

Estimation of Markov Chain Transition Probabilities and Rates from Fully and Partially Observed Data: Uncertainty Propagation, Evidence Synthesis, and Model Calibration

Nicky J. Welton, PhD, A. E. Ades, PhD

Markov transition models are frequently used to model disease progression. The authors show how the solution to Kolmogorov's forward equations can be exploited to map between transition rates and probabilities from probability data in multistate models. They provide a uniform, Bayesian treatment of estimation and propagation of uncertainty of transition rates and probabilities when 1) observations are available on all transitions and exact time at risk in each state (fully observed data) and 2) observations are on initial state and final state after a fixed interval of time but not on the sequence of transitions (partially observed data). The authors show how underlying transition rates can be recovered from partially observed data using Markov chain Monte Carlo

*methods in WinBUGS, and they suggest diagnostics to investigate inconsistencies between evidence from different starting states. An illustrative example for a 3-state model is given, which shows how the methods extend to more complex Markov models using the software WBDiff to compute solutions. Finally, the authors illustrate how to statistically combine data from multiple sources, including partially observed data at several follow-up times and also how to calibrate a Markov model to be consistent with data from one specific study. **Key words:** rate; risk; probability; transitions; Markov models; uncertainty propagation; Bayesian MCMC; evidence synthesis; calibration. (*Med Decis Making* 2005;25:633–645)*

Markov chain models are a natural approach to take when modeling the transitions of patients between discrete health states over time, for example, the progression over stages of a disease. Such models fit neatly into a decision context, whereby cost-effectiveness calculations can be integrated into simulations over the chain using a cohort or Monte Carlo approach and, in some instances, can be calculated algebraically using a matrix approach.^{1,2} There is a distinction be-

tween discrete-time Markov chains, whereby we consider transitions to occur at fixed points in time and we work with transition probabilities, and continuous-time Markov chains, whereby transitions can occur at any point in time and we work with transition rates. The use of discrete-time Markov chain models in medical decision applications dates back to the work of Beck and Pauker.¹ Hazen,^{3–5} Hazen and Pellissier⁶ and Hazen and others⁷ have championed the use of continuous-time Markov chain models based on stochastic trees and have pointed out the mathematical convenience and simplicity of these models as long as the same rates are applied over a full lifetime. A discrete-time model can approximate a continuous-time Markov model by defining a cycle length of interest (for example, yearly or 6-month intervals).

Miller and Homan⁸ recommend that in some circumstances it may be best to estimate rates from the data and then transform these into probabilities of transition over a period of time. However, as they state in

Received 10 June 2003 from MRC Health Services Research Collaboration, Bristol, United Kingdom. Revision accepted for publication 16 March 2005.

Address correspondence to Nicky J. Welton, PhD, MRC Health Services Research Collaboration, Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol BS8 2PR, UK; telephone: +44 (0)117 331 3902; e-mail: nicky.welton@bristol.ac.uk.

DOI: 10.1177/0272989X05282637

their article, they restrict attention to the case in which only a single transition is possible from each state. To estimate rates in a multistate model, a cohort study recording all state transitions and sojourn times provides the ideal source of data.⁹ However, such detailed observations may not be available. Instead, we may only know the starting state and the state reached n years later but not the route taken through the model. We refer to this data as partially observed data. One approach, using matrix decomposition, is to convert a Markov transition matrix based on observations made at 1 cycle time to a transition matrix with a shorter cycle time more suited to the model.^{10–12} This article proposes an alternative approach, which is to estimate the underlying rate matrix from partially observed data using Kolmogorov's forward equations. We start with an explanation of the forward equations approach in simple situations to introduce ideas and notation. We then give a worked example for a 3-state model, showing how this model can be used to estimate transition rates from partially observed data and from this model derive estimates of transition probabilities at any required cycle time. We adopt a Bayesian framework, using WinBUGS Markov chain Monte Carlo (MCMC) software. This software has the flexibility of a simulation-based approach to uncertainty propagation and probabilistic decision modeling but with the advantage of allowing simulation from a posterior distribution.^{9,13} We then show how the software WBDiff, an add-on to WinBUGS, can be used to extend the ideas to more complicated models in higher dimensions. The advantage of the forward equations approach is that transition rates, rather than transition probabilities, form a common basis for combining information from different studies. This approach is illustrated by an evidence synthesis involving several sources of fully and partially observed data, including a demonstration of how a Markov model can be calibrated to be consistent with a specific study.

MAPPING BETWEEN TRANSITION RATES AND PROBABILITIES IN MULTISTATE MODELS VIA FORWARD EQUATIONS

We start with the familiar 2-state situation (Figure 1). Here, there are just 2 health states, and patients progress from state 1 to state 2 at constant rate $\gamma_{1,2}$. It is not possible to make the backward transition from state 2 to state 1. It is then a standard result that the transition probability of going from state 1 to 2 in time t , $\pi_{1,2}(t)$, is $1 - \exp(-\gamma_{1,2}t)$.

This situation is, in fact, a special case of a solution to Kolmogorov's forward equations, a general result

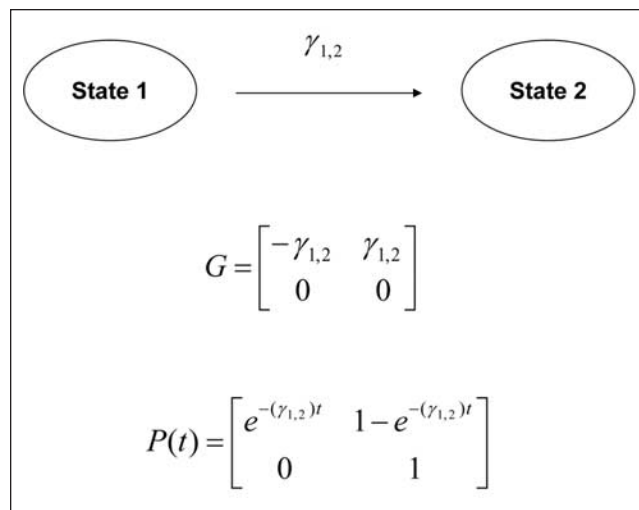


Figure 1 A 2-state Markov model, in which there is a single transition to an absorbing state. The transition rate matrix G , with elements $\gamma_{i,j}$, gives the rate at which transitions are made from state i to state j . Note that each row of G must sum to zero. The transition probability matrix $P(t)$, with elements $\pi_{i,j}(t)$, gives the probability of being in state j at time $t + s$, conditional on being in state i at time s . Note that each row of $P(t)$ must sum to 1. The $P(t)$ equations are the solution to the forward equations.

from probability theory,¹⁴ which maps between probabilities and underlying risks. In general, Kolmogorov's equations have the following solution [see Figure 1 legend for definition of $P(t)$ and G]:

$$P(t) = e^{tG} = \sum_{n=0}^{\infty} \frac{t^n}{n!} G^n. \quad (1)$$

The solution to the equations becomes increasingly complex as the number of states and possible transitions increases. However, they can (in theory at least) be evaluated using matrix decomposition (see the Appendix of Cinlar¹⁵). We have used the Linear Algebra and linalg packages within Maple 8.¹⁶ For example, Figure 2 illustrates the mapping from constant rates to probabilities for a slightly more complex model, in which there are again 2 states, but subjects can move backward from state 2 to state 1, as well as forward.

A number of technical conditions have to be met for solutions to the forward equations to exist,¹⁴ but these conditions will be readily satisfied in practical applications. It is important to note that all the models presented here are for transition rates that are constant over time, although extension to time-varying transitions is possible (see Discussion). In the next section, we illustrate how the underlying rate matrix G for a 3-state model can be estimated from data in which observations are available on only the starting states and the probability matrix $P(t)$ after a period of time t has

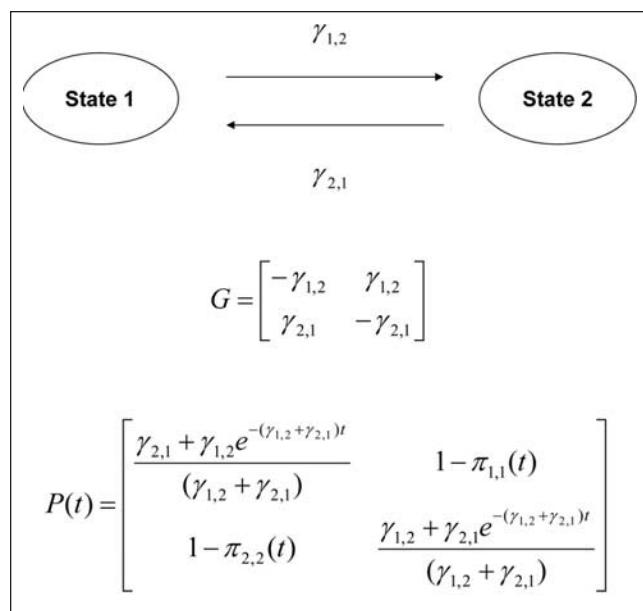


Figure 2 A 2-state Markov model with transitions in both directions. The transition rate matrix, G , and corresponding transition probability matrix, $P(t)$, are displayed.

elapsed. To set this method in context, we begin by reviewing estimation of transitions rates and probabilities from fully observed data.

ESTIMATING TRANSITION PROBABILITIES AND RATES FOR DIFFERENT DATA STRUCTURES AND PROPAGATION OF UNCERTAINTY

Observations on the Number of Transitions, Their Destination, and Time at Risk (Fully Observed Data)

The preferred data structure for estimating transition rates is to have observations on the destination state and the exact time of each transition. The observations made relate to the number of events and total time at risk rather than to the number of individuals because the same individuals can contribute observed time at risk to more than 1 state. The sufficient statistics for this type of data take the following form. Let $r_{i,j}$ be the number of transitions from state i to state j for $i \neq j$, then the total number of transitions from state i is $m_i = \sum_{j \neq i} r_{i,j}$, which, under the assumption of a continuous-time Markov chain, come from a Poisson distribution

$$m_i \sim \text{Poisson}(\lambda_i E_i),$$

where λ_i is the rate at which transitions are made from state i and E_i is the total person-years of exposure observed in state i . We give gamma priors for the λ_i

$$\lambda_i \sim \text{Gamma}(a_i, b_i),$$

where a_i and b_i can be interpreted as the prior belief on the total number of counts and total exposure, respectively.¹⁷ Typically, in the absence of prior information, we set all $a_i = .1$ and $b_i = .1$.

We next condition on having made a transition from state i . Let $\rho_{i,j}$ be the conditional probability that the destination is state j given a transition from state i occurs (for $i \neq j$). Then, under the assumption of a continuous-time Markov chain, $r_{i,1}, r_{i,2}, \dots, r_{i,i-1}, r_{i,i+1}, \dots, r_{i,k} \mid m_i \sim \text{Multinomial}(\rho_{i,1}, \rho_{i,2}, \dots, \rho_{i,i-1}, \rho_{i,i+1}, \dots, \rho_{i,k}; m_i)$.

We give a Dirichlet prior for the conditional probabilities $\rho_{i,j}$,

$$\rho_{i,1}, \rho_{i,2}, \dots, \rho_{i,i-1}, \rho_{i,i+1}, \dots, \rho_{i,k} \sim \text{Dirichlet}(d_{i,1}, d_{i,2}, \dots, d_{i,i-1}, d_{i,i+1}, \dots, d_{i,k}).$$

Typically, in the absence of prior information, we set all $d_{i,j} = 1$ (for $i \neq j$). Transitions rates $\gamma_{i,j}$ for $i \neq j$ can then be calculated as the probability of a transition from state i , multiplied that by the conditional probability of a transition to state j given a transition is made

$$\gamma_{i,j} = \begin{cases} \lambda_i \rho_{i,j} & i \neq j \\ -\lambda_i & i = j \end{cases}.$$

This model is the standard independent competing risks model.^{9,18–20} Figure 3 illustrates a simple 3-state model in which transitions between states 1 and 3 can occur directly or via state 2, and state 3 is an absorbing state. For example, state 1 could represent good health; state 2, stroke; and state 3, death. The task is now to estimate the rate matrix G from data. Suppose we observe the data displayed in Table 1 from a cohort study observing all transitions and exact timing of these events during a fixed study period. Person-years of exposure in states 1 and 2 were $E_1 = 51$ and $E_2 = 14$, respectively. Note that both censored and uncensored observations contribute to exposure. Because the times to events are observed exactly, we may estimate the rates directly. Moreover, using the model described above, the priors and likelihood are conjugate so that the posterior distribution of the rates can be derived in closed form.

$$\lambda_1 \sim \text{Gamma}(a_1 + m_1, b_1 + E_1), \quad \lambda_2 \sim \text{Gamma}(a_2 + m_2, b_2 + E_2)$$

$$\rho_{1,2} \sim \text{Beta}(d_{1,2} + r_{1,2}, d_{1,3} + r_{1,3})$$

$$\gamma_{1,2} = \lambda_1 \rho_{1,2}, \quad \gamma_{1,3} = \lambda_1 (1 - \rho_{1,2}), \quad \gamma_{2,3} = \lambda_2.$$

In this simple case, the Multinomial and Dirichlet reduce to the Binomial and Beta distributions, respectively. Although posterior means are easily obtained from the closed form solution, it is convenient to use simulation via MCMC to obtain credible intervals and

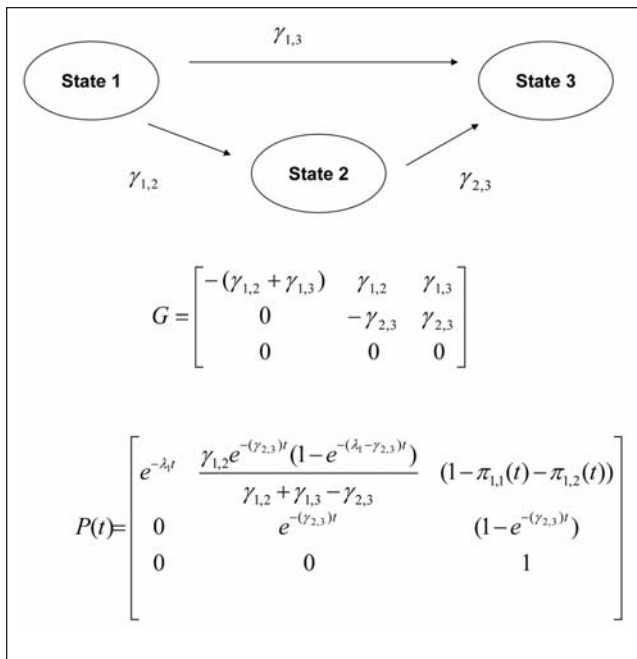


Figure 3 A 3-state Markov model with transitions from state 1 to the absorbing state 3 either directly or indirectly via state 2. The transition rate matrix, G , and corresponding transition probability matrix, $P(t)$, are displayed. For notational ease we define $\lambda_1 = \gamma_{1,2} + \gamma_{1,3}$, the sum transition rate from state 1.

Table 1 Example of Fully Observed Data for the 3-State Model Illustrated in Figure 3

Observed Exposure (person-years), E_i	State, i	Destination State, j		
		1	2	3
51	1	—	3	7
14	2	0	—	5
0	3	0	0	—

Note: The observed person-years exposure for a given state, E_i , is the time to event (or censoring) summed over all subjects who visit state i . The total number of transitions observed from state 1 was $m_1 = 10$, and the number of observed transition from state 1 to destination state 2, $r_{1,2}$, is 3. Censored observations, $r_{i,i}$, are not shown as these observations contribute to the parameter estimates only through the observed time at risk.

other posterior summaries. The results for the estimated transition rates are displayed in Table 2.

Note that the rates within each row will be correlated, via their relationship to the common λ_i , and because of the negative correlations between the ρ_{ij} on the same row. However, there is no correlation between rates on different rows of the data table.

Partially Observed Data

Suppose that the underlying process is the same as was described above, but instead of continuous observation, snapshots are taken at 2 points in time. Thus, for each subject, we know their starting state and final destination state after a period of time, but we do not know whether the destination was reached via 1 or more intermediary transitions from the initial state. For example, if a subject who was well at time t is dead at the 2-year follow-up, then he or she may have gone directly from the well to the dead state or, alternatively, he or she may have become ill before death. A similar situation arises when the cycle time used to collect the data is longer than the cycle time required in a discrete time model.^{10,12} Our approach will be to recover the rate matrix G from the data, which we regard as providing information on the cells of the matrix $P(t)$. Then, given G , we can derive transition probability matrix $P(t)$ for any t we choose.

We illustrate this with the data displayed in Table 3. Although the observed numbers of transitions are identical to Table 1, here the observations represent the distribution of subjects over states when observed after a 2-year time period has elapsed rather than representing the numbers of transitions occurring during the observation period. Only subjects recruited to the study at least 2 years ago are included. Once again, the data $r_{i,j}$ for each row i are drawn from a multinomial distribution with probabilities:

$$(r_{i,1}, r_{i,2}, \dots, r_{i,k}) \sim \text{Multinomial}(\pi_{i,1}(t), \pi_{i,2}(t), \dots, \pi_{i,k}(t); n_i).$$

We provide uninformative Gamma(1,1) priors for the unknown transition rate parameters $\gamma_{1,2}$, $\gamma_{1,3}$, and $\gamma_{2,3}$. In contrast with the fully observed data structure, whereby underlying rate parameters could be linked directly to data, in this case the data provide information on functions $P(t)$ of the underlying parameters G that are, as Figure 3 shows, highly complex. To find the posterior distributions of G , it is convenient to use Bayesian MCMC methods implemented in WinBUGS1.4 (<http://www.mrc-bsu.cam.ac.uk/bugs/>). The code required to fit this model, shown in Appendix 1, is no more than a statement of the above likelihood and priors, and the relationship between $P(t)$ and G shown in Figure 3. Similar applications of MCMC to estimate parameters from data on complex functions of parameters have been described recently^{21,22} and echo the key ideas of the Confidence Profile Method of Eddy and colleagues²³ and the Bayesian Monte Carlo methods from the environmental health risk assessment lit-

Table 2 Posterior Summary Statistics for Transition Rates for the 3-State Model (Figure 3)
Given the Fully Observed Data in Table 1, from 10 000 MCMC Simulations
after a Burn-in Period of 10 000 Simulations

	Parameter					
	$\gamma_{1,2}$	$\gamma_{1,3}$	$\gamma_{2,3}$	λ_1	γ_2	$\rho_{1,2}$
Posterior Mean	0.07	0.13	0.36	0.20	0.36	0.33
95% Credible Interval	(0.02–0.15)	(0.05–0.24)	(0.12–0.72)	(0.09–0.34)	(0.12–0.72)	(0.11–0.61)

Note: MCMC = Markov chain Monte Carlo.

Table 3 Example of Partially Observed Data:
Membership State after a 2-Year Time Cycle for the
3-State Model Illustrated in Figure 3

Numbers at Start	Initial State, i	State after t Years Have Elapsed, j		
		1	2	3
30	1	20	3	7
10	2	0	5	5
0	3	0	0	0

Note: The state membership data after 2 years for those initially in state 1, $(r_{1,1}, r_{1,2}, r_{1,3}) = (20, 3, 7)$, are assumed to be drawn from the Multinomial $(\pi_{1,1}(2), \pi_{1,2}(2), \pi_{1,3}(2); 30)$ distribution. The $\pi_{i,j}(2)$ can in turn be expressed in terms of the rate parameters, as shown in the matrix $P(t)$ in Figure 3.

erature used to inform model parameters from data on model outputs.^{24,25}

The posterior summaries are displayed in Table 4. With partially observed data, there is uncertainty as to whether those in state 3 arrived direct from state 1 or via state 2. As a result, the data on arrivals into state 3 is compatible with either $\gamma_{1,3}$ high and both $\gamma_{1,2}$ and $\gamma_{2,3}$ low, or with $\gamma_{1,3}$ low and both $\gamma_{1,2}$ and $\gamma_{2,3}$ high, giving rise to the negative correlation between parameters on different rows of G (Table 4). As there is also information on $\gamma_{2,3}$ in row 2 of Table 3, the unobserved data gives rise to the possibility of evidence inconsistency (see Model Comparison and Evidence Consistency).

Given the estimated transition rates, transition probabilities can be generated for any given cycle lengths using the relationship displayed in Figure 3 and the appropriate uncertainties carried forward within the simulation. Table 5 presents the results for cycle lengths of 2 years and 3 months, respectively. As we would expect, the probability of leaving the initial state is greater during a 2-year cycle than during a 3-month cycle.

Note that observing zero counts in empirical data does not necessarily imply that the transition rate is

Table 4 Posterior Summary Statistics and
Correlations for Transition Rates for the 3-State
Model Illustrated in Figure 3 with the Partially
Observed Data in Table 3, from 10 000 MCMC
Simulations after a Burn-in Period of 10 000
Simulations

Parameter	Posterior Summaries		Correlations		
	Mean	95% Credible Interval	$\gamma_{1,2}$	$\gamma_{1,3}$	$\gamma_{2,3}$
$\gamma_{1,2}$	0.12	(0.02–0.30)	1.00	–0.57	0.42
$\gamma_{1,3}$	0.09	(0.00–0.23)		1.00	–0.46
$\gamma_{2,3}$	0.42	(0.13–0.87)			1.00

Note: MCMC = Markov chain Monte Carlo.

zero. Briggs and others²⁶ cite this as a reason for preferring a Bayesian approach.

USING WBDiff TO OBTAIN SOLUTIONS FOR MORE COMPLEX MODELS

Although our illustration is based on a 3-state forward model, the approach is in principle completely general. The only restriction, albeit a major one, being the practical feasibility of obtaining solutions to $P(t) = e_{tG}$ in algebraic form for complex multistate models. Some solutions to more complex structures, including a 3-state model with backward transitions, can be found on our Web site at http://www.hsrc.ac.uk/Current_research/research_programmes/mpes.htm.

These solutions were obtained by writing the matrix G in the form BDB^{-1} , where D is the diagonal matrix of eigenvalues of G , and B is a matrix of eigenvectors. Maple 8 software¹⁶ was then used to solve for $P(t)$. Although solutions for $P(t)$ can be shown to exist, they become increasingly unwieldy as the dimensions and

Table 5 Posterior Summary Statistics for Transition Probabilities during 3-Month and 2-Year Cycles Based on the 2-Year Cycle Partially Observed Data in Table 3, from 10 000 MCMC Simulations after a Burn-in Period of 10 000 Simulations

Posterior Mean (95% Credible Interval)		Parameter				
		$\pi_{1,1}(t)$	$\pi_{1,2}(t)$	$\pi_{1,3}(t)$	$\pi_{2,2}(t)$	$\pi_{2,3}(t)$
3-month cycle	$t = 0.25$	0.95 (0.92–0.98)	0.03 (0.01–0.07)	0.02 (0.00–0.06)	0.90 (0.81–0.97)	0.10 (0.03–0.19)
2-year cycle	$t = 2$	0.66 (0.49–0.82)	0.13 (0.03–0.27)	0.21 (0.09–0.38)	0.46 (0.18–0.78)	0.54 (0.22–0.82)

Note: MCMC = Markov chain Monte Carlo.

complexity of the Markov model increase. In practice, algebraic solutions have to be abandoned in favor of numerical solutions, which need to be calculated within WinBUGS at each iteration of the MCMC algorithm. WinBUGS does not currently have the functionality for matrix algebra; however, the forward differential equations $P'(t) = P(t)G$ can be solved numerically within an MCMC simulation using the recently developed software WinBUGS Differential Interface, WBDiff.²⁷ This method requires the modeler only to write down the system of differential equations rather than find their solution. Appendix 2 shows how WBDiff²⁷ can be used to solve the forward equations for more complex Markov models, using the 4-state model with forward transitions for illustration.

MODEL COMPARISON AND EVIDENCE CONSISTENCY

For fully observed data, the model will give a perfect fit to the data if we have the same number of data points as parameters, because parameters from different rows of G are independent. However, for partially observed data, parameters from different rows of G are correlated, as information in rows 1 and 2 both contribute to the rate parameter $\gamma_{2,3}$. This model creates the possibility that data from different rows might be inconsistent, given a particular model, and it therefore becomes important to check the goodness of fit of the model. To illustrate, we compare the 3-parameter model presented in Figure 3 with a 2-parameter tunnel model in which $\gamma_{1,3}$, the transition rate from state 1 direct to state 3, is assumed to be zero so that all transitions from state 1 to 3 are via the tunnel state 2. For example, a tunnel model might be appropriate when modeling untreated human immunodeficiency virus disease progression through increasingly severe levels of immune compromise to acquired immunodeficiency syndrome. We can assess

whether it is reasonable to assume that $\gamma_{1,3} = 0$ by comparing the fit of the 2 models. To do this comparison, we use the posterior mean deviance statistic²⁸ (standardized by subtracting the deviance for the saturated model,²⁹ whereby predicted values are set equal to observed values). We would expect a posterior mean deviance of approximately 3 (the number of data points) for a model that fits the data well.

If these models are applied to the data in Table 3, then the 3-parameter model gives a far better fit, with posterior mean deviance 3.6 compared with 6.5 in the 2-parameter model. Figure 4 examines goodness of fit in a range of imaginary data sets in which $r_{1,3}$ is varied, but $r_{1,2} + r_{1,3}$ is held constant at 10. The remainder of the data in Table 3 is unchanged. As $r_{1,3}$ becomes smaller, the fit of both models becomes worse. This result is because the direct information on $\gamma_{2,3}$ based on the 5/10 transitions from state 2 to state 3, together with the direct information on $\gamma_{1,2}$, predicts that, from the 1→2→3 route alone, $r_{1,3}$ should be higher than that observed. However, as $r_{1,3}$ increases, this conflict disappears, and in the 3-parameter model, the extra number of observations from state 1 to state 3 can be explained by the direct transition rate $\gamma_{1,3}$. This is not the case in the 2-parameter tunnel model, whereby a conflict arises between the estimate of $\gamma_{2,3}$ obtained directly from initial state 2 and that obtained indirectly from initial state 1. Evidently, the structural relationships imposed by partially observed data place strong constraints on the parameter space.

SYNTHESIS AND CALIBRATION OF A MARKOV TRANSITION RATE MODEL FROM MULTIPLE INFORMATION SOURCES

Previous sections have looked at the analysis of single bodies of data. Although there are exceptions,^{10,11} it is uncommon to obtain all the necessary information from a single study. This final section shows how infor-

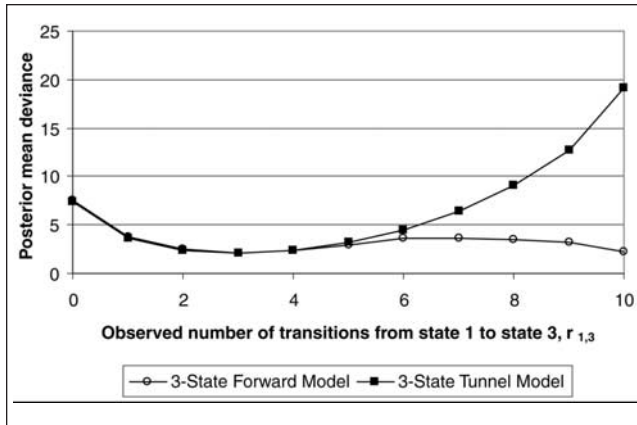


Figure 4 Comparison of the goodness of fit of the 3-parameter model illustrated in Figure 3 and a 2-parameter tunnel model in which transitions from state 1 to state 3 can only occur via state 2 (so that $\gamma_{1,3} = 0$). The posterior mean deviance is plotted for each model for 11 imaginary data sets in which the transitions from state 1 to 3, $r_{1,3} = 0, 1, \dots, 10$, $r_{1,2} = 10 - r_{1,3}$, and all other observations are as in Table 3.

mation from several studies with partial or full observation of 1 or more transitions can be statistically combined. In addition, we show how model parameters based on such multiple sources can be calibrated to be consistent with information from 1 particular study.

The illustration is based around the 3-state model with forward transitions only (Figure 3). The 3 basic parameters were $\gamma_{1,2}$, $\gamma_{1,3}$, and $\gamma_{2,3}$, or equivalently $\lambda_1(1 - \rho)$, $\lambda_1\rho$, and $\gamma_{2,3}$. To account for between-study heterogeneity in transition rates, we now assume that λ_1 will vary between studies, whereas the proportion ρ of patients leaving state 1 who go directly to state 3 remains constant across studies, as does the hazard rate $\gamma_{2,3}$ between states 2 and 3. It is further assumed that the log of the rates $\lambda_{1,i}$ across studies i are drawn from a normal random-effects model with mean Λ and variance σ^2 . The posterior distributions for Λ , σ^2 , ρ , and $\gamma_{2,3}$ will be estimated from the data, given minimally informative priors.

Many other model structures could be considered, but the above is typical of decision analyses involving multistate disease progression models. The assumption that the rate $\lambda_{1,i}$ of leaving the initial state varies between studies, whereas the distribution ρ of destination states is fixed, has been made for example in several analyses of treatments for atrial fibrillation,^{30–32} although the variation is attributed to baseline risk factors or treatment effects. Similarly, once patients have reached a clinically defined state, these studies assume a fixed rate of progression to more severe states, as with our $\gamma_{2,3}$. The log scale for the study-specific rates is of course standard practice in meta-analysis of rates.³³

Table 6 Example Data for Multiple-Source Evidence Synthesis

Study No.	No. of Transitions			Person-Years at Risk
Type A Studies	$r_{1,2}$	$r_{1,3}$	$r_{2,3}$	E
1	8			120
2	20			620
3		12		140
4		44		677
5			9	34
6			12	35
7			5	15
8			6	25
Type B Studies				
9	18	40		380
10	30	75		1169
Type C Studies	No. Reaching State 3 $r_{1,3}$	No. Starting in State 1 n	Observation Time T	
11	38	181	2	
12	15	177	1.5	
12	11	103	2	
13	60	335	5	

Note: Type A studies give observations on a single transition rate that may differ between studies. Type B studies observe first transitions from state 1 only. Type C studies give the number of participants starting in state 1 that have reached state 3 by some observation time. Study 12 has 2 groups of participants that were observed after 1.5 and 2 years, respectively.

In our fictitious illustration, 13 studies are available, comprising information of 3 types (Table 6). Each of the 8 type A studies provide a number of events r_i and a person-years exposure E_i for just 1 of the 3 rates:

$$r_i \sim \text{Poisson}(\mu_i)$$

$$\mu_i = \begin{cases} E_i \lambda_{1,i} (1 - \rho) & \text{for studies on the transition 1 to 2} \\ E_i \lambda_{1,i} \rho & \text{for studies on the transition 1 to 3} \\ E_i \gamma_{2,3} & \text{for studies on the transition 2 to 3} \end{cases}$$

There are 2 type B studies in which patients in state 1 have been observed over a total time at risk period E_i , and the number of *first* transitions $r_{1,2,i}$ and $r_{1,3,i}$ to states 2 and 3 have been recorded. These are fully observed studies, but observation ends when the patient reaches either end state. These studies contribute information on the rate of leaving state 1, and on ρ :

Table 7 Posterior Summary Results for the 3-State Model (Figure 3), Given the Multiple-Source Evidence Synthesis Data in Table 6, from 20 000 MCMC Simulations after a Burn-in Period of 10 000 Simulations

Posterior Mean (95% Credible Interval)	Parameter					
	Λ	σ	ρ	$\gamma_{2,3}$	$\gamma_{1,2}$	$\gamma_{1,3}$
(a) Mean	-2.27 (-2.62, -1.89)	0.47 (0.24, 0.84)	0.69 (0.62, 0.76)	0.299 (0.206, 0.409)	0.032 (0.021, 0.048)	0.073 (0.050, 0.105)
(b) Predictive					0.037 (0.011, 0.093)	0.083 (0.025, 0.211)
(c) Calibrated to study 13					0.016 (0.011, 0.022)	0.036 (0.027, 0.045)

Parameters: Λ = mean of normal distribution for the logs of study-specific rates $\lambda_{1,i}$ with standard deviation σ ; ρ = proportion of those leaving state 1 who go to state 3; $\gamma_{2,3}$ = transition rate from state 2 to state 3. Parameters $\gamma_{1,2}$ and $\gamma_{1,3}$ are based on either 1) the distribution of the mean rate of leaving state 1, 2) the predictive rate for a new study, or 3) calibrated to the estimated rate in study 13. MCMC = Markov chain Monte Carlo.

$$(r_{1,2,i} + r_{1,3,i}) \sim \text{Poisson}(\mu_i)$$

$$r_{1,3,i} | (r_{1,2,i} + r_{1,3,i}) \sim \text{Binomial}(\rho, r_{1,2,i} + r_{1,3,i})$$

$$\log(\mu_i) = \log(E_i) + \log(\lambda_{1,i})$$

Finally, there are 3 partially observed studies, type C, on n_i patients who start in state 1, recording the number r_i who reach state 3 in a study period T_i . It is not recorded how many go via state 2. This gives us the likelihood $r_i \sim \text{Binomial}(\pi_{1,3,i}, n_i)$ where the equations for $\pi_{1,3,i}$ are given in terms of $\lambda_{1,i}(1 - \rho)$, $\lambda_{1,i}\rho$, and $\gamma_{2,3}$ using the formulae shown in Figure 3. Study 12 comprises 2 groups of patients that were observed at 1.5 and 2 years, respectively. The likelihood for these 2 items of data share the study-specific parameter $\lambda_{1,12}$ but use a different value of t in the formula shown in Figure 3. Note that patients in the 2-year follow-up group that had reached an early end point by 1.5 years must still be analyzed in the 2-year group to avoid biased mortality estimates.

Clearly, a range of alternative models could be considered for this evidence structure, based on different epidemiological assumptions about the relationships between studies. Care must be taken, however, to ensure that model parameters are identifiable. For example, although type A studies in this example can provide direct information on $\gamma_{2,3}$, they can only provide indirect information on $\lambda_{1,i}$ and ρ , as they estimate the products $\lambda_{1,i}(1 - \rho)$ or $\lambda_{1,i}\rho$. Estimates for $\lambda_{1,i}$ are obtained, however, because ρ is informed directly by the type B studies. Similarly, given the information on ρ and $\gamma_{2,3}$ from type A and B studies, the type C studies may also be mainly providing information on the distribution of the $\lambda_{1,i}$. The WinBUGS code to fit this model is shown in Appendix 1, and posterior summa-

ries for parameters in this model, given the data list shown in Table 6, appear in Table 7.

The heterogeneity in evidence on $\gamma_{1,2,i}$ and $\gamma_{1,3,i}$ raises issues about the relationship between patient groups i , from which the historical evidence is taken, and the target population whose parameters, $\gamma_{1,2,R}$ and $\gamma_{1,3,R}$, we require. One option would be to use the mean log rate Λ , and then form the posterior distributions for $\gamma_{1,2,R}$ and $\gamma_{1,3,R}$ by calculating $(1 - \rho)\exp(\Lambda)$ and $\rho\exp(\Lambda)$, respectively, at each iteration of the MCMC simulation. An alternative would be to draw a new $\log(\lambda_{1,i})$ from the distribution $\text{Normal}(\Lambda, \sigma^2)$, and multiply this into ρ and $(1 - \rho)$ on each iteration of the MCMC simulation. This method assumes that the target population is exchangeable with the populations in previous studies. The results from these 2 possibilities are shown in Table 7.

A final, more interesting, possibility is that one particular study K best represents the target population under consideration, perhaps because it is a particularly representative or contemporary study. Analysts may then wish to calibrate the required parameters to conform to that study. In this case, our aim is to use the totality of the evidence available on $\gamma_{2,3}$ and ρ , and then have a distribution for λ_1 that is compatible with the rate at which patients in state 1 arrive in state 3 in study K . This result is accomplished very easily by forming the posterior distribution for $\gamma_{1,2,R}$ and $\gamma_{1,3,R}$ from $(1 - \rho)\lambda_{1,K}$ and $\rho\lambda_{1,K}$, respectively (Table 7).

DISCUSSION

We have attempted to provide a unified Bayesian approach to propagation of uncertainty from both fully and partially observed event history data to Markov

model parameters. The ideal form of data for estimating transition rates and probabilities would come from cohort studies in which all transitions and sojourn times are recorded (fully observed data). The literature on event history analysis (for example, see Andersen and Keiding³⁴ and Keiding³⁵) deals with this type of data. Our approach in this case is essentially identical to the standard independent competing risks analysis.

The partially observed data structure is really a degraded form of fully observed data in which the observer records the initial state and then returns to record states after a fixed interval. Our proposal is that forward equations are used to define the relation between the rate matrix and the observed data. This method can be considered an alternative to the approach presented by Craig and Sendi,¹⁰ Sendi and others¹², and Craig and others,¹¹ who map between transition probabilities on cycle lengths in the observed data and the cycle lengths required for modeling. Although the mathematical basis for the 2 approaches has much in common, we think that the use of the underlying rates as a common currency facilitates synthesis of different sources of evidence.

Our approach shares a similar philosophy with that presented by Craig and others,¹¹ who also use a Bayesian MCMC approach to model unobserved states. We feel that compared with maximum likelihood methods, the Bayesian approach is conceptually simpler, more flexible, easier to implement, and more in keeping with the decision context. It provides for the propagation of results direct from the joint posterior distributions into a decision analytic context based on either the discrete-time Markov models popularized by Beck and Pauker¹ and Sonnenberg and Beck² or the continuous-time models proposed by Hazen³⁻⁵, Hazen and Pellissier,⁶ and Hazen and others.⁷

In our partially observed data structure, we have restricted attention to the situation in which there are measurements at only 2 points in time (at the beginning and end of the study). A natural extension to this data structure is when we have repeated measurements over time, but multiple transitions between states may occur between observation times.¹⁰ Repeated observations are required to ensure identifiability in models with backward as well as forward transitions. Repeated measures are readily incorporated into our modeling framework by treating each observed distribution of states as multinomial in relation to the initial state, whereby transition probabilities depend on observation time through the solution to the forward equations.

Throughout, we have assumed that transition rates are constant over time, an assumption well known to

be unrealistic in many fields of medicine. However, the methods can be readily adapted by considering transition rates as constant within time or age strata. Then, transition rates and probabilities can be estimated separately for each age or time stratum. It would also be simple to extend the partially observed data model to include covariates, as has been done for fully observed data.^{9,20}

As noted, the technical requirements that must be met for the solutions to the forward equations to be valid are not strenuous in practical applications, and the methods may be readily applied to models with a large number of states and with both forward and backward transitions. Perhaps the most important restriction is the practical feasibility of solving $P(t) = e^{tG}$ for complex multistate models. However, the forward differential equations $P'(t) = P(t)G$ can be solved numerically within an MCMC simulation using the recently developed software WinBUGS Differential Interface.²⁷ This software is a very powerful tool that allows easy application of the methods presented here to more complex models. However, it is important to understand the evidence-consistency issues that arise when estimating transition rates from probability data in multistate models.

It is common to find a baseline Markov disease progression model that is informed from fully observed cohort studies, whereas relative treatment effect parameters come from clinical trials. The trials, however, may report (fully observed) relative risk or relative rate measures for individual transitions or (partially observed) time to clinical end points without reporting whether intermediate states were visited. Our illustrative evidence synthesis shows how partially observed and fully observed data can in principle be combined to inform an epidemiologically realistic model, though the 3 evidence types in our example by no means exhaust the huge variety of ways in which findings on disease progression can be reported. The underlying rationale of this approach to evidence synthesis is that *all* available data should be combined, including data on functions of the model parameters, to avoid an essentially arbitrary choice of which studies to include, to obtain the most precise estimates possible, and to give a realistic assessment of the uncertainty in the problem.²¹ These objectives are similar to those underlying systematic review.³⁶

Calibration of models by incorporation of field observations on model outputs is well known in risk analysis using a technique called Bayesian Monte Carlo^{24,37,38} and also in deterministic models for stock assessment.³⁹ In these disciplines, information from field data is propagated backward through the model to

inform parameters on which prior information is sparse. In our illustration, calibration of the Markov model to conform with a particular study represents a further particularly complex calculation but one that is easy to implement in a Bayesian MCMC framework. In effect, we begin with the joint posterior distribution for the parameters ρ and $\gamma_{2,3}$ based on all the data, and we then choose a distribution for λ_1 in such a way that a function of all 3 parameters, namely $\pi_{1,3}(5)$, is consistent with the likelihood provided by 1 particular study. This kind of maneuver is often adopted informally when a Markov process developed for 1 set of patients needs to be adapted to another, with faster or slower progression rates, or included in a deterministic sensitivity analysis. However, such informal adjustments fail to capture the uncertainties in the data. The present proposals represent a way of achieving the calibration within a probabilistic decision-modeling framework, correctly propagating the uncertainties both in the parameters and in the study to which the model is to be calibrated and incorporating whatever correlations between parameters are induced by the data structure.

By taking transition rates, rather than transition probabilities, as the common basis for evidence synthesis, investigators can combine information from different study designs, including partially observed studies, each recording the distribution of states at a different follow-up time. The use of forward equations may therefore help to open the way to a more formal and systematic use of available literature to inform Markov models in decision making.

ACKNOWLEDGMENTS

The Department of Social Medicine of the University of Bristol is a host center for the MRC Health Services Research Collaboration. We thank Dave Lunn and Andrew Thomas (Imperial College School of Medicine) for access to the beta-release of the WINBUGS Differential Interface (WBDiff) software. We also thank an anonymous referee for their helpful comments on earlier drafts of the manuscript.

APPENDIX 1

Files containing WinBUGS1.4 code for the analyses presented in the article, including those set out below, are available from our Web site, http://www.hsrb.ac.uk/Current_research/research_programmes/mpes.htm.

All the posterior summaries presented were obtained after a conservative burn-in period assessed using the Brooks-Gelman-Rubin diagnostic⁴⁰ tool within WinBUGS and based on a further sample large enough for the posterior standard deviation to be less than 5% of the Monte Carlo error. Markov chain Monte Carlo (MCMC) software is not strictly necessary for fully observed data because posterior distributions of the parameters can be found in closed form, and probabilistic modeling can then be carried out by simple Monte Carlo simulation. This is not the case for the partially observed data structure.

WinBUGS1.4 code for the model presented for partially observed data

```
model{
#Multinomial likelihood for observed data
  for (i in 1:2){ r[i,1:3] ~ dmulti(P[i,1:3],n[i]) }
#Find transition probabilities (for given time) in terms of rates
#using Fig. 3
  s<-(G[1,2]/(G[1,2]+G[1,3]-G[2,3]))

  P[1,1]<- exp(-(G[1,2] + G[1,3])*2)
  P[1,2]<- s*exp(-G[2,3]*2)*(1 - exp(-(G[1,2]+G[1,3]-
    G[2,3])*2))
  P[1,3]<-1 - P[1,1] - P[1,2]
  P[2,1]<-0
```

```
  P[2,2]<-exp(-G[2,3]*2)
  P[2,3]<-1 - P[2,2]
#Give exponential priors for unknown transition rate
parameters
  for (j in 2:3){G[1,j] ~ dgamma(.1,.1)}
  G[2,3] ~ dgamma(.1,.1)

#Find P(t) where t=3 months = 0.25 years
  P3[1,1]<- exp(-(G[1,2] + G[1,3])*2)
  P3[1,2]<- s*exp(-G[2,3]*.25)*(1 - exp(-(G[1,2]+G[1,3]-
    G[2,3])*2))
  P3[1,3]<-1 - P3[1,1] - P3[1,2]

  P3[2,1]<-0
  P3[2,2]<-exp(-G[2,3]*.25)
  P3[2,3]<-1 - P3[2,2]
}
```

```
#Data
list(r=structure(.Data=c(20,3,7,0,5,5), .Dim=c(2,3)), n=
c(30,10))
```

WinBUGS1.4 code for multiple-source evidence synthesis

```
model {
# PRIORS
# 'random effects' model for logs of study-specific rates lam1
of leaving # state 1. "study 14" represents predictive distr,
new study, and position
```

```

# 15 assigned to population mean
for (i in 1:14) { loglam1[i] ~ dnorm(L,P) }
loglam1[15] <- L
for (i in 1:15) {
  log(lam1[i]) <- loglam1[i]
# study-specific transition rates (1 to 2 and 1 to 3 resp.)
  gamma12[i] <- lam1[i] * (1-rho)
  gamma13[i] <- lam1[i] * rho
}
# vague priors for basic parameters
gamma23 ~ dexp(.001)
L ~ dnorm(0,.001)
sd ~ dunif(0,1)
P <- 1/(sd*sd)
rho ~ dbeta(.5,.5)

# 8 independent studies (A) informing individual transitions
for (i in 1:2){r12[i] ~dpois(mu[i])
  log(mu[i]) <- log(E12[i]) + log( (1-rho) * lam1[i] )
}
for (i in 1:2){r13[i] ~dpois(mu[i+2])
  log(mu[i+2]) <- log(E13[i]) + log( rho * lam1[i+2] )
}
for (i in 1:4){r23[i] ~dpois(mu[i+4])
  log(mu[i+4]) <- log(E23[i]) + log(gamma23)
}

# 2 studies (B) on patients starting in state 1 and observed to
states 2 & 3
for (i in 1:2) {
  n[i] <- s12[i] + s13[i]
  n[i] ~ dpois(nu[i])
  log(nu[i]) <- log(E[i]) + loglam1[i+8]
  s13[i] ~ dbin(rho,n[i])
}

```

```

# 3 studies (4 data points) reported numbers number starting
in state 1 and # entering state 3 during time T. Study 12
consists of 2 groups of # individuals with different follow-
up times, T
for (i in 1 : 4) {
  z[i] <- lam1[study[i]]-gamma23
  pi11[i] <- exp(-T[i] * lam1[study[i]])
  pi12[i] <- gamma12[study[i]] * exp(-gamma23 *
    T[i]) * ( 1- exp(- (T[i] * z[i])) ) / z[i]
  pi13[i] <- 1 - pi11[i] - pi12[i]
  r[i] ~ dbin(pi13[i],nC[i])
}
g12[1] <- (1-rho) * exp(L) #log(lam1) based on the popula-
tion mean L
g13[1] <- rho * exp(L)
g12[2] <- (1-rho) * lam1[14]
g13[2] <- rho * lam1[14]
g12[3] <- (1-rho) * lam1[13]
g13[3] <- rho * lam1[13]
}

# DATA LIST
list(
# 8 studies (A); no. events rij, person years at risk Eij.
  r12=c(8,20),      E12=c(120,620),
  r13=c( 12,44),    E13=c(140,677),
  r23=c(9,12,5,6),  E23=c(34,35,15,25),
# 2 studies (B); no. transitions sij, study observation period E
  s12=c(18,30),      s13=c(40,75),      E=c(380,1169),
#3 studies (C); no. entering state 3 r, study observation period
T, no.
# starting instate 1 nC
  r=c(38,15,11,60),      T=c(2,1.5,2,5),
  nC=c(181,177,103,335), study=c(11,12,12,13) )

```

APPENDIX 2

We illustrate the use of the freely available WBDiff²⁷ for the 4-state model in which only forward transitions are possible and state 4 is an absorbing state. The data used to illustrate the model are shown in Table 2-1. The notation used is a direct extension of that for the 3-state model described in the text, where $\pi_{ij}(t)$ are the elements of matrix $P(t)$. Writing the system of equations $P'(t) = P(t)G$ in full gives the following 6 equations that we wish to solve,

$$\begin{aligned}
 \pi'_{1,1}(t) &= \pi_{1,1}(t)\gamma_{1,1} \\
 \pi'_{1,2}(t) &= \pi_{1,1}(t)\gamma_{1,2} + \pi_{1,2}(t)\gamma_{2,2} \\
 \pi'_{1,3}(t) &= \pi_{1,1}(t)\gamma_{1,3} + \pi_{1,2}(t)\gamma_{2,3} + \pi_{1,3}(t)\gamma_{3,3} \\
 \pi'_{2,2}(t) &= \pi_{2,2}(t)\gamma_{2,2} \\
 \pi'_{2,3}(t) &= \pi_{2,2}(t)\gamma_{2,3} + \pi_{2,3}(t)\gamma_{3,3} \\
 \pi'_{2,3}(t) &= \pi_{3,3}(t)\gamma_{3,3}
 \end{aligned}$$

subject to the initial condition that $P(t) = I$. The remaining elements of $P(t)$ are found by the constraints that each row sums

to 1, and backward transitions are zero. This system of differ-

Table Appendix 2-1 Example Partially Observed Data: Membership State after a 2-Year Time Cycle for the 4-State Forward Model

Numbers at start	Initial state, <i>I</i>	State after 2 Years Have Elapsed, <i>j</i>			
		1	2	3	4
32	1	20	6	4	2
13	2	0	5	5	3
10	3	0	0	3	7
0	4	0	0	0	0

ential equations can be written into a WBDiff²⁷ module, which can be linked to WinBUGS. For inputs γ_{ij} , the module will return numerical solutions π_{ij} using the Runge-Kutta method. To do this calculation, please download the WBDiff²⁷ software from http://homepages.tesco.net/~creeping_death/.

The documentation gives instructions with an example on how to develop your own modules, including how to download the BlackBox software. The key lines of code for the 4-state forward model are

```
(*3*) nEq = 6;
(*4*) Q_11 = 0; Q_12 = 1; Q_13 = 2; Q_22 = 3; Q_23 = 4; Q_33 = 5;
(*26*) dCdt[Q_11] := C[Q_11]*theta[Q_11];
(*27*) dCdt[Q_12] := C[Q_11]*theta[Q_12] + C[Q_12]*theta[Q_22];
(*28*) dCdt[Q_13] := C[Q_11]*theta[Q_13] + C[Q_12]*theta[Q_23] + C[Q_13]*theta[Q_33];
(*29*) dCdt[Q_22] := C[Q_22]*theta[Q_22];
(*30*) dCdt[Q_23] := C[Q_22]*theta[Q_23] + C[Q_23]*theta[Q_33];
(*31*) dCdt[Q_33] := C[Q_33]*theta[Q_33];
```

The number of equations nEq = 6. The γ_{ij} parameters are in a vector theta indexed 0 to 5. The dummy variables Q_ij help map between the transition rate matrix and the parameter vector. We have called the WBDiff²⁷ module FourStateForward.odc. After compiling this, the following line is added to the file ProgramFiles/BlackBox/WBDiff/Rsrc/Grammar.odc

```
v <- "four.state.forward"(v, v, v, s, s)
"MathRungeKutta45.Install;WBDiffFourStateForward
.Install"
```

The module four.state.forward can now be called by WinBUGS running via BlackBox. The WinBUGS code is

```
model{
#ODE Code: n.grid=no. time points & grid=time points solu-
tion required for;
#dim=no. equations to solve; n.par=no. parameters;
#init=initial condition for P(t)=I;
#origin=time origin=0 here; tol=accuracy required for
solution
```

```
solution[1:n.grid, 1:dim] <- four.state.for-
ward(init[1:dim],grid[1:n.grid],theta[1:n.par], origin, tol)
#Multinomial likelihood for observed data
for (i in 1:3){ r[i,1:4] ~ dmulti(pi[i,1:4],n[i])
for(j in 1:3){
#theta=vector of rate parameters. index maps between matrix
& vector indices
theta[index[i,j]]<-G[i,j]
#pi[i,j] solves the system of ODE's, t.ind is index of grid for
time of obsn
pi[i,j]<- solution[t.ind, index[i,j]]
#Save estimated probabilities for each specified time point of
interest
for (t in 1:n.grid){P[i,j,t]<-solution[t, index[i,j]] }
} #end loop over j
#Probabilities must sum to 1
pi[i,4]<-1 - sum(pi[i,1:3])
for (t in 1:n.grid){ P[i,4,t]<-1 - sum(P[i,1:3,t]) }
} #end loop over i
#Define zero probabilities, according to the model
pi[2,1]<- 0
for (j in 1:2){ pi[3,j]<-0 }
#Priors for the transition rate parameters
for (i in 1:3){
for (j in (i+1):4){ G[i,j] ~dgamma(.5,.5) }
G[i,i]<-sum(G[i,(i+1):4])
}
}
}
#DATA
list(r=structure(.Data=c(20,6,4,2,
0,5,5,3,
0,0,3,7), .Dim=c(3,4)), n=c(32,13,10),
dim=6,origin=0,tol=1.0E-4, init=c(1,0,0,1,0,1),n.par=6,
n.grid=3,grid=c(.25,1,2),t.ind=3,
index=structure(.Data=c(1,2,3,NA,
NA,4,5,NA,
NA,NA,6,NA), .Dim=c(3,4))
)
```

REFERENCES

1. Beck JR, Pauker S. The Markov process in medical prognosis. *Med Decis Making*. 1983;3:419–58.
2. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. *Health Econ*. 1983;13:322–38.
3. Hazen G. Preference factoring for stochastic trees. *Management Science*. 2000;46:389–403.
4. Hazen GB. Factored stochastic trees: a tool for solving complex temporal medical decision models. *Med Decis Making*. 1993;13:227–36.
5. Hazen GB. Stochastic trees: a new technique for temporal medical decision modeling. *Med Decis Making*. 1992;12:163–78.
6. Hazen GB, Pellissier JM. Recursive utility for stochastic trees. *Operations Research*. 1996;44:788–809.
7. Hazen GB, Pellissier JM, Souderpandian J. Stochastic-tree models in medical decision making. *Interfaces*. 1998;28:64–80.
8. Miller DK, Homan SM. Determining transition probabilities: confusion and suggestions. *Med Decis Making*. 1994;14:52–8.
9. Parmigiani G. *Modeling in Medical Decision Making: a Bayesian Approach*. New York: Wiley; 2002.
10. Craig BA, Sendi PP. Estimation of the transition matrix of a discrete-time Markov chain. *Health Econ*. 2002;11:33–42.
11. Craig BA, Fryback DG, Klein R, Klein BEK. A Bayesian approach to modelling the natural history of a chronic condition from observations with intervention. *Statistics in Medicine*. 1999;18:1355–71.
12. Sendi PP, Craig BA, Pfluger D, Gafni A, Bucher HC. Systematic validation of disease models for pharmacoeconomic evaluations. *Journal of Evaluation in Clinical Practice*. 1999;5:283–295.

13. Fryback DG, Chinnis JO, Ulvila JW. Bayesian cost-effectiveness analysis: an example using the GUSTO trial. *Med Decis Making*. 2001;17:83–97.
14. Grimmett GDR, Stirzaker DR. *Probability and Random Processes*. 2nd ed. Oxford: Oxford University Press, 1992.
15. Cinlar E. *Introduction to stochastic processes*. Englewood Cliffs (New Jersey): Prentice-Hall; 1975.
16. Maple 8. Available from: <http://www.maplesoft.com/products/Maple8/index.shtml>. 2002.
17. Congdon P. *Bayesian Statistical Modelling*. Chichester: Wiley; 2001.
18. Cox DR, Oakes D. *Analysis of Survival Data*. London: Chapman and Hall; 1984.
19. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York: Wiley, 1980.
20. Parmigiani G, Samsa GA, Ancukiewicz M, Lipscomb J, Hasselblad V, Matchar D. Assessing uncertainty in cost-effectiveness analyses: application to a complex decision model. *Med Decis Making*. 1997;17:390–401.
21. Ades AE, Cliffe S. Markov Chain Monte Carlo estimation of a multi-parameter decision model: consistency of evidence and the accurate assessment of uncertainty. *Med Decis Making*. 2002;22:359–71.
22. Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter evidence synthesis and consistency of evidence. *Statistics in Medicine*. 2003;22:2995–3016.
23. Eddy DM, Hasselblad V, Shachter R. *Meta-analysis by the Confidence Profile Method*. London: Academic Press; 1992.
24. Dilks DW, Canale RP, Meier PG. Development of Bayesian Monte Carlo techniques for water quality model uncertainty. *Ecological Modelling*. 1992;62:149–62.
25. Patwardhan A, Small MJ. Bayesian methods for model uncertainty analysis with application to future sea level rise. *Risk Analysis*. 1992;12:513–23.
26. Briggs AH, Ades AE, Price MJ. Probabilistic sensitivity analysis for decision trees with multiple branches: use of the Dirichlet distribution in a Bayesian framework. *Med Decis Making*. In press.
27. WinBUGS Differential Interface (WBDiff). Available from: http://homepages.tesco.net/~creeping_death/.
28. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society (B)*. 2002;64:1–34.
29. McCullagh P, Nelder JA. *Generalised Linear Models*. 2nd ed. London: Chapman and Hall; 1989.
30. Naglie G, Detsky AS. Treatment of chronic nonvalvular atrial fibrillation in the elderly: a decision analysis. *Med Decis Making*. 1992;12:239–49.
31. Gage BF, Cardinali AB, Albers GW, Owens DK. Cost-effectiveness of Warfarin and aspirin for prophylaxis of stroke in patients with nonvalvular atrial fibrillation. *JAMA*. 1995;274:1839–45.
32. Greenberg M, Malenka D, Disch D. Therapeutic strategies for atrial fibrillation: the value of decision analysis. *Cardiology Clinics*. 1996;14:623–40.
33. Fleiss JL. Measures of effect size for categorical data. In: Cooper H, Hedges LV, eds. *The Handbook of Research Synthesis*. New York: Russell Sage Foundation, 1994. p 245–60.
34. Andersen PK, Keiding N. Multi-state models for event history analysis. *Stat Methods Med Res*. 2002;11:91–115.
35. Keiding N. Event history analysis and inference from observational epidemiology. *Stat Med*. 1999;18(17-18):2353–63.
36. Mulrow CD. Rationale for systematic reviews. In: Chalmers I, Altman DG, eds. *Systematic Reviews*. London: BMJ Publishing Group, 1995. p 1–8.
37. Brand KP, Small MJ. Updating uncertainty in an integrated risk assessment: conceptual framework and methods. *Risk Analysis*. 1995;15:719–31.
38. Sohn M, Small M, Pantazidou M. Reducing uncertainty in site characterisation using Bayes Monte Carlo methods. *Journal of Environmental Engineering*. 2000;126:893–902.
39. Raftery AE, Givens GH, Zeh JE. Inference from a deterministic population dynamics model for Bowhead whales (with discussion). *Journal of the American Statistical Association*. 1995;90:402–30.
40. Brooks SP, Gelman A. Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*. 1998;7:434–55.