RESEARCH ARTICLE

A hidden Markov model analysis of subject-specific seizure time-series data as a potential aid to clinical decision making

[version 1; peer review: peer review discontinued]

Martin D. King¹, Suresh Pujar^{2,3}, Rod C. Scott²⁻⁴

¹Developmental Imaging and Biophysics, UCL Great Ormond Street Institute of Child Health, London, UK

²Developmental Neurosciences Unit, UCL Great Ormond Street Institute of Child Health, London, UK

³Great Ormond Street Hospital for Children National Health Service Trust, London, UK

⁴College of Medicine, University of Vermont, Burlington, USA

V1 First published: 28 Oct 2016, 5:2592
<https://doi.org/10.12688/f1000research.9746.1>

Latest published: 28 Oct 2016, 5:2592
<https://doi.org/10.12688/f1000research.9746.1>

Peer review discontinued

At the request of the author(s), this article is no longer under peer review. [What does this mean?](#)

Any reports and responses or comments on the article can be found at the end of the article.

Abstract

Background

The seizure-count time series data acquired from three children with refractory epilepsy were used in a statistical modelling analysis designed to provide an explanation for the marked variation in seizure frequency that often occurs over time (over-dispersed Poisson behaviour). This was motivated by an expectation that a better understanding of the spontaneous shifts in seizure-activity that are observed in some cases should reduce the risk of over-treatment caused by inappropriate changes in medication.

Methods

The analyses were performed using Poisson hidden Markov models (HMMs), both Bayesian and non-Bayesian, implemented using Markov chain Monte Carlo and the expectation-maximisation algorithm, respectively. A defining feature of the models, as applied to epilepsy, is the assumed existence of two or more pathological states, with state-specific Poisson rates, and random transitions between the states. Posterior predictive simulation was used to assess the validity of the Bayesian HMMs.

Results

The results are presented in the form of state transition probability and Poisson rate estimates (i.e., the primary HMM parameters), together with information derived from these primary parameters. State-specific mean-duration (sojourn time) estimates and sojourn-time complementary cumulative probability distributions are the main focus. HMM analyses are presented for three children that differed markedly in their seizure behaviour. The first is characterised by an extreme seizure count on one occasion; the second underwent a spontaneous decrease in seizure activity during the observation

period; the third seizure-count time trajectory is characterised by a gradual change in mean seizure activity. We show that, despite their considerable differences, each of the observed seizure-count trajectories can be treated adequately using an HMM.

Conclusions

The study demonstrates that clinically relevant information can be obtained using HM modelling in three cases with markedly different seizure behaviour. The resulting subject-specific statistics provide useful clinical insights which should aid those engaged in clinical decision making.

Keywords

Refractory epilepsy , seizure-count time series , Hidden Markov model (HMM) , Bayesian analysis , Markov chain Monte Carlo (MCMC) , Poisson rates , transition probabilities , sojourn-time cumulative distributions , Prediction

Corresponding author: Martin D. King (martin.king@ucl.ac.uk)

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2016 King MD *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: King MD, Pujar S and Scott RC. **A hidden Markov model analysis of subject-specific seizure time-series data as a potential aid to clinical decision making [version 1; peer review: peer review discontinued]** F1000Research 2016, 5:2592 <https://doi.org/10.12688/f1000research.9746.1>

First published: 28 Oct 2016, 5:2592 <https://doi.org/10.12688/f1000research.9746.1>

Introduction

Epilepsy is the most common severe neurological disorder in children. Treatment with antiepileptic medication is effective in rendering approximately 70% of children seizure free. The remaining 30% often have multiple therapeutic trials and frequently are managed with multiple drugs simultaneously, increasing the likelihood of side-effects. The failure to become seizure free after 3–4 antiepileptic drug trials reduces the likelihood of an individual becoming seizure free with any subsequent trial to less than 5%¹. As a consequence, the treatment goal usually changes at this stage to modification of seizure frequency rather than seizure freedom. It can be difficult to determine whether a change in seizure frequency has occurred following a change in medication, given the acceptance that these children have periods of better seizure control and worse seizure control which are not a function of medical intervention. An understanding of the natural fluctuations of a seizure disorder in an individual would enable physicians to manage therapeutic trials more effectively and allow a more precise assessment of whether a therapeutic intervention has caused an altered seizure pattern in a way that could lead to improvements in quality of life².

Statistical process control has been used for seizure frequency monitoring³, but this approach does not enable the physician to predict future patterns. Predictive information would allow the patient and family to recognise seizure patterns that are unexpected, and suggest when medical intervention could be warranted. Patients differ enormously in their individual seizure patterns and a reliable statistical tool that can be applied to different seizure patterns is critical for the development of appropriate treatment strategies.

Hidden Markov models (HMMs) provide an approach to seizure-count time-series analysis and seizure prediction (see, for example, 4–7). The key idea is the assumed existence of two or more disease states, with random transitions between them. These disease states are hidden, but the observed data (seizure counts) provide an indicator of these hidden states. Between-state switches occur according to a probabilistic rule. The transition probabilities and state-specific Poisson seizure rates constitute the model parameters, which are estimated by fitting the observed data. A more complete but informal description of the HMM approach to modelling seizure count data is given below, together with the more formal treatment provided in Note S1. A hidden Markov model schematic is provided in Figure 1.

Figure 1. A schematic illustrating a hidden Markov model. The model consists of two time series, the first of which is an unobserved Markov chain, Z_t , while the second is the state-dependent observed process, y_t . Given an m -state hidden Markov model, on each occasion Z_t takes on a value i , $i = 1, \dots, m$, where i is a state label.

In this study we describe the application of HMMs, both Bayesian and non-Bayesian, to the analysis of seizure count data acquired from children with severe epilepsy. The Bayesian analyses were implemented using Markov chain Monte Carlo (MCMC) while the non-Bayesian HMM analyses were performed using the expectation-maximisation (EM) algorithm. Posterior predictive simulation was used to assess the validity of the Bayesian models.

We show the results obtained from a detailed analysis of the seizure data collected from three children that differed markedly in their time-dependent seizure activity. In addition to their marked dissimilarity in seizure behaviour, these children were selected because no change in medication occurred during the observation period. The model parameter estimates, taken together with the other information derived from the model, are shown to be clinically informative.

Hidden Markov models

The analyses outlined in this study were undertaken using so-called discrete-time Poisson hidden-state Markov models. The basic notion is that the observations are a discrete time-series consisting of seizure counts, observed on a number of consecutive occasions. The central premise is that the observations emanate from an underlying system that switches between states, as illustrated in Figure 1. In some applications this assumption might be based on established theory. An alternative motivation for adopting an HM model for count data is as an empirical approach to dealing with overdispersion, i.e., the condition where the count data has a variance that is greater than the expected variance under the Poisson model. Although the present application belongs to the latter category, the possibility remains that an HM process is a realistic representation of the underlying physiological process that gives rise to the observed seizure behaviour. The key idea is that the observed counts are all that is available in terms of observable data and that the underlying process is hidden.

The HMM assumes that the hidden state switches (transitions) occur according to a probabilistic rule. More specifically it is assumed that the hidden states behave as a discrete-time Markov process. A Poisson HMM assigns differing Poisson rates to each of the underlying hidden states. Thus, as the system switches between states, the rates at which the observed events occur (seizures in the present application) are altered, hence the overdispersion. The challenge is to extract information about the hidden Markov process given the observed counts. Additional information is given in Note S1, which provides a more formal description.

Local and global decoding; seizure-state duration/sojourn time

An important feature of epilepsy seizure-count HM modelling is the potential to provide information regarding the state of the subject on a given occasion. In speech recognition and other applications this is referred to as decoding. Two types of decoding algorithms are available, namely local and global. Local decoding provides state probabilities at each occasion, while global decoding determines the sequence of states that is most likely, given the observations. This paper includes decoding results based on the so-called expectation-maximisation (EM) algorithm⁸, together with

posterior seizure-state probabilities provided by Bayesian HMM analyses. The resulting seizure-state time series data were used to generate seizure-state duration (sojourn time) probability densities and associated cumulative distributions.

Parameter estimation as the main objective

The primary objective for this study is the development of statistical models that provide information that is useful from a clinical perspective. Thus the main focus is a characterisation of the seizure states that, according to the model, underlie the observed seizure behaviour. This is motivated by the belief that a model-based description of the process will provide a clue to future behaviour. From a patient management perspective seizure-state duration (sojourn time) prediction is especially relevant, prompting us to focus on sojourn-time cumulative probability distribution results. Thus the main objective is the provision of model parameter estimates. These take the form of a point estimate, usually a median or mode, together with some measure of uncertainty.

Two alternative kinds of HMM analyses have been used in this work, namely traditional likelihood analyses performed using the EM algorithm, and Bayesian analyses implemented using MCMC^{9,10}. The main reason for using these two rather different approaches was to assess robustness. Each of the seizure time series reported in this study is relatively short in duration. Parameter estimation is, therefore, compromised because a short time series will include relatively few state transitions, rendering it difficult to obtain the HMM transition probability estimates with good precision. In clinical practice, decisions must sometimes be taken within a short time frame and our approach reflects this clinical reality. By comparing the results provided by two or more differing models/analyses, an indication of robustness is obtained. We stress that, in contrast to traditional model comparison analyses, where the aim is to find the best model, the objective in this study is different. Rather than declaring a given model as being 'best', the objective is to determine whether related models yield similar parameter estimates. Agreement is taken as an indication of robustness. Sampling methods were used extensively, particularly as a mechanism for obtaining measures of uncertainty. In addition to using MCMC to perform the Bayesian analyses and thus obtain the required Bayesian credible intervals, parametric simulation¹¹ was used to generate confidence intervals associated with various EM estimates.

Bayesian analysis, Markov chain Monte Carlo simulation and posterior predictive simulation

In the Bayesian context MCMC is, in effect, a random sampling procedure for evaluating complicated integrals in high dimensional space, as required to calculate various summary statistics. It is widely used in Bayesian analysis, the Gibbs sampler being particularly prominent among the available methods. Several software packages have been developed for performing Gibbs sampling, including WinBUGS¹², which was used in the present work (version 1.4.3). Presented with the data, a 'measurement model' (likelihood) and prior distributions, WinBUGS constructs the samplers required to generate the MCMC output. In order to avoid confusion we should emphasise that the decision to use MCMC to perform the analyses is entirely unrelated to the fact that the

statistical model is based on a Markov process. MCMC is just one of several methods that might be used for HMM parameter estimation, the EM algorithm providing an alternative.

A particular concern in the present study is the adoption of meaningful HMMs. Reasonable care must be taken to ensure the validity/adequacy of these models. Posterior predictive simulation is a Bayesian model checking procedure which is especially useful in modelling applications in which standard model checking methods, based on residuals analysis for example, are problematic. Discrete-data (counts, etc.) regression modelling is an obvious example¹³. The method is particularly useful for exploring potential model failure specific to the key research questions. The basic idea underlying posterior predictive simulation is the generation of replicate datasets which, in the MCMC setting, is achieved by using the posterior sample generated under the model in question. One or more key features of the observed data are compared with the replicate datasets. If these features are not seen in a majority of the replicates, this is taken as an indication of model failure. The comparison between the simulated and observed data can be performed using some discrepancy measure or can be undertaken using graphical methods. The latter approach has been suggested as particularly useful in practice^{9,14}. In the present epilepsy seizure study the occurrence of extreme counts (either low or high) is an obvious candidate for scrutiny. The model needs to capture these extremes. It is important to note that the replicate datasets incorporate both the uncertainty specified under the model (Poisson variation in the present context) and parameter uncertainty. It might also be noted that the replicate datasets fulfil a dual role in the present study. In addition to their model checking role they provide the clinician with an indication of what might be expected in terms of future seizure pattern in the patient under consideration, assuming no indication of model inadequacy. A key feature of the posterior predictive simulation assessment method is a probabilistic measure of model inadequacy. In this paper we use Bayesian p-values, which are a common choice. Additional information is provided in [Note S2](#).

The EM algorithm and parametric simulation

Hidden Markov models have a history going back to the 1960s, long before Bayesian statistics became accepted by mainstream data analysts. Although HMM parameter estimation can be tackled using standard minimisation (maximisation) methods, the EM algorithm is commonly adopted, based on well-established literature. The monograph by Zucchini and MacDonald (2009)⁸, provides a detailed account, together with a set of routines for implementing an HMM analysis based on both minimisation and the EM algorithm. A number of R packages also provide tools for performing HMM analyses, including the R package `msm`¹⁵.

Although maximum likelihood estimation in HM modelling is well established, the traditional approach to confidence interval and standard error estimation is to resort to asymptotic (large sample) approximation. The resulting intervals/standard error estimates can be inaccurate, however, and an alternative strategy is to avoid asymptotic approximation by adopting a sampling method. Zucchini and MacDonald (2009) [8, page 53], suggest the parametric bootstrap. In this study we have adopted a sampling approach referred to as parametric simulation [11, Chapter 2].

Methods

Subjects

Young Epilepsy is the United Kingdom's leading provider of education, treatment and residential care for children and young people with severe epilepsy. Located in Surrey just outside Lingfield, it has on-site 24-hour medical services. All young people have very close supervision by carers throughout the days and nights and these carers compile seizure diaries if seizures are observed. (Seizure diaries are used in similar organisations to *Young Epilepsy* and have been used in previous studies to monitor seizure frequency and severity¹⁶.) Evaluation of the diaries plays a major role in carers and medical staff assessing whether there have been substantial changes in seizure frequency and or severity. Thus, much detail and attention is given to their completion and accuracy. Total weekly seizure frequencies over a period of 1 year (starting in September 2004) were ascertained from the diaries for all the participants.

The project was registered at Young Epilepsy as a case note review. All applications for registration are reviewed by a formally constituted internal committee consisting of physicians, nurses and lay people. Questions related to ethics that are raised at that committee are referred to the East Surrey Ethics Committee. The local committee raised no ethical questions about our study. In addition the data collection was deemed to be audit and, under the charity's official rules, consent was not required. The data were not formally blinded because the study does not take the form of a clinical trial. Instead, the focus is the development of a statistical model, the objective being parameter estimation as an aid to clinical decision making. Thus the purpose is parameter estimation as distinct from hypothesis testing. That said, the analyses were performed by MDK after extracting the seizure time-series data from the original clinical records, thus removing all sources of identification. Prof Rod Scott and Dr Suresh Pujar had access to the clinical information as part of routine clinical care, but took no part in the development of the statistical models or data analysis.

Three children with epilepsy were selected for this study based on an examination of seizure frequency charts, subject to the restriction that no change in medication had occurred during the observation period (inclusion criterion). The three cases were chosen because they differed markedly in their time-dependent seizure behaviour, each presenting a particular challenge from a statistical modelling perspective. The central point to note regarding the choice of patients is the focus of this paper, namely seizure-count time series modelling, aimed at the provision of individual-specific parameter estimates, as opposed to group-average estimation. Our premise is that the analytical procedure can be applied to any individual, recognising that the model details must be tailored to the individual under investigation. This premise is supported by the demonstration that the method yields useful statistical information when applied to patients with disparate clinical presentations.

Bayesian models

Two distinct Bayesian seizure-count time-series models were adopted in this study, namely a log-linear HMM and a Poisson-gamma HMM. The details are as follows. It might be noted that Congdon (2006)¹⁷ provides relevant information on Bayesian HMMs, including their implementation in WinBUGS. It might

also be noted that HMMs are a class of mixture model, and that mixture model parameter estimation is not always straight-forward due to a phenomenon known as label switching. The label switching problem is discussed briefly in [Note S3](#).

Log-linear HMM

The following statistical model descriptions are given using an informal notation based on the WinBUGS distributions used in their implementation. This serves the dual purpose of providing a more intelligible outline for a non-mathematical readership and, at the same time, gives an outline of the key components of the WinBUGS code used to implement the models. The observed seizure counts (y_t), conditional on state (Z_t), are assumed to have a Poisson distribution. In the J -state case and assuming time-independent state-specific Poisson rates (and noting that several model parameters are completely defined within the hierarchical set of equations that follow and do not require further definition)

$$y_t \sim \text{Poisson}(\lambda_{Z_t}), t = 1, \dots, N \quad (1)$$

$$\log(\lambda_l) \sim N(0, \epsilon_p), l < \lambda_l < u \quad (2)$$

$$\log(\lambda_j) \sim N(0, \epsilon_p), \lambda_{j-1} < \lambda_j < u, j = 2, \dots, J \quad (3)$$

$$Z_t \sim \text{Categorical}(P_{Z_{t-1}, J}), t = 2, \dots, N \quad (4)$$

$$Z_1 \sim \text{Categorical}(P_{0,1,J}) \quad (5)$$

$$P_{i,1,J} \sim \text{Dirichlet}(\mathbf{1}), i = 0, \dots, J \quad (6)$$

where the symbol $\sim$ indicates 'distributed as', $N(., .)$ is the normal distribution with the specified mean and precision, t is an observation occasion label, N is the number of observations, ϵ_p is a small value, l and u are lower and upper constraints, $P_{Z_{t-1}, J}$ is a transition matrix row, $P_{0,1,J}$ is a J -length vector of probabilities with sum equal to one, j is the state label, J the number of hidden states and $\mathbf{1}$ is a J -length vector of ones. $\text{Categorical}(p)$ is the distribution of the N mutually exclusive states, as determined by the class probability vector p . (It is implemented in WinBUGS by the `dcat` distribution (page 352 in [18](#)), which 'returns' one of the J states at each MCMC iteration). The transition matrix row priors are given as Dirichlet distributions mainly for clarity, although the majority of the simulations performed in this study adopted an equivalent $\text{Gamma}(\alpha, 1)$ construction (pages 235 and 351 in [18](#)). In those cases where the state-specific Poisson rate parameters are a function of time the observation model becomes

$$y_t | Z_t, (\lambda_1, \dots, \lambda_J) \sim \text{Poisson}(y_t | \lambda_{Z_t}), t = 1, \dots, N \quad (7)$$

$$\log(\lambda_j) = \lambda'_j + f_t, j = 1, \dots, J \quad (8)$$

$$\lambda'_l \sim N(0, \epsilon_p), l < \lambda'_l < u \quad (9)$$

$$\lambda'_j \sim N(0, \epsilon_p), \lambda'_{j-1} < \lambda'_j < u, j = 2, \dots, J. \quad (10)$$

Thus, f_t adds a time-dependent offset to each of the Poisson rate parameters in the log domain.

Among the approaches adopted for incorporating time dependence in the state-specific Poisson means is the introduction of quadratic

dependence ($f_t = \beta_1 t + \beta_2 t^2$), the introduction of a temporal random effects term and the incorporation of a first order autoregressive term (abbreviated AR(1)). The AR(1) and random effect model details are given in [Note S4](#).

Poisson gamma HMM

The following Poisson-gamma model was adopted as an alternative to the log-linear model in those cases where the state-specific Poisson rates are time-independent. Using Poisson rate-parameter prior assignments based on George *et al.* (1993)¹⁹ and Gelfand and Smith (1990)²⁰, the observation model becomes

$$y_t \sim \text{Poisson}(\lambda_{Z_t}), t = 1, \dots, N \quad (11)$$

$$\lambda_j \sim \text{Gamma}(\alpha_j, \beta_j), j = 1, \dots, J \quad (12)$$

$$\alpha_j \sim \text{Exp}(1.0), j = 1, \dots, J \quad (13)$$

$$\beta_j \sim \text{Gamma}(0.1, 1.0), j = 1, \dots, J \quad (14)$$

$$Z_t \sim \text{Categorical}(P_{Z_{t-1}, J}), t = 2, \dots, N \quad (15)$$

$$Z_1 \sim \text{Categorical}(P_{0, J}) \quad (16)$$

$$P_{i, J} \sim \text{Dirichlet}(\mathbf{1}), i = 0, \dots, J. \quad (17)$$

Sojourn-time statistics

Hidden-state sojourn times (also referred to as dwell times) are key to the clinical problem. Many parents and patients recognise the fluctuating nature of epilepsy, but find it difficult to judge how long the child will be in any given state. Sojourn-time statistics enable the clinician to provide this important clinical information. The sojourn-time density for the j th state, $j = 1, \dots, J$ is given by²¹

$$d_j(s) = P(Z_{t+s+1} \neq j, Z_{t+s-v} = j, v = 0, \dots, s-2 | Z_{t+1} = j, Z_t \neq j) \quad (18)$$

where J is the number of states, $s \in \{1, \dots, M_j\}$, and M_j is the j th state sojourn time upper limit. Thus the sojourn time, s , is the number of observation occasions $t+1, \dots, t+s$, during which the unobserved process remains in the j th state, conditional on being in some other state on occasions t and $t+s+1$. Sojourn time probability distributions were derived from the MCMC output obtained for the transition matrix diagonal elements by substitution into the geometric distribution. Specifically, probability densities were generated using $g(P_{ii}^{(k)})$, $k = 1, \dots, K$, where $g(P_{ii}) = P_{ii}^{s-1}(1-P_{ii})$ and $P_{ii}^{(k)}$, $k = 1, \dots, K$ is the Gibbs marginal posterior sample obtained for the i th diagonal element of the transition matrix. (K is the MCMC sample size and (k) the iteration label. This is based on the standard approach to evaluating the marginal posterior distribution of a variable that is a function of the model parameters; see [22](#), for example.) To provide the results in a form that will be more familiar and accessible to many clinicians, the duration probability statistics are given as complementary cumulative (probability) distributions, i.e., the probability that the sojourn time is greater than a given value. It might be noted that these complementary cumulative probabilities are equivalent to the survivor function that arises in survival analysis.

Implementation

Gibbs sampling, convergence analysis and posterior predictive simulation. Gibbs sampling was performed using WinBUGS/GeoBUGS^{12,23,24}, which was downloaded from <http://www.mrc-bsu.cam.ac.uk/bugs>. The `dcat` (categorical distribution) and `ddirch` (Dirichlet distribution) are among the WinBUGS distributions used to implement the statistical models outlined above. Truncated normal samples were generated using the `WBDev` function `djl.dnorm.trunc`²⁵. Three parallel chains were generated in each analysis, each chain consisting of 5000 samples after thinning. (Thinning is the name given to the common practice whereby a specified proportion of the MCMC output is discarded, the remaining samples being stored for subsequent processing. It has been discussed by a number of analysts, including Carlin and Louis (2009) (Section 3.4.5 in [10](#)). The reason for thinning the MCMC output is to produce a chain with reduced sample autocorrelation. The aim is to reduce both the storage and post-simulation CPU demands of an analysis without suffering much loss in precision. A thinning factor of 10 was used in the majority of the simulations. Thus, 1 in 10 samples was stored and used in subsequent calculations.) The first of each set of three chains was started at an arbitrary position in parameter space, while the other two were started at over-dispersed positions. A burn-in set of samples was acquired prior to storing each chain of 5000 samples. These burn-in samples were also discarded.

Convergence tests are an essential part of any MCMC analysis. Nevertheless, given the exploratory nature of this study in which many models were investigated for each of the three cases, a complete set of convergence tests on all models and all parameters (i.e., both primary and derived parameters) was not undertaken. Instead, the convergence tests were restricted to the log-linear model MCMC output obtained for the primary parameters, i.e., the Poisson rates and transition matrix elements. This provides protection against a complete failure to achieve an adequate sampling of parameter space. With this safeguard in place, a comparison of the statistics provided by the alternative hidden Markov models adopted in each case, e.g. log-linear compared with Poisson-gamma, provides a mechanism for assessing the robustness of the results.

Informal convergence assessment was performed by inspecting overlaid trace plots for visual signs of failure. This was followed by semi-formal analyses (see [26](#) for a review of methods), which were performed using three convergence test procedures, namely the Gelman-Rubin shrink factor diagnostic and associated shrink factor plots (which is based on an ANOVA-like assessment of the between-chain and within-chain variances), the Geweke Z-score diagnostic and Z-score plots (based on spectral density variance estimation and a Z-score comparison of chain segments), and the Raftery-Lewis diagnostic procedure (which provides a variety of data, including an estimate of the number of iterations required to obtain a given quantile to some specified accuracy, taking into account the correlation between samples). These convergence analyses were performed using the R CODA (Convergence Diagnosis and Output Analysis software) package²⁷ (R version 2.15.2, CODA version 0.16-1).

As stated in the Introduction, posterior predictive simulation was used extensively to assess the capacity of the various Bayesian models to account for the observed seizure-count time-series data. Bayesian p-values were adopted as a measure of model inadequacy. The procedure was implemented as follows. At each MCMC iteration a simulated dataset (commonly referred to as a replicate dataset) was generated using the parameter values obtained at that iteration. The resulting replicate datasets and the observed data were used to generate a Bayesian p-value estimate, as given by the proportion of replicates satisfying [equation 19](#), based on the chosen discrepancy measure. For example, among the tests that were performed is the capacity of a given model to capture the observed maximum seizure rate for the subject in question. In this case $p_B = \#(T(y^{rep}) \geq T(y))/K$, where K is the number of MCMC iterations and $T(y)$ is the number of occasions on which the observed maximum rate occurred. Additional information is given in [Note S5](#), [Note S6](#) and [Note S7](#), while [Note S2](#) provides a brief introduction to posterior predictive simulation.

The EM algorithm and parametric simulation. The majority of the analyses undertaken in this study were performed using Bayesian statistical models, as outlined above. In some instances, however, traditional (non-Bayesian) HMM EM analyses were also undertaken for comparison, the results of which are shown for one of the three cases. These EM calculations, which were restricted to those analyses where the state-specific seizure rates (i.e., Poisson rates) are time invariant, were implemented using the R routines provided by Zucchini and MacDonald (2009)⁸. Parametric simulation [11, Chapter 2] was used to construct the associated confidence intervals. (Parametric simulation is a random sampling procedure in which simulated data are generated using the fitted model. Parameters of interest are recalculated from each simulated dataset. Confidence intervals and other statistics can be constructed using the resulting parameter sample distributions.)

Results

MCMC convergence

All overlaid trace plots were effectively ideal in their visual appearance except for some Poisson rate-parameter plots which showed random and very short-lived excursions to high values. But these excursions were observed in all three chains when they did occur, suggesting satisfactory coverage of the posterior distribution. The Gelman-Rubin shrink factors and associated plots were mostly exemplary; the remainder were all satisfactory. The Geweke Z-scores were also satisfactory, although the Geweke Z-score plots invariably showed one or more scores greater than 2.

The Raftery-Lewis analyses indicated that in two of the three cases, a sample of size 5000 (after thinning) was sufficient to provide the majority of 0.025 quantile estimates with an accuracy of ± 0.005 , or better, with probability 0.95. Thus the resulting nominal 95% credible intervals have a true coverage of between 94% and 96%, with probability 0.95. In a few instances the accuracy was slightly less, but mostly better than ± 0.0055 . In Case 1 this level of accuracy was not achieved for every parameter. Nevertheless the accuracy was at least ± 0.01 , which provides nominal 95% intervals with a true coverage of between 93% and 97%, with probability 0.95.

Dataset 1. Weekly seizure count data for each of the three cases presented in this report

<http://dx.doi.org/10.5256/f1000research.9746.d138204>

Case 1. No trend with an extreme seizure rate on a single occasion

The seizure count time series observed in the first case is unremarkable except for a single occasion when the rate was close to double that observed on the majority of occasions ([Figure 2](#)). The observed variance suggests overdispersion while a comparison of a 2-state HMM and a standard Poisson model is ambiguous, AIC selecting the 2-state HMM, while BIC suggests that the standard Poisson model is adequate. We note, however, that under the Poisson distribution with the observed mean value, the probability of observing a seizure count of 13 or higher in a series of 53 observations is close to 0.05, providing justification for an examination of the 2-state HMM. Three separate 2-state HMM analyses were undertaken, the first using the EM algorithm combined with parametric simulation confidence interval estimation, the second and third using the MCMC implementation of the Bayesian log-linear and Poisson-gamma models, respectively.

Posterior predictive simulation was performed as part of each MCMC analysis. This provides a powerful method for assessing the potential inadequacy of the model under consideration, as determined by a lack of capacity to capture key features of the data. Visual inspection of graphical displays and Bayesian p-values, both derived from the simulated data, are among the tools used for model evaluation. Similarly, parametric simulation was used to produce equivalent non-Bayesian p-values using the EM parameter estimates. As stated in the Introduction, and taken up in the Discussion, the aim of this exercise is not to compare the alternative models/analyses with a view to selecting one as being preferred. Instead, we have adopted 3 distinct analyses (the two Bayesian analyses and the non-Bayesian analysis using the EM algorithm) as a mechanism for assessing robustness. Broadly, the HMM parameter estimates and associated uncertainty intervals obtained from the three analyses should agree, and each should yield satisfactory predictive simulation results. The parameter estimates are examined in the

Figure 2. Case 1. Observed seizure count time series.

next section. None of the posterior predictive simulation tests led to the suggestion that the Case 1 2-state HM models under consideration fail in their capacity to capture the observations. Similarly, a graphical inspection of the replicate time series generated from this model assessment exercise (Figure 3) gives no indication of serious model inadequacy. The conclusion is that it is reasonable to use these models as a basis for characterising the observed seizure behaviour. The details are given in Note S5.

Case 1. Primary parameter estimates

The posterior marginal sample obtained for each parameter of interest provides an estimate of its value (e.g., median and mode) together with a measure of uncertainty. In this paper uncertainty in the primary parameter estimates is summarised using 95% posterior intervals. (Narrower intervals are given for some secondary parameters.) Similarly, parametric simulation was used to generate 95% confidence intervals for the EM parameter estimates. Three sets (EM and two Bayesian model sets) of 2-state HMM primary parameter median estimates (i.e., transition probabilities and Poisson rates) are listed in Table 1, together with their 0.025 and 0.975 quantiles. Unsurprisingly, given only 53 observations, the posterior/confidence intervals obtained for the transition probabilities are large. The Poisson rate-parameter estimates (λ_1 and λ_2) are similarly obtained with a relatively low precision but are, nevertheless, informative from the clinical perspective. The

agreement between the three analyses is satisfactory, indicating a degree of robustness in the results. In particular, parametric simulation based on the EM solution is unrelated to MCMC parameter estimation based on a Bayesian model. Thus the level of agreement between the EM and MCMC results supports a claim of robustness.

Case 1. Hidden-state (seizure-state) probabilities and mean-duration statistics

An important feature of epilepsy seizure-count hidden Markov modelling is the potential to provide information regarding the seizure state of the subject on a given occasion. As stated in the Introduction, considerable insight into the underlying random process can be obtained using the EM algorithm with local and/or global decoding. Local decoding provides information on the state probabilities at each occasion while global decoding provides state-sequence probabilities. Related information is provided by the marginal posterior distribution obtained for the hidden states, Z_t , $t = 1, \dots, N$ (i.e., the hidden seizure state at each occasion) given by the Bayesian analyses. In addition, mean duration statistics can be derived from the transition probability estimates.

Figure 4 shows the EM local decoding results, with superimposed error bars indicating the 0.025 and 0.975 quantiles given by the Poisson distribution with mean values taken from the EM solution.

Figure 3. Posterior predictive simulated time series plots. These posterior predictive time series were produced by sampling (evenly spaced samples) the posterior predictive (y^{rep}) output generated by the 2-state Poisson-gamma HMM. These can be interpreted as plausible clinical seizure patterns, and could be used during a clinical consultation to give a sense of what might be expected in the future. Isolated extreme values are observed in a few of the time series, suggesting that a single extreme observation should not necessarily lead to a change in treatment because this is an expected, albeit uncommon, event.

Table 1. Case 1. 2-state HMM parameter statistics.

Parameter	Algorithm - Model		
	EM/par sim	Poi-gamma	log-linear
λ_1	1.0 (0.0, 2.87)	1.28 (0.31, 4.09)	1.28 (0.27, 3.95)
λ_2	5.33 (4.48, 6.46)	5.31 (4.49, 6.48)	5.30 (4.49, 6.35)
P_{11}	0.031 (0.0, 0.61)	0.23 (0.01, 0.75)	0.23 (0.01, 0.70)
P_{12}	0.97 (0.39, 1.00)	0.77 (0.25, 0.99)	0.77 (0.31, 0.99)
P_{21}	0.15 (0.0, 0.56)	0.17 (0.04, 0.66)	0.17 (0.04, 0.59)
P_{22}	0.84 (0.44, 1.0)	0.83 (0.34, 0.96)	0.83 (0.41, 0.96)
d_1	1.03 (1.00, 1.27)	1.29 (1.12, 1.57)	1.29 (1.12, 1.57)
d_2	6.52 (4.24, 11.1)	5.77 (4.00, 8.74)	5.78 (3.99, 8.66)

2-state HMM Poisson parameter (λ) and transition probability (P) median estimates are listed, as given by the EM algorithm (combined with parametric simulation), the Bayesian Poisson-gamma model MCMC analysis and the Bayesian log-linear model MCMC analysis. The corresponding 0.025 and 0.975 quantiles are given in brackets. The hidden-state mean-duration (d) median estimates (given by $1/(1 - P_i)$, $i = 1, 2$) are also listed together with their 0.25 and 0.75 quantiles. The mean duration 95% intervals are large, due to uncertainty in the transition probabilities from which these are derived, coupled with the inflation that occurs as $1 - P_i$ tends to zero. We have, therefore, chosen to list the 0.25 and 0.75 mean duration quantiles, since these do provide useful clinical information.

Hidden-state credibility time-series plots derived from the Poisson-gamma model marginal posterior samples are included in the figure. The agreement between the local decoding results and the Bayesian hidden-state credibility trajectory is striking, given the very different nature of the analyses on which these two results are based. The hidden-state mean-duration median estimates are included in Table 1, together with their 0.25 and 0.75 quantiles. These statistics are derived from the transition matrix diagonal elements, and their imprecision is, in part, a reflection of the lack of precision in the

transition probabilities. The uncertainty in the mean duration estimate becomes inflated as the transition probability approaches unity. Nevertheless, the mean-duration 50% intervals are not so large as to render these estimates meaningless in the clinical context.

Case 1. Sojourn-time statistics

The main motivation for the HM modelling exercise outlined in this paper is the provision of information on the expected duration of each seizure state. Accordingly, the Case 1 2-state HMM transition-matrix posterior samples were used to calculate the sojourn-time probability densities, together with the associated cumulative distributions ($F(s)$, where s is sojourn time) and complementary cumulative distributions ($1 - F(s)$). The cumulative distributions (i.e., cumulative probabilities) provide an estimate of the probability that each seizure state will persist for up to some specified time, while each complementary cumulative distribution gives an estimate of the probability that a given state will persist for longer than the specified time. (As stated in the Methods section, this complementary cumulative probability is equivalent to the survivor function that is well known in survival analysis.) We have chosen to show the complementary cumulative distributions rather than the probability densities because they provide sojourn-time information in a form that is more easily communicated among clinicians. The Case 1 2-HMM log-linear model complementary cumulative probability distributions are shown in Figure 5.

In order to provide a clear clinical insight into the information that can be extracted from these distributions, and taking 2 weeks as an example, the median estimate obtained for the probability that each of the sojourn times exceeds 2 weeks ($\text{Prob}(s > 2)$) is 5.1% (25% and 75% posterior quantiles: 1.2%, 13%), and 68% (25% and 75% posterior quantiles: 56%, 78%) for States 1 and 2, respectively. Two points emerge from an inspection of Figure 5. Firstly, the estimation error is considerable. Secondly, regardless of the estimation error,

Figure 4. Case 1 (single extreme-value case) 2-state time-series plots. (a) Local decoding results. Local decoding was performed to determine which of the two states is most likely at each occasion. The estimated mean seizure rate ($\lambda_1=1.03$, $\lambda_2=5.28$) associated with each of the two states is shown by the trapezoidal-shaped line, with superimposed error bars showing the 0.025 and 0.975 quantiles given by the Poisson distribution. The observed seizure counts are shown as filled circles. (The global decoding results are identical.) **(b)** hidden-state credibility trajectories derived from the Bayesian Poisson-gamma 2-HM model. State credibility is defined as $P(Z_t = i | \text{data}) = \#(Z_t^{(i)} = i) / N_{\text{sim}}$ (see Note S5 for an explanation of the notation used in this expression). The trajectories given by the log-linear model are visually indistinguishable. **(c)** Plot indicating which of the two states is the most credible on each occasion, i.e., the state with the largest value of $P(Z_t = i | \text{data})$. The log-linear and Poisson-Gamma plots are identical.

Figure 5. Case 1. Low and high seizure-state sojourn-time complementary cumulative distributions, as given by the 2-state log-linear model. The centre curve (full line) shows the median estimate; the surrounding curves (dashed lines) show the 0.1, 0.25, 0.75 and 0.9 quantiles. From a clinical perspective these curves suggest that the patient is expected to be in a high seizure state more often than a low seizure state. The probability that a low seizure state persists for more than 10 weeks is negligible, whereas the median curve indicates that a high seizure state lasting more than 10 weeks would not be surprising. If a change in treatment led to a low seizure state for more than 10 weeks, then it would be clinically justifiable to conclude that the treatment change had been effective.

a considerable range of sojourn times must be expected; this is an inescapable consequence of the stochastic nature of the processes and is unrelated to parameter estimation error. Periods of high seizure frequency lasting many weeks are not unexpected under these HM models. Thus the occurrence of a long period of higher than average seizure activity does not necessarily indicate a change in the underlying pathology or a permanent deterioration. An awareness that clinicians should expect to see a considerable variation in the duration of unusual seizure activity might minimise excessive treatment.

Case 2. Changepoint behaviour. 3-state model results

The most striking feature of the time series data acquired from the second patient is an abrupt fall in seizure rate after the 23rd week (Figure 6), despite no change in treatment. The need to capture this change point, together with the occurrence of two occasions with an unusually high seizure rate, renders this second case somewhat more challenging than the first. Three alternative HMMs were examined in their capacity to capture the observations, namely a 3-state HMM with constant state-specific Poisson parameters (3-HM), a 3-state HMM with AR(1) time dependence in the Poisson parameters (3-HM-AR) and a 3-state HMM based on state-specific rate parameters that include a temporal random effect term with a conditional autoregressive prior distribution (3-HM-CAR). (The distinction between an autoregressive likelihood and an autoregressive prior is important.) Posterior predictive simulation was performed to examine the capacity of these models to capture important features of the data. Visual inspection of the plots given in Figure 6 indicates that the simple 3-state HMM appears to account for the main features of the data although the posterior predictive residuals show some structure. Adding AR(1) dependence offers no substantial improvement (graphical output produced using the AR1 model is not included in this report), while the 3-HM-CAR residuals show less departure from the ideal distribution about zero. Despite these observations, posterior predictive tests suggest that the simple 3-state models are adequate. Details are given in Note S6. Parameter estimates given by three 3-HM model analyses are listed in Table 2. An important

finding emerges from the model assessment exercise, namely that the random effect term that is incorporated into the 3-HM-CAR model tends to dominate the fitted time dependence, hidden-state transitions making a relatively small contribution to the seizure rate changes. Thus, although the 3-HM-CAR model provides an important standard against which the simpler models can be compared, in Case 2 we regard the 3-HM-CAR model as uninteresting from a clinical perspective (see Note S6 for additional information).

Case 2. Mean-duration statistics

The Case 2 hidden-state mean-duration statistics are listed in Table 2 together with the associated Poisson-rate parameter and transition probability statistics. The model parameter estimates together with the corresponding hidden-state trajectory plots (this hidden-state graphical output is not included in this report) indicate that transitions between adjacent states have a non-negligible probability both before and after the change-point and that the system is not held in a high-activity state prior to the changepoint and a low activity state afterwards.

Among the other points that emerge from the analysis is the considerable uncertainty in the hidden-state trajectory underlying the observed count data. The identity of the hidden state is particularly uncertain on some occasions. The fact remains that, under the Poisson distribution, a relatively high count can arise by chance even when the system is in a low seizure state, and cannot be attributed with certainty to an underlying high seizure-rate state. Thus, a series of consecutive occasions during which the seizure counts are consistently high does not rule out the possibility of underlying transitions between high and lower activity states. Furthermore, a sequence of consecutive occasions with no change in the most probable state does not preclude the possibility that transitions have occurred. Finally, Table 2 includes the 3-HM-AR Poisson rates and transition-probability estimates, showing that these differ substantially from those obtained with the simple 3-HM models. It is not surprising that introducing time dependence into the Poisson rate parameters substantially alters the transition statistics.

Figure 6. Case 2. Observed time-series data and posterior predictive simulation results. (Top row, column **a**) Time series data collected from a patient (Case 2) with two distinct seizure activity phases. (Columns **b** and **c**) Log-linear model posterior predictive simulation results generated using two alternative 3-state HMMs. The first is a standard 3-state HMM, while the Poisson rate parameters of the second model include an additional random effects term. Each top-row plot shows 10 replicate posterior-predictive time series generated by resampling the posterior predictive simulation output at evenly spaced MCMC iterations. The associated residuals plots are given in the bottom row.

Case 2. Other models

Additional analyses were performed using a number of 2-state and single-state models. Although the 2-HM-CAR model yields very satisfactory posterior predictive simulation results, it happens to exhibit the same random effect dominance described above for the 3-HM-CAR model. For that reason, in Case 2, it is uninteresting from a clinical perspective. Similarly, the single state-CAR model is dismissed as lacking in terms of the clinical insight it provides in Case 2, although it is not ruled out by the posterior predictive simulation analysis. The posterior predictive simulation results obtained with the 2-HM and 2-HM-AR models indicate that these models tend to produce Bayesian p-values that are low, although these do not fall below 0.05 (Bayesian p-value details not included). Based on these observations, the following section focusses on the sojourn time statistics derived from the time-independent log-linear 3-state model.

Case 2. Sojourn-time statistics

As in the previous case, we again resort to using the posterior marginal samples obtained for the transition matrix diagonal elements, in conjunction with the geometric distribution, as a mechanism for generating the state sojourn-time probability densities and their associated complementary cumulative distributions. **Figure 7** shows the results given by the log-linear 3-HM model. Taking an

example duration-of-interest equal to 2 weeks, the median estimate obtained for the probability that the sojourn time exceeds 2 weeks ($\text{Prob}(s > 2)$) is 46% (25% and 75% posterior quantiles: 12%, 71%), 25% (25% and 75% posterior quantiles: 6%, 50%) and 59% (25% and 75% posterior quantiles: 31%, 74%) for States 1, 2 and 3, respectively. Although the median curves might give the impression of a difference between States 2 and 3, the estimation error is too large to allow any definite conclusion other than the inescapable fact that, putting estimation error aside, the duration of each of the states is expected to exhibit considerable variability due to the stochastic nature of the process.

Case 3. Quadratic behaviour

A striking feature of the time series collected from Case 3 is the appearance of a quadratic time dependence in the underlying seizure activity (**Figure 8**). It is well known, however, that autoregressive processes can give rise to data that are deceptive in their shape, taking on a profile that appears deterministic. The preceding three 3-state HMMs (i.e., 3-HM, 3-HM-AR, and 3-HM-CAR models) were each investigated for their capacity to capture the observations, together with an HMM supplemented with a second-order polynomial (3-HM-quad). The corresponding 2-state models were also examined, together with the simple single-state quadratic model. Details concerning the various model adequacy tests that were

Table 2. Case 2. 3-state HMM parameter statistics.

Parameter	Algorithm - Model		
	log-linear	Poi-gamma	3-HM-AR
λ_1	3.70 (0.08, 5.26)	3.22 (0.003, 5.10)	0.25 - 0.32
λ_2	6.55 (3.95, 13.0)	5.64 (3.78, 12.56)	3.1 - 4.1
λ_3	12.73 (10.94, 17.24)	12.57 (10.83, 16.68)	10.5 - 12.3
P_{11}	0.68 (0.05, 0.96)	0.53 (0.03, 0.95)	0.27 (0.01, 0.75)
P_{12}	0.18 (0.005, 0.80)	0.27 (0.008, 0.84)	0.42 (0.03, 0.87)
P_{13}	0.095 (0.004, 0.57)	0.11 (0.004, 0.61)	0.21 (0.01, 0.80)
P_{21}	0.20 (0.01, 0.75)	0.20 (0.01, 0.75)	0.12 (0.01, 0.48)
P_{22}	0.50 (0.03, 0.92)	0.55 (0.03, 0.92)	0.74 (0.10, 0.95)
P_{23}	0.19 (0.01, 0.82)	0.16 (0.006, 0.80)	0.12 (0.004, 0.63)
P_{31}	0.066 (0.002, 0.42)	0.057 (0.002, 0.36)	0.028 (0.001, 0.14)
P_{32}	0.15 (0.01, 0.80)	0.12 (0.008, 0.75)	0.11 (0.01, 0.30)
P_{33}	0.77 (0.05, 0.95)	0.80 (0.08, 0.95)	0.86 (0.65, 0.97)
d_1	3.11 (1.54, 6.28)	2.11 (1.35, 4.94)	.
d_2	2.00 (1.33, 3.44)	2.22 (1.42, 3.79)	.
d_3	4.33 (2.26, 7.05)	5.12 (3.17, 8.12)	.

Case 2 3-state HMM parameter statistics are listed, as given by an MCMC implementation of the Bayesian log-linear and Poisson-gamma models, and the 3-HM-AR model. The first two data columns list the Poisson parameter (λ) and transition probability (P) median estimates, together with the 0.025 and 0.975 quantiles, which are given in brackets. The mean duration estimates (d), as given by $1/(1-P_i)$, $i = 1, 2, 3$, are also listed. The mean duration 95% intervals are large, due to uncertainty in the transition probabilities from which these are derived, coupled with the inflation that occurs as $1 - P_{ii}$ tends to zero. We have, therefore, chosen to list the 0.25 and 0.75 mean duration quantiles, since these do provide useful clinical information. The third data column provides the 3-HM-AR model statistics. In this case the Poisson rates vary with time and are therefore listed as ranges, as given by the median minimum and maximum estimates.

Figure 7. Case 2 seizure-state sojourn-time complementary cumulative distributions. Sojourn-time complementary cumulative distributions are shown, as given by the log-linear 3-HM model. The centre curve in each graph (full line) shows the median estimate; the surrounding curves (dashed lines) show the 0.1, 0.25, 0.75 and 0.9 quantiles. These curves, taken in combination with the parameter estimates given in Table 2, indicate that a treatment-induced reduction in the duration of the intermediate and high frequency states would be beneficial, in addition to any intervention that reduces the Poisson rates.

Figure 8. Case 3 time-series data. The observed seizure counts in the third patient give the appearance of an underlying quadratic trend in seizure activity.

performed are given in [Note S7](#). Here we focus on the 2-state quadratic model (2-HM-quad), which was found to be adequate (Bayesian p-values of 0.21 and 0.34 were obtained for the two curvature tests defined in [Note S7](#)). The transition matrix and mean state-duration statistics derived from this model are included in [Table 3](#).

Case 3. Sojourn-time statistics

[Figure 9](#) shows the Case 3 sojourn-time complementary cumulative probability distributions given by the 2-HM-quad model. If, as in the previous cases, we take a 2-week duration-of-interest, the median estimates obtained for the probability that the sojourn time exceeds 2 weeks ($\text{Prob}(s > 2)$) are 25% (25% and 75% posterior quantiles: 14%, 39%) and 51.8% (25% and 75% posterior

Table 3. Case 3. 2-HM-quad model parameter statistics.

Parameter	posterior median	posterior interval
λ_1	8.1 - 13.7	5.5, 11.6; 10.2, 18.3
λ_2	13.2 - 22.4	9.9, 18.1; 19.2, 26.6
P_{11}	0.50	0.12, 0.84
P_{12}	0.50	0.16, 0.88
P_{21}	0.28	0.08, 0.67
P_{22}	0.72	0.33, 0.92
d_1	2.0	1.6, 2.7
d_2	3.6	2.6, 5.0

The Poisson rate (λ), transition probability (P) and mean duration (d) median estimates are listed in the first data column. The rates, which are time dependent, are given as ranges, based on the median minimum and maximum estimates. The second data column provides the Poisson rate and transition-probability posterior 0.025 and 0.975 quantiles, and the mean duration 0.25 and 0.75 quantile estimates.

quantiles: 38%, 64%) for States 1 and 2, respectively. The considerable uncertainty in these estimates is a reflection of the limited amount of data available for analysis. This lack of precision must be taken into consideration if sojourn times are to be used in clinical decision making, albeit in conjunction with other information. Putting this imprecision aside, the results do serve to illustrate the important fact that, at any point in time, an additional and considerable uncertainty exists regarding imminent behaviour, due to the fundamentally stochastic nature of the process.

Figure 9. Case 3 seizure-state sojourn-time complementary cumulative distributions. Complementary cumulative distributions are shown for the low and high seizure-rate states, as given by the 2-HM-quad model. The centre curve in each graph (full line) shows the median estimate; the surrounding curves (dashed lines) show the 0.1, 0.25, 0.75 and 0.9 quantiles. The median curve shown in the right-hand plot suggests that high seizure-state episodes lasting more than 15 weeks are low in probability. The patient and physician could agree that if a future high seizure episode lasts more than 15 weeks, for example, a treatment change could be considered. The converse is that a treatment change would not be recommended unless this condition is met, minimising treatment modifications to those that are most appropriate.

Discussion

The major finding in the current study is that clinically important patient-specific information can be extracted from seizure frequencies recorded over time using HM modelling. This statistical approach was useful across a variety of seizure patterns with a small but clinically appropriate number of observations, suggesting that it should generalise to much of the population with medically intractable epilepsy.

The fluctuating nature of seizure frequencies in people with medically intractable epilepsy is increasingly being recognised (see 3, for example). Without this recognition patients with epilepsy frequently seek medical advice during times of increased seizure frequency and are often prescribed either higher doses of current antiepileptic drugs or new antiepileptic drugs are initiated. This can lead to treatment with excessive doses or polytherapy which results in an increase in the number of side effects^{3,28}. An unfortunate consequence is that quality of life is then impacted by the combined effects of the underlying etiology, the side effects of medication and the inability to completely prevent seizures. A way to prevent overtreatment is to understand the nature of the spontaneous changes that occur in seizure activity and thus determine whether a given altered seizure pattern should be rejected as a reason for a change in medication. We show that clinically relevant parameters can be obtained using HM modelling in three cases representing very different seizure frequency patterns.

In each of the three cases presented in this paper, the observed data are overdispersed relative to the Poisson distribution, suggesting the presence of at least two seizure states (low and high frequency states). The state-specific mean seizure frequencies and sojourn-time statistics are useful for informing patients and their parents of what might be expected in the future, assuming that the biological or environmental processes influencing seizure frequency remain unchanged. Case 1 is an example in which seizure fluctuations over the course of a year were largely stable. According to the HMM results, the high-frequency state dominates, the estimated probability of a low-frequency episode lasting longer than 2 weeks being about 5% (median estimate of the complementary cumulative probability). If this patient were to be treated during a high frequency state, a change to a low frequency state would often be interpreted as a positive therapeutic effect. The return to a high frequency state is frequently considered to be the end of the 'honeymoon period'. But, given the absence of any change in treatment during the study period in this individual, the observed differences in weekly seizure counts must be attributed to the underlying pathophysiology. It should be noted that, due to the relatively short duration of the time series used in this analysis, the precision of the 2-week complementary cumulative probability estimate given above is low (the 75% quantile is 13%).

The other two cases appear more complex, but similar information can be derived from the seizure data. Despite the absence of a change in therapy, Case 2 underwent a marked change in seizure activity during the observation period. A 3-state HMM was used in the analysis of these data, the results suggesting that the highest seizure-rate state dominated the first half of the observation period, after which the lower activity states were dominant (hidden-state

plots not shown). Given the data acquired during the first half of the observation period, one might have predicted that the pattern would have continued. The seizure pattern did change, however, and a question for the clinician is whether the patient might revert to the high frequency state. According to the simple 3-state models (the two Bayesian 3-HM models and corresponding AR analysis), the probability of a transition from the intermediate state to the high frequency state is in the range 0.12 to 0.19 (based on median estimates; the posterior/confidence intervals are given in Table 2). Similarly, the probability of a transition from the lowest to highest activity state is non-negligible. This is important information for families who need to be prepared for a return to the highest frequency state, and should be informed about the possibility of this happening.

Case 3 appears to exhibit a quadratic time dependence in seizure rate, which can be captured by including linear and quadratic terms in the log-linear expression for the HMM Poisson rates. Despite the time dependence in the resulting state-specific rates, the hidden state transitions remain an important source of seizure-frequency variation. The transition probabilities and sojourn-time statistics thus retain their clinical significance. It should be noted that, while the HM modelling approach outlined in this paper is general in its application, and expected to apply to any seizure-count time series that might be encountered, the model details must be tailored to each individual. Thus, although simple HMMs with time independent Poisson rates offer the advantage of simplicity, cases will arise where more complex models are required. Case 3 (quadratic case) provides an example where simple 2- and 3-state HMMs were demonstrably inadequate. Although we concluded that, for differing reasons, the AR and CAR models had little to offer in the present three cases, we envisage cases where these will be useful for dealing with subtle but non-ignorable trends in the state-specific rates. Furthermore, the CAR (random effect) HMMs provided a flexible and useful standard against which to judge simpler models in their capacity to produce satisfactory posterior predictive simulation residuals.

This paper demonstrates that HMMs provide a characterisation of case-specific seizure patterns in a manner that is relevant to both the clinician and their patients. Clinical trials in patients with epilepsy tend to use average seizure frequencies over a time period of about 6 weeks. However, this approach may miss important information about changes in seizure pattern that could be captured using some form of HMM. This raises the question regarding the use of seizure pattern as an outcome in future drug trials, rather than simply averaging seizure frequency. Clearly, a typical clinical trial aims to provide population statistics, which appears to be very different to the present work where the focus is parameter estimation and decision making as it applies to the individual. The present results show that one cannot expect to obtain subject-specific HMM parameter estimates with good precision, given a single observation period that is restricted to a few weeks. It might be argued that a solution to this problem is to adopt a pooled-data (population modelling) analysis. This is a well-documented approach to improving the precision of subject-specific estimates when the data are sparse^{29,30}. Thus, population modelling might be considered in the present subject-specific estimation context, in addition to its obvious application

in clinical trials. Unfortunately, we do not have the information required to undertake a meaningful pooled-data analysis. The central problem is the marked difference between individuals regarding their seizure behaviour. Although patient heterogeneity has been dealt with in other applications by including covariates in a pooled-data model, this requires a comprehensive set of covariates to account for important differences between cases. It assumes a detailed knowledge of the relationship between the covariates and the model parameters³⁰. We do not have the data required to construct a general pooled-data model involving covariates. In particular, we do not have an exogenous variable for defining 'homogenous' sub-populations.

A final point needs to be addressed concerning the relatively poor precision of the parameter estimates reported in this paper. Some statisticians might argue that statistical modelling should not be attempted with short time series data. As a rejoinder we would suggest that important clinical decisions are currently made without any objective way of evaluating potential changes in seizure activity over time. This has a negative impact on the ability of the clinician to provide high quality and safe care to their patients. Thus, although we recognise that the model parameter estimates are compromised due to the necessity to work with short time series, the information provided by the models is far superior to anything currently available. Re-stated, a parameter estimate with a wide probability/confidence interval is better than pure guesswork.

Among other objections that might be raised regarding the approach outlined in this paper is the level of user input it requires. This precludes its use in some clinical settings. We have emphasised that the details of the statistical model (and model adequacy tests) must be tailored to each subject-specific dataset. While the lack of a black box implementation rules out the possibility of using HM modelling as a routine clinical tool, that does not preclude its use in a tertiary referral centre dealing with the small proportion of patients with severe and intractable epilepsy. In that setting the assistance of a data analyst is a realistic possibility. For those who choose to reject this proposition, the results presented in this paper do, at the very least, suggest the need for some caution in responding to an apparently sudden increase in seizure activity.

An interesting issue that arises in this work is an uncertainty regarding the interpretation that might be attached to a given HMM result, beyond the immediate clinical interpretation outlined above. There is a distinction between using a result as a convenient description of the observations, versus the supposition that it provides a real description of the neuronal processes underlying the observed seizures, thus providing real mechanistic insight. Some might reject the second of these perspectives, based on a premise that the neuronal processes are entirely deterministic and that stochastic models happen to capture the observations. A stochastic model might therefore be seen as a pragmatic solution to the absence of information on the underlying deterministic control system(s). We note that the discrimination between stochastic and deterministic behaviour in neuronal systems has been addressed (see, for example, 31,32). It might also be noted, however, that the very notion of tests designed to distinguish stochastic and deterministic behaviour might be questionable. Werndl (2009)³³, for example, states that every stochastic process is observationally equivalent to a deterministic system, and

that many deterministic systems are observationally equivalent to stochastic processes. Regardless of these more philosophical questions, for the clinician faced with the problem of patient management, a model that merely provides a characterisation of the seizure time-series data remains useful, especially if it is reliable in its predictions. It does not matter if the model merely provides an empirical description of the observed seizure pattern, unrelated to the underlying pathophysiological process.

Comments on the models and methods, including model assessment and convergence tests

To keep this paper to a reasonable length we have not shown a comprehensive set of model assessment results, as obtained for every model and every case. Instead the content of the paper is designed to illustrate some of the modelling possibilities. The majority of posterior predictive simulation results are based on tests epileptologists will consider relevant to the clinical problem, for example, the number of occasions with an extreme seizure count. A related issue is whether or not the models should be regarded as definitive. We do not regard model selection with a view to finding the 'best' model as being among the objectives. In this context there is a need to distinguish between modelling the data and modelling the process that gives rise to the data. In the latter setting model comparison and the search for the best model has a greater relevance. From a clinical perspective interest might be focused on modelling the data while those engaged in basic science research may be more interested in modelling with a view to shedding light on the underlying pathophysiology. It is our view that a clinician whose concern is patient management will be primarily interested in empirical statistical models that provide a useful characterisation of the observed time series and an insight into future seizure activity. Given this objective, we do not seek to find the best model. Instead, a comparison of the results derived from differing models is undertaken as a mechanism for assessing robustness. If different models yield consistent parameter estimates this is taken as an indicator that the estimates are sound. In that sense we do not necessarily regard alternative models as competing. This is the reason for showing the results given by both the Poisson-gamma and log-linear Bayesian models where applicable. As shown, we obtained reasonable agreement between the two models. Given the differing form of the likelihood and priors under these two models, together with the Poisson parameter ordering constraint that is incorporated into the log-linear model, the similarity in the parameter estimates suggests that these are robust, i.e., the results exhibit a desirable insensitivity to the priors and are predominantly determined by the data. Similarly, in those cases in which the EM algorithm was adopted and combined with parametric simulation (an example is given in Table 1), agreement between the resulting confidence intervals and the posterior marginal distributions derived from the corresponding Bayesian analyses is an indication of robustness. Having taken this approach to assessing the reliability of the statistics obtained for a given subject, we have attached less importance to convergence testing and simulation accuracy calculations. For example, agreement between the MCMC results given by a Bayesian model and those produced using parametric simulation based on an EM solution would be unexpected if effective convergence to a stationary distribution had not been achieved in the former case. That said, MCMC convergence test results are reported, as obtained for key model parameters.

Conclusion

In conclusion, we have presented results obtained by fitting HMMs to the seizure data collected from three paediatric epilepsy patients. We suggest that these models provide useful clinical insights. In particular, we propose that sojourn time cumulative distributions are informative and a useful aid to clinical decision making in epileptology. Thus the HM modelling approach has the potential to make a major contribution to patient management, and could ultimately maximise quality of life.

Data availability

F1000Research: Dataset 1. Weekly seizure count data for each of the three cases presented in this report, [10.5256/f1000research.9746.d138204](https://doi.org/10.5256/f1000research.9746.d138204)³⁴

Supplementary material

Note S1

Hidden Markov models. A discrete first order Markov process is a stochastic process in which the state probabilities on a given occasion depend only on the preceding state and, given that previous state, are independent of the more distant history. The term ‘conditional independence’ is used in this context. To express this a little more formally, and using the notation $Z_1, Z_2, Z_3, \dots$ to represent a sequence of states, then

$$\Pr(Z_{t+1} | Z_t, \dots, Z_1) = \Pr(Z_{t+1} | Z_t).$$

Key to a hidden Markov model is the probability of a transition between any two states, $Z_t = i$ and $Z_{t+1} = j$, i.e., the single-step transition probabilities (this description is restricted to so-called homogeneous chains which assume the transition probabilities to be time independent)

$$\gamma_{ij} = \Pr(Z_{t+1} = j | Z_t = i).$$

The count (y_t) observed on a given occasion is assumed to have a Poisson distribution with a state-specific rate, λ_{z_t} . Thus, given an HMM with m states and associated Poisson parameters $\lambda_z, z = 1, 2, \dots, m$,

$$y_t | Z_t, (\lambda_1, \dots, \lambda_m) \sim \text{Poisson}(y_t | \lambda_{z_t}), t = 1, \dots, T$$

Note S2

Posterior predictive simulation and Bayesian p-values. Posterior predictive simulation has been used extensively in this study, and it is appropriate to comment on this approach to model assessment. Gelman is among those who have championed this method (see, for example, [13,14,35–37](#)) and we find the arguments appealing. A central proposition is that no statistical model should be regarded as correct in every detail (in essence, all models are wrong; see [37](#), for example) and that what matters is model adequacy as it relates to the research question and resulting statistical

Author contributions

MDK conceived, designed and performed the analyses, in addition to writing the statistical modelling parts of the manuscript. RS formulated the clinical application and wrote the clinical parts of the manuscript. SP collected the clinical data, compiled the seizure diaries and approved the manuscript.

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

inferences. From that perspective, model inadequacies that have little impact on the statistical inferences and resulting conclusions can be ignored. For this reason our posterior predictive simulation analyses focused on the capacity of each model under consideration to capture the extremes in seizure activity. From a clinical perspective this is the primary requirement of any candidate model.

Bayarri and Castellanos (2007)³⁸, provide useful additional information regarding this model checking procedure, listing the choices that must be made. These are (1) the discrepancy measure or test statistic used to measure incompatibility between the observed data and the model, (2) the distribution of the discrepancy measure/statistic, under the model (in the present context this is provided by the posterior predictive distribution), and (3) a method for measuring conflict between the observed statistic (value) and this distribution; a measure of surprise. Tail areas (posterior predictive p-values) are a commonly used measure. In this paper we refer to these tail areas as Bayesian p-values; these were calculated as outlined in the following paragraph. Posterior predictive p-values (Bayesian p-values, p_B) were obtained for a given model by performing posterior predictive simulation and using selected test statistics to compare the observed and replicate data. Thus, given some suitable statistic, the observed value is compared with its posterior predictive distribution. Following Gelman *et al.* (2004) (page 162 in [9](#)) a Bayesian p-value is defined as the probability that replicate posterior predictive data (y^{rep}) might be more extreme than the observed data based on the chosen discrepancy measure, $T(y, \theta)$. Thus

$$p_B = \Pr(T(y^{rep}, \theta) \geq T(y, \theta) | y)$$

taken jointly over the posterior predictive distribution of y^{rep} and the posterior distribution of θ . An estimate of p_B is given by the proportion of L simulations satisfying

$$T(y^{rep(l)}, \theta^l) \geq T(y, \theta^l), l = 1, \dots, L. \quad (19)$$

In those instances where this yields a p-value greater than 0.5 the result is given as $1 - p_B$. Thus improbable simulation test results are indicated by Bayesian p-values near zero. The majority of discrepancy measures adopted in this study do not depend on unknown parameters in which case $T(y, \theta) = T(y)$.

Note S3

Label switching, label sorting and identifiability constraints. HMMs are a class of mixture model, the term ‘mixture model’ referring to statistical models consisting of a mixture of probability distributions. In fact, one approach to HMM parameter estimation is to treat it as a standard mixture model problem and to use a conventional minimisation routine to find the maximum log-likelihood⁸. The recognition that HMMs are mixture models is relevant to the present study because it is well known that a so-called label-switching problem arises in mixture model parameter estimation, and it follows that the problem must arise in HMM applications. The central issue is a ‘symmetry’ in the likelihood caused by invariance to permutation of the mixture component labels. Re-expressed, a complication arises because relabelling the mixture components (swapping the component labels) has no effect on the likelihood. In fact, all permutations of the component labels produce the same likelihood. This might be regarded as a lack of identifiability.

A number of solutions have been suggested, including the incorporation of identifiability constraints to prevent label switches (page 465 in 9; 39, including the discussion), and the application of relabelling algorithms⁴⁰. In the present study, an identifiability constraint was incorporated into the log-linear model in the form of an ordering constraint on the Poisson rates, an approach which is expected to work well when the component Poisson parameter distributions are well separated. No such constraints are included in the Poisson-gamma model. Instead post-simulation label sorting is applied to the MCMC output. We adopted a simple approach in which this was achieved by ordering the Poisson rate parameters by magnitude and subsequently sorting the transition matrix parameters as dictated by the rate-parameter label changes. This is clearly related to the order constraints included in the log-linear model. Thus, relabelling an entire set of parameters based on the sorted order of a single parameter is expected to yield valid results when the chosen parameter consists of component distributions that are well separated. More sophisticated approaches have been suggested (see, for example, 40,41). We acknowledge that label switching is especially important when mixture-model/HMM parameter estimation is performed using MCMC because it has implications regarding the need for convergence to a stationary distribution. In fact, convergence in mixture-model MCMC and the related matter of order constraints has been the subject of some controversy (see, for example, 39,42,43), one that we have summarised briefly in a previous paper⁴⁴. One or other of the more advanced methods for dealing with label switching might be explored in a future study, given the weakness of the well-separated distribution assumption in some cases.

Note S4

Temporal random effects and first order autoregressive models.

In some children with epilepsy there is an apparent time dependence in seizure activity, in addition to the observed state-switching behaviour. A temporal random effects term ($f_i = \gamma_i$ in equation 8) with an intrinsic Gaussian conditional autoregressive (CAR) prior distribution^{45,46} (also see Example 9 in 24 for additional information) was therefore adopted as one approach to incorporating time dependence into the state-specific Poisson rate parameters. In the time series context the CAR prior takes the form

$$\gamma_i | \gamma_v, i \neq v \sim N(\bar{\gamma}_i, w_i^2 / m_i) \quad (20)$$

where $\bar{\gamma}_i = \frac{1}{m_i} \sum_{v \in \delta_i} \gamma_v$ is a variance parameter, and m_i is the number of ‘adjacent’ observations, i.e., 2 for all but the first and last occasions. δ_i denotes the set of occasion labels belonging to those occasions that precede and follow the i th occasion. Thus, the prior conditional mean is the average of the adjacent observations. A complication arises in the present implementation, however, because the CAR random effects sum to zero and must be accompanied by an intercept term (see equation 23), the latter of which must be assigned a flat prior. As a consequence of the ordering constraint that is imposed on the log-linear Poisson rates, a change of variables is required to transform the Gibbs sampler rate parameters to the constrained log domain. This change-of-variables transformation was implemented, as follows, based on the zeros trick⁴⁷ (noting the altered definition of λ' in the following):

$$\log(\lambda_{jt}) = \lambda'_{jt}, j = 1, \dots, J, t = 1, \dots, N \quad (21)$$

$$\lambda'_{it} = L + (U - L)e^{\lambda''_{it}} / (1 + e^{\lambda''_{it}}) \quad (22)$$

$$\lambda''_{it} = c + \gamma_i \quad (23)$$

$$\lambda'_{jt} = \lambda'_{j-1,t} + a_j, j = 2, \dots, J \quad (24)$$

$$a_j \sim N(0, \epsilon), 0 < a_j < U', j = 2, \dots, J \quad (25)$$

$$c \sim \text{flat} \quad (26)$$

$$\gamma_i \sim \text{car.normal} \quad (27)$$

$$X_i = 0 \quad (28)$$

$$X_i \sim \text{Poisson}(g_i) \quad (29)$$

$$g_i = -\log(w_i) \quad (30)$$

$$w_i = (U - L)e^{\lambda''_{it}} / (1 + e^{\lambda''_{it}})^2 \quad (31)$$

where λ'_{jt} is the j th Poisson rate in the log domain at the t th observation time and λ''_{it} is the corresponding unconstrained Gibbs sampler parameter for the first state, L is a lower constraint, U and U' are upper constraints, `car.normal` is the intrinsic Gaussian CAR prior (defined in equation 20 and provided by the GeoBUGS `car.normal` distribution) and `flat` is an improper uniform prior

(provided by the WinBUGS dflat distribution, see page 345 in 18). w_t is the likelihood weighting required under the change-of-variables rule.

An AR(1) model was also investigated as an approach to incorporating time dependence into the state-specific Poisson rates. In this case the random coefficients in the preceding change-of-variables treatment were replaced with the following autoregressive parameters (thus $f_t = \gamma_t$ in equation 8 becomes autoregressive):

$$\gamma_t = \phi \gamma_{t-1}, t = 2, 3, \dots, N \quad (32)$$

$$\phi \sim N(0, 0.01), 0 \leq \phi < 1 \quad (33)$$

$$\gamma_1 \sim N(0, 0.01). \quad (34)$$

(It might be noted that Congdon (2001)⁴⁸ discusses autoregressive models of temporal correlation, and their implementation in WinBUGS.)

Note S5

Case 1. 2-state HMM posterior predictive simulation and Bayesian p-value results. Two strategies were adopted as a mechanism for assessing HMM performance. The first was based on Bayesian p-value generation using tests designed to probe the capacity of the model in question to capture crucial features of the data. The following results were obtained for Case 1. Tests based on observing 1 occasion with 13 seizures ($T(y) = \#\{y_{t \in \{1 \dots N\}} > 12\}$, $p_B = \#(T(y^{rep}) \geq T(y) = 1)/K$, where K is the number of MCMC iterations): log-linear model $p_B = 0.15$, Poi-gamma $p_B = 0.15$, EM $p = 0.13$. (Throughout this paper the notation $t \in \{n_1 \dots n_2\}$ is used to indicate that the occasion label, t , takes on the set of values n_1 to n_2 . Notation of the type $\#\{y_{t \in \{n_1 \dots n_2\}} > C\}$ is used to indicate the number of occasions within the specified set of occasions on which a given condition is satisfied.) Tests based on observing 2 occasions with no seizures ($T(y) = \#\{y_{t \in \{1 \dots N\}} < 1\}$, $p_B = \#(T(y^{rep}) \leq T(y) = 2)/K$; i.e., test with sign reversal): log-linear model $p_B = 0.5$, Poi-gamma $p_B = 0.49$, EM $p = 0.37$. These tests are based entirely on a comparison of the observed and simulated data, using a discrepancy measure based solely on the data, i.e., based on a comparison of y and y^{rep} only. It is common practice, however, to construct discrepancy measures that depend on the data and unknown model parameters. An example of the latter is the test based on $T(y, \theta) = \max(y_t - \lambda_{z_t})$ where λ_{z_t} is the Poisson rate parameter conditional on state. The resulting log-linear model Bayesian p-value is 0.14. Although some of these Bayesian p-values are not far from the tail-area of probabilities, we do not reject the 2-state HMM as inadequate.

The second strategy uses visual inspection of the replicate datasets generated by simulation (and residual plots) to look for systematic differences between the observations and the replicate data. The usual practice in which the results are displayed as overlaid profiles is not helpful in this instance because it obscures the key information, namely, the proportion of predictive time series lacking one or more extreme periods of seizure activity (count > 12). For this reason a subset (obtained using evenly spaced MCMC samples) of 10 time series are shown as separate plots (Figure 3). Several of these time series fail to mimic the observation that an unusually high seizure count was observed on one occasion (Figure 2), and this

is reflected in the associated p-values given above (0.13 to 0.15). Taken together, the Bayesian p-values and posterior predictive time series plots indicate that the 2-state HMM is not demonstrably inadequate. The one occasion with an observed count greater than 12 might indeed be a low probability event.

Note S6

Case 2. Posterior predictive simulation and Bayesian p-value results. A number of tests were constructed to examine the capacity of each of the three models (3-HM, 3-HM-AR, 3-HM-CAR) to capture the observations, paying particular attention to the change point. These include a test on the mean during the interval 22 to 27 weeks (i.e., $T(y) = \bar{y}_{t \in \{22 \dots 27\}}$, selected because it yielded the smallest/least favourable p-values), maximum difference in weekly seizure counts during weeks 1 to 20 ($T(y) = \max(y_{t \in \{1 \dots 20\}}) - \min(y_{t \in \{1 \dots 20\}})$), maximum difference in weekly counts over the entire observation period ($T(y) = \max(y_{t \in \{1 \dots N\}}) - \min(y_{t \in \{1 \dots N\}})$), high seizure count based on observing two occasions with a count > 18 ($T(y) = \#\{y_{t \in \{1 \dots N\}} > 18\}$), low seizure count based on observing 5 occasions with count < 3, ($T(y) = \#\{y_{t \in \{1 \dots N\}} < 3\}$), and the difference between the mean counts in the first and last ten observations, ($T(y) = \bar{y}_{t \in \{1 \dots 10\}} - \bar{y}_{t \in \{N-9 \dots N\}}$). Satisfactory results were obtained for all the tests, each of the models giving p_B values that were mostly well removed from the tail of their distributions. The lowest p_B value of 0.2 occurred in the test on the mean seizure count during the interval 22 to 27 weeks, as given by the 3-HM model. Thus, the Bayesian p-values indicate that the simplest of the three models (the 3-HM model) is not demonstrably inadequate. That said, visual inspection of the posterior predictive simulation plots given in Figure 6 indicates that the 3-HM model yields residuals that retain a degree of dependence on time, while the 3-HM-CAR residuals appear more random. The 3-HM-AR residuals appear somewhat less satisfactory than the 3-HM-CAR residuals (not shown).

A final point that emerged from the model assessment exercise is that the process captured by the random effects term that is incorporated into the 3-HM-CAR model dominates the fitted seizure time dependence, with hidden-state transitions playing a minor role. This behaviour renders this model uninteresting from a clinical perspective in this case, although it remains important as a standard against which the simpler models can be compared. It should be noted that this observation does not exclude the possibility that a random effect HMM might be useful in other cases, and provide the required clinical insight.

Note S7

Case 3. Posterior predictive simulation and Bayesian p-value results. Inspection of the posterior predictive simulated data and associated residual plots obtained for the third case (Figure S1) suggests that the 3-HM fails to capture fully the underlying curvature (similarly the 3-HM-AR model is not entirely satisfactory; plots not shown), while the 3-HM-CAR model gives the appearance of being adequate. Further investigation of the 3-HM-CAR model indicates, however, that it exhibits the same complete dominance of the random effect term that was observed in the previous case. The 2-HM-CAR model exhibits the same behaviour and these two models are dismissed as clinically uninteresting. The posterior predictive time series and residuals plots generated using the 3-HM-quad

model (Figure S2) provide little suggestion of model inadequacy, a conclusion that is supported by the Bayesian p-values given by two curvature tests ($T(y) = \bar{y}_{t \in \{1 \dots 5\}} - \bar{y}_{t \in \{21:25\}}$ and $T(y) = \bar{y}_{t \in \{N-4 \dots N\}} - \bar{y}_{t \in \{21:25\}}$) designed to detect systematic model departure from the true underlying trend. These tests yielded Bayesian p-values of 0.23 and 0.36, respectively. We note that a more standard procedure would use a single test constructed by combining these two tests, $T(y) = y_{start} - 2 * y_{mid} + y_{end}$, for example, but we have chosen to examine the two components separately.

The preceding results indicating a preference for the 3-HM-quad model over the alternative 3-state models motivates an examination of the two-state and single-state quadratic models. The 2-HM-quad model yields $p_B = 0.21$ and 0.34, respectively, for the two curvature tests, which are satisfactory. These results are broadly consistent with the residuals plots generated by posterior predictive simulation

(Figure S3), although these do appear marginally less random than those derived from the 3-HM-quad model, suggesting a degree of inadequacy in the 2-HM-quad model. Nevertheless, we focus on this model because of its relative simplicity and the useful clinical insight it provides. The hidden-state trajectories are shown in Figure S4, these indicating the occurrence of multiple transitions. These plots, together with the associated time-dependent Poisson rates (Table 3), indicate that hidden-state switching and quadratic dependence in the state-specific Poisson rates both make important contributions to the observed time dependence. The single-state quadratic model yields $p_B = 0.039$ and 0.11, respectively, for the two curvature tests, indicating model inadequacy. This is consistent with the residuals plots generated by posterior predictive simulation which indicate the presence of structure as opposed to complete randomness (these residuals plots are not included in this report).

Figure S1. Case 3 time-series data. (Top row, column **a**) The observed seizure-count time series. This plot is a duplicate of Figure 8 and is shown here to assist comparison with the simulated time-series data. (Columns **b** and **c**) Posterior predictive simulation results generated using two alternative 3-state HMMs. The first assumes time independent Poisson parameters, while the rate parameters of the second model include a temporal random effect term. The top row plots (columns **b** and **c**) each show 10 predictive time series generated by sampling the posterior predictive simulation output at evenly spaced intervals. The associated residuals plots are given in the bottom row.

Figure S2. Case 3 3-HM-quad model posterior predictive simulated data plots. (a) 10 predictive time series, as obtained by sampling the MCMC output at evenly spaced intervals, and (b) associated residuals plots.

Figure S3. Case 3 2-HM-quad posterior predictive simulated data plots. (a) 10 predictive time series, as obtained by sampling the MCMC output at evenly spaced intervals, and (b) associated residuals plots.

Figure S4. Case 3 seizure-state time series plots. These plots show which state is most credible at each occasion, as given by $P(Z_t = i | \text{data}) = \# \{Z_t^{(i)} = i\} / N_{\text{sim}}$ derived from the 2-HM-quad model. (a) $P(Z_t = i | \text{data})$ trajectory, (b) time series plot indicating which state has the largest value for $P(Z_t = i | \text{data})$ on each occasion (dashed-curve), superimposed on the observed seizure count time series (solid line). The *most credible state* curve shows the median estimate of the time-dependent seizure rate associated with that state. It serves to illustrate the magnitude of the change in mean seizure activity that is attributable to the quadratic time dependence, relative to the change caused by between-state transitions.

References

1. Kwan P, Brodie MJ: **Early identification of refractory epilepsy.** *N Engl J Med.* 2000; **342**(5): 314–319.
[PubMed Abstract](#) | [Publisher Full Text](#)
2. Haut SR: **Seizure clustering.** *Epilepsy Behav.* 2006; **8**(1): 50–55.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Pujar S, Calvert S, Cortina-Borja M, et al.: **Statistical process control (SPC)—a simple objective method for monitoring seizure frequency and evaluating effectiveness of drug interventions in refractory childhood epilepsy.** *Epilepsy Res.* 2010; **91**(2–3): 205–213.
[PubMed Abstract](#) | [Publisher Full Text](#)
4. Hopkins A, Davies P, Dobson C: **Mathematical models of patterns of seizures. Their use in the evaluation of drugs.** *Arch Neurol.* 1985; **42**(5): 463–467.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Albert PS: **A two-state Markov mixture model for a time series of epileptic seizure counts.** *Biometrics.* 1991; **47**(4): 1371–1381.
[PubMed Abstract](#) | [Publisher Full Text](#)
6. Wang P, Puterman ML: **Analysis of longitudinal data of epileptic seizure counts - a two-state hidden Markov regression approach.** *Biometric J.* 2001; **43**(8): 941–962.
[Publisher Full Text](#)
7. Wong S, Gardner AB, Krieger AM, et al.: **A stochastic framework for evaluating seizure prediction algorithms using hidden Markov models.** *J Neurophysiol.* 2007; **97**(3): 2525–2532.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Zucchini W, MacDonald IL: **Hidden Markov models for time series: an introduction using R.** Boca Raton, Florida: Chapman & Hall/CRC; 2009.
[Reference Source](#)
9. Gelman A, Carlin JB, Stern HS, et al.: **Bayesian data analysis.** 2nd ed. Boca Raton, Florida: Chapman & Hall/CRC; 2004.
[Reference Source](#)
10. Carlin BP, Louis TA: **Bayesian methods for data analysis.** 3rd ed. Boca Raton, Florida: Chapman & Hall/CRC; 2009.
[Reference Source](#)
11. Davison AC, Hinkley DV: **Bootstrap methods and their application.** Cambridge: Cambridge University Press; 1997.
[Publisher Full Text](#)
12. Lunn DJ, Thomas A, Best N, et al.: **WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility.** *Stat Comput.* 2000; **10**: 325–337.
[Publisher Full Text](#)
13. Gelman A, Goegebeur Y, Tuerlinckx F, et al.: **Diagnostic checks for discrete data regression models using posterior predictive simulations.** *Appl Stat.* 2000; **49**: 247–268.
[Publisher Full Text](#)
14. Gelman A: **Exploratory data analysis for complex models.** *J Comput Graph Stat.* 2004; **13**: 755–779.
[Publisher Full Text](#)
15. Jackson CH: **Multi-state models for panel data: The msm package for R.** *J Stat Softw.* 2011; **38**(8): 1–28.
[Reference Source](#)
16. Haut SR, Hall CB, Masur J, et al.: **Seizure occurrence: precipitants and prediction.** *Neurology.* 2007; **69**(20): 1905–1910.
[PubMed Abstract](#) | [Publisher Full Text](#)
17. Congdon P: **Bayesian statistical modelling.** 2nd ed. Chichester: John Wiley; 2006.
[Reference Source](#)
18. Lunn D, Jackson C, Best N, et al.: **The BUGS book - a practical introduction to Bayesian analysis.** Boca Raton, Florida: CRC Press; 2013.
[Publisher Full Text](#)
19. George EI, Makov UE, Smith AFM: **Conjugate likelihood distributions.** *Scand J Statist.* 1993; **20**: 147–156.
[Reference Source](#)
20. Gelfand AE, Smith AFM: **Sampling-based approaches to calculating marginal densities.** *J Am Stat Assoc.* 1990; **85**(410): 398–409.
[Publisher Full Text](#)
21. Bulla J, Bulla I: **Stylized facts of financial time series and hidden semi-Markov models.** *Comput Stat Data Anal.* 2006; **51**(4): 2192–2209.
[Publisher Full Text](#)
22. Gelfand AE, Hills SE, Racine-Poon A, et al.: **Illustration of Bayesian inference in normal data models using Gibbs sampling.** *J Am Stat Assoc.* 1990; **85**(412): 972–985.
[Publisher Full Text](#)
23. Lunn D, Spiegelhalter D, Thomas A, et al.: **The BUGS project: Evolution, critique and future directions.** *Stat Med.* 2009; **28**(25): 3049–3067.
[PubMed Abstract](#) | [Publisher Full Text](#)
24. Thomas A, Best N, Lunn D, et al.: **GeoBUGS user manual.** 2004.
[Reference Source](#)
25. Lunn D: **WinBUGS development interface (WBDev).** *ISBA Bulletin.* 2003; **10**(3): 10–11.
[Reference Source](#)
26. Cowles MK, Carlin BP: **Markov chain Monte Carlo convergence diagnostics: a comparative review.** *J Am Stat Assoc.* 1996; **91**(434): 883–904.
[Publisher Full Text](#)
27. Plummer M, Best N, Cowles K, et al.: **CODA: Convergence diagnosis and output analysis for MCMC.** *R News.* 2006; **6**(1): 7–11.
[Reference Source](#)
28. Schmidt D, Elger C, Holmes GL: **Pharmacological overtreatment in epilepsy: mechanisms and management.** *Epilepsy Res.* 2002; **52**(1): 3–14.
[PubMed Abstract](#) | [Publisher Full Text](#)
29. Gelman A, Bois F, Jiang J: **Physiological pharmacokinetic analysis using population modeling and informative prior distributions.** *J Am Stat Assoc.* 1996; **91**(436): 1400–1412.
[Publisher Full Text](#)
30. Wakefield J: **The Bayesian analysis of population pharmacokinetic models.** *J Am Stat Assoc.* 1996; **91**(433): 62–75.
[Publisher Full Text](#)
31. Chang T, Schiff SJ, Sauer T, et al.: **Stochastic versus deterministic variability in simple neuronal circuits: I. Monosynaptic spinal cord reflexes.** *Biophys J.* 1994; **67**(2): 671–683.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Schiff SJ, Sauer T, Chang T: **Discriminating deterministic versus stochastic dynamics in neuronal activity.** *Integr Physiol Behav Sci.* 1994; **29**(3): 246–261.
[PubMed Abstract](#) | [Publisher Full Text](#)
33. Wernli C: **Are deterministic descriptions and indeterministic descriptions observationally equivalent?** *Stud Hist Philos M P.* 2009; **40**(3): 232–242.
[Publisher Full Text](#)
34. King M, Pujar S, Scott R: **Dataset 1 in: A hidden Markov model analysis of subject-specific seizure time-series data as a potential aid to clinical decision making.** *F1000Research.* 2016.
[Data Source](#)
35. Gelman A, Meng X-L: **Model checking and model improvement.** In: Gilks WR, Richardson S, Spiegelhalter DJ, editors. *Markov chain Monte Carlo in practice.* London: Chapman & Hall; 1996; 189–201.
[Reference Source](#)
36. Gelman A, Meng X-L, Stern H: **Posterior predictive assessment of model fitness via realized discrepancies.** *Statist Sinica.* 1996; **6**(4): 733–808.
[Reference Source](#)
37. Gelman A: **Comment: Bayesian checking of the second levels of hierarchical models.** *Statist Sci.* 2007; **22**(3): 349–352.
[Publisher Full Text](#)
38. Bayarri MJ, Castellanos ME: **Bayesian checking of the second levels of hierarchical models.** *Statist Sci.* 2007; **22**(3): 322–343.
[Publisher Full Text](#)
39. Richardson S, Green PJ: **On Bayesian analysis of mixtures with an unknown number of components.** *J R Statist Soc B.* 1997; **59**(4): 731–792.
[Publisher Full Text](#)
40. Stephens M: **Dealing with label switching in mixture models.** *J R Statist Soc B.* 2000; **62**(4): 795–809.
[Publisher Full Text](#)
41. Celeux G, Hurn M, Robert CP: **Computational and inferential difficulties with mixture posterior distributions.** *J Am Stat Assoc.* 2000; **95**(451): 957–970.
[Publisher Full Text](#)
42. Jasra A, Holmes CC, Stephens DA: **Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling.** *Stat Sci.* 2005; **20**(1): 50–67.
[Publisher Full Text](#)
43. Geweke J: **Interpretation and inference in mixture models: simple MCMC works.** *Comput Stat Data Anal.* 2007; **51**(7): 3529–3550.
[Publisher Full Text](#)
44. King MD, Gadian DG, Clark CA: **A random effects modelling approach to the crossing-fibre problem in tractography.** *Neuroimage.* 2009; **44**(3): 753–768.
[PubMed Abstract](#) | [Publisher Full Text](#)
45. Shaddick G, Wakefield J: **Modelling daily multivariate pollutant data at multiple sites.** *Appl Stat.* 2002; **51**(3): 351–372.
[Publisher Full Text](#)
46. Wakefield JC, Best NG, Waller L: **Bayesian approaches to disease mapping.** In: Elliott P, Wakefield JC, Best NG, et al., editors. *Spatial epidemiology: methods and applications.* Oxford: Oxford University Press; 2000; 104–127.
[Publisher Full Text](#)
47. Spiegelhalter D, Thomas A, Best N, et al.: **WinBUGS user manual.** MRC Biostatistics Unit, Institute of Public Health, Cambridge; 2003.
[Reference Source](#)
48. Congdon P: **Bayesian statistical modelling.** Chichester: John Wiley; 2001.
[Reference Source](#)

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research