

7-5-2016

Modeling Imaginary Worlds: Version 4 of the AMCP Format for Formulary Submissions

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. Modeling Imaginary Worlds: Version 4 of the AMCP Format for Formulary Submissions. *Inov Pharm*. 2016;7(2): Article 11. <http://pubs.lib.umn.edu/innovations/vol7/iss2/11>

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

Modeling Imaginary Worlds: Version 4 of the AMCP Format for Formulary Submissions

Paul C. Langley

College of Pharmacy University of Minnesota

Note: An earlier version of this paper was submitted to the AMCP in December 2015 as a response to their publication of a draft for public comment of Version 4 of the AMCP Format for Formulary Submissions. There was no response received from the AMCP. This earlier version is available from the author.

Abstract

The question of demarcation between normal science and pseudoscience is critical to the discovery of new facts. The core elements supporting progress in science are: (i) empirically evaluable coherent theories and (ii) the testing of hypotheses through experimentation or systematic observation. If modeled or simulation-based claims for cost-effectiveness are to be accepted as a credible input to health care decision making then they must conform to these standards. Claims should be testable, falsifiable and replicable. If not then they are best seen as pseudoscience. This assessment of the latest version of the Academy of Managed Care Pharmacy (AMCP) Format for Formulary Submissions (Version 4.0; April 2016) concludes that, in their recommendations for cost-effectiveness modeling, the proposed standards do not meet those of normal science. Rather, in common with previous versions of the AMCP Format, the modeling framework proposed not only puts to one side the issue of testable claims, but supports the modeling of imaginary worlds or thought experiments where claims are immune to falsification. In consequence, the payer or other recipient of a modeled or simulated claim that follows the AMCP Format has no idea, in the absence of observation or experimentation, whether the claim is right or even if it is wrong. The claims are potentially misleading, possibly harmful, but to an unknown extent. They have no place in evidence-based medicine.

Introduction

In April 2016, the Academy of Managed Care Pharmacy (AMCP) launched version 4 of their *Format for Formulary Submissions*¹. The purpose of the *Format* is to provide a framework to support the submission of clinical and cost-effectiveness claims to health care purchasers in their assessment of new and competing pharmaceutical products and devices. The *Format* 'is designed to maintain a high standard of objectivity and credibility' in the information provided to manufacturers and at the same time streamlines the process. The adoption of the *Format* is regarded as best practice for the formulary review process.

Given the widespread acceptance of the *Format* a legitimate question is whether or not the standards proposed for modeled or simulation-based cost-effectiveness and cost-utility claims are credible: do they meet the standards of normal science?

If a modeled cost-effectiveness claim is to meet the standards of normal science then it has to (i) involve the construction of an empirically evaluable coherent theory and (ii) facilitate the testing of hypotheses through experimentation or observation. These are standards that have been in place since the 17th century and demarcate science from pseudoscience; the demarcation between natural selection and intelligent design². More specifically: Do the standards support the construction of testable claims for product impact, claims that have the potential to provide meaningful feedback to a formulary committee as part of ongoing disease area and therapeutic reviews? Are the claims presented capable not only of evaluation but of falsification? Is there a potential for the claims to be re-evaluated and replicated in other target patient populations? Can the claims be generalizable?

The purpose of this paper is to point to a fatal epistemological error in the AMCP *Format*, one that has characterized, not only all previous versions of the AMCP recommendations for formulary submissions but also, unfortunately, those considered exemplar formulary submission guidelines. These guidelines include those from the National Institute for Health and Care Excellence (NICE) in the UK, the Pharmaceutical Benefits Advisory Committee (PBAC) in Australia, PHARMAC in New Zealand, the Health Quality Information Authority (HQIA) in Ireland and the

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

Canadian Agency for Drugs and Technologies in Health (CADTH)^{3 4 5 6 7}. The error is that none of these guidelines meet the standards for 'normal science': they support the construction of modeled product claims for clinical and cost-effectiveness that fail to generate testable and reproducible hypotheses for the anticipated impact of products in health care systems. It is not a question of generalizability; the absence of testable claims means they are not, by definition, generalizable.

As such, these models or simulations are best considered as imaginary worlds or thought experiments. The recipient of the submission has no idea, in the absence of experimentation or observation, of whether the claims made are right or even if they are wrong. While the author(s) of the modeled claims may justify their construct on the grounds that it reflects their perception of reality, the claims have the potential to be misleading and even harmful, but to an unknown and unknowable extent. As such, they should be put to one side both by manufacturers and health system decision makers.

The Standards of Normal Science

The requirement for testable hypotheses in the evaluation and provisional acceptance of claims made for products and devices is unexceptional. Since the 17th century it has been accepted that if a research agenda is to advance, if there is to be an accretion of knowledge, there has to be a process of discovering new facts. Indeed, as early as the 16th century Leonardo da Vinci (1452 – 1519) in notes that appeared posthumously in 1540 for his *Treatise on Painting* (published in 1641) clearly anticipated the standards for the scientific method which were widely embraced a century later in rejecting thought experiments that fail the test of experience. By the 1660s, the scientific method, following the seminal contributions of Bacon, Galileo, Huygens and Boyle, had been clearly articulated by associations such as the Academia del Cimento in Florence (1657) and the Royal Society in England (founded 1660; Royal Charter 1662) with their respective mottos *Provando e Riprovando* (prove and again prove) and *nullius in verba* (take no man's word for it).

In the early 20th century standards for empirical assessment were put on a sound methodological basis by Popper in his advocacy of a process of 'conjecture and refutation'^{8 9}. Hypotheses or claims must be capable of falsification; indeed they should be framed in such a way that makes falsification likely. Life becomes more interesting if claims are falsified because this forces us to reconsider our models and the assumptions built into those models. This leads, then, to the obvious point that claims or models should not be judged on the realism or reasonableness of assumptions or on whether the model 'represents' reality. It is worth reflecting on

Popper's critique of induction: 'never in science are inferences drawn from mere observational experience to the prediction of future events'¹⁰. Or, to put it simply: not all swans are white.

It is not as though the issue of testable claims has not been raised before. This dichotomy between predictive validation and claims that a model is 'realistic' and that its claims should be taken at face value were recognized some 10 years ago in the WellPoint formulary submission guidelines in their focus on an outcomes based formulary^{11 12}. The guidelines, first issued in 2005, were explicit as to the need for predictive claims as an input to regular and ongoing disease area and therapeutic class reviews. The guidelines were designed to set standards for new product submissions as well as submissions to support disease area and therapeutic class reviews. A requirement in new submissions was that claims for costs and outcomes should be in a form that allowed evaluation in the short term. Those making a submission were asked to submit a protocol that detailed how these claims were to be assessed and reported back to the formulary committee. In the context of a life cycle perspective on drug products, initial claims were seen as provisional and subject to ongoing reviews to capture the impact of new products through comparative assessments. Claims that could not be verified were to be rejected, to include modeled claims and simulations that failed to generate testable hypotheses.

More recently, in a supplement to *the Journal of Medical Economics (JME)*, the case was put forward that if claims for the impact of products and devices on costs and outcomes in health care systems are to be accepted then they should meet the standards expected in 'normal science'^{13 14 15 16 17}. The only acceptable modeled claims in formulary submissions are those that are testable in a timeframe relevant to the needs of a formulary committee. If claims do not meet this standard they should be rejected. The supplement, given the experience of modeled claims made to NICE and the current situation in the US and other developed economies, proposed a new research agenda that focused on the testable impact of products and devices on patient outcomes, resource utilization and the costs of health care delivery. Key elements in this proposed new research agenda were that formulary submissions claims should be evaluated and reported on in a timeframe that is meaningful to the committee. In practice, this would mean 2-3 years with results reported on as part of ongoing disease area and therapeutic reviews. To support this process the supplement recommends that formulary submissions be accompanied by a protocol detailing how the claims are to be evaluated and a short list of questions a formulary committee should ask of a manufacturer's

submission. At the same time the supplement points to the ready availability of 'big data' to support claims evaluation.

Finally, it is important to point out that there has been increasing concern expressed in the last few years over the ability to replicate claims from pivotal phase 3 clinical trials. As noted in a recent editorial in *Nature*, applicants to the National Institutes of Health (NIH) are now required 'to explain the scientific premise behind their proposals and defend the quality of their experimental design'¹⁸. More recently, Camerer et al in their evaluation of laboratory experiments in economics find, of the 18 studies considered, an effect size in the same direction in only 11 replications with on average a replicated effect size of 66% of the original¹⁹. As the authors point out 'the deepest trust in scientific knowledge comes from the ability to replicate empirical findings', although rarely carried out in the social sciences.

At the same time concern has been expressed over the practice of outcomes switching where, in the process of reviewing evidence for the initial primary and secondary outcomes targets detailed in the phase 3 protocol, the data are 'reconfigured' to capture more statistically acceptable endpoints²⁰.

Given that indirect comparisons utilizing techniques such as network analysis are a key input to modeled or simulated claims for comparative cost-effectiveness, the lack of replication in a significant proportion of phase 3 trials and the presence of outcomes switching raises doubts as to whether or not these indirect clinical efficacy claims should be taken at face value. It would appear to be more appropriate to consider such claims as simply working hypotheses. While it is worth noting that the potential impact of non-replicability and outcomes switching are not considered in the latest *Format* in the reporting of clinical claims, these claims should be evaluated as part of any formulary submission.

If it is difficult to replicate claims from phase 3 clinical trials, then it is impossible to replicate claims that are untestable. In the case of modeled or simulated claims that adhere to the standards suggested by the AMCP Format, it is not the question of replication in other target populations but of not having a replicable claim in the first place. It is difficult to see how the support for non-testable modeled outcomes squares with the claim by the AMCP that decision models can provide 'benchmarks against which the product's future performance can be measured'. If it is not, as the AMCP goes on to say, the concern that models may be perceived as 'black boxes', rather it is the concern that they may be 'empty boxes'.

Cost-Outcomes Standards in the 2016 AMCP Format

The 2016 AMCP *Format* recommendations are not intended to specify methods for assessing clinical benefit, harms or economic impact. The proviso is that they should meet accepted standards of evidence based medicine and health technology assessment. There is, in addition, a strong recommendation that formulary submissions include evidence from comparative effectiveness research studies, although this evidence may not be available at new product launch. Manufacturers are expected to articulate a value argument to justify expected expenditures for the product 'in the context of its anticipated effects on the clinical evidence, health outcomes, and the economic consequences for the healthcare system'. In respect of the economic benefits, there should be a summary in terms of (i) cost per unit; (ii) potential clinical benefits (including quality of life) and potential economic benefits (including savings or cost offsets).

The AMCP *Format* sees the intent of economic modeling to 'quantify for the healthcare system the risk-benefit tradeoff of the product and its economic value'. Decision based cost effectiveness models are seen as both effective and central to the case made for the product. In model development, manufacturers are asked to consider recommendations published by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR). The analytic framework recommends including clinical events, life expectancy and quality adjusted life years (QALYs) 'with the latter two outcomes primarily relevant for lifetime analyses'. There is no prescription for the modeling framework with manufacturers free to choose between decision trees, Markov or cohort models and patient level or discrete event simulation models. The simplest feasible modeling approach is recommended. A payer perspective in modeling is recommended with a time horizon 'appropriate to the disease being studied'. Multiple timeframes are recommended for chronic disease e.g., 5-year, 10-year and lifetime' and adjusted for time preference. In presenting a base-case analysis expected clinical and economic outcomes are to be estimated for each strategy (model) together with incremental costs and effectiveness. Differences in the absolute risk of events are to be calculated and healthcare offsets vs. drug costs should be presented, as should clinical risk-benefit tradeoffs.

Consensus in Modeled and Simulated Claims

There is nothing in the AMCP *Format* Version 4 (and earlier versions) that addresses the question of the importance of meeting the standards of normal science in models or simulations to support comparative clinical and cost-effectiveness claims. To accept the position taken in the AMCP *Format* is to accept a relativist position in the

philosophy of science. Rather than subscribing to the position that the standards of normal science are the only standards to apply in health care decisions, the relativist believes that all perspectives are equally valid²¹. In their advocacy of the equivalence or symmetry principle health care decisions are to be understood sociologically. No one body of evidence is superior to another. Results of a simulation are on an equal basis with those of a RCT. For the relativist, the success of a scientific research program, in this case one built on models and simulations, rests not on its ability to generate new knowledge but on its ability to mobilize the support of the community. Basing decisions on models and simulations underpins the consensus view that evidence is constructed, never discovered. Instead of coming to grips with reality science is about rhetoric, persuasion and authority. Truth is consensus.

If we accept the relativist opinion and argue that decisions in healthcare are most appropriately based on models and simulations then we have to address the possibility that simulations can fail. Simulations or models are accepted because in the consensus view, the view of the authorities in the discipline, the ability to capture the critical or similar features of the reality of a decision is all that is required. If the simulated input conditions and the simulated core mechanism correspond to reality, the sufficient condition character of the simulation assures us that the output is necessarily entailed and predictions must correspond to reality²².

At the same time there is unlikely ever to be agreement on correspondence, sufficiency and necessary entailment. The flexibility allowed in constructing simulations means that simulations, by their nature, can always fail. Simulations can also be 'tailored' to generate the required endpoints. Rather than capturing the essence of a reality, the simulation captures the perception of the essence held by the authors of the simulation. Unless there is a process of independent assessment through experimentation or observation, there is no way in which the likelihood of failure can be judged. To argue that readers will compare one simulation with another and attempt to rebut competing simulated claims through a comparison of a model's structure or its assumptions is unlikely to be a useful exercise unless this is linked to the empirical evaluation of competing claims.

These conclusions hold irrespective of how much we attempt to build up the appearance of being scientifically rigorous in the validation of models and simulations. We can claim that they 'adequately reflect reality', we can apply deterministic and probabilistic sensitivity analyses, we can produce incremental cost effectiveness ratios (ICER) cloud diagrams, we can apply thresholds, we can argue for face and content

validity and we can even introduce an analysis of the value of perfect information. Nevertheless, the fundamental objection still applies: if there are no testable predictions the simulation fails the standards of normal science.

Quality Adjusted Life Years

The AMCP *Format* recommends that clinical events, life expectancy and quality adjusted life years (QALYs) all be assessed in the cost-effectiveness analysis. The last two endpoints are seen as primarily relevant for lifetime analyses. Following the standards of the NICE reference case preference estimates are to be derived from either patients or the general population. As well as direct preference elicitation, six possible instruments are referenced. There is no recommendation for any individual instrument or any suggestion that cost-per-QALY thresholds could be usefully applied to QALY claims.

While a QALY may be seen, as the AMCP notes, as a universal health outcomes measure there is no discussion of how a particular QALY instrument is chosen. Instruments vary in the number of health dimensions included, the number of levels captured within each dimension and their severity. They also differ in the populations surveyed to elicit preferences for health states, how the preference score is derived and how the preference data are translated into a preference score. There is no discussion over recent claims that the preferences expressed over hypothetical health states are inconsistent with the assumptions of multiattribute utility theory²³.

The issue of the relevance of incremental cost-per-QALY claims to health care decision making is not addressed. Given the recommendation that modeled claims should track the course of a chronic disease, there is no discussion of how formulary committees should factor 'discounted' comparative lifetime cost-per-QALY claims' into their decision making. Lifetime (or long-term) QALY based claims are obviously not evaluable. As such they fail to meet accepted standards for credibility in normal science. Indeed, even if QALYS were put in an evaluable form (e.g., evaluable within a 2-year time frame), the fact is that no one 'collects' QALYS (with a choice of six flavors to choose from) as elements in administrative claims or electronic medical records. If evaluable QALY claims are to be assessed then a prospective observational or experimental study would have to be underwritten and implemented. As this is a less than likely proposition, this creates a further barrier to assessing QALY claims (let alone linking such claims to future comparative effectiveness assessments). In short, health care systems are not interested in QALYs. Single payer systems may mandate QALY outcomes in modeled claims but, at least in the case of NICE, these support threshold pricing

negotiations where there is no intention of ever evaluating the claims.

ISPOR and Normal Science

In failing to recognize the standards of normal science, the AMCP is not alone. Looking back over the past decade or more ISPOR in its commitment to developing standards has emphasized the role of assumptions rather than hypotheses in model development. Certainly, the notion of validation has been addressed, but rather than emphasizing the fundamental role of prediction and testable hypotheses, this has been seen as a preferred, yet not essential element in validation^{24 25 26}. While establishing a degree of belief in a model, it may seem odd if that model is incapable of generating testable claims. The point is that predictive validation is not just one type of validation. Predictive validation (or assessment) stands alone; it is the only basis on which a model can be judged. The risk, for both manufacturers and health care decision makers, is that the modeling is seen as so inherently elegant that it supplants, as Ellis and Silk note, the need for data and testing²⁷. Unfortunately, this avoids the critical question: what potential observational evidence would persuade you that the theory is wrong and lead you to abandoning it? If there is none then it is not a scientific theory.

Implications of Accepting the AMCP Format for Formulary Submissions

It is unclear from the AMCP recommendations whether the results of modeled claims, the 'anticipated effects' are ever intended to be put in a form that is testable. Indeed, the authors of the AMCP *Format* appear unconcerned as to whether or not claims are put in a form that are testable or whether or not it is appropriate that they capable of being tested, let alone reproduced. There seems to be little if any concern that there should be feedback to a formulary committee from claims evaluation as part of ongoing disease area and therapeutic class reviews. Some of the language used could be interpreted to mean that concrete claims could be evaluated but there is nothing in the recommendations to distinguish modeled claims that are potentially testable and meet the standards of normal science from claims that are the outcomes of modeled imaginary worlds or thought experiments. Note, in particular, the recommendations that the modeled time horizon should be 'appropriate to the disease being studied' and that for chronic diseases this could extend to the patient's life time.

This failure to recognize and follow the standards of normal science will, inevitably, raise doubts as to credibility of the AMCP *Format* and its relevance for health care decision making. As well, doubt will be cast on not only of a large number of pharmacoeconomic studies, but on global

formulary guideline standards and the contribution of formulary submissions to product placement. A project is currently underway at the College of Pharmacy, University of Minnesota to evaluate the credibility of published cost-effectiveness claims. To date, modeled cost-effectiveness claims published in two journals, *PharmacoEconomics* and the *JME*, have been assessed and the results published^{28 29}. Both of these reviews asked: (i) whether the model was capable of generating evaluable claims; (ii) whether the author(s) attempted to generate evaluable claims; (iii) whether the author(s) suggested how the claims might be evaluated; and (iv) whether the author(s) cautioned readers as to the implications of generating non-evaluable claims for the credibility of the analysis? Of the 63 papers reviewed in the two publications, none met these requirements.

Conclusions and Recommendations

The 2016 AMCP *Format* (as in previous versions) suffers from a major and fatal epistemological error: it fails to support the standards of normal science. If, on the one hand, the authors of the format agree that this is a desirable standard to aim for, then the document needs to make this quite clear. If they take a relativist position then that is the end of the argument. If, on the other hand, the authors of the 2016 AMCP *Format* accept the need that in models and simulation for cost-outcomes claims they need to subscribe to the standards of normal science then this is probably best achieved by stating quite equivocally that claims made for the impact of pharmaceutical products and devices should be framed to ensure (i) that they are capable of empirical evaluation and replication; and (ii) that they should be capable of being evaluated in a relatively short period of time; a period consistent with the requirements of a formulary committee and as inputs to ongoing disease area and therapeutic class reviews. There is no middle ground.

In either event, following the WellPoint guidelines and the recommendations of the *JME* supplement, the AMCP recommendations should include a requirement that an evaluation protocol accompany claims for product impact. The *JME* supplement has provided a possible framework linked to questions that a formulary committee should ask. Developing such a protocol should be relatively straightforward given the experience in the drafting of protocols to support randomized clinical trials (RCT) and observational studies in comparative effectiveness research. Unless otherwise agreed by the parties, the content of the protocol should remain confidential and the request for a protocol should be at the discretion of the payer. Unless otherwise agreed, the costs of implementing the validation protocol should be borne by the manufacturer.

The concerns raised in this assessment of the latest AMCP Format in respect of evaluable and replicable product claims have been addressed in the recently published *University of Minnesota Social and Administrative Pharmacy Program Proposed Guidelines for Formulary Evaluations*³⁰. Unlike the latest AMCP Format these guidelines emphasize the need to conform to the standards of normal science in submitting comparative claims for new products. A key element in the guidelines is the role of a protocol to be submitted to detail how the claims are to be evaluated in a timeframe that is meaningful to a formulary committee. The guidelines point to the responsibility of the manufacturer to underwrite claims for both clinical and cost-effectiveness outcomes, potentially linking these assessments to subsequent comparative effectiveness disease area and therapeutic class reviews over the lifetime of the product.

In focusing on claims that are evaluable in the short-term, the emphasis in model building will likely move away from those models which attempt to capture the natural course of a disease and more abstract reference case approaches. Unless they can adapt to generating short-term testable claims, cohort and Markov and discrete event simulation models are likely to be supplanted by models that are based on RCTs or extrapolate from them in the short term. It is also likely that claims for product impact will be in more disaggregated terms with claims for short-term clinical outcomes and resource utilization supplanting broad-based 'value claims' for incremental cost-effectiveness. While this does not discount claims based on patient reported outcomes or quality of life, cost-per-QALY claims are unlikely to take center stage with reference case frameworks disappearing. This does not mean, however, that the re-drafted AMCP recommendations should discourage evaluable PRO and QALY claims.

Whether the AMCP and the authors of the AMCP *Format* are prepared to accept these recommendations is an open question. Given how much has been invested over the past 15 years by the AMCP and ISPOR in recommendations for modeling claims and standards respectively, the risk is that with this sunk intellectual capital, there will be a reluctance to accept that a change in direction is required. If this is the case, then it is incumbent upon the AMCP to justify how untestable claims generated by imaginary worlds contribute to effective formulary decision making in, for example, precision medicine and support the evidentiary standards being demanded by Federal government agencies such as the Centers for Medicare and Medicaid Services for quality outcomes and metrics³¹. With the arguments presented here, subscribing to the standards of normal science to drive a new research agenda offers the only sound basis for

evidence driven decision making and for new frameworks for generating claims.

References

1. Academy of Managed Care Pharmacy. AMCP Format for Formulary Submissions, Version 4.0, April 2016
2. Pigliucci M. *Nonsense on Stilts: How to tell science from bunk*. Chicago: Chicago University Press, 2010
3. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal. London: NICE, April 2013
4. New Zealand. Pharmaceutical Management Agency. Prescription for Pharmacoeconomic Analysis: Methods for cost-utility analysis (v 2.1). PHARMAC, 2012. <https://www.pharmac.govt.nz/assets/pfpa-final.pdf>
5. Ireland. Health Information and Quality Authority. Guidelines for the Economic Evaluation of Health Technologies in Ireland. Dublin: HIQA; 2014
6. Australian Government. Department of Health. *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (version 4.4)*. Canberra: July 2013
7. Canadian Agency for Drugs and Technologies in Health. Guidelines for the Economic Evaluation of Health Technologies: Canada (3rd Ed). CADTH: Ottawa, 2006
8. Popper KR., *The logic of scientific discovery*. New York: Harper, 1959.
9. Lakatos I, Musgrave A (eds.). *Criticism and the growth of knowledge*. Cambridge: University Press, 1970.
10. Popper KR. *Objective knowledge (Rev. Ed.)* Oxford: Clarendon Press, 1979.
11. 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
12. 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.
13. Langley P. The status of modeled claims. *J Med Econ*. 2015;18(12):991-992
14. Langley P. Validation of modeled pharmacoeconomic claims in formulary submissions. *J Med Econ*. 2015;18(12):993-999
15. Schommer J, Carlson A, Rhee G. Validating pharmaceutical product claims: questions a formulary committee should ask. *J Med Econ*. 2015;18(12):1000-1006

16. Belsey J. Predictive validation of modeled health technology assessment claims: lessons from NICE. *J Med Econ*. 2015;18(12):1007-1012
17. Wasser T, Haynes K, Barron J, Cziraky M. Using 'big data' to validate claims made in the pharmaceutical approval process. *J Med Econ* 2015;18(12):1013-1019
18. Editorial. *Nature*. 2016;529:256.
19. Camerer CF, Dreber A, Forsell E et al. Evaluating replicability of laboratory experiments in economics. *Science*. 10.1126/science.aaf0918 (2016)
20. Centre for Evidence Based Medicine. COMPare Project. Oxford, 2016. <http://compare-trials.org/results>
21. Wootton D. *The Invention of Science*. New York: Harper, 2015.
22. Grim P, Rosenberger R, Eason RE et al. How simulations fail. *Synthese*. 2011 DOI 10 1007/s11229-011-9976-
23. Beresniak A, Medina-Lara A, Auray JP et al. Validation of the underlying assumptions of the quality-adjusted life years outcome: Results from the ECHOUTCOME European Project. *Pharmacoeconomics*. 2015;33:61-69
24. 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
25. 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
26. 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
27. Ellis G, Silk J. Defend the integrity of physics. *Nature*. 2014 (516): 321323; see also A Frank and M Gleiser. A crisis at the edge of physics. *New York Times*, 7 June, 2015.
28. Langley PC. Imaginary Worlds: Modeled claims for cost-effectiveness published in *Pharmacoeconomics* January 2015 to December 2015. *INNOVATIONS in Pharmacy*. 2016;7(2):No.9
29. Langley PC, Rhee TG. 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):No.16
30. College of Pharmacy. University of Minnesota. Program in Social and Administrative Pharmacy. *Guidelines for Formulary Evaluations* [proposed]. June 2016 (Version 1.0) http://www.pharmacy.umn.edu/sites/pharmacy.umn.edu/files/minnesota_guidelines_july_2016-11.pdf
31. Rajkumar R, Conway PH, Tavenner M. CMS – engaging multiple players in payment reform. *JAMA*. 2014;311(10):1967-8