# Chapter 16
# A Whole-System Approach to Identify the Sources of Variation in Patient Flow

**Nasim Arbabzadeh, Mohsen A. Jafari, and Kian Seyed**

**Abstract** The main objective of this paper is to develop a quantitative framework to identify the main sources of variation in patient flow. Since 1983, under Health Care Financing Administration (HCFA)'s system, generally referred to as the Prospective Payment System (PPS), each hospital inpatient is classified into one of around 500 Diagnosis-Related Groups (DRGs), and the hospital is paid the amount that HCFA has assigned to each DRG. In other words, irrespective of what the hospital charges for, it will be paid only a fixed price for each DRG through major reimbursement plans. Therefore, it is logical to expect that by reducing the within DRG discrepancies, hospitals can cut cost and improve patient safety and satisfaction. In order to reach this goal the first step is to identify the main sources of variations. In this paper, we apply classical quality/process control tools and well known data mining methods to determine significant factors affecting the patient sequence among tens or hundreds of potential factors.

## 16.1 Introduction

During their hospital stay, patients may experience redundant steps and procedures that may lead to unnecessary excessive expenses, lower Quality of Care (QoC) and customer dissatisfaction. The excessive costs are often covered by hospitals or paid by individual patients since insurance companies have standard payment plans ranging from the infamous charge master or fee-for-service (FFS) price list to bundled payment systems such as diagnosis-related groups (DRGs), with various

N. Arbabzadeh • M.A. Jafari (✉) • K. Seyed
Industrial and Systems Engineering Department, Rutgers, The State University of New Jersey, 96 Frelinghuysen Rd., CoRE Building, Room 201, Piscataway, NJ, USA
e-mail: majafari@gmail.com

forms of "discounts off charges" and "per diems" somewhere in between [1, 2]. Regardless of who pays for these excessive and unnecessary expenses, the adverse societal impacts and negative business consequences are immense. In this paper, we focus on the patient flow process in a hospital with DRG based payment system for its inpatient claims.

Renewed focus on quality measurement and improvement and on medical-error reduction has heightened interest in paying for performance, rather than just reimbursing providers for services rendered. Private Pay for Performance (P4P) programs for hospitals usually pays bonuses as an incentive above the agreed-upon reimbursement rate. A more rational reimbursement system, which rewards quality of care rather than simply doing more to patients, is the short-term goal of paying for performance. The longer-term goal is also to make the health care system more efficient. It has become clear that under existing reimbursement structures, current market forces are insufficient to ensure either higher-quality or more-cost-effective care [2]. P4P programs can be seen as additional incentives for hospitals to seek to improve their patient flow processes which can be attained through our variation reduction framework.

Since 1983, under Health Care Financing Administration (HCFA)'s system, generally referred to as the Prospective Payment System (PPS), each hospital inpatient is classified into one of around 500 Diagnosis-Related Groups (DRGs), and the hospital is paid the amount that HCFA has assigned to each DRG. Thus all hospitals treating all patients who fall into a particular DRG may charge whatever they charge based upon their patients' courses of treatment, but each will be paid the same. One limitation to this methodology is that individual DRG categories often combine subgroups of patients with predictably different expected resource costs. HCFA has repeatedly improved the DRG definitions since 1984; in fact a new DRG system, called Medicare Severity DRGs (MS-DRGs), was adopted in October 1, 2007 which replaced 538 DRG system with 745 new MS-DRGs [1]. This enhancement, while necessary, does not fully account for differences in illness severity associated with substantial disparities in providers' costs.

The fact is only a part of these disparities is attributed to the patient profile including his/her demographics, medical history, medication, physical exams, and so on; these are uncontrollable factors in patient flow. There are also controllable factors that influence patient's experience from hospital admission to discharge. These include, but not be limited to, the order of treatments patient receives, medical procedures, current medications, received resources including physicians, nurses, technicians, transporters, and administrative work. These sources of variability could severely impact patient safety, QoC, professional satisfaction, and hospital revenue. The potential reduction in costs and increase in QoC and patient safety and satisfaction will be too rewarding to ignore. All these tools become handier especially when the regular normal operation of hospital is affected by an external incident varying from highway crashes to earthquakes and terrorist attacks. It's in such situations that having a managed patient flow can be of great help to the hospital management to increase patient care and lower the number of fatalities.

This article is organized as follows. Section 16.2 presents the literature survey. In Sect. 16.3 we present the formulation of our problem. The data to test our procedure and the results of applying our methodology are discussed in Sect. 16.4. Conclusions are presented in the final section.

## 16.2   Literature Survey

A number of researchers have used queueing models to study various aspects of the patient flow process. McClean et al. (2005) use phase-type distributions to carry out model-based clustering of patients using the time spent by the patients in hospital. They cluster patients into classes on the basis of the number of phases involved. Cadez et al. (2003) presented a new methodology for exploring and analyzing navigation patterns on a web site [3]. They partition site users into clusters such that users with similar navigation paths through the site are placed into the same cluster. Their proposed method clusters users by learning a mixture of first-order Markov models using the Expectation-Maximization algorithm. In this paper, we have used their proposed model to cluster patient sequences in the hospital.

## 16.3   Technical Approach

Patient flow is not a single datum but a pattern or a sequence of steps. Unlike classical statistics where singular or array of data is used, we need to work with flow patterns and ordered data. In this paper, we use a mixture of first-order Markov models to model patient flow. Each patient is admitted to an inpatient floor with an initial diagnosis determined by the admitting physician. After patient is discharged, her chart is reviewed by coders and a DRG is assigned based on the primary (definitive final diagnosis) and other diagnoses together with treatments, resources and procedures utilized towards treating patient's condition during her stay. For each DRG certain level of resources (treatments, diagnostic tests, procedures, etc.) are assigned and required. From admission to discharge, a patient goes through a sequence of steps both in terms of her condition and the utilized resources, treatments and procedures. Throughout this paper we will refer to this sequence of steps as *patient flow* vector and denote it by $\overrightarrow{S}_i$ which is defined as follows:

$$\overrightarrow{S}_i = [S_{i1}, S_{i2}, \cdots, S_{ij}, \cdots, S_{in}]', i = 1, 2, \cdots, m \ and \ j = 1, 2, \cdots, n \qquad (16.1)$$

where $\overrightarrow{S}_i$ is a $n \times 1$ ordered vector with jth element, $S_{ij}$, as the state of patient i at step j ($j = 1, 2, \ldots, n$). $S_{ij}$ takes on values ($s_{ij}$) from among N ($n = 1, 2, \ldots, N$) possible patient states. Therefore the sequence $[S_{i1}, S_{i2}, \cdots, S_{ij}, \cdots, S_{in}]'$ indicates that patient i first was at state $s_{i1}$, then $s_{i2}$, and so on. In our model, the last state

is always $x_n$, which is "discharged" state. The nature and definition of these states can be different according to the level of granularity of the problem, i.e. the level of detail at which patient flow is observed. They can be as aggregated as generic states that any patient may go through during a hospital stay (like admission, inpatient floor stay, and discharge), or they can be very detailed including all the steps in each of the above mentioned high level states.

As we mentioned earlier there can be several sources of variability that are intrinsic to all healthcare delivery systems. We have categorized these sources into three groups:

(i) Unique characteristics of each patient (patient profile), including demographics, medical history and other health conditions upon admission. $\overrightarrow{X}_i$ defines these characteristics:

$$\overrightarrow{X}_i = [X_{i1}, X_{i2}, \cdots, X_{ik}, \cdots, X_{ip}]', i = 1, 2, \cdots, m \ and \ k = 1, 2, \cdots, p \quad (16.2)$$

where $\overrightarrow{X}_i$ is a $p \times 1$ vector whose kth element, $X_{ik}$, represents the kth explanatory variable quantifying the kth characteristic of patient i.

(ii) Hospital resources, including medical and non-medical (overhead) staff {direct (nurse, tech, doctor) and indirect (unit secretary, housekeeping) labor and overhead labor}, major equipment, units and their functionalities (hospital factor). We denote by $\overrightarrow{Z}_i$ these characteristics:

$$\overrightarrow{Z}_i = [Z_{i1}, Z_{i2}, \cdots, Z_{i1}, \cdots, Z_{iq}]', i = 1, 2, \cdots, m \ and \ l = 1, 2, \cdots, q \quad (16.3)$$

where $\overrightarrow{Z}_i$ is a $q \times 1$ vector whose lth element, $Z_{il}$, represents the lth explanatory variable quantifying the lth hospital resource on patient i. Depending on the attribute which they quantify, $X_i$ and $Z_i$ can be of both types of explanatory variables: continuous or categorical.

(iii) Random noise denoted by $\varepsilon_i$ which are assumed to be i.i.d. random variables with mean zero and standard deviation $_i$. There are always un-assignable causes, which are usually grouped under random noise. Since random noise is statistically un-controllable, it is imperative to reduce its effect as much as possible. Any significant reduction in un-controllable variations will increase "process capability" and improve the process, which will in turn lead to significant cost reductions.

Furthermore, we assume that reentry of patient i to the hospital is a new admission with an updated $\overrightarrow{X}_i$ vector due to the new set of treatments that he received during his most recent stay. Then a historical data set of size $m$, containing $m$ vectors of $\overrightarrow{S}$, $\overrightarrow{X}_i$, and $\overrightarrow{Z}$ defines patient paths, patient characteristics and hospital resources of $m$ observed patients categorized under a specific DRG during a given time interval.

We intend to determine the number of clusters defined on the basis of sampled data collected on $\overrightarrow{S}$. We also intend to link $\overrightarrow{X}$, and $\overrightarrow{Z}$ to $\overrightarrow{S}$ in order to determine significant factors that lead to clusters within a DRG. Finally by controlling the

important attributes and reducing their variation we expect to see a reduction in the variations inherent in the patient flow process. Sections 16.3.1, 16.3.2, 16.3.3, 16.3.4, and 16.3.5 explain the steps of our algorithms in details.

### 16.3.1   Data Collection

With the current practices and adoption of EMR technology it is safe to assume that there are sufficient medical and personal data on patients, which can be mined and inferences can be made from. For example, CPT (Current Procedural Terminology) and HCPCS (Healthcare Common Procedure Coding System) codes are numbers assigned to every task and service a medical practitioner may provide to a patient including medical, surgical and diagnostic services. In principle, the data supporting CPT codes exist in hospitals (either collected real time using RFID or other RTLS technologies, or with some time lags entered by medical staff). Only in rare cases, the above data categories are all in a single database and is easily accessible; in majority of hospitals they are scattered in different databases, and data transfer and data fusion will be necessary. The data accessibility problem, however, is outside of the scope of this article. We will assume that this data exists and can be accessed for patient samples at different times.

### 16.3.2   Brainstorming

This step requires expert opinion to extract, filter, and transform data into meaningful quantifiable variables that we can further feed into our statistical engine. For this purpose, we should build multidisciplinary teams whose members will bring different perspectives and knowledge about the problem [4]. It is important to ensure that the core team and extended members include individuals that have direct contact with the process. The team should be brought together to hold brainstorming sessions for two important tasks:

1. Defining the state space of patient flow vector ($\overrightarrow{S}$): Medical judgment should be used to construct states, which both exhibit the necessary independence and make sense in terms of the delivery of care. A state space must be constructed in a manner that results in state definitions, which are mutually exclusive and collectively exhaustive [5]. This is essential to ensuring that Markov modeling of patient flow is valid.
2. Quantifying vectors of patient profile and hospital resources ($\overrightarrow{X}$, and $\overrightarrow{Z}$): To perform this task, one must try to identify as many potential variables as possible. One of the well-known tools to identify the potential causes of an event is the fishbone diagram also known as Ishikawa diagram or cause-and-effect diagram [6]. In this diagram, causes are usually grouped into major categories to identify these sources of variation.

Finally, we need to translate these potential causes into quantifiable random variables of either continuous or categorical type. The easiest case is when there are only two classes, such as variable gender with classes of "male" or "female". Examples for categorical type are gender, severity of illness, and nurse's level of expertise.

### 16.3.3 Sequence Clustering

At this step, we apply a mixture of fist-order Markov models to model patient flow sequences. We assume that the flow of each patient in the data set, $\vec{S}_i$, is generated independently (the traditional i.i.d. assumption). Statisticians refer to such a model as a mixture model with R components (R is the number of clusters). We apply Expected Maximization (EM) method to train our Once the model is trained, we can use it to assign each patient to a cluster or fractionally to the set of clusters. A mixture model for $\vec{S}$ with R components has the form:

$$p\left(\vec{S}\,\middle|\,\theta\right) = \sum_{r=1}^{R} p\left(c_r\,\middle|\,\theta\right) \cdot p_r\left(\vec{S}\,\middle|\,c_r,\theta\right) \tag{16.4}$$

where $c_r$ is the cluster assignment for a given patient, $p(c_r|\theta)$ is the marginal probability of the rth cluster $(\sum_r p(c_r|\theta) = 1)$ and $p_r\left(\vec{S}\,\middle|\,c_r,\theta\right)$ is the statistical model describing the distribution for the variables for patients in the rth cluster, and denotes the parameters of the model. We further assume that each model component is a first-order Markov model capturing the sequence of steps taken by a patient to some degree. Then, the EM method is used to train the parameters of the mixture model with known number of components R, given training data $d_{train} = \left\{\vec{S}_1, \vec{S}_2, \cdots, \vec{S}_M\right\}$ such that the following equation holds:

$$\theta^{ML} = \arg\max_\theta p\left(d_{train}\,\middle|\,\theta\right) = \arg\max_\theta \prod_{i=1}^{M} p\left(\vec{S}_i\,\middle|\,\theta\right) \tag{16.5}$$

$\theta^{ML}$ are the maximum likelihood or ML estimates of the model parameters.

In this paper, we have used Microsoft Sequence Clustering algorithm (SQL Server Analysis Services or SSAS) to carry out the sequence analysis. Microsoft SQL Server provides us with the membership assignment of each patient. Therefore, having a training data set of size M, we can run the sequence clustering algorithm and obtain the *vector of class memberships*, denoted by $\vec{Y}$, as follows:

$$\vec{Y} = [Y_1, Y_2, \cdots, Y_M]' \tag{16.6}$$

where $Y_i$ is the class membership of patient i, and can accept values of 1, 2, …, R. Later, we will feed this vector into the Variable Selection module.

### 16.3.4   Variable Selection

In this step, we will use a well-known classifier, namely random forest, to identify the most important variables which significantly affect the patient flow sequences. Random forest (or random forests) is an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the class's output by individual trees [7]. The data-set used for training comes in records of the form $(\overrightarrow{Q}, \overrightarrow{Y})$ for each data-point, where $\overrightarrow{Q}$ denotes a vector of observed characteristics (also referred as features or factors) and $\overrightarrow{Y}$ denotes a group label (also called target variable). In our application, $\overrightarrow{Q}$ is a $(p+q) \times 1$ vector of $\begin{bmatrix} \overrightarrow{X}_{p\times 1} \\ \overrightarrow{Z}_{q\times 1} \end{bmatrix}$ which contains the information of patient profile and hospital resources, i.e. the explanatory variables, and $\overrightarrow{Y}$ is the *vector of class memberships*, i.e., the output of the sequence clustering algorithm.

In order to perform the classification task we will use the *randomForest* package available in R software [8]. The input to the software will be feature vector $\overrightarrow{Q}_{(p+q)\times 1} = \begin{bmatrix} \overrightarrow{X}_{p\times 1} \\ \overrightarrow{Z}_{q\times 1} \end{bmatrix}$ and vector of *class memberships* $\overrightarrow{Y}$.

Random forests can be used to rank the importance of variables. There are two criteria based on which the Breiman's random forest calculates the importance of variables: *Gini importance* which calculates the mean Gini gain produced by $Q_i$ over all trees, and *permutation accuracy importance* which is the mean decrease in classification accuracy after permuting $Q_i$ over all trees. The variable importance plot gives a relative ranking of significant features, and absolute values of the importance scores should not be interpreted or compared over different studies. We consider, the first B variables as the most important variables where $B < p+q$. We will refer to the vector of important variables as $\overrightarrow{Q}'_{B\times 1} = \begin{bmatrix} \overrightarrow{X}'_{p'\times 1} \\ \overrightarrow{Z}'_{q'\times 1} \end{bmatrix}$, and define $\overrightarrow{X}'_i$, and $\overrightarrow{Z}'_i$ as follows:

$$\overrightarrow{X}'_i = [X_{i1}, X_{i2}, \cdots, X_{ik}, \cdots, X_{ip'}]', i = 1, 2, \cdots, m \text{ and } k = 1, 2, \cdots, p' \quad (16.7)$$

$$\overrightarrow{Z}'_i = [Z_{i1}, Z_{i2}, \cdots, Z_{i1}, \cdots, Z_{iq'}]', i = 1, 2, \cdots, m \text{ and } l = 1, 2, \cdots, q' \quad (16.8)$$

where $p \leq p'$, and $q \leq q'$.

### 16.3.5   Monitoring and Controlling Important Variables

Monitoring and controlling of important variables is the last step in our model. In the previous steps we established a relationship between patient flow sequences and process attributes, and identified those attributes that affect the patient flow process

significantly. In this step we investigate how and why these attributes affected patient flow. For this purpose, questions must be asked to find the assignable causes of variations and then a proper corrective action must be taken to eliminate them. To maintain the gained improvement and be able to detect future assignable variations, advanced statistical tools such as single-variable or multivariate control charts can be used. Using control charts is an ongoing activity over time to bring continuous improvements to the process.

## 16.4 Numerical Experimentation

In this section, we illustrate the performance of our algorithm using a simulated data set. For confidentiality reasons and also for the lack of sufficient real data at this time, we will demonstrate our model using simulated data. But the data generation will closely mimic the true real life process. We assume that patient flow sequences of cases under DRG type xxx can at most have six steps (N = 6). Seven factors have been identified as the potential causes of variation two of which are patient profile-related attributes (p = 2), and five are hospital resources (q = 5). The definitions of these variables can be found in Appendix.

To simulate expert opinion correctly, we assume to have a priori knowledge that $Z_1$, $Z_2$, and $Z_5$ are the significant variables and the rest of the attributes may not affect patient sequences significantly. Furthermore, we assume that we are given expert opinion on particular relationship between process attributes ($Z_1$, $Z_2$, and $Z_5$) and patient flow ($\overrightarrow{S}$). According to this prior knowledge, we know that exhaustively there exist 13 distinct patient sequences. It means that, assuming the patient flow process is a stable and stationary process without any chaotic behaviors, the expected path of a given patient falls into the set of 13 sequences. We model this relationship with a multinomial logit function regressing the transition probabilities on the value of significant attributes. The model is given by:

$$P\left(S_{ij} = W \middle| S_{i(j-1)} = V\right) = f\left(Z_1, Z_2, Z_3\right); W, V = 1, 2, \cdots, N \qquad (16.9)$$

where $P(S_{ij} = W | S_{i(j-1)} = V)$ is the probability that patient i is in state W at step j, given that he was in state V at step j-1. This definition comes from our assumption that the patient transfer between states follows a Markov model. $f$ is a multinomial logit function, and is defined as follows:

$$P\left(S_{ij} = W \middle| S_{i(j-1)} = V\right) = e^{\beta_0^W + \beta_1^W z_1 + \beta_0^W z_2 + \beta_0^W z_5} / 1 + \sum_{w=1}^{N} e^{\beta_0^w + \beta_1^w z_1 + \beta_0^w z_2 + \beta_0^w z_5}$$

$$W = 1, 2, \cdots, N - 1 \qquad (16.10)$$

$$P\left(S_{ij} = W \middle| S_{i(j-1)} = V\right) = 1 / 1 + \sum_{w=1}^{N} e^{\beta_0^w + \beta_1^w z_1 + \beta_0^w z_2 + \beta_0^w z_5}, W = N \qquad (16.11)$$
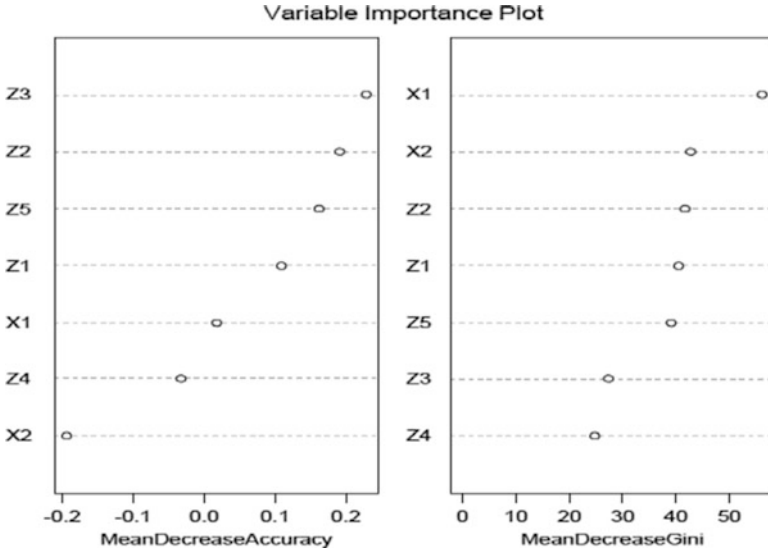
**Fig. 16.1** Variable importance plot, case I- incomplete data set

We estimate the parameters of the logit model ($\overrightarrow{\beta}^{w} = [\beta_0^w, \beta_1^w, \beta_2^w, \beta_3^w]$). To generate a simulated data set, we start with an initial set, $d^{(initial)} = \left\{ \overrightarrow{S}, \overrightarrow{X}, \overrightarrow{Z} \right\}$, and use the logit model to estimate the transition probabilities matrix as a function of significant attributes. In our example, the initial set included only 3 distinct sequences out of the set of 13 original sequences. Then, we keep fine-tuning the parameters by adding new sequences until no further improvement is gained in our estimations and the built model completely captures the original relationship between $\overrightarrow{S}$ and $\overrightarrow{X}, \overrightarrow{Z}$. The variable importance plot can verify this gradual improvement.

Following the above approach, at each iteration a training data set of 1,000 cases was generated and fed into the statistical engine. Figures 16.1 and 16.2 show the variable importance plots for two cases: case I- an incomplete data set including 7 distinct sequences, case II- a complete data set including all the13 distinct sequences.

As it can be seen in Fig. 16.1, if the incomplete data set is fed into the engine, we would mistakenly be led into the conclusion that either the set of variables $Z_2$, $Z_3$, and $Z_5$, according to the permutation accuracy importance, or the variables $Z_2$, $X_1$, and $X_2$, according to Gini importance, were the significant factors affecting the patient flow sequences. While if the complete data set is fed into the algorithm according to both measures of variable importance, $Z_1$, $Z_2$, and $Z_5$ are correctly identified as the most important variables. The values of importance measures for $X_1$, $X_2$, $Z_3$, and $Z_4$ are close to zero meaning that relatively speaking their effects on patient flow are trivial.
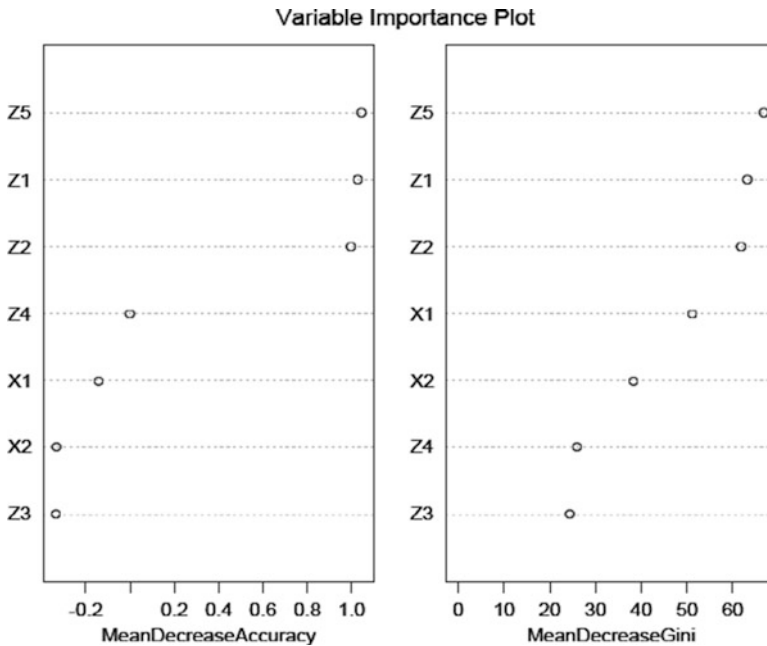
**Fig. 16.2** Variable importance plot, case II- complete data set

In summary, by using the random forest classifier we have been able to identify the significant factors that truly impact patient flow. With this valuable information, the hospital management should focus his efforts and money to improve these attributes, which can consequently improve and facilitate patient flow in the hospital. Finally, to maintain the acquired improvements, the use of multinomial or multiattribute control charts is suggested to constantly monitor and control the important attributes and be alerted if a disturbance occurs in the patient flow process [9].

Note that in our example all the important variables are hospital resource-related attributes. In case a patient profile attribute is identified as a significant variable one should use other alternative solutions to control the process. One solution would be the use of robust optimization methods to control such a process since we cannot control or change statistical distributions of patient profile attributes into our favor [10].

## 16.5   Conclusions

In this paper, we have proposed a novel framework to identify the sources of variations in the patient flow process. The main idea is that by reducing the variations of these single processes we will be able to reduce the variation of patient

sequences. Our simulated results show that having a statistically large historical data set, the classifier can correctly determine the important variables, which truly had relationships with patient sequences. We further suggest the use of statistical control charts to maintain the gained improvements. The hospital management can use this valuable information to improve the quality of its patient flow process which consequently improve patient and staff satisfaction and results in a better cost management.

## A.1  Appendix

Patient profile variables are as follows:

$$\text{Patient's age, } X_1 = \begin{cases} 1, & 0 \le age < 15 \\ 2, & 15 \le age < 30 \\ 3, & 30 \le age < 50 \\ 4, & 50 \le age \end{cases}$$

$$\text{Patient's gender, } X_2 = \begin{cases} 1, & Female, \\ 2, & Male, \\ 3, & Intersex. \end{cases}$$

Hospital resource-related variables are:

$$\text{Care taker's level of expertise: } Z1 = \begin{cases} 1, & Licensed\ Practical\ Nurse \\ 2, & Registered\ Nurse \\ 3, & Advanced\ Practice\ Nurse \end{cases}$$

$$\text{Work shift of the nurse, } Z_2 = \begin{cases} 1, & Morning\ shift, \\ 2, & Evening\ shift, \\ 3, & Night\ shift. \end{cases}$$

$$\text{The doctor's ranking: } Z_3 = \begin{cases} 1, & Physician, \\ 0, & Physician\ assistant. \end{cases}$$

$$\text{Availability of bedside monitoring tools: } Z_4 = \begin{cases} 1, & Available, \\ 0, & Otherwise. \end{cases}$$

$$\text{X\_Ray machine types: } Z1 = \begin{cases} 1, & Type\ M1\ maintained\ monthly, \\ 2, & Type\ M2\ maintained\ semiyearly, \\ 3, & Type\ M3\ maintained\ yearly. \end{cases}$$

## References

1. Lynk, W.J.: One DRG, one price? The effects of patient condition on price variation within DRGs and across hospitals. Int. J. Health Care Finance Econ. **1**(2), 111–137 (2001)
2. Nichols, L.M., O'Malley, A.S.: Hospital payment systems: will payers like the future better than the past? Health Aff. **25**(1), 81–93 (2006)

3. Cadez, I., Heckerman, D., Meek, C., Smyth, P., White, S.: Model-based clustering and visualization of navigation. Data Min. Knowl. Discov. 399–424 (2003)
4. McHugh, M., Dyke, K. V., McClelland, M., Moss, D.: Improving Patient Flow and Reducing Emergency Department Crowding: A Guide for Hospitals. AHRQ Publication. 11(12)-0094 (2011)
5. Weiss, E.N., Cohen, M.A., Hershey, J.C.: An iterative estimation and validation procedure for specification of semi-Markov models with application to hospital patient flow. Oper. Res. **30**(6), 1082–1104 (1982)
6. Ishikawa, K. (Translator: J. H. Loftus): Introduction to Quality Control, 448 p (1990). ISBN 4-906224-61-X OCLC 61341428
7. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
8. Liaw, A., Wiener, M.: Classification and regression by randomForest. R News **2**(3), 18–22 (2002)
9. Topalidou, E., Psarakis, S.: Review of multinomial and multiattribute quality control charts. Qual. Reliab. Eng. Int. **25**, 773–804 (2009)
10. Wu, C.F.J., Hamada, M.S.: Experiments: Planning, Analysis and Optimization. Wiley (2000)
11. McClean, S., Faddy, M., Millard, P.: Markov model-based clustering for efficient patient care. In: Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05) (2005).