

7-8-2016

Imaginary Worlds: The Status of Simulation Modeling in Claims for Cost-Effectiveness In Diabetes Mellitus

Paul C. Langley

University of Minnesota, langley@maimonresearch.com

Taeho Greg Rhee

University of Minnesota, Twin Cities, rhee0041@umn.edu

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

Recommended Citation

Langley PC, Rhee TG. Imaginary Worlds: The Status of Simulation Modeling in Claims for Cost-Effectiveness In Diabetes Mellitus. *Inov Pharm.* 2016;7(2): Article 17. <http://pubs.lib.umn.edu/innovations/vol7/iss2/17>

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

Imaginary Worlds: The Status of Simulation Modeling in Claims for Cost-Effectiveness In Diabetes Mellitus

Cover Page Footnote

None

Imaginary Worlds: The Status of Simulation Modeling in Claims for Cost-Effectiveness In Diabetes Mellitus

Paul C Langley, PhD and Taeho Greg Rhee, PhD(c)
College of Pharmacy, University of Minnesota, Minneapolis, MN

Abstract

Over the past 20 years a number of simulations or models have been developed as a basis for tracking and evaluating the impact of pharmacological and other interventions in type 1 and type 2 diabetes mellitus. These models have typically tracked the natural course of these diseases generating long-term composite claims for cost-effectiveness. These claims can extend over the lifetime of the modeled patient cohort. Set against the standards of normal science, however, these claims lack credibility. The claims presented are all too often either immune to failure or are presented in a form that is non-testable. As such they fail to meet the key experimental requirements of falsification and replication. Unfortunately, there is a continuing belief that long-term or lifetime models are essential to decision-making. This is misplaced. The purpose of this review is to argue that there is a pressing need to reconsider the needs of health system decision makers and focus on modeled or simulated claims that are meaningful, testable, reportable and replicable in evaluating interventions in diabetes mellitus.

Keywords: diabetes claims, cost-effectiveness modeling, simulation, credibility, replication, imaginary worlds, scientific method, pseudoscience

Introduction

Over the past 20 years, considerable time and effort has been put into building simulation models to support the evaluation of competing interventions in type 1 and type 2 diabetes mellitus. These simulations are seen as critical inputs to informing health care decision makers for disease states which have a significant long-term, if not lifetime, impact on both patients and health care payers. The IMS CORE diabetes model, for example, has been 'used extensively to evaluate the cost-effectiveness of new therapies ... and to inform reimbursement decisions, public health issues, resource planning, clinical trial design, and optimal patient strategies'¹

The purpose of this review is to argue that the faith placed by decision makers in, and their apparent acceptance of claims made by these various models are misplaced. Rather than these models being seen as a robust and credible framework for evaluating the merits of competing products, notably in the claims made for the cost-effectiveness of new therapies, the models may, to an unknown extent, be misleading. This follows from the apparent disregard for the standards of normal science; specifically, the ability of these models to generate testable predictions to support falsification and replication².

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

Credibility and the Standards of Normal Science

The requirement for testable hypotheses in product and device impacts 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 and if models are to generate meaningful hypotheses, then these hypotheses must be empirically evaluated. This argument has been well documented by Wootton in his recent reassessment of the use of language in the idea of the scientific revolution in *The Invention of Science*³. 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. A process described in a recent issue of *Science* as one where 'The deepest trust in scientific knowledge comes from the ability to replicate empirical findings directly and independently, whether through reanalyzing original data or by creating new data'⁴. Unfortunately, in both the natural and social sciences replication is rarely carried out. Where attempts to replicate clinical claims have been undertaken, for example in psychology, false positives seem to be the norm in over 60% of replicated studies⁵.

The key question is whether or not published simulated or modeled claims for comparative cost-effectiveness are credible? If claims are testable in a timeframe and for a target population that creates a meaningful and evidence-based feedback to a formulary committee, then the model meets the standards expected in normal science. It is not a question of the sophistication of the model. A further issue is whether

or not modeled claims can be replicated. As Popper notes, 'non-reproducible single occurrences are of no significance to science'³. If they can be reproduced, they contribute to the discovery of new facts, establishing the credibility of the claims made cumulatively. The issue is one of demarcation. Are we to regard models and simulations as science or pseudoscience? Are they designed to discover new facts or to generate constructed evidence? If models and simulations are not designed to support credible and empirically evaluable theories and hypotheses then they are best seen as pseudoscience in their construction of imaginary worlds⁶.

Standards for Cost-Effectiveness Modeling in Diabetes

A recent review by Henriksson et al. is important because it points to a lack of appreciation of the need to apply the standards of normal science to simulations or modeled claims in diabetes⁷. The purpose of their review was to evaluate the capabilities of diabetes models. In this case in respect of type 1 diabetes although many of the models could equally accommodate type 2 diabetes simulations. Model capabilities were assessed in term of the application of ISPOR modelling standards^{8,9}.

The Henriksson review identified 13 simulations or models in type 1 diabetes. Ten of these models were Markov-based, two used discrete event simulation and one was the Archimedes object-orientated mathematical model. The cycle lengths for the Markov and discrete event simulation (with one exception) were annual with time horizons ranging from 8 years to a lifetime. Four of the models were flexible to a lifetime simulation, one model only considered a lifetime horizon, with the timeframes for the other models ranging from 8 years to 100 years (8, 10, 20, 33, 36, 40, 100 years). The most common outcomes were: direct costs, life years, quality-adjusted life years, incremental cost effectiveness ratios and cumulative incidence of complications. Nine of the models had apparently gone through some form of validation, to include external validation, internal validation and cross validation. Details on the form of validation were sparse with the authors mentioning only the CARDS trial at the Fourth Mount Hood Challenge, which simulated 9-year microalbuminuria, retinopathy, and neuropathy outcomes.

Following ISPOR good practice standards, Henriksson et al identified a range of key attributes for best practice in type 1 diabetes models. These attributes included the type of simulation, the treatment of uncertainty, choice of maximum time horizon, the inclusion of microvascular and macrovascular health states, adverse events, costs and QALYs. As far as the time horizon the model was concerned the authors maintained that "Theoretically, a lifetime horizon should be supported in models of chronic and progressive diseases like T1DM; however any model with a maximum

time horizon of ≥ 30 years is likely sufficient in practice". The authors see quality adjusted life years (QALYs) as the gold standard as they "facilitate comparisons across interventions and are the most important measure of benefit for many new interventions".

The issue of testable claims does not arise; standards for falsification and replication are not an issue. The focus is on inputs and the ability of the simulation or model to replicate the natural course of a chronic disease state. There seems to be no consideration of the evidentiary standards that a payer might demand when claims for formulary listing are presented as discounted cost-per-QALYs over a 30-year time horizon. A timeframe that ensures claims are immune to falsification.

This lack of consideration of the possibility of meeting the standards of normal science and developing models that generate testable claims also characterizes two earlier reviews of models in type 2 diabetes^{10,11}. Both reviews see modeled claims extending over mid- to long-term horizon as appropriate to meeting the information needs of decision makers. The question of the credibility of a claim and the possibility of falsification and replication are, again, not an issue.

It is worth pointing out, however, that in the more comprehensive of the two reviews, Tarride et al., where 17 type 1 and type 2 models were considered a number of limitations in respect of the RCTs used to populate the models were noted. These included the model structure and assumptions, the generalizability of RCT results, possible changes in treatment patterns since RCT trials were initiated and the generalization of models across treatment settings and countries. The authors' response, rather than any reference to the standards of normal science, is to suggest, given the number of models and their inherent variability, that a reference case to capture clinical and utility data should be developed for all new models. This has not occurred.

Discussion

The recommendation by Tarride et al for a reference standard raises the question of the justification for simulated claims. Does the fact that long-term or lifetime simulated claims can be published, even though they are immune to failure, point to an acceptance of what may be described as a relativist position where no body of evidence is superior to another? Is the success of a research program focused on constructing and comparing type 1 and type 2 diabetes simulations reflective of a consensus opinion that in models or simulations evidence is never discovered, merely constructed? Instead of coming to grips with reality and

applying the standards of normal science to the discovery of new facts through experimentation, is the reality of science about rhetoric, persuasion and authority?

More to the point: can simulations fail? Or, in the argument for a reference or benchmark simulation, are we attempting to capture the critical or similar features of a 'diabetic' reality? Are we being asked to subscribe to a belief that if the simulated input conditions and the simulated core mechanism of the diabetes model correspond to reality, the sufficient condition character of the simulation assures us that the output is necessarily entailed and predictions for the target diabetes population must correspond to reality? The point surely is that simulations can fail. They can fail in their correspondence to input conditions, to the core mechanism and to their predictions. Indeed, the lack of correspondence and multiple failures to meet benchmarks set by randomized clinical trials have been amply demonstrated in the Mount Hood challenges^{12 13}. Indeed, it is always possible to reverse engineer a simulation to generate a required outcome! If we accept that simulations can fail, then it seems paradoxical that we persist in publishing and comparing simulated output claims that are immune to falsification. Or, if the claims are within a timeframe that is open to empirical evaluation, failing to provide evidence that the claims meet the standards of normal science. It seems pointless to ask for a reference case model if, as in the case of the reference case required by the National Institute for Health and Care Excellence, it merely entrenches modeled or simulated claims that are immune to failure¹⁴.

As a case study, consider a recent validation of the IMS Core diabetes model. Although the authors recognize the potential

role and power of predictive validation, this is put to one side on the grounds that the time frame for completing their analysis precludes this approach. Putting predictive validation to one side, as noted above, is to put to one side the standards of normal science. There is no consideration given to ability of the IMS CORE model to generate testable claims in target populations; of the need to distinguish claims that are potentially evaluable from those that, in the time frame of the simulation, are immune to failure.

Conclusions

To the extent that the supporters of simulated diabetes models believe that the model can be defended on the grounds that it 'reflects reality', or is a 'reasonable representation' of what is 'out there' (whatever those terms actually mean), it is worth reflecting on Popper's statement that 'never in science are inferences drawn from mere observational experience to the prediction of future events'¹⁵. Or, to put it in more practical terms: not all swans are white. Reporting simulated model results for 30 and 40 year time frames is clearly a challenge to credibility, if not credulity. Attempting to justify a model's credibility on the 'realism' of its assumptions is not acceptable.

If simulations or models in type 1 and type 2 diabetes are to be seen as credible, then there needs to be a commitment to move away from generating claims that are immune to failure and to focus instead on developing claims that can be empirically evaluated. This does not mean competition between models to assess their correspondence to benchmarks, but to develop claims that can be independently assessed in target diabetes populations.

References

- ¹ McEwan P, Foos V, Palmer JL et al. Validation of the IMS CORE Diabetes Model. *Value in Health*. 2014;17:714-724
- ² Langley PC. Validation of modeled pharmacoeconomic claims in formulary submissions. *J Med Econ*. 2015;18(12):993-999
- ³ Wootton D. *The Invention of Science*. New York: Harper Collins, 2015
- ⁴ Camerer et al., *Science* 10:1126/science.aaf0918 (2016)
- ⁵ Open Science Collaboration. Estimating the reproducibility of psychological science. *Science* 349.aac4716 (2015)
- ⁶ Langley PC. Supporting formulary decisions: The discovery of new facts or constructed evidence. *INNOVATIONS in Pharmacy*. 2016;7(2): Article 15
- ⁷ Henriksson M, Jindal R, Sternhufvud C et al. A systematic review of cost-effectiveness models in type 1 diabetes mellitus. *Pharmacoeconomics*. 2016. DOI 10.1007/s40273-015-0374-8

- ⁸ 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
- ⁹ 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
- ¹⁰ Yi Y, Phillips Z, Bergman G et al. Economic models in type 2 diabetes, *Curr Med Res Opin*, 2010;26(9):2105-18
- ¹¹ Tarride JE, Hopkins R, Blackhouse G et al. A review of methods used in long-term cost-effectiveness models of diabetes mellitus treatment. *PharmacoEconomics*. 2010;28(4):255-77
- ¹² Mount Hood 4 Modeling Group. Computer modeling of diabetes and its complications: a report on the Fourth Mount Hood Challenge Meeting. *Diabetes Care*. 2007;30(6):1638-46
- ¹³ Palmer AJ, Clarke P, Gray A et al. Computer modeling of diabetes and its complications: a report on the Fifth Mount Hood challenge meeting. *Value Health*. 2013;16(4):670-85
- ¹⁴ Langley PC. Sunlit Uplands: The genius of the NICE reference case. *INNOVATIONS in Pharmacy*. 2016;7(2): Article 12
- ¹⁵ Popper KR. *Objective Knowledge* (Rev. Ed.) Oxford: Clarendon Press, 1979.