

7-8-2016

Supporting Formulary Decisions: The Discovery of New Facts or Constructed Evidence?

Paul C. Langley

University of Minnesota, langley@maimonresearch.com

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

Recommended Citation

Langley PC. Supporting Formulary Decisions: The Discovery of New Facts or Constructed Evidence?. *Inov Pharm*. 2016;7(2): Article 15. <http://pubs.lib.umn.edu/innovations/vol7/iss2/15>

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

Supporting Formulary Decisions: The Discovery of New Facts or Constructed Evidence?

Paul C Langley, PhD

College of Pharmacy University of Minnesota, Minneapolis MN

Abstract

A critical question, given the growing importance of more targeted therapies to support personalized and precision medicine, is the credibility of the evidence base to support formulary decisions and pricing. On the one hand, for those who subscribe to the reference case model of the National Institute of Health and Care Excellence (NICE) in the UK, the decision rests upon the creation of modeled or simulated imaginary worlds and the application of threshold willingness-to-pay cost-per-QALY thresholds. On the other hand, for those who subscribe to the standards of normal science, the decision rests upon the ability to evaluate competing claims, both clinical and cost-effective, in a timeframe that is meaningful to a formulary committee. If we subscribe to the scientific method where the focus is on the discovery of new facts, untestable claims for clinical benefit and cost-effectiveness, such as created claims for lifetime cost per-quality-adjusted discounted life years (QALYs), are properly relegated to the category of pseudoscience. We have no idea, and will never know, whether the claims are right or even if they are wrong. If formulary decisions are to respect the standards of normal science then there has to be a commitment to claims evaluation. A willingness to accept new products provisionally, subject to an agreed protocol to support the evaluation of clinical and cost-effectiveness claims. This dichotomy between the standards of normal science and pseudoscience is explored in the context of published claims for cost-effectiveness and recommendations for product pricing in the US.

Keywords: ICER, cost-effectiveness modeling, pseudoscience, credibility, imaginary worlds, scientific method

Introduction: Demarcation

If we are to distinguish credible from non-credible claims made for pharmaceutical products and devices, we need to address what Karl Popper has called the 'demarcation problem'^{1,2}. What distinguishes those claims that meet the standards of normal science from those that might be described as pseudoscientific? Although science is not a monolithic activity, spanning the spectrum from hard science to soft science, the common element is the construction of empirically evaluable theories and hypotheses. From this perspective, unless we want to reject the standards of normal science, many pharmacoeconomic studies cross the line between science and pseudoscience. We have studies, on the one hand, that meet the required standards with hypotheses being evaluated through experimentation and systematic observation, while on the other hand we have modeled or simulated claims that clearly fail the criteria of normal science and are best seen as pseudoscience in their construction of imaginary worlds³.

reference case⁴. The limitations of the constructed reference case approach to supporting clinical acceptance of pharmaceutical products and devices is well established^{5,6}. Although there is the appearance of modeled or simulated claims embracing scientific standards in input assumptions and the construction of the core mechanism corresponding to the natural course of a chronic disease, the fact is that the modeled or simulated claims are entirely synthetic. The degree of belief in the models rests upon an acceptance of its correspondence to reality, a correspondence that necessarily entails its claims for product performance⁷. The Achilles heel is that simulations can be challenged and competing simulation constructed. Even in the context of the NICE single technology assessment process, contracted Evidence Review Groups (ERGs) can dispute the imaginary world's model or simulation presented by a manufacturer, and in turn have their 'new world' challenged by the NICE Advisory Group.

In the US, the equivalent construct is found in the standards recommended by the Academy of Managed Care Pharmacy (AMCP) Format for Formulary Submissions⁸. A recent commentary on the latest Version 4 of the AMCP Format details the apparent lack of concern with establishing the value case for pharmaceutical products and devices on a robust evidence-based platform, instead relying upon modeled or simulated worlds that track the natural course of a disease⁹. These imaginary constructs generate long-term or lifetime claims for clinical endpoints and incremental cost-per-quality adjusted life year (QALY) claims that are clearly impossible to evaluate. Indeed, one interpretation is that there was no intention of ever generating testable claims.

Corresponding author: Paul C Langley, PhD
Adjunct Professor
College of Pharmacy, University of Minnesota
Tel: 520-577-0436
Email: langley@maimonresearch.com

Probably the best example of an imaginary world is the National Institute for Health and Care Excellence (NICE)

Formulary committees in receipt of an AMCP *Format* driven submission, with competing product claims expressed in lifetime cost-per-QALY terms, suitably discounted over 10, 20 or more years, are expected to take the claim at face value as an input to determine the appropriate formulary tier position and price.

If we cast our net a little wider, recent reviews of cost-effectiveness studies published in two journals, *PharmacoEconomics* and the *Journal of Medical Economics*, concluded that none of the published papers met the standards of normal science in generating claims for competing products that could be evaluated experimentally or through systematic observation^{10 11}. Both reviews pointed to the acceptance of QALYs as the preferred endpoint with claims expressed in incremental cost per QALY terms. The fact that lifetime cost-per-QALY claims were impossible to evaluate and were never intended to be evaluated was not explored. The further point that no health care system collects QALYs on a regular basis was not raised.

The purpose of this review is to explore the extent to which the reference case approach to establishing value for competing pharmaceutical products, devices and interventions is still pursued in the US. The particular focus will be on modeled claims for cost-effectiveness and consequent threshold driven recommendations for product pricing undertaken by the Institute for Clinical and Economic Review (ICER). The ICER has been chosen, not only because of its high profile, but also because the value and claims methodology closely follows the NICE scoping and product submission process, to include a reliance upon a reference case standard.

Imaginary Worlds in Cost Effectiveness Analysis

Imaginary modeled worlds are extremely popular in cost-effectiveness analysis. Whether the popularity is a reflection of a lack of interest in developing evaluable claims for pharmaceutical products and devices that can be traced back to the release of the draft Australian guidelines in 1990, the acceptance of the reference case cost per QALY models mandated by NICE and a number of other health care systems or merely the continuation of a trend that can be traced back to the 1980s is an open question¹². In any event, there is widespread acceptance of the imaginary world approach to establishing the value of competing pharmaceutical products and devices. This acceptance embraces assessment agencies in New Zealand and Ireland as well as acceptance by professional groups, specifically the International Society for Pharmacoeconomics and Outcomes Research (ISPOR)^{13 14 15 16 17}. This willingness to embrace the construction of imaginary worlds may simply reflect belief in equivalence, where, with sufficient correspondence between the imaginary world and some perception of reality, the

untestable conclusions for comparative product value are necessarily entailed¹⁸. Truth, in this belief system, is by consensus.

If we accept that those advocating reference case modeling are not interested in developing model frameworks that have the potential to generate testable hypotheses, then this puts those models in the pseudoscience category. The predictions of the imaginary world, expressed all too often in lifetime cost per QALY terms are clearly never intended to be open to empirical evaluation. Claims from the model are immune to falsification.

A curious feature of the imaginary world modeling standard is that while, on the one hand, it attempts to correspond to reality in the representation of the course of a chronic disease through the application of, for example, Markov processes, on the other hand the methodology typically rejects any attempt, for example, to introduce adherence and persistence behavior and the impact of discontinuation, other than that reported in randomized clinical trials (RCTs) on the relative clinical benefits of competing products. The modeling is 'of this world' yet 'not of this world' (or a future world). The assumptions build into this imaginary world methodology clearly preclude testable claims. The models are simply commercial products.

ICER Case Studies

Three ICER reports have been selected for review. These are: (i) Mepolizumab in Severe Asthma; (ii) Insulin Degludec in Diabetes; and (iii) Novel Combination Therapies in Genotype 1 Chronic Hepatitis C Infection^{19 20 21}. While this is not a systematic review and selection, the reports chosen are stated by their authors as following the ICER value methodology. The focus here is on the cost-outcomes modeling and the application of willingness to pay thresholds to constructed cost-per-QALY lifetime outcome claims.

I: Mepolizumab in Severe Asthma

In March 2016, ICER through the California Technology Assessment Forum, released the final report for mepolizumab (Nucala®; GlaxoSmithKline) for the treatment of severe asthma with eosinophilia. The proposed model generated estimates of cost-effectiveness for the product utilized a Markov cohort framework, cycle length of 2-weeks, with three primary health states: asthma non-exacerbation state, an asthma exacerbation state and death. The asthma exacerbation state included three mutually exclusive sub-categories: an asthma related event that required an oral corticosteroid burst, asthma related-ED visit or asthma related hospitalization. The model structure was similar to other published asthma cost-effectiveness models where the intervention was modeled against the standard of care.

The base case for the mepolizumab model was a lifetime horizon; again, apparently consistent with other models. Given the uncertainty around duration of treatment, the impact of multiple treatment time horizons was captured through a sensitivity analysis. The treatment effect observed during the product randomized controlled trials (RCTs) trials was assumed to be consistent throughout the duration of the model time horizon. All costs and outcomes were discounted at 3% with direct healthcare costs only considered.

The hypothetical population characteristics for the model mirrored the mepolizumab exacerbation RCT population. There was an assumed allowance for inappropriate dosing linked to adverse event costs and disutility. Utilities were assigned to all health states in the model. For the non-exacerbation health states these were mapped by algorithm from the St George's Respiratory Questionnaire to the EQ-5D. For the exacerbation, health states utilities were assumed to be the same across treatment strategies. Given a dearth of utility data for these health states, the actual scores were taken as averages from the exacerbation health states. Over a lifetime treatment horizon the model yielded an estimated 23.96 avoided exacerbations (non-discounted) per patient receiving mepolizumab with the standard of care versus standard of care alone. The incremental cost per exacerbation averted was \$24,626. There was an estimated gain of 1.53 QALYs relative to standard of care. This resulted in a cost-effectiveness estimate of \$385,546 per QALY gained.

Clearly, these modeled claims are not intended to be evaluated. Apart from modifying parameters within the model and producing other estimates for lifetime exacerbations avoided, lifetime direct costs and lifetime cost-per-QALY, the outcomes with their high level of precision, are immune to failure. Applying cost-effectiveness thresholds of \$50,000, \$100,000 and \$150,000, ICER concluded from their model that to achieve a cost effectiveness threshold of \$150,000 the price per mepolizumab vial would need to be \$8932 (\$12,116) annually, a 63% discount from the \$2,500 WAC (\$32,500 per annum). Higher discounts would have to apply for lower cost-effectiveness thresholds.

II: Insulin Degludec in Diabetes

In March 2016, ICER again through the California Technology Assessment Forum, released the final report for insulin degludec (Tresiba®, Novo Nordisk A/S) for the treatment of diabetes. To generate hypothetical predictions for the cost-effectiveness of insulin degludec compared to insulin glargine U100, ICER utilized the UKPDS OM2 model to evaluate the differential treatment effects for those 30 years and over with any duration of diabetes and on a platform which is claimed to represent the typical trajectory for diabetes patients. The hypothetical target populations for the analysis were persons aged 18 years and over with either type 1 or

type 2 diabetes mellitus (DM). In the latter group hypothetical patients taking basal only insulin were considered separate from those taking basal-bolus insulin. The demographic and clinical characteristics of the hypothetical patient cohorts were from the respective insulin degludec RCTs.

The time horizon for the cost-effectiveness analysis was the hypothetical cohort lifetime extrapolated up to 70 years from the short term trial results of 6 to 12 months. The primary comparator for insulin degludec was insulin glargine U100. The model excluded insulin switching or other changes in therapy. The analysis focused on the hypothetical avoidance of hypoglycemic events, other adverse events, direct medical costs and reductions in health related quality of life. As the UKPDS OM2 does not include hypoglycemia as an outcome, the ICER modeling included hypoglycemia together with costs and disutilities for each event in a hypothetical sub-model. Each hypoglycemic event differentiated as mild/moderate daytime, mild/moderate nocturnal and severe was modeled from a patient level meta-analysis of insulin degludec clinical trials. Each event was assigned an associated cost and disutility. Disutilities were literature based. QALYs and costs were discounted at 3% per year.

The lifetime hypothetical base case analysis yielded similar rates of hypoglycemic events in the type 1 DM population for the two therapies. In the type 2 DM hypothetical population there were QALY differences but not differences in costs, as these were assumed to apply only to severe hypoglycemic events. For type 2 DM patients on basal only therapy, the base case assumptions resulted in a cost/QALY ratio of \$353,020 for insulin degludec compared to insulin glargine. For those on basal-bolus therapy the ratio was \$166,644/QALY. The report concluded that to achieve a cost-effectiveness ratio of \$150,000 per QALY gained there would need to be a cost reduction of 8% and 2% respectively in the basal only and basal-bolus populations.

III: Novel Combination Therapies in Chronic Hepatitis C

A multistate Markov model was constructed to determine the hypothetical cost-effectiveness of six treatment regimens for HCV genotype 1. Two of therapies were interferon based (peg-interferon + ribavirin [PR] and sofosbuvir + PR) and the remaining four interferon free therapies (sofosbuvir + ribavirin; simeprevir + sofosbuvir; ledipasvir/sofosbuvir 8/12; ledipasvir/sofosbuvir 12). The model was designed to capture the net costs, health benefits and ICERs of these therapies, to include how these might hypothetically vary if treatments were delayed to a later stage in disease progression. The model was intended to portray the lifetime progression of HCV based on the fibrosis stage of the disease, together with the impact of regression of liver damage after successful treatment. The comparative effectiveness of treatments was

measured primarily in terms of QALYs together with the incidence of serious HCV-related complications.

All results were presented for the patients' lifetime and discounted at 3%. Separate analyses were conducted for treatment-naïve and experienced patients, and whether patients have cirrhosis. Results were also presented for the treatment strategies of 'treat all' or 'wait until more advanced disease'. Results were also presented for a mixed cohort of treatment naïve and experienced patients. Health benefits were adjusted for the discontinuation rates reported in the various clinical trials.

The hypothetical patient population was assumed to weigh 75kg and be 60 years of age. These criteria were based on adjusted data from the 2010 National Health and Nutrition Examination Survey (NHANES). The distribution of patients across fibrosis stages was based on empirical assessments of individuals with known HCV infection. The model did not distinguish patients by viral concentrations, sex or rates, although the study report admitted that these might affect treatment outcomes and disease progression.

Lifetime costs were focused on the direct medical costs and covered drug costs, treatment related costs, other health care costs and adverse event costs. The unit costs (with a sensitivity band) were not expected to change over the patients' lifetime, although they were discounted. Utility values were estimated from the HCV literature, encompassing 10 HCV states (including death) and 5 utilities for states after sustained virologic response per Markov cycle.

Lifetime discounted QALYs, costs and ICERs were calculated for each regimen in comparison with the next least costly regimen. For completeness, each treatment option was also compared alternatively with no treatment as well as peg-interferon as a universal historical control. A range of scenarios in which alternative discontinuation rates, the distribution of the patient cohort by fibrosis stage, different cost of care gradients and the cohort's age were altered were also presented.

The base-case result, which reflects the values of inputs that were believed to be most accurate and relevant, for treatment naïve and a treat all strategy with peg-interferon yielded an ICER for peginterferon of \$11,385 compared to no treatment (a QALY gain of 1.51 years). The next lifetime lowest cost strategy of ledipasvir/sofosbuvir 8/12 generated an ICER of \$20,132 and a QALY gain of a further 1.41 years. All other sofosbuvir-based regimens were dominated with the exception of ledipasvir/sofosbuvir 12 strategy which was only slightly more effective (approximately 3 additional QALY weeks) but with an ICER of \$283,927.

Discussion

In none of the cost-utility models presented for the three case studies was any consideration given to the possibility of generating and assessing evaluable predictions. The claims presented were, in fact, immune to failure as they were expressed in lifetime cost-per-QALY terms. The choice of a modeled or simulated Markov or similar process to consider hypothetical intervention options, stages of disease and literature based estimates of costs and QALYS was justified by reference to these techniques as accepted standards in cost-effectiveness analysis. There was no discussion of the relevance of these standards to the criteria a formulary committee in the US might apply to competing interventions and the choice between competing treatment strategies. Rather, the NICE reference case model was accepted as appropriate for constructing competing claims and driving formulary decisions. The possible need to conform to the standards of normal science in developing evaluable hypotheses, evaluation and possible replication was not considered. This is an unfortunate oversight given increasing concerns with the ability to replicate RCT based claims and the possibility of outcome switching in presenting phase 3 RCT results^{22 23 24}. Indeed, it should not be forgotten that just as the construction of cost-per-QALY claims rest on the acceptance of models driving indirect comparisons to support assumptions for comparative treatment effects, the indirect treatment effect claims make implicit assumptions as to the replicability of the component individual RCT claims. The possibility of reviewing the robustness and replicability of the individual phase 3 RCT claims was not addressed.

The absence of testable claims means, for example, that there is no attempt within the report to suggest how the competing claims for mepolizumab plus the standard of care versus standard of care alone might be assessed and reported back to a formulary committee as part of an ongoing disease area and therapeutic review. There was no consideration of the possibility, first put forward in the WellPoint formulary guidelines in 2005, that to support formulary assessment, manufacturers might submit a protocol alongside their submission to propose how the claims might be evaluated and the results reported as part of ongoing disease area and therapeutic class reviews^{25 26 27}. Standards that have been recently put forward in the University of Minnesota Social and Administrative Pharmacy Program proposed Guidelines for Formulary Evaluations.²⁸ The key points in the proposed guidelines are: (i) to present claims for competing pharmaceutical products that are evaluable; (ii) to submit a protocol to detail how the claims are to be evaluated; (iii) for the manufacturer to underwrite the evaluation; and (iv) to report claims back to a formulary committee as input to ongoing disease area and therapeutic class reviews and subsequent replication.

It is also not clear how a health system, who may not have a notional threshold willingness-to-pay or even consider applying one, would interpret these results. Attempts to negotiate with GSK for a potential discount for mepolizumab, for example, could presumably be challenged by (i) the degree of belief or otherwise in relevance of the modeled claims when these are patently constructed and immune to failure and (ii) the appropriateness of assumed willingness-to-pay thresholds. The only basis for the recommendations for unit price discounts for both mepolizumab and insulin degludec are the two constructed simulations.

Competing Simulations

Just as imaginary worlds yield imaginary claims, competing imaginary worlds can yield competing imaginary claims. In the absence of claims being subject to experimentation or systematic observation, there is no way to judge their relative merits. It is entirely possible to envisage a situation where competing manufacturers commission competing simulations to support their respective products. The sheer complexity of the typical simulation, the assumptions made to justify both the structure of the model in choice of mathematical framework, the choice of stages of disease and transition probabilities between health states, together with the selection of unit costs and the trawling of the literature to cobble together utilities for the various disease states, all admit to the feasibility of reverse engineering to generate a required endpoint. This holds irrespective of appeals to ISPOR standards and the application of sensitivity analyses and the exploration of various scenarios.

If competing products yield competing claims, with the claims expressed as lifetime cost-per-QALY increments (justified in turn by an appeal to the standard of the NICE reference case), then the only way to judge the merits of the claims is, presumably, to subject the competing simulations to expert appraisal. In the UK, this is achieved by engaging evidence review groups (ERGs), usually from academic institutions to report on their assessment of the merits of a submitted model and to suggest, if they judge it appropriate, modifications to the model or to develop an entirely new model. Their deliberations are in turn assessed by the NICE Advisory Panel who may accept, reject or suggest modifications. At no stage, of course, is there any question of evaluable claims, experimentation or even replication. Whether such an appraisal will ever occur in the US is a moot point given the lack of interest in QALYs by health care systems.

Adherence and Persistence

An odd feature of the lifetime imaginary reference case methodology is the treatment of adherence and persistence. Apart from 'involuntary' discontinuation through death or discontinuation rates based on RCT experience, the reference

case standards mandated by NICE say nothing about how the modeled or simulated claims should capture non-compliance with therapy. This is a curious oversight in lifetime cost-per-QALY models as the evidence would suggest that patients are typically non-compliant with chronic therapy in a relatively short period of time. In general terms, within approximately 3 years of an index prescription, less than 30% of patients are likely to have persisted with therapy and a significant proportion of these are likely to be non-adherent (medical possession ratio < 0.8). If this is the case then it seems pointless to construct lifetime cost-per-QALY models which assume that patients remain adherent (and hence persistent) with medications over 10, 20 or even 30 years.

Why Choose QALYs?

A further odd feature of current standards for models or simulations is the choice of QALYs as the primary endpoint, with summary cost-per-QALY claims. Aficionados of utility measurement have a choice between direct measures of health states such as the variants of the rating scale, standard gamble and time trade off as well as opposed to a number of multi-attribute health status classification systems with preference scores²⁹. The latter including the Quality of Well-being Scale (QWB), the various incarnations of the Health Utilities Index (HUI), the various EuroQol measures (EQ-5D) and constructs drawn from the SF family of instruments (e.g., SF-6D).

While it is understandable if the QALY (e.g., the EQ-5D) is mandated, as it is with the NICE reference case (and enshrined in legislation), the situation in the US is quite different. With the failure some years ago of attempts to advocate a QALY reference case standard in the US and the apparent lack of interest, as noted above, by the overwhelming majority of health systems in QALY outcomes, to express comparative claims in QALY terms seems unhelpful³⁰. Indeed the chance that a QALY quality metric will be mandated in the US is effectively zero. This was made abundantly clear in the *Affordable Care and Patient Protection Act (2010)* which mandates that the Patient Centered Outcomes Research Group (PCORI) exclude discounted cost-per-QALY or similar discounting measures and threshold values for priority setting in health care³¹.

Of course, expressing outcomes in cost-per-QALY terms, apart from the choice of QALY, adds a further level of immunity to claims. If claims expressed in cost-per-QALY terms are open to evaluation then, in the absence of any health system capturing QALYs as part of routine data collection means that rather than relying on data vendors to support claims evaluation there will have to investment in a specific prospective study. Realistically, this is unlikely to occur.

Challenges

For those advocating and constructing imaginary worlds to support value claims in formulary decisions, the risk is that the claims are open to challenge. In the absence of claims expressed in evaluable terms, offering the opportunity for hypothesis testing and replication, those advocating imaginary world are in no different a position from those advocating intelligent design. There is no basis for rejecting any claims that are made. Different models created by different intelligent designers can compete, but not on the basis of hypotheses being evaluated. Model assumptions, inputs and mechanisms can be challenged, but without any necessary resolution outside of existential appeals to the standards proposed by professional groups.

Consider a recent systematic review of modeled claims published in the *Journal of Medical Economics* in the 12 months from January 2015 to December 2015. The review identified 31 cost-effectiveness papers. None met the standards of normal science in generating testable predictions. Fifteen present lifetime cost-per-QALY models. A total of 30 of the papers were funded by manufacturers with 29 supporting untestable claims for the manufacturers' product. Clearly, these could be challenged by constructing competing models, but no one appears to have thought it worthwhile. The papers are probably best seen as marketing exercises.

Conclusions

By definition, imaginary worlds yield imaginary claims. If the standards of normal science are applied as the basis for separating science from pseudoscience are accepted, then modeled or simulated claims for pharmaceutical products and devices that are immune to failure are clearly in the epistemological pseudoscience space.

However, it is not the intention here to discourage the construction of imaginary worlds to support value claims for pharmaceutical products and devices. After all, there may be health care systems (and this is certainly true of single payer systems in the UK, Ireland and New Zealand) who mandate

reference case submissions. There may also be health systems in the US willing to accept imaginary world based value claims as 'useful' inputs to formulary decisions. But there may also be health systems that fail to see the merits of constructed value claims.

Meeting the standards of normal science, setting the stage for claim evaluation, it always a challenge. The challenge, however, is more in the mindset of the analyst and the willingness to put the pursuit of pseudoscience to one side. If we have the skills and resources to build complex imaginary worlds, claiming to track the key features of a chronic disease over 20 or 30 years, then surely it would not be too much of a challenge to create modeled or simulated evaluable claims that focus on the short-term. Perhaps then cost-effectiveness models and simulations might contribute to the national three-part CMS aim of better care and better health with smarter spending³².

References

¹ Pigliucci M. *Nonsense on Stilts: How to tell science from bunk*. Chicago: University of Chicago Press, 2010

² Popper KR. *Objective Knowledge (Rev Ed)*. Oxford: University Press, 1986

³ Langley PC. Validation of modeled pharmacoeconomic claims in formulary submissions. *J Med Econ*. 2015;18(12):003-999

⁴ National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal*. London: NICE, April 2013

- ⁵ Belsey J. Predictive validation of modeled health technology assessment claims: lessons from NICE. *J Med Econ.* 2015;18(12):1007-1012
- ⁶ Langley PC. Sunlit Uplands: The genius of the NICE reference case. *INNOVATIONS in Pharmacy.* 2016;7(2): Article 126
- ⁷ Grim P, Rosenberger R, Eason RE et al. How simulations fail. *Synthese.* 2011 DOI 10.1007/s11229-011-9976-
- ⁸ Academy of Managed Care Pharmacy. AMCP Format for Formulary Submissions, Version 4.0, April 2016
- ⁹ Langley PC. Modeling imaginary worlds: Version 4 of the AMCP Format for Formulary Submissions. *INNOVATIONS in Pharmacy.* 2016;7(2): Article 11
- ¹⁰ Langley P. *Imaginary Worlds: Modeled Claims for Cost-Effectiveness Published in PharmacoEconomics* January 2015 to December 2015. *INNOVATIONS in Pharmacy.* 2016;7(2): Article 9
- ¹¹ Langley PC. *Imaginary worlds: a systematic review of the status of modeled cost-effectiveness claims published in the Journal of Medical Economics* from January 2015 to December 2015. *INNOVATIONS in Pharmacy.* 2016;7(2): Article 16
- ¹² Langley PC. Great Expectations: Cost-utility models as decision criteria. *INNOVATIONS in Pharmacy.* 2016;7(2): Article 14
- ¹³ New Zealand. Pharmaceutical Management Agency. Prescription for Pharmacoeconomic Analysis: Methods for cost-utility analysis (v 2.1). PHARMAC, 2012.
- ¹⁴ Ireland. Health Information and Quality Authority. Guidelines for the Economic Evaluation of Health Technologies in Ireland. Dublin: HIQA; 2014
- ¹⁵ 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
- ¹⁶ 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
- ¹⁷ Caro JJ, Eddy DM, Kan H et al. Questionnaire to assess relevance and credibility of modeled studies for informing health care decision making: An ISPOR-AMCP-NPC good practice task force report. *Value Health.* 2014;(17):174-182
- ¹⁸ Wootton D. *The Invention of Science.* New York: Harper Collins, 2015
- ¹⁹ California Technology Assessment Forum. Mepolizumab (Nucala, GlaxoSmithKline plc.) for the treatment of severe asthma with eosinophilia: effectiveness, value, and value-based price benchmarks. ICER, Final Report, 2016 http://icer-review.org/wp-content/uploads/2016/03/CTAF_Mepolizumab_Final_Report_031416.pdf
- ²⁰ California Technology Assessment Forum. Insulin Degludec (Tresiba, Novo Nordisk A/S) for the treatment of diabetes: effectiveness, value, and value-based price benchmarks. ICER March 14 2016 http://icer-review.org/wp-content/uploads/2016/03/CTAF_Degludec_Final_Report_031416.pdf
- ²¹ California Technology Assessment Forum. The comparative clinical effectiveness and value of novel combination therapies for the treatment of patients with genotype 1 chronic hepatitis C infection. ICER, Final Report, 2015 http://icer-review.org/wp-content/uploads/2016/01/CTAF_HCV2_Final_Report_013015.pdf
- ²² Editorial. *Nature.* 2016;529:256.

- ²³ Camerer CF, Dreber A, Forsell E et al. Evaluating replicability of laboratory experiments in economics. *Science*. 10.1126/science.aaf0918 (2016)
- ²⁴ Centre for Evidence Based Medicine. COMPare Project. Oxford, 2016. <http://compare-trials.org/results>
- ²⁵ 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
- ²⁶ 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.
- ²⁷ Schommer J, Carlson A, Rhee G. Validating pharmaceutical product claims: questions a formulary committee should ask. *J Med Econ*. 2015;18(12):1000-1006
- ²⁸ Langley PC. Nullius in Verba: The University of Minnesota Social and Administrative Pharmacy Program proposed Guidelines for Formulary Evaluations. *INNOVATIONS in Pharmacy*. 2016;7(2): Article 10
- ²⁹ Drummond MF, Sculpher MJ, Torrance GW et al. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd Ed. Oxford University press, 2005
- ³⁰ Drummond M, Brixner D, Gold M et al. Towards a consensus on the QALY. *Value Health*. 2009;12(Supp 1):S31-S35
- ³¹ Persad G. Priority setting, cost-effectiveness and The Affordable Care Act. *Am J Law Med*. 2015;41:119-166
- ³² Rajkumar R, Conway PH, Tavenner M. CMS 0- engaging multiple players in payment reform. *JAMA*. 2014;311(19):1967-8