
Article:
Bermejo, I. orcid.org/0000-0001-9105-8088, Tappenden, P. and Youn, J.H. (2017) Replicating Health Economic Models: Firm Foundations or a House of Cards? Pharmacoeconomics. ISSN 1170-7690

https://doi.org/10.1007/s40273-017-0553-x

Replicating health economic models – firm foundations or a house of cards?

Bermejo I, Research Associate, School of Health and Related Research, University of Sheffield

Tappenden P, Reader in Health Economic Modelling, School of Health and Related Research, University of Sheffield

Youn J, Research Fellow, Division of Population Health, Health Services Research & Primary Care, University of Manchester

Abstract

Health economic evaluation is a framework for the comparative analysis of the incremental health gains and costs associated with competing decision alternatives. The process of developing health economic models is usually complex, financially expensive and time consuming. For these reasons, model development is sometimes based on previous model-based analyses: this endeavour is usually referred to as model replication. Such model replication activity may involve the comprehensive reproduction of an existing model or "borrowing" all or part of a previously developed model structure. Generally speaking, the replication of an existing model may require substantially less effort than developing a new de novo model by bypassing, or undertaking in only a perfunctory manner, certain aspects of model development such as the development of a complete conceptual model and/or comprehensive literature searching for model parameters. A further motivation for model replication may be to draw on the credibility or prestige of previous analyses which have been published and/or used to inform decision-making. The acceptability and appropriateness of replicating models depends on the decision-making context: there exists a trade-off between the "savings" afforded by model replication and the potential "costs" associated with reduced model credibility due to the omission of certain stages of model development. This paper provides an overview of the different levels of, and motivations for, replicating health economic models, and discusses the advantages, disadvantages and caveats associated with this type of modelling activity. Irrespective of whether replicated models should be considered appropriate or not, complete replicability is generally accepted as a desirable property of health economic models, as reflected in critical appraisal checklists and good practice guidelines. To this end, the feasibility of comprehensive model replication is explored empirically across a small number of recent case studies. Recommendations are put forward for improving reporting standards to enhance comprehensive model replicability.

Key points for decision makers

- Model replicability is generally perceived to be an indicator of the quality of published models.
- Model replication is associated with both advantages and disadvantages. Replication may be quicker and less expensive than developing a de novo model, however model authors should be aware that these "savings" may impact upon the credibility of the model.
- Our pilot study indicates that even amongst a very small sample of studies, the majority of the models considered could not be fully replicated.

INTRODUCTION

Health economic evaluation represents a framework for the comparative analysis of the incremental health gains and costs associated with competing decision alternatives. Given that a single source of evidence, such as a randomised controlled trial (RCT), rarely provides all of the relevant information necessary to estimate costs and health outcomes for all decision alternatives, mathematical models are typically required [1]. However, the process of health economic model development is commonly complex, financially expensive and time consuming. For these reasons, the development of a model may draw on previous model-based analyses. The extent to which this is done varies between analyses: sometimes the new model is based only on an existing model structure, whilst in other instances the new model may be developed with the intention of fully reproducing the functionality and results of a previous model. This broad endeavour is usually termed model replication.

The replication of models is generally undertaken using a publication and/or other written documentation describing the existing model. Consistent with the basic principles of the scientific method [2], complete replicability, at least in principle, is generally perceived to be a desirable property of a model and may be considered as one facet of the quality of a model-based analysis. This view is well supported within the literature. For example, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force on Good Research Practices states that *"the description of the model should be sufficiently detailed that the model can be replicated mathematically"* [3]. Similarly, the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist, which has been adopted by many journals to improve the quality of reporting of economic evaluation studies, recommends the reporting of all assumptions and model parameters that would be needed by a reader to potentially reprogram the model and to replicate its findings [4]. This general sentiment is also mirrored within other critical appraisal checklists and related literature on the quality assessment of health economic models [5-7].

However, these requirements for complete replicability of models are not typically fully enforced. Given that the focus of such recommendations lies in the potential rather than the actual replication of models, authors and publishers must inevitably apply discretion in reaching judgements about whether model-based health economic analyses can be considered truly replicable. The extent to which this criterion is met within published analyses is currently unknown and is the subject of an ongoing research study [8]. This paper provides an overview of the different levels of, and motivations for, replicating health economic models, and discusses the advantages, disadvantages and caveats associated with this type of modelling activity. In addition, the feasibility of comprehensive model replication is explored empirically across a small number of recent case studies. Recommendations are put forward for improving reporting standards to enhance health economic model replicability.

MOTIVATIONS, ADVANTAGES AND DISADVANTAGES OF MODEL REPLICATION

Purposes of model replication

There are several motivations for replicating published models and their associated analyses. For example, a model author may wish to evaluate a new intervention or assess an existing intervention against different comparators using a model structure which has been previously defined and agreed as being appropriate as part of an earlier decision-making process. Alternatively, a model author adopting one particular modelling methodology may wish to reproduce a published model using an alternative approach in order to compare and contrast the results. In other instances, model authors may seek to replicate an existing model structure and accompanying parameter inputs in order to revise certain structural assumptions that are deemed inappropriate for the current decision problem, to incorporate alternative and/or newer evidence, and/or to adapt the model parameters to reflect the same decision problem within a different geographical health care jurisdiction. Model replication may also be undertaken retrospectively for the purposes of cross-validation: for example, the structure and/or parameters of a new model may be manipulated to determine whether it can produce similar results to previously published analyses. Across all of these situations, there is a necessary underlying assumption that the conceptual basis of the published model structure is adequate to address the current decision problem, or at least that it does not prohibit any necessary adaptation required to fully address the current decision problem.

Levels of model replication

Broadly speaking, model replication can be thought of in terms of a spectrum of activity which is dependent both on the purpose of the replication exercise and the extent to which that replication is implemented. At one end of the spectrum, the model author may attempt to comprehensively replicate a published model in its entirety, with the intention of reproducing the exact results presented in a study publication or report (e.g. an incremental cost-effectiveness ratio [ICER] for a given economic comparison or set of comparisons). There are several examples of comprehensive model replication in the literature. For example, Woods and Rizzo replicated a previous model of antidepressant therapies in order to revise the key assumptions around success rates and treatment duration which influenced the ICER [9]. In another example, Smolen et al. replicated a published model of onabotulinumtoxinA for the treatment of chronic migraine, originally developed by Batty et al. [10], for the purposes of future adaptation and expansion [11].

At the other end of the spectrum, the model author may attempt to replicate only the published model structure, without fully mimicking the decision problem in which that structure was originally applied. In such instances, the goal is not to reproduce the published ICER, and the replicator cannot fully ensure that the original model has been accurately replicated without doing so. Rather, the intention is to "borrow" the published model structure to address a different decision problem to that for which it was

initially developed. There are numerous examples of "borrowing" existing model structures within the published literature as well as within independent health technology assessment reports and pharmaceutical company submissions to health care reimbursement agencies. For example, the general hybrid decision tree/Markov structure developed for the assessment of glycoprotein IIb/IIIA antagonists for the treatment of non-ST elevation acute coronary syndromes (ACS) developed by Palmer et al. [12] (later cited in Briggs et al. [1]) was subsequently used to inform appraisals of other products within the National Institute for Health and Care Excellence (NICE) Technology Appraisal Programme, including ticagrelor for the treatment of ACS [13] and bivalirudin for the treatment of ST-segment-elevation myocardial infarction [14]. Similarly, the School of Health and Related Research (ScHARR) Multiple Sclerosis (MS) Model [15], which was originally developed to inform NICE's appraisal of beta interferon and glatiramer acetate, despite various modifications and adaptations [16], is cited as the source of the general model structure used by the manufacturers of products in all subsequently completed NICE appraisals of disease-modifying therapies for MS [17-21]. A further example is the York Psoriasis Model: this model was initially developed within NICE Technology Appraisal 103 [22] and was explicitly adopted by the manufacturers in subsequent appraisals of infliximab [23], adalimumab [24], and ustekinemab [25].

Alternatively, the model author may attempt to replicate a published model for use in part of their de novo model structure. This involves replicating the published model structure and inputs for use as a "sub-model" within the broader structure of the new model (although in some instances it may be possible to simply use the original model outputs as input parameters to the new model without fully replicating its structure). For example, within their economic analysis of oxaliplatin and capecitabine for the adjuvant treatment of colon cancer, Eggington et al. [26] replicated a model of treatments for metastatic colorectal cancer [27] to inform downstream costs and health outcomes for relapsed patients (in this case, the original model was available hence it was unnecessary to fully rely on the reporting of the study publication). In another example, Kearns et al. [28] produced a model simulating joint costs and outcomes of service reconfiguration options for type 2 diabetes and depression which involved replicating the underlying structure and inputs of a published depression model [29] and the United Kingdom Prospective Diabetes Study (UKPDS) outcomes model (version 2) [30-32] based on the study publications.

The advantages and disadvantages of model replication

Irrespective of the extent to which it is applied, model replication may be associated with two clear advantages. Firstly, replicating an existing model may require considerably less effort, time and expense than developing a de novo model as a consequence of bypassing, or undertaking in only a perfunctory manner, certain stages of model development such as conceptual modelling and/or extensive literature searching for parameters. Secondly, the use of a previously accepted model structure and previously

identified evidence sources may increase a decision-maker's confidence in the analysis of the new technology. In this sense, it is not only the model structure and inputs which are being "borrowed", but also the credibility or perceived prestige of the original model arising from its publication in a peer reviewed journal or from its acceptance and use by decision-makers. Model replication may however also be associated with some important disadvantages. In particular, two key underlying assumptions associated with the use of replicated models should be noted: (i) the existing model structure is suitable to address the new decision problem, and; (ii) the existing model was suitable to address the decision problem for which it was originally intended. Importantly, neither of these assumptions necessarily hold true.

The first assumption may be unsound for older publications of models which fare poorly when judged against currently accepted views of best practice, for models which do not adhere to current economic reference cases, and for those which reflect outdated theories of disease natural history or outdated treatment pathways and/or evidence. For example, in the NICE technology appraisal of bosutinib for previously treated chronic myeloid leukaemia (CML) [33], the company used an approach to extrapolate overall survival from a surrogate relationship between major cytogenetic response and overall survival that was based on an analysis undertaken within a previous appraisal for another treatment for CML [34]. The Appraisal Committee concluded that there was considerable uncertainty whether a study that had assessed imatinib escalation for CML could plausibly apply to bosutinib in the third-line setting. A further example of questionable model replication can also be found in the NICE technology appraisal of multiple biologic treatments for rheumatoid arthritis (RA). In this appraisal [35], all six companies submitted models which, based on previous appraisals [36, 37], made the structural assumption that Health Assessment Questionnaire Disease Index (HAQ DI) increased linearly while on conventional treatments. Conversely, the Assessment Group modelled non-linear HAQ DI trajectories using a latent class approach based on a recent relevant study by Norton et al. [38]. Finally, in the NICE appraisal of olaparib for relapsed ovarian cancer, the company justified their semi-Markovian model structure through reference to the previous appraisal of bevacizumab for the first-line treatment of ovarian cancer, although the exact definitions of health states differed between the models. The ERG had serious concerns regarding the restrictive assumptions adopted within the "borrowed" structure and its inability to generate plausible overall survival predictions which reflected the empirical results of the trial used to inform the model's parameters (an issue which was most likely caused by the adopted structure). The Appraisal Committee noted that the company's structure was "unconventional" and expressed concerns regarding the plausibility of the extrapolation of overall survival data generated using the company's adopted structure.

The assumption that the replicated model was suitable to address the decision problem for which it was originally intended may also be invalid as neither the prior acceptance and use of a model by a decision-

maker nor the journal peer review process are perfect signals of model quality. Consequently, there is a risk that what is perceived to have been suitable before is not suitable now, and may never have been suitable in the first place. It should further be recognised that when adopting an existing structure it would be unwise to entirely bypass the development of a de novo conceptual model. On the contrary, the new conceptual model should be compared with models identified through literature searches and should be discussed with clinical and economic experts to ensure its appropriateness. Failure to do so may lead to the loss of an important point of model validation [39] and can impose restrictions on the available choices about how evidence can best be used to inform the current decision problem. Thus, whilst the intention of model replication may be to borrow credibility from an existing structure, its unthinking use may produce the converse result.

FEASIBILITY OF COMPREHENSIVE MODEL REPLICATION – A PILOT STUDY

If a model developer intends to replicate a model and its results, the advantages of such activity can only be realised if the published model is actually fully replicable. In order to assess the feasibility of comprehensive model replication, we undertook a small pilot study in which we attempted to fully reproduce five recent economic analyses published in Pharmacoeconomics between August and November 2016 [40-44]. For each study, the feasibility of replicating the analyses was explored through consideration of published information relating to model structure, assumptions and model parameter values (see Table 1) within the full study publication and any supplementary material available online, but excluding information reported in other papers or reports. As one of the included models was developed by one of the authors of this paper (PT), the replication of this model was undertaken independently by a different author (IB).

Across the five case studies, our model replication efforts were met with limited success: two models were fully replicated (Versteegh [44] and Tappenden et al. [41]), two models were unequivocally not replicable (Oddershede et al. [43] and Davies et al. [42]), and one model was potentially replicable although the replication process failed in this instance (Elvidge et al. [40]). In the two instances whereby the models were reproduced, there were discrepancies between the replicated model results and those reported within the study publications.

The model reported by Davies et al. was based on analyses of the published CORE diabetes model (McEwan et al. [45]) which has been previously used to inform a number of reimbursement decisions. This paper included a model structure diagram but only parameter values relating to the baseline characteristics of the model population and treatment effect parameters. Resource use and cost parameters were reported as supplementary material, however, the event probabilities and risk equations were not. Consequently, full model replication failed. The paper by Oddershede et al. [43] was principally an economic evaluation alongside a clinical trial (EEACT), but included long-term

modelling beyond the observed study period using generalised linear models (GLMs) [46]. These GLMs were calculated from individual patient data and no details were provided on their covariates or coefficients; as such, it was not possible to replicate the published results. Within the paper reported by Elvidge et al. [40]), the reporting of the model structure and parameter inputs was generally clear, however the model structure diagram did not exactly match the implemented model, a finding which was determined through subsequent personal communication with the authors. As a consequence of this ambiguity, the replication of the Elvidge et al model was abandoned. With respect to the model by Tappenden et al. [41], replication was possible, as the description of the model structure was clear and the list of parameters used in the model was fully reported. However, some reported event probabilities were subject to rounding errors which led to differences between the published results and those generated using the replicated model. Based on the information provided in the paper and the supplementary material, the analyses reported by Versteegh [44] were simple to reproduce, although minor additional assumptions around other-cause mortality and half-cycle correction were required. However, whilst the modelled incremental health gains were broadly similar between the replicated model and the original publication, the incremental costs remained consistently lower in the replicated model, thereby also affecting the accuracy of the ICERs. A second author (JY) independently attempted to replicate the Versteegh model; however, this further analysis produced similar discrepancies to the first replicated model.

Table 1: Feasibility of comprehensive model replication across five published case study models

[INSERT TABLE 1 HERE]

BARRIERS AND FACILITATORS

Barriers to comprehensive model replication

Several barriers may hinder or preclude comprehensive model replication. In particular, these include:

- Inadequate and/or insufficient reporting of the model and/or its inputs: This may relate to the exclusion of certain model parameters or a lack of correspondence between the implemented model and the reported model. Insufficient reporting of the model may lead to a situation whereby analysts must make assumptions about the model structure and/or its parameters; this can lead to discrepancies in the replicated model results.
- The use of ambiguous language: The description of the implemented model may be ambiguous and open to multiple interpretations; this may affect all forms of model replication and may lead to discrepancies in replicated model results.
- Confidentiality of model inputs: Confidentiality of input parameters represents a particular problem in some UK-analyses undertaken to inform the NICE Technology Appraisal Programme as manufacturers of a product may offer a confidential price discount as part of a

Patient Access Scheme (PAS). Unless the ICERs generated using the list prices of the products are available, it is unlikely that full model replication will be possible. In addition, the published model may have been parameterised using datasets that are not publicly available. Comprehensive replicability of the health economic model may also require full disclosure of all statistical models used to derive the health economic model parameters.

- Reporting limits: Journal publication limits may restrict full reporting of models. This is likely to be an issue for highly complex models and/or those which feature a large number of parameters. In addition, the implementation of complex model assumptions and algorithms, for example, patient-level simulations involving multiple competing events and updating of events, may be technically challenging to reproduce and may require additional assumptions which deviate from those applied in the original model.
- Errors in original publications: Unidentified programming errors in the original model may lead to difficulties in replicating published model results.

Alongside the barriers described above, other practical difficulties may also arise. For example, if a different modelling methodology is adopted for comparison purposes, e.g. converting a published Markov model to a discrete event simulation, parameter inputs may need to be re-estimated to reflect the data structures of the new methodology. This may require additional assumptions as well as new statistical analyses. In instances whereby the original structure is used but the data inputs are updated, some aspects of the structural arrangement and model assumptions may be revised to reflect the structure of the new data and the characteristics of the target population. In certain instances, the original structure of the published model may have been heavily influenced by the nature of the event risk data; if the model structure is to be replicated and subsequently adapted to reflect costs and outcomes for a different population, the existing structure may no longer be able to accommodate the new data. Finally, the complexity of some models might render their full replication based solely on a published paper unrealistic. Modellers interested in replicating such models would be best advised to approach the authors of the replicated model.

Suggestions for improving model replicability

A recent survey of health economic stakeholders suggested that across all stakeholders and countries, making health economic models available in an open format was considered to be beneficial [47]. The provision of open source models would circumvent problems around model replicability. This is however not universal practice; in the absence of such a policy, the following suggestions may serve to improve the replicability of published models.

- The stronger enforcement of transparent reporting standards such as the CHEERS checklist [4]) may increase the number of models which are fully replicable. Most importantly, publications

describing models should include a comprehensive description of the model structure and a full list of model assumptions and input parameters. The use of supplementary appendices may be helpful in this regard. Wherever possible, model authors should seek to lift confidentiality of all input parameters prior to publication. Ensuring model replicability is principally the concern of the model author, the main incentive for which is the increased credibility of the model through its transparent reporting.

- Model authors should ensure that the description of the model structure, assumptions and parameter sources presented in written documentation reflects the implemented model. Specifically, authors should ensure that diagrams and text are not ambiguous or open to multiple interpretations. For highly complex models, the use of separate diagrams for parts of the model may be helpful.
- Publishers could require submitting authors to include a statement in any written documentation confirming that the model and its results could, in principle, be reproduced by a competent analyst.
- Improving reporting of exact data sources and relevant analytic methods used to analyse those data may increase the transparency of parameter estimation. This includes not only citing studies which report the general specifications of the dataset, but also providing information about how the data have been used within the model.

CONCLUSIONS

Model replication can take several forms, ranging from comprehensive reproduction of the model structure and its results to the borrowing of a previously accepted model structure. Comprehensive replicability is generally perceived to be a desirable property of health economic models, however, the acceptability and appropriateness of using previously published models depends on the decision-making context; when this type of modelling activity is pursued, model authors should be mindful of any important aspects of the model development process which have been omitted, the relevance of the replicated model to the current decision problem, as well as any criticisms levied against the original model. Our pilot study indicates that from only a very small sample of economic modelling studies, comprehensive model replication is not always possible; further analyses are required to clarify the extent of this problem.

Data Availability Statement

The replicated models generated during the current study are available from the corresponding author on reasonable request.

Compliance with Ethical Standards

IB, PT and JY did not receive any funding to support the drafting of this manuscript and have no conflicts of interest to declare. PT attempted to replicate four of the five models included in the pilot study. IB replicated the remaining model. JY replicated one of the models. IB, PT and JY were involved in writing the manuscript.

References

1. Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. 1st ed. New York: Oxford University Press; 2006.
2. Popper K. The logic of scientific discovery. 1st ed. London: Routledge; 1992.
3. Weinstein M, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe M, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. Value in Health. 2003;**6**(1):9-17.
4. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D. Consolidated health economic evaluation reporting standards (CHEERS) - explanation and elaboration: A report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value in Health. 2013;16(2):231-50.
5. Eddy D. Technology assessment: the role of mathematical modeling Assessing medical technology. 1st ed. Washington DC: National Academy Press; 1985.
6. Akehurst R, Anderson P, Brazier J, Brennan A, Briggs A, Buxton M, et al. Decision analytic modelling in the economic evaluation of health technologies. Pharmacoeconomics. 2000;17(5):443-4.
7. Eddy D, Hollingworth W, Caro J, Tsevat J, McDonald K, Wong J, . Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. Medical Decision Marking. 2012;32(5):733-43.
8. McManus E. How reproducible are economic decision-analytic models? A systematic review and replication study - study outline. 2016; Available from: http://www.clahrcprojects.co.uk/impact/publications/how-reproducible-are-economic-decision-analytic-models-systematic-review-and
9. Woods S, Rizzo J. Cost-effectiveness of antidepressant treatment reassessed. British Journal of Psychiatry. 1997;170:257-63.
10. Batty A, Hansen R, Bloudek L, Varon S, Hayward E, Pennington B, et al. The cost-effectiveness of onabotulinumtoxinA for the prophylaxis of headache in adults with chronic migraine in the UK. Journal of Medical Economics. 2013;16(7):877-87.
11. Smolen L, Klein T, Kelton K. Replication of a published Markov chronic migraine cost-effectiveness analysis model for purposes of early phase adaptation and expansion. Value in Health. 2015(83):A19.
12. Palmer S, Sculpher M, Philips Z, Robinson M, Ginnelly L, Bakhai A, et al. Management of non-ST-elevation acute coronary syndromes: how cost-effective are glycoprotein IIb/IIIA antagonists in the UK National Health Service? International Journal of Cardiology. 2005;100(2):229-40.
13. AstraZeneca Ltd. Ticagrelor for the treatment of acute coronary syndromes - manufacturer's submission to the National Institute for Health and Care Excellence. 2011; Available from: https://www.nice.org.uk/guidance/TA236/documents/acute-coronary-syndromes-ticagrelor-astrazeneca4

14. The Medicines Company. Bivalirudin (Angiox®) for the treatment of ST-segment elevation myocardial infarction - manufacturer's submission to the National Institute for Health and Care Excellence. 2011; Available from: https://www.nice.org.uk/guidance/TA230/documents/myocardial-infarction-persistent-stsegment-elevation-bivalirudin-the-medicines-company2

15. Chilcott J, McCabe C, Tappenden P, O'Hagan A, Cooper N, Abrams K, et al. Modelling the cost effectiveness of interferon beta and glatiramer acetate in the management of multiple sclerosis. Commentary: evaluating disease modifying treatments in multiple sclerosis. British Medical Journal. 2003;326(7388):522.

16. Allen F, Montgomery S, Maruszczak M, Kusel J, Adlard N. Convergence yet Continued Complexity: A Systematic Review and Critique of Health Economic Models of Relapsing-Remitting Multiple Sclerosis in the United Kingdom. Value in Health. 2015;18(6):925-38.

17. Biogen Idec Ltd. Natalizumab (Tysabri®) for the treatment of adults with highly active relapsing remitting multiple sclerosis - manufacturer's submission to the National Institute for Health and Care Excellence. 2007; Available from: https://www.nice.org.uk/guidance/TA127/documents/multiple-sclerosis-natalizumab-manufacturer-submissions-biogen-idec-uk-and-elan-pharma-international-ltd-joint-development-agreement-confidential-information-removed2

18. Gani R, Giovannoni G, Bates D, Kemball B, Hughes S, Kerrigan J. Cost-effectiveness analyses of natalizumab (Tysabri) compared with other disease-modifying therapies for people with highly active relapsing-remitting multiple sclerosis in the UK. Pharmacoeconomics 2008;26(7):617-27.

19. Genzyme. Teriflunomide for the treatment of relapsing remitting multiple sclerosis in adults - manufacturer's submission to the National Institute for Health and Care Excellence. 2013; Available from: https://www.nice.org.uk/guidance/TA303/documents/multiple-sclerosis-relapsing-teriflunomide-evaluation-report4

20. Novartis Pharmaceuticals UK Ltd. Fingolimod for the treatment of relapsing remitting multiple sclerosis in adults - manufacturer's submission to the National Institute for Health and Care Excellence. 2011; Available from: https://www.nice.org.uk/guidance/TA254/documents/manufacturer-submission-from-novartis2

21. Biogen Idec Ltd. Dimethyl fumarate for the treatment of adult patients with relapsing remitting multiple sclerosis - manufacturer's submission to the National Institute for Health and Care Excellence. 2013; Available from: https://www.nice.org.uk/guidance/TA320/documents/multiple-sclerosis-relapsingremitting-dimethyl-fumarate-evaluation-report2

22. Woolacott N, Hawkins N, Mason A, Kainth A, Khadjesari Z, Bravo Vergel Y, et al. Efalizumab and etanercept for the treatment of psoriasis - assessment report submitted to the National Institute for Health and Care Excellence. 2007; Available from: https://www.nice.org.uk/guidance/TA103/documents/psoriasis-efalizumab-and-etanercept-assessment-report2

23. Schering-Plough Ltd. Remicade® in the treatment of psoriasis in England and Wales - manufacturer's submission to the National Institute for Health and Care Excellence. 2007; Available from: https://www.nice.org.uk/guidance/TA134/documents/psoriasis-infliximab-scheringplough-ltd2

24. Abbott Laboratories Ltd. Adalimumab (Humira®) for the treatment of moderate to severe plaque psoriasis - manufacturer's submission to the National Institute for Health and Care Excellence. 2007; Available from: https://www.nice.org.uk/guidance/TA146/documents/psoriasis-adalimumab-appraisal-consultation-manufacturer-submission-abbott-laboratories-ltd2

25. Janssen-Cilag Ltd. Ustekinumab (Stelara) for the treatment of moderate to severe plaque psoriasis in England & Wales - manufacturer's submission to the National Institute for Health and Care Excellence. 2009; Available from: https://www.nice.org.uk/guidance/TA180/documents/psoriasis-ustekinumab-manufacturer-submission-janssencilag2

26. Eggington S, Tappenden P, Pandor A, Paisley S, Saunders M, Seymour M, et al. The clinical and cost-effectiveness of oxaliplatin and capecitabine for the adjuvant treatment of colon cancer: systematic review and economic evaluation. British Journal of Cancer. 2006;95(9):1195-201.

27. Tappenden P, Jones R, Paisley S, Carroll C. The cost-effectiveness of bevacizumab in the first-line treatment of metastatic colorectal cancer in England and Wales. European Journal of Cancer. 2007;43(17):2487-94.

28. Kearns B, Rafia R, Leaviss J, Preston L, Brazier J, Palmer S, et al. The cost-effectiveness of changes to the care pathway used to identify depression and provide treatment amongst people with diabetes in England: a model-based economic evaluation. BMC Health Services Research. 2017;17(1):78.

29. Tosh J, Kearns B, Brennan A, Parry G, Ricketts T, Saxon D, et al. Innovation in health economic modelling of service improvements for longer-term depression: demonstration in a local health community. BMC Health Services Research. 2013;13(150).

30. Clarke P, Gray A, Legood R, Briggs A, Holman R. The impact of diabetes-related complications on healthcare costs: results from the United Kingdom Prospective Diabetes Study (UKPDS Study No. 65). Diabetic Medicine.20(6):442-50.

31. Alva M, Gray A, Mihaylova B, Clarke P. The effect of diabetes complications on health-related quality of life: the importance of longitudinal data to address patient heterogeneity. Health Economics. 2014;23(4):487-500.

32. Hayes A, Leal J, Gray A, Holman R, Clarke P, et al. UKPDS outcomes model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82. Diabetologia. 2013;56(9):1925-33.

33. National Institute for Health and Care Excellence. Bosutinib for previously treated chronic myeloid leukaemia. 2013; Available from: https://www.nice.org.uk/guidance/ta299

34. National Institute for Health and Care Excellence. Dasatinib, high-dose imatinib and nilotinib for the treatment of imatinib-resistant chronic myeloid leukaemia (CML) (part review of NICE technology appraisal guidance 70), and dasatinib and nilotinib for people with CML for whom treatment with imatinib has failed because of intolerance. 2012; Available from: https://www.nice.org.uk/guidance/ta299

35. Stevenson M AR, Tosh J, Simpson E, Everson-Hock E, Stevens J, et al. Adalimumab, etanercept, infliximab,certolizumab pegol, golimumab, tocilizumab and abatacept for the treatment of rheumatoid arthritis not previously treated with disease-modifying antirheumatic drugs and after the failure of conventional disease-modifying antirheumatic drugs only: systematic review and economic evaluation. Health Technol Assess. 2016;20(35).

36. National Institute for Health and Care Excellence. Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a TNF inhibitor. 2013; Available from: https://www.nice.org.uk/guidance/ta195

37. National Institute for Health and Care Excellence. Adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis. 2013; Available from: https://www.nice.org.uk/guidance/ta130

38. Norton S, Sacker A, Dixey J, Done J, Williams P, Young A. Trajectories of functional limitation in early rheumatoid arthritis and their association with mortality. Rheumatology. 2013;52(11):2016-24.

39. Tappenden P, Chilcott J. Avoiding and identifying errors and other threats to the credibility of health economic models. Pharmacoeconomics. 2014;32(10):967-79.

40. Elvidge J, Bullement A, Hatswell A. Cost effectiveness of characterised chondrocyte implantation for treatment of cartilage defects of the knee in the UK. Pharmacoeconomics. 2016;34(11):1145-59.

41. Tappenden P, Ren S, Archer R, Harvey R, Martyn St-James M, Basarir H, et al. A model-based economic evaluation of biologic and non-biologic options for the treatment of adults with moderately-to-severely active ulcerative colitis after the failure of conventional therapy. Pharmacoeconomics. 2016;34(10):1023-38.

42. Davies M, Glah D, Chubb B, Konidaris G, McEwan P. Cost effectiveness of IDegLira vs. alternative basal insulin intensification therapies in patients with type 2 diabetes mellitus uncontrolled on basal insulin in a UK setting. Pharmacoeconomics. 2016;34(9):953-66.

43. Oddershede L, Walker S, Stöhr W, Dunn D, Arenas-Pinto A, Paton NI, et al. Cost effectiveness of protease inhibitor monotherapy versus standard triple therapy in the long-term management of HIV patients: Analysis using evidence from the PIVOT trial. Pharmacoeconomics. 2016;34(8):795-804.

44. Versteegh M. Impact on the Incremental Cost-Effectiveness Ratio of Using Alternatives to EQ-5D in a Markov Model for Multiple Sclerosis. Pharmacoeconomics. 2016;34(11):1133-44.

45. McEwan P, Foos V, Palmer J, Lamotte M, Lloyd A, Grant D, et al. Validation of the IMS CORE diabetes model. Value in Health. 2014;17(6):714-24.

46. Barber J, Thompson S. Multiple regression of cost data: use of generalised linear models. Journal of Health Services Research & Policy. 2004;9(4):197-204.

47. Dunlop W, Mason N, Kenworthy J, Akehurst R. Benefits, challenges and potential strategies of open source health economic models. Pharmacoeconomics. 2017;35(1):125-8.

Table 1: Feasibility of comprehensive model replication across five published case study models

Authors	Title	Disease area	Model type	Model structure diagram reported	List of assumptions reported	Full parameter list reported	Was full replication successful?
Elvidge et al [40]	Cost effectiveness of characterised chondrocyte implantation for treatment of cartilage defects of the knee in the UK	Knee defects	Markov	Yes	Yes. Provided in text	Yes. Supplementary appendix provides survival model parameters and distributions.	No. The model structure diagram does not exactly match states in implemented model. This led to some difficulty in fully mapping the transition probability inputs to the state transition matrix.
Versteegh [44]	Impact on the incremental cost-effectiveness ratio of using alternatives to EQ-5D in a Markov model for multiple sclerosis	Multiple sclerosis	Markov	Yes	Yes. Provided in text and supplementary material	Yes, excluding life tables.	Yes, but with discrepant results.
Tappenden et al [41]	A model-based economic evaluation of biologic and non-biologic options for the treatment of adults with moderately-to-severely active ulcerative colitis after the failure of conventional therapy	Ulcerative colitis	Markov	Yes	Yes	Yes, excluding life tables.	Yes. Minor differences between published and replicated results, likely due to rounding errors.
Oddershede et al [43]	Cost effectiveness of protease inhibitor monotherapy versus standard triple therapy in the long-term management of HIV patients: Analysis using evidence from the PIVOT trial	HIV treatment	EEACT with long-term modelling	No	No	No. Restricted to resource use parameters.	No. GLM equations were not reported, hence survival and QALY estimates could not be replicated for either treatment group
Davies et al [42]	Cost effectiveness of IDegLira vs. alternative basal insulin intensification therapies in patients with type 2 diabetes mellitus	Type 2 diabetes mellitus	Patient-level simulation	Yes	No	No. Event probabilities not reported.	No.

	uncontrolled on basal insulin in a UK setting						

Abbreviations: EEACT, economic evaluation alongside a clinical trial; EQ-5D, EuroQol 5 dimensions questionnaire; HIV, human immunodeficiency virus; GLM: generalised linear model; QALY: quality-adjusted life year