

7-5-2016

Nullius in verba: The University of Minnesota Social and Administrative Pharmacy Program proposed Guidelines for Formulary Evaluations

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. Nullius in verba: The University of Minnesota Social and Administrative Pharmacy Program proposed Guidelines for Formulary Evaluations. *Inov Pharm*. 2016;7(2): Article 10. <http://pubs.lib.umn.edu/innovations/vol7/iss2/10>

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

Nullius in verba: The University of Minnesota Social and Administrative Pharmacy Program proposed Guidelines for Formulary Evaluations

Paul C. Langley, PhD

College of Pharmacy University of Minnesota

Abstract

The University of Minnesota Social and Administrative Pharmacy Program proposed Guidelines for Formulary Evaluations (GFE) are designed to focus on the credibility of clinical and cost-outcomes claims in formulary decision making. The last few years have witnessed increasing concern over the credibility of trial based efficacy claims, with a surprisingly high proportion of claims falling short. At the same time cost-outcomes claims, where comparative clinical claims are a key input, have been presented where the claims made are not open to experimental or observational assessment. This follows from standards recommended for modeled and simulated claims. In the absence of cost-outcomes claims being presented in an evaluable form, it is impossible not only to replicate the claim or to judge whether or not it is credible. Claims for product value based on such claims are unacceptable. The guidelines proposed here are designed to overcome these limitations and support an evidence base that is both credible and replicable for formulary decisions. This is achieved by focusing on short-term modeled or simulated claims for cost-effectiveness. The requirement for modeled or simulated claims that are evaluable within a time frame that is meaningful for formulary decisions marks a major departure from format submissions already in play, not only in the US but in other developed economies. Rather than subscribing to the gold standard of long term or lifetime cost-per-QALY claims a short-term time horizon of no more than 2-years is recommended. This allows products to be provisionally placed on formulary but subject to a protocol that supports an evaluation to be reported back to a formulary committee as part of ongoing disease area or therapeutic class reviews. The place of the product and its contracted price can then be reviewed.

Acknowledgements: These proposed Guidelines for Formulary Evaluation (GFE) were developed as part of the University of Minnesota Social and Administrative Pharmacy Program. I would like to thank Professor Jon Schommer and Professor Brian Isetts for their support and encouragement.

Keywords: credibility, replication, formulary standards, evaluation, scientific method

Introduction

Over the past 25 years increasing attention has been given to standardizing modeled or simulated claims for the comparative clinical and cost-effectiveness claims for pharmaceutical products. The result has been the publication of literally hundreds of modeled or simulated claims, together with submissions to formulary committees. Unfortunately, in the overwhelming majority of instances the modeled claim has failed to meet the standards of normal science: claims have been presented that are either untestable or presented in untestable terms^{1,2}. This is seen most obviously in claims presented for chronic disease interventions where the

modeled outcomes are in long-term or lifetime cost-per-QALY terms. On the standards of normal science such claims should either be rejected or, if possible, recast in a form that allows them to be evaluated experimentally or observationally.

This fundamental methodological flaw has characterized, not only the various versions of the Academy of Managed Care Pharmacy (AMCP) *Format for Formulary Submissions* in the US but also those exemplar formulary submission guidelines including the National Institute for Health and Care Excellence (NICE) in the UK, the Pharmacy Benefits Advisory Committee (PBAC) in Australia and the Canadian Agency for Drugs and Technologies in Health (CADTH)^{3,4,5,6}. The problem is that none of these guidelines, including the latest version 4 of the AMCP *Format for Formulary Submissions* published in April 2016 meet the standards for 'normal science': they support the construction of modeled claims for clinical and cost-effectiveness that fail to generate testable and reproducible hypotheses (or claims) for the anticipated impact of products in health care systems. As such, they are best considered as imaginary worlds or thought experiments.

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

The recipient of the submission has no idea 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, even harmful, in supporting formulary decisions but to an unknown and unknowable extent.

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 prepared 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, Descartes, 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' and the demarcation of science from nonscience^{7,8}. 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 their assumptions or on whether the model 'represents' reality. It is worth reflecting on Popper's comment on induction: 'never in science are inferences drawn from mere observational experience to the prediction of future events'⁹. Or, to put it simply: not all swans are white.

If we accept the standards of normal science then it is possible to make a distinction between science and pseudoscience. As Pigliucci points out, while the designation 'science' captures a range of disciplines ranging from the 'hard' sciences of controlled experiments to the more 'soft' sciences of non-laboratory assessments, the three core elements of science that characterize and link these disciplines are: (i) it is an investigation of nature; (ii) it is committed to the construction of empirically verifiable

theories; and (iii) it is further committed to testing through observation or experimentation¹⁰. Failure to meet these core elements, notably the ability or commitment to 'produce and test hypotheses based on systematically collected empirical data [via experiment or observation]' is what differentiates science from non-science or pseudoscience. These boundaries are not hard and fast. A field can fall into the pseudoscience category yet practitioners may not see this as a problem even though they refuse to accept the testability criterion.

Outcomes Based Formulary

It is not as though the issue of testable claims has not been raised before in submissions to formulary committees. 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 validation 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 or put to one side, to include non-testable modeled claims and simulations.

More recently, in a supplement to the *Journal of Medical Economics*, 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 or recast in an acceptable format. 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.

The fundamental point is that models or simulations can fail irrespective of claims made that if the simulation is sufficient in its correspondence to reality then it necessarily entails the claims made. The fact is that it is entirely possible to construct competing simulations that come to diametrically opposed conclusions¹⁸. Unless claims are presented as evaluable hypotheses there is no experimental basis for judging the worth of a simulated claim let alone the merits of competing claims.

The purpose of the *Guidelines for Formulary Evaluation* (GFE) proposed here is to argue that the present standards for formulary submissions should be put to one side¹⁹. Current standards, as exemplified in the US with the latest version 4 of the *AMCP Format for Formulary Submissions*, fail to meet the standards of normal science. Their emphasis on the construction of evidence through modeled and simulated claims lacks credibility. There is no commitment to produce and test hypotheses. Rather, as detailed in the GFE, the focus should be on submissions that support short-term evaluable modeled or simulated claims; claims that can be evaluated within a relatively short time frame and reported back to formulary committees as part of ongoing disease area and therapeutic class reviews.

Credibility and Replication

The need to set new standards for formulary submissions and evaluations is, if we accept the importance of subscribing to the standards of normal science, long overdue. There is now increasing concern regarding the limited and, in many cases, dubious evidence base that has supported formulary decisions. This has stemmed, in part, from the inability to replicate results reported for pivotal clinical trials^{20 21}. At the same time, there are well documented instances of outcomes switching where the primary and secondary outcomes claims presented to support formulary listing are inconsistent with the primary and secondary outcomes initially proposed in the original phase 3 protocols²². The re-configuration of outcomes claims and the absence of replication both impact the credibility of clinical claims.

If doubt is expressed as to the credibility of individual clinical claims, then further doubts are raised if disparate clinical

claims are bundled together and re-calibrated to generate indirect comparisons for comparative efficacy for classes of competing therapies.

If, to take a further step, these indirect claims are a key input to modeled or simulated cost-effectiveness claims, credibility is further lessened if claims for cost-effectiveness are presented in a form that prohibits experimentation. In this case, we are not even at first base: there is no way in which these cost-effectiveness claims can be evaluated let alone replicated.

The absence of evaluable claims places a formulary committee in a difficult situation; (i) they can take the non-evaluable claims on board on the basis that some evidence (even if we can't judge its merits) is better than none; (ii) they can take it on board on the grounds that it represents the professionally agreed standard for modeled or simulated yet untestable claims by groups such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and agencies such as the AMCP and NICE; or (iii) they can reject the claim on the grounds that it is purely speculative and lacks any credibility^{23 24}. They might also add that, under (i) and (ii) acceptance of a modeled or simulated claim lays them open to challenge if they are asked to justify a formulary placement and pricing decision.

The purpose of the GFE is to support option (iii). To ensure that claims supporting formulary decisions are credible. Manufacturers and others who submit claims are asked to justify the claims made by presenting those claims in an evaluable form, supported by an assessment protocol. All claims made for products should be treated as provisional. The decision then rests with the formulary committee either to provisionally accept a product subject to the claims being evaluated and reported back to the committee or to put formulary review to one side until a claims assessment has been presented. If an assessment has already been made then these results can be reported to the committee as part of the submission.

Quality and Affordability of Care

Requiring claims for products to be evaluable is essential if clinical and pricing decisions are to be driven by a credible evidence base. Feedback is essential if the credibility of claims is to be maintained. This does not mean evidence from one or more phase 3 trials, but evidence from assessing those claims in the target patient population. The guidelines proposed are designed to support value and care co-ordination initiatives in the delivery of health care to target populations. Consistent with the objectives of the *Patient Protection and Affordable Care Act* the guidelines are designed to support quality outcomes and metrics in health care delivery²⁵. The

guidelines provide a basis for a credible, robust and replicable evidence base in formulary decisions. If such an evidence base is in place then it can usefully support a shift to population-based value payment systems that focus on not just personalized but precision medicine in matching individuals to products.

Evaluating Formulary Submissions

The first part of the GFE is concerned with the credibility and replication of clinical claims for the product and its comparators. All submissions are by invitation from the formulary committee or the health system. Manufacturers and others making a submission are asked, in the first place, to provide a full description of the product and its comparators (with respective product inserts). This should be accompanied by a comparative clinical assessment, reporting both direct and indirect comparisons between the product and comparator therapies.

For each product and comparators the target populations (including potential sub-populations) should be identified and profiled. This description should include an epidemiological profile together with a demographic and clinical profile; the latter to detail co-morbidities. In addition, with the increasing emphasis on personalized or precision medicine, target populations may further be described in terms of their molecular taxonomy or pathology. If this is a consideration, the molecular algorithm should be detailed together with the choice of diagnostic or test device. The place of the product and comparators in therapy should be described together with treatment guidelines supporting product placement.

With increasing concern being expressed over the pricing of drugs, for example in the anticancer area for end-of-life metastatic treatments, a critical input to a formulary decision is the launch price (wholesale acquisition cost) of a product and the current price of comparator therapies for the target indications. The GFE asks manufacturers to indicate what the anticipated launch price of the drug is expected to be (or the current price if it has already been launched). Where a drug has been launched, information should be provided on whether or not there have been price changes since launch. At the same time, for comparison purposes, manufacturers are asked to provide anticipated launch or current prices in other global markets: Canada, the European Union and Australia.

A particular emphasis is placed on the credibility of product and comparator claims. Those making the submission are asked to detail (and provide summary protocols) regarding the original primary and secondary outcomes that the pivotal phase 3 trials were intended to address. Respondents are then asked to detail (i) if there have been any amendments to

the original trial protocol resulting in outcomes switching and (ii) whether the published claims are consistent with the initial trial protocol. Any discrepancies are to be noted. In addition, in respect of the claims made, respondents are asked whether there have been any attempts to replicate published claims for both primary and secondary outcomes and whether the claims have been supported. This requirement applies to both the product and comparators. In addition, if indirect comparative outcome claims have been presented as part of the submission, the respondent is asked whether or not these have been evaluated and replicated.

The respondent is also asked to detail the incidence of significant adverse events for both the product and comparators for (i) the pivotal phase 3 trials and (ii) the post-market entry period. Where there has been a reported replication of the primary and secondary claims, the adverse profile should be reported and matched against those reported for the pivotal trials.

Finally, the respondent is asked to address the issues of adherence and persistence to the index prescription for the product and comparators. For each of the product and comparators evidence is required for adherence and persistence behavior (i) during the pivotal phase 3 trials and (ii) for a period up to 4 years from product launch. In the latter case the evaluation should include evidence for switching between the product and comparators as well as between the comparator products. This review of adherence and persistence should report on descriptive studies as well as those that have attempted to assess the determinants of this behavior in the therapeutic product class. The review should report on the extent to which patterns of adherence and persistence have modified primary clinical claims for the product and comparators in the target population or sub-populations, together with the results of any interventions to impact adherence and persistence behavior.

Cost-Outcomes Claims

The issues of credibility and replication apply equally to cost-outcomes claims as they do to clinical claims. In the second part of the GFE it is proposed that cost-outcomes claims should only be considered credible (and acceptable) if: (i) the claims are evaluable; and (ii) the proposed claims evaluation provides feedback in a meaningful time horizon. Modeled or simulated claims that fail to meet these criteria could be recast in evaluable terms or simply rejected. These would include claims that profess to capture the natural course of a disease, employing Markov or discreet event techniques. These are of no interest if they generate untestable claims for long-term or lifetime cost-per-QALY outcomes. There is no restriction on the type of cost-outcomes claims that can be submitted. Outcomes can be expressed in clinical

terms or as patient reported outcomes (PROs). The GFE does not encourage claims expressed as QALYs. It needs to be demonstrated that QALY and cost-per-QALY claims are a unique feature of the intervention that should be captured and reported in a meaningful time frame. Otherwise, QALY claims should be put to one side.

Where patient reported outcomes are used they should meet accepted standards for their measurement properties with a statement indicating whether or not interval differences are clinically meaningful.

The critical issue is the modeled time frame. As a rule of thumb, a model time-horizon that extends beyond 2 years is unlikely to be acceptable. Long-term or lifetime simulated cost-outcomes claims are considered irrelevant, and potentially misleading, for formulary decisions.

From an evaluation perspective, it is important that claims made are capable of being assessed either from existing data sets (e.g., administrative claims linked to EMRs) to capture resource utilization impacts as well as clinical endpoints or from a prospective observational study to capture outcome such as generic utilities that are not typically captured in EMRs. Those making the submission have the choice, therefore, of tailoring their claims to available data from third party vendors or underwriting a prospective study. Whether the modeled or simulated claims are presented as outcomes from a simple decision model or as a more complex structure such as a Markov model or a discrete event simulation is at the discretion of those preparing the submission. Similarly, it is at the discretion of those preparing the submission whether they present claims as incremental cost-outcomes ratios with supporting sensitivity analyses and probabilistic claims. The focus is on the claims being presented in evaluable terms.

Again, it is also entirely at the discretion of those making the submission whether they want to support their choice of model and the assumptions driving the model by reference to quality checklists such as CHEERS²⁶. The focus on generating testable predictions means that a model is not judged on the correspondence of its input assumptions and core mechanism to some perception of reality. The issue of the sufficiency of correspondence to the real world necessarily entailing the claims made is irrelevant. The merits of a model rest upon its ability to generate testable claims and the results of the assessment of those claims.

Modeled claims should accommodate anticipated adherence and persistence behavior. In many chronic disease states product utilization data point to relatively low rates of adherence and persistence by the end of the first 12 months

from the index prescription. Those presenting a submission should make clear what the anticipated adherence and persistence behavior for the product is expected to be and whether or not it offers advantages over comparator products. Modeled claims should not be based on a possible minority of patients introduced to a new therapy who are assumed to be adherent (e.g., medical possession ratio ≥ 0.8) and persistent over a modeled time horizon of, say, 2 years.

In situations where a manufacturer has already submitted a protocol elsewhere and is underwriting an evaluation this should be reported as part of the submission. In this case, as long as the formulary committee agrees that the protocol is acceptable, then the manufacturer should commit to reporting to the committee once the other study is complete. There is no requirement for any evaluation to be undertaken in the health system's patient population.

Finally, it is not the intention of GFE to support modeled cost-utility claims that are designed to demonstrate that a product or device is 'cost-effective' because it generates incremental claims that are below a notional community willingness to pay threshold. If a formulary committee is prepared to base its decision making on willingness to pay thresholds that is their decision and they should inform those making submissions accordingly.

Claims Assessment Protocol

The claims assessment protocol proposed in the final section of the GFE is the critical link between the claims presented as a result of models or simulations and the subsequent assessment of those claims. The format for a protocol submission, which follows from an earlier protocol standard (PROST), is detailed in Table 1.

The GFE protocol format applies to both retrospective and prospective studies where patients are tracked from existing data sources as well as to prospective observational studies (including randomized trials) that involve reporting from human subjects.

On submission, the protocol is reviewed by the formulary committee. The questions the formulary committee should address are detailed in Table 2 under six subject areas: the objectives of the proposed study; the context; the target population; the claims to be evaluated; the study design and study implementation and reporting.

Discussion

The primary purpose of the GFE is to move the focus of attention in formulary evaluations to standards that are consistent with the scientific method: an acceptance of experimentation and the discovery of new facts in clinical and

cost-effectiveness claims. A shift in focus from an acceptance that constructed evidence driven by models and simulations generating untestable claims is the cornerstone of formulary decision making. This shift in focus does not put aside decision modeling or Markov and discreet event simulation constructs. Rather, it asks that these techniques support short-term evaluable modeled claims that can be assessed and the outcomes reported back to formulary decision makers in a meaningful time frame. A requirement that is not found in the *AMCP Format for Formulary Submissions*.

The issue is one of credibility. Formulary decisions should be evidence driven, transparent and replicable. If clinical claims based on pivotal phase 3 trials show an unfortunate tendency toward non-replicability then claims for comparative cost-effectiveness, which rely upon those same clinical claims as a major input to models or simulations, are in an even more precarious position. Unlike phase 3 claims, where replication is protocol driven, the absence of testable claims from cost-effectiveness models and simulations fail even to get to first base: there are no testable claims to evaluate, let alone replicate.

Over the past 25 years literally hundreds if not thousands of modeled cost-effectiveness claims have been presented and accepted as inputs to formulary decisions. The fact that these typically put forward untestable claims points to a possible alternative explanation for this apparent rejection of the scientific method. The emphasis on models and simulations may represent an acceptance, whether fully articulated or not, of the relativist or equivalence program in science²⁷. This 'strong program' argues that the content of science is explained sociologically. It rejects the notion of superior evidence in favor of the belief that evidence is never discovered, it is constructed within a particular social community. No one body of evidence is superior to another. The success of a research program is judged by its ability to receive the support of that particular community. Truth is consensus. Science is about 'rhetoric, persuasion and authority'. Indeed, a relativist, if we accept this interpretation, may see the fact of a model's untestableness as the characteristic that makes them so valuable, if not essential, to making the best possible policy and clinical care decisions. Modeling, in this belief system, is seen as filling gaps left behind after other forms of evidence have been exhausted, thus providing essential value.

Judged by the standards of normal science, the relativist position is unacceptable and best characterized as pseudoscience. In the absence of experimentation we have no idea whether the claims for clinical benefit or cost-effectiveness are right or even if they are wrong. The fact that the author of an 'untestable' simulation believes the

correspondence of the model to reality is sufficient and that the claims are necessarily entailed is no defense. It puts to one side any notion of progress, the discovery of new facts. It is this commitment to experimentation and observation that underpins these proposed standards for formulary evaluation. If this is unacceptable then it is difficult to see how we can continue to justify the creation and publication of untestable modeled or simulated cost-effectiveness claims without, in the near future, someone pointing out that the emperor has no clothes.

Conclusions

The standards for product evaluation proposed here mark a significant departure from accepted standards for formulary submissions and recommended standards for the modeling and validation of claims. The reason for this is quite simple: current standards lack credibility. Their acceptance by manufacturers, academic groups, professional associations and recipients of modeled claims is beside the point. After all, it took the best part of 1500 years to overthrow the straightjacket of Aristotelian philosophy. Judged by the criteria of normal science, the acceptance of the scientific method, current standards fail to support the process of 'systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses' (OED)²⁸. To the contrary, the standards currently in place suggest an alternative interpretation: a relativist position that rejects any appeal to superior evidence. Science, in this view, is not a way of coming to grips with reality. Evidence is never discovered, it is always 'constructed'. Truth is consensus. In contrast, the standards proposed here for formulary evaluations clearly reject a relativist position. GFE offers a practical guide to formulary evaluations. With a commitment to the standards of normal science, formulary decisions can rest on a credible evidence base. An evidence base that recognizes the need for replication of claims, the fundamental importance of experimentation and observation, and the discovery of new facts to ensure value in the delivery of health care.

References

1. Langley P. Imaginary Worlds: Modeled Claims for Cost-Effectiveness Published in PharmacoEconomics January 2015 to December 2015. *INNOVATIONS in Pharmacy*. 2016;7(2):No.9
2. 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):No.16

3. Academy of Managed Care Pharmacy. Format for Formulary Submissions. AMCP, April 2016
4. National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal*. London: NICE, April 2013
5. Australian Government. Department of Health. *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (version 4.4)*. Canberra: July 2013
6. Canadian Agency for Drugs and Technologies in Health. *Guidelines for the Economic Evaluation of Health Technologies: Canada (3rd Ed)*. CADTH: Ottawa, 2006
7. Popper KR., *The logic of scientific discovery*. New York: Harper, 1959.
8. Lakatos I, Musgrave A (eds.). *Criticism and the growth of knowledge*. Cambridge: University Press, 1970.
9. Popper KR. *Objective knowledge (Rev. Ed.)* Oxford: Clarendon Press, 1979.
10. Pigliucci M. *Nonsense on Stilts: How to tell Science from Bunk*. Chicago: University of Chicago Press, 2010
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. Grim P, Rosenberger R, Eason RE et al. How simulations fail. *Synthese*. 2011 DOI 10.1007/s11229-011-9976-7
19. 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
20. Editorial. *Nature*. 2016;529:256
21. Camerer CF, Dreber A, Forswell E et al. Evaluating replicability of laboratory experiments in economics. *Science*. 10.1126/science.aaf0918 (2016)
22. Centre for Evidence Based Medicine. COMPare Project. Oxford, 2016. <http://compare-trials.org/results>
23. 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
24. 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
25. Rajkumar R, Conway PH, Tavenner M. CMS – engaging multiple players in payment reform. *JAMA*. 2014;311(19):1967-8
26. Husereau D, Drummond M, Petrou S et al. Consolidated health economics evaluation reporting standards (CHEERS) statement. *J Med Econ*. 2013;16(6):713-19
27. Wootton D. *The Invention of Science*. New York: Harper Collins, 2015
28. Oxford English Dictionary: <http://www.oed.com>

Table 1: GFE Protocol Submission Format

Section	Required Content
Title	Proposed Assessment Protocol for Evaluating Clinical and Cost-Outcome Claims for [product]
Abstract	Structured summary of protocol assessment objectives: (i) claims to be evaluated; (ii) target population; (iii) proposed study design; (iv) data sources; (v) timelines for evaluation; and (vi) reporting standard
Context	A statement of the context of the study to include: (i) a systematic review of clinical and modeled cost-effectiveness claims for the product and comparators in the relevant time horizon; (ii) a systematic review of adherence and persistence patterns for the product and comparators in the target population; and (iii) a statement of the contribution this study is expected to make in evaluating competing interventions in this disease or therapy area.
Claims	A statement of the claims that the proposed study will be evaluating: (i) clinical outcomes; (ii) adverse event and safety outcomes; (iii) resource utilization outcomes; (iv) direct medical costs; and (v) cost-effectiveness.
Target Population	Characteristics of target population (age, gender, ethnicity, race) to include sub-groups and clinical markers (presence/exclusion of co-morbidities; stage of disease)
Data	Description of proposed data source(s) and any permissions required to access data/target population (e.g., IRB approvals, EMR approvals, access to administrative claims data) and confidentiality requirements
Study Design	Rationale and description of study design, to include: (i) rationale for choice of study design; (ii) concurrence with good practice guidelines; (iii) treatment comparators; (iv) timeframe; (v) statistical and analysis plan; (vi) data sources; (vii) statistical hypotheses; (viii) description of statistical methods; (ix) sample size; (x) procedures to minimize bias; (xi) quality assurance; (xii) confidentiality; (xiii) data handling; and (xiv) reporting.
Budget	A detailed budget and a commitment to underwriting the evaluation.
Interim Reporting	Proposed schedule for interim reporting of study implementation and progress.
Final Report	Proposed submission data for final report and proposed report outline.

Table 2: GFE: Questions a Formulary Committee Should Ask

Section	Question
Objectives	Has the protocol provided a summary of the study objectives? Does the protocol abstract provide a summary of (i) population targeted; (ii) comparators; (iii) timeframe; (iv) study design; (v) data sources; (vi) endpoints; and (vii) reporting format?
Context	Has the protocol provided a systematic review to establish the context for the claims to be validated and the contribution this validation may have to an overall assessment of comparative treatment benefits and harms in the population targeted and indication for the product?
Target Population	Has the protocol provided a rationale for the population to be targeted, to include (i) demographic characteristics; (ii) clinical status; (iii) disease stage; (iv) co-morbidities?
Claims	Has the protocol provided a summary of the modeled claims for cost-effectiveness to be evaluated? Has the protocol identified claims for (i) clinical benefits; (ii) adverse events and safety; (iii) resource utilization; (iv) direct medical costs; and (v) cost-effectiveness?
Study Design	Has the protocol provided a rationale for the study design (e.g., prospective effectiveness trial, cohort/ observational study) Were alternative study designs proposed? Were risks of bias and other limitations described?
Implementation and Reporting	Has the protocol given a timetable for interim and final reporting of results?