A Bayesian framework for describing and predicting the stochastic demand of home care patients

Raffaele Argiento · Alessandra Guglielmi · Ettore Lanzarone · Inad Nawajah

R. Argiento · E. Lanzarone
CNR–IMATI, Via Bassini 15, 20133 Milan, Italy
e-mail: raffaele@mi.imati.cnr.it

E. Lanzarone
e-mail: ettore.lanzarone@cnr.it

A. Guglielmi · I. Nawajah (✉)
Dipartimento di Matematica, Politecnico di Milano, P.zza Leonardo da Vinci 32, 20133 Milan, Italy
e-mail: inad.nawajah@mail.polimi.it

A. Guglielmi
e-mail: alessandra.guglielmi@polimi.it

1 Introduction

This paper concerns the prediction of demand from patients assisted by home care (HC) service. HC refers to any type of care given to a patient at his/her own home rather than in hospital or other health care facilities. Nurses, physicians, and other professional figures bring all the necessary facilities at patients' homes and therein provide care to patients. Hence, the main benefit of HC is the reduction of hospitalization rate. This significantly increases the quality of life for the assisted patients, as they remain at home, and at the same time it yields relevant cost savings for the entire health care system, as hospitalization costs are avoided (Jones et al. 1999; Hollander and Miller 2002; Comondore et al. 2009). This service is a relevant and growing sector in the health care domain of western countries, because of the population ageing, the increase in chronic pathologies, the introduction of innovative technologies, and the continuous pressure of governments to contain health care costs.

Resource planning is crucial for operating in HC organizations. Human and material resources have to be properly managed in order to avoid process inefficiencies, treatment delays and low quality of service (Matta et al. 2014). However, management of HC organizations is quite complex; complexity mainly holds because of the large number of assisted patients, the synchronization of resources at patients' home, and the service delivery in a usually large territory. Moreover, random events affect the delivery of service and mine the feasibility of plans (Lanzarone and Matta 2012; Lanzarone et al. 2012; Matta et al. 2014). Patients' conditions, resource unavailability and the duration of operators' transfers in the territory are the main randomness sources. All of them cause revisions of patients' care pathways, of scheduled visits and, consequently, of HC resource plans. Indeed, the most relevant randomness source is related to changes in patients' conditions, which consequently determine a demand for visits different from the planned one in terms of number, frequency, and duration (Lanzarone et al. 2010). Other frequent unexpected events are changes in patients' visiting hours, operator and material resource unavailabilities, scheduling problems, and random events delaying the transportation of human and material resources.

In the majority of HC providers, patients are always followed by nurses and, in some cases, by other operators. Hence, nurses manage the care pathway of patients, provide the largest number of visits, and deal with emergencies and other variations of service demand (Matta et al. 2014). Several HC providers pursue the continuity of care, usually for nurses (Borsani et al. 2006; Matta et al. 2014). Continuity of care means that a patient is assigned to only one operator for each category, named the *reference* or the *principal* operator, who follows the entire patient care pathway during the sojourn in charge of the HC facility and preferably provides all of the visits pertinent to his/her category. This is considered to be an important indicator of HC service quality, because information loss among operators is avoided and patients receive care from the same operator rather than continuously developing new relationships with new persons (Woodward et al. 2004; Anderson et al. 2012). However, continuity of care makes resource planning even more difficult, because the nurse-to-patient assignments have to be maintained over future variations of

patients' conditions and operators' availabilities. For this reason, some other HC providers do not take the continuity of care into account.

Continuity of care needs affordable estimates of the amount of visits required by each patient in short-term and middle-term planning horizons to avoid operational problems. More in general, the ability of estimating patient health progression and the demand for visits during the care pathway is a fundamental requirement for developing robust decision support tools for HC resource planning.

This paper focuses on the estimation of patients' demands, which, as just mentioned, is fundamental for planning human and material resources. However, estimates are computed independently from the specific planning issues they will be used for, based on historical observations of patients' visits and health conditions. Indeed, we propose and validate a Bayesian framework for representing and predicting the demand evolution of HC patients over time. We define a generalized linear mixed model with autoregression over time. The inference is based on the posterior density of parameters, obtained through a Markov chain Monte Carlo simulation scheme.

In general, the Bayesian approach to population modeling is particularly appealing from a scientific perspective, as it allows prior information (*prior to see the data*) to be incorporated. It suggests that any scientific inference is based on two parts: one depends on the scientist's subjective opinion and understanding of the phenomenon before experimentation, and the other depends on the observations obtained from the experiment itself. In cases in which complexity derives from multiple sources of evidences, frequentist statistical methods have difficulties; for instance, when the statistical model is described in terms of many parameters, it can be quite difficult computing estimates and comparing estimators. On the contrary, under the Bayesian approach, it is straightforward that both estimation and comparison are driven by one single probability law, i.e., the posterior distribution of the parameters. As underlined in Spiegelhalter et al. (2004), Ch. 1, the Bayesian perspective can be more flexible than traditional statistical methods, more efficient using all available evidence, and more useful, since it provides predictions of observable quantities through probabilities. Of course, external evidence (prior belief) must be introduced with caution and used in a clear, explicit and transparent manner. Very often the prior information is translated into probabilistic assumptions that are subjective. However, the subjective interpretation fits very well problems in health care, where there are so many human and organizational sources of randomness (see Aven and Eidesen 2007). Another advantage of the Bayesian approach is its focus on observable quantities; in this context, prediction is quite straightforward: it is given by the conditional distribution of the quantity of interest, on background information (including assumptions and suppositions) and on data. Of course, the accuracy of the model needs to be addressed, for instance comparing predictions to observed data.

The paper is organized as follows: the literature dealing with the stochastic modeling of patients' conditions in health care facilities and, in particular, in HC is presented in Sect. 2. The general features of an HC provider and the typical structure of HC patients' dataset are described in Sect. 3, whereas the Bayesian model is detailed in Sect. 4. Then, the application of the proposed model to a

relevant real case is presented in Sect. 5: the dataset of such real case is described in Sect. 5.1; posterior inference of the model parameters is reported in Sect. 5.2; goodness-of-fit, predictions for patients already in charge and for newly admitted patients are shown in Sects. 5.3–5.5, respectively. Finally, conclusions and some possible suggestions for future work are then given in Sect. 6.

2 Related literature

Several studies that deal with the development of stochastic models for patients' conditions evolution in health care facilities can be found in the literature. Different approaches are adopted, both frequentist and Bayesian, spreading from Markov models to Bayesian hierarchical models. In the following, we briefly revise this literature, pointing out the main features of each proposed model.

Concerning frequentist approaches, Markov models have been used to study the hospitalization of geriatric patients (Taylor et al. 1996; McClean et al. 1998) or the natural history of hepatitis-C, in order to determine compensatory funds for patients who acquired the pathology through blood transfusions (Krahn et al. 2004). Faddy and McClean (2005) adopt Markov chain models based on phase-type distributions to model changes in geriatric patients' health conditions and to evaluate the impact of covariates. Alagoz et al. (2005) develop an empiric natural-history model to predict changes in laboratory values and clinical characteristics of patients with end-stage liver disease, showing that cubic splines can generate quantitative natural histories and are also useful for developing clinically robust microsimulation models of other diseases. More recently, Gómez-Batiste et al. (2012) describe conceptual innovations in palliative care epidemiology and the methods to identify patients in need of palliative care, and Blanco-Encomienda (2013) propose a multi-state Markov model to estimate the cost of care provisioning to elderly people, in order to help governments in efficiently and effectively allocating resources. Considering the specific case of HC patients, Lanzarone et al. (2010) propose a frequentist patient model, which provides estimates on the major variables of interest for an HC provider: how many patients are followed up in the course of time and, for each of them, the amount of required visits. The model integrates a care pathway model on one hand, and a cost model on the other. After identifying state variables and patient classes, the patient's evolution is described through the sequence of values assumed by the state variables, which is modeled as a Markov chain. Then, an empirical probability distribution of the cost is associated with each possible state, where the cost is intended as the number of visits to the patient in a fixed number of days. This group of papers suggests that frequentist approaches are widely adopted in the general health care context, but not in the HC environment, apart from Lanzarone et al. (2010), to the best of our knowledge. However, these models are quite different from application to application, as they need to adapt to the specific context and ad-hoc configurations. For instance, model in Lanzarone et al. (2010) fits the data well, but it includes empirical distributions and ad-hoc configuration.

Of course, the Bayesian approach has been adopted for modeling the stochastic behavior of patients of several health care facilities, both for medical and management purposes, as several studies can be found in the medical statistical literature. For instance, Pauler and Finkelstein (2002) adopt a Cox proportional hazard model, under the Bayesian perspective, to predict the disease progression in prostate cancer recurrence. Bergamaschi et al. (2000) analyze the risk of secondary progression in multiple sclerosis patients by using two different approaches: a Cox proportional hazards model and a Bayesian latent-variable model. Berzuini and Allemani (2004) assess the effect of a highly active antiretroviral therapy on the course of the acquired immune deficiency syndrome by means of a Bayesian model, in which the sequence of longitudinal cell count observations and the associated time to the syndrome are jointly modelled at an individual subject's level as depending on the treatment. Verotta (2005) compares alternative estimation methods for the analysis of clinical human immunodeficiency virus data, underlining that Bayesian models incorporate prior knowledge into the models themselves, thus avoiding some of the model simplifications introduced when the data are analyzed using other methods. Guglielmi et al. (2010) propose a Bayesian hierarchical generalized linear model to analyze the survival probabilities after acute myocardial infarction in Milan, Italy, by considering both clinical registries and administrative databases. More recently, Carreras et al. (2012) consider Markov models to describe the natural history of specific diseases, with particular attention to the cervical cancer, and propose two Bayesian models for carrying out a probabilistic sensitivity analysis on the transition probabilities. From a management point of view, some papers deal with Bayesian techniques in order to predict the demand for care. For example, Congdon (2001) implements a Bayesian generalized linear model, with Poisson outcome, which is used to predict patients' traffic from home to hospital, in order to facilitate the reconfigurations of emergency hospital services. Marshall et al. (2002) model patients' duration of stay to facilitate resource management of geriatric hospitals by using phase-type distributions conditioned on a Bayesian belief network. Moreover, Marshall et al. (2005) present an overview of such modeling technique in comparison with other methods, with particular attention to their impact and suitability in managing hospital services. All these papers point out the relevance and the flexibility of the Bayesian modeling approach, which can fit several clinical and health care problems. However, in spite of these benefits, the Bayesian approach has not been considered so far when modelling HC patients data. Owing to this lack, the aim of this paper is to propose a Bayesian model to predict the demand for visits in future periods. Indeed, we want to investigate the potentialities of such an approach in the HC context, where Bayesian statistics has not been applied yet.

On the other hand, papers dealing with robust approaches for managing resources in health care, which require patients' demands estimation, underline the importance of such estimation models. For instance, the patients' stochastic model of Lanzarone et al. (2010) was adopted for the implementation of robust nurse-to-patient assignment approaches for HC, e.g., a stochastic programming model (Lanzarone et al. 2012), an analytical policy (Lanzarone and Matta 2012) and a

cardinality–constrained model (Carello and Lanzarone 2014). The Bayesian model we propose here is intended also to be used for this kind of applications.

3 Care pathway of HC patients

The care pathway of a HC patient usually includes *admission*, *care supply* and *discharge* (see Lanzarone et al. 2010; Matta et al. 2014). The admission phase consists of the preliminary and the multidimensional assessments. During the first visit, an operator, generally a nurse, collects personal data and other information about the patient's clinical, functional and social conditions. Then, a multidisciplinary team evaluates the patient's conditions and tests his/her functional abilities. The assessment defines the patient's needs on which the service will be designed and provided. The service supply starts after the multidisciplinary team develops the Therapeutic project (ThP), which includes detailed information about type and frequency of the required visits, and all other operational activities executed by HC operators at patient's home (Asquer et al. 2007; Matta et al. 2014). During the care supply phase, the patient receives the service from the HC operators as prescribed by his ThP. The ThP is periodically assessed in order to check its adequacy for the patient's needs, which may change depending on clinical conditions, as well as the social environment. The revision period is usually one month: at each periodic revision, the ThP can be reconfirmed or modified to address the new patient's needs. However, a revision can be introduced before the end of the month in case of sudden variation of patient's condition. Finally, the patient is discharged when he/she recovers, needs a different kind of service (such as hospitalization), or dies.

Providers also assign a category to each patient, usually named Care Profile (CP), based on the type and the number of visits required by the ThP, which are the most significant factors affecting costs. Indeed, once a ThP is defined, a CP is assigned to the patient at the same time. Then, a patient's CP may change along with the time according to his/her ThP modifications. From a modeling point of view, the CP is a categorical variable that evolves over time and represents the status of patient's conditions and requirements. The classification of patients via CP is common to several HC providers of different countries (Matta et al. 2014); based on the specific country or region, the number and typologies of the CPs may vary, but such type of classification is always present. Moreover, several registries, collecting social and clinical covariates of patients, are usually available.

From a management point of view, the time is divided into discrete slots (e.g., the day or the week). At each time slot, each patient in charge receives a certain number of visits from the different types of operators.

3.1 Formalization of the care pathway and assumptions

Patient conditions can be described by a set of representative state variables (e.g., pathology, age and social conditions), which must be observable and measurable at low cost throughout the patient health care history. Consequently, the patient's care pathway can be described as a sequence of multidimensional states. As in

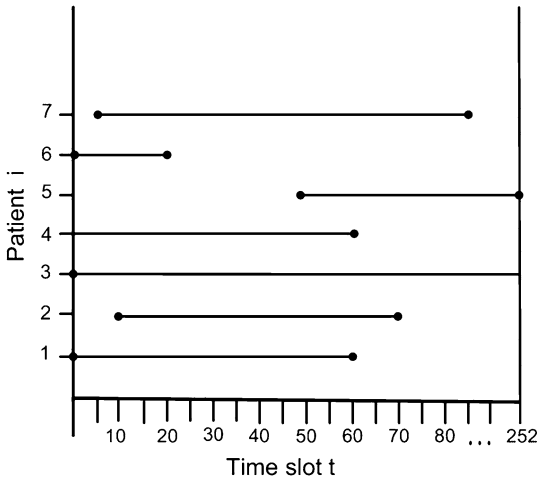

Fig. 1 Examples of entry–exit service times for some patients. Patient 2 enters the service at $t = 10$ and exit at $t = 70$, patient 6 enters at the initial slot $t = 0$, and patient 5 exits at the last slot $t = 25$. The absence of a point at the beginning or the ending of the path means that the patient is also in service out of the considered slots. Patient 4 is already in service before $t = 0$ and patient 3 remains in service after $t = 252$

Lanzarone et al. (2010), we assume that CP is the only state variable, since it actually summarizes all of the other significant variables. For example, the pathology, which appears to be a significant factor, is not directly taken into account because included in the choice of ThP and, then, in the value of CP assigned to the patient.

In our analysis, we consider a sequence of time slots t (with $t = 1, \ldots, T$) in which several patients i (with $i = 1, \ldots, n$) are assisted. Moreover, the week is considered as the reference period for the length of all time slots t. In other words, each time slot is assumed equal to 1 week. This reflects the "true" assignment phase of the planning process, very often carried out over a weekly basis.

We denote by $T_L(i)$ and $T_U(i)$ the time slots when patient i enters and exits the service, respectively, as schematized in Fig. 1. We assume that each patient enters and exit only once during his/her care pathway, i.e., we do not consider cases in which a patient is temporarily discharged and reenters the service after a pause. This sometimes happens in the HC services, but such cases are usually limited.

The provided service is described in terms of the number of visits provided by nurses to patient i at time slot t, which is denoted with $N_{i,t}$. We focus on nurses because they are in charge of following HC patients' care pathways and they provide the largest number of visits to patients. Hence, they meet short-term demand variations, manage emergencies, and deal with highly uncertain workloads. Finally, continuity of care, which strongly requires the estimation of future demands to be effectively implemented, is usually pursued for nurses. However, we remark that our model does not take into account continuity of care (see Sect. 4.1); in fact, the nurse-to-patient assignments are made after the predictions are computed. Moreover, we do not take into account the duration of visits in the model. In fact, for planning purposes, several HC providers assume that all visits have the same duration, including the travel times to reach and leave the patient's home (Matta et al. 2014). Hence, the number of visits is a valid indicator of the workload amount required by patients.

Finally, the knowledge of the specific nurse who provides each visit is not included in the analysis. This is an issue to take into account when the estimation model will be applied in order to derive the planning. However, with this choice, the care pathway of a patient is independent from the specific nurse who provides each visit. This is an admissible assumption, as all nurses have been equally and adequately skilled to provide the assigned visits.

Each patient i is thus characterized by the sequence of $N_{i,t}$ and $CP_{i,t}$ for each $t \in \{T_L(i), \ldots, T_U(i)\}$. In case of time slots longer than one day, as in our case, the CP may change during t according to the ThP modifications. In this case, we assume $CP_{i,t}$ as the CP assumed for the largest part of the time slot. This is an admissible approximation because time slots are taken equal to only 1 week (i.e., $CP_{i,t}$ is the CP assumed for at least four days over seven).

Summing up, data observed for each patient i at each slot $t \in \{T_L(i), \ldots, T_U(i)\}$ are:

1. Number of visits $N_{i,t}$ received by patient i from nurses during time slot t—count data;
2. Care Profile $CP_{i,t}$ assumed by patient t at time slot t—categorical variable assuming N_{cp} integer values s (with $s = 1, \ldots, N_{cp}$).

Moreover, in our framework, we take into account two social and clinical registry covariates, whereas others are neglected:

1. age_i: age of patient i at $t = T_L(i)$, expressed in terms of a normalized age as follows:

$$age_i = \frac{age_{patient_i} - age_{mean}}{age_{max} - age_{min}}$$

   where $age_{patient_i}$ is the age in years of patient i at $t = T_L(i)$, while age_{mean}, age_{min} and age_{max} are the mean, the minimum and the maximum ages of patients in the dataset, respectively.
2. sex_i: gender of patient i, expressed in terms of a binary variable equal to 0 if male, or 1 if female.

These covariates do not depend on t. We use the notation $x_i = (age_i, sex_i)$, and $\boldsymbol{x} = (x_1, \ldots, x_n)$.

4 Bayesian model for HC patient's demand

In this section, we present the Bayesian model describing the patients' demand evolution over time in full details (Sect. 4.1). Moreover, based on this model, we also provide the posterior predictive probabilities of the demand for visits in future time slots (Sect. 4.2).

4.1 The model

A Bayesian model is traditionally described by means of the likelihood function and the prior distribution. By likelihood function we mean the conditional distribution of

the response variable, which is $N_{i,t}$ in our case, given all the parameters and covariates. Two main aspects have to be taken into account, keeping in mind the specific problem and dataset we consider: modeling the distribution of $N_{i,t}$ as a function of the covariates on the one hand, and of the number of visits at the previous week $N_{i,t-1}$ on the other hand. The former dependence is modeled by using a generalized linear model; in particular, one of the most straightforward choices for count data is the Poisson regression. The latter dependence is taken into account here by an autoregressive model of lag 1 (i.e., a Markov model) where, in the logarithmic scale, the mean value of $N_{i,t}$ depends on the mean value of $N_{i,t-1}$. Accordingly, we model the number of visits $N_{i,t}$ for patient i at time slot t as a Poisson random variable with expected value $\lambda_{i,t}$. Then, the evolution of $\lambda_{i,t}$ over the discrete time slots is determined according to a Markov Chain (with continuous values), as mentioned above. Note that $CP_{i,t}$ is not a random variable in this model, but it is considered as a fixed covariate.

Let $\boldsymbol{N}_i = (N_{i,T_L(i)}, N_{i,T_L(i)+1}, \ldots, N_{i,T_U(i)-1}, N_{i,T_U(i)})$ the demand vector of patient i, and assume that $\boldsymbol{N}_1, \ldots, \boldsymbol{N}_n$ are conditionally independent. We propose the following generalized linear model for any $i = 1, \ldots, n$:

$$\begin{aligned} N_{i,t}|\lambda_{i,t} &\sim \text{Pois}\ (\lambda_{i,t}) \qquad T_L(i) \le t \le T_U(i) \\ \log(\lambda_{i,T_L(i)}) &\sim \mathcal{N}(\gamma_1 age_i + \gamma_2 sex_i + \gamma_3[CP_{i,T_L(i)}], \sigma_0^2) \\ \log(\lambda_{i,t})|\lambda_{i,t-1} &\sim N(\alpha[CP_{i,t}]\log(\lambda_{i,t-1}) + \beta[CP_{i,t}], \sigma^2) \quad T_L(i) < t \le T_U(i), \end{aligned} \tag{1}$$

where Pois $(\lambda_{i,t})$ denotes the Poisson distribution with mean value $\lambda_{i,t}$, and $\mathcal{N}(\mu, \sigma^2)$ is the Gaussian distribution with mean μ and variance σ^2. The first line in (1) is the generalized linear model, whereas the last one sets up the autoregression of lag 1. In this formulation, $\lambda_{i,t}$ is a latent variable that represents the hidden health status of patient i at time slot t, and is responsible for the demand for visits (i.e., the bigger is parameter $\lambda_{i,t}$, the worse are patient's conditions, and the higher is the expected number of visits). Parameters $\alpha_s := \alpha[CP_{i,t} = s]$, $\beta_s := \beta[CP_{i,t} = s]$ and $\gamma_{3,s} := \gamma_3[CP_{i,t} = s]$ describe the random–effects as a function of the value of $CP_{i,t}$. Moreover, α_s represents the first order auto–regressive coefficient as a function of $CP_{i,t}$.

Summing up, (1) is a generalized linear mixed model with two fixed-effects parameters (γ_1 and γ_2) and three random-effects parameters ($\gamma_{3,s}$, α_s and β_s). The random-effects parameters take into account the similarity of health state conditions (i.e., the CP) on patients' demand for visits $N_{i,t}$: two different patients with the same CP display the same contribution in terms of $\gamma_{3,s}$, α_s and β_s. On the other hand, the fixed-effects parameters do not depend on any grouping factor. For a comprehensive review of Bayesian modeling, see Gelman et al. (2013). Generalized linear mixed models are an extension of generalized linear models that include both fixed and random-effects (see Fahrmeir and Tutz 1994; Zeger and Karim 1991). The overall model proposed in this paper is a generalization of the one proposed in Giardina et al. (2011) for longitudinal binary data.

All parameters are included in vector $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_{N_{cp}}, \beta_1, \ldots, \beta_{N_{cp}}, \gamma_1, \gamma_2, \gamma_{3,1}, \ldots, \gamma_{3,N_{cp}}, \sigma, \sigma_\alpha, \sigma_\beta, \sigma_{\gamma_3})$. According to the Bayesian approach, the knowledge

about model parameters is summarized by a probability distribution, called the *prior density*, i.e. the second ingredients of a Bayesian model. The prior belief is then updated through the data, computing the conditional distribution of the parameters given the data, i.e. the *posterior distribution*, by means of Bayes' theorem. In this paper, we adopt a noninformative approach to elicit the prior (see, for instance, Jackman 2009, Sect. 2.1). In particular, a marginal prior centered at 0 with large variance is chosen for the fixed effects parameters γ_1 and γ_2, i.e., the Gaussian distribution with 0 mean and 1000 variance. As for as the standard deviation parameters are concerned, a standard noninformative choice (Gelman 2006) is to assume a uniform prior over a *large* interval $(0, \sigma_{max})$ and then to perform a robustness analysis, decreasing σ_{max} as long as the same posterior inference is obtained. Summing up, all parameters are assumed conditionally independent and:

$$\begin{aligned}
\gamma_1, \gamma_2 &\overset{\text{iid}}{\sim} \mathcal{N}(0, 1000) \\
\alpha_s | \sigma_\alpha^2 &\overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\alpha^2), \quad s = 1, \ldots, N_{cp} \quad \text{with } \sigma_\alpha \sim \mathrm{U}(0, 5) \\
\beta_s | \sigma_\beta^2 &\overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\beta^2), \quad s = 1, \ldots, N_{cp} \quad \text{with } \sigma_\beta \sim \mathrm{U}(0, 2) \\
\gamma_{3,s} | \sigma_{\gamma_3}^2 &\overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{\gamma_3}^2), \quad s = 1, \ldots, N_{cp} \quad \text{with } \sigma_{\gamma_3} \sim \mathrm{U}(0, 15) \\
\sigma &\sim U(0, 5).
\end{aligned}$$

Differently from the others, σ_0^2 is assumed as a fixed constant value (equal to 3) to avoid identifiability issues.

As mentioned before, the statistical analysis is performed by means of the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{N}, \boldsymbol{x})$ of parameter vector $\boldsymbol{\theta}$. Denoting by $\boldsymbol{\lambda} = (\lambda_{it})$ all the latent variables representing the health status of all patients, it follows that:

$$\mathcal{L}(\boldsymbol{N}, \boldsymbol{\lambda}|\boldsymbol{x}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{t=T_L(i)}^{T_U(i)} \mathcal{L}(N_{i,t}|\lambda_{i,t}, x_i) \mathcal{L}(\lambda_{i,t}|\lambda_{i,t-1}, x_i) \mathcal{L}(\lambda_{i,T_L(i)}).$$

Hence, based on Bayes' Theorem, the posterior density is:

$$\begin{aligned}
\pi(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{N}, \boldsymbol{x}) &\propto \pi(\boldsymbol{\theta}) \mathcal{L}(\boldsymbol{N}, \boldsymbol{\lambda}|\boldsymbol{x}) \\
&\propto \pi(\boldsymbol{\theta}) \prod_{i=1}^{n} \prod_{t=T_L(i)}^{T_U(i)} \mathcal{L}(N_{i,t}|\lambda_{i,t}, x_i) \mathcal{L}\left(\lambda_{i,t}|\lambda_{i,t-1}, x_i\right) \mathcal{L}\left(\lambda_{i,T_L(i)}\right).
\end{aligned} \tag{2}$$

Observe that $\pi(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{N}, \boldsymbol{x})$ is not available in a closed analytical form. Therefore, a Markov Chain Monte Carlo (MCMC) algorithm is used to get samples from it. MCMC algorithms are simulation schemes aiming at approximating a target distribution, in which a Markov chain with invariant distribution equal to the target one is built. Under suitable conditions of the chain, the ergodicity theorem yields the convergence of the algorithm (see Robert and Casella 2004).

In this paper, a Gibbs sampler scheme is adopted to perform the MCMC sampling from the posterior distribution. Open source software is used: the Gibbs scheme is implemented by JAGS 3.4.0 (Plummer 2003), running the code from an R interface (www.r-project.org) by means of the *rjags* package.

4.2 Prediction of patients' demands

A distinctive feature of the Bayesian approach is the feasibility of the statistical analysis from a predictive point of view. In this study, our goal is to predict the demand for visits at future time slots, given *covariates* and data.

As discussed in Sect. 1, this is highly important for HC decision makers, who are interested in improving the service efficiency. In fact, the predictions of patients' demands allow HC managers to make robust decisions (e.g., the number operators needed in next weeks and the optimal operator-to-patient assignments under continuity of care) that are supported by the prediction of the demand for visits from patients in charge. For instance, since HC managers have the whole predictive distributions of each patient's demand at their disposal, it is easy to compute through them the predictive distributions of each nurse's workload. Then, in addition to the expected value, it is easy to get the (predictive) probability that, in a future week, a nurse's workload exceed any fixed threshold (e.g., the weekly working time without overtime).

Assume we are at time slot t; then, the posterior predictive probability of $N_{i,t+1}$ at the next time slot $t+1$ is given by:

$$\begin{aligned}&\mathcal{L}\left(N_{i,t+1}=k|\boldsymbol{x},\mathbf{N_1},\ldots,\mathbf{N_n}\right)\\&=\int\mathcal{L}\left(N_{i,t+1}=k|\lambda_{i,t+1}\right)\mathcal{L}\left(\mathrm{d}\lambda_{\mathrm{i,t+1}}|\lambda_{\mathrm{i,t}}\right)\pi(\mathrm{d}\lambda_{\mathrm{it}}|\mathbf{N_1},\ldots,\mathbf{N_n}),k=0,1,\ldots\end{aligned}\tag{3}$$

where $N_{i,t+1}$ and $\lambda_{i,t+1}$ are the number of nurse visits and the patient's health status at time slot $t+1$, respectively. Once again, evaluation of (3) is given through the same MCMC strategy, adding a new step when $\lambda_{i,t+1}$ must be drawn.

Analogously, we are able to compute the posterior predictive distribution for a newly admitted patient; in this case, no information is available from previous time slots. Let us denote by i^* the new patient with covariate vector x^* and CP trajectory CP^*; then the predictive distribution of i^* is computed integrating the conditional joint distribution of $\boldsymbol{N}_{i^*}=(N_{i^*,T_{L(i^*)}},\ldots,N_{i^*,T_{U(i^*)}})$ and $\boldsymbol{\lambda}_{i^*}=(\lambda_{i^*,T_{L(i^*)}},\ldots,\lambda_{i^*,T_{U(i^*)}})$ as in (1) with x^* and CP^*, with respect to the posterior distribution of $\boldsymbol{\theta}$. From a computational point of view, in our MCMC scheme, we simulate trajectories of i^* according to (1), where parameter vector $\boldsymbol{\theta}$ assumes values from the posterior simulated sample.

5 Application to a real HC provider

We apply the proposed methodology to a relevant real case. We consider the data of one of the largest Italian public HC providers. This operates in the north of Italy, covering a region of about 800 km^2, with about 1,000 patients assisted at the same time. This provider has been chosen because its data have been already analyzed in other works and a comparison among approaches is thus facilitated (Lanzarone et al. 2010). Moreover, human resource organization and patient classification adopted by this provider respect the general features common to several HC providers, which have

been detailed in Sect. 3. In particular, the three phases of the care pathway (i.e., admission, care supply and discharge) are present, a ThP is assigned to each patient, and the revision of the ThP (periodic or secondary to sudden patient's variations) is implemented. Besides, the assignment of a CP is also implemented, based on the normative of the regional health care system. For these reasons, the HC provider we consider is representative of a general class of providers, and general conclusions extensible to other HC facilities can be derived from this real case analysis.

Patients of this HC provider are grouped into two categories (palliative and non-palliative) and each category includes a certain number of CPs, as reported in Table 1, with a total of 14 CPs (Lanzarone et al. 2010). CPs related to palliative care refer to a homogeneous class of terminal patients whose pathology is in a terminal state. On the contrary, for non-palliative care, each CP includes a wide range of patients in terms of age, pathology and social context, even if patients characterized by similar levels of demand are generally classified by the same CP.

5.1 Description of the dataset

We started the analysis considering the same dataset as in Lanzarone et al. (2010). This includes 7,677 patients assisted over a period of 252 weeks, from January 2004 to March 2008. Due to the high similarity among some CPs in terms of health behavior, pathology and number of weekly visits, we regrouped them as shown in Table 1, thus obtaining nine CPs (i.e., $N_{cp} = 9$ in the analysis). According to the assumptions, only patients without any interruption of service (e.g., hospitalization periods with an interruption of the HC assistance) are analyzed. Moreover, the provider is divided into three divisions and the analysis refers to the largest one.

Summing up, we have 3,095 patients with an overall number of 87,555 observations between $T_L(i)$ and $T_U(i)$ and an average number of weekly visits (i.e.,

Table 1 Classification of CPs into palliative and non-palliative care, with a further division for non-palliative ones into extemporary and integrated home care: CPs according to the HC provider (Lanzarone et al. 2010) and regrouped for our analysis

Types of care	CPs given by the provider	CPs regrouped for this work
Extemporary care—non-palliative care with very low frequency of visits	1	1
	15	9
Integrated home care—non-palliative care with medium-high intensity	10	8
	9	7
	2, 12	2
	3, 13	3
	4, 14	4
	5	5
Palliative care offered to terminal patients, generally affected by oncological diseases	6, 7, 8	6

The list of CPs within each category is reported in an ascending order of complexity and expected demand for visits

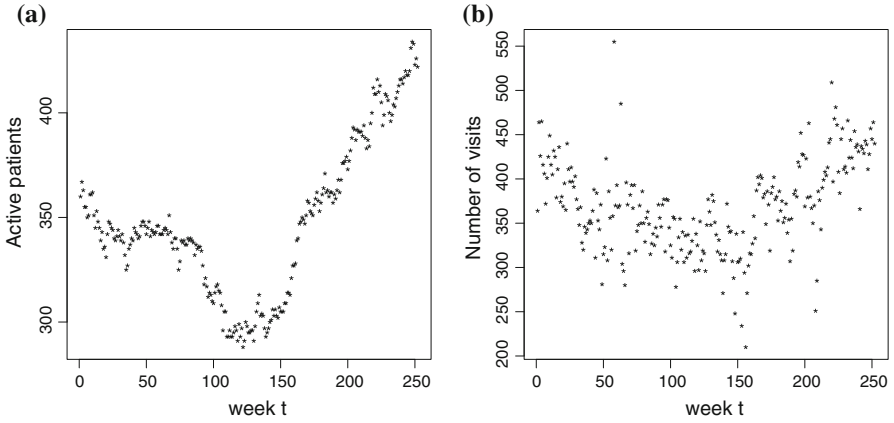

Fig. 2 Number of active patients (**a**) and total number of visits provided by nurses (**b**) for all weeks $t = 1, \ldots, 252$

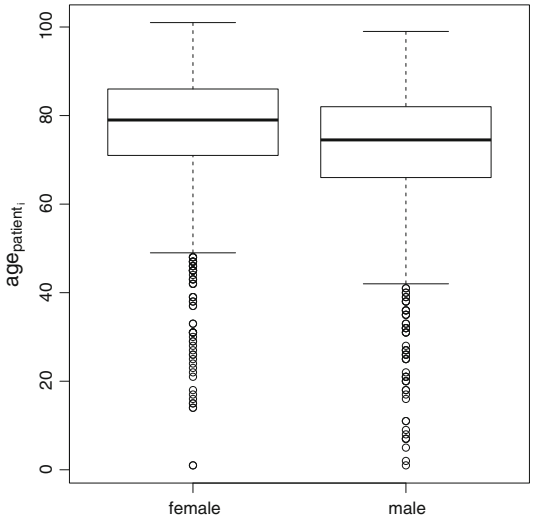

Fig. 3 Box plot of ages for all patients included in the dataset by gender. *Points* denote outliers, i.e. datapoints less than 1.5 times the interquartile range below the first quartile

the ratio between 92,945, the overall number of observed weekly visits, and 87,555 observations) equal to 1.06. The number of active patients in the dataset and the weekly total number of observed visits along the 252 weeks are plotted in Fig. 2. The dataset includes 1,306 males (42.2 %) and 1,789 females (57.8 %). The age ranges from a minimum of 1 year to a maximum of 101 years, while the empirical mean and standard deviation are 78.8 and 14.0 years, respectively. The box plot of age by gender is given in Fig. 3. Finally, the total number of observations between $T_L(i)$ and $T_U(i)$ and the average number of visits, grouped by CP, age and gender, are reported in Tables 2 and 3, respectively.

Table 2 Number of observations and average number of visits grouped by $CP_{i,t}$

	$CP_{i,t} = 1$	$CP_{i,t} = 2$	$CP_{i,t} = 3$	$CP_{i,t} = 4$	$CP_{i,t} = 5$
Total no. of obs	39,768	4,055	4,402	4,319	6,550
Average no. of visits	0.40	1.54	1.71	1.68	2.99

	$CP_{i,t} = 6$	$CP_{i,t} = 7$	$CP_{i,t} = 8$	$CP_{i,t} = 9$
Total no. of obs	7,430	16,704	1,910	2,417
Average no. of visits	2.22	0.99	0.38	1.13

Table 3 Number of observations and average number of visits grouped by age and gender

Age	Male		Female	
	No. of obs	Average no. of visits	No. of obs	Average no. of visits
≤ 50	1,732	1.29	825	1.38
$(50, 60]$	1,737	1.09	2,439	0.89
$(60, 70]$	4,782	1.22	5,245	1.55
$(70, 80]$	12,417	1.10	18,350	1.02
$(80, 90]$	11,540	0.87	20,739	0.94
≥ 90	1,794	0.66	5,955	1.43

5.2 Posterior distributions of model parameters

The MCMC has been run with 255,000 iterations, a burn-in of 5,000, and a thinning of 50 iterations. Thus, the final sample size consists of 5,000 samples. Standard convergence diagnostics, as Geweke and the two Heidelberger-Welch diagnostics (Plummer et al. 2006) have been checked, together with traceplots, autocorrelations and MC error/posterior standard deviation ratios for all the parameters, indicating that convergence has been achieved. Moreover, several independent chains gave the same posterior means for all parameters. Fig. 4 displays traceplots of all parameters α_s and $\gamma_{3,s}$ $(s = 1, \ldots, 9)$; analysis of such plots confirms good mixing and convergence of the chain, since samples are not auto-correlated and there is no clear trend.

Detailed results for each model parameter in $\boldsymbol{\theta}$ are reported in terms of their posterior 95 % credibility interval (CI). Figure 5 displays CIs and posterior medians of parameters α_s, β_s and $\gamma_{3,s}(s = 1, \ldots, 9,)$ that depend on the CP, whereas posterior quantiles of the other parameters not depending on the CP (i.e., γ_1, γ_2 and σ^2) are reported in Table 4.

Results show that the random-effects parameters are significantly different with respect to the CP. This means that the numbers of visits $N_{i,t}$ strongly depend on patients' classification. This agrees with the clinical evidence because, for the HC provider we are considering, the division among CPs is done on purpose to take into

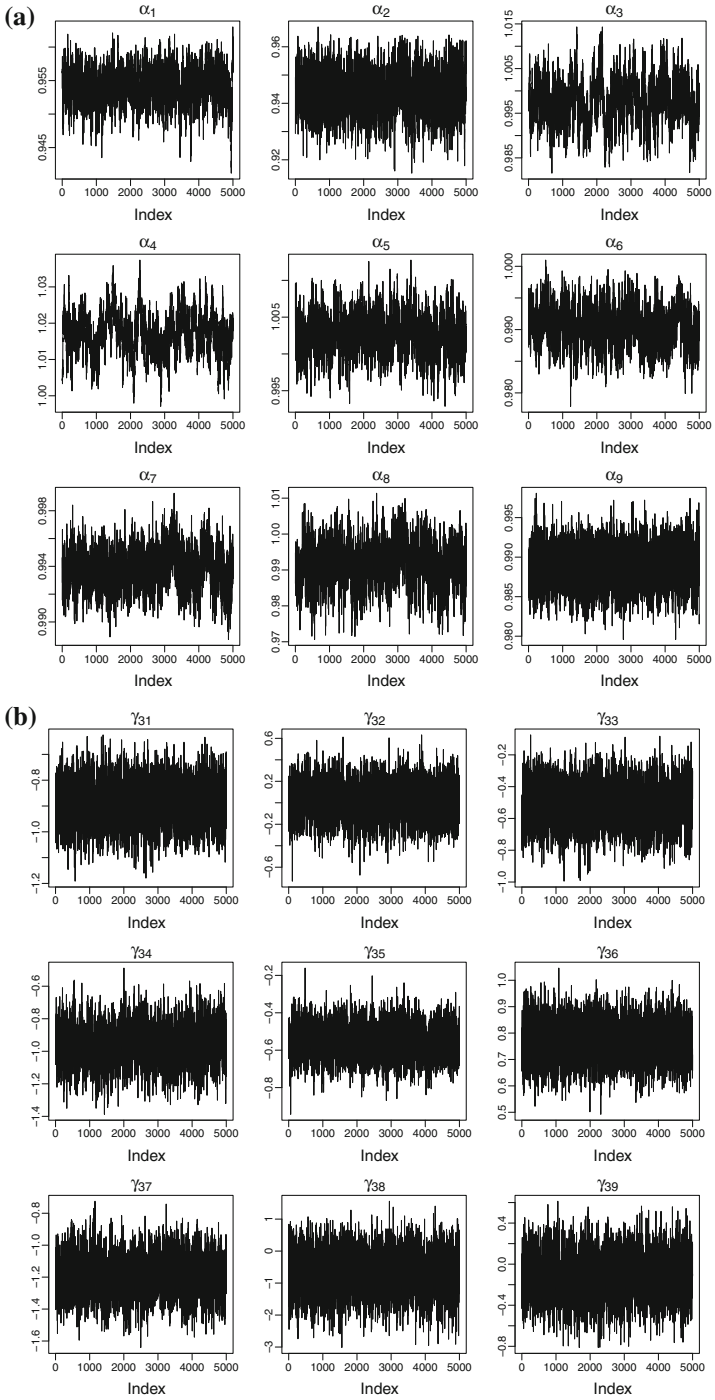

Fig. 4 Traceplots of parameters α_s (**a**) and $\gamma_{3,s}$ (**b**) with $s = 1, \ldots, 9$

account the differences of patients' demands. Moreover, also Lanzarone et al. (2010) found the CP to be significant for the number of visits (i.e., a cost distribution in terms of the number of weekly visits was assigned to each CP). As far as the evolution in time is concerned, $CP = 1$, $CP = 2$ and $CP = 6$ look like rather different from all others. This is coherent with the features of such profiles and the dataset itself, since $CP = 1$ represents low-intensity patients, $CP = 6$ palliative patients who form a well separated class, and $CP = 2$ denotes patients with a low demand (even if not the lowest). For instance, from Fig. 5a, posterior CIs of α_1 and α_2 are similar, but different from all the others, while, on the other hand, if we consider CIs of parameters β_s in Fig. 5b, β_1, β_2 and β_6 are rather different among themselves, but also differ from all other β_s. If we look at the CIs of γ_{3s} in Fig. 5c, which determine the average initial number of visits, it is clear that $\gamma_{3,6}$ assumes the largest values; this points out that, when palliative patients (i.e., with $CP = 6$) enter the study, they need much more visits than other patients, a conclusion that matches observed data. Moreover, observe that $\gamma_{3,8}$ has a much larger posterior variance since very few patients were assigned to $CP = 8$. We remark that the majority of patients in the dataset were assigned to $CP = 1$ and $CP = 7$ (see Table 2) for many time slots t of the observed period. This is the reason why CIs of parameters for these groups are generally smaller than the others.

As far as time-invariant covariates are considered, the posterior distribution of γ_1 is mostly constrained on positive values, whereas the credibility interval of γ_2 is quite symmetric around the null value, even if the posterior probability that $\gamma_2 < 0$ is larger than 0.5 (0.658). This means that the initial demand for visits strongly increases with age, and that patient's gender does not significantly affect this demand.

In conclusion, the posterior density of model parameters show a good fit of the autoregression generalized linear mixed model to the dataset and allow the model to catch clinical evidence.

5.3 Bayesian goodness-of-fit

In this section, we focus our attention on patients who are in charge at a certain time slot t and we want to predict the number of nurse visits at the following slot $t + 1$. As mentioned above, in our Bayesian formulation the prediction of nurse visits $\{N_{i,t+1} \ \forall i, t\}$ is straightforwardly obtained by means of posterior predictive distributions; see (3).

For goodness-of-fit purposes, we divided the dataset into a training set and a testing set according to a cross-validation approach. Patients who are in charge of the provider at week t are in the testing set, whereas all the others are in the training set. We computed the posterior density of model parameters again, considering only the training set. Then, we computed the predictive distributions (3) for patients in the testing set by means of the posterior density of parameter $\boldsymbol{\theta}$ obtained under the training set. Finally, we checked the predictions with the corresponding observed data of the testing set. Indeed, the number of visits in the following week is

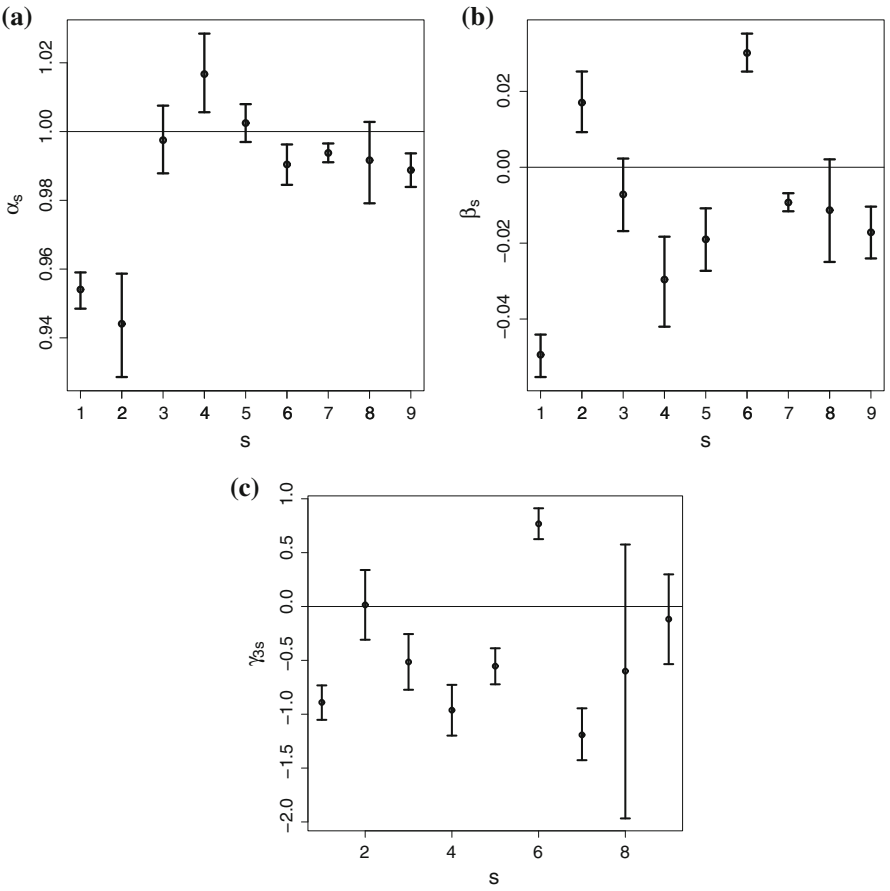

Fig. 5 Posterior 95 % credible intervals and medians of α_s, β_s and $\gamma_{3,s}$, with $s = 1, \ldots, 9$

Table 4 Posterior quantiles of the fixed-effects parameters

	2.5 %	50 %	97.5 %
γ_1	1.401	1.905	2.404
γ_2	−0.165	−0.030	0.107
σ	0.121	0.125	0.130

evaluated in advance by the model and compared with the real observation once it is available.

This validation procedure was applied at some of the 252 weeks of the dataset. For each one of them, we computed the posterior predictive distribution of the number of nurse visits for all the patients in charge at that week. Then, the accuracy of the predictions in a week is evaluated in terms of the Mean Absolute Error (MAE), which is defined as:

Table 5 MAE_{t+1} at the four analyzed weeks

$t+1$	100	150	176	235
MAE_{t+1}	0.455	0.530	0.499	0.473

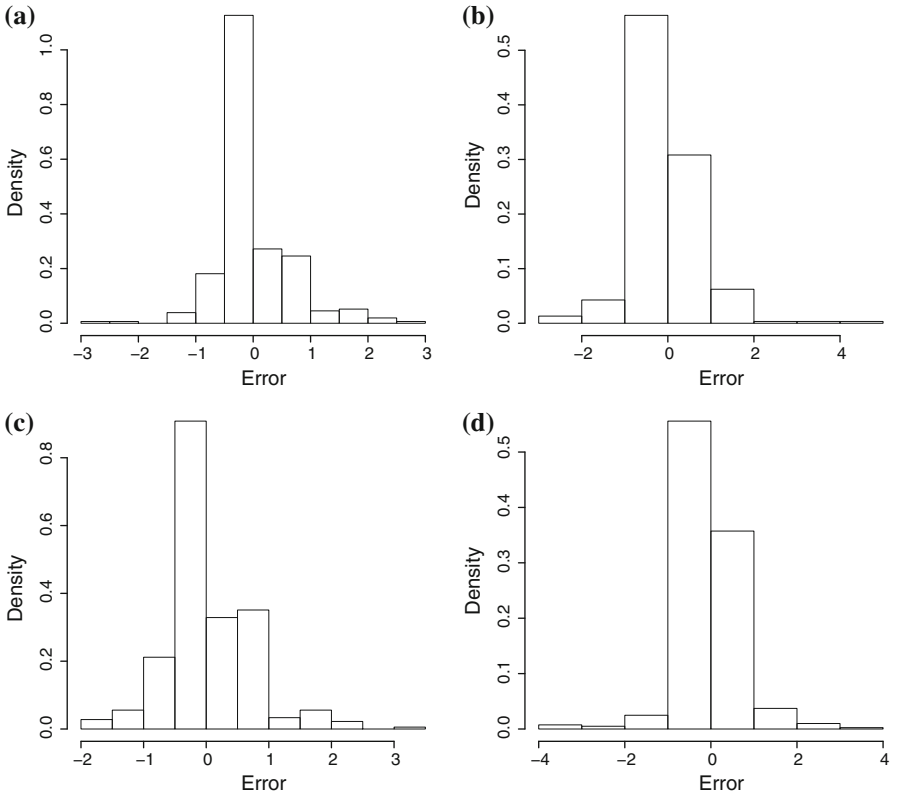

Fig. 6 Sample histograms of the errors for predictions at $t+1 = 100$ (**a**), $t+1 = 150$ (**b**), $t+1 = 176$ (**c**), and $t+1 = 235$ (**d**)

$$MAE_{t+1} = \frac{1}{m_t}\sum_{i=1}^{m_t}\left|n_{i,t+1} - \hat{N}_{i,t+1}\right|,$$

where m_t is the number of patients in charge at week t (i.e., the testing set), $n_{i,t+1}$ their observed numbers of nurse visits at week $t+1$, and $\hat{N}_{i,t+1}$ the corresponding Bayesian predictions of nurse visits at week $t+1$ given information at t. Here, $\hat{N}_{i,t+1}$ is taken as the mean of the predictive distribution. If a patient exits between time t and time $t+1$, the observed number $n_{i,t+1}$ is equal to 0. The lowest is MAE_{t+1}, the highest is the accuracy of our prediction at time $t+1$.

We computed the *MAE* at four different weeks, considering patients at $t = 99$, $t = 149$, $t = 175$ and $t = 234$. The number of patients in charge at those weeks are $m_{99} = 309$, $m_{149} = 305$, $m_{175} = 359$ and $m_{234} = 403$, respectively. The MAEs are

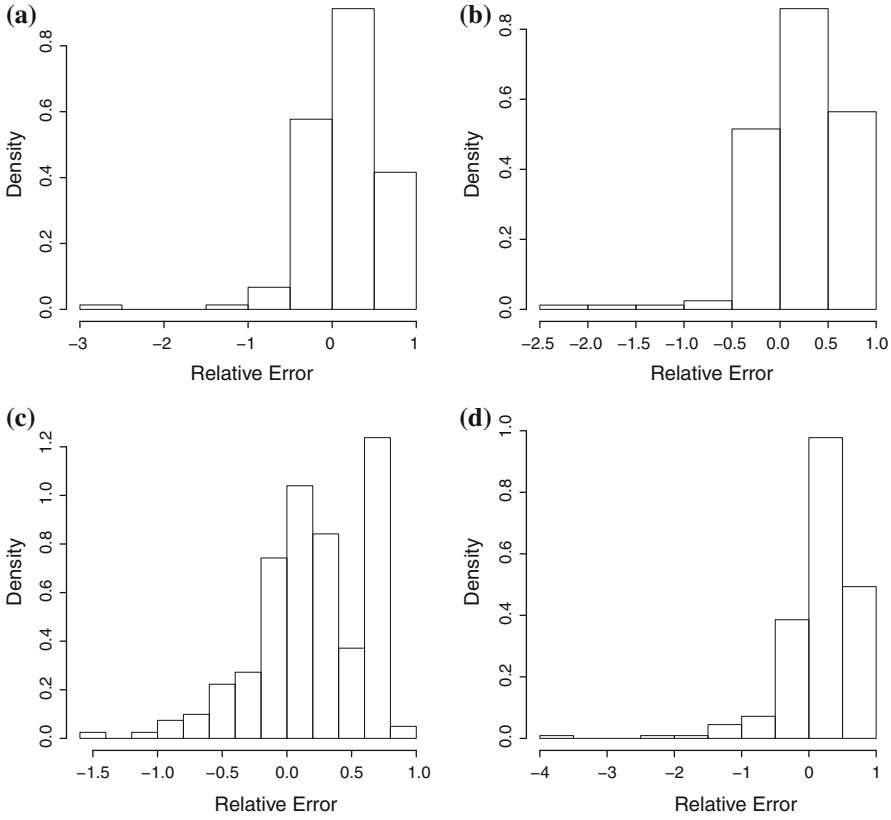

Fig. 7 Sample histograms of the relative errors for predictions at $t+1=100$ (**a**), $t+1=150$ (**b**), $t+1=176$ (**c**), and $t+1=235$ (**d**)

displayed in Table 5. The largest value is equal to 0.530 at week $t+1=150$, thus showing good fit of the model in predicting future nurses' visits.

To assess the accuracy of the predictions and to check the presence of asymmetric errors, Fig. 6 displays the sample histograms of the errors for all patients $i=1,\ldots,m_t$, where the error is defined as $n_{i,t+1}-\hat{N}_{i,t+1}$. The figure shows that accurate predictions are provided by the model, since small errors are observed in all 4 weeks. At each week, the error is null for the majority of patients (i.e., 59, 52, 51 and 52 % at $t+1$ equal to 100, 150, 176 and 235, respectively) and between -1 and 1 for a large percentage of them (i.e., 91, 92, 93 and 94 % at $t+1$ equal to 100, 150, 176 and 235, respectively). However, Fig. 6 points out a general underestimation, since positive errors are more frequent than negative ones.

In addition, we also analyzed the relative error for each patient i, defined as $(n_{i,t+1}-\hat{N}_{i,t+1})/n_{i,t+1}$. Figure 7 displays the sample histograms of the relative errors, which show limited relative errors even if worse than the errors in Fig. 6.

The Mean Absolute Errors were also computed separately for each CP:

Table 6 MAE^s_{t+1} at the four analyzed weeks, and average values among these weeks

$CP = s$	MAE^s_{100}	MAE^s_{150}	MAE^s_{176}	MAE^s_{235}	Average
1	0.471	0.538	0.487	0.452	0.487
2	0.184	0.324	0.440	0.405	0.338
3	0.604	0.577	0.362	0.648	0.548
4	0.282	0.798	0.500	0.504	0.521
5	0.737	0.615	0.717	0.746	0.704
6	0.781	0.762	0.813	0.489	0.711
7	0.350	0.402	0.398	0.289	0.350
8	0.478	0.337	0.432	0.000	0.312
9	0.219	0.432	0.351	0.457	0.365

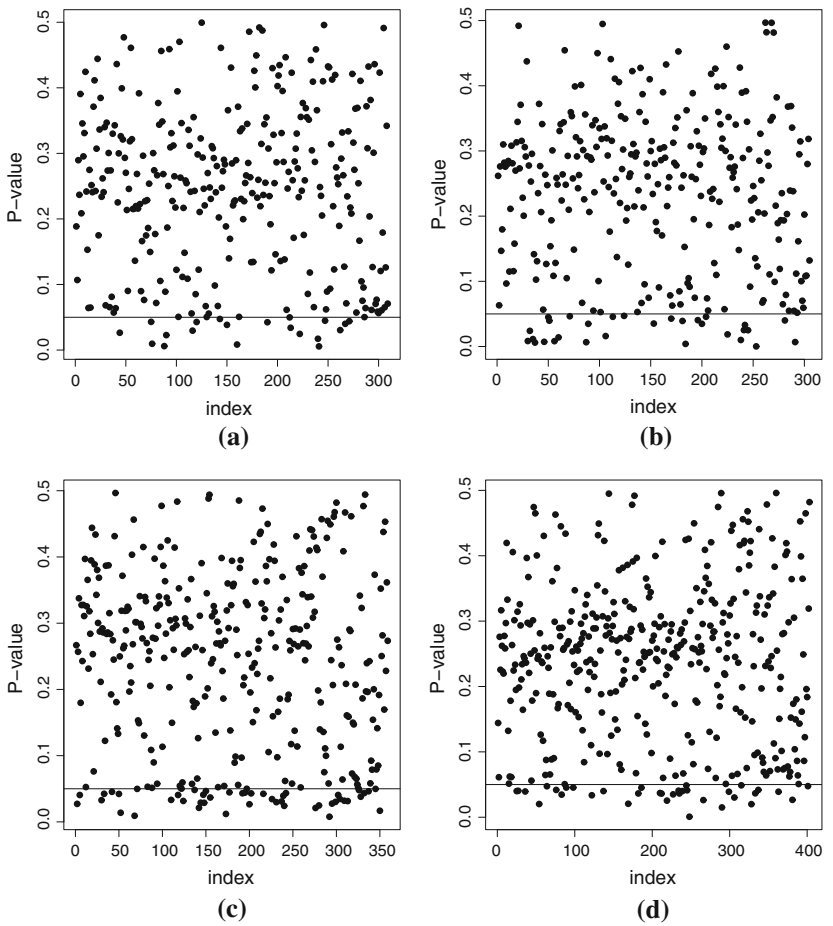

Fig. 8 Posterior predictive p values at week $t + 1 = 100$ (**a**), $t + 1 = 150$ (**b**), $t + 1 = 176$ (**c**) and $t + 1 = 235$ (**d**)

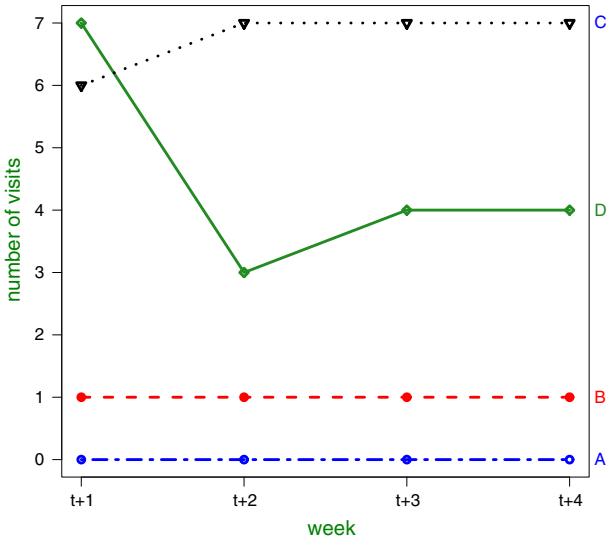

Fig. 9 Prediction at four future weeks for some patients

where m_t^s is the number of patients with $CP = s$ at week t. Table 6 shows that the largest MAE_{t+1}^s values are obtained for $CP = 5$ and $CP = 6$, which correspond to the profiles that require the highest number of visits. On the other hand, profiles with lower nominal demand yield smaller errors. This CP stratification contributes to maintain limited the relative errors.

In addition, we took into account the posterior predictive p values, according to Gelman et al. (2013). Posterior predictive p values is a tool for carrying out model evaluations and comparisons, which has become fairly popular in Bayesian goodness-of-fit, partly in consequence of its easy implementation by MCMC methods. In the words of Gelman et al, $N_{i,t+1}^{new}$ at week $t+1$ denotes the i-th "replicated data that could have been observed, or, to think predictively, as the data we would see tomorrow if the experiment that produced $n_{i,t+1}$ today were replicated with the same model and the same value of the parameter that produced the observed data" see (Gelman et al. 2013, Sect. 6.3). In particular, we computed

$$p\,\text{value}_{i,t+1} = \min\left\{P(N_{i,t+1}^{new} > n_{i,t+1}|data), P(N_{i,t+1}^{new} \leq n_{i,t+1}|data)\right\},$$

in terms of the predictive distribution of $N_{i,t+1}^{new}$.

A value close to 0 denotes that the model is inadequate for fitting this observation. To simplify, observations with posterior predictive p values less than 0.05 (say) are classified as unusual or *outlier*. Figure 8 displays posterior predictive p values for all patients at four different weeks. The percentage of outlying patients are 6, 9, 7, 8 %, obtained at week $t = 100, 150, 176$ and 235, once again showing a good fit of our model to the data.

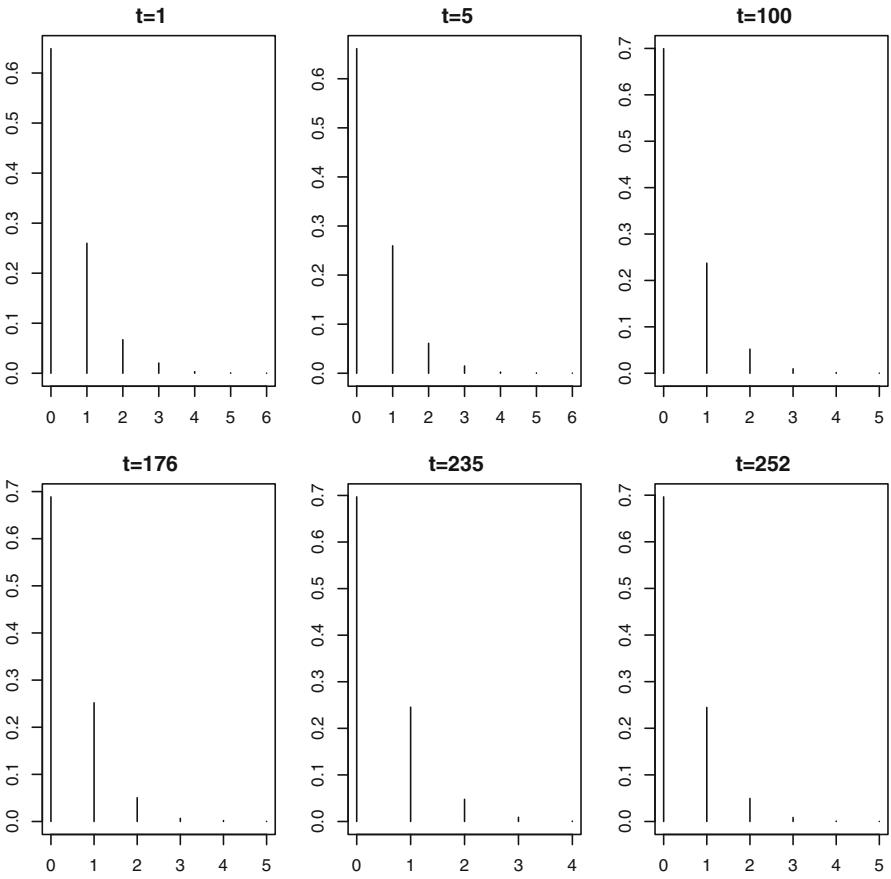

Fig. 10 Posterior predictive probabilities of the number of nurse visits at $t = 1$, $t = 5$, $t = 100$, $t = 176$, $t = 235$ and $t = 252$ for a new patient entered at $t = 1$

5.4 Prediction at future weeks for a patient already in charge

It is very important for the decision maker in HC to predict the number of visits required by a patient already in charge over future time periods (not only the next one). In the long term resource planning, this information will help in dimensioning the workforce or assigning operators to patients, in order to avoid future operator surcharges. So, the HC management can make a decision (e.g. the number of operators needed in next weeks, or the nurse-to-patient assignment) by means of the predicted number of visits in future weeks.

To show how we can use our model to this aim, we computed posterior predictive distributions for some patients when time is $t + 1,\ t + 2,\ t + 3,\ t + 4$, by iterating (3). In this calculation, we must fix the CP evolution in the future periods, since care profile is considered as a fixed covariate in our model. In the absence of further information, the most suitable choice is to assume that the CP remains fixed and

equal to the current one at time t. This assumption is legitimate since the revision period of CP is usually one month (=4 weeks).

As an example, in Fig. 9 we plotted the number of future nurse visits to four patients. Patients A and B correspond to $CP = 1$, while patients C and D to $CP = 6$.

5.5 Prediction for a newly admitted patient

We follow the approach described at the end of Sect. 4.2 for simulating the posterior predictive distribution of a hypothetical female patient i^*. She is 75 years old (close to the sample average age of patients in the dataset), who is admitted in the service at $t = 1$ and maintains CP equal to 1 at each week. We simulated the whole trajectory of this patient, i.e. $N_{i^*,t}$ for each $t = 1, \ldots, 252$. In Fig. 10 we display the posterior predictive probabilities of $N_{i^*,t}$ at different time slots t. Observe that, since $CP = 1$, all these distributions are concentrated around zero, i.e., this patient is not burdensome in terms of nurse visits. This prediction provides an immediate evaluation of the increased nurse workload after the assignment of a new patient and an updated overview of the workloads.

6 Conclusions

Home care providers need suitable skills and tools in order to predict the demand evolution of the patients in charge for supporting the delivery of care. In this paper, we have first explored the application of a Bayesian model to the HC context, in order to predict the demand for visits from patients in charge. The Bayesian approach is now widely recognized as a proper framework for analyzing patients demand in health care (Aven and Eidesen 2007; Spiegelhalter et al. 2004). In addition, we have applied the model to a dataset from one of the largest public Italian HC providers. The data consist of nurse visits of 3,095 patients who were in service over 252 weeks.

This paper is a first attempt to propose a Bayesian model to predict the nurse visits' demand. Here the main aim is to explore the extent to which Bayesian statistical analysis can and should be incorporated into HC research, for the purpose of assisting rational health care decision-making. The approach fits well in the HC context, and the results from the application to a relevant real case validate the model. Hence, the relevance of the model in the practice is guaranteed.

A distinctive feature of the Bayesian approach is the feasibility of performing the statistical analysis from a predictive point of view. This means that, on one hand, we can predict the number of nurse visits at future weeks for patients already in charge; on the other hand, we can predict the whole trajectory of a new patient entering the service. Both predictions could help the long-term resource planning, e.g., the assignment of an operator to a patient under continuity of care, in order to avoid future operator surcharges. In particular, we have computed Bayesian estimates of all the parameters, as well as posterior predictive probabilities, through a MCMC algorithm.

Comparing our approach with the frequentist model in Lanzarone et al. (2010), we can observe that the main advantage of our Bayesian model is its simpler

mathematical formulation. In fact, Lanzarone et al. (2010) provide a care pathway model and a cost function joined together, while the Bayesian model presented here is simply specified by assigning the conditional distribution of a set of random variables according to two levels of hierarchy (the likelihood and the prior). In this way, our model is easier to apply, and allows a higher degree of flexibility. On the other hand, the approach in Lanzarone et al. (2010) adopts a Markov chain evolution model to estimate the future evolution of CP, whereas in this paper the CP is considered as a fixed covariate and its evolution along with time is not taken into account. Both models involve the classification of the patients according to CPs; in addition, the Bayesian model includes two additional covariates which were neglected in Lanzarone et al. (2010), i.e., patient's age and gender. Finally, note that, in this paper, we have assumed that each patient enters and exit only once during his/her care pathway, whereas a state *pause* was adopted in Lanzarone et al. (2010) to consider such event. In terms of results, both models show good validation outcomes even if the results are not directly comparable.

To conclude, we believe that the Bayesian approach is an effective and promising tool for dealing with the randomness in the HC setting, and that there are potential areas of development for such an approach in HC. For instance, our future work will deal with a Bayesian joint modeling and prediction of the number of visits and the care pathway evolution. In this way, there will be no need to assume the CP as fixed.

Acknowledgments I. Nawajah acknowledges the Italian Ministry of Foreign Affairs and Hebron University for the support through the Italian project E-PLUS "Enhancement of the Palestinian University System".

References

Alagoz O, Bryce CL, Shechter S, Schaefer A, Chang CCH, Angus DC, Roberts MS (2005) Incorporating biological natural history in simulation models: empirical estimates of the progression of end-stage liver disease. Med Decis Mak 25(6):620–632

Anderson LH, Flottemesch TJ, Fontaine P, Solberg LI, Asche SE (2012) Patient medical group continuity and healthcare utilization. Am Manag Care 18(8):450–457

Asquer G, Borsani V, Matta A (2007) Analisi della struttura organizzativa degli erogatori del servizio di assistenza domiciliare. Politiche Sanitarie 8(3):95–117

Aven T, Eidesen K (2007) A predictive Bayesian approach to risk analysis in health care. BMC Med Res Method 7(1):7–38

Bergamaschi R, Romani A, Tonietti S, Citterio A, Berzuini C, Cosi V (2000) Usefulness of Bayesian graphical models for early prediction of disease progression in multiple sclerosis. Neurol Sci 21(2):819–823

Berzuini C, Allemani C (2004) Effectiveness of potent antiretroviral therapy on progression of human immunodeficiency virus: Bayesian modelling and model checking via counterfactual replicates. Roy Statist Soc Series C (Appl Stat) 53(4):633–650

Blanco-Encomienda FJ (2013) A multi-state Markov model for projecting health care spending. Eur. Sci. J. 9(21):745–752

Borsani V, Matta A, Beschi G, Sommaruga F (2006) A home care scheduling model for human resources. Third IEEE/International conference on service systems and service management, Troyes, France 1:449–454

Carello G, Lanzarone E (2014) A cardinality-constrained robust model for the assignment problem in home care services. Eur J Oper Res 236(2):748–762

Carreras G, Baccini M, Accetta G, Biggeri A (2012) Bayesian probabilistic sensitivity analysis of Markov models for natural history of a disease: an application for cervical cancer. Ital J Public Health 9(3):1–10
Comondore VR, Devereaux P, Zhou Q, Stone SB, Busse JW, Ravindran NC, Burns KE, Haines T, Stringer B, Cook DJ et al (2009) Quality of care in for-profit and not-for-profit nursing homes: systematic review and meta-analysis. Brit Med J 339:b2732
Congdon P (2001) The development of gravity models for hospital patient flows under system change: a Bayesian modelling approach. Health Care Manag Sci 4(4):289–304
Faddy M, McClean S (2005) Markov chain modelling for geriatric patient care. Methods Inf Medicine-Methodik der Information in der Medizin 44(3):369–373
Fahrmeir L, Tutz G (1994) Multivariate statistical modelling based on generalized linear models. Springer, New York
Gelman A (2006) Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and Draper). Bayesian Anal 1(3):515–534
Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian data analysis, 3rd edn. Chapman & Hall/CRC Press, Boca Raton, FL
Giardina F, Guglielmi A, Quintana FA, Ruggeri F (2011) Bayesian first order auto-regressive latent variable models for multiple binary sequences. Stat Model 11(6):471–488
Gómez-Batiste X, Martínez-Muñoz M, Blay C, Espinosa J, Contel JC, Ledesma A (2012) Identifying needs and improving palliative care of chronically ill patients: a community-oriented, population-based, public-health approach. Curr Opin Support Palliat Care 6(3):371–378
Guglielmi A, Ieva F, Paganoni AM, Ruggeri F (2010) A Bayesian random-effects model for survival probabilities after acute myocardial infarction. Chil J Stat 3(1):15–29
Hollander M, Chappell N, Havens B, McWilliam C, Miller JA (2002) Substudy 5: study of the costs and outcomes of home care and residential long term care services: a report prepared for the health transition fund, Health Canada, National evaluation of the cost-effectiveness of home care
Jackman S (2009) Bayesian analysis for the social sciences. Wiley, New York
Jones J, Wilson A, Parker H, Wynn A, Jagger C, Spiers N, Parker G (1999) Economic evaluation of hospital at home versus hospital care: cost minimisation analysis of data from randomised controlled trial. Brit Med J 319(7224):1547–1550
Krahn M, Wong JB, Heathcote J, Scully L, Seeff L (2004) Estimating the prognosis of hepatitis C patients infected by transfusion in Canada between 1986 and 1990. Med Decis Mak 24(1):20–29
Lanzarone E, Matta A (2012) A cost assignment policy for home care patients. Flex Serv Manuf J 24(4):465–495
Lanzarone E, Matta A, Scaccabarozzi G (2010) A patient stochastic model to support human resource planning in home care. Prod Plan Control 21(1):3–25
Lanzarone E, Matta A, Sahin E (2012) Operations management applied to home care services: the problem of assigning human resources to patients. IEEE Trans Syst Man Cyber A 42(6):1346–1363
Marshall A, Vasilakis C, El-Darzi E (2005) Length of stay-based patient flow models: recent developments and future directions. Health Care Manag Sci 8(3):213–220
Marshall AH, McClean SI, Shapcott C, Millard PH (2002) Modelling patient duration of stay to facilitate resource management of geriatric hospitals. Health Care Manag Sci 5(4):313–319
Matta A, Chahed S, Sahin E, Dallery Y (2014) Modelling home care organisations from an operations management perspective. Flex Serv Manuf J 26(3):295–319
McClean S, McAlea B, Millard P (1998) Using a Markov reward model to estimate spend-down costs for a geriatric department. Oper Res Soc 49(10):1021–1025
Pauler DK, Finkelstein DM (2002) Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. Stat Med 21(24):3897–3911
Plummer M (2003) Jags: a program for analysis of bayesian graphical models using gibbs sampling. In: Proceedings of the 3rd international workshop on distributed statistical computing (DSC 2003), pp 20–22
Plummer M, Best N, Cowles K, Vines K (2006) Coda: convergence diagnosis and output analysis for mcmc. R News 6(1):7–11. http://CRAN.R-project.org/doc/Rnews/
Robert CP, Casella G (2004) Monte Carlo statistical methods. Springer, New York
Spiegelhalter D, Abrams K, Myles J (2004) Bayesian approaches to clinical trials and health-care evaluation. Wiley, Hoboken
Taylor G, Mclean S, Millard P (1996) Geriatric-patient flow-rate modelling. Math Med Biol 13(4):297–307

Verotta D (2005) Models and estimation methods for clinical hiv-1 data. Comput Appl Math 184(1):275–300
Woodward CA, Abelson J, Tedford S, Hutchison B (2004) What is important to continuity in home care? Perspectives of key stakeholders. Soc Sci Med 58(1):177–192
Zeger SL, Karim MR (1991) Generalized linear models with random effects; a Gibbs sampling approach. J Am Sat Assoc 86(413):79–86