R you still using Excel? The advantages of modern software tools for health technology assessment

Devin Incerti[1]

Howard Thom[2]

Gianluca Baio[3]

Jeroen P. Jansen[1,4]

[1]Innovation and Value Initiative, Los Angeles CA, USA

[2]Department of Population Health Sciences, Bristol Medical School, University of Bristol, Canynge Hall, BS8 2PS, UK

[3]Department of Statistical Science, University College London, London WC1E 6BT, United Kingdom

[4]Stanford University School of Medicine, Department of Health Research & Policy – Epidemiology, Stanford CA, USA

Corresponding author: Devin Incerti, PhD, Innovation and Value Initiative, 11100 Santa Monica Boulevard, Suite 500, Los Angeles, CA 90025, devin.incerti@thevalueinitiative.org

What is relevant health technology assessment?

Economic models are used in health technology assessments (HTAs) to evaluate the cost-effectiveness of competing medical technologies and inform the efficient use of health care resources. Historically, these models have been developed with specialized commercial software (such as TreeAge) or more commonly with spreadsheet software (almost always Microsoft Excel). Although these tools may be sufficient for relatively simple analyses, methodological and computational advances now allow for models that are increasingly sophisticated and realistic but cannot be reasonably computed in Excel. Related fields such as statistics, economics, and machine learning utilize many of these advances and consequently make heavy use of modern programming languages such as R, Python, MATLAB, and Julia.[1-4] Yet, in the context of HTA, there seems to be an ongoing preference among manufactures preparing submissions and reviewers for economic models in Excel.

The choice of software is not just an academic exercise, but puts unnecessary constraints on the analysis that may ultimately limit its credibility and relevance for HTAs. In this article, we consider four criteria that economic models should strive to meet and argue that many of these are unobtainable without the use of modern software:

1. **Clinical realism**: A model should reflect the state of evidence, the current understanding of the disease, and be accepted by clinical experts.

2. **Quantifying decision uncertainty***:* A model should be capable of quantifying decision uncertainty and informing prioritization of future research.

3. **Transparency and reproducibility:** Resources should exist so that a model can be completely understood, reproduced, and pressure tested.

4. **Reusability and adaptability***:* It should be possible to easily update a model to reflect new clinical evidence or adapt it for a new market, indication, or intervention.

Why Excel is insufficient and modern software can help

Clinical realism

The most important criteria for any economic model are arguably the extent to which they do justice to the available clinical evidence and reflect the clinical and biological understanding of the disease of interest. However, the limitations of Excel often produce artificial separation between parameter estimation based on the clinical evidence and model simulation. As a result, economic models often attempt to fit "square pegs into round holes" because the model structure used for the cost-effectiveness analysis is frequently not aligned with the statistical model used for parameter estimation. As shown in **Figure 1**, the statistical and economic models should be completely integrated.

Consider economic models for the evaluation of oncology drugs, which cover over 40% of submissions to the National Institute for Health and Care Excellence (NICE) in the UK. A standard model structure is based on 3 health states (pre-progression, progression, and death) with time in each health state typically simulated using a state transition model (STM) such as a Markov model or a partitioned survival model (PSM).[5-7] Once the structure of the economic model is determined, the first step in a cost-effectiveness analysis is to estimate the parameters of the chosen model. As outlined by the NICE decision support unit (DSU), the health state probabilities in the STM should ideally be estimated using a multi-state model, and the health state probabilities in a PSM require estimation of the progression free survival (PFS) and overall survival (OS) curves using flexible survival models (e.g., parametric models, splines, fractional polynomials) to fit the data. STMs may be more appropriate than PSMs if it is necessary to extrapolate outcomes beyond available follow-up time.[5] In some cases, the relevant clinical evidence base is not limited to a single trial, but consists of many studies and a formal evidence synthesis, such as a network-meta analysis (NMA), needs to be performed.[8] Using a STM or PSM framework in the context of an evidence synthesis based on published summary data is arguably less straightforward; nonetheless, the evidence synthesis model should follow the structure of the economic model as closely as possible.

The second step in a cost-effectiveness analysis is to use the parameter estimates from step one to simulate disease progression and compute relevant outcomes such as costs and quality-adjusted life-years (QALYs) given the model structure of choice. Although spreadsheet software

can be used to simulate outcomes with PSMs and STMs in simple cases, it quickly becomes cumbersome and oftentimes infeasible if there is a need to simulate outcomes capturing patient heterogeneity, to model multi-line sequential treatment strategies, to quantify decision uncertainty (see next section), and/or to simulate disease progression from a fitted multi-state statistical model. For instance, in the multi-state case, if transition probabilities depend on time since treatment initiation, then outcomes are simulated based on a *time-inhomogeneous Markov model* structure using a complex estimator known as the Aalen-Johansen estimator.[9,10] On the other hand, if (more plausibly) transition probabilities depend on time since entering an intermediate state (e.g., the progressed state), then it would be necessary to simulate outcomes based on a *semi-Markov model* structure using an individual-level simulation.[11]

R and its extensive collection of packages allows users to perform both parameter estimation (including evidence synthesis) and the subsequent simulation of disease progression and outcomes in one software environment, so that the economic model does not need to be unnecessarily simplified due to the constraints of Excel.[12] For instance, the *flexsurv* package can fit parametric models and splines; *survHE* can fit both Bayesian and maximum likelihood survival models; *mstate, flexsurv, and msm* packages can estimate multi-state models; and *metafor*, *netmeta* and *mada* can be used for meta-analysis. Furthermore, R packages such as *R2OpenBUGS*, *rjags*, and *rstan* allow Bayesian evidence synthesis to estimate model parameters based on multiple source of evidence. Once parameters have been estimated, the analyst could use R packages built to simulate specific economic models, such as the R package *hesim*, that simulates PSMs and STMs from fitted statistical models and computes costs and QALYs for cost-effectiveness analysis.

Using the same software for evidence synthesis and simulation modeling is especially beneficial in a Bayesian framework. Posterior simulation techniques such as Markov Chain Monte Carlo (MCMC) ensure that parameter uncertainties and correlations are fully reflected in the simulation of model outcomes without the need for additional parametric assumptions to represent parameter uncertainty in the economic model.[17] In oncology, when the standard three state model structure is used, the simulated posterior distribution of the fitted statistical model can be used to directly simulate disease progression.[18] A close relationship between the statistical model for

parameter estimation and the economic model structure has been demonstrated in several disease areas outside oncology as well.[19-21]

In complex economic models, a subset of parameters often cannot be estimated using standard statistical techniques. The unknown parameters can, however, be estimated using model calibration, a technique that identifies parameter values that maximize the fit between model output and observed data. R's package system is again advantageous as it contains a large suite of packages that allow for efficient model calibration. Numerical optimization can be implemented using a number of general-purpose functions in the *stats* (i.e., *optim)* and *optimx* packages, with genetic algorithms using *genalg*, or with differential evolution using *DEoptim*. The *lhs* R package provides methods for Latin hypercube sampling (LHS), which is used for model calibration by randomly sampling sets of parameters from a multidimensional distribution and identifying the best fitting parameter set. It is also possible to estimate the full distribution of the parameters using Bayesian calibration with the large number of R packages for Bayesian inference (http:://cran.r-project.org/web/views/Bayesian.html).[13] A fully calibrated economic model can be readily used to compute model outcomes as the statistical analysis and model structure are perfectly integrated. This is especially common in infectious disease modeling where compartmental models are used to model the spread of disease.[14-16]

Quantifying decision uncertainty

Simulating an economic model becomes even more complex if the model inputs are treated as uncertain, as has been requested by HTA bodies such as NICE.[22] Probabilistic sensitivity analysis (PSA) is typically used to quantify decision uncertainty, which requires randomly sampling the model parameters from their joint probability distribution, and simulating the model for each random sample. Although a PSA can be performed in Excel using Visual Basic for Applications (VBA), other programming languages offer some advantages. For example, R has validated functions for sampling from univariate (e.g., uniform, beta, normal, lognormal, gamma) and multivariate (e.g., multivariate normal, Dirichlet) distributions commonly used for PSA.[23] Moreover, R packages such as *BCEA* and *hesim* have functions to produce a number of standard representations of decision uncertainty such as cost-effectiveness planes, cost-effectiveness acceptability curves (CEACs), and the cost-effectiveness acceptability frontier (CEAF).[24-26]

PSA and uncertainty quantification can aid research prioritization through value of information analysis. The value of removing all uncertainty around all parameters is the expected value of information (EVPI), the value for a subset of parameters is the expected value of perfect partial information (EVPPI), and the expected value of sample information (EVSI) is the value of reducing only some uncertainty in the parameters through a specific research design. However, EVPPI and EVSI require nested Monte Carlo simulation for accurate and precise estimation but this is infeasible in Excel for any but the simplest of models. Conversely, there are efficient regression and approximation methods implemented in the R packages *BCEA* and *EVSI* and the R Shiny web applications SAVI (http://savi.shef.ac.uk/SAVI/) and BCEAweb (https://egon.stats.ucl.ac.uk/projects/BCEAweb/) [27-29] Although PSA results can be generated from a model using any programming language including VBA, R is advantageous because it facilitates an integrated approach where both the model and EVPPI estimation are implemented together.

Implementation and analyses of PSAs in languages such as R can be made more efficient by using efficient programming techniques such as vectorization, linking to compiled languages (e.g., Fortran, C/C++), or through parallel computing. For example, Krijkamp et al. show that a vectorized microsimulation model in R reduced run time by 97%.[30] In cases where vectorization is not feasible, the R package Rcpp can be used to link C++ code to R as is done in both the IVI-RA individual patient simulation and the *hesim* package. The performance implications are significant with McEwan et al. reporting that a PSA in C++ completed in 3.8 hours while the same analysis in Excel's VBA took 14.5 days, a 246-fold improvement.[31] Significant speed improvements can also be achieved through the R package *parallel*, where multiple tasks are carried out simultaneously, as they are for PSA. Parallel computation is essential for the nested Monte Carlo estimation of EVPPI and EVSI for complex models where regression or approximation techniques are infeasible.

Transparency and reproducibility

An important advantage of script based programming languages is that analyses can be performed using reproducible scripts.[32] While some may view this as an added complexity, we believe that it improves transparency considerably by making the entire analysis completely reproducible including:

1. The statistical models used to estimate model parameters.
2. The simulation of model outcomes using the economic model.
3. Analysis of model output including computation of quantities of interest such as incremental cost-effectiveness ratios and summaries of the PSA.

To further increase transparency, such analyses can be embedded within reproducible documents. For instance, the R packages *knitr* and *Sweave* can be used to create PDF, html or docx documents in which each figure, table, and number cited in the text is based on code run in a script. A useful application is the creation of dynamic reports such as the documentation for the IVI-RA model (https://innovationvalueinitiative.github.io/IVI-RA/model-description/model-description.pdf) and the SAVI and BCEAweb web-applications. Reproducible notebooks (e.g., R notebooks, Juypter notebooks) with code embedded within text are a second use case. We have provided a complete example of a notebook that encompasses steps 1-3 above at https://innovationvalueinitiative.github.io/modern-software-HTA/ in which a semi-Markov model is used to evaluate the cost-effectiveness of a heath technology intervention.

While these replicable scripts can help, we believe that the future of cost-effectiveness modeling lies in web apps, in which graphical interfaces are used to run script-based models. An advantage is that decision makers can tailor analyses to their local population by modifying the characteristics of the target population or using parameters based on data relevant to the local setting. Users can also easily pressure test the sensitivity of results to baseline assumptions quickly through the point and click interface. These web apps are becoming more common and examples include complex analyses of PSAs such as SAVI and BCEAweb to interfaces for bespoke models (https://innovationandvalueinitiative.shinyapps.io/ivi-ra-expert/).

Reusability and adaptability

Reproducible scripts typically involve running a series of operations, most of which are based on functions written in specialist libraries such as R's large collection of open-source statistical packages. The open-source nature improves collaboration as model developers can make improvements to existing code or add new modules. In addition, the package system enhances reusability, which ultimately improves consistency of decision-making and reduces effort---giving more time for careful analysis---for model adaptations to other countries, indications, or

interventions. For example, EVPPI can be estimated for any economic model using the *BCEA* R package so that model developers do not need to write their own function.

One potential drawback is that a programming error (i.e., a bug) in a package's function could be propagated to every analysis conducted using that function. Although a concern, this can also happen with Excel if the same spreadsheet template is used for multiple models such as an adaption to a new country or indication. In addition, we believe that packages actually reduce bugs. Packages are carefully developed with authors typically spending far more time testing than a developer of a single model would and the open-source nature of packages allowing users to inspect and correct errors in the code. Software best practices like unit tests [33,34] (where each individual function is tested to ensure it obtains the correct result) and continuous integration [35,36] (where the packages is installed on an external server and all tests are re-run) enhance quality as well. These formal testing processes also enhance code adaptability as developers can make modifications or add new features without worrying that they have created unintended errors in existing code.

A path forward for health technology assessment

Programming languages like R, Python, Matlab, and Julia have active user and developer communities and are becoming increasingly well suited for the development of economic models for HTA. Nonetheless, there is a general lack of experience in the HTA community with this software. It is perhaps partly for this reason that HTA bodies have tended to favor economic models developed using Excel. For instance, while the National Institute for Health and Care Excellence (NICE) accepts models written in R, permission must be granted to submit models using other software. [37] Likewise, use of specialized software including R may be accepted by other HTA bodies, but typically requires pre-approval.[38-40] To our knowledge, no HTA bodies explicitly state that they will accept models developed in Python, Matlab, or Julia. So while R seems to be an acceptable option among many HTA bodies, ambiguity over the acceptance of programming languages by HTA bodies is still likely a significant barrier to adoption.

One reason for this lack of experience is that there is currently insufficient training and guidance on how to efficiently implement some of the more common types of economic models. It is therefore critical to train the next generation of decision scientists and health economists in state-

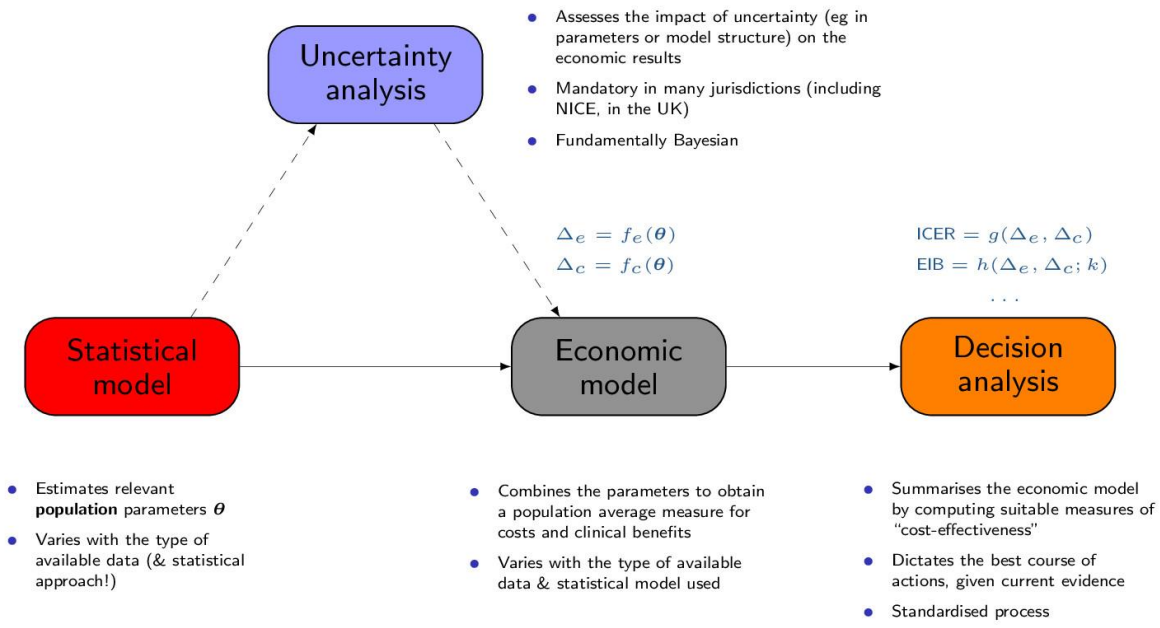
of-the methods and the software required to implement those methods. The HTA community can make steps toward these ends by:

- Developing university courses, workshops such as those by the Decision Analysis in R for Health Technologies in Health (DARTH) team (http://darthworkgroup.com/), and public webinars such as those provided by the ISPOR Student Network.
- Writing tutorial papers such as those by Williams et al. [41] in multi-state modeling and Krijkamp et al. [30] for developing microsimulation models.
- Making code freely available on repositories like those created on GitHub by DARTH and the Innovation and Value Initiative and thoroughly documenting that code.
- Encouraging pharmaceutical companies and decision-makers such as NICE to favor script-based programming languages when developing economic models.

Communication between decision-makers and modelers must of course be a two-way street. Modelers need to make the benefits of programming languages and web apps and the adverse consequences of not using them clear. On the other hand, decision-makers need to clarify their needs, which can, in turn, help modelers build appropriate software tools. Ultimately, both decision-makers and modelers should strive to develop models that are sufficiently accurate for decision-making in an HTA context. While Excel may be adequate in some circumstances, others may require software that facilitates methodologies beyond the capabilities of Excel. However, even if Excel-based models can produce results that are accurate enough, they still limit transparency, reproducibility, modifiability, computational efficiency, and use of modern software testing techniques.

We have focused on the advantages of R because it has the most existing code applicable to health economic modeling. But no single software tool is a panacea, and in our view, guidelines should therefore not dictate the use of certain software over other, equally valid, alternatives. Instead, model developers should be encouraged to improve upon existing tools and to adopt new tools when deemed beneficial, which can help foster an environment that encourages innovation. Such an environment will help ensure that coverage and pricing decisions confer greatest possible benefit and capture all scientific uncertainty, thus enabling correct prioritization of future research.

Figure 1. Integration of statistical and economic models for cost-effectiveness analysis



1. R Core Team. R: A language and environment for statistical computing. In: R Foundation for Statistical Computing, Vienna, Austria; 2014:http://www.R-project.org/.
2. Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: A fresh approach to numerical computing. *SIAM review.* 2017;59(1):65-98.
3. MATLAB User's Guide. The mathworks. *Inc, Natick, MA.* 1998;5:333.
4. Van Rossum G, Drake Jr FL. *Python tutorial.* Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands; 1995.
5. Woods B, Sideris E, Palmer S, Latimer N, Soares M. *NICE DSU Technical Support Document 19: Partitioned Survival Analysis for Decision Modelling in Healthcare: A Critical Review.* Decision Support Unit, ScHARR, University of Sheffield;2017.

6. Glasziou PP, Simes RJ, Gelber RD. Quality adjusted survival analysis. *Statistics in medicine.* 1990;9(11):1259-1276.
7. Siebert U, Alagoz O, Bayoumi AM, et al. State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force–3. *Medical Decision Making.* 2012;32(5):690-700.
8. Dias S, Ades A, Welton NJ, Jansen JP, Sutton AJ. Network meta-analysis for decision-making. In: John Wiley & Sons; 2018.
9. Aalen OO, Johansen S. An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics.* 1978:141-150.
10. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine.* 2007;26(11):2389-2430.
11. Fiocco M, Putter H, van Houwelingen HC. Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models. *Statistics in Medicine.* 2008;27(21):4340-4358.
12. Jalal H, Pechlivanoglou P, Krijkamp E, Alarid-Escudero F, Enns E, Hunink MGM. An Overview of R in Health Decision Sciences. *Medical Decision Making.* 2017;37(7):735-746.
13. Menzies NA, Soeteman DI, Pandya A, Kim JJ. Bayesian Methods for Calibrating Health Policy Models: A Tutorial. *Pharmacoeconomics.* 2017;35(6):613-624.
14. Drabo EF, Hay JW, Vardavas R, Wagner ZR, Sood N. A Cost-effectiveness Analysis of Preexposure Prophylaxis for the Prevention of HIV Among Los Angeles County Men Who Have Sex With Men. *Clin Infect Dis.* 2016;63(11):1495-1504.
15. Eaton JW, Johnson LF, Salomon JA, et al. HIV treatment as prevention: systematic comparison of mathematical models of the potential impact of antiretroviral therapy on HIV incidence in South Africa. *PLoS medicine.* 2012;9(7):e1001245.
16. Chowell G, Nishiura H. Transmission dynamics and control of Ebola virus disease (EVD): a review. *BMC Med.* 2014;12:196.
17. Baio G. *Bayesian methods in health economics.* Chapman and Hall/CRC; 2012.
18. Dias S, Ades A, Welton NJ, Jansen JP, Sutton AJ. Network Meta-Analysis of Survival Outcomes. In: *Network meta-analysis for decision-making.* John Wiley & Sons; 2018:288-312.
19. López-López JA, Sterne JA, Thom HH, et al. Oral anticoagulants for prevention of stroke in atrial fibrillation: systematic review, network meta-analysis, and cost effectiveness analysis. *bmj.* 2017;359:j5058.

20. Dias S, Welton NJ, Sutton AJ, Ades A. *NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials.* Decision Support Unit, ScHARR, University of Sheffield;2011.
21. Welton NJ, McAleenan A, Thom HH, et al. Screening strategies for atrial fibrillation: a systematic review and cost-effectiveness analysis. 2017.
22. Claxton K, Sculpher M, McCabe C, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health economics.* 2005;14(4):339-347.
23. Jalal H, Pechlivanoglou P, Krijkamp E, Alarid-Escudero F, Enns E, Hunink MGM. An Overview of R in Health Decision Sciences. *Med Decis Making.* 2017;37(7):735-746.
24. Black WC. The CE plane: a graphic representation of cost-effectiveness. *Medicial Decision Making.* 1990;10(3):212-214.
25. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health economics.* 2001;10(8):779-787.
26. Barton GR, Briggs AH, Fenwick EA. Optimal cost-effectiveness decisions: the role of the cost-effectiveness acceptability curve (CEAC), the cost-effectiveness acceptability frontier (CEAF), and the expected value of perfection information (EVPI). *Value Health.* 2008;11(5):886-897.
27. Strong M, Oakley JE, Brennan A, Breeze P. Estimating the expected value of sample information using the probabilistic sensitivity analysis sample: a fast, nonparametric regression-based method. *Medical Decision Making.* 2015;35(5):570-583.
28. Heath A, Manolopoulou I, Baio G. Estimating the expected value of partial perfect information in health economic evaluations using integrated nested Laplace approximation. *Statistics in medicine.* 2016;35(23):4264-4280.
29. Heath A, Manolopoulou I, Baio G. Efficient Monte Carlo Estimation of the Expected Value of Sample Information Using Moment Matching. *Medical Decision Making.* 2018;38(2):163-173.
30. Krijkamp EM, Alarid-Escudero F, Enns EA, Jalal HJ, Hunink MM, Pechlivanoglou P. Microsimulation Modeling for Health Decision Sciences Using R: A Tutorial. *Medical Decision Making.* 2018;38(3):400-422.
31. McEwan P, Bergenheim K, Yuan Y, Tetlow AP, Gordon JP. Assessing the relationship between computational speed and precision. *Pharmacoeconomics.* 2010;28(8):665-674.
32. Hollman C, Paulden M, Pechlivanoglou P, McCabe CJP. A comparison of four software programs for implementing decision analytic cost-effectiveness models. 2017;35(8):817-830.

33. Wickham H. testthat: Get started with testing. *The R Journal.* 2011;3(1):5-10.
34. Hamill P. *Unit test frameworks: tools for high-quality software development.* O'Reilly Media, Inc.; 2004.
35. Duvall PM, Matyas S, Glover A. *Continuous integration: improving software quality and reducing risk.* Pearson Education; 2007.
36. Travis CI. https://travis-ci.org/. Accessed August 13, 2018.
37. National Institute for Health and Care Excellence (NICE). *Guide to the processes of technology appraisal.* 2014.
38. Australian Government Department of Health. *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee.* 2016.
39. Canadian Agency for Drugs and Technologies in Health. *Submission Guidelines for the CADTH Common Drug Review.* 2014.
40. Norwegian Medicines Agency. Guidelines for the submission of documentation for single technology assessment of pharmaceuticals. In:2018.
41. Williams C, Lewsey JD, Briggs AH, Mackay DF. Cost-effectiveness analysis in R using a multi-state modeling survival analysis framework: a tutorial. *Medical Decision Making.* 2017;37(4):340-352.

Funding acknowledgement

HT received financial support from the UK Medical Research Council Hubs for Trials Methodology Research (HTMR) Collaboration and innovation in Difficult and Complex randomised controlled Trials In Invasive procedures (ConDuCT-II) to complete this work.