Taylor & Francis
Taylor & Francis Group

ORIGINAL RESEARCH

Check for updates

Comparison of the expected rewards between probabilistic and deterministic analyses in a Markov model

Xuanqian Xie[a], Man Wah Yeung[a], Zhuoyu Wang[b], Myra Wang[a], Olga Gajic-Veljanoski[a], Vivian Ng[a] and Andrei Volodin[c]

[a]Evidence Development and Standards, Health Quality Ontario, Toronto, Canada; [b]Carrefour de l'innovation et de l'évaluation en santé, Centre de recherche du Centre hospitalier de l'Universite de Montreal, Montreal, Canada; [c]Department of Mathematics and Statistics, University of Regina, Regina, Canada

ABSTRACT

Objectives: In Markov models that evaluate the cost-effectiveness of health-care technologies, it is generally recommended to use probabilistic analysis instead of deterministic analysis. We sought to compare the performance of probabilistic and deterministic analysis in estimating the expected rewards in a Markov model.

Methods: We applied Jensen's inequality to compare the expected Markov rewards between probabilistic and deterministic analysis and conducted a simulation study to compare the bias and accuracy between the two approaches.

Results: We provided mathematical justification why probabilistic analysis is associated with greater Markov rewards (life-years and quality-adjusted life-years) compared with deterministic analysis. In our simulations, probabilistic analyses tended to generate greater life-years, bias, and mean square error for the estimated rewards compared with deterministic analyses. When the expected values of transition probabilities were the same, weaker evidence derived from smaller sample sizes resulted in larger Markov rewards compared with stronger evidence derived from larger sample sizes. When longer time horizons were applied in cases of weak evidence, there was a substantial increase in bias where the rewards in both probabilistic and deterministic analysis were overestimated.

Conclusion: Authors should be aware that probabilistic analysis may lead to increased bias when the evidence is weak.

1. Introduction

In recent years, there has been a shift in economic modeling from the use of deterministic analysis to probabilistic analysis for evaluating the cost-effectiveness of health-care technologies. Deterministic analysis considers only the expected values of model inputs – that is, point estimates of costs and clinical inputs. Probabilistic analysis considers model inputs as probability distributions (i.e., sampling distributions of the sample mean, not the distributions of the samples) to account for parameter uncertainty [1]. Model outputs (costs and effects) are subsequently also distributions, obtained by Monte Carlo simulations. In each run, one set of results is generated based on the values of the model inputs randomly drawn from their pre-specified distributions.

The use of probabilistic methods was mainly advocated to represent parameter uncertainty, termed as 'probabilistic sensitivity analysis' (PSA) [2]. The results of a PSA can be summarized as a cost-effectiveness acceptability curve (CEAC) or a scatter plot, both of which show the probability of a health-care technology being cost-effective compared with an alternative [3]. Many advantages have been cited for PSA. Firstly, standard sensitivity analysis becomes unwieldy when many parameters in a decision analytic model are estimated with uncertainty. PSA can also account for uncertainty from the joint distributions of correlated

parameters (e.g., from a regression analysis) [4]. Furthermore, it can be a starting point to measure decision uncertainty in terms of value of information, which can quantify the probability and consequences of erroneous decision-making through additional calculations of either net health benefit or net monetary benefit [1,5]. From a policy perspective, it may be important to not only generate results as point estimates (deterministic), but also understand the uncertainty associated with the results, and ultimately the decision being made. Recently, some guidelines state that probabilistic analysis provides 'best estimates' [3] and 'less biased estimates' [6] of costs and outcomes compared with a deterministic analysis in a non-linear model (e.g., Markov model). This encourages the use of the probabilistic approach over the deterministic approach. Guidelines from some institutions recommend probabilistic analysis not only for sensitivity analyses, but also for the reference case to estimate the expected cost and clinical outcomes [6,7]. These recommendations differ from the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Modeling Good Research Practices Task Force, which is not definitive in its recommendation of using deterministic versus probabilistic analysis [4].

Therefore, we sought to understand why probabilistic analysis is less biased than deterministic analysis in the Markov model. We considered the references provided in the guidelines [3,6] and also conducted a literature search.

CONTACT Xuanqian Xie ✉ shawn.xie@hqontario.ca 🖃 shawn.xie@hqontario.ca, 130 Bloor Street West, 10th floor, Toronto, ON M5S 1N5, Canada

Although several publications discuss the advantages of probabilistic analysis over deterministic analysis [5,8–11], we did not find any simulation studies that compared the performance (i.e., bias and accuracy) of the two in estimating the expected cost and effectiveness in Markov models. In the real world, evidence for decision-making is often not strong, and the true distributions of parameters of interest may not be adequately understood from the samples obtained (e.g., due to small sample size studies that inform input model parameters). Therefore, it may be challenging to assign proper distributions for these parameters in a probabilistic analysis. Furthermore, when the sample size is small and the variance is large, the probabilistic analysis may yield extreme values that tend to bias the results (i.e., the expected mean value of the model outputs) in some types of non-linear models (e.g., $f(x) = \frac{1}{x}$). The deterministic analysis uses point estimates as model input parameters; consequently, there is a smaller chance of observing extreme results which potentially could yield a less biased result. To better understand the performance of both the probabilistic and deterministic analysis in a Markov model, especially when model input parameters are informed by weak evidence, we conducted a simulation study and compared these two analytical approaches for estimating important Markov rewards such as expected life-years.

Section 2 provides an example of a probabilistic analysis generating greater life-years using weaker evidence (e.g., data derived from smaller sample sizes). Section 3 introduces justifications of why the Markov rewards (life-years and quality-adjusted life-years [QALYs]) in a probabilistic analysis are greater than in a deterministic analysis, and why weaker evidence in a probabilistic analysis generally leads to larger rewards compared with stronger evidence. In Section 4, we present our simulations to compare the bias and accuracy of the expected reward (the model output) in probabilistic versus deterministic analysis. In Section 5, we examine how the time horizon impacts the bias of the estimated rewards in probabilistic and deterministic analyses. We examine the limitations in Section 6 and discuss our findings in Section 7. In the Supplementary Materials, we provide additional explanation for the estimation of the true estimand and bias for both analytical approaches for a non-linear model in a simulation study.

2. Favorable results for weaker evidence in probabilistic analysis

Probabilistic analysis is a powerful tool in decision-making although its interpretation is not straightforward. Attention should be given to the level of uncertainty associated with parameter inputs, as they influence the expected rewards generated probabilistically from a Markov model.

For example, a model input that uses a treatment effect derived from weak evidence (e.g., based on a small sample size) generally leads to misleadingly favorable results in terms of life-years and QALYs in a probabilistic analysis. We illustrate this in an example below.

We developed a simple Markov model that included three states: healthy, disease, and dead (see Figure S1 in the Supplementary Materials). All patients are initially in the healthy state and only one-way transitions are allowed in the model (i.e., people can only move to worse health states and cannot move back to better health states). Thus, there are three possible transitions: P_{12} (from healthy state to disease state), P_{13} (from healthy state to dead state), and P_{23} (from disease state to dead state). We assumed the transition probabilities are constant. We used the duration in the surviving states (i.e., life-years) as the reward. We set the cycle length at 1 year, the time horizon to 100 years, and did not apply discounting and half-cycle correction.

Using this model, we compared the standard of care (Trt A) with a new treatment (Trt B). Both treatments have the same expected value of P_{12}, 0.07. However, Trt B has weaker evidence generated from a cohort with a smaller sample size [distribution of P_{B12}: Beta ~ (7, 93)], while Trt A has stronger evidence [distribution of P_{A12}: Beta ~ (70, 930)]. The distributions of Beta ~ (7, 93) and Beta ~ (70, 930) can be found in Figure S2 in the Supplementary Materials. We also assumed that Trt A and Trt B have the same parameters for the remaining two transition probabilities, P_{13}: Beta ~ (30, 970) and P_{23}: Beta ~ (50, 950). Probabilistic analysis shows that Trt B has a greater expected reward than Trt A (24.38 versus 24.15 life-years over a time horizon of 100 Markov cycles). Furthermore, the probability of Trt B having a greater reward than Trt A was 54% in the probabilistic analysis. In simulations where the evidence for Trt B was even weaker, the difference in reward between Trt B and Trt A was even greater. If we test other transition probabilities for P_{12}, P_{13}, and P_{23}, we arrive at a similar result: weaker evidence is associated with greater rewards. The next section describes the reasons for this finding.

3. Justifications for why probabilistic analyses generate greater life-years and QALYs than deterministic analyses

In linear relationships, we can use a straight line to represent the relationship between the independent and dependent variables. For every unit of change in the independent variable, the dependent variable has a corresponding proportional change. This relationship does not occur in a non-linear relationship, such as Markov models. Therefore, the expected rewards (i.e., model outputs) of a probabilistic or deterministic analysis in a Markov model will be different.

Elbasha and Chhatwal characterized the conditions for upward or downward bias using a three-state Markov model and derived rewards (e.g., life-years, QALYs, lifetime disease costs) using mathematical expressions for model inputs (e.g., transition probabilities) [12]. Further, the authors used Jensen's inequality to evaluate the curvature of each reward function with model parameters. They demonstrated that life-years and QALYs are convex functions of transition probabilities. Based on these findings, we can deduce that life-years and QALYs estimated using probabilistic analysis are greater than those using deterministic analysis. The justification is provided below.

If a function is convex on an interval, for all x_1 and x_2 in this interval and $0 \le \lambda \le 1$, the following inequality holds:
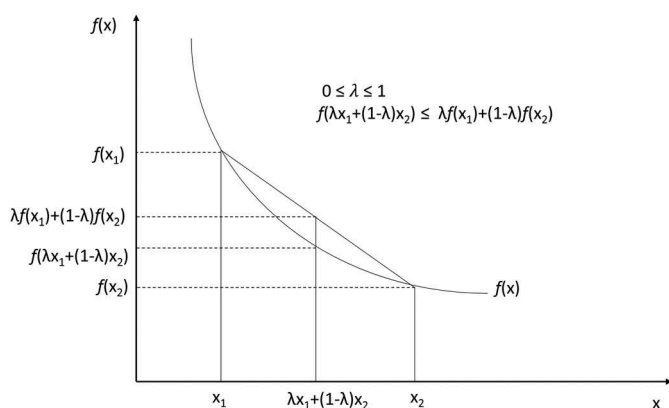
Figure 1. Convex function of Jensen's inequality.

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \qquad (3.1)$$

Figure 1 illustrates the concept of convexity using a simple geometric interpretation. If a function is convex, the graph of this function is below the secant line.

Mercer then extended Jensen's inequality, as shown below [13]:

Let $f(x)$ be a convex function. Let $0 < x_1 \leq x_2 \leq \ldots \leq x_n$ and w_k be positive weights for x_k, then:

$$f\left(\sum w_k x_k\right) \leq \sum w_k f(x_k) \qquad (3.2)$$

Let x be the transition probability and $f(x)$ be the life-years in a Markov model. Let us assume that n samples are simulated in the probabilistic analysis, and let each simulation have the same weight (i.e., 1/n) so that the sum of weights equals 1. When n is large, $\sum w_k x_k$ would equal the expected value of x (i.e. E(x)). Then, $f(\sum w_k x_k)$ would equal to $f(E(x))$, which represents the expected rewards in the deterministic analysis. The right side of the inequality expression 3.2, $\sum w_k f(x_k)$, is equal to the average reward in the probabilistic analyses in n simulations. Based on inequality 3.2, we can conclude that the rewards in life-years and QALYs of a Markov model from deterministic analysis are less than those from probabilistic analysis, irrespective of the distributions used for transition probabilities.

When model inputs are obtained from strong evidence, the parameter uncertainty is small and is concentrated around the mean value, so the rewards (life-years and QALYs) derived from probabilistic analysis will be close to the rewards derived from deterministic analysis. However, if model inputs are obtained from weak evidence, the range of model inputs is wider, so the differences in rewards between probabilistic and deterministic analyses can be larger. Therefore, weak evidence generally leads to greater or equal life-years and QALYs compared with strong evidence.

4. Assessing the bias and accuracy of probabilistic versus deterministic analysis in a Markov model: a simulation study

To estimate clinical and economic outcomes, economic models often include many model inputs including costs, health utilities, and transition probabilities. With so many different parameters in a simulation, it is difficult to predict how the model inputs of interest will impact the many model outputs. Using transition probabilities only (i.e., no costs and utilities), we can estimate the duration of time spent in selected states. In a disease progression Markov model, we can treat the duration in the surviving states (i.e., life-years) as the reward. The life-year generally has a strong correlation with other outcomes in the economic evaluation. For instance, greater life-years generally correspond to greater costs and QALYs. In our simulation study, we generated transition probabilities and chose life-years as the reward. However, we discuss other model outcomes (QALYs and costs) at the end of this section.

In the simulation described below, we compared probabilistic and deterministic analysis for estimating the reward in a Markov model.

4.1 Markov model

We used the same model (Figure S1 in the Supplementary Materials) as described in Section 2.

4.2 Data generation

We defined the true transition probabilities as 0.07, 0.03, and 0.05 for P_{12}, P_{13}, and P_{23}, respectively. We assumed that these three transition probabilities are obtained from three independent cohorts with binomial data. Let X denote the number of events of interest from a given sample size cohort (n) for a given true transition probability. We considered three scenarios with different cohort sizes: n = 100 for P_{12}, P_{13}, and P_{23} (i.e., weak evidence); n = 500 for P_{12}, P_{13}, and P_{23} (i.e., moderate evidence), and n = 2,000 for P_{12}, P_{13}, and P_{23} (i.e., strong evidence). We then estimated the transition probabilities for the deterministic analysis (i.e., p = X ÷ n) and the probabilistic analysis (i.e., p ~ Beta (X, n − X)). Since X is a random variable, it can result in different transition probabilities (i.e., input parameter estimation) in different simulations. We calculated the rewards for both types of analyses. Note that the rewards of the probabilistic analysis were from the average value of 2,000 simulations. We repeated the deterministic and probabilistic analysis 5,000 times (i.e., 5,000 repetitions). Figure 2 outlines the steps of our simulation study in assessing the performance of probabilistic analysis and deterministic analysis.

4.3 Methods

We used SAS 9.4 to generate datasets for the deterministic and probabilistic analysis and calculated the true reward. The SAS code for the simulation study is available upon request.

4.4 Estimand

The estimand of interest is the total reward (life-years) over 100 cycles (interpreted as the expected survival time for 100 years of follow-up) generated by the deterministic and probabilistic analysis for the Markov model. We obtained the true value of the reward, 23.83 life-years (the estimand), over 100 cycles based on the true P_{12}, P_{13}, and P_{23} [14]. In the main
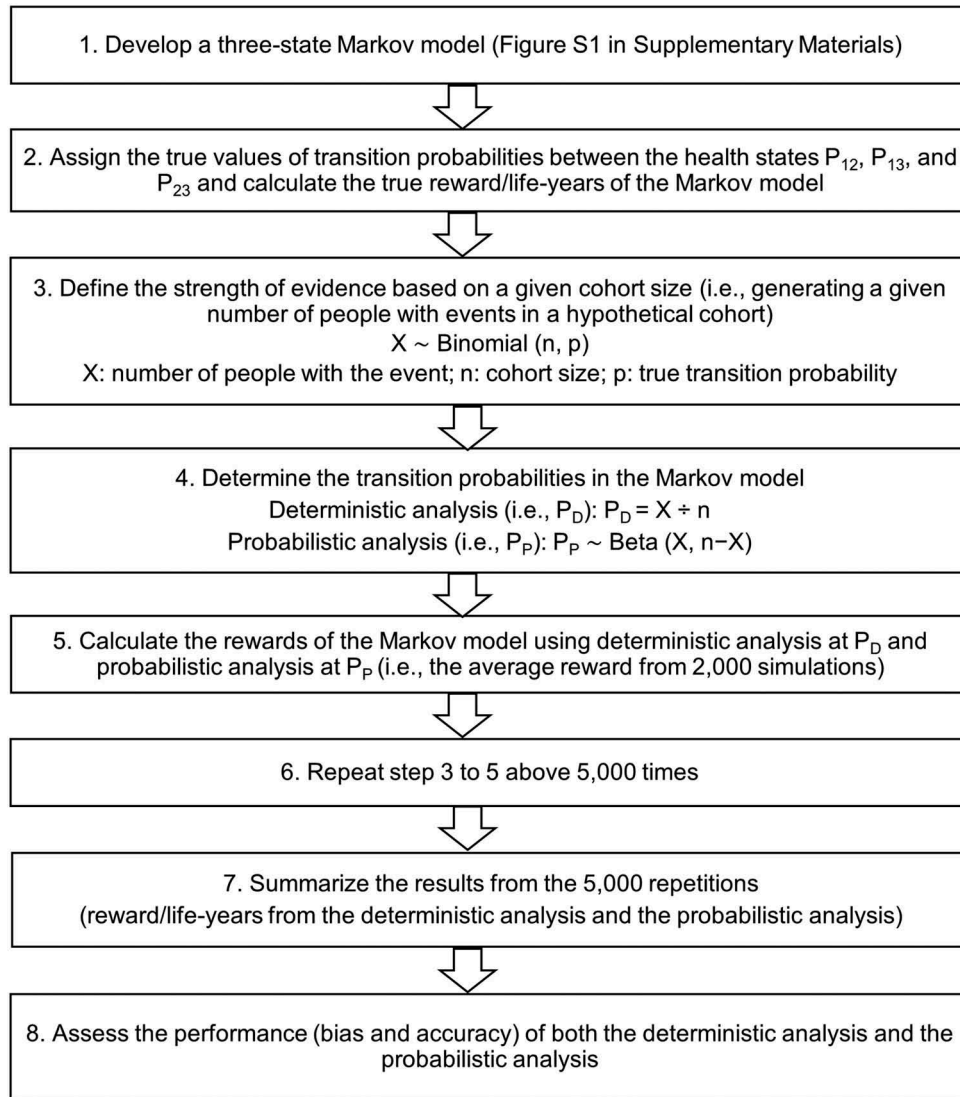
Figure 2. The Process of Assessing the Bias and Accuracy of Probabilistic Analysis versus Deterministic Analysis in Markov Model.

analysis, we used the fixed true transition probabilities in the Markov model. In the sensitivity analysis, we used the true random transition probabilities. We also discussed how to define the true estimand as a fixed variable or a random variable in a non-linear model in Supplement 2.

4.5 Performance measures

The main performance measures are bias and precision. We compared the bias and mean square error (MSE) of the rewards from the deterministic and probabilistic analysis [14,15]

$$\delta = \overline{\hat{\beta}} - \beta \qquad (3.1)$$

$$MSE = \left(\overline{\hat{\beta}} - \beta\right)^2 + \frac{1}{B-1} \times \sum_{i=1}^{B} \left(\widehat{\beta_i} - \overline{\hat{\beta}}\right)^2$$

$$= \delta^2 + \left(SE\left(\hat{\beta}\right)\right)^2 \qquad (3.2)$$

where δ is the bias, β is the true reward of the population in life-years, $\overline{\hat{\beta}}$ is the sample mean reward estimated from deterministic or probabilistic analysis, B is the number of simulated

samples (i.e., 5,000), $\widehat{\beta_i}$ is the estimated reward in the simulated sample 1, 2, 3… B, and $SE\left(\hat{\beta}\right)$ is the empirical standard error of the reward in the simulated samples (calculated as the standard deviation of all samples).

4.6 Main results

The results comparing the rewards from the deterministic and probabilistic analysis are presented in Table 1. Although the simulation generated unbiased estimates of P_{12}, P_{13}, and P_{23}, both deterministic and probabilistic analysis overestimated the reward. Further, the probabilistic analysis is associated with larger bias and MSE than the deterministic analysis. The bias and MSE reduced as the sample size of the binomial data increased. When the hypothetical cohort size was large (e.g., 2,000, in the case of strong clinical evidence), the bias in both analyses was small. However, when the sample size of the hypothetical cohort was 100 (in the case of relatively weak evidence), the bias in both analyses was large. In particular, the bias in the probabilistic analysis was even larger.

Table 1. Comparison of Deterministic and Probabilistic Analysis in a Markov Model: a Simulation Study.

	True Reward	Deterministic Analysis	Probabilistic Analysis
Sample size of 100 for estimating the number of events from binomial data with true P_{12}, P_{13}, and P_{23}.			
Reward, mean (SD)	23.83	27.36 (10.01)	30.77 (10.85)
Bias	–	3.52	6.93
MSE	–	112.66	165.79
Sample size of 500 for estimating the number of events from binomial data with true P_{12}, P_{13}, and P_{23}.			
Reward, mean (SD)	23.83	24.43 (3.49)	25.08 (3.62)
Bias	–	0.60	1.25
MSE	–	12.53	14.67
Sample size of 2,000 for estimating the number of events from binomial data with true P_{12}, P_{13}, and P_{23}.			
Reward, mean (SD)	23.83	24.00 (1.64)	24.16 (1.66)
Bias	–	0.17	0.33
MSE	–	2.71	2.85

Abbreviations: MSE, mean square error; SD, standard deviation.
Note: We first developed a simple Markov model that included three states: healthy, disease, and dead. There are three possible transitions: P_{12} (from healthy state to disease state), P_{13} (from healthy state to dead state), and P_{23} (from disease state to dead state). The total reward was the life-years over 100 cycles.

4.7. Sensitivity analyses results

- Using the same model (Figure S1 in the Supplementary Materials), we also examined the one-way transition Markov model using other underlying true values of P_{12}, P_{13}, and P_{23}, including using a high value for the transition probability from healthy to disease state (P_{12}).
- We conducted additional analyses to define the multinomial distribution for P_{12} and P_{13} (the joint distribution to replace the independent binomial data in the main analysis).
- We re-structured the model with the clinical event (i.e., P_{12}) conditional on being alive in the healthy state.

In summary, the bias and MSE in the deterministic analysis are generally smaller than in the probabilistic analysis and in the sensitivity analyses above.

We used the random true transition probabilities [P_{12}: Beta ~ (7, 93); P_{A13}: Beta ~ (3, 97); and P_{23}: Beta ~ (5, 95)], which resulted in a larger true estimand (26.43 life-years) compared with using the fixed true transition probabilities in the main analysis. The true distributions of P_{12}, P_{13}, and P_{23}, the method used to estimate transition probabilities, and cohort sizes affected the performance of the deterministic and probabilistic analyses. In general, the deterministic analysis showed smaller bias and MSE for smaller to moderate cohort sizes (e.g., ≤2,000). When cohort sizes were large, both analyses tended to result in estimands close to the true estimand.

4.8 Other model outcomes

When estimating model outputs of QALYs and costs, we need to incorporate more parameters. Similar to life-years, rewards in terms of QALYs are also convex functions of P_{12}, P_{13}, and P_{23} [12]. Elbasha and Chhatwal proved that there is no heterogeneity bias introduced in measuring cost and effectiveness if

heterogeneity only exists in disease cost and utilities [12]. Thus, probabilistic analyses estimating QALYs are likely to lead to greater bias and larger MSE compared with deterministic analyses. This is supported by our preliminary analyses of simulations (data not shown). Total disease cost is concave (i.e., negative of a convex function) with respect to P_{12} (disease progression) but is convex with respect to mortality (P_{13} and P_{23}). The concave and convex effects from different parameters may partially be canceled out by each other, but the impact on the direction of bias is not obvious. However, our preliminary simulation results showed that probabilistic analyses also resulted in greater bias and larger MSE compared with deterministic analysis in estimating disease cost.

5. Impact of time horizon on bias in a Markov model: a simulation study

Time horizon may substantially impact the cost-effectiveness in economic evaluations. Wisløff et al. 2014 reviewed cost–utility studies published in 2010 and found the magnitude of QALYs gained was strongly associated with the time horizon [16]. The mean QALYs gained were 0.04, 0.17, and 0.43 for studies with a time horizon of ≤ 1 year (62 studies), >1 and ≤5 years (55 studies), and >5 years (190 studies), respectively. Kim et al. 2017 reviewed US-based economic studies published between 2005 and 2014. This study reported that 71% of studies (552 out of 782) used a long-term time horizon (>5 years) and 25% (198 out of 782) used a short-term (≤5 years) time horizon [17]. For studies with multiple time horizons, the extension of the time horizon yielded more favorable cost-effectiveness results in most cases (19 out of 23 studies). Currently, most economic evaluation guidelines recommend applying a long-term time horizon to capture the long-term cost and effectiveness [3,6,7,18]. However, the accuracy of outcomes derived probabilistically in a long-term model versus a short-term model has not been well investigated. Below, we introduce an example and then describe how our simulation study examines the impact of time horizon on bias for life-years in a Markov model.

Continuing with the three-health state disease model example, Trt A from Section 2 remains the standard care treatment for this section. We assumed there is another new treatment (Trt C) where the P_{12} is from weaker evidence with an expected value of 0.08 [i.e., Beta ~ (4, 46)]. This transition probability point estimate is higher than that of Trt A, 0.07 [i.e., Beta ~ (70, 930)]. The higher P_{12} means a higher chance of moving from the healthy state to the disease state, and therefore leads to a lower reward (e.g., life-year and QALY). Again, we assumed that P_{13} and P_{23} transitions were the same for both Trt C and Trt A. Given a time horizon of 20 cycles, the reward for Trt A is greater than Trt C (14.47 versus 14.43 life-years). However, if we extend the time horizon to 100 cycles, the reward of Trt A is less than Trt C (24.15 versus 24.20 life-years). Interestingly, Trt A had a lower expected disease progression rate than Trt C (0.07 versus 0.08 annually), but Trt C resulted in greater life-years than Trt A in the long term (100 cycles).

We aimed to understand the reasons leading to these results through simulation. When the evidence is strong, the bias in
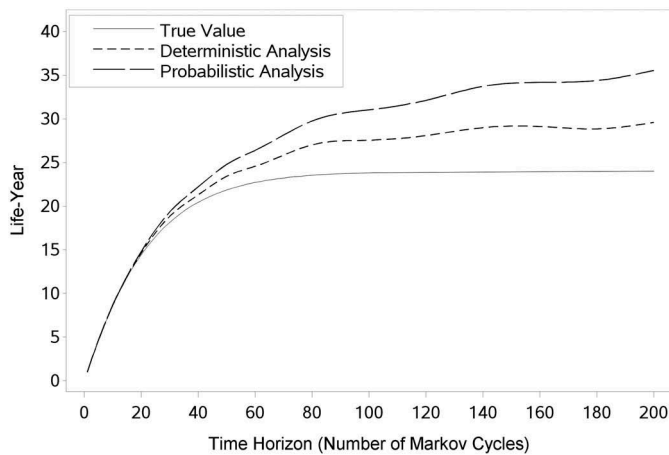
Figure 3. The rewards in deterministic and probabilistic analyses versus time horizon.

both deterministic and probabilistic analyses will be relatively small. Therefore, we examined how time horizon impacts the magnitude of bias given weaker evidence (i.e., hypothetical cohort size of 100) for P_{12}, P_{13}, and P_{23}. We used the same model and true transition probabilities from Section 4. Figure 3 shows the reward for different time horizons (i.e., the number of Markov cycles). In brief, the bias, which is the difference in reward between the deterministic or probabilistic analysis and the true value, dramatically increases as the time horizon increases. Note that the increase is not proportional. Given a time horizon of 20 cycles, the biases from the deterministic and probabilistic analyses were 0.19 and 0.38, respectively; given 50 cycles, the biases were 1.57 and 2.91, respectively; and given 100 cycles, the biases were 3.73 and 7.21, respectively. In summary, when the evidence is weak, a Markov model run over the long term is often associated with large bias and overestimates the reward.

6. Limitations

Our study has several limitations. Firstly, many aspects, including the methodological approaches for defining the true estimand and generating random variables, the sample size of studies used to inform model input parameters, and the function of the model, can affect the results of simulation studies that aim to evaluate the performance (bias and precision) of deterministic and probabilistic analyses. We only selected some typical situations in the simulations, so this may not characterize the direction and magnitude of bias in probabilistic and deterministic analysis for other distributions and/or more complex conditions. Secondly, model outcomes from probabilistic analysis are more informative than point estimates from a deterministic analysis. For example, the expected value of perfect information from outputs of probabilistic analysis provides the upper bound of the net health or net monetary benefits of future research, which is very important for decision-making [1,5]. It can also be extended to calculate the expected value of perfect information of a parameter (or parameters). Our simulation study did not reflect this important advantage of probabilistic analysis.

7. Discussion

To our knowledge, this is the first simulation study to compare the bias and accuracy of the rewards generated by probabilistic versus deterministic analysis in Markov models, and also the first that explores the impact of time horizon on the magnitude of bias. Although our examples are drawn from simple one-way disease progression Markov models with three health states, they illustrate cautionary tales of when probabilistic and deterministic results can greatly diverge.

With the shift toward using probabilistic analyses in the fields of decision analysis and health economics, we encourage modelers to consider the limitations of estimating the expected values in probabilistic analyses when there is weak evidence. We are proponents of conducting both probabilistic and deterministic analyses in economic evaluations to explore possible discordance in the expected costs and effectiveness between these two approaches. Our study indicates that if discordance occurs, deterministic analysis may offer less biased estimates in effectiveness.

In general, a lifetime model should be considered for economic evaluations [3,6,7,18]. However, sometimes existing evidence for an intervention is over the short term only and is weak, while long-term consequences are unknown. Since our simulation illustrated the potential for substantially increased bias in a long-term model, researchers need to balance the potential benefits and limitations of long-term disease modeling to justify the selection of the most appropriate time horizon when the evidence is weak.

Key issues

- The average life-year rewards and QALYs in a probabilistic analysis are generally larger than in a deterministic analysis.
- Based on the same expected values of transition probabilities, weaker evidence results in larger life-years and QALYs compared with stronger evidence in a probabilistic analysis.
- Bias and mean square error (MSE) of calculated rewards (life-years) are generally smaller in a deterministic analysis than a probabilistic analysis. Bias and MSE reduce as the sample size increases in both deterministic and probabilistic analyses.
- When the evidence is weak, longer time horizons are generally associated with substantially increased bias and overestimate the life-years in both probabilistic and deterministic analyses.
- Researchers need to balance the potential benefits and limitations of a long-term model to select the appropriate time horizon in Markov models.

Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties. The opinions expressed in this publication do not necessarily represent the opinions of Health Quality Ontario. No endorsement is intended or should be inferred.

Reviewer Disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

Author contribution statement

Xuanqian Xie conceived the study idea and simulated the data. Xuanqian Xie and Man Wah Yeung designed the study and drafted the manuscript. Zhuoyu Wang, Myra Wang, Olga Gajic-Veljanoski, Vivian Ng, and Andrei Volodin provided important intellectual content, critically revised the manuscript, and interpreted the results. All authors approved this version and agree to be accountable for all aspects of the work.

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Briggs A, Sculpher M, Claxton K. Decision modelling for health economic evaluation. Oxford (UK): Oxford University Press; 2006.
   •• **A seminal health economic modeling textbook that introduces probabilistic and deterministic analyses and the expected value of perfect information.**
2. Briggs A. Probabilistic analysis of cost-effectiveness models: statistical representation of parameter uncertainty. Value Health. 2005;8(1):1–2.
3. National Institute for Health and Care Excellence. Methods for the development of NICE public health guidance. 3rd ed. London (UK): The Institute; 2012.
4. Briggs AH, Weinstein MC, Fenwick EA, et al. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM modeling good research practices task force–6. Value Health. 2012;15(6):835–842.
   • **An article that provides guidance on how to properly define distributions in probabilistic analyses.**
5. Claxton K. Exploring uncertainty in cost-effectiveness analysis. Pharmacoeconomics. 2008;26(9):781–798.
6. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: canada. 4th ed. Ottawa (ON): The Agency; 2017. p. 76.
7. Institute For Clinical And Economic Review. ICER's reference case for economic evaluations: principles and Rationale. 2018;15. https://icer-review.org/wp-content/uploads/2018/07/ICER_Reference_Case_July-2018.pdf
8. Thompson KM, Graham JD. Going beyond the single number: using probabilistic risk assessment to improve risk management. Hum Ecol Risk Assess. 1996;2(4):1008–1034.
9. Doubilet P, Begg CB, Weinstein MC, et al. Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach. Med Decis Making. 1985;5(2):157–177.
10. Griffin S, Claxton K, Hawkins N, et al. Probabilistic analysis and computationally expensive models: necessary and required? Value Health. 2006;9(4):244–252.
11. Claxton K, Sculpher M, McCabe C, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ. 2005;14(4):339–347.
12. Elbasha EH, Chhatwal J. characterizing heterogeneity bias in cohort-based models. Pharmacoeconomics. 2015;33(8):857–865.
   •• **A study that demonstrated life-years and QALYs are convex functions of transition probabilities in a Markov model.**
13. Mercer AMD. A variant of Jensen's Inequality. J Inequalities Pure Appl Math. 2003;4(4): 1-2. Article 73.
   •• **A study that provides proof for an extension of Jensen's inequality.**
14. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Stat Med. 2019;38(11):2074–2102.
   • **An article that provides guidance on designing simulation studies to evaluate statistical methods.**
15. Burton A, Altman DG, Royston P, et al. The design of simulation studies in medical statistics. Stat Med. 2006;25(24):4279–4292.
16. Wisloff T, Hagen G, Hamidi V, et al. Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. Pharmacoeconomics. 2014;32(4):367–375.
17. Kim DD, Wilkinson CL, Pope EF, et al. The influence of time horizon on results of cost-effectiveness analyses. Expert Rev Pharmacoecon Outcomes Res. 2017;17(6):615–623.
18. European Network For Health Technology Assessment. Methods for health economic evaluations: A guideline based on current practices in Europe. 2015;100. https://www.eunethta.eu/wp-content/uploads/2018/03/Methods_for_health_economic_evaluations.pdf