

8-24-2016

Imaginary Worlds: The European Network for Health Technology Assessment (EUnetHTA) Recommendations for Health Economic 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. Imaginary Worlds: The European Network for Health Technology Assessment (EUnetHTA) Recommendations for Health Economic Evaluations. *Inov Pharm*. 2016;7(3): Article 3. <http://pubs.lib.umn.edu/innovations/vol7/iss3/3>

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

Imaginary Worlds: The European Network for Health Technology
Assessment (EUnetHTA) Recommendations for Health Economic
Evaluations

Cover Page Footnote

None

Imaginary Worlds: The European Network for Health Technology Assessment (EUnetHTA) Recommendations for Health Economic Evaluations

Paul C. Langley, PhD

College of Pharmacy University of Minnesota, Minneapolis MN

Abstract

The European Network for Health Technology Assessment (EUnetHTA) guidelines for health economic evaluations represent a consolidated view of non-binding recommendations for assessments of the relative effectiveness of pharmaceuticals or other health technologies. EUnetHTA views itself as the scientific and technological backbone of the development of health technology assessment in the European Union and among its member states and other partners. Unfortunately, the standards for health technology assessment proposed by EUnetHTA do not meet the standards of normal science. They do not support credible claims for the clinical and comparative cost-effectiveness of pharmaceuticals. In rejecting the standards of normal science the guidelines put to one side the opportunity not only to re-assess and replicate clinical and cost-effectiveness claims but to provide meaningful feedback on claims assessment to health care decision makers. The purpose of this review is to make the case that, in failing to support standards for experimentation, EUnetHTA is advocating its partners support the creation of modeled or simulated imaginary or false worlds. While EUnetHTA is not alone in recommending the construction of imaginary worlds to support formulary decisions, there is still the opportunity to revisit these recommendations and decide whether or not to encourage a scientifically rigorous approach to health technology assessments - to abandon a commitment to intelligent design in favor of natural selection.

Keywords: EUnetHTA, pseudoscience, economic evaluations, imaginary worlds, scientific method

Introduction

A hallmark of health technology assessments is the commitment to constructing evidence for the impact of products and devices through modeled or simulated imaginary worlds where the claims made are impossible to evaluate^{1 2 3}. This is a long standing commitment with hundreds of papers published in leading journals all subscribing to imaginary world scenarios^{4 5 6}. At the same time, over the past 25 years, a number of countries have established guidelines to support technical reviews of pharmaceutical products and devices^{7 8 9 10}. These also support the construction of imaginary worlds to make the case for formulary placing and pricing^{11 12}. This is seen most clearly in the National Institute for Health and Care Excellence (NICE) reference case and its mandating lifetime cost-per-quality adjusted life year (QALY) models in chronic disease¹³.

At the same time professional and other groups, such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the Academy of Managed Care Pharmacy (AMCP), have also supported the construction of imaginary worlds through recommendations for standards in modeling clinical and cost-outcomes claims^{14 15}.

Over the past few months, however, there has been an effort to point out that, set against the standards of normal science in its commitment to hypothesis testing and ongoing experimentation, modeled and simulated claims in health technology assessments typically fail to meet accepted scientific standards. Rather than a commitment to developing testable clinical and cost-effectiveness claims, supporting replication in target populations and providing feedback to health system decision makers, the commitment to non-testable claims makes any application of the standards and processes or normal science impossible. The result, at least in the US, is that claims based on models or simulations play little, if any role, in health care decision making.

The purpose of this review is to consider whether the methods proposed by EUnetHTA for health economic evaluations meet the standards of normal science. The question raised is whether the standards proposed are consistent with, or even recognize, the potential for evaluation, falsification and replication in clinical and comparative effectiveness claims. This question applies both to the recommendations for health technology assessment as well as the associated relative effectiveness assessments^{16 17 18}.

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 EUnetHTA RECOMMENDATIONS

The purpose of the EUnetHTA guideline is twofold: firstly, to set a general framework for how to conduct economic evaluations and, secondly, to increase the transferability of economic evaluations among EUnetHTA partners. The

guideline itself is based on a review of guidelines developed by 25 of the 33 countries in the EUnetHTA collaborating partnership (28 EU member states together with Norway, Switzerland, Russia and Turkey). EUnetHTA was established in 2006 and is seen as the 'scientific and technological backbone of the health technology assessment development in the European Union and its member states'¹⁹. The present recommendations follow from, in European Union terminology, the Joint Action 2 (JA2) aim to bring collaboration to a higher level in establishing a sustainable health technology assessment structure as detailed in the May 2016 Final Technical Report²⁰. Although the focus here is on the recommendations for economic evaluations and relative effectiveness assessments (REAs), it should be noted that these are elements in the health technology assessment CORE model domains which also encompass issues of ethics, organizational aspects, patients and social aspects, and legal aspects as dimensions of value²¹.

Methods for Health Economic Evaluations

The main recommendations for what EUnetHTA describes as the reference case are:

- Type of analysis: results to be presented as both a cost-effectiveness (CEA) and cost-utility analysis (CUA), with a cost-minimization analysis (CMA) where no difference between an intervention and its comparator (or a cost-consequences analysis [CCA])
- Clinical evidence from a systematic review of the literature, with RCTs the most appropriate source
- Time horizon: long enough to reflect all important relevant differences in costs and outcomes between the technologies being compared
- Use of Models: decision model, where choice of model depends on research question
- Perspective: health care and possibly societal perspective
- Costs: resource utilization in units with country specific costs
- Outcome measure: natural units (including life years) and as QALYs
- Discounting: base case discount rate between 3 and 5 per cent
- Results: CEA and CUA presented as absolute and incremental values
- Uncertainty: deterministic and probabilistic sensitivity analysis

Although the focus of these reference case recommendations is to improve transferability of results between EUnetHTA partners through a common report structure, little detail is given as to how an economic evaluation is to be undertaken.

As they stand, they represent a 'common view' on which the various EUnetHTA partners find basic agreement. The recommendations reflect clearly the impact of the NICE reference case, although they stop short of the more specific requirements for a NICE reference case model including recommendations for cost-QALY thresholds.

Relative Effectiveness Assessment

To clear up any initial confusion: the term relative effectiveness assessment does not apply to the formulation of comparative clinical and cost-effectiveness claims in a form that is amenable to empirical evaluation and replication in target populations. Rather, the term applies to: *the extent to which the effects observed in clinical studies are likely to reflect the expected results when a specific intervention is applied to the population of interest*. The relative effectiveness of interventions is to be judged from 'suitable' clinical trial data; from pragmatic trials that have the 'noise of practice' rather than from 'explanatory' trials conducted within a strict trials setting. Given this, the assessor of relative effectiveness should always indicate the likeliness that the available evidence is applicable to the decision problem.

A completed relative value assessment is intended to provide a set of recommendations for the selection and assessment of clinical endpoints, broadly categorized as mortality, morbidity, clinical status, symptoms, function and health related quality of life (HRQoL). In the case of the clinical endpoints, the relative effectiveness assessment is regarded as a measure of how a patient feels, functions or survives. The clinical endpoint should be reproducible and valid, to facilitate comparisons across studies and jurisdictions. At the same time, in chronic disease they should be long-term or final endpoints with all-cause mortality used where relevant as it is seen as the most unbiased endpoint. Overall survival is the preferred endpoint in survival analyses. Extrapolation from intermediate to final endpoints should be underpinned by a clear biological or medical rationale or a strong or validated link, otherwise the focus should be on intermediate endpoints. HRQoL measures are not considered adequate as the primary endpoint in a relative value assessment.

Discussion

The Standards of Normal Science

The requirement for testable hypotheses to support the discovery of new facts is unexceptional. Since the 17th century it has been recognized that if a research agenda is to advance, if there is to be an accretion of knowledge and if models are to generate meaningful hypotheses, then these hypotheses must be such that they can be empirically evaluated. This position has been well documented by Wootton in his reassessment of the use of language in the idea of the scientific revolution in *The Invention of Science*²².

If a model fails to generate testable or measurable hypotheses, then it should be seen as simply a construct to support the exploration of imaginary worlds or thought experiments and not part of a meaningful research program; a program that underpins the notion of progress in the accumulation of knowledge. Consider the motto of the Royal Society, first meeting in 1660 and a Royal Charter in 1662, *nullius in verba* – ‘take no man’s word for it’.

Replication and Claims Assessment

By 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’^{23 24}. 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. At the same time there is the need for replication. This has been brought home in the last few years by attempts to replicate clinical trial and laboratory claims. There has been increasing concern expressed over the presence of repetitive flaws and the need for guidelines to improve experimental reproducibility. 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’²⁵.

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 original²⁶. As the authors note: the deepest trust in scientific knowledge comes from the ‘ability to replicate empirical findings’ although rarely carried out in the social sciences.

Unfortunately, if these EUnetHTA recommendations reflect a common European view on health economic evaluations, then they represent a commitment to the creation of modeled or simulated imaginary worlds. There is, apparently, no concern for the standards of normal science and the commitment to evaluable clinical and cost-outcome claims, experimentation and replication. There is no concept of a research program in economic evaluations to support an accumulation and exchange of knowledge between health care systems as to the impact of pharmaceuticals and devices in target patient populations.

As far as can be ascertained, relative effectiveness assessments are not intended to support prospective experimentation: the discovery of new facts through experimentation and replication of comparative clinical claims in target populations. Although those developing a

relative effectiveness assessment are asked, in the case of clinical endpoints, that they be reproducible and valid, this refers to the choice of endpoints, in a hierarchy of endpoints. The focus is on measures that may support potential replication, not on the role of ongoing replication studies to assess the merits of clinical claims in target populations.

While this rejection of experimentation may reflect an implicit recognition that there are limited data for actually evaluating and replicating claims across even a handful of the partners under the EUnetHTA umbrella or simply a failure to appreciate that the standards of normal science are equally applicable in health technology assessments is a moot point. The key point is that there is no commitment to hypothesis testing. Existing clinical data are taken at face value, pragmatic trials may be assigned a greater weight in systematic reviews and the assessor’s role is to judge the applicability of reported results to the ‘likely’ impact of an intervention on target populations in specific jurisdictions. Presumably these ‘likely’ claims, assessed treatment benefits, are then an input to a decision model to judge, in turn, the ‘likely’ cost-effectiveness of the intervention in the target population in one or more of the EUnetHTA partners’ jurisdictions. There are no recommendations for these presumed ‘likelihoods’ to be explored. This is unfortunate because the standards proposed for relative effectiveness assessments have the potential to provide a robust platform both for clinical claims assessment and replication, as well as a platform for evaluable cost-outcomes claims.

The EUnetHTA Reference Case Recommendations

The EUnetHTA recommendations for a base-case reference model to support the transferability of results between consortium partners rests upon a methodology that supports the creation of imaginary worlds. Claims for comparative cost-outcomes are, given the recommended standards, impossible to evaluate. The recommendations endorse the creation of lifetime cost-per-QALY claims in order to capture, for possibly decades into the future, all relevant differences in costs or outcomes, appropriately, discounted, between the technologies being compared. Partners to the EUnetHTA consortium are encouraged to build their own versions of the base case with the result that the various partners can then compare their various imaginary world constructs and their claims for comparative, constructed cost-per-QALY estimates.

Unfortunately, apart from the issue of choosing between the various generic QALY measures as the outcome ‘gold standard’, the principal objection to a cost-per-QALY reference case as the basis for transferability of modeled claims is that no health system collects QALY measures as a standard in, for example, electronic medical records or

administrative records. The only way a QALY claim could be evaluated (and assuming it was presented in a form that could be reported on in a 2-3 year timeframe) would be through a comparative prospective observational study. Expressing claims in QALY-terms, therefore, sets up an immediate barrier to claims evaluation and replication²⁷.

The commitment in the EUnetHTA recommendation to the construction of imaginary worlds puts to one side the standards of normal science. Rather than recommendations focused on the acceptance of hypothesis testing, falsification and replication in claims made for pharmaceutical products and devices, at no stage in the guidelines is there any mention of the potential for developing testable claims for product impact or the feasibility of evaluating these claims through hypothesis testing, falsification and replication. These could involve claims, not only for clinical outcomes, but adherence and resource utilization in target patient populations. To all intents and purpose, if we subscribe to the EUnetHTA reference case, the standards of normal science appear to have been put to one side in favor of non-evaluable claims based on models or simulations²⁸.

Certainly, attempts to evaluate comparative claims for cost-effectiveness in many countries are made difficult by the absence of data. The EUnetHTA partners are not as well served as the United States in access to big data. This is not an excuse for asking for formulary decisions to be based on constructed simulations that create non-evaluable claims. The fundamental objection to simulations driving decision making is that there is no opportunity, notably in long term or lifetime simulations, to generate feedback. Health systems that have relied on reference case modeling and thresholds to support formulary decisions have no idea, in the absence of experimental data, whether those claims are right or even if they are wrong. We also have little guidance as to whether, in selecting RCT claims as inputs to models or simulations, those claims meet the standards for replication. If not, then further uncertainty is cast on the validity of a reference case model or simulation. Health system decision makers are, therefore, in an unenviable position. If they accept the EUnetHTA reference case, then decisions could be challenged on the grounds that they lack testable predictions and experimental confirmation. If they insist on an appropriate evidence base, then they should require manufacturers to provide this, putting to one side the reference case simulation standards.

Relativism and Comparative Claims

Acceptance of decisions driven by simulated or modeled reference cases is to accept a relativist position. 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 valid. 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. Indeed, there is no objection to basing clinical claims on simulations in lieu of funding the more expensive RCT. For the relativist, and this applies to the subject area and community of 'pharmacoeconomists', 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 EUnetHTA 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.

Unfortunately, there are no recognized criteria for choosing between modeled or simulated claims that generate untestable predictions. They are accepted because their proponents believe it is possible to capture the critical or similar features of the reality of a decision. 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 corresponded to reality²⁹. But simulations can be challenged; simulations can fail and simulations can be 'reverse engineered' to generate required comparative effectiveness claims and still meet recommended guidelines. Absent the ability to evaluate claims, we have no basis for claiming that one simulation, and its support for a sponsor's product is superior to another. The claims are immune to failure.

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 ICER cloud diagrams, we can apply thresholds and we can even introduce an analysis of the value of perfect information. Nevertheless, the fundamental objection still applies: there are no testable predictions and, as such, the simulation fails the standards of normal science. They are appropriately viewed as pseudoscience; sharing the platform with intelligent design and not natural selection³⁰.

Conclusions

If we put to one side the EUnetHTA reference case recommendations for health technology evaluations, is there an acceptable methodology to support health system decision making? Recently, the Program in Social and

Administrative Pharmacy, University of Minnesota proposed guidelines for formulary evaluation that were designed to meet the standards of normal science³¹. The key points in these guidelines are: (i) clinical and cost-outcomes claims should be evaluable; (ii) claims should be evaluable in a time horizon that is meaningful to health care decision makers (e.g., feedback in 1-2 years); (iii) if possible, claims should be evaluable from existing data sources (e.g., electronic medical records); (iv) a protocol detailing how the claims will be evaluated should accompany formulary submissions; and (v) the protocol-based evaluation should be underwritten by the manufacturer.

Whether EUnetHTA is prepared to withdraw their current reference case recommendations for health technology assessment is an open question. Possibly too much has been invested in formulating common-core guidelines to support constructed, imaginary world claims to pull back now. Against this, there may be individuals within the EUnetHTA partnership who may accept the arguments put forward here and accept the need to put health care decision making on a defensible evidence base.

References

- ¹ Langley P. The status of modeled claims. *J Med Econ*. 2015;18(12):991-992
- ² Langley P. Validation of modeled pharmacoeconomic claims in formulary submissions. *J Med Econ*. 2015;18(12):993-999
- ³ Langley PC. Modeling imaginary worlds: Version 4 of the AMCP Format for Formulary Submissions. *INNOVATIONS in Pharmacy*. 2016;7(2):11 <http://pubs.lib.umn.edu/innovations/vol7/iss2/11>
- ⁴ Langley PC. Imaginary Worlds: Modeled claims for cost effectiveness published in *PharmacoEconomics* January 2015 to December 2015. *INNOVATIONS in Pharmacy*. 2016;7(2):9 <http://pubs.lib.umn.edu/innovations/vol7/iss2/9>
- ⁵ Langley PC. Imaginary Worlds: Modeled claims for cost-effectiveness published in the *Journal of Medical Economics* January 2015 to December 2015. *INNOVATIONS IN Pharmacy*. 2016; 7(2):16 <http://pubs.lib.umn.edu/innovations/vol7/iss2/16>
- ⁶ Langley PC. Imaginary Worlds: Modeled Claims for cost-effectiveness published in *Value in Health* January 2015 to December 2015. *INNOVATIONS in Pharmacy*, 2016 ; 7(2):18 <http://pubs.lib.umn.edu/innovations/vol7/iss2/18>
- ⁷ National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal*. London: NICE, April 2013
- ⁸ Canadian Agency for Drugs and Technologies in Health. *Guidelines for the Economic Evaluation of Health Technologies: Canada* (3rd ed.). Ottawa: CADTH, 2006.
- ⁹ Australian Government. Department of Health. *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (version 4.4)*. Canberra: July 2013.
- ¹⁰ New Zealand. Pharmaceutical Management Agency. Prescription for Pharmacoeconomic Analysis: Methods for cost-utility analysis (v 2.1). PHARMAC, 2012.
- ¹¹ Langley PC. Na domhain shamhlaíochta: formulary submission guidelines in Ireland and the standards of normal science. *Curr Med Res Opin*. 2016;32(5): DOI:10.1080/03007995.2016.1190699
- ¹² Langley PC. He ao pohewa: The PHARMAC Prescription for Pharmacoeconomic Analysis in New Zealand and the standards of normal science. *INNOVATIONS in Pharmacy*. 2016; 7(2):13 <http://pubs.lib.umn.edu/innovations/vol7/iss2/13>
- ¹³ Langley PC. Sunlit uplands: The genius of the NICE reference case. *INNOVATIONS in Pharmacy*. 2016;7(2):12 <http://pubs.lib.umn.edu/innovations/vol7/iss2/12>
- ¹⁴ 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
- ¹⁵ 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
- ¹⁵ 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
- ¹⁶ European Network for Health Technology Assessment (EUnetHTA). Methods for health economic evaluations. May 2015. <http://eunetha.eu/outputs/eunetha-methodological-guideline-methods-health-economic-evaluations>

- ¹⁷ European Network for Health Technology Assessment (EUnetHTA), Guideline: Levels of Evidence – Applicability of evidence for the context of a relative effectiveness assessment. November 2015. <http://eunethta.eu/outputs/levels-evidence-applicability-evidence-context-relative-effectiveness-assessment-amended-ja1>
- ¹⁸ Academy of Managed Care Pharmacy. *Format for Formulary Submissions (Version 4)*. AMCP: April 2016
- ¹⁹ Kristensen FB. The current and future lifecycle approach to decision making on health technologies. *Value & Outcomes Spotlight*. 2016;2(3):12-13
- ²⁰ European Network for Health technology Assessment (EUnetHTA) Joint Action 2 (2012 – 2015) Final Technical report. Executive Summary. May 2016
http://eunethta.eu/sites/5026.fedimbo.belgium.be/files/EUnetHTAJA2_FinalTechnicalReport_Executive%20Summary_May2016.pdf
- ²¹ European Network for Health Technology Assessment (EUnetHTA). CORE Model. 2015. <http://eunethta.eu/hta-core-model>
- ²² Wootton D. *The Invention of Science*. New York: Harper Collins, 2015
- ²³ Popper KR., *The logic of scientific discovery*. New York: Harper, 1959.
- ²⁴ Lakatos I, Musgrave A (eds.). *Criticism and the growth of knowledge*. Cambridge: University Press, 1970.
- ²⁵ Editorial. *Nature*. 2016;529:256.
- ²⁶ Camerer CF, Dreber A, Forswell E et al. Evaluating replicability of laboratory experiments in economics. *Science*. 10.1126/science.aaf0918 (2016)
- ²⁷ Langley PC. Great Expectations: Cost-utility models as decision criteria. *INNOVATIONS in Pharmacy*. 2016; 7(2):14
<http://pubs.lib.umn.edu/innovations/vol7/iss2/14>
- ²⁸ Langley PC. Supporting Formulary Decisions: The discovery of new facts or constructed evidence? *INNOVATIONS in Pharmacy*. 2016;7(2):15 <http://pubs.lib.umn.edu/innovations/vol7/iss2/15>
- ²⁹ Grim P, Rosenberger R, Eason RE et al. How simulations fail. *Synthese*. 2011 DOI 10.1007/s11229-011-9976-7
- ³⁰ Pigliucci M. *Nonsense on stilts: How to tell science from bunk*. Chicago: University of Chicago Press, 2010
- ³¹ College of Pharmacy. University of Minnesota. Program in Social and Administrative Pharmacy. Guidelines for Formulary Evaluation [proposed]. June 2016 (version 1.0)
http://www.pharmacy.umn.edu/sites/pharmacy.umn.edu/files/minnesota_guidelines_july_2016-11.pdf