Winter 12-4-2009

Intensifying Patient Flow Management by Using Data Mining Methods

Jouni Paavilainen

Jussi Myllärniemi

Jaani Väisänen

INTENSIFYING PATIENT FLOW MANAGEMENT BY USING DATA MINING METHODS

Jouni Paavilainen[1], Jussi Myllärniemi[2], and Jaani Väisänen[3]
Department of Business Information Management and Logistics,
Tampere University of Technology, Finland
[1]jouni.paavilainen@tut.fi; [2]jussi.myllarniemi@tut.fi; [3]jaani.vaisanen@tut.fi

Abstract

Health care organizations need means to process the patient flow-related data into meaningful information in order to understand and manage complex patient flow networks. In this paper we present a theoretical framework of the patient flow management process and examine how this process could be intensified with the help of data mining methodology. Based on the empirical findings in a case organization, a Finnish special care hospital, the potential of suitable data mining methods, like sequence analysis, is evident. Using the data mining methodology, huge amounts of data could be processed into relevant and reliable information to support decision-making. As a result, the overall quality and effectiveness of the patient flow network would be improved.

Keywords: Health Care, Patient Flow, Patient Flow Network, Patient Flow Management, Data Mining, Sequence Analysis

Introduction

Throughout the world, the pressure for intensifying health care services is high. Demand for the services is increasing rapidly due to e.g. ageing, while even the current level of the health care costs is considered to be too high in many countries. [1] Therefore, in future, more services should be offered with existing resources. Based on our previous researches concerning the health care network in Tampere, Finland [2] [3], the health care related expenses could be efficiently controlled by intensifying the management of patient flows. The main function of a health care system is to provide patient flows, the functionality of which thus affects directly the cost of the system [4]. Consequently, the intensified management of the patient flows could result in remarkable benefits.

Despite the importance of the matter, there is a lack of broad understanding when it comes to the formation of the whole health care network and its patient flows [1]. Thus, the biggest challenge is to clarify the overview of the complex network formed by various organizations and units with different responsibilities and functions. It is claimed that instead of concerning the whole network, individual health care units currently develop their own operations usually separately from each other [5].

But how can the broader understanding of patient flow network be achieved in order to manage the service system better? It is found that health care organizations record huge amounts of patient flow-related data, but do not have the means to analyze and process it into relevant information for management and development purposes [6]. Therefore, there is a need for analyzed information which could be utilized in the decision-making concerning the development of a whole health care system. It seems that data mining methods could be an efficient approach to tackle the challenge, as they are designed to process large amounts of data and find notable correlations among it [7]. Regardless of this, the utilization of data mining in this field is not systematically studied [8].

In this paper, we consider patient flow management in Finnish health care environment and especially in Hatanpää hospital, which is the municipal special health care provider in the city of Tampere. The hospital serves citizens mainly in surgeries and orthopedics as well as in internal and infectious diseases. The hospital management has recognized challenges in the patient flow management. The challenges are related to both internal operations of the hospital but also to the whole health care network of Tampere. This paper concentrates on the challenges that could be resolved by processing the patient flow-related data into useful information more effective than today.

The paper is organized as follows. In the theoretical part, we form the concept of patient flow management based on the existent literature and our earlier research, as well as present the concept of data mining. After that, we examine the patient flow management practices in Hatanpää hospital; what kind of patient flow-related data is recorded and which are the current methods for analyzing it. Finally, we examine via an example how the use of data could be intensified with the help of data mining methods. To conclude, the point of interest of this paper can be stated as follows: how patient flow-related data is refined into information at the moment in order to manage processes that provide patient flows, and how this could be intensified by using data mining methods.

Theoretical Framework

In this chapter, we examine the concepts of a patient flow and a patient flow network, and based on them, present a theoretical framework for intensifying the patient flow management. At the end of the chapter, we present the data mining methodology, which could be used when processing data into information in order to intensify the patient flow management.

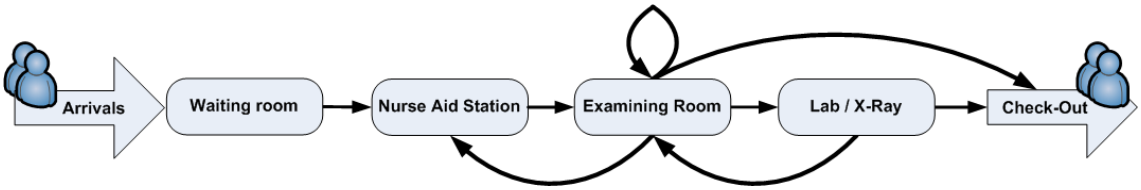

FIGURE 1. Patient flow for patients at a family practice clinic [4]

Patient Flow Management

The term patient flow refers to the movement of patients between different operations of health care units. Four common characteristics are mutual to all patient flows: (1) an entrance, (2) an exit, (3) a path connecting the entrance to the exit, and (4) random health care elements along the path. So, between entrance and exit, there are a set of units that the patient may encounter and, at the same time, exploit units' resources (e.g., beds, examining rooms, physicians, nurses, and medical procedures). This implies that patient flows can be depicted as a network. [4] Figure 1 is an illustration of this kind of network. In this example, the setting is quite simple with only few units and a quite little variety of patients. Considering a whole hospital with dozens of units and hundreds of patient groups, the overall network becomes much more complex.

It is important to dissect the whole network instead of separate units. To improve the flow, hospitals must view the problem in terms of an interdependent system rather than individual departments. The units that actually deliver the care in hospitals are tightly linked. Therefore, understanding patient flow requires looking at the whole care system, not just the isolated unites. Any individual department that often improves flow by itself in its area in fact relegates the problem to other dependent departments. [5] So, it can be stated that an understanding of patient flow is needed to manage a health care facility. Patient flow through different health care units is equivalent to the demand for services of these units [4]. Thus, an effective resource allocation and capacity planning are contingent upon the patient flows.

Studies reveal that it is a common but incorrect assumption that patient flows are a result of randomness. When analyzed, the variation caused by the structure of the service system itself is much bigger that the one caused by the randomness of patient arrivals and their diagnosis. [5]. Hence, once a health care organization has a grip of its patient flow network, the variation of demand can be managed. Consequently, quantitative tools, like forecasting and queuing models, can be used to help decision makers to estimate what resources (e.g. medical staff and equipment) will be required. [4] That is, resource allocation can be done much more accurately and flexibly when the operation is considered as a network of patient flows. [9]

It is now stated that the understanding of patient flow network is important, but how can this intellect be achieved? The concept of knowledge plays a key role in managing the health care services. Hence, one must have a wide knowledge about the system to steer it efficiently. [10] The term knowledge here refers to the outcome of human action which takes place in decision-making situations. This knowledge is acquired from information, which in turn is processed from data. [11] As mentioned earlier, a huge amount of data related to patient flows is recorded every day by the healthcare organizations, but this data seems to be used only for periodical simple reporting. Thus, the data should be processed into information and knowledge more efficiently than today.

At a general level, the data cultivation into information and knowledge is an iterative process [11] [12] [13]. Using this general theoretical framework as a basis, the authors have made a framework for intensifying the patient flow management in their earlier research [3]. This framework is illustrated in Figure 2. The intensification of patient flow management can be seen as an iterative process. This process proceeds as follows:

Hospital resources (1.) include all the tangible and intangible capital the hospital possesses. The former includes e.g. staff, facilities, equipment and materials. The latter means intellectual capital meaning e.g. the expertise and processes of the hospital. These resources **make it possible to provide (2.)** chosen care services for the customers, that is, produce patient flows including all the needed services. The amount and quality of the resources dictates how good the

patient flows can be arranged.

From an operational perspective, **patient flows (3.)** can be thought of as the movement of patients through a set of locations in a health care facility. Concerning the patient flows, a huge amount of data **is recorded (4.)**. This data is related to things like sources and destinations of the patients, queuing, utilization of support services, and durations of different phases.

Data related to patient flows (5.) can be divided into two different categories: operational

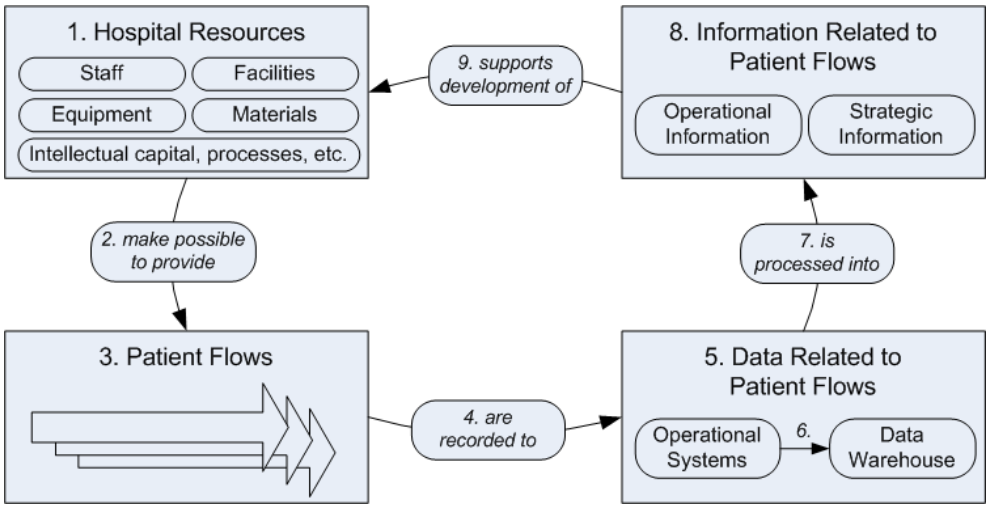

FIGURE 2. A theoretical framework for intensifying patient flow management [3]

data in operational IT systems and the data stored in **(6.)** different kinds of data warehouses in order to use it at a later date. To utilize the data in monitoring and management of the flows, it must **be processed into (7.)** information. This can be done by using various methods, including:

- Basic reporting using e.g. Microsoft Excel
- Measuring pre-defined key-figures
- Traditional mathematical and statistical methods
- Different kinds of data mining methods, e.g. clustering, segmentation and associate analysis
- Modelling and simulation methods.

Information related to patient flows (8.) can also be divided into two categories, though the difference between them is quite flexible. Operational information is related to daily processes and can be used to monitor and manage them. Strategic information for its part is used by the top management in order to make middle and long term strategic decisions. In other words, information related to patient flows **supports the development of (9.)** hospital resources. As stated earlier, effective resource allocation and capacity planning are contingent upon patient flows.

As said, the understanding of the patient flow network offers many possibilities to develop the system. One can identify and fix the problems – like bottlenecks – of the current state of the system as well as forecast the future demand in order to anticipate what resources will be required [4]. So, the achieved information about the network helps decision makers to manage patient flows better, which lead to better flows. This comes with great significance, for potential benefits of better flows include better clinical outcomes, improved patient safety, greater patient and staff satisfaction, and improved financial performance [5].

Data Mining

Data mining is defined as "…the process of selection, exploration, and modeling of large quantities of data to discover regularities or relations that are at first unknown with the aim of obtaining clear and useful results for the owner of the database" [14]. Considering the definition above, data mining would seem to be a useful tool in examining patient flow related data for the following reasons:

- The amount of data in the patient flows is seemingly large (tens of thousands of observations)
- Considering the amount of different units and the total number of all possible patient flows, the data structure is rather complex
- The underlying patterns in such a large data mass are most likely unknown at first.

The advantage of data mining over any traditional statistical techniques is that in statistics, there is usually a hypothesis based on some kind of a priori knowledge, and this hypothesis is then tested on the data in order to conclude whether the hypothesis is true or not [15]. In data mining, no prior knowledge is needed [14], which does not restrict the user into dealing with assumptions or preconceptions that might not be known

beforehand in the first place. Therefore, the authors propose that different data mining techniques would provide significant improvement over traditional statistical techniques in step 7 of the Figure 2.

In the light of framework illustrated in chapter 2.1., the point of interest of this study can be stated as follows: **How patient flow related data is refined into information in order to manage processes that provide patient flows, and how this could be intensified using data mining methods.**

Patient Flow Management in Hatanpää Hospital

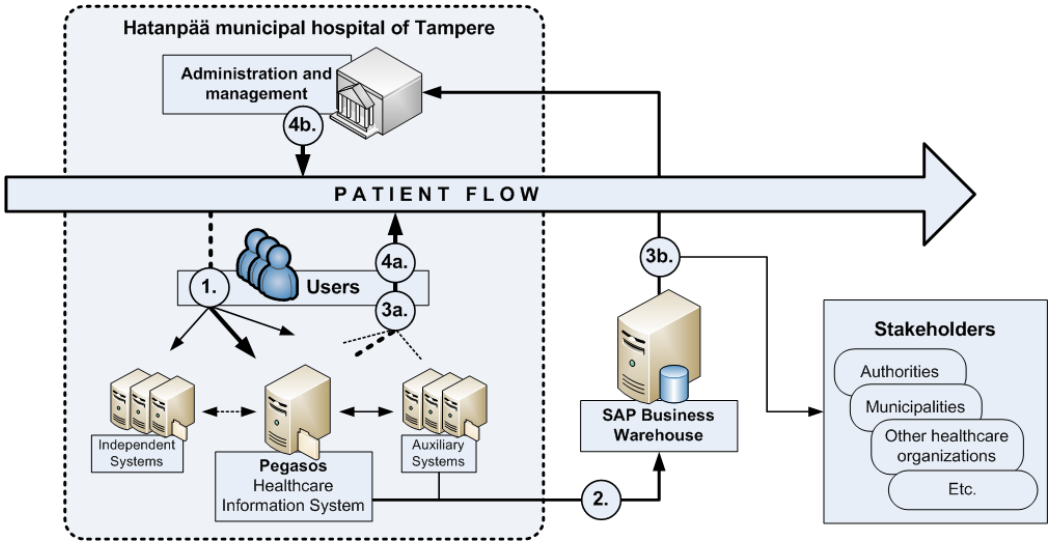

FIGURE 3. Patient flow management in Hatanpää hospital from the information flow point of view [3]

researchmade in Hatanpää hospital [3] is presented. First, we examine how patient flow related data is recorded for the moment and how it is used for patient flow management purposes. The actual process is illustrated in the light of the theoretical framework presented in the previous chapter. After this, the main challenges of the current process are discussed.

The Overall Process

According to our research, patient flow management in Hatanpää hospital can be seen as an iterative process which is illustrated in Figure 3.

The first step in the real process takes place, when users working with patients (e.g. doctors, nurses and receptionists) record patient flow related data to different operational information systems. The main system is called Pegasos provided by Logica. Besides it, there are a group of auxiliary systems as well as independent systems, which all have different levels of interfaces with it. This step corresponds to the step 4 in the theoretical model presented in Figure 2. The substantial content of the recorded data is described later in this chapter.

In **the second step** selected data is collected from Pegasos and the auxiliary systems in order to move them to the data warehouse (DW) provided by SAP. In this point, the data is structured and harmonized using EDM-system. This step corresponds to the step 6 in the theoretical model.

In **the third step**, the data related to patient flows is processed into useful information. In case of daily used operational information, this is done by getting the desired data straight from the operational systems by the users (3a). In case of strategic data used by the hospital administration and management, the desired information is generated using the report functionalities of the data warehouse (3b). Also different stakeholders get their information through the warehouse. This step as a whole corresponds to the step 7 in the theoretical model.

In **the fourth step**, the adopted information is used in the hospital's decision-making to support the development of hospital resources, in the first place its intellectual capital. The improved resources make it possible to react to the challenges of turbulent health care environment and to provide better patient flows. This step as a whole corresponds the steps 9 and 2 in the theoretical model.

Challenges in Processing the Recorded Data into Strategic Information

Many challenges and remarks related to the patient flow management process have been identified in Hatanpää hospital. Development is needed in every step presented in the model at figure 3. In this

paper we will focus on the step 3b, where data is processed into strategic information for the hospital management and administration. For understanding the challenges in this step, we first shortly describe the data the hospital records.

The substantial content of the recorded data related to patient flows is illustrated in figure 4. During the life cycle of a patient flow that typically starts from the entrance of a patient, the data is recorded into many different information systems. Though the data is mainly used in patients' health care, it can also be used in managing the whole hospital. Besides collecting the basic parameters of the patient, such as age, gender and diagnosis, the hospital unit records all kind of the basic information about patient flows, like volumes of the patient groups, prices of care, durations of stays, and patients' former and coming health care unit. This data is used to form some health care unit and patient group specific parameters. For example, average duration of stay is derived from patient

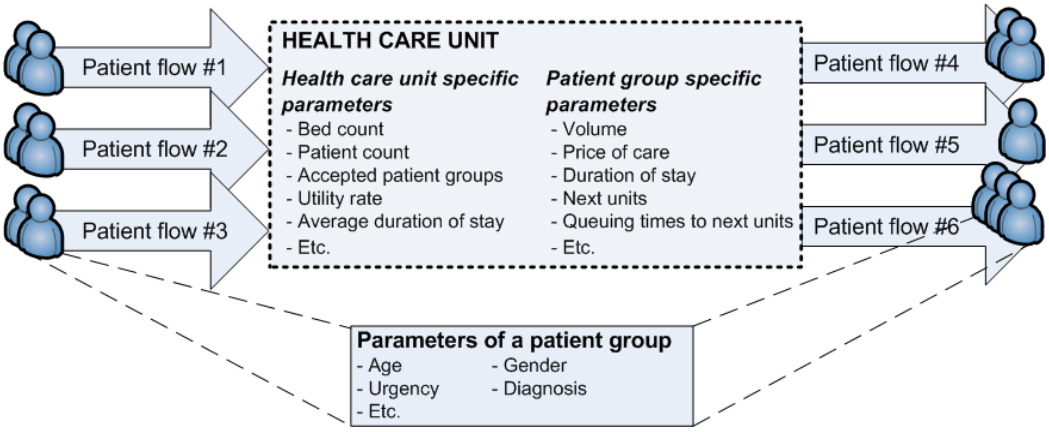

FIGURE 4. The substantial recorded data related to patient flows

specific parameters. On the other hand, some health care unit specific parameters are primary parameters, like bed counts, that are recorded for managing the operations of the hospital.

The health care unit specific parameters together with patient related parameters offer for the management a key to observe and manage the efficiency of the hospital. However, in the complex and turbulent environment with multiple changing parameters, it is difficult and challenging to gain the overall understanding of the service system the hospital offers. Thus, the transformation of the data into the strategic information for supporting the decision-making is a challenging task.

According to our research [3], Hatanpää hospital suffers of insufficient use of the patient flow-related information that is mainly caused by ineffective information utilizing methods. The decision-making is usually based on the simple periodical reports, which do not reveal the overall state of the system. Thereby, the hospital needs new methods and better means for decision-making in order to control and manage patient flows more efficiently. The hospital has indications, like huge data mass and the data warehouse, for optimization and even forecasting the operations, but so far data mining methods have not been used for achieving, for example, better understanding of patients' behavior and hospital's own operations. This has lead to problems, like un-optimized resource allocation and inefficient capacity utilization.

Challenges in the data processing are related to the hospital's internal operation, but also to the inter-organizational activities in the health care network of Tampere. The organizations in the network share enormous amount of data and the shared data structure is so complex that the data processing with traditional methods, like Microsoft Excel, is almost impossible. [3] Limitations of data processing methods of the hospital together with the complex health care environment inflict deficiencies of reporting and measuring the right things. In other words, there is a need of versatile and processed information as supporting the decision-making. One must analyze the patient flows and after that, optimize them by using the gathered information over them.

The current incapability of the process can be illustrated through the following example. When a patient enters the patient flow, patient-related data is being recorded into the Pegasos system (see Fig. 3). At this moment, the system generates pre-determined reports that examine the patient flow at unit-level. This means that the reports only reveal the patient flow parameters (e.g. the duration of the stay) in a specific unit, such as an x-ray unit or examination room, but they do not examine the dependencies between the parameters from separate units. The above setting is examined in the next chapter, in which a selected data mining method is used to examine the whole network instead of

single units. This is done in order to optimize the entire system, not just single units separately.

Potential of a Selected Data Mining Method in Hatanpää Hospital

Currently, as mentioned, the pre-determined reports related to patient flows examine the patient flow only at unit-level. That is, they only reveal the patient flow parameters in a specific unit, but they do not examine the dependencies between the parameters from separate units. The authors propose utilizing a data mining method called **sequence analysis** to improve the usefulness of the patient flow data in order to help the hospital administration making better and more informed decisions in their management activities.

Sequence analysis is an extension of a popular association rules technique [16] where the analyst is faced with a task of finding which items occur together the most in a dataset. This technique is often referred to as market basket analysis as it is usually used in order to find out which items customers purchase in stores [14] [17]. A number of probabilistic figures can be used to gain more insight into the occurrences of the different combinations (rules) of items in a dataset. These probabilistic figures can be applied into sequence analysis techniques as well.

In its most simplistic form, an association rule is a logical statement between two occurrences, such as patient visiting the examination room and visiting the x-ray room [14]. More formally, we can write such a rule as follows:

$$examination \rightarrow x\text{-}ray$$

Most commonly used probabilistic figures describing association rules are support and confidence [14]. The support of a rule is the probability for that specific rule to occur in the dataset, which can be written as follows:

$$support\{examination \rightarrow x\text{-}ray\} = Prob(examination \rightarrow x\text{-}ray) = Prob(examination) * Prob(x\text{-}ray)$$

In the example above, support would reveal the proportion of the patients entering both the examination and x-ray rooms from all the patients recorded in the system. Confidence of a rule is the conditional probability of the event B if event A has occurred. By following the example above, we can write the confidence of a rule as follows:

$$confidence\{examination \rightarrow x\text{-}ray\} = support\{examination \rightarrow x\text{-}ray\} / support\{x\text{-}ray\}$$

In this case, confidence would reveal the probability of a patient entering the x-ray room if he or she also enters the examination room during the hospital stay. By examining the confidences of items in a patient flow, the hospital management would be able to identify the most strained units that may act as bottlenecks in patient flows. By following our example above, the hospital administration could find out that there is a 90 % probability of a patient having a need to visit the x-ray room if the same hospital visit also includes a visit to the examination room. By utilizing such information the positioning and recruiting of the hospital personnel can be optimized by putting emphasis to the most loaded units.

Association rules, however, do not take into account the order in which the patient goes through the patient flow in the hospital. In order to optimize the patient flows, one first needs to describe them. To this end, the authors propose the use of sequence analysis. Sequence analysis takes into account the temporal order in which the items occur in the dataset [14]. In our example it is probable that the visit to the examination room precedes the visit to the x-ray room. By utilizing sequence analysis it is possible to describe the whole patient flow from the first event (entering the hospital) to the last (leaving the hospital). By using the support and confidence figures, it is possible to identify e.g. the most frequent patient flows by examining the binomial probability of entering into a certain unit or not entering it.

Once the most crucial patient flows have been identified and mapped, the next logical step would be optimizing the patient flows so that the operations of the hospital would better serve its customers. At this point the target hospital has yet to begin the patient flow identification, so any optimization activities are purely on a theoretical level, and is heavily dependent on the hospital management's vision, but different data mining techniques could provide answers for the following patient flow-related issues:

- Predicting the patient flows of individual patients based on their symptoms, urgency, and demographic variables
- Identifying possible jams and blockages in patient flows in advance
- Identifying possible alternatives to patient flows for aforementioned situations
- Shorten any waiting periods during a patient's stay in the hospital
- Decrease the cost of a single patient for the hospital.

Conclusions and Future Research

At the beginning of this paper, we showed the importance of understanding the health care system as a network of patient flows in order to manage it

better. Then we presented a theoretical framework for intensifying the patient flow management. Besides, we presented the data mining methodology, which could be used to boost the step where the data is processed into information in order to intensify the management activities. In the empirical part of the paper we examined the patient flow management process in Hatanpää hospital through the presented theoretical framework. After that, we focused on the present challenges in processing the recorded data into strategic information, and brought into attention an example of the current, imperfect process. Finally, we proposed utilizing a data mining method called sequence analysis to improve the use of the patient flow data in order to help the hospital administration making better and more informed decisions.

Based on the literature and our earlier research, our hypothesis was that patient flow management could be intensified by using selected data mining methods. The example presented in chapter 4 revealed that data mining indeed could be a very suitable approach for processing huge amounts of patient flow-related data into useful information. This information helps to understand a complex service system and can, thereby, be used to support decision-making as well as the communication about the patient flow-related matters to all interest groups. As a result, the overall quality and effectiveness of the patient flow network would be improved. This means better clinical outcomes, improved patient safety, greater patient and staff satisfaction, and improved financial performance. On these grounds, we recommend that data mining methods would be more utilized in the referred field.

Despite the many opportunities of data mining in this field, there are also many challenges to conquer. One challenge is the ethicality of patient-related data that hinders the utilization of some data. Furthermore, one of the most common challenges is the quality of data. Insufficient and false records of data influence straightly to relevance and reliability of utilized information. One problem is also the organizational culture of the health care organizations. It has not risen to the level where patient flow-related information could be used effectively in developing the hospital activities.

The scientific contribution of this paper is the clarification of the theoretical concept of patient flow management process (Fig. 2) and its applicability to the real life process (Fig. 3), as well as the illustration of the potential of the data mining methods to intensify the process. The paper has also pragmatic significance as it propose a useful tool-set to the end-users for analyzing the actual data and consequently, new kind of processed information to support the decision-making. Although the study is conducted in a single hospital, the results can be well generalized to other hospitals.

Further research is, however, needed. In future research, the discovered advantages and also challenges of the presented sequence analysis method should be empirically tested. Also the potential offered by other data mining methods, such as neural networks, should be explored.

References

[1] J. Kujala, P. Lillrank, V. Kronström and A. Peltokorpi, "Time-based management o f patient processes," Journal of Health, Or ganisation and Management, vol. 20, pp. 512-524, 2006.

[2] J. Paavilainen and S. Toivonen, "Tampere en yliopistollisen sairaalan logistiset proses sit ja tilaresurssit vuoteen 2020 / Tampere University Hospital's Logistic Processes and Space Resources till 2020," Tampere University of Technology, Tampere, Finlan d, 2008.

[3] J. Väihinpää, J. Paavilainen, J. Myllärnie mi and K. Valkama, "Potilasvirtojen hallin nan tehostaminen Hatanpään sairaalassa / Intensifying Patient Flow Management in Hatanpää Hospital," The City of Tampere, Tampere, Finland, 2009.

[4] M. J. Côté, "Understanding Patient Flow," Decision Line, vol. 3, pp. 8-10, 2000.

[5] C. Haraden and R. Resar, "Patient Flow i n Hospitals: Understanding and Controllin g It Better," Frontiers of Health Services Management, vol. 20, pp. 3-15, 2004.

[6] L. Delesie and L. Croes, "Operations rese arch and knowledge discovery: A data mi ning method applied to health care manag ement," International Transactions in Oper ational Research, vol. 7, pp. 159-170, 200 0.

[7] E. Turban, R. Sharda, J. E. Arondon and D. King, Business Intelligence: A Manag erial Approach. Upper Saddle River (NJ): Pearson Prentice Hall, 2008,

[8] M. W. Isken and B. Rajagopalan, "Data mining to support simulation modeling of patient flow in hospitals," Journal of Me dical Systems, vol. 26, pp. 179-197, 2002.

[9] J. M. H. Vissers, "Patient flow-based allo cation of inpatient resources: A case study, " European Journal of Operational Resear ch, vol. 105, pp. 356-370, 1998.

[10] H. Laihonen, "Terveysjärjestelmän johtamis en tietovirrat / The Knowledge Flows of

Health System Management," 2009.
[11] C. W. Choo, Information Management for the Intelligent Organization: The Art of Scanning the Environment, 3rd ed., Medford (NJ): Information Today, 2002.
[12] B. Gilad and T. Gilad, "A systems approach to business intelligence," Business Horizons, vol. 28, pp. 65-70, 1985.
[13] E. Vitt, M. Luckevich and S. Misner, Business Intelligence: Making Better Decisions Faster. Redmond (WA): Microsoft Press, 2002, pp. 202.
[14] P. Giudici, Applied Data Mining: Statistical Methods for Business and Industry. New York: Wiley, 2003.
[15] D. R. Anderson, D. J. Sweeney and T. A. Williams, Statistics for Business and Economics. 10th ed., Mason (OH): Thomson South-Western, 2009, pp. 1018.
[16] S. Kudyba, Data Mining and Business Intelligencea Guide to Productivity. Hershey (PA): Idea Group Publishing, 2001.
[17] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2nd ed. Amsterdam: Elsevier, 2006, pp. 770.