Evaluation of Markov models with discontinuities

J. Pérez-Martín, MEng,[1] Iñigo Bermejo, PhD,[2] and Francisco J. Díez, PhD[1]

February 7, 2019

[1]Department of Artificial Intelligence, Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain

[2]School of Health and Related Research, University of Sheffield, UK.

Financial Disclosures/Conflicts of Interest:

There are no financial conflicts of interest to disclose.

Address correspondence to J. Pérez-Martín, Dept. Artificial Intelligence, UNED. Juan del Rosal 16, 28040 Madrid, Spain. E-mail: jperezmartin@dia.uned.es.

Abstract

Background: Several methods, such as the half-cycle correction and the life-table method, were developed to attenuate the error introduced in Markov models by the discretization of time. Elbasha and Chhatwal have proposed alternative "corrections" based on numerical-integration techniques. They present an example whose results suggest that the trapezoidal rule, which is equivalent to the half-cycle correction, is not as accurate as Simpson's 1/3 and 3/8 rules. However, they did not take into consideration the impact of discontinuities.

Objective: To propose a method for evaluating Markov models with discontinuities.

Design: Applying the trapezoidal rule, we derive a method that consists in adjusting the model by setting the cost at each point of discontinuity to the mean of the left and right limits of the cost function. We then take from the literature a model with a cycle length of 1 year and a discontinuity on the cost function and compare our method with other "corrections" using as gold standard an equivalent model with a cycle length of 1 day.

Results: As expected, for this model the life-table method is more accurate than assuming that transitions occur at the beginning or the end of cycles. The application of numerical integration techniques without taking into account the discontinuity causes large errors. The model with averaged cost values yields very small errors, especially for the trapezoidal and the 1/3 Simpson rules.

Conclusion: In the case of discontinuities, we recommend applying the trapezoidal rule on an averaged model because this method has a mathematical justification and in our empirical evaluation it was more accurate than the sophisticated 3/8 Simpson rule.

1 Introduction

Markov models are the most popular modeling framework in health technology assessment (Beck and Pauker, 1983; Sonnenberg and Beck, 1993). They represent the state of a patient as a set of mutually exclusive and exhaustive health states, such that the patient is in one and only one of them at a time. All the possible events are represented by transitions from one state to another. Every state has an associated cost and an effectiveness value. The total cost and effectiveness, accumulated over time, determine the net benefit of each intervention.

In economic evaluations of health technologies most Markov models use a discrete-time approach, i.e., the time horizon is divided into a finite number of intervals of the same length, called *cycles*. In the classical presentation of these models, transitions can only occur at the boundary between consecutive intervals. The evaluation of a Markov model applies transition matrices to calculate for each cycle the probability that the patient is in one or another state. In the case of monotonically decreasing costs—which typically happens when patients are dying progressively—the assumption that transitions occur at the end of each cycle overestimates the cost, and assuming that they occur at the beginning underestimates it. The *half-cycle correction* (HCC) (Sonnenberg and Beck, 1993; Naimark et al., 2008) tries to minimize the error by adding a cycle of half duration at the beginning of the process, which can be interpreted as assuming that transitions occur exactly in the middle of each cycle. Current guidelines for economic evaluation of health technologies recommend using this correction (Siebert et al., 2012). However, in recent years there has been a controversy about how to interpret and apply this method, and whether it should be replaced with a different approach, the *life-table method*, which averages the probabilities of state occupancy at the boundaries of each interval (Naimark et al., 2008, 2013, 2014; Barendregt, 2009, 2014). (Some authors present life tables as a way of implementing the HCC, but we refer to them as a different method for the sake of clarity.)

Recently Elbasha and Chhatwal (2016a,b) have proposed alternative *within-cycle correc-*

tions based on numerical-integration techniques, which assume that cost and utility evolve continuously over time. These authors present an example whose results "suggest that the standard HCC method and the trapezoidal rule are not as accurate as Simpson's 1/3 and 3/8 rules" (Elbasha and Chhatwal, 2016a).

However, some medical models include abrupt changes in costs at certain points in time; for example, when expensive drug treatments are provided only for a limited time. In this case, the application of numerical integration may cause large errors because these techniques are very sensitive to the values of cost chosen for the points of discontinuity. Intuitively, a discontinuity is an abrupt change in the value of a function, as in Figure 2. For a mathematical definition, see Sec. 2.1.2. In this paper we propose a new method for evaluating Markov models with discontinuities, based on the application of the trapezoidal rule. At the points of discontinuity, instead of setting the cost to the left or the right limit of the cost function, we set it to the average of the limits. We then take a model from the literature—in fact, the model that made us aware of the problem of discontinuities—and, using a slightly modified version of it as a gold standard, compare several "corrections". Our experiments show that the direct application of numerical integration may cause large errors, while the averaged model returns very accurate results, especially when evaluated with the trapezoidal and 1/3 Simpson rules.

2 Methods

2.1 Evaluation of Markov models

A discrete Markov model consists of a (usually finite) set of states, together with a probability distribution $P_0(s)$ for the states at time 0 and a transition function that describes the dynamics of the system and allows to compute the distribution $P_t(s)$ for any point in time after 0—see Eq. 3 below. This is sometimes called *state membership function* because it indicates the probability that the system (the patient) is in state s at time t. Markov models used for economic evaluation also contain functions that specify the cost and the effectiveness for every state at each moment, $c(s,t)$ and $e(s,t)$ respectively, and discount functions for them. In this paper we focus on how to calculate the costs; the calculation of effectiveness is analogous.

Let $c(s,t)$ be the cost function and $\gamma(t)$ the corresponding discount function; necessarily, $\gamma(0) = 1$. The instantaneous discounted cost is

$$c(t) = \sum_{s} P_t(s) \cdot c(s,t) \cdot \gamma(t) \tag{1}$$

and the total cost is

$$C = \int_{t_0}^{t_f} c(t) \cdot dt \; . \tag{2}$$

In most models the cost function $c(t)$ decreases monotonically because the cohort that enters the model requires fewer and fewer resources as patients progressively die.

2.1.1 Evaluation of discrete-time Markov models

Discrete-time Markov models divide the time into a finite number of intervals of the same length, τ, called *cycles*. When the model is evaluated for a limited number of cycles, h (the horizon), the time points that delimit the intervals are $\{0, \tau, 2\tau, \ldots, h\tau\}$. The probability of being in state s at time $i\tau$ is

$$P_i(s) = \sum_{s'} P_i(s \mid s') \, P_{i-1}(s') \; . \tag{3}$$

where $P_i(s \mid s')$ is the *transition matrix*. When the transition matrix is time-independent, we can drop the subindex and just write $P(s \mid s')$. (In order to simplify the notation, we have written $P_i(s)$ instead of $P_{i\tau}(s)$.) This way we obtain a set of probability distributions, $\{P_0(s), P_1(s), \ldots, P_h(s)\}$. Applying Equation 1 we can obtain the cost at the points that separate the cycles, $\{c(0), c(\tau), \ldots, c(h\tau)\}$.

When the probability function $P_t(s)$ and the cost function $c(s,t)$ vary smoothly, so does the instantaneous discounted cost, $c(t)$. Then the trapezoidal rule approximation says that the cost accrued in the i-th cycle is

$$\int_{i\tau}^{(i+1)\tau} c(t) \cdot dt \approx \frac{c(i\tau) + c((i+1)\tau)}{2} \cdot \tau \;, \tag{4}$$

and consequently the cost

$$C_{TR} = \frac{\tau}{2} \cdot c(0) + \tau \sum_{i=1}^{h-1} c(i\tau) + \frac{\tau}{2} \cdot c(h\tau) \;. \tag{5}$$

is a good approximation of the total cost, C.

2.1.2 Evaluation of discrete-time Markov models with discontinuities

In mathematics, a function $c(t)$ is continuous at a point t' when

$$\lim_{t \to t'^{+}} c(t) = \lim_{t \to t'^{-}} c(t) = c(t') \;; \tag{6}$$

otherwise, there is a discontinuity at t'.

When $c(s,t)$ is discontinuous at the boundaries of the i-th interval and continuous inside, Equation 4 should be replaced with

$$\int_{i\tau}^{(i+1)\tau} c(t) \cdot dt \approx \tau \cdot \left(\lim_{t \to i\tau^{+}} c(t) + \lim_{t \to (i+1)\tau^{-}} c(t) \right) \;. \tag{7}$$

If we define

$$c^{*}(s,t) = \frac{1}{2} \left(\lim_{t \to i\tau^{-}} c(s,t) + \lim_{t \to i\tau^{+}} c(s,t) \right) \;, \tag{8}$$

and use this value instead of $c(s,t)$ when applying Equation 1 to compute the instantaneous cost, then Equation 5 is still a good approximation of the total cost. This means that if the cost function $c(s,t)$ is discontinuous at $i\tau$ for a state s, we can apply the trapezoidal rule provided that in the model we set the cost for $c(s, i\tau)$ to the average of the left and right limits of $c(s, t)$. This is the "correction" we propose for the evaluation of Markov models with discontinuities.

2.2 Empirical evaluation

2.2.1 Our gold-standard

In order to compare several methods for the evaluation of Markov models, other authors have used synthetic examples (Soares and Canto e Castro, 2012; Naimark et al., 2013; Barendregt, 2014; Elbasha and Chhatwal, 2016a,b). In this paper we use a slightly modified version of a real-world model, built by Chancellor et al. (1997) to determine the incremental cost-effectiveness ratio (ICER) of two interventions for HIV: monotherapy, which only applies zidovudine, and combination therapy, which adds lamivudine for two years, until it becomes ineffective for clinical reasons. The model has a cycle length of 1 year and was evaluated for a horizon of 20 years. It is now obsolete for clinical practice because there are more effective treatments for HIV, but it is still useful for pedagogic purposes. In particular, this model is studied as an example in the book of Briggs et al. (2006). An Excel version of the original model is available at `www.herc.ox.ac.uk/downloads/decision-modelling-for-health-economic-evaluation`. We have reimplemented it as a Markov influence diagram (MID) (Díez et al., 2017)—see Fig. 1—using OpenMarkov, an open source tool developed by our group (Arias et al., 2017). The model is available at `www.probmodelxml.org/networks`.

The inaccuracy introduced by the discretization of time into cycles increases with the cycle length: when it approaches 0, all the methods studied in this paper converge to the same results as a continuous-time model (Soares and Canto e Castro, 2012; Naimark et al., 2008, 2013; Barendregt, 2014; Chhatwal et al., 2016; Elbasha and Chhatwal, 2016b). This should be our gold standard for comparing the accuracy of different methods. However, it

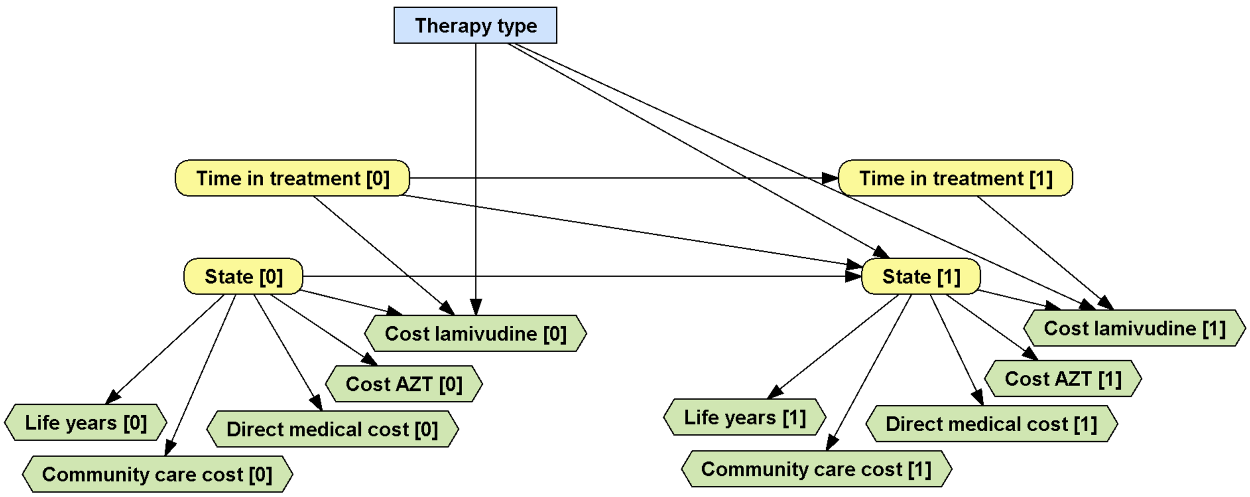

Figure 1: A Markov influence diagram for the HIV model of Chancellor et al. (1997). The rectangular node represents the decision of applying either monotherapy or combined therapy. Rounded rectangles represent variables that are not under the direct control of the decision maker. Hexagons represent costs and effectiveness; their value only depends on the state of the patient, except for the cost of lamivudine, which also depends on the choice of therapy and the time in treatment. This node induces a discontinuity because lamivudine is withdrawn at the end of the second year.

is impossible to adjust the transition *rates* so that they lead to exactly the same transition *probabilities* as in the Markov model (Chhatwal et al., 2016). So we used as gold standard a Markov model with a cycle length of 1 day leading to annual transition probabilities very close to the original ones, and then used the model with these new annual probabilities to compare the different "corrections". Figure 2 shows that the cost function, calculated with the daily-transitions model, has a discontinuity at $t = 2$, when the patients in the combination therapy arm stop receiving lamivudine.

2.2.2 Comparison of different approximations

As mentioned in the introduction, we have applied three "classical" approaches. Two of them are based on the assumption that transitions occur at the beginning or at the end of each cycle. The third is the life-table method (Barendregt, 2009, 2014), which first averages the occupancy probabilities at the boundaries of the cycle and then calculates the discounted costs. Another classical approach is the HCC, which cannot be applied to this model because

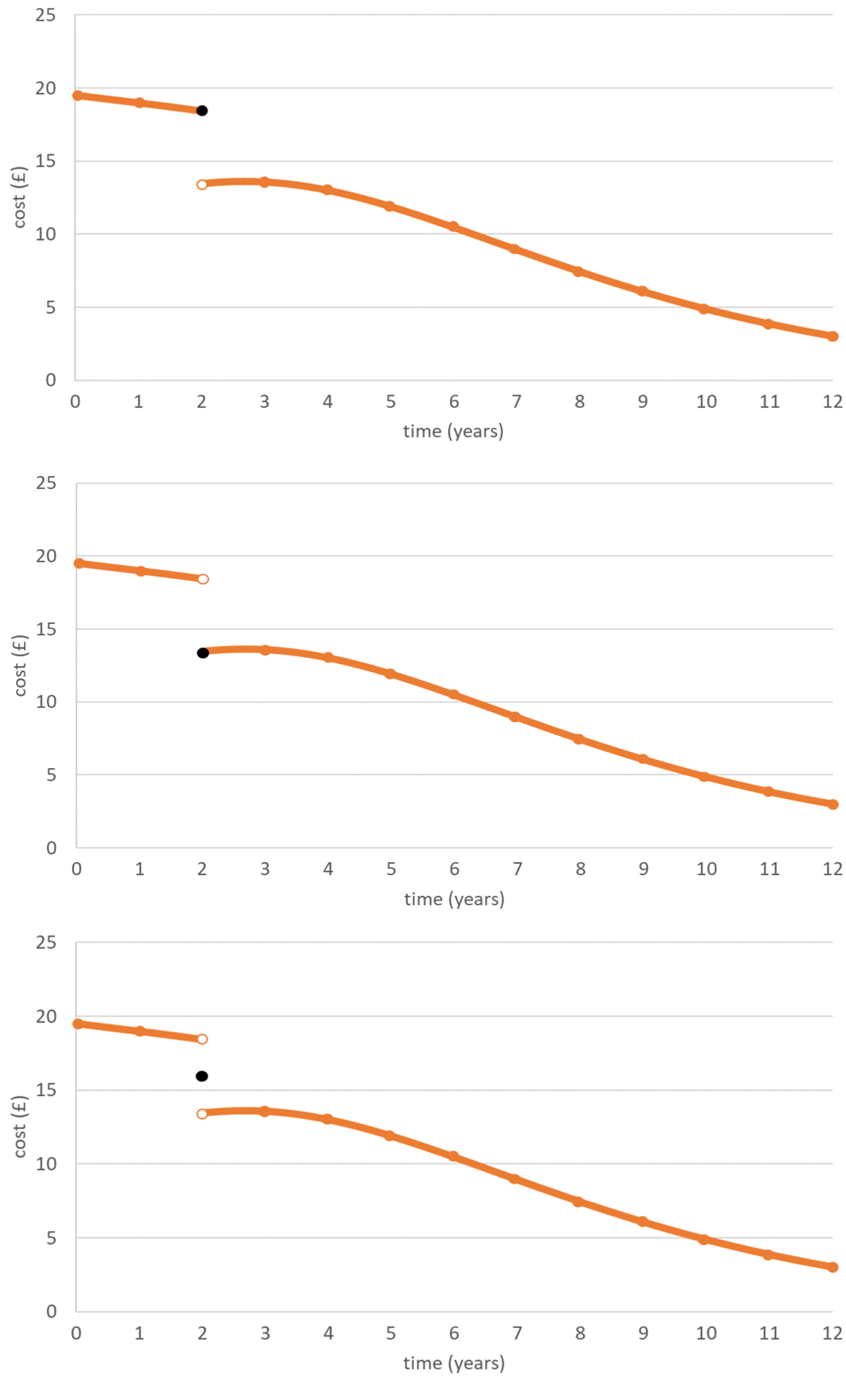

Figure 2: Instantaneous cost for combination therapy obtained from our gold standard, i.e., the model with a cycle length of one day. It has a sharp discontinuity at $t = 2$. The upper plot results from assuming that combination therapy is applied throughout the interval [0, 2 years), and the plot in the middle throughout the interval (0, 2 years]. The lower plot is obtained from the model in which the costs at $t = 2$ are the average of the limits.

it assumes that the cost function for each state is constant. These methods are explained in more detail in the Appendix.

We have also applied three of the numerical-integration methods proposed in (Elbasha and Chhatwal, 2016b): the trapezoidal rule and two Simpson rules, called 1/3 and 3/8. (We did not study the Riemann-sums rules because they are equivalent to the "classical" techniques that assume that transitions occur at the beginning or the end of each cycle.) These methods are based on the instantaneous cost values, $\{c(0), c(\tau), \ldots, c(h\tau)\}$, which correspond to the dots in the three plots in Figure 2. Unlike the classical approaches, numeric integration is sensitive to the value of the cost function at $t = 2$, and for this reason we have examined the three cases shown in Figure 2.

3 Results

Table 1 summarizes the results obtained with each method. As expected, the approaches that assume that transitions occur at the beginning or the end of each cycle give higher errors than the life-table method.

We also observe that numerical integration is very sensitive to the value of the cost functions at the point of discontinuity because these methods try to estimate the value of the cost, $c(t)$, around the point $t = 2$ by "propagating" it towards its left and its right. Thus, taking the left limit (as in the upper plot in Figure 2) overestimates the cost of combination therapy and, in turn, increases the ICER by 13% with respect to the gold standard. Reciprocally, taking the right limit underestimates the cost and reduces the ICER by a similar amount. In contrast, the estimates obtained from the averaged model (lower plot in Figure 2) are much more accurate. The first of them, based on the trapezoidal rule, is justified by the analysis in Section 2.1.2. The second and the third are based on the two Simpson rules. Even though we have no mathematical justification for them, in this particular example the 1/3 rule is more accurate than the trapezoidal rule. However, the

Method	ICER (£/QALY)	Percentage error (%)
Gold standard (cycle length = 1 day)	6,543	
Classical approaches		
transitions at the beginning of cycle	6,680	2.09
transition at end of cycle	6,401	−2.17
life tables	6,539	0.06
Numerical integration		
– **Left limit at** $t = 2$ **years**		
trapezoidal	7,440	13.71
1/3 Simpson	7,146	9.22
3/8 Simpson	7,603	16.20
– **Right limit at** $t = 2$ **years**		
trapezoidal	5,640	−13.80
1/3 Simpson	5,945	−9.14
3/8 Simpson	5,579	−14.73
– **Average of limits at** $t = 2$ **years**		
trapezoidal	6,540	−0.05
1/3 Simpson	6,545	0.03
3/8 Simpson	6,590	0.72

Table 1: Impact of different within-cycle correction methods in the ICER and the percentage error with respect to gold standard. (ICER: incremental cost-effectiveness ratio. QALY: quality-adjusted life year. HCC: half-cycle correction.)

estimate made by the 3/8 rule is much worse that the others. This result might be surprising at first sight because this rule is in general more accurate than the others. The explanation is that in the presence of a discontinuity the attempt to approach the non-linearity of the function as faithfully as possible leads to an "overadjustment" that results in the opposite effect.

With respect to computational efficiency, the evaluation of the gold standard model, with daily transitions, required 22.64 hours, while the classical approaches and the numerical-integration methods applied to a model with yearly transitions only took one or two seconds (1.40 s on average). Each evaluation has been made using the algorithm described in Díez et al. (2017).

4 Discussion

The usual way of presenting discrete-time Markov models states that transition can only occur at the boundaries between cycles, i.e., at $t = 0, \tau, 2\tau, 3\tau \ldots$ When transitions are allowed at $t = 0$, most authors say that "transitions occur at the beginning of the cycle"; when they are not allowed, it is said that "transitions occur at the end". Both assumptions lead to inaccuracies when computing accumulative outcomes, such as cost and effectiveness.

HCC was proposed by Sonnenberg and Beck (1993) as a method for obtaining more accurate estimates. Naimark et al. (2008) offered two analytical justifications of HCC with didactic purposes. Their idea was criticized as a "kludge" in a paper entitled "The half-cycle correction: banish rather than explain it" (Barendregt, 2009), which argued that HCC should be replaced with a more accurate technique, the *life-table method*. A few years later Naimark et al. (2013) proposed several modifications aimed at "redeeming the kludge", but again Barendregt (2014) criticized their work severely, to the point that Naimark et al. (2014) surrendered and accepted that "the standard approach to the HCC is flawed and should be abandoned".

In our opinion, the problem was that in (Sonnenberg and Beck, 1993) and (Naimark et al., 2008) the HCC was not justified as the application of the trapezoidal rule to the instantaneous discounted cost but as its application to the state-occupancy probabilities in order to subsequently calculate the cost and effectiveness accrued in each cycle. That was the source of several mathematical inconsistencies and made the HCC impossible to apply when the cost function is discontinuous (Barendregt, 2009, 2014). Another problem of the traditional way of presenting the HCC is the assumption that transitions can only occur at the boundary between cycles—an idea repeated again and again in the literature. However, Elbasha and Chhatwal (2016a,b) showed that (in the absence of discontinuities) the trapezoidal rule returns exactly the same result as the HCC—see Sec. A.2 in the Appendix—but the interpretation is different, because that rule does not assume that transitions only occur at the boundaries between cycles or halfway through each cycle.

Elbasha and Chhatwal also argued that more sophisticated numerical-integration techniques, such as Simpson 1/3 and 3/8 rules, generally give more accurate results than the trapezoidal rule when the function of interest is not linear inside each cycle. Unfortunately they did not take into account one of Barendregt's criticism of the HCC: its inaccuracy when the model has discontinuities, a problem which also affects numerical-integration approaches. We faced it when evaluating the HIV model: the cost function has a severe discontinuity at the end of the second year, when lamivudine becomes clinically ineffective and is withdrawn.[1] Common sense says that it does not matter whether it is withdrawn just one second before $t = 2$ or one second later, so a model in which lamivudine is applied in the interval $[0, 2)$ should give the same results as if it is applied in the interval $[0, 2]$. A continuous-time model would be insensitive to this modeling decision, but numerical integration is very sensitive to the cost at $t = 2$. In our example, the difference between withdrawing lamivudine just before or after $t = 2$ is higher than 27%; the error with respect to the gold standard is $\pm 13\%$. If the ICER estimated is close to the willingness-to-pay threshold, this error may lead to making a wrong decision.

Table 1 also shows that the errors were reduced when applying numerical-integration techniques on a model in which the cost at $t = 2$ is the average between administering lamivudine and not administering it. The application of the trapezoidal rule to this model is justified by the algebraic analysis in Section 2.1.2 and results in an error of only -0.05% in the ICER. The 1/3 Simpson rule also gives a small error, 0.03%. These are smaller than the 0.06% of the life-table method and much smaller than the 0.72% of the 3/8 Simpson rule.

A limitation of our study is that we have only studied one model. Further studies are necessary in order to determine to what extent the qualitative results obtained generalize to other models. However, our analysis serves at least as a warning that in the case of discontinuities numerical-integration techniques may give wrong results because they were

[1] The cost function $c(s, t)$ is discontinuous only for the three states in which the patient is alive. When the patient is dead (fourth state) there is no discontinuity.

designed for continuous functions.

Most of the issues arising from discretization into cycles can be mitigated (if we have data) by choosing an appropriate cycle length (Soares and Canto e Castro, 2012). Likewise, most problems arising from discontinuities could be addressed by choosing the appropriate within-cycle correction—based on the domain knowledge.

Finally, we should mention that some costs are incurred continuously over time while others occur at specific points in time. For example, in the case of cochlear implants, there is the upfront cost of the device and the surgery (Pérez-Martín et al., 2017). These costs should be accounted for separately, excluding them in any within-cycle correction, as explained by Elbasha and Chhatwal (2016a).

5 Conclusion

In many cases it is not possible to build a continuous-time Markov model or to shorten the cycle length of a given discrete-time model (Chhatwal et al., 2016), but so-called "within-cycle corrections" may give very good approximations (Soares and Canto e Castro, 2012; Elbasha and Chhatwal, 2016a,b). However some models have discontinuities in costs due to the withdrawal of expensive therapies. (There might also be discontinuities in the effectiveness, but we have not found any example.) In this case, the direct application of those "corrections" may lead to significant errors. We have proved mathematically that the trapezoidal rule—formally equivalent to the half-cycle correction but with a different interpretation—yields a good approximation also when the cost function has discontinuities at the boundaries between cycles but is constant or varies smoothly within each cycle, provided that it is applied on a model in which the value of the instantaneous cost function at each point of discontinuity is set to the average of the left and right limits, as dictated by Equation 8. In the real-world model we have studied, the $1/3$ Simpson rule also gave a good result, even though we did not have mathematical justification for it, but the $3/8$ Simpson rule, which is more accurate for models without discontinuities, introduced significantly larger errors. As a conclusion, when a model has discontinuities we recommend building an averaged model and applying the trapezoidal rule instead of more sophisticated numerical-integration techniques.

6 Acknowledgments

This work has been supported by grant TIN2016-77206-R of the Spanish Government, co-financed by the European Regional Development Fund. JP received a predoctoral grant from the Spanish Ministry of Education, as well as travel support for his stay at the University of Sheffield. We thank the reviewers of the Medical Decision Making journal for their comments and corrections.

A Appendix: A comparison of classical methods

In recent years we have witnessed an interesting debate: HCC vs. the life-table method (LT). Even though the controversy seems to be closed after the insightful analyses of Elbasha and Chhatwal (2016a,b), we hope that this appendix can still shed some light on the issue. Here "classical methods" are those not based on numerical integration.

A.1 Absence of within-cycle transitions

The usual way of presenting discrete-time Markov models states that transition can only occur at the boundaries between cycles, i.e., at $t = 0, \tau, 2\tau, 3\tau \ldots$ It implies that the state of the system does not change within a cycle. When transitions are not allowed at $t = 0$, most authors say that "transitions occur at the beginning of each cycle". In this case the probability in the interval $[i\tau, (i+1)\tau)$ is $P_i(s)$. With these assumptions and approximations, the total cost is

$$C_E = \tau \sum_{i=0}^{h-1} \sum_{s} P_i(s) \cdot c_i(s) \cdot \gamma(i\tau) \ , \tag{9}$$

where the subindex E stands for "end". When the instantaneous cost function (cf. Eq. 1) decreases monotonically, this assumption leads to an overestimation of the total cost, i.e.,C_E is higher than the true cost, C, given by Equation 2.

When transitions at $t = 0$ are allowed, it is said that "transitions occur at the beginning of the cycle". Therefore the probability in the interval $(i\tau, (i+1)\tau]$ is $P_{i+1}(s)$ and the total cost is

$$C_B = \tau \sum_{i=0}^{h-1} \sum_{s} P_{i+1}(s) \cdot c_i(s) \cdot \gamma(i\tau) \ , \tag{10}$$

where the subindex B stands for "beginning". When the cost function decreases monotonically, this assumption leads to an underestimation of the total cost, so that we have $C_B < C < C_E$.

A.2 The half-cycle correction

In an attempt to obtain a better approximation, Sonnenberg and Beck (1993) introduced the half-cycle correction, which consists in including a cycle of length $\tau/2$ at the beginning of the process, in which no transition has yet occurred, and another cycle of the same length at the end, so that the total duration of the process is $h\tau$. This is equivalent to assuming that the transitions occur at the time points $\{0.5\tau, 1.5\tau, 2.5\tau, \ldots\}$, i.e., halfway through each cycle (Sonnenberg and Beck, 1993; Naimark et al., 2008). The approach is justified in the literature as a method for approximating the *state-occupancy probabilities*—the vertical axis in Figure 10 in (Sonnenberg and Beck, 1993), reproduced as Figure 3 in (Naimark et al., 2008), clearly shows it. Those figures also show that the HCC still assumes that transitions only occur at certain points in time, which makes the occupancy probability constant between consecutive points.

The approximate state-occupancy probabilities are multiplied by the cost of each state to compute the cost accrued in each interval. This calculation makes sense when the cost is constant, but encounters a problem when there is an abrupt discontinuity at the boundary between two intervals. In the HIV example there is a discontinuity at $t = 2\tau$, so it is not clear which costs this method should use for the second interval, which extends from 1.5τ to 2.5τ .

In practice, HCC is implemented using (the equivalent of) Equation 5, whose first term on the right-hand side can be interpreted as the cost accrued in a cycle of length $\tau/2$ in which no transition has yet occurred. However, that equation stems from applying the trapezoidal rule the *instantaneous costs*, not to the occupancy probabilities. This inconsistency was severely criticized by Barendregt (2009, 2014), who proposed the life-table method as a better alternative.

A.3 The life-table method

This method, being consistent with the arguments put forward by the advocates of HCC, first estimates the average state-occupancy probabilities inside each cycle by averaging the probabilities at its boundaries and then calculates the costs:

$$\int_{i\tau}^{(i+1)\tau} c(t) \cdot dt \approx \sum_s \frac{P_i(s) + P_{i+1}(s)}{2} \cdot c_i(s) \cdot \gamma(i\tau) \cdot \tau \ , \tag{11}$$

where $c_i(s)$ is the cost inside the i-th interval, again assumed to be constant. This implies that

$$C_{LT} = \sum_{i=0}^{h-1} \sum_s \frac{P_i(s) + P_{i+1}(s)}{2} \cdot c_i(s) \cdot \gamma(i\tau) \cdot \tau \ , \tag{12}$$

This is called the *life-table method* because it is based on the procedure that demographers use it to estimate life expectancy (Barendregt, 2009).

It is easy to check that

$$C_{LT} = \frac{C_E + C_B}{2} \ .$$

We have already seen that when the total-cost function decreases monotonically, C_E is an overestimation of the true cost and C_B is an underestimation, so their arithmetic mean, C_{LT}, is expected to be closer to the true value than if we assumed that all transitions occur either at the beginning or at the end of each cycle.

Please note that this method is insensitive to discontinuities that occur at the boundary between intervals, which explains the small error it returned for the HIV model—see Table 1.

A.4 A comparison of HCC and LT

The justification of HCC implicitly assumes that the instantaneous cost for each state is time independent; this allows us to write $c(s)$ instead of $c(s,t)$. Under this assumption, HCC computes exactly the same value as the trapezoidal rule, C_{TR}. Equation 5 stems from Equation 4, which can be rewritten as

$$\int_{i\tau}^{(i+1)\tau} c(t) \cdot dt \approx \sum_s \frac{P_i(s) \cdot \gamma(i\tau) + P_{i+1}(s) \cdot \gamma((i+1)\tau)}{2} \cdot c(s) \cdot \tau \ . \tag{13}$$

Comparing this expression with Equation 11, we can see that they only differ in the way of applying the discounts.[2] When the cycle length τ is short or the discount function decreases slowly, then $\gamma(i\tau) \approx \gamma((i+1)\tau)$ and $C_{LT} \approx C_{TR}$. The main difference is that Equation 13, implicitly used by the trapezoidal rule and HCC, applies the correct discount at each boundary of the cycle, while LT applies the same discount, $\gamma(i\tau)$, from the beginning to the end of the cycle.

Therefore we disagree with *some* of the arguments claiming the superiority of LT over HCC.[3] For example, Barendregt (2009) argued that the standard HCC method is incompatible with discounting. Naimark et al. (2013), who were initially strong advocates of the

[2] Therefore, the assertion of Elbasha and Chhatwal (2016b) that the trapezoidal rule gives the same result as LT is true only when there is no discount. In fact, Section 9.2.1 in (Gray et al., 2011) shows an example in which the two methods yield different numerical results.

[3] In these remarks we agree with Elbasha and Chhatwal (2016a,b), except for minor details that do not deserve a discussion here.

HCC, finally agreed with him. However, the computation of C_{TR} clearly uses the discounted costs (cf. Eqs. 1 and 5).

Barendregt (2009) also said: "I know of very few relevant Markov models in medical decision making where QALY weights and unit costs are constant across all cycles". However, HCC does not require that the cost function $c(t)$, given by Equation 1, be constant. In our analysis, the derivation of Equation 5 only required that $c(s,t)$ be constant for every state s, as it is in many models—the HIV examined in this paper is rather an exception.

In turn, Naimark et al. (2013) said that "the standard approach to the HCC assumes that the state membership curve is declining and monotonic." However, when Equation 5 is justified as the application of the trapezoidal rule, it does not require that assumption.

In summary, we claim that there was nothing wrong in the application of the HCC in the absence of discontinuities. On the contrary, C_{TR}, given by Equation 5, is generally more accurate than C_{LT} , given by 12. The problem was in the way of explaining and justifying the method.

References

Arias, M., Artaso, M. A., Bermejo, I., Díez, F. J., Luque, M., and Pérez-Martín, J. (2017). Advanced algorithms for medical decision analysis. Implementation in OpenMarkov. In *Proceedings of the 16th Conference on Artificial Intelligence in Medicine (AIME 2017)*, Vienna, Austria.

Barendregt, J. J. (2009). The half-cycle correction: banish rather than explain it. *Medical Decision Making*, 33:961–970.

Barendregt, J. J. (2014). The life table method of half-cycle correction: getting it right. *Medical Decision Making*, 29:500–502.

Beck, J. R. and Pauker, S. G. (1983). The Markov process in medical prognosis. *Medical Decision Making*, 3:419–458.

Briggs, A., Claxton, K., and Sculpher, M. (2006). *Decision Modelling for Health Economic Evaluation*. Oxford University Press, New York.

Chancellor, J. V., Hill, A. M., Sabin, C. A., Simpson, K. N., and Youle, M. (1997). Modelling the cost effectiveness of lamivudine/zidovudine combination therapy in HIV infection. *Pharmacoeconomics*, 12:54–66.

Chhatwal, J., Jayasuriya, S., and Elbasha, E. H. (2016). Changing cycle lengths in state-transition models: Challenges and solutions. *Medical Decision Making*, 36:952–964.

Díez, F. J., Yebra, M., Bermejo, I., Palacios-Alonso, M. A., Arias, M., Luque, M., and Pérez-Martín, J. (2017). Markov influence diagrams: A graphical tool for cost-effectiveness analysis. *Medical Decision Making*, 37:183–195.

Elbasha, E. H. and Chhatwal, J. (2016a). Myths and misconceptions of within-cycle correction: a guide for modelers and decision makers. *Pharmacoeconomics*, 34:13–22.

Elbasha, E. H. and Chhatwal, J. (2016b). Theoretical foundations and practical applications of within-cycle correction methods. *Medical Decision Making*, 36:115–131.

Gray, A. M., Clarke, P. M., Wolstenholme, J., and Wordsworth, S. (2011). *Applied Methods of Cost-effectiveness Analysis in Healthcare*. Oxford University Press, New York.

Naimark, D. M. J., Bott, M., and Krahn, M. (2008). The half-cycle correction explained: two alternative pedagogical approaches. *Medical Decision Making*, 28:706–712.

Naimark, D. M. J., Kabboul, N. N., and Krahn, M. D. (2013). The half-cycle correction revisited: redemption of a kludge. *Medical Decision Making*, 33:961–970.

Naimark, D. M. J., Kabboul, N. N., and Krahn, M. D. (2014). Response to "The life table method of half-cycle correction: getting it right". *Medical Decision Making*, 34:283–285.

Pérez-Martín, J., Artaso, M. A., and Díez, F. J. (2017). Cost-effectiveness of pediatric bilateral cochlear implantation in Spain. *The Laryngoscope*.

Siebert, U., Alagoz, O., Bayoumi, A. M., Jahn, B., Owens, D. K., Cohen, D. J., and Kuntz, K. M. (2012). State-transition modeling: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. *Medical Decision Making*, 32(5):690–700.

Soares, M. O. and Canto e Castro, L. (2012). Continuous time simulation and discretized models for cost-effectiveness analysis. *Pharmacoeconomics*, 30:1101–1117.

Sonnenberg, F. A. and Beck, J. R. (1993). Markov models in medical decision making: A practical guide. *Medical Decision Making*, 13:322–339.