Production Planning & Control

ISSN: 0953-7287 (Print) 1366-5871 (Online) Journal homepage: http://www.tandfonline.com/loi/tppc20

A patient stochastic model to support human resource planning in home care

Ettore Lanzarone , Andrea Matta & Gianlorenzo Scaccabarozzi

To cite this article: Ettore Lanzarone , Andrea Matta & Gianlorenzo Scaccabarozzi (2010) A patient stochastic model to support human resource planning in home care, Production Planning & Control, 21:1, 3-25, DOI: 10.1080/09537280903232362

To link to this article: http://dx.doi.org/10.1080/09537280903232362

Published online: 14 Oct 2009.

Production Planning & Control
Vol. 21, No. 1, January 2010, 3–25

A patient stochastic model to support human resource planning in home care

Ettore Lanzarone[a]*, Andrea Matta[a] and Gianlorenzo Scaccabarozzi[b]

[a]*Dipartimento di Meccanica, Politecnico di Milano, Milan, Italy;* [b]*Dipartimento Interaziendale della Fragilità ASL/AO di Lecco, Lecco, Italy*

(*Received 2 April 2009; final version received 17 July 2009*)

A large number of variables and unpredictable events affect the quality of home care (HC) services. Changes in patient clinical and social conditions and troubles in service organisation are only some examples that can make the management of HC activities quite difficult. The estimation of patient requirements would support HC providers in human resource planning before the care execution, thus improving the service efficiency. This article proposes a stochastic model to represent the patient's care pathway; on the basis of historical data of an HC structure, the model provides predictions on the major variables of interest: how many patients are followed up in the course of time and, for each of them, the duration of care and the amount of required visits. The predicted variables of interest provide information about the future workload of each operator. This becomes a useful support tool for human resource planning in the medium and short terms. Numerical results prove the applicability of the proposed stochastic model in practice.

Keywords: resource planning; stochastic model; home care; Markov chain; patient pathway

1. Introduction

Home care (HC) includes medical, paramedical and social services delivered to patients at their home. The main purpose of HC is to alleviate the pain of the patients and to improve or sustain their health and quality of life. HC services are generally provided to patients with complex ailments, such as the elderly and the terminally ill; however, other categories of patients can be involved, such as children, post-surgery patients and those stricken by cerebral ictus (Davies and Dale 2003, Chahed *et al*. 2006, 2009). The main benefit of HC is the decreased hospitalisation rate, which leads to significant cost savings in the entire health care system (Jones *et al*. 1999, Hollander *et al*. 2002).

Resource planning is crucial for operating in HC organisations, since several human and material resources have to be properly managed in order to avoid process inefficiencies, treatment delays and low quality of service (Chahed *et al*. 2006, 2009). Unfortunately, resource planning in HC organisations is quite complex, even when compared to industrial organisations. The complexity is mainly due to the large number of assisted patients, synchronisation of resources and service delivery in an often vast territory. Moreover, random events that affect the delivery of service mine the feasibility of resource plans. The major sources of randomness are the patient conditions, resource availability and the duration of operator transfers in the territory. All of these sources can cause the revision of both the patient care pathway and/or the HC resource planning. Furthermore, HC providers typically suffer from their staff's lack of suitable skills, methodologies and tools for managing logistic and organisational activities to support the care delivery (Castelnovo *et al*. 2006, Borsani *et al*. 2006). Thus, in order to solve the numerous problems emerging from the daily disruptive events, plans are frequently modified in real time (Borsani *et al*. 2006).

The possibility of estimating the patients' health progression during their care pathway could help HC organisations in developing decisional support tools for robust resource planning. The literature presents various studies focused on developing stochastic models for representing patient conditions in health care systems. Some models evaluate the evolution of patient conditions and provide estimates about the necessary resources (Taylor and McClean 1996, Taylor *et al*. 1997, McClean *et al*. 1998, McClean and Millard 1998, Taylor *et al*. 2000, Congdon 2001, Marshall *et al*. 2002, Krahn *et al*. 2004, Marshall and McClean 2004, Marshall *et al*. 2005, Koizumi *et al*. 2005, McClean and Millard 2007). Other models deal with studying the evolution of certain pathologies or investigate the cost-effectiveness of specific therapies in treated patients (Bennett *et al*. 1997, Flessa 1999,

*Corresponding author. Email: ettore.lanzarone@polimi.it

ISSN 0953–7287 print/ISSN 1366–5871 online

DOI: 10.1080/09537280903232362
http://www.informaworld.com

Bergamaschi *et al*. 2000, Verbeek *et al*. 2001, Pauler and Finkelstein 2002, Emparan *et al*. 2003, Kousignian *et al*. 2003, Remák *et al*. 2003, Berzuini and Allemani 2004, Husted *et al*. 2005, Alagoz *et al*. 2005, Magherini *et al*. 2005, Altman and Petkau 2005, Verotta 2005). However, none of them deal with HC systems. This article aims to fill this lack of HC patient modelling, by proposing an approach for resource planning.

This work proposes an approach for developing stochastic models to support resource planning in HC organisations by describing the stochastic evolution of patients in charge, and providing estimates on the number of visits required by the patients during their stay. These estimates can help HC providers in assigning the workload to their operators, in accepting new customers, and in optimally planning their human resources; in current practice, the workload is generally assigned without considering the variability of the patient requests in their care pathway; thus, HC plans are vulnerable to disruptive events (Borsani *et al*. 2006). The proposed patient model consists of two components: a care pathway model and a cost model. The pathway model describes the patient stochastic evolution over time and the cost model then describes the necessary resources for the patient as a function of his particular care pathway.

Model development closely depends on the availability of data, as shown in the literature (Taylor and McClean 1996, Taylor *et al*. 1997, Bennett *et al*. 1997, McClean *et al*. 1998, McClean and Millard 1998, Flessa 1999, Bergamaschi *et al*. 2000, Taylor *et al*. 2000, Congdon 2001, Verbeek *et al*. 2001, Pauler and Finkelstein 2002, Marshall *et al*. 2002, Emparan *et al*. 2003, Remák *et al*. 2003, Kousignian *et al*. 2003, Krahn *et al*. 2004, Berzuini and Allemani 2004, Marshall and McClean 2004, Husted *et al*. 2005, Magherini *et al*. 2005, Marshall *et al*. 2005, Verotta 2005, Alagoz *et al*. 2005, Koizumi *et al*. 2005, Altman and Petkau 2005, McClean and Millard 2007). In fact, historical data analysis is often made difficult by partial lack of information (censored data) and by the high variability in the observed patients' clinical behaviour. In front of these limitations this article also shows that, for operation planning purposes, the stochastic evolution of patients can be captured by models that are easily applied. By taking into account the standard data available in HC organisations, it is possible to avoid the details of clinical and psycho-social patient conditions (not always economically available at regular intervals) for obtaining accurate workload estimates.

In light of the evidence that only a small percentage of the proposed models are implemented in practice, the implementation of the model in real situations is crucially important (Kellogg and Walczak 2007). The stochastic model presented was therefore applied to a real case. Historical data for one of the largest public Italian HC providers (Dipartimento Interaziendale della Fragilità of the Azienda Sanitaria Locale (ASL) and the Azienda Ospedaliera (AO) di Lecco, hereafter called ASL Lecco) were used to develop and validate the model. Moreover, a software application containing the model is currently used by the provider to assign workloads to the operators, taking into account the variability in patient requests. This HC provider is representative of a general class of providers (Chahed *et al*. 2006, 2009) in terms of organisation and resource planning; therefore, the proposed approach can be considered general and applicable to other similar structures.

2. Stochastic model of the patient requested workload

In this section, the general structure of the model is explained. After a literature review (Section 2.1), the development of the stochastic model concerning the patients health progression is described (Section 2.2).

2.1. *Related literature*

In the literature, various stochastic models were developed to quantify the resources requested by a patient as a function of his conditions, as well as to make plans for health services.

Bayesian techniques were used to predict patient traffic from their home to the hospital, in order to facilitate reconfigurations of the emergency hospital services (Congdon 2001, Marshall *et al*. 2002, 2005). Marshall *et al*. analysed different models to evaluate patient traffic (Marshall *et al*. 2005). Markovian models were used to study the hospitalisation of geriatric patients (Taylor and McClean 1996, Taylor *et al*. 1997, McClean *et al*. 1998, McClean and Millard 1998, Taylor *et al*. 2000, McClean and Millard 2007) or the natural history of hepatitis-C, in order to determine the allocation of compensatory funds to patients who acquired the pathology through blood transfusions (Krahn *et al*. 2004). Queuing modelling was used by Koizumi *et al*. in order to analyse the congestion of mental health facilities (Koizumi *et al*. 2005). The Coxian distribution was used by Marshall and McClean in order to identify patient characteristics for predicting the duration of hospitalisation (Marshall and McClean 2004). Gardiner *et al*. combined a Markov chain with a mixed-effects model to estimate the net present value of medical care costs over a finite period of time as a function of patient characteristics (Gardiner *et al*. 2006).

Other models studied the evolution of certain pathologies. Bayesian models were proposed for forecasting the disease progression in prostate cancer recurrence (Pauler and Finkelstein 2002), in multiple sclerosis (Bergamaschi *et al*. 2000), in end-stage liver disease (Alagoz *et al*. 2005) or in immunodeficiency virus (Berzuini and Allemani 2004, Verotta 2005). Markovian models were also adopted in order to study frontal lobe patients (Magherini *et al*. 2005), breast cancer patients (Verbeek *et al*. 2001), HIV infected patients (Kousignian *et al*. 2003), patients affected by psoriatic arthritis (Husted *et al*. 2005) and by multiple sclerosis lesions (Altman and Petkau 2005). Markov chain modelling was also used to estimate the cost-effectiveness of specific therapies in patients affected by particular pathologies (Bennett *et al*. 1997, Emparan *et al*. 2003, Remák *et al*. 2003). Flessa used simulation to model the malaria epidemiology for supporting decision-makers of malaria-control programmes from both a clinical perspective and a management perspective (Flessa 1999).

2.2. *Patient model*

The proposed model integrates a care pathway model and a cost model, in which the cost is expressed in terms of number of requested visits. The model provides estimates on the main variables of interest for an HC provider: the patient care duration (CD), the number of cared patients and the number of requested visits within a time frame. The input variables for the model are the patient conditions, upon their admission into the HC facility and during the execution of their care pathway and care delivery information, e.g. patient type and the number and typology of provided visits along the care pathway execution.

We propose a multistep approach to develop the patient stochastic model (Figure 1):

(1) Identification of state variables and patient classes: the variables having a significant impact on the CD and on the number of visits are assumed as the model state variables. These can be identified by using statistical techniques on historical data and by interviewing nurses and physicians working within the HC organisation. Furthermore, a classification based on the initial patient conditions is necessary to simplify and to apply the approach.

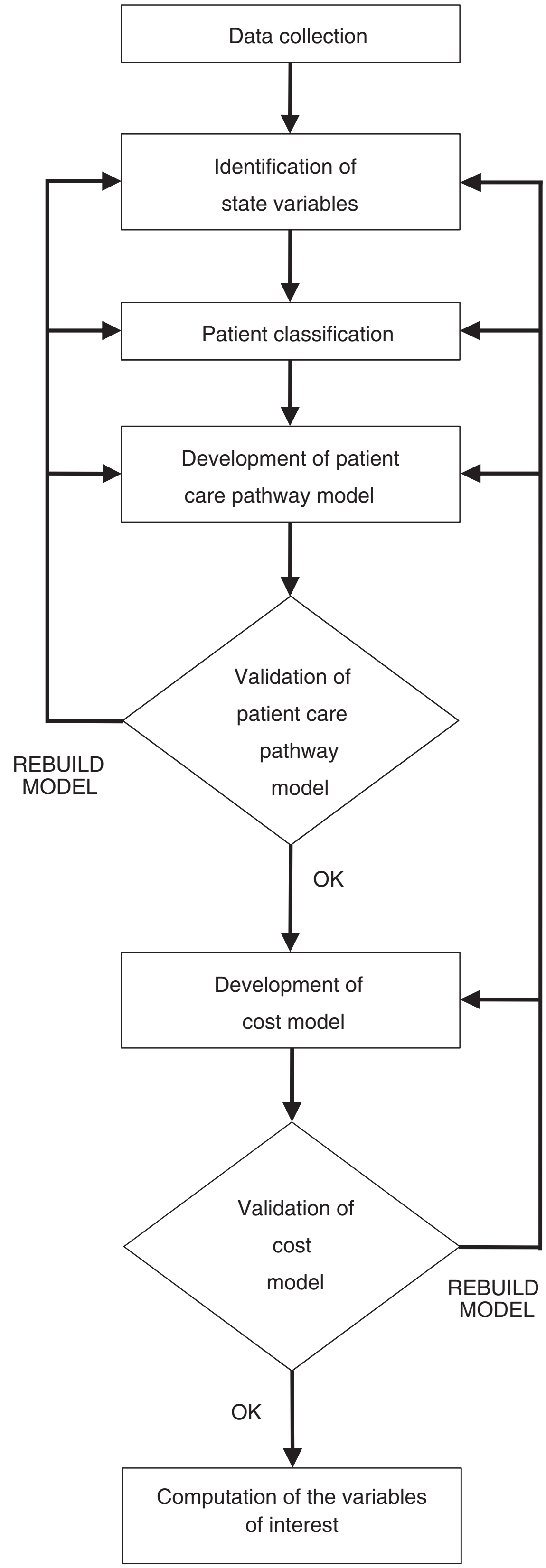

Figure 1. Flow chart of the proposed multistep approach for developing the patient stochastic model.

(2) Development of the patient care pathway model: the patient evolution is described with the sequence of values assumed by the state variables. The probability of a change of state in discrete time is estimated from historical data for each patient class.
(3) Validation of the pathway model: the pathway model is cross-validated analysing the CD and the number of cared patients over time.
(4) Development of the cost model: an empirical probability distribution of the cost is fitted from historical data and is associated with each possible patient state. This cost is the number of visits requested by the patient in fixed number of days.
(5) Validation of the entire model: the entire model (patient pathway and cost) is cross-validated by analysing the number of requested visits by a patient during his care pathway.

These steps are described more fully in the remainder of this section.

2.2.1. *Identification of model state variables and patient classes*

Patient conditions can be described along with the value assumed by a set of representative state variables (e.g. the pathology, the age and the presence of a caregiver), which have to be observable and measurable at low cost throughout the patient care pathway. Consequently, the care pathway can be described as a sequence of states.

Statistical analysis on the measurable input variables must be performed in order to identify the significant variables contributing to the CD and the number of delivered visits. Our experience suggests the adoption of non-parametric statistical tests. Moreover, health specialists working within the HC structure can corroborate the definition of the model assumptions and the interpretation of the statistical analysis. The resulting significant variables $x_1, x_2, \ldots, x_n$ are the assumed state variables of the model and are included as the components of the model state vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]$. If some state variables are continuous, they are discretised in a finite number of intervals. In addition to the values of $\boldsymbol{x}$, which relate to the cared patients, two additional values are introduced: a state *End*, which describes patients who have been discharged permanently (as a consequence of death or definitive recovery) and a state *Pause*, which describes the patients temporary discharged (as a consequence of events such as a hospitalisation period). The possible values of $\boldsymbol{x}$ (values related to the cared patients, *End*, *Pause*) define the state space $\boldsymbol{\Omega}$.

Moreover, patients admitted to the HC service present different initial characteristics in terms of pathology, age, domestic context and other clinical, psychological and social conditions. Consequently, they have different care pathway evolution and, even if two different patients are in the same state, they can have different future evolution depending on their personal characteristics. However, patients having some similarities in their care pathway can be grouped into classes so that the patients belonging to the same class are described by the same stochastic evolution model. This classification must be based only on the initial patient characteristics, which is the only information available when the patient is admitted.

Since it is not possible in practice to consider all the patient characteristics, some simplifications have to be introduced in order to reduce the amount of classes or groups. In practice, the patient classification is usually simplified by service refunding. Indeed, most national and regional policies require that HC providers classify the patients into classes depending on the costs incurred in delivering the service. Moreover, the providers generally use a more detailed profile classification than what is requested by the public authorities to better control the use of resources. Thus, it is important to assess whether the adopted classification represents an appropriate choice for the model development.

2.2.2. *Patient care pathway model*

The sequence of the patient states $\boldsymbol{x}(k)$ (with $k = 1, 2, \ldots$) along his care pathway is the result of a stochastic process $\boldsymbol{X}(k)$. In this section, the development of a model to describe the random vector $\boldsymbol{X}(k)$ is presented.

The care pathway model is modelled by using a Markov chain in which the states are the possible values of the state vector $\boldsymbol{x} \in \boldsymbol{\Omega}$. These states and the associated costs are the same for all the patients, whereas the transition matrixes are differentiated and based on the patient classification. Thus, a specific transition probability matrix is obtained for each class. Markov chain modelling has been selected for its simplicity and low cost in applying the model. Simulation can also be used; however, this technique necessitates expensive data analysis that is not often counterbalanced by better results.

Beginning from the initial value assigned, when a patient is classified and admitted into the service, the value of the state vector $\boldsymbol{x}$ may change during the care pathway, depending on the class. The Therapeutic

Project (ThP) for each patient is periodically revised (usually every month), but changes depending on the patient conditions can occur each day when a modification of the care is necessary. Starting from the initial value $\boldsymbol{x}(1)$, the patient state vector $\boldsymbol{x}$ is assumed to evolve in discrete time from day k to day $k+1$, according to the transition probability matrix $\boldsymbol{P}=[p_{ij}]$.

According to the frequentistic approach, the transition probability matrix of each class is obtained from historical data for the care pathways as follows:

$$p_{ij} = \begin{cases} \frac{d_i - \sum_{j \neq i} n_{ij}}{d_i} & i = j;\quad i \neq 0 \\ \frac{n_{ij}}{d_i} & i \neq j;\quad i \neq 0 \end{cases} \tag{1}$$

where p_{ij} is the transition probability from state $\boldsymbol{x_i}$ to state $\boldsymbol{x_j}$, n_{ij} is the total number of transitions from state $\boldsymbol{x_i}$ to state $\boldsymbol{x_j}$ and d_i is the total number of days spent by all the patients of the class in the state $\boldsymbol{x_i}$. In contrast, the state *End* (named $\boldsymbol{x_0}$) is defined as an absorbent state.

Consequently, the state probability vector $\boldsymbol{\pi}(k)=[\pi_i(k)]$, which contains the probabilities of state $\boldsymbol{x_i}$ at day k, can be obtained starting from the initial value $\boldsymbol{\pi}(1)$.

The proposed Markov chain is homogeneous, i.e. the transition probabilities are assumed to be invariant with the time; this assumption has already been reported in the literature (McClean *et al.* 1998). A further assumption of Markov chain modelling is the geometric distribution of sojourn times in each state. The accuracy of these assumptions will be verified during the validation process section.

2.2.3. *Cost model*

A cost in terms of the number of visits N_i requested by the patient in a fixed period is ascribed to each patient state $\boldsymbol{x_i}$. For each sojourn in state $\boldsymbol{x_i}$, N_i is defined as the ratio between the number of visits performed during the sojourn and the number of days spent in state $\boldsymbol{x_i}$ scaled to the reference period. The empirical distribution of N_i is defined from historical data by adopting a frequentistic approach. For each state $\boldsymbol{x_i}$, the empirical distribution is defined as follows:

$$P\,[a < N_i \le b] = \frac{s_{i(ab]}}{s_i} \tag{2}$$

where $s_{i(ab]}$ is the total number of days spent by the patients of all the classes in state $\boldsymbol{x_i}$ with $a < N_i \le b$ and s_i is the total number of days spent by all the patients of all the classes in state $\boldsymbol{x_i}$. Thus, a uniform distribution of visits along the sojourn in each state is assumed. Conversely, state *End* and *Pause* are characterised by zero visits.

Moreover, it could be necessary to differentiate N_i based on the day the care was given. Historical N_i values are plotted as a function of the day of care for each possible state vector $\boldsymbol{x_i}$. The values are grouped when a clear difference can be observed, and non-parametric statistical analysis is used to assess the significance of the group on N_i. Therefore, a specific distribution of N_i is considered for each significant group. The notation $P[a \le N_i < b|\alpha_i;\beta_i]$ will be adopted to indicate that N_i refers to a day of care $k \in [\alpha_i;\beta_i)$.

In addition to the total number of visits, the same procedure can also be used to determine the empirical distribution of visits requested from a specific category of operator (nurse, physician, etc.). In this case, distributions are extracted from historical data by only considering the visits performed by a certain category of operator.

2.2.4. *Computation of the variables of interest*

The state probability vector $\boldsymbol{\pi}(k)$ allows for estimating the CD and the number of patients in charge (n). For each patient class (the index referring to the class is omitted for a more fluent reading), the probability that CD, including temporary interruptions of care, is equal to or less than k days is calculated as follows:

$$P[CD \le k] = \pi_0(k+1). \tag{3}$$

An implicit assumption is that $\pi_0(1)=0$. The expected value and standard deviation of CD are then calculated from the above distribution.

For each patient class, the expected number of patients $n(k)$ still undergoing HC treatment after a number k of days, not including new arrivals of patients, is calculated as follows:

$$n(k) = n_0 \cdot (1 - P[\mathrm{CD} \le k-1]) \tag{4}$$

where n_0 is the initial number of patients within the examined class. The total number of patients in the care of the HC provider is the sum of $n(k)$ over all the different patient classes.

The state probability vector $\boldsymbol{\pi}(k)$ and the cost distributions together allow for an estimate of the number of visits requested by a patient during his care pathway. The probability that the number of visits V_{pat} (required by a patient in the reference period) in the day of care k assumes a value between a and b is:

$$P[a < V_{\text{pat}} \le b\,|\,k] = \sum_{i \in \Omega} \left\{\pi_i(k) \cdot P[a < N_i \le b\,|\,\alpha_i;\,\beta_i]\right\} \tag{5}$$

with $k \in [\alpha_i; \beta_i)$ for each state $\boldsymbol{x_i}$. The expected value and the standard deviation of V_{pat} are then calculated from the above distribution.

The total number of required visits can be used to support resource planning in the short and medium terms. Indeed, the number of visits, along with the time necessary for providing the visit at patient home, gives information about the workload required for patient care and, when considering all of the patients, for the entire structure.

Once the set of assisted patients is defined, the proposed model also allows for predicting the individual operator's workload. After the assignment of a new patient to an operator, the distribution of the number of visits V_{op} of the operator is determined by the convolution product between the patient workload distribution V_{pat} and the operator workload distribution prior to the new assignment. From these estimates, it is possible to develop algorithms that optimally manage human resources in HC organisations.

2.2.5. *Validation*

Validation is necessary in order to verify the adherence of the model to the described situation. Particularly, it verifies the consistency of the hypothesis at the basis of the model.

CD, $n(k)$ and V_{pat} are analysed in the validation process. The first day of care is considered as the initial day of observation ($k = 1$) for each patient.

Model random sub-sampling cross-validation is performed by randomly splitting the historical data of patients into two groups several times (Brusco and Steinley 2009). For each split, data from the training set (Set A) are used to develop the model, both transition matrixes and cost distributions, while data from the testing set (Set B) are used to verify it. The results from each split are then averaged. The advantages of this method are that the proportion between the training and the validation set does not depend on the number of random splits (as is the case in the K-fold cross-validation); furthermore, not as many splits are necessary as it happens in the leave-one-out cross-validation.

The first validation only concerns the care pathway model and does not consider the cost distributions. This validation allows for quantifying the impact of the approximations introduced by the use of a homogeneous Markov chain. $P[\text{CD} = k]$, $P[\text{CD} \leq k]$ and $n(k)$ are computed for each patient class and are compared between Sets A and B. In the training Set A, they are determined by the method reported in Equations (3) and (4); in the testing Set B, they are directly calculated from the historical data. This approach only allows validating the transition to state *End*; nevertheless, other transitions cannot be evaluated with the available data without including the associated cost model. If this validation is unsuccessful, it is necessary to revise either the Markov chain or the classification.

A second validation concerns the overall model, including cost distributions. The number of visits V_{pat} required by a patient in the day k of care is computed for each patient class and compared between Sets A and B. In the training Set A, V_{pat} is determined by the method reported in Equation (5); in the testing Set B, it is directly calculated from historical data. If the CD is lower than k, V_{pat} at day k is assumed to be null. If this validation is unsuccessful, it is necessary to revise either the cost model or the entire care pathway model.

3. Model application to a real case

The proposed model was applied to a real case, considering the historical data of one of the largest public Italian HC providers (ASL Lecco).

3.1. *Characteristics of the HC provider*

In the analysed provider, the care pathway of a patient includes admission, care supply and discharge. The admission phase consists of the following:

- *Preliminary assessment*: during the first visit, an operator, generally a nurse, collects personal data and other information regarding the patient's clinical, functional and social conditions.
- *Multidimensional assessment*: a multidisciplinary team evaluates the patient's conditions and tests his functional abilities. The most common testing methods are the Karnofsky Performance Status (KPS) Scale (Mor *et al.* 1984), Activities of Daily Living (ADL) Scale (Katz *et al.* 1970), Instrumental Activities of Daily Living (IADL) Scale (Lawton and Brody 1969) and the Global Evaluation Functional Index (GEFI) (Cucinotta *et al.* 1989).

The assessment defines the patient needs on which the service will be designed and provided (Chahed *et al.* 2006, 2009). The service supply starts after the multidisciplinary team develops the ThP, which includes detailed information about the kind and the frequency of the required visits and all other operational activities executed by HC operators at the patient's home (Asquer *et al.* 2007). During the care supply phase, the patient receives the service from the HC provider as

Table 1. Classification of CPs according to ASL Lecco. The assigned number of each CP adopted in this study is also reported.

Types of care	Assigned number to the CPs
Extemporary Care: *provided to patients who need care with a very low frequency of visits*	1, 15
Integrated Home Care: *characterised by a medium–high care intensity* (*CP are ordered for increasing values of expected number of visits during the ThP duration*)	10, 9, 2, 3, 4, 5, 12, 13, 14
Palliative Care: *offered to terminal patients generally affected by oncological diseases* (*CP are ordered for increasing values of expected number of visits during the ThP duration*)	8, 7, 6

prescribed by his ThP. The ThP is periodically assessed in order to check its adequacy for the patient needs, which may change depending on clinical conditions, as well as the social environment within his home. The revision period is usually 1 month: at each periodic revision, the ThP can be reconfirmed or modified to address the new needs. Finally, the patient is discharged when he recovers, he needs a different kind of service (such as hospitalisation), or he dies.

The analysed provider assigns a category, called the care profile (CP), to each patient on the basis of the costs estimated from his ThP. Indeed, once a ThP is defined, a CP is assigned to the patient based on the type and the number of required visits, which are the most significant factors affecting costs. The patient's CP may change with time according to his ThP modifications.

The provider includes 14 CPs that are related to as many cost levels and are grouped into three main categories of service (Table 1). CPs related to the Palliative Care refer to a homogeneous class of terminal patients whose pathology is in a terminal state. In the other cases, each CP includes a large range of patients in terms of age, pathology and social context. However, patients belonging to the same class are characterised by similar ThPs.

The organisation of the analysed provider started in 2003 and continues today with a very few changes. Nevertheless, 2003 was a transition year. This is because the provider was changing patient evaluation standards for the admission. Thus, in order to obtain homogeneous historical data for model development, only data representing patients admitted since January 2004 and discharged before the data collection (April 2008) were taken into account for carrying out the model. Moreover, patients whose data were incomplete or contained errors were excluded. Therefore, the analysed sample included 7277 patients.

3.2. *Identification of state variables and patient classification*

ASL Lecco, like the other HC providers, collects a large amount of data related to each patient care pathway that considers both the patient's personal information and the provided service. In particular, first pathology, clinical state, functional and social patient conditions are stored; moreover, the series of CPs and the type, the duration and the number of provided visits are also stored in the provider database.

The multidimensional indexes (e.g. KPS, ADL and IADL) could constitute the state vector $\boldsymbol{x}$, since they represent the patient conditions from the clinical, functional and psycho-social points of view. However, the multidimensional assessment is provided only at the beginning of the ThP to reduce costs and consequently it cannot be used to describe the patient evolution. In contrast, the CP, together with the number of visits, is continuously updated. Specialists of the HC structure suggested that the patient pathology and the caregiver availability are the most representative information reflecting the patient's social and clinical conditions. For this reason, the potentially significant variables relevant to the patient care pathway are the age, the pathology, the First CP (FCP), the indexes (KPS, ADL, IADL, GEFI) and the caregiver availability at the patient's home.

In order to identify which of these data constitute the state vector $\boldsymbol{x}$, a statistical analysis was performed to identify the patients' significant information that influences the total number of visits and the CD. Only the FCP, the pathology and the KPS index were statistically significant (null P-value). Since HC specialists define the ThP and the CP by taking into account the clinical conditions and the multidimensional assessment, the assignment of a CP to a patient is strictly related to his pathology and KPS index. These observations suggested to consider the CP as the only state variable of the Markov chain that represents the patients during their care pathway. Thus, the state vector is reduced to a scalar, which is the CP.

Therefore, the state space Ω includes the 14 CPs, plus the *End* and *Pause* states.

In this real case, the set of CPs was also used to build the patient classification, according to the provider practice. Particularly, the classification is based only on the initial patient characteristics and the patients were consequently classified according to their FCP. Thus, 14 possible classes are included in the model (Table 1).

3.3. *Markov chain*

The Markov chain of each patient class includes 16 states that correspond to possible values of $x \in \Omega$. While the states are the same for each class, the transition probability matrixes are different: 14 transition probability matrixes are computed, each one corresponding to a specific FCP (Table 2).

In this real case, all of the transition probabilities do not seem to depend on the day of care, and the transition matrixes were assumed to be homogeneous. In order to assess this assumption, the initial matrixes (obtained only from data for the first 60 days of care) were also computed. The similarity between the homogeneous and the initial matrixes was then evaluated. The homogeneous and the corresponding initial matrix did not significantly differ in all of the classes.

It is important to notice that the homogeneity assumption only refers to the transitions from a CP to another and not to the number of visits required by the patients.

3.4. *Cost model (weekly requested visits)*

The week was considered as the reference period to calculate the number of requested visits N_i and V_{pat}. In fact, the planning process in the analysed provider is carried out over a weekly horizon. Therefore, it is useful to consider visits on a weekly basis.

Even if each CP is characterised by a nominal range of weekly requested visits (according to HC provider protocols), there is high variability within this range. Thus, the entire empirical distribution of the weekly number of requested visits was constructed to include this variability.

Each distribution of N_i (Equation (2)) was discretised in 81 intervals, each corresponding to 0.25 visits, so that the first two intervals are $N_i = 0$ and $0 < N_i \leq 0.25$, while the last two are $19.5 < N_i \leq 19.75$ and $N_i > 19.75$. The N_i associated with each interval was calculated with respect to the upper bound of the interval (20 for the last one). Nevertheless, the results do not significantly change even if other interval values are considered.

The described statistical analysis was performed to differentiate N_i on the basis of the day of care (Table 3).

3.5. *Model validation*

Six FCPs were the most important in terms of total visits provided to all of the patients belonging to these classes, which cover about 86% of the total workload of the analysed provider (Table 4). These six FCPs include 5973 patients, while the remaining eight FCPs do not include enough data to obtain a reliable model. Nevertheless, each of the eight remaining FCPs represents a very small workload percentage for the provider so that a certain error does not significantly affect the overall workload prevision. Therefore, the validation analysis is reported relative to the six most important FCPs in terms of workload (Table 4). Validation outcomes are reported for all of these six FCPs, while detailed results are reported only for the most important FCP of each type of care. The number of visits requested from all the types of operator was studied and 10 random splits were carried out.

As for the Markov chain, comparisons between training Set A and testing Set B are reported (Figures 2–4). The results showed good superposition between the model outcomes of Set A and the real data of Set B: the distributions and the general decreasing trend of $n(k)$ are well reproduced. Only the local oscillations of the trends between two close days are not captured. As a matter of fact, the absence of memory of the Markov chain imposes results in an exponential decreasing trend that cannot describe the oscillations.

A large deviation was found at the beginning of the care pathway for FCP = 1. This FCP includes a large number of pathologies and clinical conditions which are similar only in terms of care pathway, while other patient classes include very similar patients (e.g. Palliative Care classes). A source for this error is that some providers (including ASL Lecco) develop the ThP in two phases (Asquer *et al.* 2007): a preliminary ThP is first established for a testing period, usually 1 month, during which the provider has the possibility to better customise the ThP based on emerging needs. A more accurate profile assignment would avoid this testing period, improve the patient classification and allow for a more reliable model description. Another source could be associated with the absence of memory of Markov chain modelling, thus reflecting a high initial deviation from the exponential decreasing trend

Table 2. Transition probability matrixes of the most important FCPs (in terms of workload (Table 4)) for each ASL Lecco type of care (Table 1). CP = 0 refers to the *End* state and CP = 11 to the *Pause* state.

FCP = 1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0.0059	0.9913	1×10^{-4}	0.0002	0.0001	0.0002	0.0003	7×10^{-6}	1×10^{-5}	0.0002	1×10^{-5}	0.0016	7×10^{-6}	2×10^{-5}	0	1×10^{-5}
2	0.0054	0.0043	0.9779	0.0008	0.0008	0.0016	0.0004	0.0004	0	0.0043	0	0.0039	0	0	0.0004	0
3	0.0064	0.0041	0.0012	0.9782	0.0006	0.0012	0.0003	0	0	0.0052	0	0.0029	0	0	0	0
4	0.0059	0.0016	0.0023	0.0039	0.9793	0.002	0	0	0	0.0031	0	0.0016	0.0004	0	0	0
5	0.0097	0.0008	0.0008	0.0027	0.0019	0.9763	0.0024	0	0	0.0024	0	0.0027	0	0.0003	0	0
6	0.0236	0	0	0	0	0	0.9733	0.0009	0.0006	0	0	0.0017	0	0	0	0
7	0.0042	0	0	0	0	0	0.0052	0.9864	0.001	0	0	0.0031	0	0	0	0
8	0.0016	0.0033	0	0	0	0	0.0016	0.0033	0.9901	0	0	0	0	0	0	0
9	0.0066	0.0016	0.0004	0.0002	0.0005	0.0001	0.0003	0	0	0.9875	0	0.0026	0.0001	0	0	0
10	0.0144	0	0.0048	0	0	0	0	0	0	0	0.9808	0	0	0	0	0
11	0	0.0022	0.0004	0.0004	0.0003	0.0008	0.0006	0.0001	2×10^{-5}	0.0006	4×10^{-5}	0.9943	7×10^{-5}	2×10^{-5}	2×10^{-5}	0.0001
12	0.0017	0	0	0	0	0.0017	0	0	0	0	0	0.0026	0.9922	0.0009	0.0009	0
13	0.0093	0.0031	0	0	0	0	0	0	0	0.0031	0	0	0.0062	0.9782	0	0
14	0.0079	0	0	0	0	0	0	0	0	0	0	0	0	0.004	0.9881	0
15	0.0051	0	0	0	0	0	0	0	0	0	0	0.0051	0	0	0	0.9898

FCP = 5	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0.007	0.9898	0.0002	0.0002	0	0.0003	0.0002	0	0	1×10^{-4}	1×10^{-4}	0.002	1×10^{-4}	0	0	1×10^{-4}
2	0.0084	0.0004	0.982	0.0003	0.0004	0.0001	0.0004	0	0	0.0036	0.0005	0.0038	0	0	0	0
3	0.008	0.0003	0.0009	0.9766	0.0016	0.0005	0	0	0	0.0085	0.0001	0.0033	0	0.0001	0	0
4	0.0048	0.0007	0.0018	0.0046	0.9773	0.0014	0.0003	0	0	0.0059	0.0002	0.0031	0	0	0	0
5	0.0069	0.0004	0.0011	0.0021	0.0026	0.979	0.0006	2×10^{-5}	0	0.0032	3×10^{-5}	0.0035	0.0002	0.0001	0.0001	3×10^{-5}
6	0.0239	0	0	0	0	0	0.9733	0.0016	0	0	0	0.0013	0	0	0	0
7	0.0062	0	0	0	0	0	0.0031	0.9846	0.0015	0	0	0.0031	0	0	0	0.0015
8	0	0	0	0	0	0	0.0227	0	0.9318	0	0	0.0455	0	0	0	0
9	0.01	0.0005	0.0003	0.0007	0.0001	0.0002	0.0002	4×10^{-5}	0	0.9836	0.0003	0.004	7×10^{-5}	0	4×10^{-5}	0
10	0.0031	0.001	0.0003	0	0	0.0003	0.0007	0	0	0.0003	0.9917	0.0021	0.0003	0	0	0
11	0	0.0007	0.0003	0.0004	0.0007	0.0021	0.0005	4×10^{-5}	3×10^{-5}	0.0013	5×10^{-5}	0.9936	8×10^{-5}	7×10^{-5}	7×10^{-5}	0.0001
12	0.0083	0.0005	0	0	0	0.0005	0	0	0	0	0.001	0.0052	0.9828	0	0.0016	0
13	0.0157	0	0	0	0	0.0031	0	0	0	0.0031	0	0.0063	0.0126	0.9591	0	0
14	0.008	0	0	0	0	0.0011	0	0	0	0.0034	0	0.0023	0.0046	0	0.9805	0
15	0.0095	0	0	0	0	0	0	0	0	0	0	0.0048	0	0	0	0.9857

(*continued*)

Table 2. Continued

FCP = 6	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
2	0	0	0.9889	0	0.0111	0	0	0	0	0	0	0	0	0	0	0
3	0	0.0333	0	0.9667	0	0	0	0	0	0	0	0	0	0	0	0
4	0.0036	0	0	0	0.9964	0	0	0	0	0	0	0	0	0	0	0
5	0.009	0	0.009	0	0	0.964	0	0	0	0	0	0.018	0	0	0	0
6	0.0244	0	0	2E-05	0	0	0.9708	0.0027	0.0002	0	0	0.002	0	0	0	0
7	0.0041	0	0	0	0	0	0.0045	0.9879	0.0014	0	0	0.0021	0	0	0	0
8	0.0031	0	0	0	0	0	0.0009	0.0012	0.9924	0.0002	0	0.0022	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0.9848	0	0.0152	0	0	0	0
10	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625
11	0	0	0	0	0	0.0005	0.0172	0.0049	0.0014	0	0	0.9759	0	0	0	0
12	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625
13	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625
14	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625
15	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625

Table 3. Total number of visits (average value ± SD) as a function of the CP and the day of care.

CP	Day of care	Number of visits (N_i) Mean	SD
1	1-30	1.345	1.365
	31-inf	0.536	0.639
2	1-30	2.826	1.116
	31-240	2.285	1.014
	241-inf	2.126	0.963
3	1-inf	2.561	1.158
4	1-inf	2.755	1.382
5	1-30	4.291	2.237
	31-90	4.086	2.322
	91-330	4.664	2.290
	331-inf	5.439	2.497
6	1-30	6.509	3.013
	31-inf	5.971	2.913
7	1-30	3.531	1.499
	31-inf	2.969	1.648
8	1-inf	1.482	1.093
9	1-30	1.935	1.099
	31-180	1.539	0.864
	181-inf	1.363	0.891
10	1-30	1.277	1.098
	31-inf	0.590	0.744
12	1-30	2.347	1.108
	31-240	1.923	1.101
	241-inf	1.547	0.974
13	1-inf	2.810	1.185
14	1-inf	2.707	1.388
15	1-30	0.921	1.188
	31-60	1.354	1.265
	61-inf	0.864	0.915

Table 4. Historical data of workload (number of provided visits) for patients belonging to each FCP, determined as the product of the number of patients and the average number of visits, both obtained from historical data. Bold values refer to the six most important FCPs.

FCP	Number of patients N_p	Average number of visits N_v	Average workload $(N_p \cdot N_v)$	Workload percentage (%)
1	**1156**	**21.85**	**25258**	**10.3**
2	*348*	*46.71*	*16255*	*6.6*
3	**539**	**38.87**	**20950**	**8.5**
4	**568**	**36.07**	**20488**	**8.4**
5	**1081**	**52.20**	**56423**	**23.0**
6	**1588**	**38.33**	**60865**	**24.8**
7	*134*	*50.18*	*6724*	*2.7*
8	*25*	*58.52*	*1463*	*0.6*
9	**1041**	**25.56**	**26609**	**10.8**
10	*15*	*36.73*	*551*	*0.2*
12	*69*	*31.09*	*2145*	*0.9*
13	*47*	*32.43*	*1524*	*0.6*
14	*25*	*28.96*	*724*	*0.3*
15	*641*	*8.41*	*5393*	*2.2*
TOT	7277		245372	100

of $n(k)$ in Set B, or to the homogeneity assumption that seems unsuitable for this particular FCP.

Considering all the splits, the Error (Err), Absolute Error (AbsErr) and Root Mean Squared Error (RMSE) of the CD mean value, standard deviation and skewness were calculated (Table 5) for the six most important FCPs (Table 4), according to Equations (6)–(8).

$$\mathrm{Err} = \frac{1}{10} \cdot \sum_{\text{division}} (\mathrm{value}_{\text{Set A}} - \mathrm{value}_{\text{Set B}}) \tag{6}$$

$$\mathrm{AbsErr} = \frac{1}{10} \cdot \sum_{\text{division}} |\mathrm{value}_{\text{Set A}} - \mathrm{value}_{\text{Set B}}| \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{10} \cdot \sum_{\text{division}} (\mathrm{value}_{\text{Set A}} - \mathrm{value}_{\text{Set B}})^2} \tag{8}$$

The CD mean value was reproduced by the model with small errors for the six FCPs analysed. The mean value was often underestimated except for FCP = 5, which presented a slight overestimation. Moreover, a low RMSE indicated a reliable superposition between the Markov chain outcomes and the provider historical data.

With regard to the CD standard deviation, the Markov chain captured the high variability of the data, which is reflected in the high standard deviation of Set B. Nevertheless, the standard deviation was always lower in Set A. In four cases, Err and AbsErr differed only in sign, underlining that the standard deviation of Set A was lower than in Set B in all of the random splits. This is related to the local oscillations between two near days that were not captured by the model (Figure 2). Furthermore, the CD skewness was well reproduced without a clear overestimation or underestimation.

With regard to the validation of the entire model, including the cost model of the patient number of visits, comparisons between training Set A and testing Set B are reported (Figures 5–7). The results showed good superposition between Sets A and B. Particularly, the cumulative distribution had a good superposition in all six FCPs analysed.

The differentiation of N_i over time (Table 3) allowed the relevant sudden variations which may occur (especially in the first months of care) to be considered, which improves the model prediction.

The scatter plots demonstrating the cumulative number of visits between training Set A and testing Set B are also reported (Figure 8). Both the data

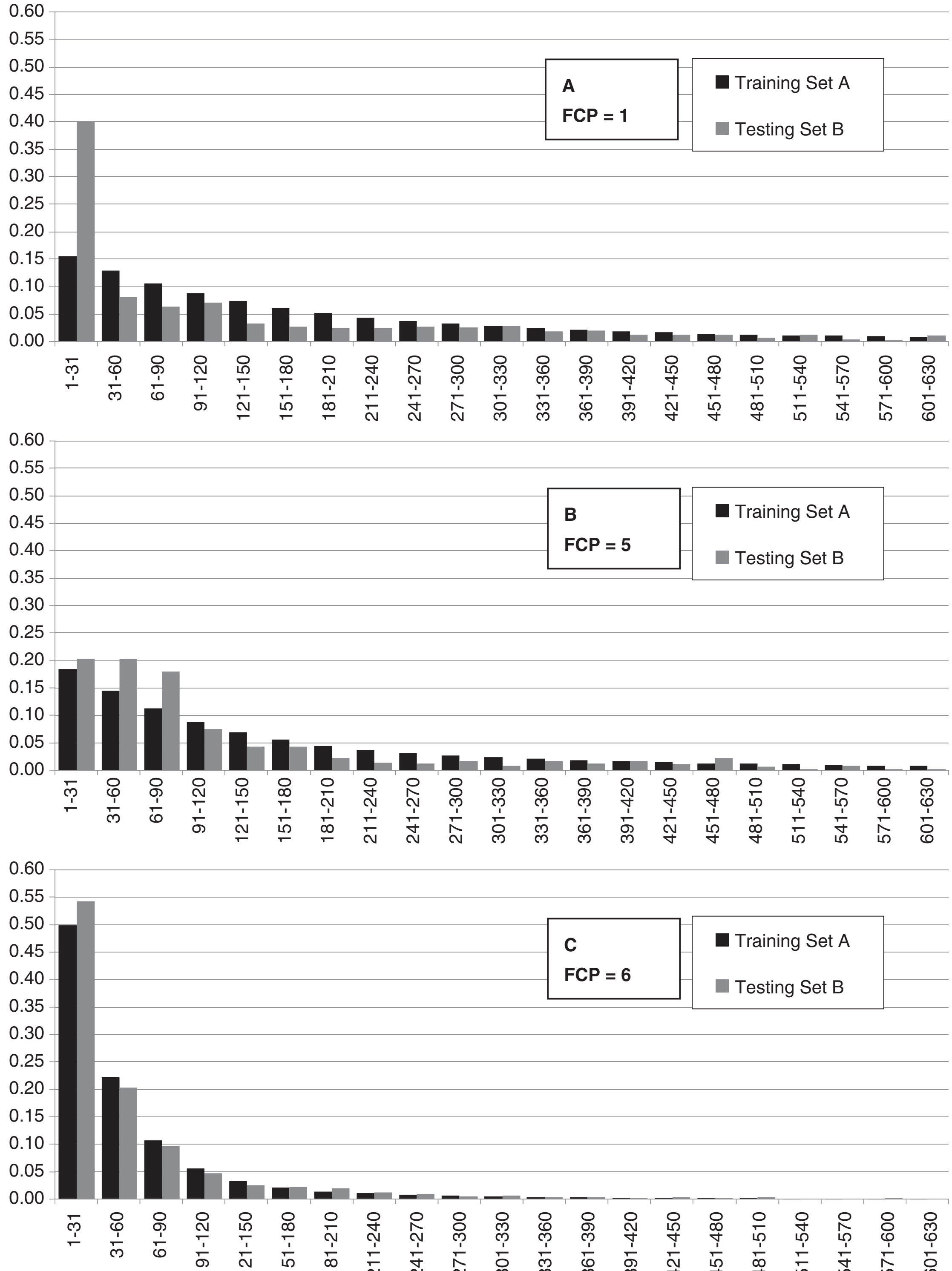

Figure 2. CD distributions relative to one of the random splits for the most important FCPs (in terms of workload (Table 4)) of each ASL Lecco type of care (Table 1). Abscissa values are grouped every 30 days for a better visualisation even if each day was considered in the model.

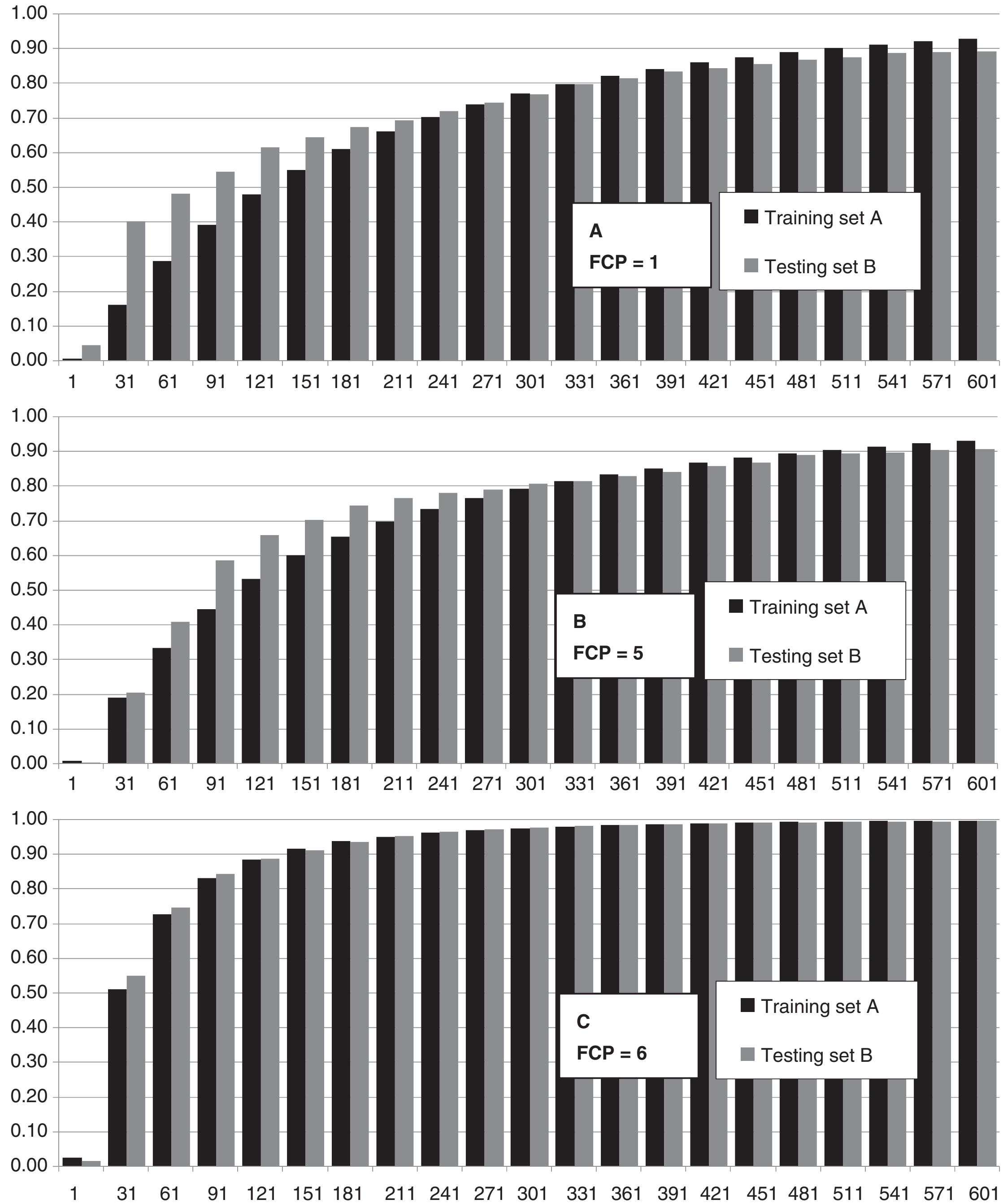

Figure 3. CD cumulative functions relative to one of the random splits for the most important FCPs (in terms of workload (Table 4)) of each ASL Lecco type of care (Table 1). Abscissa values are grouped every 30 days for a better visualisation even if each day was considered in the model.

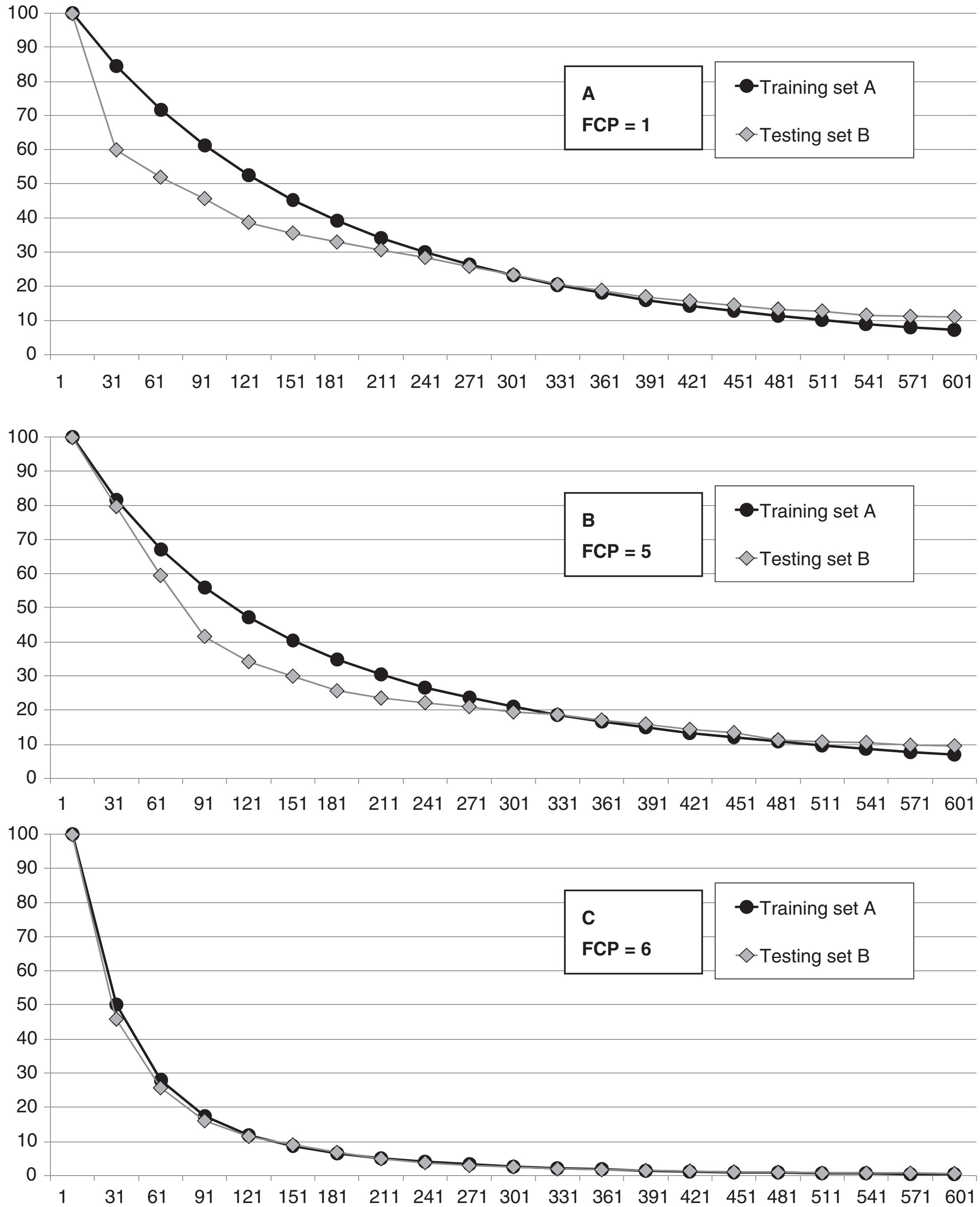

Figure 4. Number of patients $n(k)$ still in charge after k days relative to one of the random splits for the most important FCPs (in terms of workload (Table 4)) of each ASL Lecco type of care (Table 1). Abscissa values are grouped every 30 days for a better visualisation even if each day was considered in the model.

of the random splits and the average values are considered. The plots presented a cloud of points on the diagonal that shows the superposition between the model outcomes and the provider historical data.

Moreover, a comparison between the predicted and the actual number of visits, supplied by one of the three divisions of ASL Lecco, was performed for the 35 weeks after the data collection for the model

Table 5. CD mean value, SD and skewness averaged on the 10 random splits. Err (Equation (6)), AbsErr (Equation (7)) and RMSE (Equation (8)). The six most important FCPs in terms of workload (Table 4) are considered.

	FCP = 1	FCP = 3	FCP = 4	FCP = 5	FCP = 6	FCP = 9
Mean – Set A	205.26	169.32	159.81	195.36	57.48	184.04
Mean – Set B	206.98	171.99	162.39	193.16	58.49	184.98
Err	−1.71	−2.66	−2.58	2.20	−1.02	−0.94
AbsErr	15.76	21.41	15.69	14.94	2.75	14.88
RMSE	22.81	26.80	18.70	17.51	3.70	17.07
Mean SD – Set A	237.03	206.60	192.32	230.11	83.27	254.38
Mean SD – Set B	283.52	252.50	221.75	267.07	106.54	272.15
Err	−46.49	−45.90	−29.43	−36.96	−23.27	−17.77
AbsErr	46.49	48.77	29.43	36.96	23.27	21.74
RMSE	49.52	56.71	38.81	41.56	25.26	29.48
Mean skewness – Set A	2.31	2.48	2.59	2.31	4.12	2.46
Mean skewness – Set B	1.90	2.83	2.40	2.28	5.62	2.21
Err	0.41	−0.35	0.19	0.03	−1.51	0.25
AbsErr	0.41	0.50	0.25	0.15	1.51	0.25
RMSE	0.47	0.88	0.29	0.20	1.59	0.30

implementation (marked as Week 0). Each time, the number of visits in the following week was evaluated in advance by the model and compared with the real datum after it was available (Table 6). The results showed small errors during each week and for the overall number of visits.

3.6. *Discussion*

An overview of the model outputs for the six most important FCPs is reported (Table 4) in terms of the number of patients $n(k)$ still in charge after k days and the average number of visits V_{pat} requested along the time, which were extracted from a random split (Figure 9). According to the observed demands of the patients, the outcomes show the differences among the various FCPs. Palliative patients (FCP = 6) show a higher number of requested visits at the beginning of their care pathway along with shorter CDs. In contrast, extemporary care patients (FCP = 1) had the lowest number of visits within the HC period. Integrated HC patients had intermediate needs.

The results show that the hypothesis of an absence of memory connected to Markov chain modelling does not represent a problem, since there was good superposition between the model and the real data. The literature asserts that this hypothesis fits the data in the presence of terminal patients where no evolution of the pathology can be observed (Lee *et al*. 2004, Lee and Zelen 2008). According to this observation, our results showed that better superposition between Sets A and B was found in FCP = 6 (terminal patients of the Palliative Care). Furthermore, we showed that the model presented also adapts to other patient categories with limited errors that become concentrated at the beginning of the care pathway. Thus, by coupling a homogeneous Markov chain to time-dependent number of visit distributions, it was also possible to appropriately describe other patient categories in addition to terminal patients.

The main exploitation of the model is within the workload planning and distribution among the operators while preserving the continuity of care (once an operator was assigned to a patient, he follows the entire ThP of this patient) and balancing the number of requested visits among the operators.

The results of the model, in terms of distributions of CD and number of visits along the time, can be used for robust workload planning that allows for the generation of different future scenarios of visits requested by each patient for the next period. Moreover, the model allows for balancing the workload on the basis of the future expected number of visits. In contrast, if only the actual conditions were considered, the decreasing number of visits would not be captured and a large overestimation (attributed to a virtual overload not corresponding to reality) would be obtained.

4. Conclusion

Home Care providers need suitable skills, methodologies and tools for managing logistic and organisational activities in order to predict the evolution of the

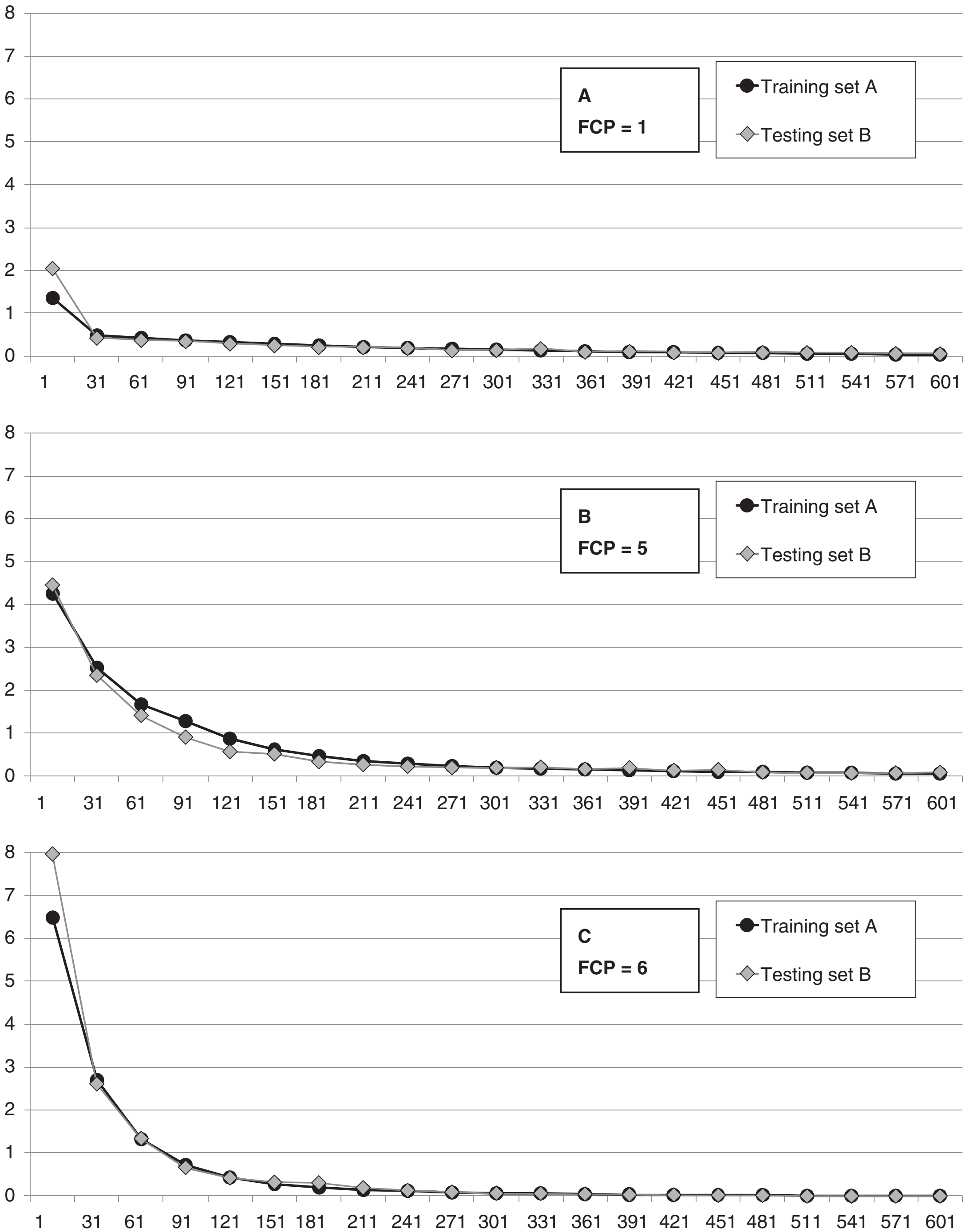

Figure 5. Average number of visits requested by a patient along the time relative to one of the random splits for the most important FCPs (in terms of workload (Table 4)) of each ASL Lecco type of care (Table 1). Abscissa values are grouped every 30 days for a better visualisation even if each day was considered in the model.

patients in charge and to support the delivery of care (Borsani *et al*. 2006, Castelnovo *et al*. 2006), while to date no literature study has dealt with workload planning in HC systems.

In particular, this article proposes an approach to develop a stochastic model for studying the changes in clinical, functional and social conditions for patients receiving an HC service. The model consists of a

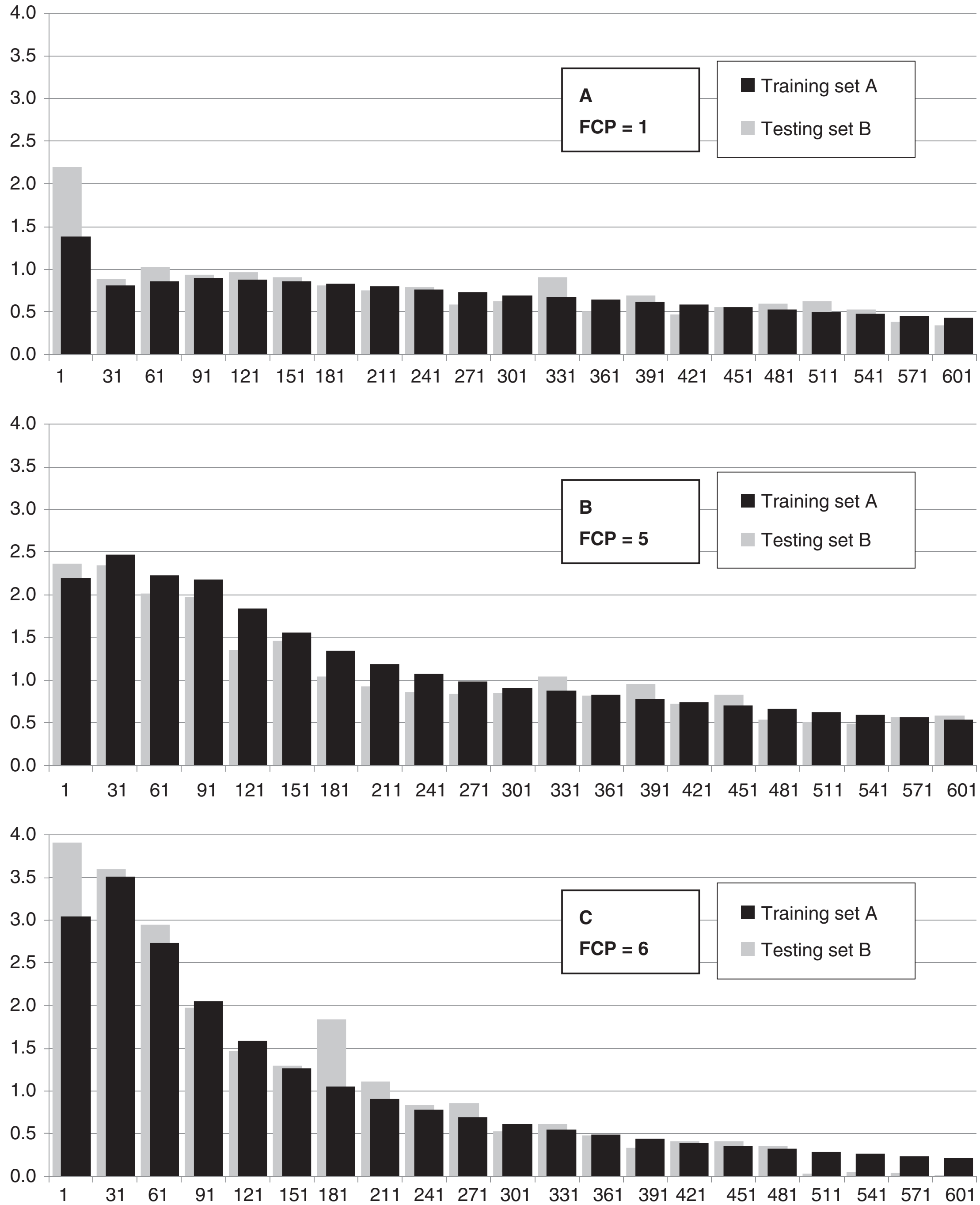

Figure 6. Standard deviation of the number of visits requested by a patient along the time relative to one of the random splits for the most important FCPs (in terms of workload (Table 4)) of each ASL Lecco type of care (Table 1). Abscissa values are grouped every 30 days for a better visualisation even if each day was considered in the model.

Markov chain coupled to the number of visits distribution for each state. These models performed well in the analysis of a real case and we believe that they can be applied to other HC providers. However, either the patient model or the cost model can be developed with other techniques, if they result to perform better in other specific cases; in these cases the proposed procedure for developing the patient stochastic model represents a general template for building reliable models.

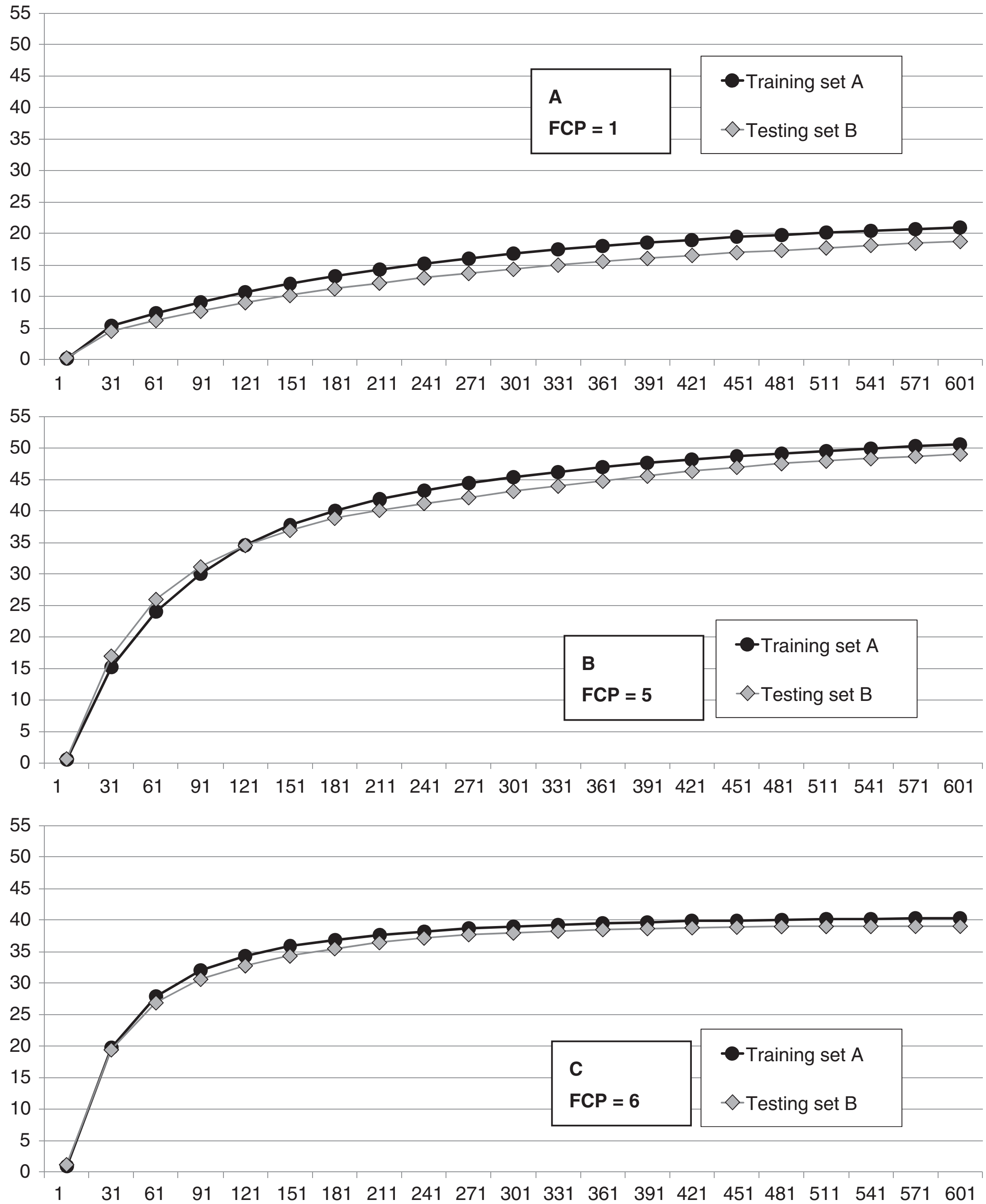

Figure 7. Cumulative number of visits requested by a patient along the time relative to one of the random splits for the most important FCPs (in terms of workload (Table 4)) of each ASL Lecco type of care (Table 1). Abscissa values are grouped every 30 days for a better visualisation even if each day was considered in the model.

The proposed model provides estimates on the CD and the number of requested visits over a period of time. This information could help the short-term resource planning, during the assignment of an operator to a patient, in order to avoid future operator surcharges. In the medium term, the HC management can make decisions (e.g. the number of operators needed in the next months) that are supported by the estimates of the number of patients and the workload related to them.

The model was validated by considering the real case of one of the largest public Italian HC providers,

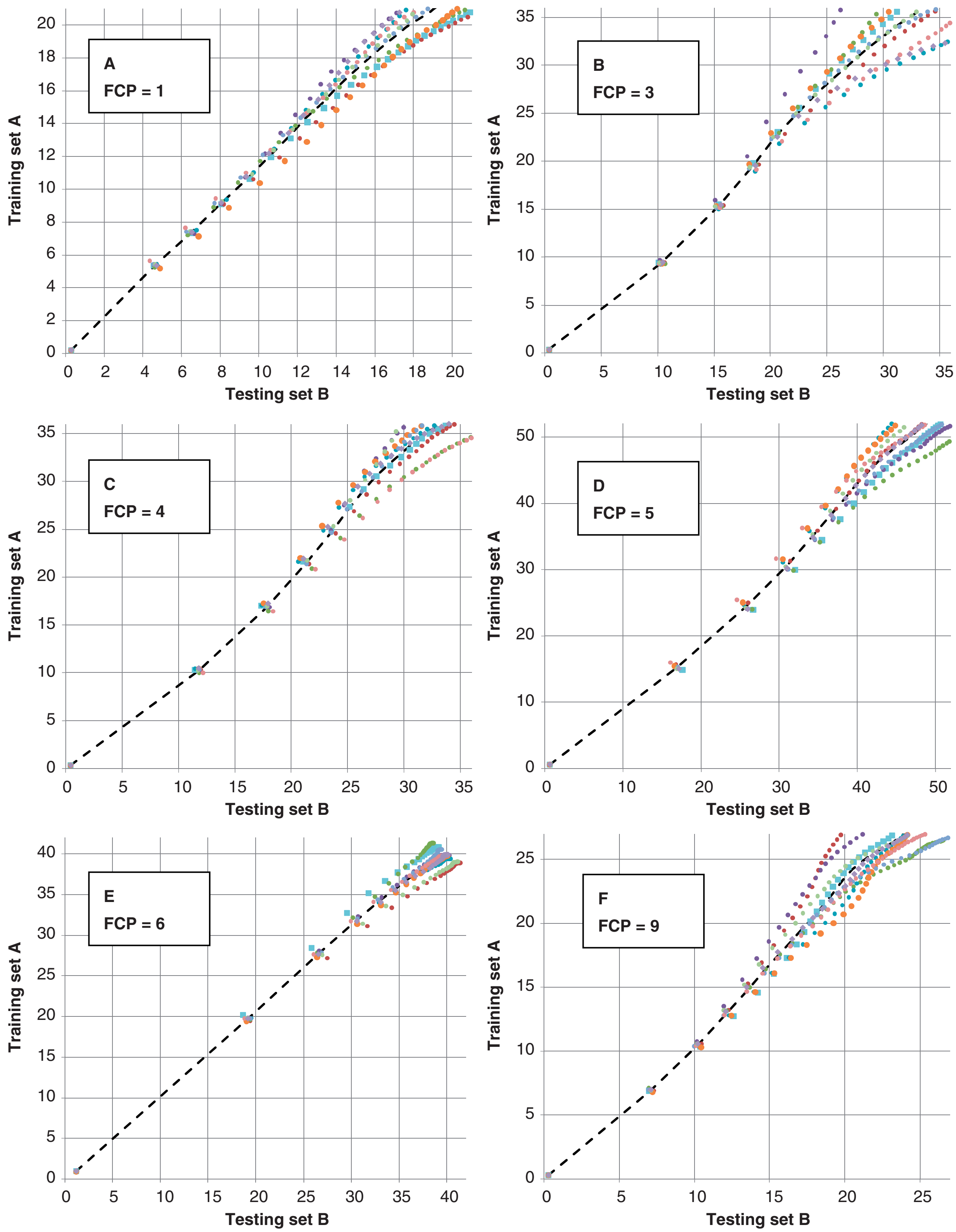

Figure 8. Scatter plot of the cumulative number of visits requested by a patient computed from Set A as a function of the corresponding data in Set B. The plots refer to the six most important FCPs in terms of workload (Table 4). Points of all the random splits are reported, while the line represents the average values on all the splits. Values are plotted every 30 days for a better visualisation even if each day was considered in the model.

Table 6. Comparison between the predicted and the actual number of visits supplied by the nurses in one of the three divisions of ASL Lecco for the 35 weeks following the data collection for the model implementation (marked as Week 0).

	Total			Palliative Care patients			Non-Palliative Care patients		
Week	Model	Real	Error (%)	Model	Real	Error (%)	Model	Real	Error (%)
0	794.3	845	−6.0	136.4	133	2.5	657.9	712	−7.6
1	807.4	871	−7.3	139.1	118	17.9	668.2	753	−11.3
2	810.6	892	−9.1	133.0	131	1.5	677.6	761	−11.0
3	821.1	819	0.3	136.4	139	−1.9	684.7	680	0.7
4	782.5	805	−2.8	103.6	107	−3.2	678.9	698	−2.7
5	767.7	867	−11.5	98.2	101	−2.8	669.6	766	−12.6
6	788.7	852	−7.4	111.4	120	−7.2	677.3	732	−7.5
7	790.6	801	−1.3	116.0	129	−10.1	674.5	672	0.4
8	791.0	779	1.6	119.3	117	2.0	671.7	662	1.5
9	796.6	781	2.0	111.4	104	7.1	685.2	677	1.2
10	750.0	827	−9.3	99.8	82	21.8	650.2	745	−12.7
11	752.6	745	1.0	110.2	103	7.0	642.3	642	0.1
12	742.7	767	−3.2	108.8	111	−2.0	633.9	656	−3.4
13	749.4	781	−4.0	114.0	94	21.3	635.4	687	−7.5
14	741.6	756	−1.9	104.1	91	14.5	637.4	665	−4.2
15	760.8	819	−7.1	113.1	126	−10.3	647.7	693	−6.5
16	761.2	788	−3.4	115.2	113	1.9	646.1	675	−4.3
17	763.2	784	−2.7	101.2	100	1.2	662.0	684	−3.2
18	748.6	766	−2.3	103.2	94	9.8	645.4	672	−4.0
19	745.5	704	5.9	97.4	79	23.3	648.1	625	3.7
20	735.9	798	−7.8	85.7	78	9.9	650.2	720	−9.7
21	729.9	785	−7.0	88.6	105	−15.6	641.3	680	−5.7
22	740.0	804	−8.0	101.2	119	−15.0	638.8	685	−6.8
23	730.9	815	−10.3	109.6	115	−4.7	621.3	700	−11.2
24	739.7	825	−10.3	104.0	108	−3.7	635.7	717	−11.4
25	744.8	764	−2.5	101.6	99	2.6	643.2	665	−3.3
26	762.7	807	−5.5	102.0	91	12.1	660.6	716	−7.7
27	789.6	798	−1.1	117.6	108	8.9	672.0	690	−2.6
28	785.4	824	−4.7	139.0	126	10.3	646.4	698	−7.4
29	768.4	805	−4.5	121.3	119	1.9	647.1	686	−5.7
30	795.7	816	−2.5	134.1	113	18.7	661.6	703	−5.9
31	809.3	826	−2.0	123.0	111	10.8	686.4	715	−4.0
32	810.4	852	−4.9	116.4	111	4.9	694.0	741	−6.3
33	815.4	781	4.4	125.3	106	18.2	690.1	675	2.2
34	799.2	829	−3.6	127.6	121	5.5	671.6	708	−5.1
35	806.9	827	−2.4	122.0	116	5.2	684.8	711	−3.7
TOT	27830.3	29005	−4.1	4090.8	3938	3.9	23739.4	25067	−5.3

which was directly involved during the phases of model development, validation and application.

Numerical results confirmed that the proposed model provides fairly accurate estimations. They showed how the stochastic evolution of a HC patient can be easily captured by this simple and easily applicable model. Only the standard available data, defined and collected by ASL Lecco (and by each HC provider), were considered. This avoided the requirement for unavailable details of clinical and psycho-social patient conditions and allowed accurate workload estimations to be obtained.

Finally, the model was developed as a simple software application. It was integrated into the managerial software already being used by the analysed provider. It is currently in use to support HC decisions with patient to operator assignment.

A future development involves the utilisation of model outcomes for a robust assignment of patients to the operators, thus preserving the continuity of care

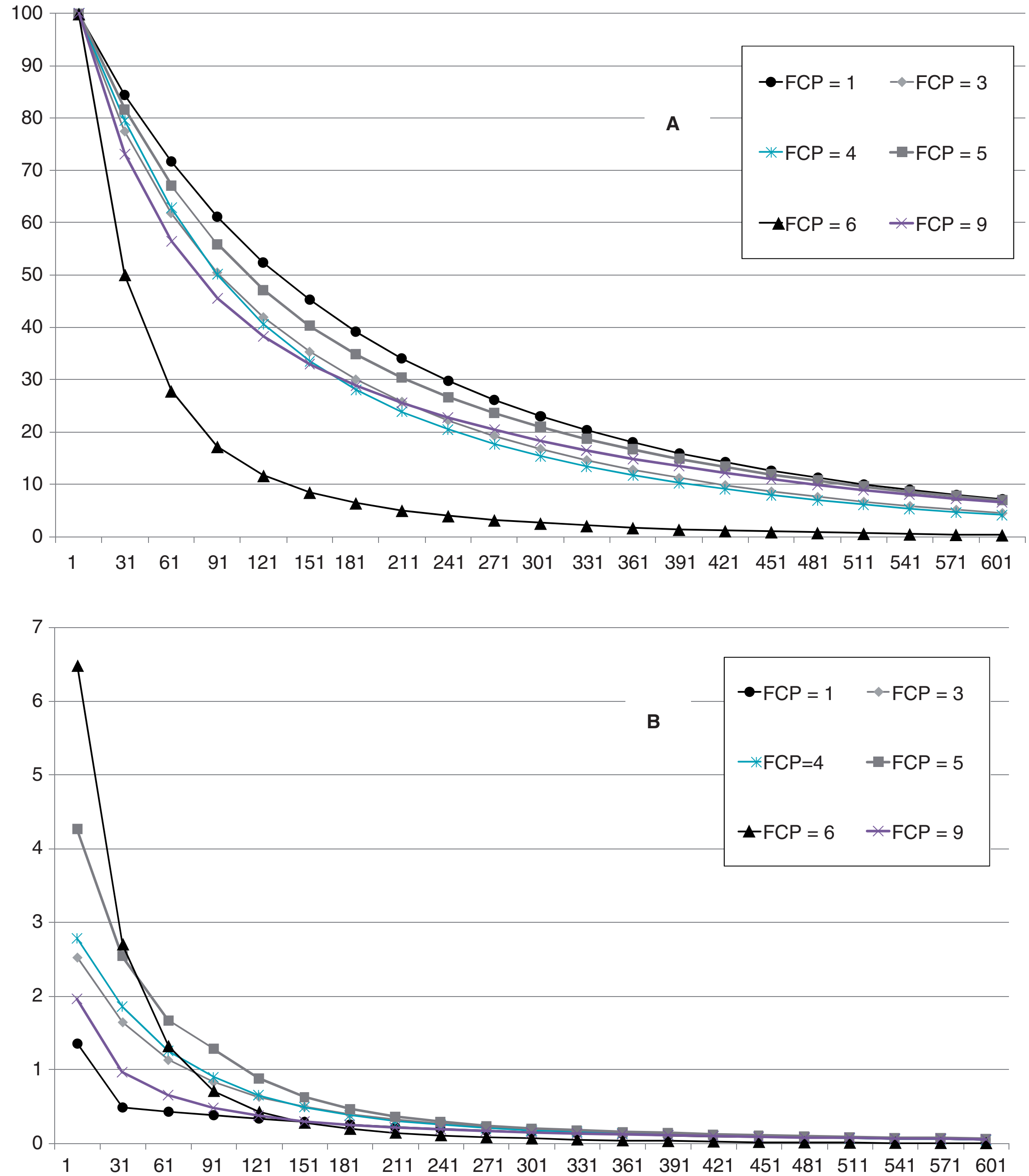

Figure 9. Overview of model outputs in terms of the number of patients $n(k)$ still in charge after k days (A) and the average number of visits requested by a patient along the time (B). Data refer to a random split for the six most important FCPs (in terms of workload (Table 4)).

(once a patient is assigned to an operator, he will follow his entire care pathway) and balancing the workload among the operators. For this purpose, stochastic programming could be adopted, where the inputs are the model outputs. Moreover, the definition of an optimal policy assignment could also be developed.

Other future work will focus on the comparison with other techniques (such as simulation or regression methods). The main goal is to improve the obtained results and determine the most appropriate technique in satisfying the HC requirements.

Acknowledgements

This research was partially funded by Fondazione LuVI, under the project 'Development of a supporting tool to service planning in home care', and by Sapio Life.

Notes on contributors

Ettore Lanzarone, MS, PhD, was born in Milan, Italy, in 1979. He received his MS degree (cum laude) in Biomedical Engineering and PhD degree in Bioengineering in 2004 and 2008, respectively, from the Politecnico di Milano, Milan, Italy, working on pulsatile devices for temporary replacement of the cardiac functionality. At present he works as a research fellow at the Department of Mechanical Engineering of the Politecnico di Milano dealing with stochastic models to estimate the future clinical conditions of patients for the planning of the human and material resources in health structures, particularly home care services.

Andrea Matta, MS, is a researcher at the Department of Mechanical Engineering of the Politecnico di Milano, Milan, Italy. He received his Master Degree in Industrial Engineering in 1997 from the Politecnico di Milano. Since 1997, he has been developing his research activity at the Department of Mechanical Engineering of the Politecnico di Milano. His research area includes analysis and design of production and service systems. He currently teaches the following courses at the Politecnico di Milano: Manufacturing, Integrated Production Systems and Virtual Manufacturing.

Gianlorenzo Scaccabarozzi, MD, is the director of the Dipartimento Interaziendale della Fragilità ASL/AO di Lecco, Lecco, Italy. He is member of the scientific board of numerous associations for Palliative Care. He participated with different working groups at the Italian Health Ministry and the Regione Lombardia; he has been a member of the National Board for Essential Levels of Assistance at the Italian Health Ministry. He is external professor for the Palliative Care Master at the Università degli Studi di Milano Bicocca, Milan, Italy. He received the silver medal of merit for Public Health (from the President of the Italian Republic).

References

Alagoz, O., *et al.*, 2005. Incorporating biological natural history in simulation models: empirical estimates of the progression of end-stage liver disease. *Medical Decision Making*, 25, 620–632.

Altman, R.M. and Petkau, A.J., 2005. Application of hidden Markov models to multiple sclerosis lesion count data. *Statistics in Medicine*, 24 (15), 2335–2344.

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

Bennett, W.G., *et al.*, 1997. Estimates of the cost-effectiveness of a single course of interferon-2b in patients with histologically mild chronic hepatitis C. *Annals of Internal Medicine*, 127, 855–865.

Bergamaschi, R., *et al.*, 2000. Usefulness of Bayesian graphical models for early prediction of disease progression in multiple sclerosis. *Neurological Sciences*, 21 (52), 5819–5823.

Berzuini, C. and Allemani, C., 2004. Effectiveness of potent antiretroviral therapy on progression of human immunodeficiency virus: Bayesian modelling and model checking via counterfactual replicates. *Journal of the Royal Statistical Society, Series C*, 53 (4), 633–650.

Borsani, V., *et al.*, 2006. A home care scheduling model for human resources. *In: Proceedings of ICSSSM06* (*International conference on service systems and service management*), Troyes, France, 25–27 October. IEEE, 449–454.

Brusco, M.J. and Steinley, D., 2009. Cross validation issues in multiobjective clustering. *British Journal of Mathematical and Statistical Psychology*, 62 (2), 349–368.

Castelnovo, C., *et al.*, 2006. A multi agent architecture for home care services, In: M. Tavakoli and H.T.O. Davies, eds. *Reforming health systems: analysis and evidence*. Aldershot, UK, Ashgate Publishing, 135–151.

Chahed, S., *et al.*, 2009. Analysis of organisational models of health home care service providers. Paper submitted to *International Journal of Healthcare Technology and Management*.

Chahed, S., *et al.* 2006. Operations management related activities in home health care structures. *In: Proceedings of INCOM (Information Control Problems in Manufacturing) conference*, Saint-Etienne, France, 17–19 May, Vol. 3, 641–646. Available from: http://www.emse.fr/incom06.

Congdon, P., 2001. The development of Gravity models for hospital patient flows under system change: a Bayesian modelling approach. *Health Care Management Science*, 4, 289–304.

Cucinotta, D., Angelin, A., and Godoli, G., 1989. Proposta e validazione di un semplice indice per la valutazione funzionale globale dell'anziano: il GEFI. *Giornale Gerontologia*, 38, 31–36.

Davies, C. and Dale, J., 2003. Paediatric home care for acute illness: I. GPs' and hospital-at-home staff views. *International Journal of Health Care Quality Assurance*, 16 (7), 361–366.

Emparan, C., *et al.*, 2003. Cost-effectiveness analysis of Basixilimab induction and calcineurin-sparing protocols in 'old to old' programs using Markov models. *Transplantation Proceedings*, 35 (4), 1324–1325.

Flessa, S., 1999. Decision support for malaria control programmes – a system dynamics model. *Health Care Management Science*, 2, 181–191.

Gardiner, J.C., *et al*., 2006. A dynamic model for estimating changes in health status and costs. *Statistics in Medicine*, 25, 3648–3667.

Hollander, M., *et al*., 2002. *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.

Husted, J.A., *et al*., 2005. Description and prediction of physical functional disability in psoriatic arthritis: a longitudinal analysis using a Markov model approach. *Arthritis & Rheumatism*, 53 (3), 404–409.

Jones, J., *et al*., 1999. Economic evaluation of hospital at home versus hospital care: cost minimisation analysis of data from randomised controlled trial. *British Medical Journal*, 319, 1547–1550.

Katz, S., *et al*., 1970. Progress in development of the index of ADL. *Gerontologist*, 10 (1), 20–30.

Kellogg, D.L. and Walczak, S., 2007. Nurse scheduling: from academia to implementation or not? *Interfaces*, 37, 355–369.

Koizumi, N., Kuno, E., and Smith, T.E., 2005. Modelling patient flows using a queuing network with blocking. *Health Care Management Science*, 8, 49–60.

Kousignian, I., *et al*., 2003. Markov modelling of changes in HIV-specific cytotoxic T-lymphocyte responses with time in untreated HIV-1 infected patients. *Statistics in Medicine*, 22 (10), 1675–1690.

Krahn, M., *et al*., 2004. Estimating the prognosis of hepatitis C patients infected by transfusion in Canada between 1986 and 1990. *Medical Decision Making*, 24, 20–29.

Lawton, M.P. and Brody, E.M., 1969. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*, 9, 179–186.

Lee, S., Huang, H., and Zelen, M., 2004. Early detection of disease and scheduling of screening examinations. *Statistical Methods in Medical Research*, 13, 443–456.

Lee, S.J. and Zelen, M., 2008. Mortality modeling of early detection programs. *Biometrics*, 64, 386–395.

Magherini, A., *et al*., 2005. Time ordering in frontal lobe patients: a stochastic model approach. *Brain and Cognition*, 58 (3), 286–299.

Marshall, A.H., *et al*., 2002. Modelling patient duration of stay to facilitate resource management of geriatric hospitals. *Health Care Management Science*, 5, 313–319.

Marshall, A.H. and McClean, S.I., 2004. Using coxian phase-type distributions to identify patient characteristics for duration of stay in hospital. *Health Care Management Science*, 7, 285–289.

Marshall, A., Vasilakis, C., and El-Darzi, E., 2005. Length of stay-based patient flow models: recent developments and future directions. *Health Care Management Science*, 8, 213–220.

McClean, S.I., McAlea, B., and Millard, P.H., 1998. Using a Markov reward model to estimate spend-down costs for a geriatric department. *Journal of the Operational Research Society*, 49, 1021–1025.

McClean, S.I. and Millard, P.H., 1998. A three compartment model of the patient flows in a geriatric department: a decision support approach. *Health Care Management Science*, 1, 159–163.

McClean, S.I. and Millard, P.H., 2007. Where to treat the older patient? Can Markov models help us better understand the relationship between hospital and community care? *Journal of the Operational Research Society*, 58, 255–261.

Mor, V., *et al*., 1984. The Karnofsky performance status scale. An examination of its reliability and validity in a research setting. *Cancer*, 53 (9), 2002–2007.

Pauler, D.K. and Finkelstein, D.M., 2002. Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. *Statistics in Medicine*, 21 (24), 3897–3911.

Remák, E., *et al*., 2003. A Markov model of treatment of newly diagnosed epilepsy in the UK. An initial assessment of cost-effectiveness of topiramate. *The European Journal of Health Economics*, 4, 271–278.

Taylor, G. and McClean, S., 1996. Geriatric-patient flow-rate modeling. *IMA Journal of Mathematics Applied in Medicine and Biology*, 13, 297–307.

Taylor, G., McClean, S.I., and Millard, P.H., 1997. Continuous-time Markov models for geriatric patient behaviour. *Applied Stochastic Models and Data Analysis*, 13, 315–323.

Taylor, G.J., McClean, S.I., and Millard, P.H., 2000. Stochastic models of geriatric patient bed occupancy behaviour. *Journal of the Royal Statistical Society, A 163 Part*, 1, 39–48.

Verbeek, H.J.M., *et al*., 2001. Routine follow-up examinations in breast cancer patients have minimal impact on life expectancy: a simulation study. *Annals of Oncology*, 12 (8), 1107–1113.

Verotta, D., 2005. Models and estimation methods for clinical HIV-1 data. *Journal of Computational and Applied Mathematics*, 184 (1), 275–300.