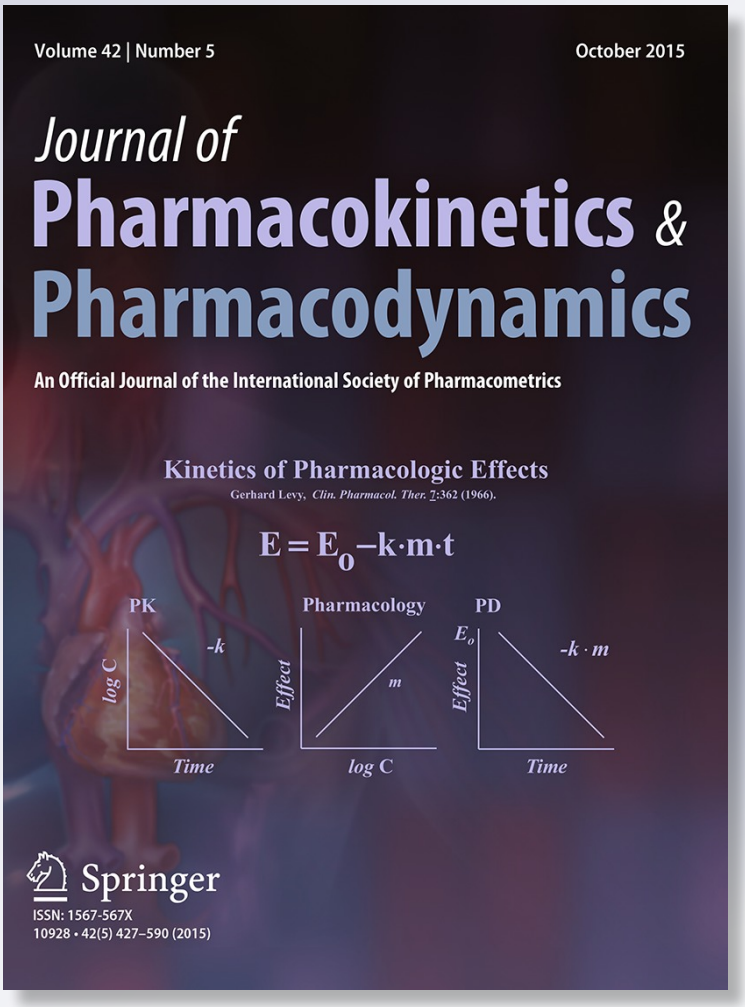
Bayesian population modeling of drug dosing adherence

Kelly Fellows, Colin J. Stoneking & Murali Ramanathan

Journal of Pharmacokinetics and Pharmacodynamics

ISSN 1567-567X
Volume 42
Number 5

J Pharmacokinet Pharmacodyn (2015)
42:515-525
DOI 10.1007/s10928-015-9439-8



Springer

J Pharmacokinet Pharmacodyn (2015) 42:515–525
DOI 10.1007/s10928-015-9439-8

ORIGINAL PAPER

Bayesian population modeling of drug dosing adherence

Kelly Fellows[1,2] · **Colin J. Stoneking**[2] · **Murali Ramanathan**[1]

Received: 28 July 2015 / Accepted: 19 August 2015 / Published online: 29 August 2015

Abstract Adherence is a frequent contributing factor to variations in drug concentrations and efficacy. The purpose of this work was to develop an integrated population model to describe variation in adherence, dose-timing deviations, overdosing and persistence to dosing regimens. The hybrid Markov chain–von Mises method for modeling adherence in individual subjects was extended to the population setting using a Bayesian approach. Four integrated population models for overall adherence, the two-state Markov chain transition parameters, dose-timing deviations, overdosing and persistence were formulated and critically compared. The Markov chain–Monte Carlo algorithm was used for identifying distribution parameters and for simulations. The model was challenged with medication event monitoring system data for 207 hypertension patients. The four Bayesian models demonstrated good mixing and convergence characteristics. The distributions of adherence, dose-timing deviations, overdosing and persistence were markedly non-normal and diverse. The models varied in complexity and the method used to incorporate inter-dependence with the preceding dose in the two-state Markov chain. The model that incorporated a cooperativity term for inter-dependence and a hyperbolic parameterization of the transition matrix probabilities was identified as the preferred model over the alternatives. The simulated probability densities from the model satisfactorily fit the observed probability distributions of adherence, dose-timing deviations, overdosing and persistence parameters in the sample patients. The model also adequately described the median and observed quartiles for these parameters. The Bayesian model for adherence provides a parsimonious, yet integrated, description of adherence in populations. It may find potential applications in clinical trial simulations and pharmacokinetic-pharmacodynamic modeling.

Keywords Adherence · Compliance · Dosing patterns · Runs · Circular distribution

✉ Murali Ramanathan
murali@buffalo.edu

[1] Department of Pharmaceutical Sciences and Neurology, State University of New York, 355 Kapoor Hall, Buffalo, NY 14214-8033, USA

[2] Seminar for Statistics, Department of Mathematics, ETH Zurich, Zurich, Switzerland

Introduction

Lack of patient adherence to prescribed dosing regimens is a potentially preventable contributing factor in numerous hospitalizations and emergency room visits [1]. Incorporating good models for patient adherence in pharmacokinetics and pharmacodynamics (PK–PD) models could enhance clinical trial simulations.

Dr. Gerhard Levy conducted several early seminal investigations into patient adherence using PK–PD principles that have proven elegant and insightful [2, 3]. His work highlighted the impact that patient-specific and drug-specific characteristics have on plasma drug concentration and drug efficacy when doses are missed. He advocated for simulations to determine whether missed doses would cause plasma concentrations to drop below the therapeutic range for drugs with known concentration-effect relationships [2, 3]. Subsequent research applied these concepts to integrate adherence data from electronic medication event monitoring systems (MEMS) technology to PK [4] and population PK and PD modeling [5, 6].

Dr. Levy's work on patient compliance explains the relationship between a drug's PK/PD properties and the consequences of missing one or several doses of that drug. For example, a drug with a shallow effect versus concentration slope and a long half-life would be less affected by a missed dose, and could be considered a "forgiving" drug [2]. He introduced the concept of the medication noncompliance impact factor; a ratio used to determine how much of an impact a missed dose would have on plasma concentration. Larger noncompliance impact factors indicate that a missed dose will have a more pronounced effect [2]. Although it is understood that less frequent dosing is associated with improved adherence to the dosing regimen, this alone does not make for a better drug. With more frequent dosing there is less concern for sub-therapeutic plasma concentrations, a more relevant measure of noncompliance, following a single missed dose. This leads to the conclusion that adjustments to dosing regimens can be useful in reducing the risk for sub-therapeutic concentration levels following a missed dose [2, 3]. The "forgiving drug" concept has become more relevant and applicable for clinical trial design in the personalized medicine era, because it can help inform the need for adherence monitoring [1]. In clinical trials, poor adherence not only undermines the power to detect drug efficacy, but also the ability to detect adverse events. The adverse consequences of poor patient adherence can be mitigated with the use of "forgiving drugs" [1]. Additionally, adherence data has been integrated with pharmacogenetic outcomes in a clinical trial setting [7].

In clinical trial and practice settings, patient adherence data can be obtained using several methods including patient-reported adherence, the use of medication possession ratios within e-prescribing systems, and from MEMS. MEMS technology provides the most detailed information of individual dosing events as it provides a time stamp at which medication container openings occur. MEMS monitoring is considered a gold standard for obtaining objective surrogate measures of dosing events. However, there is an inherent assumption that medication opening results in medication administration. The information from MEMS devices has been previously used to conduct simulations of drug concentrations and efficacy [8, 9]. In our recent work, dosing failures and successes were modeled using a two-state Markov chain, a von Mises (VM) distribution was used to model dose-timing deviations, a Weibull distribution was used to model persistence, and a Poisson distribution was used to model the frequency in which an overdose would occur. Additionally, we introduced the patient cooperativity index (PCI) an odds ratio calculated using the parameters from the Markov chain that can be used to describe the variations in dosing patterns between two patients with the same level of overall adherence. The goal of this research is to use Bayesian Markov chain–Monte Carlo (MCMC) techniques to extend our previous model for use in population modeling.

Methods

Adherence data sets

The data used for model assessment were the observed dosing regimens for 207 patients on once-daily hypertension treatment. Data on dose-timing, number of dose units taken for each dose, and dosing intervals were extracted from the Pharmionic Knowledge Center (PKC) database (http://www.iadherence.org/www/pkc.adx), a compilation of electronically monitored events that serve as an objective surrogate for a patient's dosing history.

Dosing and persistence data for each patient were obtained from the calendar plots. The dosing interval data was extracted from the dosing time graphs [10].

Modeling individual dosing profiles

The hybrid Markov chain–VM model described in detail in [10] was used. The key features of this model are described below. The individual model consisted of four inter-dependent components:

1. A two-state Markov chain was used to model the occurrence of dosing failures and dosing successes.
2. The dose-timing deviations were modeled with a VM distribution.
3. Persistence was modeled with a Weibull distribution.
4. The frequency of overdosing events was modeled with a Poisson distribution. The time between overdosing events was exponentially distributed as a direct consequence of the Poisson model for overdosing event frequency.

Modeling the two-state Markov chain

Individual level model

We modeled the short-range dependence of dosing using a two-state time-homogeneous Markov chain with transition matrix A:

$$\boldsymbol{A} = \begin{matrix} S \\ F \end{matrix} \begin{bmatrix} p_{SS} & p_{SF} \\ p_{FS} & p_{FF} \end{bmatrix} = \begin{matrix} S \\ F \end{matrix} \begin{bmatrix} p_{SS} & 1 - p_{SS} \\ p_{FS} & 1 - p_{FS} \end{bmatrix}$$

The probability of a success followed by a success at the next dosing event was denoted by p_{SS} and the probability of a failure followed by a success at the next dosing event was denoted by p_{FS}.

Maximum-likelihood estimates [11] of p_{SS} and p_{FS} were obtained from N_{SS}, N_{FS}, N_{FF} and N_{SF}, the frequencies of

success followed by success, failure followed by success, failure followed by failure and success followed by failure events, respectively, using:

$$\widehat{p}_{SS} = \frac{N_{SS}}{N_{SS} + N_{SF}}$$

$$\widehat{p}_{FS} = \frac{N_{FS}}{N_{FS} + N_{FF}}$$

The PCI was defined as:

$$PCI = \left(\frac{p_{FS}}{1 - p_{FS}}\right) \Big/ \left(\frac{p_{SS}}{1 - p_{SS}}\right)$$

The PCI represents an odds ratio of the odds of a failure followed by a success, to the odds of a success followed by a success. Patients with low PCI values are less likely to take a dose following an omission, whereas those with high PCI values are more likely to take a dose following an omission.

Bayesian population modeling

Four alternative approaches were compared for modeling the two-state Markov chain component of the model. The dose-timing deviation, overdosing and persistence components of all four models were identical.

The count matrix, N_i, which contains: $N_{SS}(i)$, the count of successes that are followed by successes in subject i; $N_{SF}(i)$, the count of successes that are followed by failures, $N_{FS}(i)$, the count of failures that are followed by successes; and $N_{FF}(i)$, the count of failures that are followed by failures:

$$N_i = \begin{bmatrix} N_{SS}(i) & N_{SF}(i) \\ N_{FS}(i) & N_{FF}(i) \end{bmatrix}$$

Let $N_{Sum}(i) = N_{SS}(i) + N_{SF}(i) + N_{FS}(i) + N_{FF}(i)$. The relative frequency matrix, Q_i contains the relative frequencies corresponding to the cells of the count matrix N_i.

$$Q_i \equiv \frac{N_i}{N_{Sum}(i)} = \begin{bmatrix} q_{SS}(i) & q_{SF}(i) \\ q_{FS}(i) & q_{FF}(i) \end{bmatrix}$$

The transition matrix A_i used for defining the Markov chain, is generated by normalizing each entry of N_i to the appropriate row counts, such that:

$$A_i = \begin{matrix} S \\ F \end{matrix} \begin{bmatrix} p_{SS}(i) & p_{SF}(i) \\ p_{FS}(i) & p_{FF}(i) \end{bmatrix} = \begin{matrix} S \\ F \end{matrix} \begin{bmatrix} p_{SS}(i) & 1 - p_{SS}(i) \\ p_{FS}(i) & 1 - p_{FS}(i) \end{bmatrix}$$

Model 1, Dirichlet prior

For Model 1, the counts of $N_{SS}(i)$, $N_{SF}(i)$, $N_{FS}(i)$, and $N_{FF}(i)$ in N_i were modeled using a multinomial distribution:

$$N_i \sim Multinomial\,(Q_i,\ N_{Sum}(i))$$

$$Q_i \sim Dirichlet\,(\alpha)$$

The counts $N_{SS}[i]$, $N_{SF}[i]$, $N_{FS}[i]$ and $N_{FF}[i]$ were viewed in this model as resulting from the trials of a categorical event with four possible outcomes. The Dirichlet is a vector distribution of probabilities that sum to unity. In this case, the Dirichlet vector is of order 4.

Model 2, independent beta priors for each row

$N_{FF}[i]$ was modeled using a binomial distribution that was a function of $p_{FF}[i]$, the probability of a failure following a failure and the total number of failures $N_F[i] = N_{FS}[i] + N_{FF}[i]$. $N_{FS}[i]$ and $N_{SF}[i]$ were modeled analogously.

$$N_{FF}[i] \sim Binomial(p_{FF}[i],\ N_F[i])$$

$$N_{FS}[i] \sim Binomial(1 - p_{FF}[i],\ N_F[i])$$

$$N_{SF}[i] \sim Binomial(p_{SF}[i],\ N_S[i])$$

The value of $N_{SS}[i]$ was obtained by difference since:

$$\begin{aligned} N_{Sum}[i] &= N_{Total}[i] - 1 \\ &= N_{FF}[i] + N_{FS}[i] + N_{SF}[i] + N_{SS}[i] \end{aligned}$$

The individual rows of the transition matrix were modeled with Beta distributions.

$$p_{FF}[i] \sim Beta(a_{FF},\ b_{FF})$$

$$p_{SF}[i] \sim Beta(a_{SF},\ b_{SF})$$

The values of the two remaining elements of the transition matrix were obtained by difference:

$$p_{FS}[i] = 1 - p_{FF}[i]$$

$$p_{SS}[i] = 1 - p_{SF}[i]$$

Model 3, inter-dependent beta priors

In Model 3, the counts $N_{FF}[i]$, $N_{SF}[i]$, $N_{FS}[i]$ were modeled with binomial distributions as in Model 2.

However, the transition matrix was modeled with an alternative re-parameterization adapted from a strategy for dependent discrete Markov chains proposed by Devore [12], Rudolfer [13] and Islam and O'Shaughnessy [14]. The elements of the transition matrix for subject i were parameterized using the overall adherence $p_S(i)$, first-order auto-correlation term, $r[i]$.

$$A_i = \begin{bmatrix} p_S(i) + p_F(i)r(i) & p_F(i) - p_F(i)r(i) \\ p_S(i) - p_S(i)r(i) & p_F(i) + p_S(i)r(i) \end{bmatrix}$$

We reasoned that this re-parameterization would be better equipped to model transitions matrices where the

row probabilities were dependent. In Model 3, there are two parameters, $p_S(i)$ and $r(i)$, subject to the constraints: $p_F(i) = 1 - p_S(i)$ and $-1 \leq r(i) \leq 1$.

The overall adherence $p_S(i)$ can range from 0 to 1 and the correlation term $-1 \leq r(i) \leq 1$. The priors for parameters, $p_S(i)$ and $r(i)$ were both Beta distributions but with support $[0, 1]$ and $[-1, 1]$, respectively.

$$p_S(i) \sim Beta(a_P, b_P)$$

$$r(i) \sim Beta(a_r, b_r) \text{ over } [-1, 1].$$

Model 4, helix-coil model

The parameterization of this model consisted of hyperbolic functions of two parameters, $s(i)$ and $\sigma(i)$, that underlie for the helix-coil transition model in polymer physics. Previous work from our group had demonstrated promise of this approach for adherence modeling [15].

$$A_i = \begin{bmatrix} \frac{s(i)}{1 + s(i)} & \frac{1}{1 + s(i)} \\ \frac{\sigma(i)s(i)}{1 + \sigma(i)s(i)} & \frac{1}{1 + \sigma(i)s(i)} \end{bmatrix}$$

The priors for $s(i)$ and $\sigma(i)$ were:

$$s(i) = e^{z(i)}$$

$$z(i) \sim Normal(a_s, b_s)$$

$$\sigma(i) \sim Gamma(a_\sigma, b_\sigma)$$

The $z(i)$ is an intermediate dummy variable and b_s represents the precision of the normal distribution.

Dose-timing deviation distribution

The dose-timing deviation of the ith dose, δ_i, was defined as the difference between the actual dosing time and the nearest prescribed dosing time.

Individual level model

We assume that the PDF of the dose-timing deviations δ after transformation to angular coordinates is distributed according to the VM PDF function $VM(\mu, \kappa)$:

$$2\pi\frac{\delta}{\tau} \sim VM(\psi, \omega)$$

The VM distribution describes angular random variables and its PDF $p(\theta)$ at angular position θ radians is:

$$p(\theta) = \frac{e^{\kappa cos(\theta - \psi)}}{2\pi I_0(\omega)} \quad for\ \psi - \pi \leq \theta \leq \psi + \pi$$

The $\psi(i)$ is the mean and $\omega(i)$ is a measure of how concentrated the dose-timing deviation angles are around the mean; I_o is the Bessel function of order zero.

The R circular statistics package [16] was used to obtain maximum likelihood estimates for ψ and ω, the mean and concentration parameters for the VM distribution of dose-timing deviation angles for each subject.

Bayesian population modeling

The prior for the VM location parameter $\psi(i)$ was a VM distribution. Both a gamma prior and log-normal prior were investigated for modeling the VM concentration parameter. In pilot experiments, the log-normal distribution provided a better fit. Therefore, the prior for the VM concentration parameter $\kappa(i)$ was assumed to follow a log-normal distribution.

$$\psi(i) \sim vonMises(a_\psi, b_\psi)$$

$$\omega(i) \sim LogNormal(a_\omega, b_\omega)$$

The a_ω and b_ω represent the mean and precision of the log-normal prior, respectively.

Modeling overdosing events

Individual level model

The distribution of over-dosing events was modeled with a Poisson distribution. Maximum-likelihood estimation was used to calculate Poisson rate parameter $\lambda(i)$. The over-dosing fraction (OF) was defined as the fraction of dosing events in which the patient was taking more doses than prescribed (>1 each day).

Bayesian population modeling

Both gamma and exponential priors were evaluated for λ, the rate parameter of the overdosing distribution. From exploratory probability–probability plots, it was apparent that the exponential distribution provided a better and more parsimonious fit to the data distribution. The exponential distribution was selected for the subsequent experiments.

$$\lambda(i) \sim Exponential(a_\lambda).$$

Long-term persistence

Individual level model

Persistence was defined as the period from the date of the first successful dosing event to the last successful dosing event. For once-daily dosing, $N_{Total}[i]$, the total number of observed dosing events is equivalent to the persistence of subject i.

Persistence is a time-to-event variable and has to be handled in a manner that appropriately addresses censored data. As noted in our previous report, there were three sub-groups in the PKC adherence dataset. Two sub-groups had a 6-month observation period but disparate overall adherence values of 68.5 and 92.8 %, respectively. The third sub-group was followed up for 9-months and had 92.5 % adherence. The three groups were simultaneously modeled with appropriate censoring time. Subjects whose persistence values were ±2 days of (6 or 9-month) their respective sub-groups observation period were considered censored. The *dinterval* function in RJAGS, which creates indicator variable from the observed times and the censoring time (observation period) as inputs was used.

Bayesian population modeling

The persistence $N_{Total}[i]$ was modeled using a Weibull distribution with shape parameter υ and scale parameter κ:

$$N_{Total}[i] \sim Weibull(\upsilon, \kappa)$$

The Weibull density with the accelerated life parameterization for a random variable x is defined as:

$$Weibull(\upsilon, \kappa) = \upsilon\kappa x^{\upsilon-1} e^{-\kappa x^{\upsilon}}$$

The Weibull distribution has to be reparameterized for MCMC modeling because υ, the Weibull shape parameter, shows poor mixing (http://sourceforge.net/p/mcmc-jags/discussion/610036/thread/d5249e71/) [17]:

$$\kappa = \frac{1}{\varrho^{\upsilon}}$$

$$\xi = e^{\varrho}$$

The full re-parameterization was

$$Weibull(\upsilon, \varrho) = \frac{\upsilon}{\varrho}\left(\frac{x}{\varrho}\right)^{\upsilon-1} e^{-(x/\varrho)^{\upsilon}}$$

We refer to ξ as the log-scale parameter. A gamma distribution was used as the prior for υ, the shape parameter whereas a normal distribution was used as the prior for ξ.

$$\upsilon \sim Gamma(a_{\upsilon}, b_{\upsilon})$$

$$\xi \sim Normal(a_{\xi}, b_{\xi})$$

Markov chain–Monte Carlo implementation

Mathematica (Wolfram Research, Champaign, IL) was used for estimating individual level parameters. Both SPSS (IBM, Armonk, NY) and the R statistics program were used for exploratory statistical analyses [16, 18].

Bayesian analysis was conducted using the well-established MCMC method in RJAGS software [19, 20]. To allow the MCMC chains to converge we employed a conservative burn-in phase of 50,000 runs before evaluating any statistics. The MCMC algorithm was implemented for three chains each with 200,000 runs and thinning interval of 2.

Uninformative conjugate priors were used for the hyper-parameter distributions for which closed forms were available. Uninformative closed form priors with compatible support were identified from numerical experiments for a subset of hyper-parameters lacking closed form conjugate priors.

The MCMC runs were analyzed with the convergence diagnosis and output analysis (CODA) package [20]. Add-on code for analyzing the VM distribution in RJAGS was developed by Colin Stoneking and Klaus Oberauer (Department of Cognitive Psychology, University of Zurich, Switzerland).

Convergence and mixing were assessed using multiple approaches. First, visual inspection of parameter trace plots was conducted for evidence of poor mixing. The Gelman–Rubin–Brooks plot, which shows the evolution of Gelman and Rubin's shrink factor as the number of iterations increases [21], and the Gelman and Rubin multiple sequence diagnostic multivariate scale reduction factor were also examined [21–23]. The Gelman–Rubin statistic was less than 1.05 [21, 23].

The two-state Markov chain Models 1–4 were compared using the deviance information criterion (DIC) [24] as computed in CODA [24, 25]:

$$DIC = \overline{D} + p_D$$

where $\overline{D}$ is the mean deviance and p_D is the effective number of parameters.

For visual predictive checks, the empirical density functions were computed using the "density" function in R with default settings. The density plots from RJAG simulations were compared to the empirical density functions. The empirical density functions were truncated outside of support to enable meaningful direct comparison with the corresponding simulated density function. Raw histograms were visually overlaid on the simulations and densities to provide a visual reference.

The Bayesian analyses were conducted on a MacBook Pro laptop computer running the OS X Yosemite operating system version 10.10.3.

Results

The distribution of the model parameters for the 207 subjects on the once-daily treatment for hypertension are summarized in Fig. 1 and Table 1. The majority of the parameters in the hybrid Markov chain–VM model were markedly non-normal as evidenced by the skewness values and the histogram profiles. This provided the motivation to investigate MCMC methods for population modeling.

Visual examination of the iteration trace plots did not identify evidence for poor mixing of the chains for any of the parameters in each of the models. The Gelman–Rubin–Brooks plots for all of the parameters in each of the parameters converged to values less than 1.05 and the Gelman and Rubin multivariate potential scale reduction factors approached unity.

Table 2 compares the DIC of the four competing models for the two-state Markov chain component of the adherence model. The DIC is an extension of model selection criteria such as the Akaike information criterion (AIC) that is suitable for MCMC methods [22, 24]. Lower values of deviance indicate a more parsimonious model. The rank

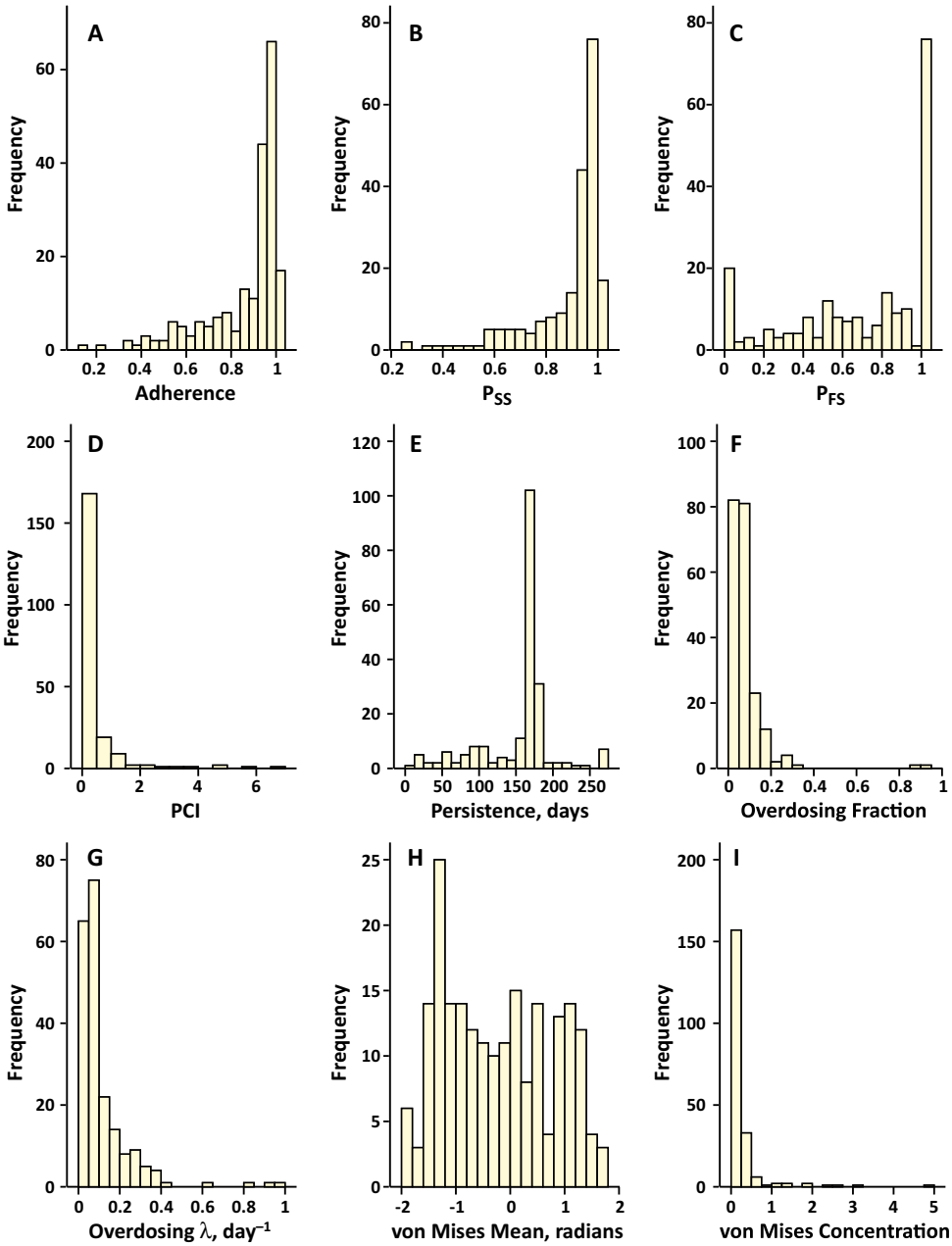


Fig. 1 Histograms for key adherence variables and model parameters of the hybrid Markov chain–von Mises model at the individual level. The histograms in **a**–**d** show overall adherence fraction, p_{SS}, p_{FS} and PCI, respectively, which are the key variables and derived parameters of the two-state Markov model. The histogram in **e** shows the apparent persistence values. The values are not corrected for censoring. The histograms in **f** and **g** show the fraction of overdosing events and the Poisson rate parameter for overdosing. The histograms in **h** and **i** show the mean and concentration parameters of the von Mises distribution for dose timing deviations

Table 1 Mean, standard deviation and percentiles for the model parameters

	Definition	Mean	SD	Skewness	2.5 %	5 %	10 %	25 %	Median	75 %	90 %	95 %	97.5 %
p_S	Adherence	0.865	0.174	−1.71	0.390	0.475	0.570	0.797	0.943	0.982	0.994	1.00	1.00
p_{SS}	Probability of a success followed by a success	0.892	0.148	−2.08	0.435	0.571	0.671	0.870	0.956	0.985	0.994	1.00	1.00
p_{FS}	Probability of a failure followed by a success	0.690	0.339	−0.817	0.000	0.000	0.0704	0.475	0.818	1.00	1.00	1.00	1.00
PCI[a]	Patient cooperativity index	0.361	0.906	4.48	0.000	0.000	0.000	0.0000	0.0080	0.320	1.00	1.54	3.56
N_{Total}	Persistence[a]	155	0.339	−0.935	18.0	44.6	82.6	155	169	170	180	223	270
OF	Overdosing fraction	0.0754	0.0980	5.81	0.000	0.0044	0.0080	0.0290	0.0590	0.0880	0.151	0.191	0.253
λ	Poisson rate parameter	0.111	0.135	3.46	0.000	0.00427	0.00979	0.0337	0.0710	0.1430	0.257	0.340	0.395
ψ	von Mises mean	−0.243	0.987	0.19	−1.87	−1.59	−1.45	−1.17	−0.327	0.589	1.13	1.36	1.49
ω	von Mises concentration	0.287	0.503	5.87	0.0334	0.0456	0.0710	0.110	0.172	0.249	0.420	1.00	1.84

[a] The PCI statistics shown only reflect those values that are finite. PCI can be indeterminate if terms in the denominator are zero. The persistence statistics are not adjusted for censoring

Table 2 Deviance information criterion (DIC) for the four competing two-state Markov chain models

Model	DIC
Model 1	13385
Model 2	2433
Model 3	2487
Model 4	2405

order of the DIC values was Model 1 > Model 3 > Model 2 > Model 4. Thus Model 4 represents the preferred model. We surmise that Model 4 builds in inter-dependence across rows of the two-state Markov chain transition matrix, which could provide greater flexibility for modeling adherence data sets.

The distribution of the hyper-parameters for the dose-timing deviation, overdosing and persistence components of the adherence model were very similar across Models 1 through Model 4. This is not surprising given that these particular components are only weakly dependent on the two-state Markov chain component. Table 3 summarizes the distribution of the hyper-parameters for dose-timing deviation, overdosing and persistence components from Model 4.

Table 4 summarizes the distribution of the hyper-parameters for the two-state Markov chain components for Models 1–4. The two-state Markov chain parameter estimates for Model 1 showed large standard deviations relative to the mean as assessed by the coefficient of variation (CV) compared to two-state Markov chain parameter estimates for the other models. Because Model 1 was markedly inferior to the other models we did not pursue refinements to hyper-parameter distribution assumptions of Model 1. The CV values of hyper-parameter distribution for the other models were satisfactory—majority were <15 % and all were <25 %.

We conducted the equivalent of a visual predictive check by conducting simulations in RJAGS for several observed variables and parameters in Fig. 1. Figure 2 summarizes the results. The dashed green line represents the results from the Bayesian simulations based on Model 4 and the red line represents the empirical density function for the variable. The histograms from Fig. 1 are provided as an approximate visual reference. Figure 2 demonstrates that Model 4 and the modeling strategies for the dose timing deviations and overdosing components are satisfactorily fit data. Figure 3 plots the observed and predicted quartile values from Model 4 for overall adherence fraction, p_{SS}, p_{FS}, the fraction of overdosing events, the Poisson rate parameter for overdosing, and the mean and concentration parameters of the VM distribution for dose timing deviations. The proximity of the values to the 45° line of equality provides additional evidence in support of Model 4.

Table 3 Mean, standard deviations and credibility intervals for the hyper-parameters of the dose-timing deviation, overdosing and persistence components of the adherence model

Component	Hyper-parameter	Prior	Mean	SD	CV %	SEM	2.5 %	25 %	50 %	75 %	97.5 %
Dose timing	a_ψ	von Mises	2.86	0.0764	2.7	1.48×10^{-4}	2.71	2.81	2.86	2.91	3.01
Dose timing	b_ψ	von Mises	1.44	0.137	9.5	2.66×10^{-4}	1.18	1.35	1.44	1.53	1.72
Dose timing	a_ω	Log normal	−1.73	0.0583	3.4	1.12×10^{-4}	−1.84	−1.77	−1.73	−1.69	−1.61
Dose timing	b_ω	Log normal	1.43	0.140	9.9	2.55×10^{-4}	1.16	1.33	1.42	1.52	1.71
Persistence	υ	Gamma	2.56	0.217	8.5	1.23×10^{-3}	2.15	2.41	2.55	2.70	2.99
Persistence	ξ	Normal	5.53	0.0555	1.0	3.44×10^{-4}	5.43	5.49	5.53	5.57	5.65
Overdosing	a_λ	Exponential	9.04	0.628	6.9	1.15×10^{-3}	7.85	8.61	9.03	9.46	10.3

The estimates from Model 4 are shown

CV % coefficient of variation of parameter estimate as a percentage, *SEM* standard error of the mean adjusted for autocorrelation

Table 4 Mean, standard deviations and credibility intervals for the hyper-parameters of the two-state Markov chain parameters for Models 1–4

	Prior	Mean	SD	CV %	SEM	2.5 %	25 %	50 %	75 %	97.5 %
Model 1										
a_{FF}	Beta	2.68	4.43	165	4.67E−02	1.77E−02	0.264	9.86E−01	3.13	15.3
a_{FS}	Beta	2.78	4.59	165	4.35E−02	1.79E−02	0.270	1.01	3.23	16.1
a_{SF}	Beta	2.70	4.40	163	4.06E−02	1.80E−02	0.267	1.00	3.16	15.4
a_{SS}	Beta	3.36	5.44	162	5.39E−02	1.94E−02	0.315	1.22	3.98	19.3
b_{FF}	Beta	0.697	1.98	284	1.51E−02	1.18E−07	4.36E−03	5.84E−02	0.472	5.77
b_{FS}	Beta	0.649	1.85	285	1.43E−02	1.14E−07	3.84E−03	5.21E−02	0.436	5.35
b_{SF}	Beta	0.653	1.90	291	1.48E−02	1.11E−07	3.75E−03	5.07E−02	0.424	5.44
b_{SS}	Beta	0.120	0.614	514	4.51E−03	1.43E−08	3.75E−04	5.34E−03	4.66E−02	0.957
Model 2										
a_{FF}	Beta	0.606	9.04E−02	15	5.84E−04	0.447	0.543	0.600	0.663	0.801
a_{SS}	Beta	4.53	0.578	13	1.81E−03	3.47	4.13	4.50	4.90	5.74
b_{FF}	Beta	1.52	0.212	14	1.12E−03	1.15	1.38	1.51	1.66	1.98
b_{SS}	Beta	0.542	5.00E−02	9	1.70E−04	0.449	0.508	0.540	0.575	0.646
Model 3										
a_r	Beta	9.31	1.14	12	9.66E−03	7.23	8.51	9.26	10.0	11.7
a_P	Beta	0.562	5.24E−02	9	1.83E−04	0.466	0.526	0.560	0.596	0.671
b_r	Beta	5.90	0.768	13	6.55E−03	4.51	5.36	5.86	6.39	7.52
b_P	Beta	3.54	0.439	12	1.42E−03	2.74	3.23	3.52	3.83	4.46
Model 4										
a_σ	Gamma	0.667	7.94E−02	12	4.66E−04	0.525	0.612	0.663	0.718	0.836
a_s	Normal	3.03	0.121	4	2.31E−04	2.79	2.94	3.02	3.11	3.27
b_σ	Gamma	0.696	0.165	24	1.17E−03	0.416	0.579	0.682	0.797	1.06
b_s	Normal	0.361	4.16E−02	12	9.20E−05	0.284E	0.332	0.359	0.388	0.447

CV % coefficient of variation of parameter estimate as a percentage, *SEM* standard error of the mean adjusted for autocorrelation

Discussion

In this research, we conducted Bayesian population analyses of a hybrid Markov chain–VM model for analyzing adherence data using the MCMC method. The Bayesian model satisfactorily described the distribution of adherence, dose-timing deviations, overdosing and persistence in the data set of 207 patients on once-daily hypertension treatment.

Linear and non-linear mixed effect modeling methods are widely used for population modeling in

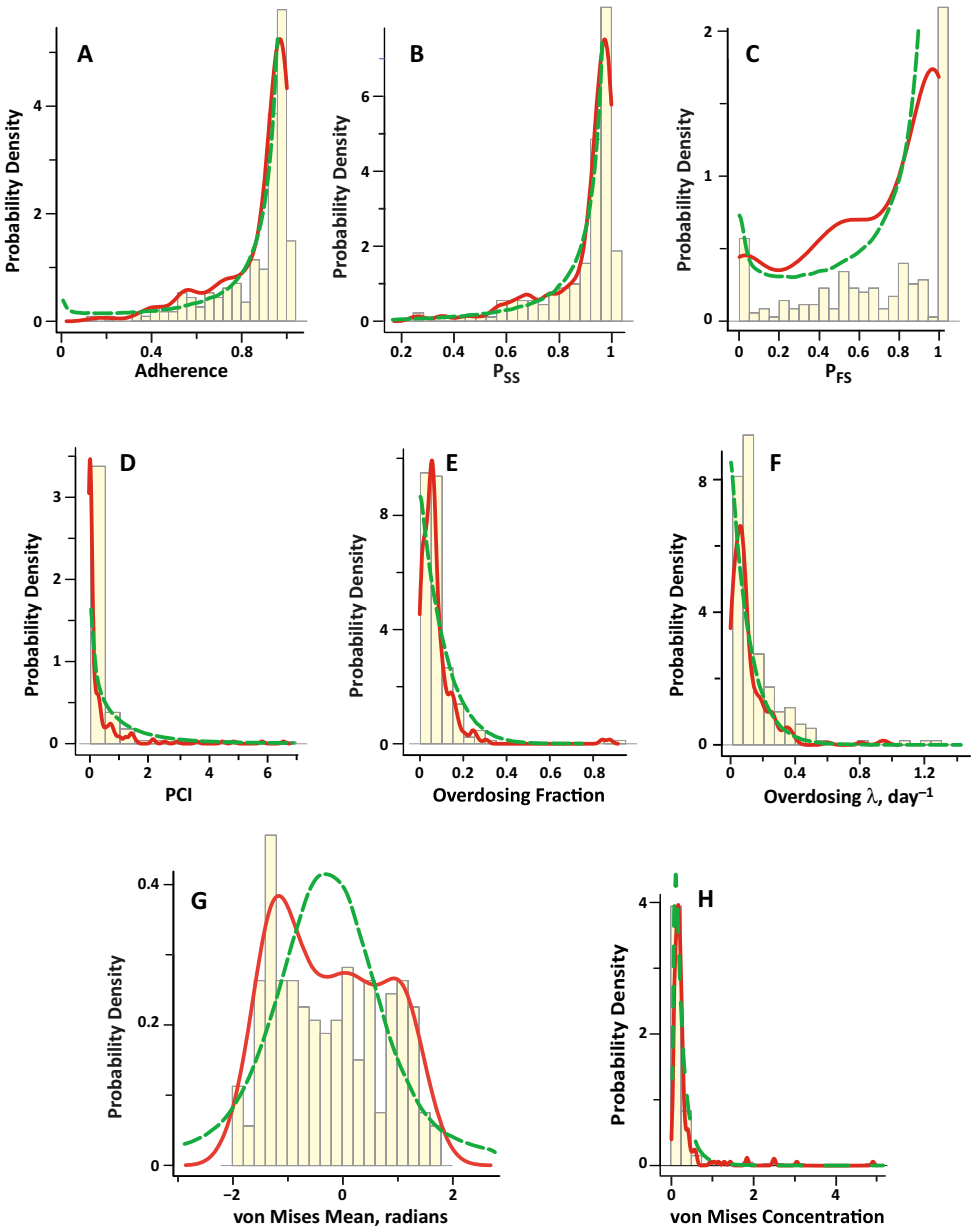


Fig. 2 Visual predictive check. The *dashed green lines* in **a**–**h** shows the simulations of the estimated density functions from the Bayesian population modeling for the key variables and parameters of the hybrid Markov chain–von Mises model. The *red lines* show the empirical density function for the same variables. The corresponding histograms are manually placed in the background to provide an approximate visual reference to the model fit from Model 4 and the empirical density function. The simulated and empirical densities and histograms in Fig. 1a–d show overall adherence fraction, p_{SS}, p_{FS} and PCI, respectively, which are the key variables and derived parameters of the two-state Markov model. The simulated and empirical densities and histograms in **e** and **f** correspond to the fraction of overdosing events and the Poisson rate parameter for overdosing, respectively. The simulated and empirical densities and histograms in **g** and **h** are for mean and concentration parameters of the von Mises distribution for dose timing deviations. The *y*-axis scale should not be used for the histograms (Color figure online)

pharmacometrics. Bayesian MCMC methods however are particularly well suited for extending our hybrid Markov chain–VM adherence model to the population setting because these methods enable inclusion of the diverse distributions, which can be combined as necessary for the integrated adherence model.

The most challenging aspect of the population modeling was characterizing the distribution of the two-state Markov-chain transition matrix. We used a systematic approach of evaluating four competing approaches varying in complexity and in the nature of the prior distributions. We also compared two different parameterizations of the transition probabilities in Model 3 and 4. Model 4 was identified as the best model based on the DIC. However, Model 3, which also contains built-in inter-dependence across rows of the transition matrix, was inferior to both Model 4 and 2. We attribute this apparent discrepancy to distribution characteristics of the Model 3 parameters, which may not have been modeled adequately by the available closed form distributions. The beta distribution was the only closed form distribution available for modeling the two parameters in Model 3 given their limited domain of support. Our selection of priors implicitly assumes that a single population generates the underlying transition matrix. Mixture distributions are necessary if the underlying adherence patterns exhibit clustering in different subgroups of subjects, e.g., in "adherent" and "non-adherent" subgroups. Bayesian nonparametric approaches,

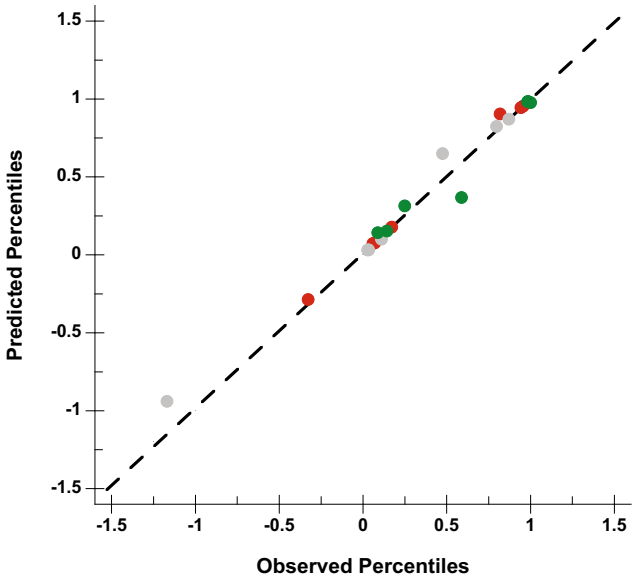


Fig. 3 Observed and predicted quartiles. The figure plots the observed median (*red circles*), quantiles corresponding to the 25th (*gray circles*) and the 75th (*green circles*) percentiles versus the corresponding predicted values of the median, and the quantiles corresponding to the 25th and 75th percentiles from Model 4 for overall adherence fraction, p_{SS}, p_{FS}, the fraction of overdosing events, the Poisson rate parameter for overdosing, and the mean and concentration parameters of the von Mises distribution for dose timing deviations. The *dashed line* is the 45° line of equality (Color figure online)

particularly those based on Dirichlet processes, are emerging as an alternative to Bayesian parametric approaches in application areas [26, 27] where the number of mixture components is not known. Mixture distributions and nonparametric approaches may potentially prove useful for large-scale population-wide adherence datasets from diverse diseases and across therapeutic classes.

We used an objective operational definition of persistence based on the last successful dosing event during the observation period. This operational definition does not penalize for drug holidays, which are modeled using the Markov chain component of the model. In real life settings, prolonged periods of non-adherence may occur that make it difficult to distinguish between persistence and the re-initiation of the dosing regimen. To address the impact of the observation period on persistence we used methods of survival modeling that consider right censoring of data. Additionally, it was necessary to re-parameterize the Weibull distribution to facilitate estimation in the Bayesian framework.

As noted in Methods, the parameterization of Model 4 was motivated by the helix-coil model which is used to describe DNA melting and protein unfolding in biophysics and is itself based on the one-dimensional Ising model in physics. Girard et al. described a multi-state Markov model coupled with normality assumptions for adherence modeling [28]. The potential similarities between the adherence modeling problem and the helix-coil model were first identified in early work from our group [15]. In the helix-coil model, the parameter σ is interpreted as a measure of cooperativity. Values of σ less than unity indicate positive cooperativity, whereas values greater than unity indicate negative cooperativity. One difference between the biophysical and adherence setting is that the distribution of σ in patient populations contains values that span both positive and negative cooperativity; negative cooperativity generally does not occur frequently in the biophysical setting. Another distinction is that the helix-coil model implicitly assumes an exponential Boltzmann distribution of polymer chains, whereas our Model 4 incorporates the Gamma distribution for σ and the lognormal distribution for s.

We used the DIC, a widely used measure of model discrimination in Bayesian modeling. The DIC is analogous to the AIC familiar to modeling and pharmacometrics researchers [29]. Based on the recommendations of Burnham and Anderson, models having AIC values within 1–2 units of the model with lowest AIC merit consideration, whereas models with AIC values >3 U from the model with the lowest AIC are considered substantially weaker and can potentially be ruled out [30]. All of our models differed by >25 units from each other and from Model 4. This margin of DIC is generally considered sufficient to rule out models that are not competitive [24, 31].

We envision potential applications of our modeling strategy in clinical trial simulations and population PK–PD modeling. Currently, resampling approaches of MEMS traces are used to incorporate adherence into modeling. However, the availability of an integrated Bayesian model that spans adherence, overdosing and persistence could substantially enhance these empirically driven methods because they provide greater control over simulations and more detailed insights into the underlying factors driving drug concentration variability.

In conclusion, this manuscript acknowledges Dr. Gerhard Levy's important scientific contributions to PK–PD and also to our understanding the consequences of inadequately addressing medication adherence issues.

Acknowledgments This paper is a tribute and acknowledgement of the numerous seminal contributions of Dr. Gerhard Levy to the pharmaceutical sciences. My research interests in adherence modeling were sparked by an incidental conversation with Dr. Gerhard Levy in the corridor many years ago. We take this opportunity to congratulate and felicitate Dr. Levy on his 50-year track record of scientific accomplishments. We are grateful to Colin Stoneking and Klaus Oberauer (Department of Cognitive Psychology, University of Zurich, Switzerland) for kindly providing the add-on code for analyzing the VM distribution. This research is not funded. Support from the National Multiple Sclerosis Society (RG4836-A-5) to the Ramanathan laboratory is gratefully acknowledged.

References

1. Blaschke TF, Osterberg L, Vrijens B, Urquhart J (2012) Adherence to medications: insights arising from studies on the unreliable link between prescribed and actual drug dosing histories. Annu Rev Pharmacol Toxicol 52:275–301. doi:10.1146/annurev-pharmtox-011711-113247
2. Levy G (1993) A pharmacokinetic perspective on medicament noncompliance. Clin Pharmacol Ther 54(3):242–244
3. Levy G, Zamacona MK, Jusko WJ (2000) Developing compliance instructions for drug labeling. Clin Pharmacol Ther 68(6):586–591. doi:10.1067/mcp.2000.110976
4. Rubio A, Cox C, Weintraub M (1992) Prediction of diltiazem plasma concentration curves from limited measurements using compliance data. Clin Pharmacokinet 22(3):238–246
5. Vrijens B, Goetghebeur E (2004) Electronic monitoring of variation in drug intakes can reduce bias and improve precision in pharmacokinetic/pharmacodynamic population studies. Stat Med 23(4):531–544. doi:10.1002/sim.1619
6. Vrijens B, Goetghebeur E (1999) The impact of compliance in pharmacokinetic studies. Stat Methods Med Res 8(3):247–262
7. Savic RM, Barrail-Tran A, Duval X, Nembot G, Panhard X, Descamps D, Verstuyft C, Vrijens B, Taburet AM, Goujard C, Mentre F, Group ACS (2012) Effect of adherence as measured by MEMS, ritonavir boosting, and CYP3A5 genotype on atazanavir pharmacokinetics in treatment-naive HIV-infected patients. Clin Pharmacol Ther 92(5):575–583. doi:10.1038/clpt.2012.137
8. Kenna LA, Labbe L, Barrett JS, Pfister M (2005) Modeling and simulation of adherence: approaches and applications in therapeutics. AAPS J 7(2):E390–E407. doi:10.1208/aapsj070240
9. Feng Y, Gastonguay MR, Pollock BG, Frank E, Kepple GH, Bies RR (2011) Performance of Cpred/Cobs concentration ratios as a metric reflecting adherence to antidepressant drug therapy. Neuropsychiatr Dis Treat 7:117–125. doi:10.2147/NDT.S15921
10. Fellows K, Rodriguez-Cruz V, Covelli J, Droopad A, Alexander S, Ramanathan M (2015) A hybrid Markov chain–von Mises density model for the drug-dosing interval and drug holiday distributions. AAPS J 17(2):427–437. doi:10.1208/s12248-014-9713-5
11. Rajarshi MB (2012) Markov chains and their extensions. Statistical inference for discrete time stochastic processes. Springer, New York. doi:10.1007/978-81-322-0763-4_2
12. Devore J (1976) A note on the estimation of parameters in a Bernoulli model with dependence. Ann Stat 4(5):990–992. doi:10.1214/aos/1176343597
13. Rudolfer SM (1990) A Markov chain model of extrabinomial variation. Biometrika 77(2):255–264
14. Islam MN, O'shaughnessy CD (2013) On the Markov chain binomial model. Appl Math 4:1726–1730
15. Wong D, Modi R, Ramanathan M (2003) Assessment of Markov-dependent stochastic models for drug administration compliance. Clin Pharmacokinet 42(2):193–204. doi:10.2165/00003088-200342020-00006
16. Agostinelli A, Lund U (2013) Circular statistics (version 0.4-7). Department of Environmental Sciences, Informatics and Statistics, Foscari University, Venice
17. Weibull model: problem with slow mixing and effective sample (2015). http://sourceforge.net/p/mcmc-jags/discussion/610036/thread/d5249e71/
18. Team RC (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
19. Plummer M, Stukalov A, Denwood M (2015) Bayesian graphical models using MCMC: interface to the JAGS MCMC library. http://mcmc-jags.sourceforge.net
20. Plummer M, Best N, Cowles K, Vines K (2006) CODA: convergence diagnosis and output analysis for MCMC. R News 6:7–11
21. Brooks SP, Gelman A (1998) General methods for monitoring convergence of iterative simulations. J Comput Graph Stat 7:434–455
22. Gelman A, Carlin JB, Stern HS, Rubin DB (2004) Bayesian data analysis. Texts in statistical science. CRC Press, Boca Raton
23. Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Stat Sci 7:457–511
24. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). JR Stat Soc B 64(4):583–639. doi:10.1111/1467-9868.00353
25. Plummer M (2008) Penalized loss functions for Bayesian model comparison. Biostatistics 9(3):523–539. doi:10.1093/biostatistics/kxm049
26. Fellingham GW, Kottas Athanasios, Hartman BM (2015) Bayesian nonparametric predictive modeling of group health claims. Insurance 60:1–10
27. Müller P, Quintana F (2004) Nonparametric Bayesian data analysis. Stat Sci 19(1):95–110
28. Girard P, Blaschke TF, Kastrissios H, Sheiner LB (1998) A Markov mixed effect regression model for drug compliance. Stat Med 17(20):2313–2333
29. Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19:716–723
30. Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, New York
31. Zhu L, Carlin BP (2000) Comparing hierarchical models for spatio-temporally misaligned data using the deviance information criterion. Stat Med 19(17–18):2265–2278