

1-10-2017

# Dreamtime: Version 5.0 of the Australian Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee (PBAC)

Paul C. Langley

*University of Minnesota*, [langley@maimonresearch.com](mailto:langley@maimonresearch.com)

Follow this and additional works at: <http://pubs.lib.umn.edu/innovations>

## Recommended Citation

Langley PC. Dreamtime: Version 5.0 of the Australian Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee (PBAC). *Inov Pharm*. 2017;8(1): Article 5. <http://pubs.lib.umn.edu/innovations/vol8/iss1/5>

*INNOVATIONS in pharmacy* is published by the University of Minnesota Libraries Publishing.

## Dreamtime: Version 5.0 of the Australian Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee (PBAC)

Paul C Langley, PhD

Adjunct Professor, College of Pharmacy, University of Minnesota

### Abstract

*In September 2016 the Australian Department of Health published Version 5.0 of the Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee (PBAC). These guidelines, which were first published for comment in 1990, set out how to prepare a submission to list a new medicine or medicinal product on the Pharmaceutical Benefits Schedule (PBS). The guidelines give instructions on the information required by the PBAC and the Economic Sub-Committee (ESC), the most appropriate form for presenting clinical evidence and the standards for an economic evaluation. 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 claims presented to support PBS listing credible, evaluable and replicable. The review concludes that the PBAC guidelines do not meet the standards expected in normal science. The absence of empirically evaluable claims means there is no way of judging whether they are right or even if they are wrong. If the Guidelines were never intended to support experimentation and systematic observation to generate feedback to health system decision makers, then this should be made clear by the PBAC. If not, then consideration should be given to redrafting the guidelines to ensure they conform to these standards. Hopefully, future versions of the guidelines will address this issue and focus on a rigorous research program of claims assessment and feedback.*

**Keywords:** PBAC Guidelines, economic evaluations, imaginary worlds, simulations

### Introduction

Australia, in common with countries including the UK, Ireland, the Netherlands and New Zealand, has issued guidelines to support formulary submissions <sup>1 2 3 4 5</sup>. 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) <sup>6 7 8</sup>. Guidelines have also been proposed as a standard for the European Union under the EUnetHTA umbrella <sup>9</sup>. A common feature of these guidelines is to mandate a reference case where the recommended decision framework for chronic disease tracks a hypothetical cohort of patients over their lifetime. Comparative product claims are expressed as cost utility quality adjusted life years (QALYs).

Previous publications and formulary evaluation commentaries in this series have made clear that comparative product claims expressed in lifetime cost-per-QALY terms put to one side a commitment to the standards

of normal science. These standards require modeled claims or hypotheses to be credible, evaluable and replicable. <sup>10 11 12 13 14</sup>.

A lifetime cost-per-QALY model is not designed to generate evaluable claims <sup>15</sup>. It is a construct that is defended by its sufficient correspondence to 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.

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 <sup>16</sup>. This commitment to 'informing' decision makers through the construction of imaginary worlds is seen in the recently released draft of the 4<sup>th</sup> edition of the *Guidelines for the Economic Evaluation of Health Technologies: Canada* where it is made clear that 'Economic evaluations are designed to inform decisions. As such, they are distinct from conventional research activities, which are designed to test hypotheses' <sup>17</sup>.

In an effort to meet the standards of normal science, guidelines have been proposed by the Program in Social and Administrative Pharmacy at the University of Minnesota that

**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

Tel: 520-577-0436; Email: [langley@maimonresearch.com](mailto:langley@maimonresearch.com)

Web: [www.maimonresearch.net](http://www.maimonresearch.net)

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<sup>18</sup>. 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<sup>19 20</sup>.

The purpose of this commentary is to consider the economic evaluation standards of Version 5.0 of the PBAC guidelines from the perspective of normal science: 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 that supports 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 purpose of version 5.0 of the PBAC guidelines is to guide the preparation of a submission to the PBAC for the public funding of a new medicine or medicinal product under the Pharmaceutical Benefits Scheme (PBS). The structure of a major submission covers:

- Submission Executive Summary
- Section 1: Context
- Section 2: Clinical Evaluation
- Section 3: Economic Evaluation
  - Section 3A: Cost-effectiveness analysis
  - Section 3B: Cost-minimization analysis
- Section 4: Use of Medication in Practice
- Section 5: Options to present additional relevant information

The focus of this commentary is on (i) Section 2 - standards for clinical evaluation and (ii) Section 3A - standards for economic evaluation.

#### Standards for Clinical Evaluation

Section 2 of the PBAC guidelines focuses on the collation of the best clinical evidence to support the effectiveness and safety of the proposed medicine in its approved indication. This assessment has eight components:

- a systematic literature search to identify relevant clinical trials or studies (sub-sections 2.1-2.2)

- an analysis and interpretation of results of each included trial to include comparative treatment effect relative to comparators and credibility of findings (sub-sections 2.3-2.5)
- additional analyses of comparative treatment effect (in the absence of estimates from the whole trial population) (sub-section 2.6)
- an assessment of the applicability of treatment effect evidence for the target Australian population (sub-section 2.7)
- therapeutic conclusions for efficacy and safety versus comparator (sub-section 2.8)

This conclusion from sub-section 2.8 forms the basis for the economic evaluation in Section 3.

#### *Sub-section 2.1 Literature Search Methods*

This details the search methods to capture all relevant randomized trials (or nonrandomized studies) for the clinical evaluation. The focus is on randomized trials that compare the proposed medicine with the main comparator. If there are no direct randomized comparisons the search will identify randomized trials for indirect comparison. If there are no indirect comparisons, the search will focus on nonrandomized studies.

#### *Sub-section 2.2 Identification of Relevant Trials*

This section details the search for and identification of relevant trials. Steps required are to: (i) present results using a PRISMA flowchart; (ii) list all trials indicating which are included/excluded with reason for exclusion; (iii) create a master list of included trials; (iv) if applicable, identify trials used for indirect comparisons; (v) describe how the included trials are used to support the clinical claim; and (vi) attach copies of trials.

#### *Sub-section 2.3 Trial design and execution*

The focus of this sub-section is on the risk of bias, internal validity, to identify those trials with the greatest scientific rigor. Information requested, with results presented in tabular form, focuses on the design and conduct of the trial (e.g., group allocation, blinding). Risk assessments are also to be undertaken for non-randomized studies, together with efforts for mitigating risk. The ROBINS-I tool is recommended for assessment purposes.

#### *Sub-section 2.4 Trial characteristics*

This sub-section details the individual trial characteristics: (i) eligibility criteria and patient/subgroup characteristics; (ii) treatments in each arm; (iii) primary and patient relevant outcomes; (iv) minimally important difference for outcomes;

(v) non-inferiority margins and (vi) cross-reference to source documents.

#### **Sub-section 2.5 Trial results for whole trial population**

This sub-section reports the results for the whole trial population. Information is required on: (i) effectiveness results for each trial for relevant outcomes; (ii) adverse events; and (iii) cross reference to source documents.

#### **Sub-section 2.6 Trial results: additional analyses**

Additional analyses include sub-group assessments, meta-analyses, indirect comparisons and adjustments for treatment switching.

#### **Sub-section 2.7 Assessment of differences between the trial setting and the Australian setting after listing**

This sub-section explores possible differences between the observed comparative benefits and harms in the trial setting, and the benefits and harms that are likely to occur in the Australian setting after listing on the PBS.

#### **Sub-section 2.8 Interpretation of the clinical evidence**

This sub-section is designed to (i) summarize the overall clinical trial evidence and (ii) present the overall therapeutic conclusion for comparative effectiveness and safety. The summary should address (without repeating earlier assessments):

- the level of evidence
- the quality of evidence
- clinical and patient importance of effectiveness and safety outcomes
- statistical precision of the evidence
- effect size
- consistency in clinical trial results
- results

This sets the stage for the choice of economic evaluation. It is critical in determining the success of the submission. The therapeutic conclusion should be succinct detailing (i) whether the product is superior/noninferior/inferior versus the comparator and (ii) superior/noninferior/inferior in its safety versus the comparator.

#### **Standards for Economic Evaluation**

Section 3 of the PBAC guidelines sets the standards for an economic evaluation where the proposed medicine or medicinal product is substituted for the main comparator. The guidelines note that the information requested is to cover a full and transparent description of the evaluation with sensitivity analyses to capture uncertainty. The evaluation can be either a cost-effectiveness (Section 3A) or a cost

minimization analysis (Section 3B). The former applies where a therapeutically superior product involves higher costs or where an inferior product to the main comparator results in lower costs. This analysis requires a quantitative assessment of incremental costs and outcomes expressed as a cost-effectiveness ratio. The focus here is on section 3A.

#### **Section 3A: Cost-Effectiveness Analysis**

The preference of the PBAC is for economic evaluations based on results from randomized clinical trials with any adjustments or additions to account for differences in the population and setting, timeframe of analysis or outcomes of interest. Evidence standards for the cost-effectiveness analysis and submission are detailed in nine sub-sections. These are:

- Overview and rationale of the economic evaluation (Section 3A.1)
- Computational methods and structure of the economic evaluation (Section 3A.2)
- Population and setting (Section 3A.3)
- Model transition probabilities or variables, transformation and extrapolation (Section 3A.4)
- Health Outcomes (Section 3A.5)
- Health care resource use and costs (3A.6)
- Model validation (Section 3A.7)
- Results of the base case economic evaluation (Section 3A.8)
- Uncertainty analysis: model inputs and assumptions (Section 3A.9)

#### **Section 3A.1 Overview and rationale of the economic evaluation**

The information required to give an overview includes an identification of the key components of the economic evaluation (in a decision framework), a justification of the type of analysis, choice of outcome measures and the primary decision that is being addressed. The perspective in the base-case model should be that of the Australian health system; wider societal impacts can be introduced as supplementary analyses. Claims can be made based upon clinical trials. Modeled claims would either extrapolate from the trial or be developed as a stand-alone model. There needs to be a summary of steps translating from trial to model. Discounting should be applied for models that extend beyond one-year.

Key components of the economic evaluation include type of analysis, outcomes, time horizon, method(s) used to generate results, health states, cycle length, transition probabilities and software.

The preferred modeling framework is cost-utility particularly when: (i) there is a claim of incremental life-years gained; (ii) there is an improvement in quality, but not quantity, of life; and (iii) relevant randomized trials report results using a multi-attribute utility instrument (MAUI). If a cost-effectiveness model is used a case should be made for not translating outcomes to utilities. If a cost-consequences or a cost-benefit analysis is presented it should be secondary to the base case cost-utility or cost-effectiveness model.

### **Section 3A.2 Computational methods and structure of the economic evaluation**

This section details four issues that the submission should address: (i) reporting a supplementary literature review for additional clinical or epidemiological studies; (ii) reporting and justifying the model structure, and the time horizon; (iii) describing the modeling technique and (iv) providing an electronic copy of the model.

The model structure should capture all relevant health states or clinical events along a disease, ensuring it is consistent with treatment algorithms. The model structure should be informed by the results of the literature review of economic evaluations and other clinical and economic literature. The time horizon for the model should be defined and justified. It should capture all important differences in outcomes between the intervention and comparator but not to extend unnecessarily beyond this. If there are no mortality implications and only temporary quality of life or health effects a short term horizon may be appropriate. A lifetime time horizon is appropriate if a treatment affects mortality or long term/ongoing quality of life. The validity and plausibility of the time horizon is determined by the population of the model and the realism of the inputs.

### **Section 3A.3 Population and setting**

The setting of the economic evaluation should be the Australian health care system, with the modelled population: (i) representing the target Australian population indicated for use of the proposed medicine; (ii) with use consistent with the treatment guideline; and (iii) any proposed restrictions on access to the drug. Relevant patient and clinical characteristics may include age, sex, ethnicity, medical condition and severity of the medical condition, and comorbidities. The submission should detail which patient characteristics are incorporated explicitly and which are implicit (associated with use of other data) or not included. The submission should detail also how heterogeneity in patient characteristics (if relevant) are to be captured in the cost-effectiveness analysis.

### **Section 3A.4 Model transition probabilities or variables, transformation and extrapolation**

In state transition frameworks the transition probabilities should be detailed together with other modeled inputs in the base-case model. Any translation from surrogate to target clinical outcomes should be justified and described, as should any extrapolations from trial data. Alternative input data should be available to support sensitivity or scenario analyses.

### **Section 3A.5 Health Outcomes**

The submission should justify the choice of the final health outcome for comparative clinical impact, to include quality of life or utility scores. The utilities should be presented as point estimates (with corresponding standard deviations and 95% confidence intervals) for each health state in the model. If a MAUI has been used to assign utility weights, the scoring algorithm should be reviewed to determine whether the preference weights are relevant in an Australian treating environment. Any mapping of a patient reported outcome measure to preference weights has to be validated. If this is not possible, then it should be demonstrated that the outcome measures are valid for the target population. Using different published studies to obtain utility weights from the literature is discouraged.

### **Section 3A.6 Health care resource use and costs**

Relevant health care resource items include:

- medicines (direct costs of treatment and medicines used to treat adverse reactions)
- medical services, including procedures
- hospital services
- diagnostic and investigational services
- community-based services
- any other direct medical costs.

In constructing the model, the submission should define the natural units and quantify the number of natural units of the resource item provided to patients in each treatment group or for each health state patient's encounter. All unit prices and costs should be in Australian dollars with a consistent year of analysis. Future costs are valued at current prices consistent with using constant prices in the economic evaluation (there should be no allowance for inflation or other possible future price changes).

### **Section 3A.7 Model validation**

The model should be validated operationally to support the generated results. This may involve tracing patient pathways through the model to establish face validity. For external validity model traces may be compared with corresponding empirical data to check consistency of outcome. Comparing

the model outcomes against those for similar models can establish cross-validity. Validation should include both data sources used in the model (dependent validation) and data sources not used in the model (independent validation).

### **Section 3A.8 Results of the base case economic evaluation**

For the base case model, the submission should provide:

- a calculation of proposed medication cost per patient
- a stepped presentation of the cost-effectiveness results and the incremental cost-effectiveness ratio (i.e., the key steps involved in transforming the comparative clinical data into the modelled base-case estimate of incremental cost-effectiveness).
- disaggregated and aggregated costs and outcomes
- a summary of the base-case estimate of the incremental cost-effectiveness ratio

### **Section 3A.9 Uncertainty analysis: model inputs and assumptions**

The submission should provide:

- a review of the methods to capture uncertainty for inputs, translations of inputs and model structure
- a univariate sensitivity and scenario analysis
- a review of relevant multivariate analyses and any probabilistic sensitivity analyses
- a summary of the results of the uncertainty analysis

### **Discussion**

There is no requirement in the PBAC guidelines for claims for competing products to be presented as 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 apparent 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.

Although the guidelines do not mandate a reference case standard, it is clear that in the case of chronic disease that a reference standard applies. The approved analytical techniques for modeling and the presentation of results are consistent with the reference case approach and are no different from standards which explicitly take a reference case approach in other single payer health systems. This means that the PBAC guidelines face the same criticisms that those for the UK, Ireland, New Zealand and the Netherlands face: we have no idea whether the claims, irrespective of arguments that the

model has been validated (Section 3A.7), are right or even if they are wrong.

### **Constructed Evidence**

Unfortunately, in the absence of claims that are evaluable and replicable, there is no assurance that they are credible. To argue that they meet these standards, in the absence of any empirical evidence to support the claims, is to adopt a relativist position. For a relativist evidence is never discovered, only constructed within a particular social community<sup>21</sup>. In a community of health economists who accept a reference case paradigm to support claims for competing pharmaceutical products, relativists would 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 modeling is seen as the 'gold standard'. The 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)<sup>22</sup>. 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.

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 the PBAC for modeling the base case outside of short-term or trial based assessments points to a substantial misallocation of time and resources to justify a cost-utility model that makes a claim for sufficiency in an Australian population. Yet, at the end of the day an effort, with the highly prescriptive and probably unnecessarily exhaustive requirements of the guidelines which generate, unevaluable claims. As an example, the effort put into developing Markov models with utility values applied to each health state over a lifetime (Section 3A.2), particularly when, as noted below, the patient has most likely discontinued that therapy within 2 to 3 years of the index prescription.

### **Immunity to Failure**

While simulations can fail, lifetime cost-per-QALY modeled claims are immune to failure. 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 reverse engineering a model 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. As detailed below, neglect of anticipated adherence and persistence behavior (which may differ between target and comparator arms) may be 'sufficient' to underline the credibility of competing lifetime or long-term models. Indeed, any number of competing simulations could be constructed around compliance patterns yet meet PBAC standards. The same argument would hold, again detailed below, for assumption as to long term drug pricing and other direct medical cost expenditures. In the absence of evaluable claims there is no way competing models can be adjudicated and re-assessed. There is no 'progress' in our understanding of the impact of competing therapies in the Australian population.

### **Validation**

Irrespective of the extent to which model builders claim that they are a valid construct and have a pivotal role to play in informing decision makers, the validation standards proposed in the PBAC guidelines make no claim for validation that involves empirical evaluation in target treating populations. Section 3A.7 of the guidelines refers only to tracking patients through the logic of the model, justifying assumptions empirically and comparing the modeled outcomes to those from similar models. The question of the feasibility of empirical assessment does not arise.

It is worth noting that the PBAC guidelines address the issue of plausibility in the timelines for a modeled assessment (Section 3A 2.2). The validity of the lifetime horizon is not independently nominated but is determined by the population of the model, and the inputs. If the timeline is implausible, the PBAC argues, inputs that are not realistic will result in the model predicting a duration of outcomes or survival, estimates of QALYs saved and incremental cost-per-QALY savings that are themselves implausible. If this argument is applied, as noted below, to adherence and persistence with target

medicines then it leaves little scope for long-term or lifetime models. A systematic review of the compliance literature for adherence and persistence patterns in the target disease state could, presumably, be a key criterion for justifying a model's timelines. This issue is not addressed.

### **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<sup>23</sup>. Utilities are not collected on a regular basis by health care systems as part, for example, of searchable electronic medical records. There is also the question of choice of utility measure or instrument and the possible need to crosswalk from one instrument to another. As detailed above, claims expressed in lifetime cost-per-QALY terms are unevaluable; they are imaginary constructs with no claim to credibility. It is worth noting that the PBAC guidelines do not mandate a reference cost-per-QALY standard or the application of a preferred QALY instrument. It is not clear from the PBAC guidelines how competing submissions within a disease area utilizing competing quality of life instruments are to be evaluated.

The issue of following the NICE model and applying cost-per-QALY willingness-to-pay thresholds as a resource allocation and pricing tool has, apparently, not arisen. If cost-per-QALY thresholds (e.g., AUS\$50,000 per quality life year saved) are not decision criteria then the application of lifetime cost-utility modeling standards, particularly where there is no standard instrument, seems redundant.

### **Clinical Standards**

Although the clinical standards for evidence to support overall therapeutic claims for the product detailed in Section 2 of the guidelines are appropriate (if not overly prescriptive), there are two issues that are not addressed. 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 an Australian treating environment, but these are best seen as secondary considerations to the reproducibility of claims (see below) and compliance. As will be discussed below, these issues provide, along with claims for treatment effect in an Australian setting, key elements in the case for protocol driven claims assessment and reporting.

### **Long-term Uncertainty**

One possible 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 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), but that makes 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 Phase 2 results<sup>24 25</sup>. 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<sup>26</sup>. 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<sup>27</sup>. Given these concerns, it could be argued that they place a premium on the need to reproduce clinical claims in target Australian patient populations.

### **Adherence and Persistence**

A puzzling feature of lifetime reference model claims is the neglect of adherence and persistence behavior. As noted, the PBAC guideline is no exception. Nowhere in the instructions for constructing the imaginary reference case is there any guidance on including (possibly imaginary) 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<sup>28</sup>. If this is the case, then to model competing

therapies assuming full compliance over the lifetime of the patient cohort would seem pointless.

In the case of non-valvular atrial fibrillation, to give an example, a recent study by Yao et al compares adherence patterns for warfarin with those for rivaroxaban, dabigatran and apixaban and their impact on risk of stroke and major bleeding<sup>29</sup>. During a median follow up of 1.1 years, only 47.5% of NOAC patients were adherent, defined as a medication possession ratio (MPR) of  $\geq 80\%$ . Adherence to warfarin was 40.2%. Apixaban had the highest unadjusted adherence (61.9%) and dabigatran the lowest (38.5%). The rivaroxaban rate was 58.4%. Applying a multivariate logistic regression, adjusted adherence rates were 38.7% for warfarin and 47.5% for all NOACs. Higher rates of adherence were found across all treatments for those at higher risk. For those with a CHA<sub>2</sub>DS<sub>2</sub>-VASc  $\geq 4$  the warfarin adherence was 53.4% and the average for the NOACs 59.8%.

Persistence with NOACs has been reported in three recent observational studies. Forslund et al utilizing data from the administrative health register of the Stockholm region evaluated crude and adjusted persistence from the index OAC prescription. In the period April 2011 to December 2014<sup>30</sup>. At the end of the first year crude overall persistence was 88.2% and 82.9% at the end of the second year. Persistence with warfarin at the end of the first year was 85.0%, apixaban 85.9%, dabigatran 74.4% and rivaroxaban 77.4%. In the UK, Martinez et al reported on persistence with longitudinal data from the Primary Care Clinical Practice Research Datalink between January 2011 and May 2014<sup>31</sup>. Persistence with warfarin at the end of the first year was 63.6% and 79.2% for all NOACs. In Germany, Beyer-Westendorf et al reported persistence from primary care patients at 180 days of 66.0% for rivaroxaban, 60.3% for dabigatran and 58.1% for VKA. At 1-year corresponding persistence estimates were 53.1%, 47.3% and 25.45% respectively. An MPR  $\geq 0.8$  was found for 61.4% of rivaroxaban and 49.5% of dabigatran patients<sup>32</sup>.

Longer term studies suggest that by 3 years from index prescriptions no more than 30% of patients met the standard of  $\geq 80\%$  days covered. There are limited data for longer periods. Experience in Australia, for example, in the period November-December 2013 to March 2015 with records from the Pharmaceutical Benefits Scheme reported by Simons et al found that for index prescriptions in a sample of 1,471 atrial fibrillation patients with a mean age of 76 years on NOACs and 74 years on warfarin found that 62% discontinued within 12 months<sup>33</sup>. The corresponding figure for NOACs was 30%. Overall, 9% of those on NOACs failed to pick up the first repeat prescription compared to 14% of those on warfarin.

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 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. Under reasons for discontinuation death was given in 3.1% of warfarin patients and 2.8% of edoxaban patients.

It is unlikely that patterns of adherence and persistence in Australian target populations will be different from those reported in the US and Europe. Even so, estimates of adherence and persistence vary and it would be appropriate not only to include estimates in modeled claims but to evaluate assumptions in treatment settings.

**Lifetime Costs**

A further puzzling feature is the assumption that current prices and costs of resource inputs (corrected for discounting) remain unchanged over the lifetime of the hypothetical patient cohort (Section 3A.6) . The PBAC guidelines ask that no account be taken of possible inflation and, by implication, any allowance for the future prices of any resource item. Again, 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 again 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 increases ranging from 7.9%. to 35.7% <sup>34</sup>. 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 on the PBS.

Changing government policies towards drug pricing under the PBS in Australia are difficult to predict. It should not be

assumed that PBS ‘protection’ would reduce price increases overall and over the lifetime of the treating cohort. The cost of a drug would include both co-payments over and above PBS coverage as well as off-label use (e.g., in cancer protocols) where funding is absent <sup>35</sup>.

**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 <sup>36</sup>. 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 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 Australian Guidelines**

A possibly surprising conclusion, given the concerns expressed above as to the shortcomings of the latest PBAC guidelines, is that the present format can be relatively easily refocused to support claims that are credible, evaluable and replicable. The critical step is to revisit modeled study timelines. Putting in place two requirements that: (i) claims must be evaluated within a two-year timeframe from PBS listing; and (ii) that the PBAC submission should be accompanied by a protocol detailing how the claims are to be assessed would effectively overcome claims that the guidelines lack scientific rigor.

The timeline flexibility is recognized in the guidelines. Those making a submission are asked to justify the modeled timeline assumptions whether these are extrapolated from an RCT or are part of a *de novo* model framework. Redrafting would require a statement that the preferred timeline should be no more than 2-years. If a manufacturer wishes to argue that the anticipated benefits from their product against the existing standard of care will only be realized in a longer timeframe, then they should be asked to submit a protocol detailing short-term surrogate markers. This is unlikely to occur if, as the evidence would suggest across the majority of disease states, that within two years the majority (if not the overwhelming majority) of patients will have discontinued therapy. A lifetime perspective is, quite clearly, redundant if any benefits from a new product over a comparator are to occur while patients are clinically adherent to therapy.

To reinforce the preferred 2-year timeframe requirement, 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. Clearly, factoring in adherence and persistence will qualify long-term claims for comparative product performance expressed in terms of discounted QALYS and costs per QALY adjusted life years. This is unavoidable yet desirable. It does not mean that QALYs and the application of cost-utility models are necessarily abandoned but that submissions to the PBAC should make the case for their relevance within a 2-year timeframe.

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. This may be as simple as linking patients to an app that reminds them to continue their medications. Indeed, the PBAC may even reject a submission in the absence of an adherence and persistence strategy.

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 and more uncertain benefits.

The PBAC, 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 PBAC 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. Again, the present guidelines give considerable scope over the choice of decision model framework as long as long-term or lifetime timelines are put to one side.

Adopting a short term time horizon will mean that the PBAC will have to drop quality of life as the preferred outcome. In a timeframe as short as two years claims expressed in quality life years saved may make little sense. The exception here may be claims in end-stage cancer interventions or other disease states where the focus is on months rather than years in patient survival. Otherwise, focusing on clinical and cost-effectiveness claims where endpoints are defined as clinical markers will not only support claims replication from phase 3 trials but may be more relevant to patients and physicians.

In redrafting the 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. The PBAC could also consider sponsoring training programs for protocol development as rigorous and well managed protocols for either experimental or observational are the key to effective 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 the PBAC should be for the implications of including unreproducible or at least potentially dodgy clinical data as input to a speculative long-term cost-per-QALY model to support lifetime cost-per-QALY claims. A 2-year assessment timeframe would avoid promoting clinical and cost-effectiveness claims that are immune to failure. At the same time a protocol would allow a review of claims for clinical effectiveness in target Australian populations.

### 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. To argue that constructed imaginary worlds can inform decision makers is to put aside the standards of normal science in favor of claims which can never be evaluated. The latest version of the PBAC guidelines fail 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.

Unfortunately, the imaginary constructs of the *status quo* in Australia may be seen as the preferred option. Redrafting the guidelines will be seen as unrealistic; imposing an unnecessary burden on both manufacturers and the PBAC. 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 25 years of PBAC submissions, to be mutually advantageous and the least troublesome option.

### References

- <sup>1</sup> Australia. Department of Health. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (PBAC). Version 5.0. Canberra September 2016
- <sup>2</sup> United Kingdom. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London: Nice, April 2013.
- <sup>3</sup> Ireland. Health Information and Quality Authority. Guidelines for the Economic Evaluation of Health Technologies in Ireland. Dublin: HIQA; 2014
- <sup>4</sup> Zorginstituut Nederland. Guideline for economic evaluation in healthcare. Diemen: 2016
- <sup>5</sup> New Zealand. Pharmaceutical Management Agency. Prescription for Pharmacoeconomic Analysis: Methods for cost-utility analysis (v 2.1). PHARMAC, 2012.
- <sup>6</sup> Academy of Managed Care Pharmacy. *Format for Formulary Submissions (Version 4)*. AMCP: April 2016
- <sup>7</sup> 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
- <sup>8</sup> 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
- <sup>9</sup> European Network for Health Technology Assessment (EUnetHTA). Recommendations for Health Economic Evaluations. 2015
- <sup>10</sup> Langley PC. Langley PC. Supporting formulary decisions: The discovery of new facts or constructed evidence? *Inov Pharm*. 2016; 7(2): No.15
- <sup>11</sup> Langley P. Validation of modeled pharmacoeconomic claims in formulary submissions. *J Med Econ*. 2015;18(12):993-999

- <sup>12</sup> 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
- <sup>13</sup> 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
- <sup>14</sup> Langley PC. Denkbeeldige Wereld. The new Dutch guideline for economic evaluations in healthcare. *Inov Pharm*. 2017;8(1): No.4
- <sup>15</sup> Langley PC. Sunlit uplands: The genius of the NICE reference case. *InovPharm*. 2016;7(2): No, 12
- <sup>16</sup> Pigliucci M. Nonsense on stilts.: How to tell science from bunk. Chicago: University of Chicago Press, 2010
- <sup>17</sup> Canadian Agency for Drugs and Technologies in Health (CADTH). Guidelines for the Economic Evaluation of Health Technologies: Canada (4<sup>th</sup> edition; draft). Ottawa: October 20, 2016.
- <sup>18</sup> 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
- <sup>19</sup> 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
- <sup>20</sup> 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
- <sup>21</sup> Wootton D. The Discovery of Science. New York: Harper Collins, 2015
- <sup>22</sup> <https://royalsociety.org/about-us/history/>
- <sup>23</sup> Langley PC. Great Expectations: Cost-Utility Models as Decision Criteria. *Inov Pharm*. 2016;7(2): No. 14
- <sup>24</sup> Baker M. Biotech giant posts negative results. *Nature*. 2016;530:141
- <sup>25</sup> Lawton JS. Reproducibility and replicability of science and thoracic surgery. *J Thoracic Cardio Surg*. 2016;152(6):1489-1491
- <sup>26</sup> Kaplan RM, Irvin VL. Likelihood of null effects of large NHLBI clinical trials has increased over time. *Plos One*. 2015; 10(8):e0132382
- <sup>27</sup> Baker M. Is there a reproducibility crisis? *Nature*. 2016;533:452
- <sup>28</sup> Langley PC. Next Generation Sequencing Claims in Precision Medicine: Questions a Formulary Committee Should Ask. *Inov Pharm*. 2016;7(4): No.11.
- <sup>29</sup> 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
- <sup>30</sup> Forslund T, Wettermark B, Hjemdahl PO. Comparison of treatment persistence with different oral anticoagulants in patients with atrial fibrillation. *Eur J Clin Pharmacol*. 2016;72(3):329-38
- <sup>31</sup> 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
- <sup>32</sup> 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

- 
- <sup>33</sup> Simons LA, Ortiz M, Freedman SB et al. Improved persistence with non-vitamin K oral anticoagulants compared with warfarin in patients with atrial fibrillation: recent Australian experience. *Curr Med Res Opin.* 2016; Jul 27:1-14 (epub]
- <sup>34</sup> 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
- <sup>35</sup> Mellor JD, Van Koeeverden P, Yip SW et al. Access to anticancer drugs: many evidence-based treatments are off-label and unfunded by the Pharmaceutical Benefits Scheme. *Intern Med J.* 2012;42(11):1224-9
- <sup>36</sup> Langley PC. Creative Destruction: Next Generation Sequencing in Drug Development, Formulary Evaluations and Pricing. *Inov Pharm.* 2016;7(4): No. 13