



Universiteit  
Leiden

The Netherlands

## A Markov approach to characterizing the PK-PD relationship of anti-migraine drugs

Maas, H.J.

### Citation

Maas, H. J. (2007, June 5). *A Markov approach to characterizing the PK-PD relationship of anti-migraine drugs*. Retrieved from <https://hdl.handle.net/1887/12040>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/12040>

**Note:** To cite this publication please use the final published version (if applicable).

# A

## **Appendix: Analysis of responses in migraine modelling using hidden Markov models**

*V Anisimov, HJ Maas, M Danhof, OE Della Pasqua  
Published online in Stat. Med. (2007)*

*Markov-type models have been used in the analysis of disease progression. Although standard errors of model parameters are usually estimated, available software often does not permit the construction of confidence intervals around predictions of the dependent or response variable. A method is presented to calculate means and confidence intervals of model-predicted responses in time governed by a non-homogeneous hidden Markov model in continuous time. The Kolmogorov equations serve as the basis for the calculations. The method is realised in S-Plus and is applied to the prediction of headache responses in clinical studies of anti-migraine treatment. Means and confidence intervals are calculated by numerically solving differential equations that are nonlinear in the explanatory variable. Results indicate that uncertainty on predicted drug responses is larger than that on predicted placebo responses and that pain-free responses are less precisely predicted than pain relief responses. This is due to the uncertainty in the drug-specific parameters which is not present in predicted placebo responses.*

## A.1 Introduction

The outcome of clinical studies into paroxysmal diseases and disease progression is typically characterised by categorical data and multiple measurements. Acknowledging the existence of onset and offset features in most diseases, it has been shown that these data are suitably analysed using Markov model approaches. Markov analysis avoids the difficulty of making multiple comparisons between time points. Assessment of the efficacy of anti-migraine drugs often requires comparison of responses at various discrete times [1].

Over the years, a considerable effort has been put into making available software packages for Markov model analysis. While differing in flexibility and the numbers of modelling options offered, they all provide estimates of parameter precision. It usually involves standard errors calculated from the asymptotic variance-covariance matrix, which is obtained by the inverse matrix of second-order derivatives of the objective function evaluated at the maximum likelihood estimates.

Though this is insightful from the modeller's point of view, the results are often communicated graphically, plotting the modelled dependent variable (usually referred to as the response) versus time. This representation most clearly conveys the essence of disease progression, including effects that can be expected from any medical intervention.

Confidence intervals should be a part of this representation. Firstly, because it is of interest to know to what extent (a lack of) precision in the individual parameters affects the range of response predicted by the model. The narrower the intervals, the higher the precision in the estimates. Secondly, because the width of the interval is indicative of the chance of detecting a significant treatment effect under the null hypothesis.

Calculating confidence intervals for responses is a feature that is not usually present in the Markov software packages. It is true that confidence intervals for a single model parameter can be easily constructed assuming asymptotic normality. For simple functions of a parameter, such as the odds ratio, intervals can be readily approximated by scaling [2]. However, this paper addresses the calculation of confidence intervals for a special function which depends on the trajectory of a hidden Markov chain. Calculation of the errors of this function is not straightforward and to our knowledge has not been demonstrated in the literature.

The rationale for presenting this method is to enable fast and routine calculation of confidence intervals. The tools available for the derivation of confidence intervals for Markov models are limited. Only one paper was found in which confidence intervals were constructed for Markovian variables [3]. In that paper, Monte Carlo simulation of a Markov model and a linear approximation to this model were used to obtain confidence intervals. The first of these approaches is computationally expensive, the second requires that a new approximation model be built for every new or modified Markov model. Both methods are therefore not convenient for the routine evaluation of Markov models.

Our method uses the following approach to obtain mean responses and their confidence intervals by numerical calculation:

1. An S-Plus procedure is applied to obtain parameter estimates and standard errors for a hidden Markov model based on series of treatment responses observed over

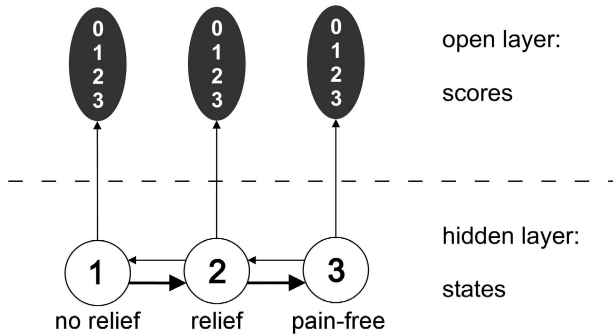
time.

2. The Kolmogorov differential equations are taken for the transition probabilities of a non-homogeneous Markov model.
3. A system of differential equations is derived for the derivatives of transition probabilities with respect to the parameters defining them. These are similar to the Kolmogorov differential equations.
4. Replacing the parameters with their estimates, a numerical algorithm calculates the transition probabilities from (2) and their derivatives from (3).
5. Mean responses over time are calculated using the transition probabilities. Using the delta method it is shown that the confidence intervals can be constructed using the derivatives of transition probabilities and the variance-covariance matrix of the estimates. Calculations were performed in a numerical S-Plus procedure.

The delta method approximates the expected values of functions of random variables when direct evaluation is not feasible [4]. The approximation usually is a truncated Taylor series centered at the mean of the variables. In the current application the functions are the transition probabilities of the Markov chain.

The approach can be applied to both homogeneous and non-homogeneous Markov models and hidden Markov models in continuous time. Non-homogeneity in these models is a feature which is necessary for describing the time-dependent effects of medical interventions on disease progression. These effects may be both linear or nonlinear in nature. Hidden Markov models [5] differ from their regular counterparts in that they contain a layer for observations that are conditionally dependent on the states in the underlying unobserved Markov chain. The observational layer may be used to account for the misclassification of scores or simply to cluster a number of scores into a smaller number of states.

Headache relief and headache-free status are the most important endpoints measured in clinical trials investigating the efficacy of serotonin receptor agonists to abort headache during migraine attacks. In this paper, an example is presented in which the algorithm is applied to calculate predicted means and confidence intervals for headache responses. Parameter estimates which are at the basis of these predictions are obtained by fitting headache scores derived from efficacy trials to a three-state hidden Markov model. Blood concentrations of anti-migraine drugs (triptans) are included in the fitting procedure as a covariate explaining the difference between the placebo response and the response after drug treatment. The hidden Markov model consists of i) a hidden layer representing the clinical state of a patient's migraine attack, which can be described as "no relief", "relief" or "pain-free", and ii) an open or observational layer that includes the distributions of headache scores for each state of the hidden layer (Figure A.1) Headache scores reflect pain intensity and can assume the values 0, 1, 2, 3 in increasing order of intensity. Input arguments for the algorithm include the parameter estimates, the elements of the asymptotic variance-covariance matrix corresponding to the transition rate matrix, covariate values and sequences of time-points for which predicted headache responses are



**Figure A.1:** Hidden Markov model for migraine attacks. Bold arrows denote transitions that are affected by the covariate (drug concentration).

desired. The algorithm then applies the Kolmogorov differential equations to the input arguments in order to obtain mean responses. In parallel to this, variances of the mean responses are calculated by applying the delta method to the transition probabilities.

The results show that the prediction of the mean responses obtained with this approach agrees well with that obtained using Monte Carlo simulations. Using the delta method, confidence intervals could be calculated in a matter of minutes. The speed of this method mainly depends on the step-size of the iterations. Monte Carlo simulations were performed in several hours. Simulation speed depends both on the step-size and the number of samples drawn from the parameter space.

Confidence intervals calculated with this method clearly reflect the amount of uncertainty expressed by the standard errors. This is demonstrated by comparing the width of the 95% confidence intervals for placebo responses with that for the drug-induced responses. The relatively large uncertainty in the drug-related parameters is translated into wide intervals around drug-induced response, whereas the intervals around placebo responses (which are not influenced by drug) are narrow.

The procedure of estimating confidence intervals uses estimators of the parameters of a hidden Markov model. For this purpose we use an S-Plus algorithm developed by Bureau *et al.* [6] [2] and later adapted for applications in migraine modelling [7]. For general information on this software, we recommend [8] and [9]. Bureau *et al.* describe a hidden Markov model applied to longitudinal measurements of binary disease outcome. Disease dynamics are characterised by a two-state Markov process the states of which denote the actual presence and absence of disease. A second layer gives the distributions of the measured responses conditional on the (latent) states. The structure allows taking into account the misclassification of disease outcomes due to measurement error.

Covariates on the transitions in the Markov process and on the distribution of responses are accommodated in a generalised regression framework. The papers demonstrate the inclusion of covariates on the Markov transitions. The current paper differs from [6] and [2] in the following aspects: the functions used for the regression of covariates are more complex, the Markov process contains three states instead of two and the purpose of the conditional distributions of the observed response is to cluster scores rather than to account for measurement error. Output from “hmm” fitting procedures can be directly used for predicting mean responses and confidence intervals.

## A.2 Methods

### A.2.1 Model

A hidden Markov model is developed to analyse headache intensity data derived from three clinical efficacy studies on the anti-migraine drug sumatriptan. Headache intensity is expressed on a four-point scale on which 0, 1, 2, 3 represent no pain, mild pain, moderate pain and severe pain, respectively. Patients enter a study when their migraine headache is at its worst and at least of moderate intensity. Assessments of headache are performed at various times after oral administration of placebo or active treatment (25, 50 or 100 mg of sumatriptan). The model is applied to predict the time course of the clinical endpoints “pain-free status” and “pain relief”. These endpoints are composites of the headache intensities. A patient is pain-free when headache intensity has decreased to 0 starting at 2 or 3. Pain relief is defined as headache intensity of at most 1, starting from 2 or 3.

Based on the aforementioned definitions, a three-state hidden Markov model is developed (Figure A.1) in which the hidden states represent the clinical state (state 1 no relief, state 2 relief, state 3 pain-free). The observed layer contains the headache intensity scores which are assumed to be distributed multinomially and are conditional on the hidden states. Starting in state 1 with a score of either 3 or 2, patients progress over time to the states with less severe headache scores (rightward transitions). However, we also assume that there is a possibility of temporary worsening of the status of the patient, that means, we also allow leftward transitions from states with less severe headache scores to the states with more severe scores. Finally, we assume that the model has a structure of birth-and-death process, which means that direct transitions from state 1 to state 3 and *vice versa* are not allowed.

The rates of transition towards states of less headache are assumed to be affected by sumatriptan concentrations. Since sumatriptan concentration data are not available for most patients in the efficacy studies, concentration data from early phase studies of sumatriptan are used to construct concentration versus time profiles. These profiles are obtained using the nonlinear mixed effects modelling program NONMEM V (Globomax LLC, MD, USA) [10], which allows prediction of so-called population-based pharmacokinetic parameters.

Using this method, the time course of drug concentration in plasma is modelled ac-

ording to a two-compartment pharmacokinetic model with combined first-order and zero-order drug input rates, as described by Cosson *et al.* [11]. The first-order absorption rate represents the initial absorption process starting  $\sim 15$  minutes after dosing and accounts for most of the drug absorbed. Utilising the pharmacokinetic parameter estimates, concentration versus time profiles are simulated and added to the headache intensity data and treated as a covariate that changes over time. The addition of a time-varying covariate renders the Markov chain non-homogeneous.

The use of pharmacokinetic data from healthy volunteers to predict concentration-effect relations in patients requires some justification. It is known that the absorption kinetics in patients having a migraine attack may be altered. In particular, gastric stasis delays the uptake of drug into the bloodstream in some patients. Comparisons between the pharmacokinetics in patients and those in volunteers are generally complicated due to infrequent sampling in patients. A pharmacokinetic analysis investigating absorption kinetics concluded that, apart from a delayed start of the zero-order absorption, absorption parameters were very similar between the two groups [11]. Considering that the initial rate of absorption does not differ between populations and that the rate, not the extent of absorption is the main determinant of the onset of response [12], it was assumed that using pharmacokinetics from volunteers would not change the concentration-effect relationships.

We assume that in the case of  $x < y$ , the time-dependent transition rates from state  $x$  to state  $y$ ,  $r_{x,y}(t)$  are promoted by triptan drugs in a concentration-dependent way:

$$r_{x,y}(t) = r_{x,y}(0) \exp\left(\frac{E_{max,xy}C(t)}{EC_{50,xy} + C(t)}\right). \quad (\text{A.1})$$

In this relation, drug concentration  $C(t)$  is incorporated into the model's covariate structure. Parameters in the transition rate matrix include baseline rates  $r_{x,y}(0)$  and drug-related parameters  $E_{max,xy}$  and  $EC_{50,xy}$ .  $E_{max,xy}$  denotes the maximum increase in transition rate  $r_{x,y}(t)$  that can be achieved by the drug.  $EC_{50,xy}$  is the concentration at which half of the maximum effect is reached and is a measure of drug potency on the transition  $r_{x,y}(t)$ .

The general form of relation is that of a hazard rate. As suggested by Cox [13], the expression in the exponent can be any known function. We use a non-linear function of variable  $C(t)$ . In pharmacology, this expression is known as the  $E_{max}$  function [14]. It is based on the empirical observation that the concentration-response relation is a hyperbola which is bounded at the top by the maximum pharmacological response. Relation (A.1) was found to better describe the drug effect than a linear expression in the exponent.

When  $x > y$ ,  $r_{x,y}(t)$  is given by:

$$r_{x,y}(t) = r_{x,y}(0). \quad (\text{A.2})$$

(A.1) and (A.2) were chosen based on the observation that with increasing dose of a triptan drug the headache pain decreases. In theory, this observation can be modelled by 1) increasing the forward transition rate in the Markov chain as a function of drug concentration, 2) decreasing the backward transition rate in the Markov chain as a function

of drug concentration or 3) a combination of both effects. The first option was chosen mainly because this mechanism is medically well-supported: triptan drugs promote pain relief by inhibiting pain-producing neural mechanisms. Furthermore, assuming a drug action on the backward transition rate of the Markov chain does not permit the description of rapid pain relief, which is often observed in migraine patients. Assuming drug effects on both the forward and backward transition rates is not feasible. Due to overparameterisation, the model parameters become unidentifiable.

According to our assumptions,  $r_{1,3}(\cdot) = r_{3,1}(\cdot) = 0$ .

Data from three clinical migraine studies are used to estimate these parameters. In addition, the distribution of scores in each of the states is assessed. Since all patients enter the studies at moderate or severe headache levels (*i.e.* “no relief” state), the initial distribution over states is fixed. Headache intensity score is treated as the dependent variable and sumatriptan concentration is treated as a time-varying covariate affecting headache intensity. The clinical studies consists of a placebo and a drug arm, the latter including three dose groups. In total, data are available from 1180 migraine attacks. 392 attacks were treated with placebo whereas 46, 44 and 698 attacks were treated with single doses of 25 mg, 50 mg and 100 mg oral sumatriptan, respectively. Headache assessments were made at 0, 0.5, 1.0, 1.5, 2.0, 4.0, 8.0, 12 and 24 hours after dosing. Data up to 8.0 hours are included in the analysis. Firstly, because an accurate prediction of headache responses is therapeutically most relevant at early time points. Secondly, because the assumption underlying the model, the Markov property, is only valid for a particular time interval [15]. After 8.0 hours the Markov property appears not tenable without losing accuracy at earlier times. In fact, later sampling times were captured to understand recurrence of headache, rather than to directly assess efficacy. Table A.1 summarises estimates and standard errors (se) of the parameters in the transition rate matrix and the estimates of the distribution of scores in the observational layer in different states. Standard errors are square roots of the diagonal elements of the asymptotic variance-covariance matrix which equals the inverted information matrix of the observed data. This matrix requires the calculation of second derivatives of the matrix of transition probabilities.

Parameter estimation is performed by the Expectation Maximisation (EM) algorithm using an open-code *C* program which is operated within S-Plus on a linux workstation (SuSE Linux 7.2 Professional, kernel version 2.4.4-4GB-SMP). A user-written model defining the covariate effect and the first and second derivatives of the transition matrix is implemented in the model specification file. A generalised regression framework allows for linear and nonlinear regression of continuous covariates as well as for regression of categorical covariates. Implementation of a new model requires coding the regression function and its first and second partial derivatives. Both time-invariant and time-varying covariates can be dealt with. Between two adjacent observations, a time-varying covariate takes on the value associated with the later observation. In the current nonlinear covariate model (A.1), the parameter  $EC_{50}$  is estimated as its natural logarithm,  $\ln(EC_{50})$ . This transformation is applied because  $EC_{50}$  values are known to be log-normally distributed. This is in contrast with  $E_{max}$  values, which are normally distributed [16]. Further details on parameter estimation in non-homogeneous hidden



**Table A.1:** parameter estimates of the transition rate matrix and distribution of scores in the HMM. Estimates were obtained by the EM algorithm.

transition	parameter (unit)	estimate	se
1-2	$r_{1,2}(0)$ ( $h^{-1}$ )	0.16	0.07
	$\ln(EC_{50})$	2.2	0.60
	$E_{max}$	1.3	0.15
2-1	$r_{2,1}(0)$ ( $h^{-1}$ )	0.08	0.16
2-3	$r_{2,3}(0)$ ( $h^{-1}$ )	0.22	0.09
	$\ln(EC_{50})$	5.0	0.62
	$E_{max}$	6.0	2.8
3-2	$r_{3,2}(0)$ ( $h^{-1}$ )	0.04	0.17

state	1	2	3
score			
no pain	0.0	0.0	1.0
mild pain	0.0	0.96	0.0
moderate pain	0.55	0.04	0.0
severe pain	0.45	0.0	0.0

Markov models in continuous time can be found in papers by Bureau *et al.* [2, 6]. These include a description of the likelihood function and the EM algorithm.

## A.2.2 Predictions

We consider the 8 parameters defining the time-dependent transition rates of the intensity matrix as the main parameters of the model (Table A.1). These parameters are contained in the vector  $\theta$ . The EM algorithm gives as output a) the estimated elements of  $\hat{\theta}$  and b) the variance-covariance matrix of  $\hat{\theta}$ . We use these estimators to calculate the predicted state probabilities of the HMM, the predicted mean headache responses and corresponding confidence intervals at any time point  $t$ . For this purpose, a newly written S-Plus routine is created. Central to this routine are Kolmogorov's differential equations which translate the transition rates of a HMM into transition probabilities.

## A.2.3 Mean responses

Consider a Markov process, non-homogeneous in time,  $X(t)$  with states  $\{1, 2, 3\}$  and transition rates at time  $t$ ,  $r_{x,y}(\theta, t)$  from state  $x$  to state  $y$ . Denote by  $P(x, y, \theta, t)$  a transition probability on time interval  $[0, t]$ :

$$P(x, y, \boldsymbol{\theta}, t) = P(X(t) = y | X(0) = x), \quad x, y = 1, 2, 3.$$

Then  $P(x, y, \boldsymbol{\theta}, t)$  can be calculated by solving the system of Kolmogorov's differential equations in the non-homogeneous case. Denote by  $\mathbf{p}(x, \boldsymbol{\theta}, t)$  the row vector with components  $P(x, y, \boldsymbol{\theta}, t)$ ,  $y = 1, 2, 3$ , and define the transition rate matrix by

$$\mathbf{R}(\boldsymbol{\theta}, t) = \begin{pmatrix} -r_1(\boldsymbol{\theta}, t) & r_{1,2}(\boldsymbol{\theta}, t) & r_{1,3}(\boldsymbol{\theta}, t) \\ r_{2,1}(\boldsymbol{\theta}, t) & -r_2(\boldsymbol{\theta}, t) & r_{2,3}(\boldsymbol{\theta}, t) \\ r_{3,1}(\boldsymbol{\theta}, t) & r_{3,2}(\boldsymbol{\theta}, t) & -r_3(\boldsymbol{\theta}, t) \end{pmatrix},$$

where  $r_x(\boldsymbol{\theta}, t) = \sum_{y \neq x} r_{x,y}(\boldsymbol{\theta}, t)$ . Then, for any initial state  $x$ , the vector  $\mathbf{p}(x, \boldsymbol{\theta}, t)$  is a solution of the system of Kolmogorov's differential equations

$$\frac{\partial}{\partial t} \mathbf{p}(x, \boldsymbol{\theta}, t) = \mathbf{p}(x, \boldsymbol{\theta}, t) \mathbf{R}(\boldsymbol{\theta}, t) \quad (\text{A.3})$$

with the initial condition  $\mathbf{p}(x, \boldsymbol{\theta}, 0) = e_x$ , where  $e_x$  is a row vector with the  $x$ -th entry equal to one and other entries equal to zeros. This system can be solved using a quick recurrent numerical procedure. Take for  $h$  some small value (in our calculations  $h = 0.001$  is sufficiently small). Denote by  $\hat{P}(x, y, \boldsymbol{\theta}, kh)$  the calculated value of  $P(x, y, \boldsymbol{\theta}, t)$  at the point  $t = kh$ . For any initial state  $x$ , let  $\hat{\mathbf{p}}(x, \boldsymbol{\theta}, kh)$  be the row vector with components  $\hat{P}(x, y, \boldsymbol{\theta}, kh)$ ,  $y = 1, 2, 3$ . To find a numerical solution of (A.3), we use the following algorithm: at  $k = 0$  start with  $\hat{\mathbf{p}}(x, y, \boldsymbol{\theta}, kh) = \mathbf{p}(x, \boldsymbol{\theta}, 0)$ . Then for any  $k \geq 0$ ,

$$\hat{\mathbf{p}}(x, \boldsymbol{\theta}, (k+1)h) = \hat{\mathbf{p}}(x, \boldsymbol{\theta}, kh) + h \hat{\mathbf{p}}(x, \boldsymbol{\theta}, kh) \mathbf{R}(\boldsymbol{\theta}, kh).$$

To predict the mean response, the pain scores conditional on the states of the Markov chain should also be taken into account. In the following, this is done for the pain relief response, thus including all scores 0 and 1. Given the initial state  $x$ , the predicted mean fraction of patients  $\hat{M}$  with responses  $\{0, 1\}$  at time point  $kh$  is:

$$\hat{M}(x, kh, \{0, 1\}) = \sum_{y=1}^3 \hat{P}(x, y, \boldsymbol{\theta}, kh) \sum_{j=0}^1 \hat{p}_y(j), \quad (\text{A.4})$$

where  $\hat{p}_y(j)$ , with  $j = 0, 1, 2, 3$ , are the probabilities of the scores in state  $y$  which are estimated together with the Markov transition rate parameters in the EM algorithm. A total mean taking into account the initial distribution  $\pi(x)$ ,  $x = 1, 2, 3$ , is

$$\hat{M}(kh, \{0, 1\}) = \sum_{x=1}^3 \pi(x) \hat{M}(x, kh, \{0, 1\}). \quad (\text{A.5})$$

If the initial conditions are fixed to  $\mathbf{p}(x, \boldsymbol{\theta}, 0) = (1, 0, 0)$ , as is the case in the migraine attack application, then  $\pi(1) = 1$  and  $\pi(2) = \pi(3) = 0$ .

### A.2.4 Confidence intervals

The procedure for calculating the confidence intervals on the mean predicted responses involves the delta method which approximates the variance of a nonlinear function of a random variable. The basic idea is to approximate the function of the estimator with a first-order Taylor approximation. The variance can then be calculated using standard rules.

#### Calculation of confidence intervals for the nonlinear function of unknown parameters

Let  $f(\boldsymbol{\theta})$  be some nonlinear differentiable function of the vector of parameters  $\boldsymbol{\theta}$ . Assume that the estimator  $\widehat{\boldsymbol{\theta}}$  of a vector  $\boldsymbol{\theta}$  has the properties:

$$\mathbf{E}[\widehat{\boldsymbol{\theta}}] \approx \boldsymbol{\theta}, \quad \mathbf{E}[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})] \approx B^2, \quad (\text{A.6})$$

where  $B^2$  is the asymptotic variance-covariance matrix corresponding to the parameters in the transition rates only.  $B^2$  is assumed to be reasonably small. Symbol  $\mathbf{a}^T$  denotes a column vector transposed to  $\mathbf{a}$ . Using the delta method we show how to construct the approximation of the variance of  $f(\widehat{\boldsymbol{\theta}})$  and its confidence intervals. Denote

$$f'_i(\boldsymbol{\theta}) = \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_i}, \quad i = 1, 2, \dots, r.$$

Let  $\nabla f(\boldsymbol{\theta})$  be the row vector with components  $(f'_i(\boldsymbol{\theta}), i = 1, 2, \dots, r)$ . Using Taylor expansion up to the first term we get

$$f(\widehat{\boldsymbol{\theta}}) \approx f(\boldsymbol{\theta}) + \nabla f(\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T = f(\boldsymbol{\theta}) + \sum_i f'_i(\boldsymbol{\theta})(\widehat{\theta}_i - \theta_i), \quad (\text{A.7})$$

where  $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \dots, \widehat{\theta}_r)$ . Denote by  $\sigma_i^2$  the diagonal elements of  $B^2$  and by  $\rho_{ij}$  all other elements. Using (A.7) we get:

$$\begin{aligned} \mathbf{E}[f(\widehat{\boldsymbol{\theta}})] &\approx f(\boldsymbol{\theta}), \\ \mathbf{Var}[f(\widehat{\boldsymbol{\theta}})] &\approx \nabla f(\boldsymbol{\theta}) B^2 \nabla f(\boldsymbol{\theta})^T = \sum_i f'_i(\boldsymbol{\theta})^2 \sigma_i^2 + 2 \sum_{i < j} f'_i(\boldsymbol{\theta}) f'_j(\boldsymbol{\theta}) \rho_{ij}. \end{aligned} \quad (\text{A.8})$$

Thus, the asymptotic  $(1 - \alpha)100\%$  confidence interval for  $f(\boldsymbol{\theta})$  can be calculated as follows:

$$\left( f(\widehat{\boldsymbol{\theta}}) - z_{1-\alpha/2} \sqrt{\mathbf{Var}[f(\widehat{\boldsymbol{\theta}})]}, f(\widehat{\boldsymbol{\theta}}) + z_{1-\alpha/2} \sqrt{\mathbf{Var}[f(\widehat{\boldsymbol{\theta}})]} \right), \quad (\text{A.9})$$

where  $z_{1-\alpha}$  is the  $(1 - \alpha)$ -quantile of the standard normal distribution:  $\mathbf{P}(\mathcal{N}(0, 1) \leq z_{1-\alpha}) = 1 - \alpha$ .

### Calculation of confidence intervals for mean responses

Given the initial state  $x$ , the mean fraction of patients with responses  $\{0, 1\}$  at time point  $t$  is:

$$M(t, \boldsymbol{\theta}, \{0, 1\}) = \sum_{x=1}^3 \pi(x) \sum_{y=1}^3 P(x, y, \boldsymbol{\theta}, kh) \sum_{j=0}^1 \hat{p}_y(j). \quad (\text{A.10})$$

Using formula (A.8) we get

$$\mathbf{Var} \left[ M(t, \hat{\boldsymbol{\theta}}, \{0, 1\}) \right] = \nabla M(t, \boldsymbol{\theta}, \{0, 1\}) B^2 \nabla M(t, \boldsymbol{\theta}, \{0, 1\})^T, \quad (\text{A.11})$$

where  $\nabla M(t, \boldsymbol{\theta}, \{0, 1\})$  is a row vector with entries  $M'_i(t, \boldsymbol{\theta}, \{0, 1\}) = \frac{\partial}{\partial \theta_i} M(t, \boldsymbol{\theta}, \{0, 1\})$ ,  $i = 1, 2, \dots, r$  (which are the partial derivatives of  $M(\cdot)$  with respect to  $\theta_i$ ) and  $B^2$  is the estimated variance-covariance matrix of the parameters  $\boldsymbol{\theta}$  included in the transition rate matrix. From (A.10) it follows

$$M'_i(t, \boldsymbol{\theta}, \{0, 1\}) = \sum_{x=1}^3 \pi(x) \sum_{y=1}^3 P'_i(x, y, \boldsymbol{\theta}, t) \sum_{j=0}^1 p_y(j). \quad (\text{A.12})$$

Thus, to compute  $\mathbf{Var} \left[ M(t, \hat{\boldsymbol{\theta}}, \{0, 1\}) \right]$  we need to compute the functions  $P'_i(x, y, \boldsymbol{\theta}, t)$ , which, for convenience, will be renamed  $\Psi_i(x, y, \boldsymbol{\theta}, t)$ . Note that  $\Psi_i(x, y, \boldsymbol{\theta}, t)$  is the derivative of  $P(x, y, \boldsymbol{\theta}, t)$  with respect to  $\theta_i$ . As at  $t = 0$  the function  $P(x, y, \boldsymbol{\theta}, 0)$  does not depend on  $\boldsymbol{\theta}$  and  $P(x, y, \boldsymbol{\theta}, 0) = \delta_x(y)$ , where  $\delta_x(y) = 1$ , if  $x = y$ , and  $\delta_x(y) = 0$  otherwise, we get the initial conditions for  $\Psi_i(x, y, \boldsymbol{\theta}, t)$ :

$$\Psi_i(x, y, \boldsymbol{\theta}, 0) = 0, x, y = 1, 2, 3, i = 1, 2, \dots, r.$$

Using again Kolmogorov's differential equations for the transition probabilities  $P(x, y, \boldsymbol{\theta}, t)$ , we now derive differential equations for the functions  $\Psi_i(x, y, \boldsymbol{\theta}, t)$ . Denote by  $\mathbf{p}(x, \boldsymbol{\theta}, t)$  and  $\boldsymbol{\psi}_i(x, \boldsymbol{\theta}, t)$  the row vectors with entries  $P(x, y, \boldsymbol{\theta}, t)$  and  $\Psi_i(x, y, \boldsymbol{\theta}, t)$ . Taking the derivative of both parts in equation (A.3) with respect to  $\theta_i$  we obtain the following equation:

$$\frac{\partial}{\partial t} \boldsymbol{\psi}_i(x, \boldsymbol{\theta}, t) = \boldsymbol{\psi}_i(x, \boldsymbol{\theta}, t) R(\boldsymbol{\theta}, t) + \mathbf{p}(x, \boldsymbol{\theta}, t) \frac{\partial}{\partial \theta_i} R(\boldsymbol{\theta}, t). \quad (\text{A.13})$$

Using the formulae for the rates  $r_{x,y}(t)$  the functions  $\frac{\partial}{\partial \theta_i} R(\boldsymbol{\theta}, t)$  can be calculated in a closed form. The system of equations (A.13) can be solved numerically and simultaneously with (A.3). The numerical procedure returns the values  $\hat{\Psi}_i(x, y, \boldsymbol{\theta}, kh)$ , enabling the calculation of  $\mathbf{Var} \left[ \hat{M}(kh, \boldsymbol{\theta}, \{0, 1\}) \right]$ .

To assure that confidence intervals on responses are confined between 0 and 1, a logit transformation is applied on  $\hat{M}(kh, \boldsymbol{\theta}, \{0, 1\})$  before calculating the confidence interval.

This transformation is of the form  $\ln(\hat{M}/(1 - \hat{M}))$  (omitting the arguments for simplicity). The logit function is the most commonly used link function. In the range 0.1 – 0.9 it is almost indistinguishable from the probit function. It is symmetric, mathematically tractable and can be interpreted as a log odds ratio. For these reasons, a logit transformation was preferred over a probit transformation and no other link functions were considered. The transformed variance is approximated as follows:

$$Var[\widehat{M}_{logit}] \approx Var[\hat{M}] \left( \hat{M}(1 - \hat{M}) \right)^{-2}.$$

Denoting the transformed function  $\widehat{M}_{logit}(kh, \boldsymbol{\theta}, \{0, 1\})$ , the interval is given by expression (A.14):

$$\left( \widehat{M}_{logit}(kh, \boldsymbol{\theta}, \{0, 1\}) - z_{1-\alpha/2} \sqrt{Var[\widehat{M}_{logit}(kh, \boldsymbol{\theta}, \{0, 1\})]}, \right. \\ \left. \widehat{M}_{logit}(kh, \boldsymbol{\theta}, \{0, 1\}) + z_{1-\alpha/2} \sqrt{Var[\widehat{M}_{logit}(kh, \boldsymbol{\theta}, \{0, 1\})]} \right), \quad (A.14)$$

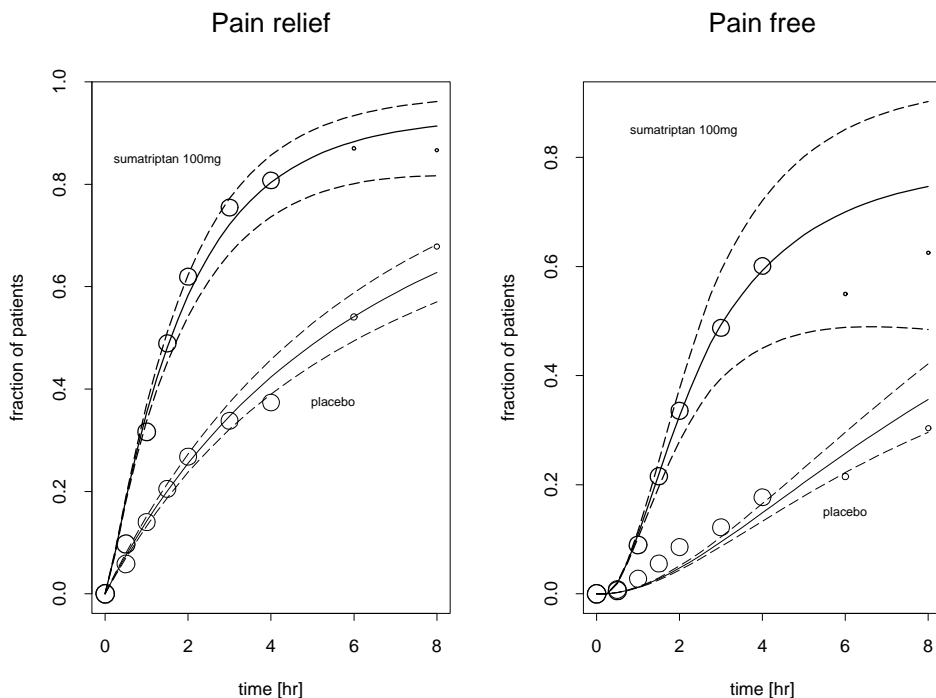
where  $z_{1-\alpha}$  is the  $(1 - \alpha)$ -quantile of the standard normal distribution:  $P(N(0, 1) \leq z_{1-\alpha}) = 1 - \alpha$ . Mean responses and intervals were then back-transformed to obtain probabilities.

### A.2.5 Performance of the algorithm

It is tested how well the confidence intervals calculated by the new algorithm agree with responses generated by the standard Monte Carlo simulation algorithm in the S-Plus module. The latter generates headache scores on the basis of the parameter estimates and a random number generator. The Monte Carlo simulations of the response versus time profiles are generated based on 1000 samples. Pain scores are simulated with between-observation intervals of 0.001 hours. Calculations are performed to assess both pain-free and pain relief responses after administration of placebo and 100 mg sumatriptan.

## A.3 Results

Based on the parameter estimates in Table A.1, the time course of mean headache relief and mean pain-free status (Figure A.2) is predicted for placebo treatment and 100 mg sumatriptan, until 8.0 hours after dosing. Predicted time courses of pain relief and pain free status are compared with those found in the original data (markers in Figure A.2). Mean responses and 95% confidence intervals are calculated using the new algorithm. A Monte Carlo simulation was performed taking into account variability in the parameters of the transition matrix. Figure A.3 compares the analytically calculated confidence intervals for placebo and sumatriptan and for two endpoints, pain-free and pain relief. There is a good agreement among confidence intervals for all placebo predictions and for the pain relief predictions. Only in the case of pain-free predictions for a high dose



**Figure A.2:** Hidden Markov model predictions of headache response after treatment with placebo or sumatriptan 100 mg: mean pain relief response (left panel), mean pain-free response (right panel), and 95% confidence intervals. Markers denote observed responses. The diameters of the markers reflect the numbers of patients included at consecutive time points. Means and 95% confidence intervals were calculated using the algorithm based on the Kolmogorov equations described in this paper.

of sumatriptan the agreement was not good. This is due to the rather large standard errors associated with parameters  $EC_{50,23}$  and  $Emax_{23}$ . Because less data were available to estimate parameters for this transition, we examined the behaviour of the confidence intervals in the hypothetical case where more patients are available to estimate the parameters (Figure A.3, lower panel). If 50% more patients would be added, then there would be a better agreement between the confidence intervals. It can also be observed that the interval based on analytical calculation is generally wider than that obtained by simulation. Thus, for a large enough number of patients, the analytically constructed intervals are more conservative.

The 95% confidence intervals obtained from calculations with the delta method are compared with those generated by a Monte Carlo simulation routine. Simulation conditions are as described in section A.2.5. The results from the two methods are broadly similar although the delta method tends to give narrower boundaries especially at early

time points (Figure A.3).

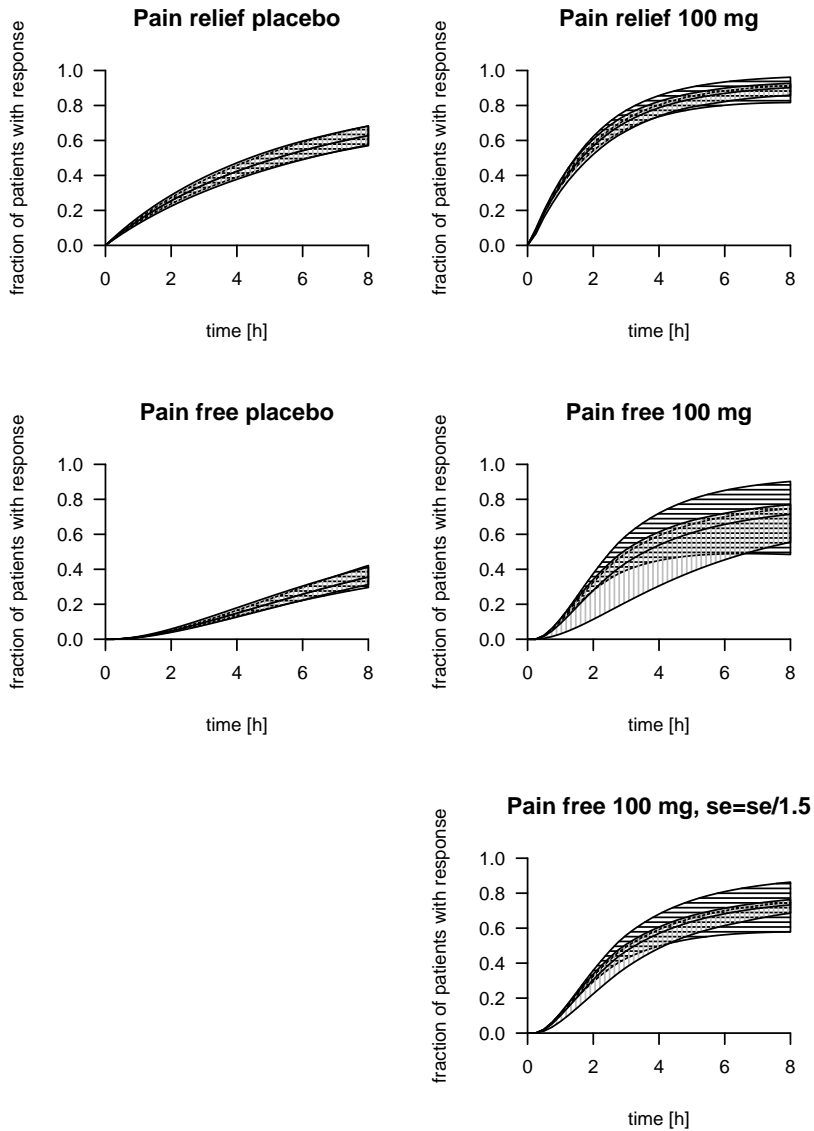
Figure A.4 illustrates the use of 95% confidence intervals in concentration versus response relationships. For three fixed times (1, 2 and 4 hours) after dose administration, pain relief responses are plotted that correspond to various oral doses of sumatriptan. These graphs are constructed by i) calculating mean placebo responses and confidence intervals for a sequence of time points including the times used in the graphs, ii) repeating these calculations for a large number of doses, iii) extracting from these calculations the concentrations and responses (including intervals) corresponding to the times of interest (1, 2 and 4 hours), iv) subtracting the placebo responses from the drug responses at these times and v) plotting for each time the pairs of concentrations and responses. The width of confidence intervals increases with time.

## A.4 Discussion

This paper describes an algorithm for the calculation of mean predicted headache responses as well as confidence intervals. These predictions are based on a hidden Markov model in continuous time. Due to time-varying drug concentrations, the Markov model is non-homogeneous. The method is based on i) Kolmogorov's differential equations which are used to obtain state-to-state transition probabilities from transition rates and ii) the delta method for approximating the derivatives of nonlinear functions with respect to their parameters. The algorithm provides a fast solution for obtaining confidence boundaries which clearly reflect the uncertainty in the parameter estimates as given by their standard errors.

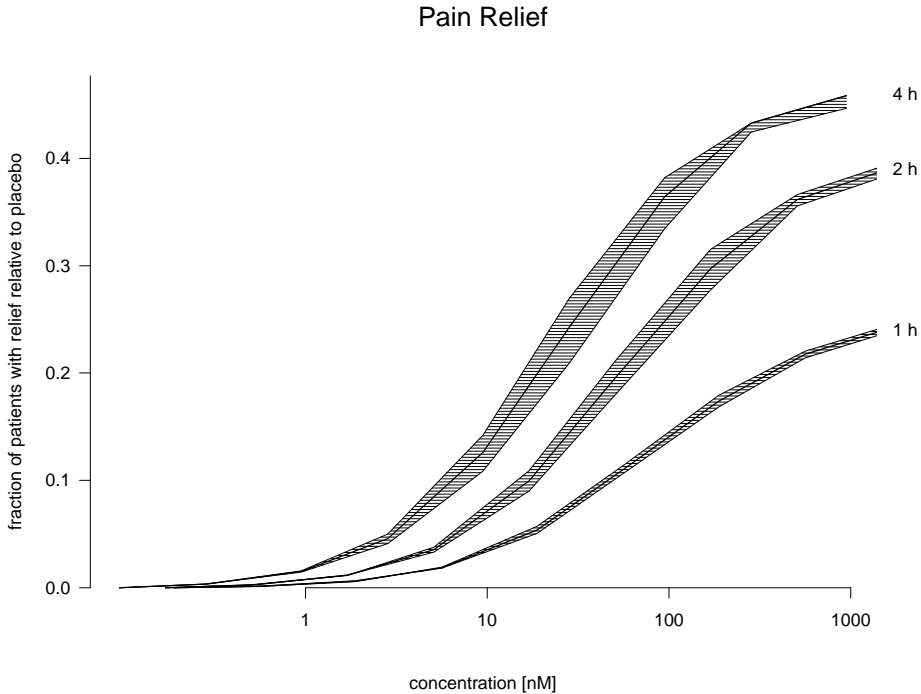
The hidden Markov model adequately predicts headache responses. Figure A.2 indicates that the mean predicted pain relief and pain-free responses largely correspond to the observed responses. After 2 hours, the 95% confidence interval on the pain-free response after sumatriptan is relatively wide. This reflects the large standard errors of the drug-related parameters  $EC_{50}$  and  $E_{max}$  on the transition from state 2 (pain relief state) to state 3 (pain-free state). Explanations for the poor precision of the pain-free parameters include a lesser availability of pain-free data (not all patients reach the pain-free state), lack of different dose groups in the data set and an intrinsically high variability in the pain-free response in a population of patients. Subject-to-subject variability in the transition rates is an important concern when modelling migraine responses. There is a need to explore random-effect models to account for the large interindividual variability. In the current hidden Markov S-Plus library the incorporation of random effects was not considered as it is not included as an option. Moreover, as our model contains a rather complicated nonlinear covariate function, adding random effects to the model could further complicate parameter estimation.

Given the transient character of Markov chains and the fact that each of the transitions in the model is associated with different uncertainties, one can expect the confidence intervals to vary over time. The cumulative nature of pain relief and pain free responses causes the width of the confidence interval to increase over time. On the other hand, the size of the interval is limited by the natural bounds of the probability measure which is



**Figure A.3:** Pain relief responses (mean and 95% confidence interval, upper two panels) and pain free responses (mean and 95% confidence interval, middle and lower panels) for sumatriptan 100 mg and placebo. Responses are calculated using the delta method based on the Kolmogorov equations (horizontal shading) and simulated using a Monte Carlo simulations (vertical shading). The lower panel shows the hypothetical effect of increasing the number of patients by 50% (equivalent to dividing the standard errors by 1.5).





**Figure A.4:** Concentration versus response relationships for sumatriptan. The three graphs give the placebo-corrected pain relief that can be expected at 1.0 hours, 2.0 hours and 4.0 hours after oral dosing. Note that a concentration of *e.g.* 10 nM represents different doses in each of the graphs, as plasma drug concentration is time-dependent [11]. Thus, for a concentration of 10 nM to be reached after 4 hours, a higher dose would be needed than for reaching the same concentration after 2 hours. The main application of these graphs is to determine the maximum response that can be gained relative to placebo. Means and 95% confidence bands were calculated using the new algorithm.

defined between zero and unity. All these factors make that assigning confidence levels to migraine responses is not so straightforward.

The calculated confidence intervals represent uncertainty arising from estimation of unknown parameters in the transition rate matrix. These include both baseline transition rates and parameters linking these rates to non-homogeneous covariates, such as drug concentrations. When link functions are nonlinear, these parameters are particularly difficult to estimate. In the current application of acute anti-migraine therapy the parameters in the link function are the potencies ( $EC_{50}$ ) and maximum effects ( $E_{max}$ ) of the drug. It can be verified in Table A.1 that the drug-related parameters are less precisely estimated than the baseline transition rates. These differences in precision are reflected in Figure A.2. The width of the confidence intervals is considerably larger in the predictions

for 100 mg sumatriptan than those for placebo.

As indicated in Figure A.4, the width of the confidence intervals depends on the value of the drug concentration. Confidence intervals are widest when the mean response is close to 50% of its half-maximum value. This is explained by the relatively large standard error of the parameter  $EC_{50,12}$  and the sensitivity of the response to  $EC_{50,12}$ , which is highest in this part of the curve. Note that the concentration at half-maximum pain relief is not equal to the value of parameter  $EC_{50,12}$ . Whereas this parameter expresses the potency of the drug for a single transition in the model, the pain relief responses are the result of individual trajectories consisting of many transitions. Moreover, pain relief responses are plotted as placebo-subtracted values, making the concentrations at half maximum response appear smaller as exposure time increases. The width of the confidence intervals is smaller than is expected on the basis of findings on variability in studies of anti-migraine therapy. This is because the intervals reflect uncertainty due to parameter estimation, rather than biological variability between patients and occasions. Although a high biological variability may affect parameter precision, the main determinant of parameter precision is the amount of data available.

In the current application standard errors in the matrix of observed scores have not been taken into account when calculating the 95% confidence intervals. However, the current algorithm can be extended relatively easily to accommodate the errors in the observed scores. The vector of parameters will be extended by adding the probabilities of the scores  $p_y(j)$  and the variance-covariance matrix  $B^2$  will be replaced by  $B^{2*}$ , which includes the (co)variances of the  $p_y(j)$ . From A.4 and A.5 it is clear that the errors in the score probabilities are linearly combined with errors in the initial distribution vector and the transition matrix. The errors in the observed scores are therefore additive to the other sources of uncertainty. This means that the same formulas can be used if uncertainties in the scores need to be included.

Time-dependent misclassification rates (score probabilities) can be part of a formal and theoretic discussion on confidence intervals. However, such rates, in the practice of longitudinal data analysis, rarely occur or have not been identified. This is likely due to 1) the difficulty of identifying such effects in longitudinal data sets and 2) the large number of records that would be needed to estimate the associated parameters, in particular when the distribution of observed scores is multinomial. It is not the aim of this manuscript to explore all possible combinations of covariates. Rather, it illustrates for the first time how confidence intervals can be easily constructed for Markov models in continuous time, while focusing on the practice of modelling clinical migraine data.

In reality, drug concentrations are measurements and therefore they are random variables. Thus, it is expected that the pharmacokinetics of a drug also add to the variability of the response. In this analysis, due to the absence of pharmacokinetics in patients, drug concentrations at specific time points were treated as constants. Moreover, the currently applied software does not allow for uncertainty in the explanatory level to be incorporated into the model. An impression can be obtained of the extent to which variability in the pharmacokinetics adds to the variability in response by recalculating the confidence levels using quantiles of the concentration profiles. For example, the lower and upper confidence levels of the response can be calculated using the 5th and 95th percentiles of

the distribution of concentration profiles, respectively.

The delta method is a fast and appealing method for calculating confidence levels. Although its results are often good, there are conditions under which the results are less accurate [4]. It has been argued that when the diagonal entries in the variance-covariance matrix are large, this method tends to underestimate variance [17]. In general, the more terms are included in the Taylor series the more accurate are the approximations. A trade-off ought to be made between precision on the one hand and feasibility of calculation on the other hand. Second and higher order partial derivatives of the parameters in the Markov transition matrix are extremely large expressions. They are prone to rounding errors and overflow. Considering the main objective of this paper, illustrating the application of the delta method to the Kolmogorov equations, the calculation of further terms of the Taylor series is not thought essential. Rather than comparing different approximations, the accuracy lost by using a first-order Taylor approximation may be more conveniently assessed by comparing the calculated confidence intervals with those obtained by a Monte Carlo analysis.

In conclusion, an approach has been developed for the calculation of mean predicted headache responses and confidence intervals, based on (hidden) Markov models. The calculated means are shown to be equivalent to those obtained from simulations. Moreover, responses are calculated considerably faster using the algorithm. To our knowledge, no other methods have been described in literature to calculate confidence intervals on responses obtained with Markov modelling. This approach will help in identifying uncertainty in predictions of clinical responses predicted by Markov models.

## References

- [1] C. Allen, K. Jiang, W. Malbecq, and P.J. Goadsby. Time-to-event analysis, or who gets better sooner? An emerging concept in headache study methodology. 1999, *Cephalalgia*, 19, 552–556.
- [2] A. Bureau, S. Shiboski, and J.P. Hughes. Applications of continuous time hidden Markov models to the study of misclassified disease outcomes. 2003, *Stat. Med.*, 22, 441–462.
- [3] D.J. Cher and L.A. Lenert. Rapid approximation of confidence intervals for Markov process decision models: Applications in decision support systems. 1994, *J.Am.Med.Inform.Assoc.*, 4, 301–312.
- [4] G.W. Oehlert. A note on the delta method. 1992, *Am.Stat.*, 46, 27–29.
- [5] L.R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. 1989, *Proc.IEEE*, pp. 257–286.
- [6] A. Bureau, J.P. Hughes, and S.C. Shiboski. An S-Plus implementation of Hidden Markov Models in Continuous Time. 2000, *J.Comp.Graph.Stat.*, 9, 621–632.

- [7] H.J. Maas, M. Danhof, and O.E. Della Pasqua. Prediction of headache response after migraine treatment using a Markov model. 2006, *Cephalalgia*, 26, 416–422.
- [8] Data Analysis Division. *S-Plus 6.0 User's Guide*. NONMEM Project Group, UCSF, Mathsoft, Seattle, WA, 2000.
- [9] R. Ihaka and R. Gentleman. R: A Language for Data Analysis and Graphics. 1996, *J.Comp.Graph.Stat.*, 5, 299–314.
- [10] S.L. Beal and L.B. Sheiner. *NONMEM Users Guide – Part I Users Basic Guide*. NONMEM Project Group, UCSF, San Francisco, 1989.
- [11] V.F. Cosson and E. Fuseau. Mixed effect modeling of sumatriptan pharmacokinetics during drug development: II. From healthy subjects to phase 2 dose ranging in patients. 1999, *J.Pharmacokinet.Biopharm.*, 27, 149–171.
- [12] A.W. Fox. Onset of effect of 5-HT<sub>1B/1D</sub> agonists: a model with pharmacokinetic validation. 2004, *Headache*, 44, 142–147.
- [13] D.R. Cox. Regression Models and Life-Tables. 1972, *J.R.Stat.Soc.B*, 34, 187–220.
- [14] N.H. Holford and L.B. Sheiner. Pharmacokinetic and pharmacodynamic modeling in vivo. 1981, *CRC Crit.Rev.Bioeng.*, 5, 273–322.
- [15] H. Hassani and A. Ebbutt. Use of a stochastic model for repeated binary assessment. 1996, *Stat.Med.*, 15, 2617–2627.
- [16] J.R. Carpenter. A method for presenting and comparing dose-response curves. 1986, *J.Pharmacol.Meth.*, 15, 283–303.
- [17] M. Moerbeek, A.H. Piersma, and W. Slob. A comparison of three methods for calculating confidence intervals for the benchmark dose. 2004, *Risk Anal.*, 24, 31–40.