

Joint Prediction of Demand and Care Duration in Home Care Patients: a Bayesian Approach

Previsione Congiunta della Domanda e della Durata di Cura in Assistenza Domiciliare: un Approccio Bayesiano

Inad Nawajah, Raffaele Argiento, Alessandra Guglielmi and Ettore Lanzarone

Abstract Home Care (HC) service is an alternative to conventional hospitalization and consists of delivering medical, paramedical and social services to patients at their homes rather than in hospitals or nursing homes. Human resource planning in HC is a difficult task and, for a good quality of planning, knowledge of future patients' demands is required. The aim of this paper is to propose a Bayesian model for predicting the number of visits required by HC patients, which is fundamental for planning human and material resources, and at the same time describing the natural history of Care Profiles. We model patients' holding times, i.e., the duration of Care Profiles, and the number of nurses' visits at each future time slot. The model has been applied to the real data of one of the largest public HC providers in Italy. We computed the estimates of all model parameters and the predictions for both new patients and patients already in the charge. Preliminary results show the applicability of the approach in the practice and good quality of predictions.

Abstract *L'Assistenza Domiciliare (HC) è un'alternativa al ricovero ospedaliero convenzionale e consiste nell'erogazione di prestazioni mediche, paramediche e sociali ai pazienti direttamente a casa loro piuttosto che negli ospedali o nelle case di cura. La pianificazione delle risorse umane in Assistenza Domiciliare è un compito difficile e, per una pianificazione di buona qualità, è richiesta la conoscenza delle domande future dei pazienti. Lo scopo di questo lavoro è proporre un modello bayesiano per prevedere il numero di visite richieste dai pazienti in HC, che risulta fondamentale per pianificare le risorse umane e materiali, e allo stesso tempo l'evoluzione dei Profili di Cura. Sono stati modellati i tempi di permanenza dei pazienti, cioè le durate dei Profili di Cura, e il numero di visite infermieristiche*

Inad Nawajah, Alessandra Guglielmi

Dipartimento di Matematica, Politecnico di Milano, P.za Leonardo da Vinci 32, 20133 Milan, Italy
e-mail: inad.nawajah@mail.polimi.it; alessandra.guglielmi@polimi.it

Raffaele Argiento, Ettore Lanzarone

Istituto di Matematica Applicata e Tecnologie Informatiche (IMATI), Consiglio Nazionale delle Ricerche (CNR), Via Bassini 15, 20133, Milan, Italy
e-mail: raffaele@mi.imati.cnr.it; etttore@mi.imati.cnr.it

per ogni intervallo di tempo futuro. Il modello è stato applicato ai dati reali di uno dei più grandi erogatori pubblici di HC in Italia. Abbiamo calcolato le stime di tutti i parametri del modello e le previsioni sia per i nuovi pazienti che per quelli già in carico. I risultati preliminari mostrano l'applicabilità pratica del metodo e l'accuratezza delle previsioni.

Key words: Home Care, Bayesian Modeling and Estimation, Generalized Linear Mixed Effects Model

1 Introduction

Many random events affect Home Care (HC) service delivery and mine the feasibility of plans. Among them, the most relevant randomness source is related to changes in patients' conditions, who consequently have a demand for visits different from the planned one [1, 2]. For this reason, each reliable resource planning should require an accurate estimation of the demand in future time periods. In the literature, only few works deal with the representation of HC patients' demand evolution along with the time and its prediction in future periods. Lanzarone et al. [1] proposed a frequentist Markovian model, whereas Argiento et al. [3] proposed a Bayesian framework, where the Care Profile (CP) is considered as a fixed covariate.

CP is the patient classification often adopted by HC providers, based on the specific requirements and the costs of the provided service. Usually, a CP is assigned to each patient at the beginning of his/her care pathway and monthly confirmed or changed; however, CP can be modified in advance in case of a sudden variation in patient's conditions.

In this paper, we propose a Bayesian model for jointly describing and predicting the evolution of HC patients' demands and CPs over time. Indeed, we generalize the model in Argiento et al. [3] considering the CP as a response variable.

2 Dataset

We consider n patients in the charge of a HC provider over a period divided into discrete time slots. Each patient i enters the service at time slot $T_L(i)$ and exits at time slot $T_U(i)$. Data observed for each patient i at each time slot $t \in \{T_L(i), \dots, T_U(i)\}$ are:

- number of visits $N_{i,t}$ required to nurses (count data);
- Care Profile $CP_{i,t}$ (categorical covariate evolving in time that assume R integer values, indexed by $r = 1, \dots, R$).
- age $age_{i,t}$.

Moreover, each patient i is characterized by the gender (sex_i – categorical variable).

3 Bayesian Model

We consider $CP_{i,t}$ as a response variable, modeled and estimated together with $N_{i,t}$. Hence, we introduce the joint distribution of $N_{i,t}$ and $CP_{i,t}$, which is then factorized as the conditional distribution of $N_{i,t}$ given $CP_{i,t}$ times the marginal distribution of $CP_{i,t}$.

In particular we assume that, for each patient i , the transitions between CPs are regulated by a multistate Markov Chain, whereas the holding times of CPs depend on all CPs durations. Indeed, let $\boldsymbol{\eta}_i = (\eta_{i,1}, \eta_{i,2}, \dots, \eta_{i,J(i)})$ be the sequence of all of the $J(i)$ different values assumed by patient i 's CP during the care pathway (i.e., the visited states). Let also $H_{i,j}$ be the number of time slots that patient i spent in his/her j -th observed CP (i.e., in $\eta_{i,j}$). In this way, each patient i is characterized by the sequences of Care Profiles $\boldsymbol{\eta}_i = (\eta_{i,1}, \dots, \eta_{i,J(i)})$, of holding times $\mathbf{H}_i = (H_{i,1}, \dots, H_{i,J(i)})$, and of number of nurse visits $\mathbf{N}_i = (N_{i,T_L(i)}, N_{i,T_L(i)+1}, \dots, N_{i,T_U(i)})$. Hence, the CP trajectory $\{CP_{i,t}, t = T_L(i), \dots, T_U(i)\}$ can be represented by two vectors $\boldsymbol{\eta}_i$ and $\mathbf{H}_i$, so that we model $\mathcal{L}(\{CP_{i,t}\}_t)$ by assigning $\mathcal{L}(\{(\eta_{i,j}, H_{i,j})\}_j)$. For each patient $i = 1, \dots, n$, we assume:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\eta}_i, \mathbf{H}_i) &= \mathcal{L}(\eta_1) \mathcal{L}(H_1 | \eta_1) \mathcal{L}(\eta_2 | \eta_1) \mathcal{L}(H_2 | H_1, \eta_2) \dots \\ &\dots \mathcal{L}(\eta_J | \eta_{J-1}) \mathcal{L}(H_J | H_1, \dots, H_{J-1}, \eta_J). \end{aligned}$$

We assume conditional independence among patients, i.e., among the trajectories $\{CP_{i,t}, N_{i,t}, t = T_L(i), \dots, T_U(i)\}$. The contribution to the likelihood of each patient i is (conditionally to the same covariates and parameters, not reported here) as follows:

$$\begin{aligned} \mathcal{L}(\{CP_{i,t}\}_t, \{N_{i,t}\}_t) &= \mathcal{L}(\{CP_{i,t}\}_t) \mathcal{L}(\{N_{i,t}\}_t | \{CP_{i,t}\}_t) \\ &= \mathcal{L}(\{CP_{i,t}\}_t) \prod_t \mathcal{L}(N_{i,t} | CP_{i,t}) \end{aligned}$$

where $\mathcal{L}(\{CP_{i,t}\}_t)$ is the law of the process described above, and $\mathcal{L}(N_{i,t} | CP_{i,t})$ the law of a *Generalized Linear Mixed Effects Model* (GLMM).

We model $N_{i,t}$ of patient i at time slot t as a Poisson distribution whose rate parameter depends on the current CP and the covariates. In particular, we propose the following formulation:

$$N_{i,t} | CP_{i,t} = r \sim \text{Pois}(\lambda_r e^{x_{i,t}^T \boldsymbol{\gamma}}), \quad r = 1, \dots, R, \quad T_L(i) \leq t \leq T_U(i) \quad (1)$$

where $x_{i,t}$ is the vector of the *covariates* of patient i at time slot t (i.e., $age_{i,t}$ and sex_i) and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$ is the corresponding regression parameter vector. Moreover, the latent variable λ_r represents the health state of patients with $CP_{i,t} = r$, which is responsible for the demand for visits (i.e., the bigger the parameter λ_r is, the worse patient's conditions are, and the higher the expected number of visits is). Here we assume:

$$\log(\lambda_r | \mu_\lambda, \sigma_\lambda^2) \stackrel{\text{iid}}{\sim} N(\mu_\lambda, \sigma_\lambda^2), \quad r = 1, \dots, R \quad (2)$$

where μ_λ and σ_λ^2 are given according to the prior beliefs on the health states $\lambda_1, \dots, \lambda_R$.

As far as the law of the CP trajectories is concerned, we assume that the holding times are distributed according to Negative Binomial distribution on $\{1, 2, \dots\}$, $NB(z, q)$, where q denotes the probability of “success” and z denotes the “predefined number of failures before success”:

$$\begin{aligned} H_{i,j} | \eta_{i,j}, H_{i,j-1}, \dots, H_{i,1} &\sim NB(z[\eta_{i,j}], q_{i,j}) \quad 1 < j \leq J(i) \\ H_{i,1} &\sim NB(z[\eta_{i,1}], q_{i,1}). \end{aligned} \quad (3)$$

Here the parameters $q_{i,j}$ are modelled through the following logit regression:

$$\begin{aligned} \text{logit}(q_{i,j}) &= \log \frac{q_{i,j}}{1 - q_{i,j}} = \beta_1[\eta_{i,j}] + (H_{i,1} + \dots + H_{i,j-1})\beta_2 \\ \text{logit}(q_{i,1}) &= \beta_1[\eta_{i,1}]. \end{aligned}$$

Model (1)-(3) results a GLMM with three fixed-effects parameters (γ_1, γ_2 and β_2) and two random effects parameters ($\beta_{1,r} = \beta_1[\eta_{i,j} = r]$ and $z_r = z[\eta_{i,j} = r]$ that depend on the current state).

Finally, we consider the transition probability matrix $\mathbf{P}$, whose elements $P_{r,s}$ denote the probability that CP moves from the r -th to s -th category. In the following, we present the Bayesian regression setup for the $R+1$ dimensional Markov chain describing the CPs, where the last state $R+1$ represents the exit from the service and is an absorbing state. The successive CP states $\boldsymbol{\eta}_j$ are modeled by a homogeneous Markov chain with states $\{1, \dots, R+1\}$. We assume that the initial state $\eta_{i,1} \sim \text{categorical}(\boldsymbol{\pi}_r)$, with $\pi_r \geq 0$ for each r and $\sum_{r=1}^R \pi_r = 1$. The transition matrix $\mathbf{P}$ between the states is here modeled with independent rows, and each row

$$(P_{r,1}, \dots, P_{r,R+1}) \sim \text{Dirichlet}(a_1, \dots, a_{R+1}).$$

Typically, in the absence of prior information, we set all parameters a_r equal to 1. Self-transitions are not allowed (i.e., $P_{r,r} = 0, \forall r = 1, \dots, R$) according to the definition of holding times, and $P_{R+1,R+1} = 1$ almost surely as $R+1$ is an absorbing state.

Finally, all parameters ($\gamma_1, \gamma_2, \mu_\lambda, \sigma_\lambda, \lambda_1, \dots, \lambda_R, P_{r,1}, \dots, P_{r,R}, \beta_{1,1}, \dots, \beta_{1,R}, \beta_2, z_1, \dots, z_R$ are a priori (conditionally) independent. The description of the model is completed by specifying the prior distributions for $\mu_\lambda, \sigma_\lambda, \gamma_1, \gamma_2, \boldsymbol{\beta}_1, \beta_2$, and $\mathbf{z}$. We adopt the following marginal prior densities:

$$\begin{aligned} \mu_\lambda &\sim \mathcal{N}(0, \sigma_\mu^2), \quad \sigma_\mu \sim U(0, 5), \quad \sigma_\lambda \sim U(0, 2), \quad \gamma_1, \gamma_2 \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1000), \\ \beta_{1,r} | \sigma_{\beta_1}^2 &\stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{\beta_1}^2), \quad r = 1, \dots, R, \quad \text{with } \sigma_{\beta_1} \sim U(0, 2), \quad \beta_2 \sim \mathcal{N}(0, 1000), \\ z_r &\stackrel{\text{iid}}{\sim} \Gamma(2, 2), \quad r = 1, \dots, R, \end{aligned}$$

i.e., a vaguely informative prior is assumed on the parameters.

4 Application to Real Case

We applied the model to one of the largest Italian HC providers; a dataset from the same provider has already been analyzed either in the frequentist and Bayesian framework [1, 3]. The week is considered as the time slot and 252 weeks - from January 2004 to March 2008 - are included in the study. The provider consists of three divisions and we refer to patients of the largest one; moreover, we include only patients who entered and left the service once within the time window, and patients whose care pathway is entirely contained in the 252 weeks. In this way, our dataset consists of 2358 patients. Patients are grouped in two main categories (palliative and non palliative), and each category includes a certain number of CPs. Fifteen CPs are present in the provider [1]; however, similar CPs are joined together as in [3], thus reducing R to 9. The model was implemented in JAGS [4] to obtain the posterior densities of parameters, considering a burn-in of 5000 iterations, a thinning of 50 iterations, and a final sample size of 5000. The chains passed the standard convergence tests.

In this study, our goal is to predict the demand for visits at future time slots. This is highly important for HC decision makers, who are interested in improving the service efficiency. In fact, predictions of patients' demands allow HC managers to make robust decisions (e.g., the number operators needed in next weeks and the optimal nurse-to-patient assignments under continuity of care) that are supported by the prediction of the demand for visits from patients in the charge. However, here we do not show results of the predictions due to lack of space.

Fig. 1 displays the 95% credibility intervals and the posterior medians of parameters $z_r, \beta_{1,r}$ and $\log(\lambda_r)$ (with $r = 1, \dots, 9$), whereas posterior quantiles of the other parameters not depending on the CP (i.e., γ_1, γ_2 , and β_2) are reported in Tab. 1.

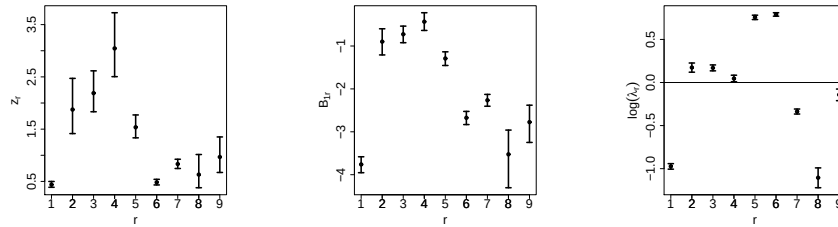


Fig. 1 Posterior 95% credible intervals and medians of $z_r, \beta_{1,r}$ and $\log(\lambda_r)$, with $r = 1, \dots, 9$.

Fig. 1 shows that the random-effects parameters are significantly different with respect to the CP. This means that the posterior distribution of the holding times $\{H_{i,j}\}$ strongly depends on patient's CP. For instance, since larger values of β_1 imply larger expected values of the holding times, it's clear from Fig. 1 that patients with CP equal to 2, 3, 4 and 5 tend to have larger holding times. This agrees with the clinical evidence because, for the HC provider we are considering, the division

Table 1 Posterior quantiles of the fixed-effects parameters.

	2.5%	50%	97.5%
(age) γ_1^*	0.906	0.980	1.057
(gender) γ_2^*	0.098	0.119	0.138
(slope) β_2	-0.012	-0.010	-0.008

among CPs is done on purpose to classify patients based on care duration and level of demand. Palliative patients show a high number of requested visits at the beginning of their care pathway (λ_6 refers to palliative patients who usually have a high demand for visits for a limited time period). Finally, we remark that larger credibility intervals in Fig. 1 are generally obtained for profiles with a smaller number of observations, as for CP=8 and CP=9.

As far as fixed-effects covariates are considered, the posterior distribution of γ_1 and γ_2 are mostly concentrated on positive values (Tab. 1), meaning that the demand for visits increases with age and it is larger for female patients. On the other hand, β_2 is a posteriori concentrated on negative values (i.e., the holding time at j increases when the summation of all holding times until $j - 1$ decreases).

Table 2 Posterior means of transition probability matrix $\mathbf{P}$.

$r \backslash s$	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
P_1	0.0000	0.0105	0.0398	0.0230	0.0274	0.0356	0.0293	0.0021	0.0063	0.8261
P_2	0.0359	0.0000	0.0417	0.0237	0.0298	0.0357	0.3231	0.0241	0.0059	0.4802
P_3	0.0300	0.0734	0.0000	0.0707	0.0478	0.0091	0.3682	0.0046	0.0046	0.3916
P_4	0.0207	0.0758	0.2149	0.0000	0.0525	0.0159	0.3065	0.0068	0.0024	0.3046
P_5	0.0073	0.0288	0.1765	0.2268	0.0000	0.0215	0.1835	0.0018	0.0036	0.3503
P_6	0.0039	0.0013	0.0013	0.0013	0.0026	0.0000	0.0013	0.0013	0.0013	0.9855
P_7	0.0592	0.0180	0.0495	0.0221	0.0192	0.0151	0.0000	0.0220	0.0014	0.7936
P_8	0.1429	0.0283	0.0863	0.0287	0.0860	0.0583	0.0859	0.0000	0.0282	0.4554
P_9	0.0302	0.0152	0.0151	0.0304	0.0154	0.0303	0.0154	0.0152	0.0000	0.8328

Posterior means of $\mathbf{P}$ (Tab. 2) show that palliative patients (CP=6) exit the study with higher probability than non palliative ones because most of the palliative patients die after a relatively short time. Moreover, also extemporary patients (CP=1 and 9) exit the service with high probability (high posterior means of $P_{1,10}$ and $P_{9,10}$), because such patients tend to maintain the same profile during the entire care pathway.

To conclude, in this paper, we have proposed a Bayesian approach for modeling and predicting the evolution of HC patients' demand over time. Here, we show that the posterior density of model parameters demonstrates a good fit of the generalized linear mixed model to the dataset, thus allowing the model to catch the clinical evidence.

References

1. Lanzarone, E., Matta, A., Saccabarozi, G.: A patient stochastic model to support human resource planning in home care. *Prod. Plan. Control.* **21**, 3-25 (2010)
2. Lanzarone, E., Matta, A., Sahin, E.: Operations management applied to home care services: the problem of assigning human resources to patients. *IEEE Trans Syst Man Cyber A.* **42**, 1346-1363 (2012)
3. Argiento, R., Guglielmi, A., Lanzarone, E., Nawajah, I.: A Bayesian framework for describing and predicting the stochastic demand for care in home care patients. *Under second review* (2014)
4. Plummer, M.: JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 20-22 (2003) <http://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf>