



Methodological Note: Reporting Deterministic versus Probabilistic Results of Markov, Partitioned Survival and Other Non-Linear Models

Edward C. F. Wilson^{1,2}

Accepted: 26 May 2021 / Published online: 14 July 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract

When making decisions under uncertainty, it is reasonable to choose the path that leads to the highest expected net benefit. Therefore, to inform decision making, decision-model-based health economic evaluations should always present expected outputs (i.e. the mean costs and outcomes associated with each course of action). In non-linear models such as Markov models, a single ‘run’ of the model with each input at its mean (a deterministic analysis) will not generate the expected value of the outputs. In a worst-case scenario, presenting deterministic analyses as the base case can lead to misleading recommendations. Therefore, the base-case analysis of a non-linear model should always be the means from a probabilistic analysis. In this paper, I explain why this is the case and provide recommendations for reporting economic evaluations based on Markov models, noting that the same principle applies to other non-linear structures such as partitioned survival models and individual sampling models. I also provide recommendations for conducting one-way sensitivity analyses of such models. Code illustrating the examples is provided in both Microsoft Excel and R, along with a video abstract and user guides in the electronic supplementary material.

Key Points for Decision Makers

Markov models are non-linear in the outputs (costs, quality-adjusted life-years [QALYs], etc.).

Deterministic analysis of non-linear models with inputs at their means will NOT generate the mean costs and QALYs.

Probabilistic analysis must therefore be used to generate the means.

These should be reported as the base-case analysis.

1 Introduction

Sackett et al. [1] defined evidence-based medicine as the “conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. [This] means integrating individual clinical expertise with the best available external clinical evidence from systematic research.”

Decision models could be defined as a structured synthesis of that “best available external clinical evidence” combined with (“best available”) information on resource use and cost. Such models then estimate the (incremental) cost-effectiveness of treatments, guiding population-level decision making as to whether they represent good value for money for routine use within a health system.

The evidence base is always uncertain, represented by the standard error and/or confidence/credibility interval around model inputs such as the treatment effect, health state utilities, and incidence of side effects. This ‘parameter uncertainty’ gives rise to ‘decision uncertainty’, that is, uncertainty as to whether the incremental cost-effectiveness ratio (ICER) is above or below the threshold, or—equivalently—which treatment is associated with the maximum net benefit [2]. Given this uncertainty, it is reasonable to base decisions on which course of action leads to the best

✉ Edward C. F. Wilson
ed.wilson@phmr.com

¹ Health Economics Group, Norwich Medical School, University of East Anglia, Norwich NR2 3PB, UK

² PHMR Ltd, Berkeley Works, Berkley Grove, London NW1 8XY, UK

outcome on average or, more explicitly, which course of action has the highest expected value [3], that is, has the highest mean net benefit.

Decision modelling fits most snugly within the framework of statistical decision theory [4, 5], to which the concept of basing decisions on expected values (the mean) alone is fundamental. A risk-neutral decision maker [6] presented with the ‘best available’ evidence and interested in maximising expected net benefit should, somewhat tautologically, make decisions yielding the maximum expected net benefit, irrespective of decision uncertainty or the results of statistical hypothesis tests [7]. It follows then, that decision analyses (economic evaluations) should seek to report the expected costs and consequences from different courses of action and base adoption recommendations on those alone. That is not to say that uncertainty is irrelevant. On the contrary, it should be used to assess whether there is an economic case for investment in further research to reduce that uncertainty via value of information analysis [7–9].

A common structure for a decision model is a Markov model, where stages of disease are broken down into several discrete health states and patients transition between them in each period with given probabilities, accruing costs and outcomes through time. However, unlike a simple decision tree, this generates a non-linear relationship between model inputs and the outputs. Thus, a single iteration of the model with each input at its mean (a deterministic analysis) will not generate an estimate of the expected (i.e. mean) costs and consequences [10]. This is also true for any non-linear model, such as a partitioned survival model (which is often implemented as a Markov model).

Deterministic analyses should therefore never be reported as the base- or reference-case analysis. Instead, the expected values from a probabilistic analysis (Monte Carlo simulation) should be reported as the primary estimate. In this paper, I illustrate the problem with a conceptual example and an applied Markov model then consider how to handle the specific issue of one-way sensitivity analyses (OWSA). I finish with a discussion and recommendations. Note that I use the terms ‘expected value’ and (arithmetic) ‘mean’ synonymously as they are mathematically identical concepts. Code for the applied example is provided in Microsoft Excel (Microsoft Corp. 2021; Redmond, WA, USA) and R [11] formats, along with video user guides and a video abstract summarising this paper (see the electronic supplementary material [ESM]). The package used to run the Markov model (rrapidMarkov) is available from GitHub [12], and the example also requires the gtools package [13]. A basic familiarity with either Microsoft Excel VBA or R is assumed.

Table 1 $E(x)^2 \neq E(x^2)$

	x	$y (=x^2)$
	1	1
	2	4
	3	9
	4	16
	5	25
Expected value	$E(x)=3$	$E(y)=E(x^2)=11$

$E(x)^2$ is 9, which is below the expected value of y , which is 11

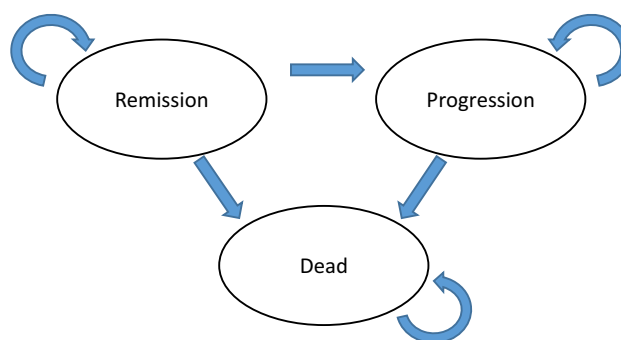


Fig. 1 Markov model structure

2 The Problem with Non-linear Models

As described in the introduction, statistical decision theory [4, 5] makes a strong case for considering the mean as the summary statistic of interest. In non-linear models, calculating results with all the inputs at their means will not yield the mean outputs. A simple example is illustrated in Table 1, where the ‘model’ is $y=x^2$. The input is x , and the output is y . The relationship between the two is simply the square. We are interested in the expected value of the output, y . Suppose x can take any integer value between 1 and 5 with equal probability. The expected value of x is therefore $(1 + 2 + 3 + 4 + 5)/5 = 3$. Evaluating the model at the mean of x therefore yields $3^2 = 9$. However, this is not equal to the mean of y , which is $(1^2 + 2^2 + 3^2 + 4^2 + 5^2)/5 = 11$. In other words, $E(x)^2 \neq E(x^2)$.

3 Applied Example: Markov Model

The logic described above applies to any non-linear model such as a Markov model. Suppose a model for disease X was divided into three stages: remission, progression, and dead (Fig. 1) with an annual transition period. For

illustrative purposes, the model is stationary (that is, transition probabilities do not vary with time).

Every year on current treatment (which I term ‘Old’), a patient has a mean 10% probability of transitioning to the progressive state, 10% probability of death, and thus 80% probability of remaining in the remission state. Uncertainty is represented as a Dirichlet (80,10,10) distribution, where the dimensions of the distribution represent probability of remission, progression, and death, respectively (Table 2). Once in the progressive state, a patient has a mean annual probability of death of 30%, and they cannot transition back

to the remission state. I assume there is less knowledge as to the transitions from this state and so it is modelled as a Dirichlet (0,7,3). (Given there is zero probability of remission, this could equally be modelled as a beta (3,7), with the probability of remission hard coded as 0%). Costs and health state utilities are assumed constant (and therefore known with certainty) for purposes of illustration.

A treatment, ‘New’ is available, which is taken as an adjunct to current care whilst the patient is in remission and costs £950 per annum (thus the per-cycle cost of remission with New is £1450). Its benefit is to slow the transition to progression (relative risk 0.7). This is modelled as a log Normal distribution, LN(ln(0.69875), 0.05). Once disease progresses, the patient ceases taking New.

The model is run for 10 cycles (years) with future costs and quality-adjusted life-years (QALYs) discounted at 3.5% per annum. For ease of illustration, we assume the only uncertainties in the model are the transition probabilities and relative risk of progression and that costs and health state utilities are known with perfect precision.

Table 3 shows a comparison between a deterministic analysis with each input at its mean and probabilistic results from 10,000, 100,000, and 1,000,000 Monte Carlo simulations. The deterministic point estimate ICER is £19,807, whilst the probabilistic analysis yields an ICER of £20,970 (from 1 million simulations). Following the logic described in the introduction, the ‘correct’ result is £20,970 per QALY gained. Of note is that the probabilistic ICER is higher than the deterministic. To emphasise this point, inputs in this example were chosen deliberately to generate an ICER either side of the UK National Institute for Health and Care Excellence (NICE) £20,000/QALY threshold [14]. In the experience of the author, deterministic analyses of Markov models tend to yield more favourable (lower) ICERs than probabilistic, but

Table 2 Model inputs

Inputs	Mean	Distribution
Probabilities		
Remission >> remission	0.8	Dirichlet (80, 10, 10)
Remission >> progression	0.1	
Remission >> dead	0.1	
Progression >> remission	0	Dirichlet (0, 7, 3)
Progression >> progression	0.7	
Progression >> dead	0.3	
RR progression with New	0.7	LogNormal (ln(0.7–0.5×0.05 ²), 0.05)
Health state utilities		
Alive	0.8	Constant
Progressive	0.5	Constant
Dead	0	Constant
Cost		
Remission	£500	Constant
Progression	£1000	Constant
Dead	£0	Constant
Drug cost, New	£950	Constant
Discount rate	3.5%	

Table 3 Deterministic and probabilistic results

	Cost (£)			QALYs			ICER (£)
	Old	New	Increment	Old	New	Increment	
Deterministic	3075.82	7304.73	4228.91	3.776	3.989	0.214	19,806.84
Probabilistic # simulations							
10,000	3161.74 (2406.25– 4118.36)	7420.90 (5914.41– 8956.51)	4259.15 (3508.16– 4838.14)	3.848 (3.050– 4.694)	4.051 (3.228– 4.898)	0.203 (0.178– 0.204)	20,976.75
100,000	3161.93 (2405.60– 4116.16)	7419.49 (5926.09– 8955.29)	4257.57 (3520.49– 4839.13)	3.847 (3.058– 4.686)	4.050 (3.235– 4.890)	0.203 (0.177– 0.204)	20,979.96
1,000,000	3160.21 (2404.82– 4111.91)	7416.95 (5922.80– 8943.42)	4256.75 (3517.98– 4831.51)	3.846 (3.058, 4.679)	4.049 (3.233– 4.884)	0.203 (0.175– 0.205)	20,969.89

Probabilistic results show mean (95% credibility limits). Costs and QALYs discounted at 3.5% per annum
 ICER incremental cost-effectiveness ratio, QALY quality-adjusted life-year

Table 4 One-way sensitivity analysis illustration of method

$R \gg R$	$R \gg P$	$R \gg D$	$P \gg R$	$P \gg P$	$P \gg D$	RR
<i>Section A: Sampled parameter values</i>						
0.799	0.099	0.102	0.000	0.815	0.185	0.690
0.794	0.121	0.084	0.000	0.756	0.245	0.708
0.797	0.073	0.130	0.000	0.694	0.306	0.674
0.825	0.073	0.102	0.000	0.646	0.354	0.693
0.815	0.097	0.088	0.000	0.804	0.196	0.666
<i>Section B: First value for RR</i>						
0.799	0.099	0.102	0.000	0.815	0.185	0.5
0.794	0.121	0.084	0.000	0.756	0.245	0.5
0.797	0.073	0.130	0.000	0.694	0.306	0.5
0.825	0.073	0.102	0.000	0.646	0.354	0.5
0.815	0.097	0.088	0.000	0.804	0.196	0.5
<i>Section C: Second value for RR</i>						
0.799	0.099	0.102	0.000	0.815	0.185	0.6
0.794	0.121	0.084	0.000	0.756	0.245	0.6
0.797	0.073	0.130	0.000	0.694	0.306	0.6
0.825	0.073	0.102	0.000	0.646	0.354	0.6
0.815	0.097	0.088	0.000	0.804	0.196	0.6

D dead, P progression, R remission, RR relative risk of progression with New, $X \gg Y$ probability of transitioning from state X to state Y

this is not necessarily the case as it is dependent on model structure.

4 One-Way Sensitivity Analyses

Conducting a probabilistic analysis provides the data required for a probabilistic sensitivity analysis by definition, and the outputs can be used to illustrate decision uncertainty using standard techniques such as cost-effectiveness acceptability curves [15]. Whilst the general direction of travel in decision modelling has been towards probabilistic analysis for some time [10], there are situations when an OWSA is still of value to enhance understanding of an analysis, for example, to determine the threshold value of a key parameter at which the adoption decision changes.

In these situations, for the reasons outlined in the introduction, it is important to show how the *expected* costs and outcomes vary with changes in the parameter of interest. However, a simple OWSA on the deterministic results does not show how the *expected* values change. It therefore risks “provid[ing] decision makers with biased and incomplete information” [16]. McCabe et al. [16] presented a promising methodological development in the area. The expected value of perfect parameter information can also be used to show where there is greatest value in eliminating uncertainty in the model [7, 8] and thus, by implication, the parameters to which the ICER is most sensitive. However, another method that generates the expected costs and outcomes is to hold the

parameter of interest at one value and run the probabilistic analysis to calculate the expected costs and outcomes, then hold it at the next value and record the expected values, repeating for all values of interest. This answers the question “what is the ICER if the parameter of interest is known to equal x with certainty?”

This can be quite burdensome, especially when using relatively slow software with heavy graphical overheads such as Microsoft Excel. However, careful coding can alleviate the majority of this. For example, if the interest is only in extreme values, such as in the construction of a tornado diagram, the analysis only needs conducting at the upper and lower extremes. VBA code can easily handle the repetitive nature of the exercise. The most pragmatic method (certainly in R) is to simply replace the set of sampled values for the parameter of interest and recalculate the model. This has the benefit of avoiding extra noise and computational expense from resampling the entire dataset. I detail this in the following.

Suppose we wish to conduct an OWSA on the treatment effect (relative risk of progression, parameter ‘RR’) between the values of 0.5 and 1.0. Table 4 section A shows the first five sets of sampled input parameters. To conduct the analysis, we simply substitute all values in column ‘RR’ with the first value (0.5) as per Table 4 section B and calculate the expected costs and QALYs from Old and New and the resulting ICER. We then repeat with the next value (0.6, Table 4 section C) and so on. The results can be presented in the conventional manner, for example as per Fig. 2. (For

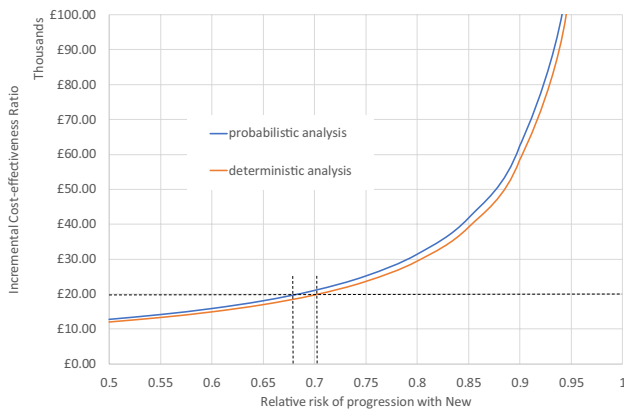


Fig. 2 One-way sensitivity analysis on relative risk of progression with New: comparison of deterministic and probabilistic results. Dotted lines indicate the threshold value for relative risk at a willingness to pay of £20,000 per quality-adjusted life-year. The deterministic analysis finds that the relative risk must be below ~0.705 for New to be cost effective, yet the probabilistic analysis finds that it must be below ~0.68. Raw data are presented in the electronic supplementary material

more information on the conduct and presentation of results of sensitivity analyses, as well as decision modelling and economic evaluation in general, I refer readers to sources such as Drummond et al. [17] and Briggs et al. [18].)

Figure 2 is plotted with both the deterministic and the probabilistic analyses (probabilistic analyses based on 10,000 simulations, data table in the Appendix in the ESM). Again, the ICER under the deterministic analysis is consistently below that of the probabilistic, with consequent implications for interpretation. Under the deterministic analysis, the ICER is below £20,000 per QALY so long as the RR of progression with New is below approximately 0.705. However, the ‘true’ value, under the probabilistic analysis, is approximately 0.68: the deterministic analysis underestimates how effective New must be in order to be cost effective at its given price.

Note that the ICER generated at $RR = 0.7$ (£20,974.74; Appendix Table in the ESM) is not the same as the base-case probabilistic ICER (£20,979.96 from 10,000 simulations; Table 3), despite using the same set of random samples. This is because the OWSA answers the question “what is the ICER if the relative risk is exactly 0.7?”, whereas the base analysis is run with samples from the distribution of RR, which has a mean of 0.7.

I suggest the method just described be used for parameters that have a non-linear relationship with the outcomes (costs and QALYs). However, it is common for a model to comprise a decision tree with Markov models attached to the terminal nodes, for example, a short-term diagnostic and treatment pathway followed by long-term prognosis. The parameters within the decision tree (e.g. prevalence of

disease, sensitivity and specificity of a diagnostic test) have a linear relationship with the outcomes and hence a deterministic one-way analysis is mathematically sound, provided the expected values of the cost and QALYs from the Markov models are substituted for the models themselves at the terminal nodes. However, if the model code is already written to conduct a one-way analysis as described, it may be more expedient to use that same code rather than writing anew.

5 Discussion

The described examples demonstrate explicitly that deterministic analyses of non-linear models, such as Markov models, lead to biased estimates of the mean cost and outcomes. Differences between deterministic and probabilistic results are not simply due to noise (Monte Carlo error) but to a genuine bias. It has been argued that the bias in most cases is ‘modest’ and so can be ignored [19]. In the opinion of the author, this may have been satisfactory when computational power limited the feasibility of conducting Monte Carlo simulations with sufficient iterations. However, a rational manufacturer should price their intervention to yield an ICER right at or just below the payer’s threshold to extract the maximum producer surplus; this is also the point of maximum decision uncertainty where small differences can swing the accept/reject decision. In the experience of the author, deterministic analyses where there are two comparators tend to underestimate the ‘true’ ICER; where there are multiple comparators, the bias can be much bigger and swing in either direction. (A systematic review is warranted to establish this empirically.) Thus, the payer is at risk of approving an intervention that is not cost effective or failing to approve a value-for-money intervention. Software tools such as R provide an order of magnitude increase in processing speed over, for example, Microsoft Excel, and high-powered clusters of cloud-based processors are cheaply available on demand, thus weakening the computational power argument.

Whilst the focus of this paper is Markov models, the same is true for other non-linear structures such as partitioned survival models and microsimulations. The latter has the added challenge of capturing both first- and second-order uncertainty via nested loops, increasing computational expense exponentially. Graphically heavy and non-vectorised software such as Microsoft Excel is extremely slow to process such analyses, so I would recommend a programming language such as R [11], which—besides being extremely fast—is also open source and can therefore be deployed freely on high-power clusters, taking full advantage of parallel computing. Indeed, even for the simple Markov model in this example, whilst R was able to process 10,000 Monte Carlo simulations in approximately 5.5 s and the OWSA in 31.5 s (using a single core without parallelisation), Excel took 36 s and 6.9 min, respectively.

Current health technology assessment (HTA) agency guidelines are not explicit as to their requirement for probabilistic analyses of non-linear models, with the notable exception of the Canadian Agency for Drugs and Technologies in Health (CADTH), which states clearly that “final results should be based on expected costs and expected outcomes. These should be estimated through probabilistic analysis” [20]. The Pharmaceutical Benefits Advisory Committee in Australia [21] and the Institute for Clinical and Economic Review (ICER) in the USA [22] both request base-case analyses and OWSAs presented as tornado diagrams, followed by probabilistic sensitivity analyses, implying the base case should be deterministic. However, the ICER requests a “deterministic base-case analysis (*if appropriate for model type*)” [23] (but without further elaboration) and that “expected values of costs and outcomes for each intervention are *also* estimated through probabilistic sensitivity analysis” [22] (emphases added). NICE guidelines (England) [14] state that “in non-linear decision models, probabilistic methods provide the best estimates of mean costs and outcomes.” Australian, US, and UK guidance thus show awareness of the issue but, in my opinion, could be more explicit and prescriptive. Greater prominence and clarity are warranted in future revisions, for example clearly expressing a preference for probabilistic analyses where possible.

6 Recommendations

Analysts should avoid presenting deterministic analyses of Markov and other non-linear models as the base- or reference-case analysis, as these provide a biased estimate of the ICER. If deterministic analyses must be presented, they should be relegated to supplementary material and couched with appropriate caveats. HTA agencies should consider emphasising the need for probabilistic analysis of non-linear models in future updates of their guidelines.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40258-021-00664-2>.

Declarations

Funding No sources of funding were used to conduct this study or prepare this manuscript.

Conflicts of interest Edward CF Wilson has no conflicts of interest that are directly relevant to the content of this article.

Availability of data and material Available as supplementary material.

Code availability Available as supplementary material.

Authors' contributions ECFW was solely responsible for all aspects of creating this manuscript.

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* (Clinical research ed). 1996;312(7023):71–2.
2. Paulden M. Calculating and interpreting ICERs and net benefit. *PharmacoEconomics*. 2020;38(8):785–807. <https://doi.org/10.1007/s40273-020-00914-6>.
3. Huygens C. De Ratiociniis in Ludo Aleae (The Value of all Chances in Games of Fortune; Cards, Dice, Wagers, Lotteries etc. Mathematically Demonstrated.). S. Keimer for T. Woodward, London. 1657 (English translation 1714). <https://math.dartmouth.edu/~doyle/docs/huygens/huygens.pdf>. Accessed 14 Apr 2020.
4. Raiffa H, Schlaifer R. Applied statistical decision theory. Boston: Harvard Business School; 1961.
5. Pratt J, Raiffa H, Schlaifer R. Introduction to statistical decision theory. Cambridge: Massachusetts Institute of Technology; 1995.
6. Arrow KJ, Lind RC. Uncertainty and evaluation of public investment decisions. *Am Econ Rev*. 1970;60(3):364–78.
7. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ*. 1999;18(3):341–64.
8. Wilson EC. A practical guide to value of information analysis. *PharmacoEconomics*. 2015;33(2):105–21. <https://doi.org/10.1007/s40273-014-0219-x>.
9. Claxton K, Cohen JT, Neumann PJ. When is evidence sufficient? *Health Aff (Millwood)*. 2005;24(1):93–101.
10. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Econ*. 2005;14(4):339–47. <https://doi.org/10.1002/hec.985>.
11. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2019. <https://www.R-project.org/>.
12. Wilson ECF. rrapidMarkov—a package for Markov models in R. 2021. <https://github.com/EdCFWilson/rrapidmarkov>. Accessed 8 Feb 21.
13. Warnes G, Bolker B, Lumley T. gtools: various R Programming Tools. R package version 3.8.2. 2020. <https://CRAN.R-project.org/package=gtools>. Accessed 14 Apr 2020.
14. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. NICE, London. 2013. <http://www.nice.org.uk/media/D45/1E/GuideToMethodsTechnologyAppraisal2013.pdf>. Accessed 14 Apr 2020.
15. Fenwick E, Byford S. A guide to cost-effectiveness acceptability curves. *Br J Psychiatry*. 2005;187:106–8. <https://doi.org/10.1192/bjp.187.2.106>.
16. McCabe C, Paulden M, Awotwe I, Sutton A, Hall P. One-way sensitivity analysis for probabilistic cost-effectiveness analysis: conditional expected incremental net benefit. *PharmacoEconomics*. 2020;38(2):135–41. <https://doi.org/10.1007/s40273-019-00869-3>.
17. Drummond M, Sculpher M, Claxton K, Stoddart G, Torrance G. Methods for the economic evaluation of health care programmes. 4th ed. Oxford: Oxford University Press; 2015.
18. Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.
19. Briggs A, Sculpher M, Claxton K. Chapter 4.1.1: Uncertainty and nonlinear models. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press; 2006. p. 78.
20. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada. 4th Edition. 2017. <https://www.cadth.ca/dv/guidelines-economic-evaluation-health-technologies-canada-4th-edition>. Accessed 14 Apr 2020.

21. Pharmaceutical Benefits Advisory Committee. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (PBAC) v5.0. 2016. <https://pbac.pbs.gov.au/>. Accessed 14 Apr 2020.
22. Institute for Clinical and Economic Review. A guide to ICER's methods for health technology assessment. 2018. http://icer-review.org/wp-content/uploads/2018/08/ICER-HTA-Guide_082018.pdf. Accessed 14 Apr 2020.
23. Institute for Clinical and Economic Review. ICER's reference case for economic evaluations: Principles and rationale. 2018. http://icer-review.org/wp-content/uploads/2018/07/ICER_Reference_Case_July-2018.pdf. Accessed 14 Apr 2020.