

# Medical Decision Making

<http://mdm.sagepub.com/>

---

## **Bias Associated with Failing to Incorporate Dependence on Event History in Markov Models**

Tanya G. K. Bentley, Karen M. Kuntz and Jeanne S. Ringel  
*Med Decis Making* 2010 30: 651 originally published online 16 April 2010  
DOI: 10.1177/0272989X10363480

The online version of this article can be found at:

<http://mdm.sagepub.com/content/30/6/651>

---

Published by:



<http://www.sagepublications.com>

On behalf of:



<http://www.smdm.org>

**Additional services and information for *Medical Decision Making* can be found at:**

**Email Alerts:** <http://mdm.sagepub.com/cgi/alerts>

**Subscriptions:** <http://mdm.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://mdm.sagepub.com/content/30/6/651.refs.html>

# Bias Associated with Failing to Incorporate Dependence on Event History in Markov Models

Tanya G. K. Bentley, PhD, Karen M. Kuntz, ScD, Jeanne S. Ringel, PhD

**Purpose.** When using state-transition Markov models to simulate risk of recurrent events over time, incorporating dependence on higher numbers of prior episodes can increase model complexity, yet failing to capture this event history may bias model outcomes. This analysis assessed the tradeoffs between model bias and complexity when evaluating risks of recurrent events in Markov models. **Methods.** The authors developed a generic episode/relapse Markov cohort model, defining bias as the percentage change in events prevented with 2 hypothetical interventions (prevention and treatment) when incorporating 0 to 9 prior episodes in relapse risk versus a model with 10 such episodes. Magnitude and sign of bias were evaluated as a function of event and recovery risks, disease-specific mortality, and risk function. **Results.** Bias was positive in the base case for a prevention strategy, indicating that failing to fully incorporate dependence on event history overestimated the prevention's predicted impact. For treatment, the bias was

negative, indicating an underestimated benefit. Bias approached zero as the number of tracked prior episodes increased, and the average bias over 10 tracked episodes was greater with the exponential compared with linear functions of relapse risk and with treatment compared with prevention strategies. With linear and exponential risk functions, absolute bias reached 33% and 78%, respectively, in prevention and 52% and 85% in treatment. **Conclusion.** Failing to incorporate dependence on prior event history in subsequent relapse risk in Markov models can greatly affect model outcomes, overestimating the impact of prevention and treatment strategies by up to 85% and underestimating the impact in some treatment models by up to 20%. When at least 4 prior episodes are incorporated, bias does not exceed 26% in prevention or 11% in treatment. **Key words:** economic analysis; cost-effectiveness analysis; decision analysis; Markov models; outcomes research; priority setting for spending. (*Med Decis Making* 2010;30:651-660)

Markov models, or state-transition models, can be used to simulate the risk of recurrent events or episodes over time for a hypothetical population. The risks of recurrent events often depend on an individual's history of prior events. For example, in

predicting health and economic outcomes associated with illicit drug abuse, the probabilities of use may change over time within and between individuals as people initiate use, recover, and potentially experience recurrent episodes. Similarly, past history of other mental health disorders such as depression may affect subsequent relapse risk.

Markov models can incorporate dependence on event history to a certain degree by adding health states that track event history and make subsequent event risk dependent on this history. As an example, one can consider a model of illicit drug use, recovery, and recurrence, which allows individuals to recover from a drug-use event and to relapse. A simpler model that includes only these 3 states (or 4 including death) would capture the lifetime patterns of recovery and relapse without allowing the probabilities of transitioning from one state to another to change as a function of prior history. On the other hand, creating a slightly more complex model by

Received 2 June 2009 from the RAND Corporation, Santa Monica, CA (TGKB, JSR); University of California, Los Angeles (TGKB); Partnership for Health Analytic Research, LLC, Beverly Hills, CA; and the Department of Health Policy and Management, University of Minnesota, Minneapolis (KMK). This article was presented at the 30th annual meeting of the Society for Medical Decision Making, October 2008, Philadelphia, Pennsylvania. This article describes work done with support from the National Institute on Drug Abuse (NIDA 1R01 DA019993). Revision accepted for publication 26 December 2009.

Address correspondence to Tanya G. K. Bentley, PhD, Partnership for Health Analytic Research, LLC, 280 South Beverly Drive, #404, Beverly Hills, CA 90212; telephone: (310) 858-9555; e-mail: tbentley@pharllc.com.

DOI: 10.1177/0272989X10363480

adding states to track the number of past drug-use episodes (e.g., episode-0, recovery-1, episode-1, recovery-2, etc.) would allow relapse risk to vary as a function of the number of tracked prior episodes.

However, incorporating dependence on higher numbers of prior episodes in Markov cohort models can result in increasing levels of model complexity. Although microsimulation (i.e., 1st-order Monte Carlo models) can more readily capture such dependence on event history, this often comes at the expense of the need for more detailed programming, difficulty in debugging, and extensive computing needs. In addition, although one may know that relapse risk for an event such as illicit drug use increases with prior history, one may not adequately model the dependence on such history in a cohort model. To the extent that the predictions of such health and economic outcomes depend on the degree to which this event history is captured, model output may be biased. This bias may differentially affect different types of strategies evaluated in a model. For example, results from a model that compares prevention with treatment programs may be inaccurately interpreted if results from one intervention are more biased—or biased in a different direction—than those from another type of intervention.

Given that the risk of subsequent event can be different up to only a limited number of prior episodes in a Markov cohort model, the aim of this article is to quantify the degree and direction of potential bias under varying model conditions and assumptions. We evaluate 2 types of interventions (a prevention strategy to decrease event/relapse risk and a treatment to increase probability of recovery), comparing outcomes in models with varying degrees of dependence on event history.

## METHODS

We used a generic episode/relapse Markov cohort model to evaluate the potential bias associated with inadequately capturing dependence on event history for subsequent risk of relapse (Figure 1). We quantified the bias when the risk of relapse was a function of 0 to 10 tracked prior event episodes, using 10 as the gold standard. This number was chosen because a model that allows risk of relapse to be different for more than 10 tracked prior episodes would be less likely to reflect available data. Likewise, because of the low proportion of individuals who would reach more than 10 episodes, any risk change caused by incorporating more tracked episodes would not be

likely to significantly affect results. The goal of the analysis was thus to quantify the difference in cohort model results when the maximum number of tracked event episodes ( $N$ ) varied from 0 to 9, compared with those from a gold standard model in which  $N = 10$ .

We built a generic Markov cohort model of reoccurring event risk (Figure 1), which allows event/relapse risk to vary over time as a function of the number of tracked prior episodes up to  $N$ . The cohort begins in an event-free, “well” state. Each annual cycle, “well” individuals face a baseline age-specific probability of an initial event occurrence. They may subsequently recover from the episode and then may relapse in future cycles. The model incorporates  $N + 1$  unique event states ( $N = 0, 1, \dots, 10$ ), where  $N$  represents the maximum number of recurrent event episodes—or relapses—tracked and incorporated in the subsequent relapse risk function. The simplest version of the model is one in which  $N = 0$  (e.g., subsequent relapse risk does not depend on prior event history) that would have 4 health states: well, event, recovery, and dead. As  $N$  increases, the model becomes more complex, adding  $N$  event states and  $N - 1$  recovery states to allow the risk of the  $i$ th recurrent event from the  $i$ th recovery to increase with the number of prior events ( $i = 1, 2, \dots, N$ ). For all values of  $i$ , the model allows individuals to have an infinite number of events; however, after  $N$  recurrences, subsequent event risk remains at a constant level such that overall relapse risk over 10 episodes is the same at each age, regardless of  $N$ . In other words, for all  $i > N$ , the probability of  $i$ th relapse equals the probability of the  $(I + 1)$ th relapse, and so forth. The full gold standard model, as shown in Figure 1, thus has 23 states: 1 well state, 11 tracked-event states (1 initial and 10 relapses), 10 recovery states, and dead.

Individuals in any event state face a 30% probability of recovery. Persons can die either from event or from other causes, and the outcome of interest is cumulative time spent with episodic event over a lifetime, with versus without a hypothetical intervention. In the base case, the model starts with a population of 12-year-olds, chosen because risk of chronic, episodic events such as drug use tends to increase in earnest in the early teenage years.<sup>1–10</sup> We modeled the probability of dying from an event using the following function of age-specific, all-cause mortality ( $\lambda_{\text{age}}$ )—based on U.S. Vital Statistics data<sup>11</sup>—and event-specific mortality ( $\lambda_{\text{event}}$ ):

$$\Pr(\text{Die}) = 1 - \exp(-(\lambda_{\text{age}} + \lambda_{\text{event}})).$$

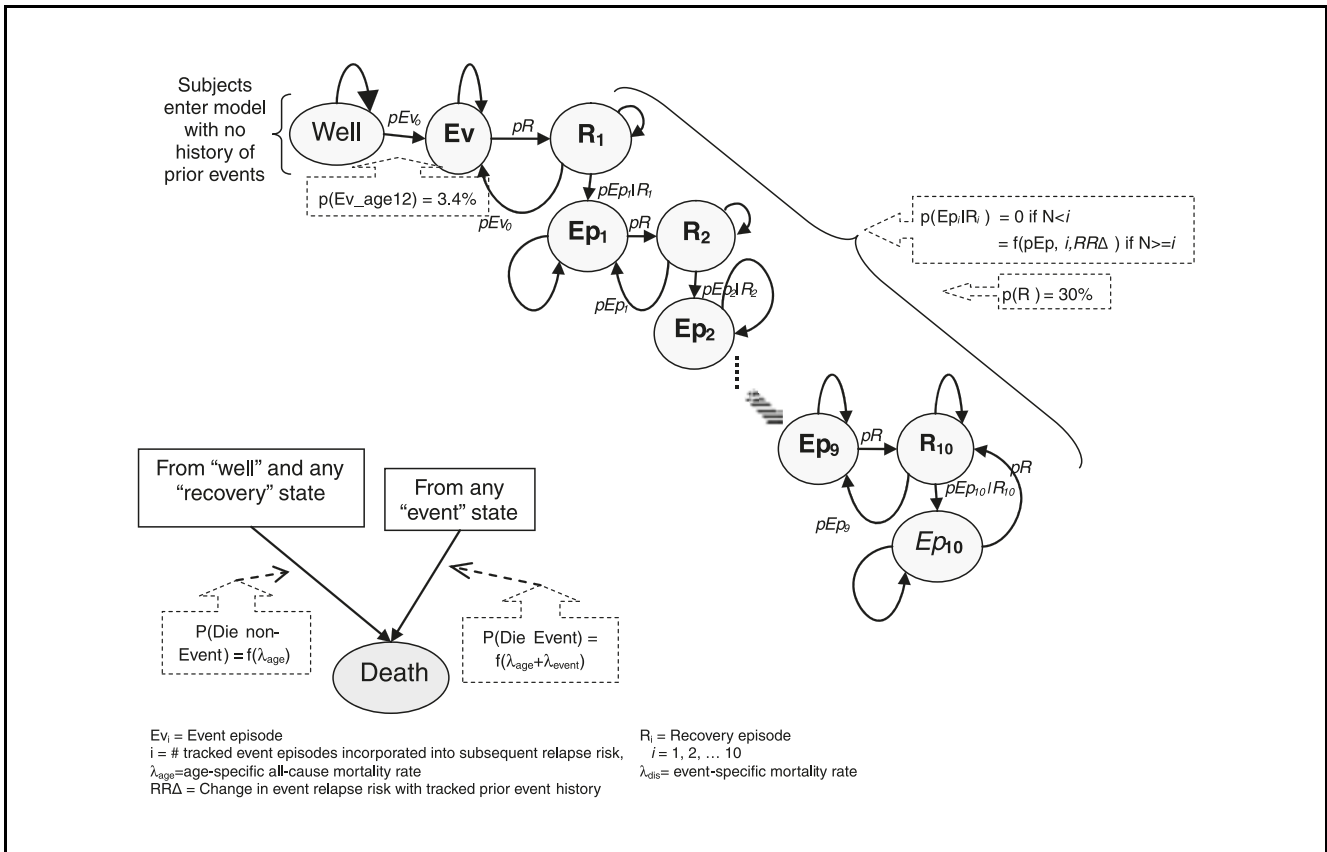


Figure 1 Schematic diagram of a generic Markov cohort model of event-history bias.

All models were run using Treeage Pro<sup>®</sup> software (Williamstown, MA). Model parameter values and ranges are shown in Table 1.

We assumed that relapse risk changes as a function of  $i$ —the number of tracked prior event episodes—and used 2 different functional forms:

Basecase: Linear:  $pRelapse_i = pEv + i^* x_{lin}, (i \leq N)$

Alternative function: Exponential:  $pRelapse_i = pEv^*(x_{exp})^i, (i \leq N),$

where  $pRelapse_i$  is the annual probability of experiencing the  $i$ th relapse,  $pEvent$  is the risk of initial event occurrence, and  $x_{lin}$  and  $x_{exp}$  are model parameters that characterize the change in relapse risk with number of prior tracked episodes. They were calculated to attain plausible average relapse risks (e.g., 8% for 12-year-olds) over 10 episodes, to reflect findings in the published literature for recurrent episodic conditions such as substance abuse and other mental health disorders.<sup>1-10</sup> Modeling the risk increase as

a linear function suggests that event risk increases by a fixed additive quantity with each tracked episode. Exponential growth yields a constant multiplicative increase with each tracked episode. Figure 2a shows the 2 risk functions given base-case and nonintervention parameter estimates of a 3.44% initial event risk and an 8% average relapse risk for 12-year-olds over 10 episodes. For both functions, relapse risk remains constant after individuals reach the maximum number of tracked episodes,  $N$  (see Figure 2b). Long-term risk therefore depends on  $N$ ; after  $N$  events, individuals may have additional relapses, but the risks for these episodes do not increase any further, remaining at a level calculated to keep overall risk constant over 10 episodes (tracked or untracked). We derived the constant relapse risk assuming a linear function of risk growth after  $N$  is reached. The base-case relapse risk for  $i > N$  is

$$\left[ 0.08 - \left( pEv + \frac{x_{lin} * N}{2} \right) \left( \frac{N + 1}{11} \right) \right] * \left( \frac{11}{10 - N} \right),$$

**Table 1** Model Parameters and Assumptions

Variable	Base-Case Estimate		Range	
Start age (y)	12		0–80	
Probability of recovery	30%		10%–45%	
Event-specific mortality rate ( $\lambda$ //I// event)	0.01		0–0.05	
Intervention				
Prevention (decrease in event/relapse risk)	10%		1%–50%	
Treatment (increase in probability of recovery)				
<b>Average event risk over 10 episodes, without intervention</b>	<b>Age</b>	<b>Estimate (%)</b>	<b>Lower (%)</b>	<b>Upper (%)</b>
	12–13	8.0	4.1	16.3
	14–18	10.3	5.8	19.2
	19–29	14.1	8.6	23.9
	30–44	7.8	3.9	16.0
	45–64	5.2	2.0	12.8
	65+	4.7	1.6	12.2
<b>Annual baseline initial event risk (<math>pEvent</math>) without intervention</b>				
	12–13	3.44	2.58	4.30
	14–18	5.73	4.30	7.16
	19–29	9.50	7.13	11.88
	30–44	3.20	2.40	4.00
	45–64	0.67	0.50	0.83
	65+	0.33	0.25	0.42

where  $x_{in}$  is as previously defined and  $pEv$  is the age-specific risk of initial event occurrence. We evaluated the impact of this assumption in sensitivity analyses by allowing relapse risk to increase to a different and lower constant level after individuals reach  $N$  tracked episodes (Figure 2c). In sensitivity analyses, relapse risk for  $i > N$  equals that of episode  $N$ :

$$pRelapse_{i>N} = pEp_N.$$

We modeled 2 types of interventions: prevention and treatment. Prevention decreases event risk by a percentage among individuals who are currently event free—including those in “well” or “recovery” states—and represents a general prevention strategy, such as a campaign to “Just Say No” to drug use. Treatment, on the other hand, was assumed to increase recovery rate by a percentage among individuals currently experiencing an event. The outcome of interest was cumulative time in event (CTE) that is saved over a lifetime, with or without each intervention. We evaluated the magnitude and sign of bias as a function of all model parameters listed in Table 1.

We define bias as the percentage change in CTE prevented over a lifetime with 2 hypothetical interventions, when models allow 0 to 9 event episodes

to influence subsequent relapse risk ( $N = 0, 1, \dots, 9$ ), as compared with a model in which  $N = 10$ :

$$\text{Bias} = \frac{CTESaved_N - CTESaved_{10}}{CTESaved_{10}} * 100\%,$$

where  $CTESaved$  is the difference in CTE between no intervention and intervention.

## RESULTS

Table 2 shows the results of our base-case analysis using the linear function of risk growth with number of tracked prior episodes, and Figure 3a shows the base-case bias results for both the linear and exponential functions. For a starting population of 12-year-olds with no history of prior events, age-specific risks of initial event and lifetime event/relapse (3.4% and 8%, respectively, for 12-year-olds), 1% event-specific mortality, and 30% chance of recovery, the model predicted that a prevention strategy that decreases event and relapse risk by 10% would reduce the cumulative time spent in event (e.g., using drugs) by on average just less than 1 y, whereas treatment that increases recovery by 10% would decrease cumulative event time by half a year (see Table 2). This benefit decreased with the number of tracked prior episodes, starting for  $N = 0$  (where results for

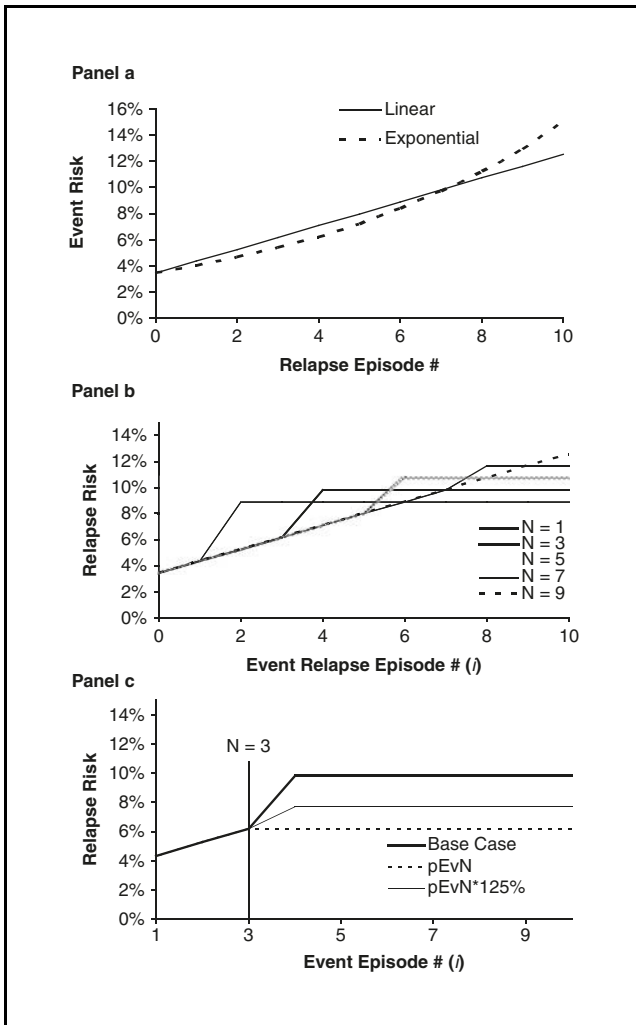


Figure 2 (a) Event risk as linear and exponential functions of relapse episode. (b) Base-case relapse risk function for 1, 3, 7, and 9 tracked prior event episodes. (c) Relapse risk for 3 tracked prior event episodes with 3 relapse risk functions.

linear and exponential risk functions will be the same) at 1.06 and 0.66 years for prevention and treatment strategies, respectively, and declining in prevention to 0.67 and 0.88 years (linear and exponential risk functions, respectively) and in treatment to 0.43 and 0.35 years. The magnitude of time in event saved was predicted to always be greater in prevention than treatment and with linear as compared with exponential risk functions. In prevention, the direction of bias associated with incorporating fewer than 10 event episodes in subsequent relapse risk was positive in the base case, indicating an overestimated intervention benefit (Figure 3a), whereas in

treatment, it was positive for all  $N \leq 2$  and was negative—indicating underestimated benefit—for  $N > 2$ . For both linear and exponential risk functions, the shape of the bias in prevention from  $N = 0$  to 10 was an inverted “U” with its mode at  $N = 1$  and in treatment was “U” shaped with its minimum at  $N = 3$ . The average bias over 10 tracked episodes was greater with the exponential than linear functions of relapse risk and with treatment than prevention strategies. With the linear and exponential risk functions in the base case, absolute bias reached 33% and 78%, respectively, in prevention and 52% and 85% in treatment.

We also evaluated the degree and sign of bias as a function of cohort start age and intervention effect, event/relapse risk, recovery risk, and event-specific mortality; the results of the latter 3 are shown in Figure 3b-d. In general, neither the sign nor shape of bias was sensitive to changes in these model parameters: Bias remained positive and overestimated benefits in all prevention models except when  $N = 0$  with high event risk and changed from positive to negative (underestimating benefits) with increasing values of  $N$  in treatment models. The magnitude of bias was differentially sensitive to varying these parameter estimates, depending on the function of relapse risk growth (linear v. exponential) and intervention type (prevention v. treatment).

For example, when the model was evaluated with double the average lifetime event and relapse risk, the effect of treatment was overestimated by more than 100% when zero prior tracked episodes were incorporated, but this bias (at  $N = 0$ ) for the linear function in prevention was  $-10\%$ —indicating a slightly underestimated effect—and remained at less than 12% even when positive (Figure 3b). When event/relapse risk was low, bias decreased with both interventions and for both the exponential and linear risk functions. The magnitude of bias was negatively associated with event-specific mortality and age and was positively associated with probability of recovery and intervention effect.

Figure 3e shows the results of varying our base-case assumption regarding estimates of relapse risk for event episodes beyond  $N$ . When we assumed that this risk remained constant at that of the last tracked prior episode ( $N$ ), the sign of the bias became negative in both prevention and treatment, and the magnitude declined dramatically, to less than 42% for any level of  $N$  and to less than 4% when at least 3 prior episodes were tracked. Overall, when considering any of these extreme model parameters, as long as at least 4 prior episodes were tracked, bias

**Table 2** Base Case<sup>a</sup> Predicted Events, Events Saved, and Event Time Saved in Prevention and Treatment Interventions by Number of Tracked Prior Episodes

Number of Tracked Episodes	CTE (y)			CTE Saved (y)		Bias (%)	
	No Intervention	Prevention	Treatment	Prevention	Treatment	Prevention	Treatment
0	10.09	9.03	9.43	1.06	0.66	20.2	51.8
1	8.02	6.84	7.52	1.18	0.50	33.4	15.8
2	6.92	5.87	6.48	1.05	0.44	19.0	1.8
3	6.50	5.55	6.07	0.95	0.43	7.8	-0.9
4	6.36	5.45	5.93	0.90	0.43	2.6	-0.7
5	6.31	5.43	5.88	0.89	0.43	0.8	-0.3
6	6.30	5.42	5.87	0.88	0.43	0.2	-0.1
7	6.30	5.42	5.87	0.88	0.43	0.1	-0.04
8	6.30	5.42	5.87	0.88	0.43	0.0	-0.01
9	6.30	5.42	5.87	0.88	0.43	0.0	-0.01
10	6.30	5.42	5.87	0.88	0.43	0.0	0.00
Average	6.88	5.93	6.42	0.95	0.46	7.6	6.1

Note: CTE = cumulative time in event; Ix =.

a. Assuming a starting population of 12-year-olds, a linear function of risk growth with number of tracked prior episodes, age-specific initial event risk and average risk over 10 episodes (3.44% and 8%, respectively, for 12-year-olds), a 30% probability of recovery, a 1% event-specific mortality, and a constant relapse risk for  $i > N$ .

never exceeded 26% in prevention and 11% in treatment.

## DISCUSSION

We assessed the tradeoffs between model bias and complexity when incorporating dependence on prior event history in subsequent event risk in a decision-analytic Markov cohort model. The use of such models is particularly valuable in that the models allow one to simulate event or disease risk over time, thus improving model validity, especially when evaluating interventions for episodic/recurrent risk events. However, a primary benefit of Markov modeling—capturing risk changes based on prior history—is limited by the very nature of such models, in which the ability to incorporate past history is restricted to the number of model health states. Model outcomes may thus be biased due to the changing risk over time that is not fully captured as individuals with past history and higher risk progress more rapidly to disease than their history-free counterparts. This article evaluates 2 interventions and compares outcomes in models with varying degrees of dependence on lifetime event history, thus quantifying the degree and direction of potential bias under varying model conditions and assumptions.

Prior research had considered the impact of between-person heterogeneity—both unobservable

and observable—on model outcomes. Kuntz and Goldie<sup>12</sup> looked at the impact of an unobservable dichotomous factor that may affect event risk and demonstrated that failure to incorporate these differential event risks would bias model outcomes, because such a model would not account for the fact that higher-risk individuals in the model would move into event states more rapidly than would those with lower risk. The magnitude of this heterogeneity bias was found to depend on baseline event risk and the relative risk of event and could be as large as 50%.<sup>12</sup> Similar conclusions were drawn in Zaric's analytic perspective of the issue.<sup>13</sup> Bentley and colleagues<sup>14</sup> similarly assessed the tradeoffs between model bias and complexity and/or data limitations when categorizing continuous risk factors in Markov models. The authors found that categorizing continuously valued risk factors in Markov models has a negligible effect on model outcomes, remaining under various assumptions at less than 4% absolute bias when at least 2 categories were used.

What had not previously been determined, however, is the bias that may occur when an individual's event history affects future risk. Because of the memoryless property of Markov models—by which event risk in each state is independent of prior states—the impact of history on future event risk cannot be captured without adding additional health states. For example, in Valenstein and colleagues' 2001 model of depression screening in primary care, transition probabilities were different for

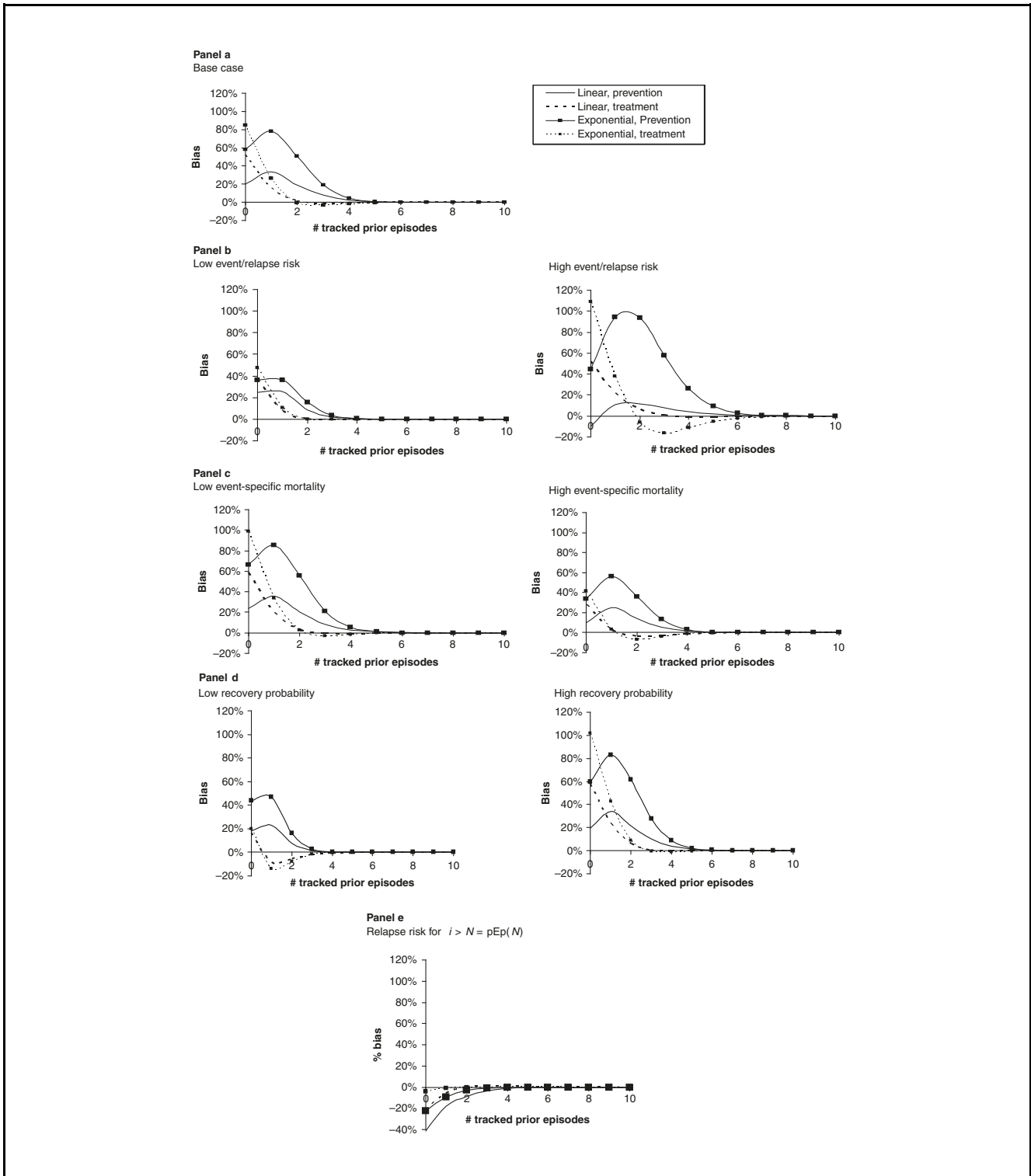


Figure 3 Event history bias in prevention and treatment interventions, with relapse risk as linear and exponential functions of tracked prior episodes assuming the base case (a) and 4 varied model parameters: event and relapse risk (b), event-specific mortality (c), probability of recovery (d), and relapse risk function (e).



patients who had never been depressed than for those who were currently depressed or had a history of depression and who were in either state but in or out of treatment.<sup>15</sup> This was accomplished in the Markov model by adding separate health states for currently depressed but not in treatment, depressed and in treatment, history of depression and in remission, and history of depression and in treatment.

We evaluated event-time saved and the degree and direction of event history bias for 2 different interventions in simplified generic Markov cohort models that incorporated varying degrees of dependence on prior episodic history. We compared results under varying model assumptions including risks of baseline event, recovery, relapse, and event-specific mortality, parameters considered important in helping modelers evaluate the impact of including greater or fewer event-history states when modeling episodic or recurrent diseases of varying prevalence and mortality. Event-history bias was defined as the percentage change in CTE that is prevented over a lifetime when models allow 0 to 9 event episodes to influence subsequent relapse risk as compared with a model that incorporates 10 such prior episodes. A positive bias thus indicates that failing to incorporate an adequate number of prior events and the full effects of event history on relapse risk overestimates the benefits gained from an intervention, whereas a negative bias suggests an underestimation.

For our base-case assumptions of a starting population of 12-year-olds with no history of prior events and a 3.44% age-specific baseline event risk, 1% event-specific mortality, 30% chance of recovery, and 8% age-specific lifetime event/relapse risk, the model predicted that a prevention strategy would reduce cumulative event time by, on average, more than double that of treatment (approximately 1.0 and 0.5 y of event time saved for prevention and treatment, respectively). This greater impact of prevention than treatment was also predicted when using an exponential function of risk growth with prior history, with 0.8 and 0.4 average years of time saved. The model projected that the bias caused by failing to incorporate prior event history would be positive—thus overestimating an intervention's effect—for prevention and that it would go from positive to negative with increasing values of  $N$  for treatment. The average bias would be greater in prevention than in treatment and when using the exponential as compared with linear risk functions. As long as at least 3 prior tracked episodes were incorporated, the model predicted that bias in either prevention or treatment would not exceed 8% with the

linear function or 19% with exponential. Even when using more extreme model parameters in sensitivity analyses, absolute bias never exceeded 26% in prevention and 11% in treatment as long as at least 4 prior episodes were tracked.

In general, we would expect event-history bias to be positive and that failing to incorporate prior event history would overestimate the impact of an intervention. This is due to the overestimation in earlier model cycles of recurrent event risk that is necessary when fewer prior tracked episodes—represented by  $N$ —are incorporated, in order to account for the increased relapse risk that should be captured for some individuals as they accumulate history of prior events (e.g., for many mental health disorders, with greater frequency of prior episodes comes greater likelihood of subsequent risk). This is exemplified in Figure 2b, which shows that the constant relapse risk assumed after  $N$  has been reached is higher at earlier episodes (values of  $i$ ), in order to maintain an assumption of constant lifetime event risk for all levels of prior history ( $N$ 's) incorporated. Thus, for smaller values of  $N$ , the model sends too many individuals at early cycles to event states (see Table 2), and the model overpredicts occurrence of earlier event episodes—represented by  $i$ —and overestimates the benefits of an intervention. A model that more fully captures event history, however (such as that for  $N = 9$ , shown in Figure 2), can isolate the increased relapse risk to the specific individuals with prior history and thus more accurately model event risk over time.

Although the model did—as expected—predict that bias would be positive in the prevention strategy, in treatment bias was predicted to be positive only for  $0 \leq N \leq 2$  and to become negative—indicating an underestimated benefit—when incorporating 3 or more prior episodes. This is likely because of 2 effects working against each other. As  $N$  increases, CTE decreases—in both nonintervention and intervention—due to the effect of overestimating lifetime average event risk at earlier values of  $N$ , as explained above and as occurred in the prevention strategy. Simultaneously in treatment, however, the intervention would—at lower values of  $N$ —both overestimate the number of people in earlier events as well as underestimate those in recovery, and this would occur to a greater extent in intervention than in nonintervention. The model consequently underestimates cumulative event time saved, and as  $N$  increases, this effect outweighs that of overestimated lifetime risk, resulting in a negative bias in the treatment intervention for higher values of  $N$ .

We find that the effect on these results of varying model input parameters is greatest when using an exponential risk function to evaluate a treatment intervention and when varying model estimates of event and relapse risk, when bias could reach close to 110% when zero prior episodes are tracked. Despite this, bias was never predicted to exceed 60% in treatment when using a linear risk function or to exceed 94% in prevention with either linear or exponential risk functions. When relapse risk for episodes beyond  $N$  was assumed to be the same as that for episode  $N$ , the model projected that absolute bias would decrease dramatically and that the sign would be negative—indicating an underestimated benefit—for prevention and treatment. Thus, when allowing risk to remain constant after  $N$  but at a lower level than that for the base case (as shown in Figure 2c), at lower values of  $N$  the model would fail to sufficiently compensate for the greater relapse risk of individuals with prior history and would underestimate benefits.

Yet whether we varied risk growth functions or model input parameters, the model predicted that absolute bias would decline rapidly with increasing  $N$ , suggesting that incorporating just 3 to 4 episodes of prior event history can greatly improve model accuracy when evaluating interventions for recurrent events over time. Given the challenges of obtaining data on the effects of multiple prior events, this knowledge offers modelers the opportunity for greater parsimony in model building, allowing the incorporation of only a few such prior episodes and easing model and results interpretation.

The results of this analysis must be considered in light of its limitations. To most clearly demonstrate the effects of failing to incorporate the full effects of prior history on model outcomes, we used generic disease-prevention models and made simplifying assumptions about the relationship between event history and relapse risk and about methods for modeling interventions. Although we evaluated results for both linear and exponential functions of relapse risk growth, these may not be clinically accurate, as the nature of such relationships between prior history and subsequent risk may be highly irregular, unpredictable, and/or nonparametric.

Similarly, this model does not consider duration or intensity of prior episodes or time since prior episode, any of which might have multiplicative effects on the bias. For example, a longer or more severe prior episode could elevate subsequent risk, such that our results could underestimate bias for

individuals with long prior episodes, or our results could similarly overestimate bias for those with shorter or less severe prior episodes. Because populations with multiple recurrences of episode/relapse conditions such as substance abuse tend to be dominated by people with either higher intensity or longer duration of use, it is likely that our results are conservative and underestimate potential bias.

We also did not allow probability of recovery to vary with the number of tracked prior episodes. Although recovery rates could decrease with number of prior episodes (e.g., people may become more addicted to a substance the more often it is used), it could be possible that the opposite scenario would play out and that history of success with prior recovery predicts similar success in the future. Therefore, at a population level, the potential impact—if any—of varying recovery with prior history is unclear. Future work should consider the possible effects of these other aspects of event history on model outcomes; evaluating the individual effects of duration and severity of prior episodes, time since previous episode, and impact of history on recovery could contribute to better knowledge of total bias.

Markov state-transition models offer valuable tools for simulating recurrent-event risks over time; yet because of the memoryless nature of Markov states, risk within each is assumed homogenous, and thus, such models' ability to capture the effects of prior history on subsequent risk over time is limited by the number of health states feasible in the model. Our analysis indicates that when using Markov models to evaluate interventions for episodic/recurrent diseases, failing to incorporate an adequate number of prior events and the full effects of event history on relapse risk can substantially overestimate predicted benefits of a prevention strategy and both overestimate and underestimate those of treatment. These potential errors should be considered by modelers when designing models and by policy makers when interpreting results of such analyses.

## REFERENCES

1. Eaton WW, Kramer M, Anthony JC, Dryman A, Shapiro S, Locke BZ. The incidence of specific DIS/DSM-III mental disorders: data from the NIMH Epidemiologic Catchment Area Program. *Acta Psychiatr Scand.* 1989;79(2):163–78.
2. Lewinsohn PM, Clarke GN, Seeley JR, Rohde P. Major depression in community adolescents: age at onset, episode duration, and time to recurrence. *J Am Acad Child Adolesc Psychiatry.* 1994;33(6):809–18.

3. Lewinsohn PM, Hops H, Roberts RE, Seeley JR, Andrews JA. Adolescent psychopathology: I. Prevalence and incidence of depression and other DSM-III-R disorders in high school students. *J Abnorm Psychol.* 1993;102(1):133–44.
4. Lewinsohn PM, Klein DN, Seeley JR. Bipolar disorder during adolescence and young adulthood in a community sample. *Bipolar Disord.* 2000;2(3 pt 2):281–93.
5. Lewinsohn PM, Rohde P, Klein DN, Seeley JR. Natural course of adolescent major depressive disorder: I. Continuity into young adulthood. *J Am Acad Child Adolesc Psychiatry.* 1999;38(1):56–63.
6. Lewinsohn PM, Rohde P, Seeley JR. Major depressive disorder in older adolescents: prevalence, risk factors, and clinical implications. *Clin Psychol Rev.* 1998;18(7):765–94.
7. Lewinsohn PM, Rohde P, Seeley JR, Fischer SA. Age-cohort changes in the lifetime occurrence of depression and other mental disorders. *J Abnorm Psychol.* 1993;102(1):110–20.
8. Lewinsohn PM, Rohde P, Seeley JR, Klein DN, Gotlib IH. Natural course of adolescent major depressive disorder in a community sample: predictors of recurrence in young adults. *Am J Psychiatry.* 2000;157(10):1584–91.
9. Lewinsohn PM, Zeiss AM, Duncan EM. Probability of relapse after recovery from an episode of depression. *J Abnorm Psychol.* 1989;98(2):107–16.
10. Perkonig A, Lieb R, Wittchen HU. Prevalence of use, abuse and dependence of illicit drugs among adolescents and young adults in a community sample. *Eur Addict Res.* 1998;4(1–2):58–66.
11. Arias E. United States life tables, 2001. National Vital Statistics Reports National Center for Health Statistics. 2004;52(14).
12. Kuntz KM, Goldie SJ. Assessing the sensitivity of decision-analytic results to unobserved markers of risk: defining the effects of heterogeneity bias. *Med Decis Making.* 2002;22(3):218–27.
13. Zanic GS. The impact of ignoring population heterogeneity when Markov models are used in cost-effectiveness analysis. *Med Decis Making.* 2003;23:379–96.
14. Bentley TGK, Weinstein MC, Kuntz KM. The effects of categorizing continuous variables in decision-analytic models. *Med Decis Making.* 2009;29(5):549–56.
15. Valenstein M, Vijan S, Zeber JE, Boehm K, Buttar A. The cost-utility of screening for depression in primary care. *Ann Intern Med.* 2001;134(5):345–60.