markov processes that are designed to support the construction of lifetime imaginary worlds. Many of the standards and processes detailed in the CADTH guidelines will be relevant to the construction of short-term claims models. Just as the reference case modeled outcomes can be expressed as total costs and outcomes, incremental costs and outcomes and incremental cost-effectiveness ratios, so short-term claims can be expressed in these terms.

In redrafting the CADTH guidelines the question of evaluating claims is not a 'one-off'. Agreeing a claims assessment protocol sets the stage for claims replication. With claims assessment within a 2-year timeframe, there is ample scope for evaluating feedback from a claims protocol. The protocol could be administered in a number of treatment settings to capture both heterogeneity in response and its determinants as well as more specific claims targeted at sub-populations; evaluations which may be better focused on modeled cost-effectiveness rather than cost-utility claims. A reassessment of secondary outcome claims (e.g., patient satisfaction, comorbidity outcomes) and more detailed assessments of the determinants of compliance behavior could be a requirement. As noted in the review of the PBAC guidelines, CADTH could consider sponsoring training programs for protocol development as rigorous and well managed protocols for either experimental or observational assessments.

Life is, of course, more interesting if clinical and cost-outcomes claims fail. Protocols should be judged on their rigor: how high is the bar for claims to be accepted? Is the protocol designed to minimize false positives? As noted above, there are ongoing concerns over the ability to replicate RCT claims, even when the claims are based on two well conducted clinical trials have been accepted by regulators. The concern of CADTH should be for the implications of including unreproducible or at least potentially dodgy clinical data as input to the modeled claims.

Conclusion

From the perspective of normal science, constructed evidence for product impact claims is not acceptable. Regardless of how decision modeling is defended by the application of validation standards, the treatment of uncertainty and the application of value of information techniques, in the last resort the model stands or falls on its ability to generate credible, evaluable and replicable predictions. The latest version of the CADTH guidelines fails to meet this standard. If these guidelines are to be seen as credible then they must abandon constructing imaginary worlds in favor of evaluable and replicable claims for comparative product performance; to reject intelligent design in favor of natural selection.

Unfortunately, the imaginary constructs of the *status quo* may be seen as the preferred option. After all, it is what everyone else does. Redrafting the guidelines will be seen as unrealistic; imposing an unnecessary burden on both manufacturers and CADTH. After all, is CADTH in a position to challenge the community of appointed pharmacoeconomic scholars? Irrespective of the potential benefits to patients and physicians, formulary and pricing decisions driven by the construction of imaginary worlds may be agreed by the parties, after 20years of CADTH submissions, to be mutually advantageous and the least troublesome option. Would Sgt. Preston, King and Rex approve?

References

1 Canada. Canadian Agency for Drugs and Technologies in Health. Guidelines for the Economic Evaluation of Health Technologies: Canada. Methods and Guidelines. Ottawa: CADTH, October 2017

2 United Kingdom. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London: Nice, April 2013.

3 Ireland. Health Information and Quality Authority. Guidelines for the Economic Evaluation of Health Technologies in Ireland. Dublin: HIQA; 2014

4 Zorginstituut Nederland. Guideline for economic evaluation in healthcare. Diemen: 2016

5 New Zealand. Pharmaceutical Management Agency. Prescription for Pharmacoeconomic Analysis: Methods for cost-utility analysis (v 2.1). PHARMAC, 2012.

6 Australia. Department of Health. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (PBAC). Canberra: September 2016

7 Academy of Managed Care Pharmacy. *Format for Formulary Submissions (Version 4).* AMCP: April 2016

8 Caro JJ, Briggs A, Siebert U et al. Modeling good research practices – Overview: A report of the ISPOR-SMDM Modeling good practices task force – 1. *Value Health*. 2012 (15:796-803

9 Eddy DM, Hollingworth W, Caro JJ et al. Model transparency and validation: A report of the ISPOR-SMDM modeling good research practices Task Force – 7. *Value Health*. 2012;15:843-850

10 European Network for Health Technology Assessment (EUnetHTA). Recommendations for Health Economic Evaluations. 2015

11 Langley PC. Supporting formulary decisions: The discovery of new facts or constructed evidence? *Inov Pharm*. 2016; 7(2): No.15

12 Langley P. Validation of modeled pharmacoeconomic claims in formulary submissions. *J Med Econ*. 2015;18(12):993-999

13 Langley PC. He ao pohewa: The PHARMAC Prescription for Pharmacoeconomic Analysis in New Zealand and the standards of normal science. *Inov Pharm* . 2016; 7(2): No. 13

14 Langley PC. Langley PC. Na domhain shamhlaíochta: formulary submission guidelines in Ireland and the standards of normal science. *Curr Med Res Opin*. 2016;32(5):1641-1644

15 Langley PC. Denkbeeldige Wereld. The new Dutch guideline for economic evaluations in healthcare. *Inov Pharm*. 2017;8(1): No. 4

16 Langley PC. Sunlit uplands: The genius of the NICE reference case. *InovPharm*. 2016;7(2): No. 12

17 Canada. Canadian Agency for Drugs and Technologies in Health. Guidelines for the Economic Evaluation of Health Technologies: Canada. Methods and Guidelines. Draft Report. Ottawa: CADTH, October 2016

18 Pigliucci M. Nonsense on stilts: How to tell science from bunk. Chicago: University of Chicago Press, 2010

19 Langley PC. Nullius in verba: Version 2.0 of the University of Minnesota, School of Social and Administrative Pharmacy Program, Proposed Guidelines for Formulary Evaluation. *Inov Pharm.*2016;7(4): No.16

20 Sweet B, Tadlock CG, Waugh W, et al. The WellPoint outcomes based formulary: enhancing the health technology assessment process. *J Med Econ* 2005;8:13-25

21 Langley PC. Recent developments in the health technology assessment process in Fulda TR, Wertheimer AI, eds. Handbook of Pharmaceutical Public Policy. New York: Pharmaceutical Products Press, 2007

22 Langley PC. Intelligent design and Imaginary Worlds in Cost-Effectiveness Claims: An overview of commentaries in *INNOVATION in Pharmacy* from July 2016 to February 2017. *Innov Pharm*. 2017;8(1):No. 22

23 Wootton D. The Discovery of Science. New York: Harper Collins, 2015

24 https://royalsociety.org/about-us/history/

25 Langley PC. Great Expectations: Cost-Utility Models as Decision Criteria. *Inov Pharm.* 2016;7(2): No. 14

26 National Institute for Health and Care Excellence. Decision Support unit. Comparing the EQ-5D-3L and 5L versions. What are the implications for cost-effectiveness estimates? NICE: March, 2017
http://www.nicedsu.org.uk/USERIMAGES/DSU_3L%20to%205L%20FINAL.pdf

27 Drummond M, Sculpher M, Torrance G et al. Methods for the economic evaluation of health care programmes. (3rd Ed). Oxford: University Press, 2005

28 Baker M. Biotech giant posts negative results. *Nature*. 2016;530:141

29 Lawton JS. Reproducibility and replicability of science and thoracic surgery. *J Thoracic Cardio Surg*. 2016;152(6):1489-1491

30 Kaplan RM, Irvin VL. Likelihood of null effects of large NHLBI clinical trials has increased over time. *Plos One*. 2015; 10(8):e0132382

31 Baker M. Is there a reproducibility crisis? *Nature*. 2016;533:452

32 Langley PC. Next Generation Sequencing Claims in Precision Medicine: Questions a Formulary Committee Should Ask. *Inov Pharm.* 2016;7(4): No.11

33 Yao X, Abraham NS, Alexander GC et al. Effect of adherence to oral anticoagulants on risk of stroke and major bleeding among patients with atrial fibrillation. *J Am Heart Assoc*. 2016;5:e003074 doi: 10.1161/JAHA.115.003074

34 Forslund T, Wettermark B, Hjemdahl P. Comparison of treatment persistence with different oral anticoagulants in patienbts with atrial fibrillation. *Eur J Clin Pharmacol*. 2016;72(3):329-38

35 Beyer-Westendorf J, Enliken B, Evers T. Real-world persistence and adherence to oral coagulation for stroke risk reduction in patients with atrial fibrillation. *Europace*. 2016 Jan 31. Pii:euv421

36 Martinez C, Katholing A, Wallenhorst C et al. Therapy persistence in newly diagnosed non-valvular atrial fibrillation treated with warfarin or NOAC. A cohort study. *Thromb Haemost*. 2015;115(1):31-9

37 Higuera L, Carlin CS, Anderson S. Adherence to disease modifying therapies for multiple sclerosis. *J Man Care Spec Pharm*. 216;22(12):1394-1401

38 Jörntén-Karlsson M, Pintat S, Molloy-Bland M et al. Patient-centered interventions to improve adherence to statins: A narrative synthesis of systematically identified studies. *Drugs*. 2016;76(15):1447-1485

39 Law MR, Cheng L, Dhalla IA et al. The effect of cost on adherence to prescription medications in Canada. *CMAJ*. 2012;184(3):297-3022

40 Kratzer J, Cheng L, Allin S et al. The impact of private health insurance coverage on prescription drug use in Ontario, Canada. *Healthcare Policy*. 2015;1)(4): 62-74

[41] Hennessy D, Sanmartin C, Ronksley P et al. Out-of-pocket spending on drugs and pharmaceutical products and cost-related prescription non-adherence among Canadians with chronic disease. *Statistics Canada Health Reports*. 2016;27(6):3-8

[42] Hartung D, Bourdette D, Ahmed S et al. The cost of multiple sclerosis drugs in the US and the pharmaceutical industry. Too big to fail? *Neurology*. 2015:84:2185-2192

[43] Langley PC. Next Generation Sequencing Claims in Precision Medicine: Questions a Formulary Committee should ask. *Inov Pharm.* 2016;7(4): No. 11

[44] Langley PC. Creative Destruction: Next Generation Sequencing in Drug Development, Formulary Evaluations and Pricing. *Inov Pharm.* 2016;7(4): No. 13