Journal of Biomedical Informatics 128 (2022) 104038

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Assessment of the feasibility of developing a clinical pathway using a clinical order log

Jungeun Lim [a,1], Kidong Kim [b,1], Minseok Song [a,*], Sooyoung Yoo [c,**], Hyunyoung Baek [c], Seok Kim [c], Somin Park [d], Woo-Jin Jeong [e]

[a] Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang, South Korea
[b] Department of Obstetrics and Gynecology, Seoul National University Bundang Hospital, Seongnam, South Korea
[c] Healthcare ICT Research Center, Office of eHealth Research and Businesses, Seoul National University Bundang Hospital, Seongnam, South Korea
[d] Dept of Quality Improvement & Process Innovation, Seoul National University Bundang Hospital, Seongnam, South Korea
[e] Dept of Otorhinolaryngology-Head & Neck Surgery, Seoul National University Bundang Hospital, Seongnam, South Korea

ARTICLE INFO

Keywords:
Process mining
clinical pathway
feasibility assessment
clinical order log
visualization

ABSTRACT

A clinical pathway (CP) is a tool for effectively managing a care process. There are several research efforts on developing clinical pathways (CPs) in the process mining domain. However, the nature of the data affects data analysis results, and patient clinical variability makes it challenging to develop CPs. Thus, it is crucial to determine candidate care processes that can be standardized as CPs before applying process mining techniques. This paper proposed a method for assessing CP feasibility regarding clinical complexity using clinical order logs from electronic health records. The proposed method consists of data preparation, activity & trace homogeneity evaluations, and process inspection using process mining. Each step consists of metrics to measure the homogeneity of processes and a visualization method to demonstrate the diversity of processes based on the log. The case study was conducted with five surgical groups of patients from a tertiary hospital in South Korea to validate the proposed method. The five groups of patients were successfully assessed. In addition, the visualization methods helped clinical experts grasp the diversity of care processes.

1. Introduction

The care process is complicated and requires interaction among doctors, nurses, patients, and others. It usually evolves differently depending on multiple factors such as the patient's characteristics, medical circumstances, medical professional's judgment, and treatment organization, even for patients with the same disease [1–4]. A clinical pathway (CP) is a tool for effectively managing a care process [1,5].

A CP aims at the standardization of a medical process [2] and tries to improve treatment quality and reduce costs by eliminating unnecessary tasks and avoiding omissions of essential ones [1,2,6,7]. Therefore, a CP offers various positive effects, such as maintaining the quality of medical care, decreasing the hospitalization period, and preventing complications [1,2,6,7]. Owing to these advantages, CPs have been used in the United States since the 1980s [8], and continual efforts have been devoted to developing and utilizing CPs for more patient groups [9–11]. However, the development of a CP is not straightforward since a CP is usually designed manually through multiple discussions based on medical guidelines and the experiences of specialists from multidisciplinary fields. Thus, considerable time and effort are required, and several medical staff should be involved.

Recently, to support CP development, various data-driven approaches have been proposed to analyze data accumulated in hospital information systems (HIS) and derive CPs [12–14]. Process mining is one of the data-driven methods aiming at the effective management of processes [15]. Several attempts were made in the process mining area to analyze medical processes. For instance, Mans et al. analyzed the care flow of kinetic oncology patients using trace clustering and fuzzy miner [16], and Lang et al. analyzed clinical processes using seven process mining techniques, namely α algorithm, α++ algorithm, heuristic miner, DWS algorithm, genetic algorithm, multi-phase miner, and region miner [13].

* Corresponding author at: 77 Cheongam-Ro, Nam-Gu, Pohang, Gyeongbuk, 37673, South Korea.
** Corresponding author at: 166 Gumi-Ro, Bundang-Gu, Seongnam, Gyeonggi, 13620, South Korea.
E-mail addresses: mssong@postech.ac.kr (M. Song), yoosoo0@snubh.org (S. Yoo).
[1] These authors contributed equally to this work and are the first authors.

https://doi.org/10.1016/j.jbi.2022.104038
Received 30 August 2021; Received in revised form 26 January 2022; Accepted 23 February 2022
Available online 3 March 2022

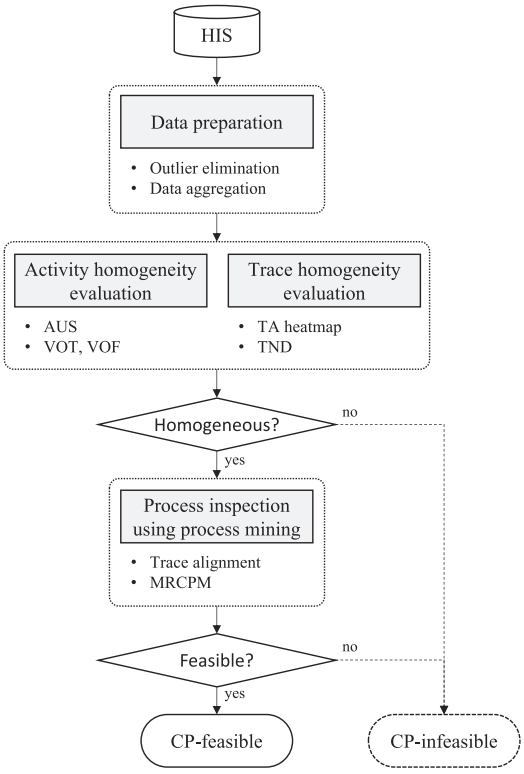

Fig. 1. The overview of the method for CP feasibility assessment. HIS: health information systems, AUO: area under the order set utilization rate curve, VOT: variation of timing, VOF: variation of frequency, TA heatmap: trace abstraction heatmap, TND: trace network density, MRCPM: matching-rate-based CP mining.

In terms of the CP development, Xu et al. used latent dirichlet allocation (LDA) to group activities with the same topics and derived a process model using fuzzy mining [17–19]. Huang et al. focused on finding major process patterns that can be considered when developing CPs and tried various methods such as temporal mining, dynamic programming, and LDA [20,12,21,22]. Chen et al. proposed a distinctive method of extracting a typical treatment process using affinity provision clustering [23] while Cho et al. suggested a greedy optimization algorithm utilizing fitness values, a metric used in conformance checking, to derive CPs [24]. These strategies allow for the discovery of more relevant patterns and the development of more accurate CPs based on evidence. However, exploring and analyzing huge amount of data still need a lot of time and efforts.

Among several kinds of medical processes, there are some processes that cannot be standardized because of their complexity. When several variations are involved in the treatment process of a disease according to patients, developing a standardized clinical pathway of such treatment process becomes extremely difficult. For example, the process models derived in the study of Lang et al. were all spaghetti-like [25]. Spaghetti-like models are useful since they shows reality. However, understanding and standardizing such processes are very challenging. That is, the less structured models are difficult to standardize as CPs. Thus, before developing a CP using a data-driven approach (i.e., process mining), it is useful to check whether a clinical process can be standardized or not by investigating the complexity of the medical data. Instead of deriving process models with process discovery algorithms and exploring the structure of the derived models, the metrics on the complexity of the medical data provide meaningful information on the derived processes.

In process mining, there are studies to evaluate the complexity of event logs [26–29]. The basic metrics are the number of unique events, unique traces, and average trace length [26–28]. Metrics related to process diversity are simple trace diversity [27], advanced trace diversity [27], trace similarity rate [28], and affinity [26]. There are some advanced metrics using the notion of entropy [29]. Such metrics are useful for assessing the diversity of general processes, however they are not suited to evaluating medical processes that contain many concurrent activities.

The focus of this paper was to develop a method for evaluating CP feasibility by assessing the homogeneity of the care processes using event logs in HIS. The method included several homogeneity evaluation metrics and a visualization method for evaluating the complexity of the data from activity and trace perspectives. To validate the proposed method, experiments were conducted with five surgical groups of patients from a tertiary hospital in South Korea. The experiment results showed that the proposed metrics indicate the degree of the feasibility of deriving CPs. It also showed that the method was applicable when a group of patients is more than a hundred.

The remaining part of this paper is organized as follows. Section 2 defined the clinical event log and CP. Section 3 explained the proposed method in detail, and Section 4 presented experiments to validate the method. Section 5 discussed the results and limitations of this study. Finally, Section 6 concluded the paper.

2. Preliminaries

In this section, a clinical event log and a CP were defined. In HIS, there is the order information that contains instructions of medical practitioners to others such as medical staff, pharmacy, and laboratory [30]. A clinical event log from HIS contains detailed information about all the required activities during the care process [31]. A CP is a care process that consists of clinical orders. Note that among several types of care processes, in this paper we focus on surgical processes.

2.1. Clinical event log

A clinical event log L is a set of traces TR, where each trace corresponds to a patient. A trace is a finite sequence of clinical events E. An event $e \in E$ is a record of clinical activities such as delivering a clinical order, performing surgery, hospitalization, and discharge. In this study, only clinical order events were considered as elements composing the clinical event log.

Definition 1. (clinical event). A clinical event is expressed as $e = (o,t)$, where o is the clinical order of e and t is the occurring time of e. The time t represents the relative dates compared to the surgery dates. For example, a day of surgery has 0 as its time value. The time value becomes 1 on the next day and -1 on the day before surgery.

Definition 2. (clinical trace). A clinical trace is expressed as $\sigma = \langle e_1, e_2, \cdots, e_n \rangle$. The sequence $\langle e_1, e_2, \cdots, e_n \rangle$ is a finite nonempty sequence of clinical events, and each event can appear more than once. The time of the sequence is non-decreasing, i.e., for $1 \leqslant i \leqslant j \leqslant n : \pi_t(e_i) \leqslant \pi_t(e_j)$, where π is a function to obtain the value of an attribute recorded for an event.

Definition 3. (clinical event log). A clinical event log is a set of traces. For m patients in the target group for assessing the feasibility of CP development, the clinical event log is expressed as $\mathscr{L} = \{\sigma_1, \sigma_2, \cdots, \sigma_m\}$, where each trace corresponds to a patient.

2.2. Clinical pathway

A typical surgical CP comprises clinical orders from the time of admission to the time of discharge, with the date of the surgery as the control point. The orders are defined daily, and the CP comprises a daily set of clinical orders. The t-th daily set of clinical orders D_t is expressed as $D_t = (t, \{o_1, o_2, \cdots, o_n\})$, where t is the relative date compared to the surgery date, similar to t in Definition 1 and $o_i \in \{o_1, o_2, \cdots, o_n\}$ is a clinical order. For example, $D_{-1} = (-1, \{o_1, o_2, o_3\})$ indicates that the

Fig. 2. Examples of the OU curve. A: OU curve of the target with high order homogeneity; B: OU curve of the target with low order homogeneity. The area of the shaded part is AUO.

orders o_1, o_2, and o_3 should be delivered on the day before the surgery.

Definition 4. (clinical pathway). A CP is expressed by $CP = \langle D_{-n}, \cdots, D_0 \cdots, D_m \rangle$, which is a combination of a daily set of clinical orders for the entire period from $-n$-th day to m-th day compared to the surgery date. For example, $CP = \langle D_{-1}, D_0, D_1 \rangle$ signifies that the patient who is in the target group of the CP should obtain D_{-1} on the day before the surgery, D_0 on the surgery day, and D_1 on the day after the surgery which is the discharge day in this case.

3. Methodology for assessing CP feasibility

This section describes the proposed method (Fig. 1)) to assess the feasibility of CP development. The method consists of data preparation, activity & trace homogeneity evaluations, and process inspection using process mining.

3.1. Data preparation

Clinical processes are inherently complex and heterogeneous. They involve a series of decisions on diseases treatment, and the complexity typically increases as more things require decision-making [32]. First, the decision depends on the clinical condition of a patient and the treatments for each patient could differ [33]. Even if the patient's condition is the same, the treatments may differ depending on the preference of healthcare provider [34]. Furthermore, the process complexity increases further because a treatment does not involve a single person in charge of the entire process but rather a group of people such as doctors, nurses, staff, and carers [35]. Because of this nature of the medical processes, it is difficult to standardize the processes of all patients into one CP. Accordingly, CP aims to standardize the majority of patients, not all patients.

Therefore, we targeted patients except outliers for CP feasibility evaluation and proposed to perform outlier elimination in data preparation. For outlier elimination, the sequence length [36], i.e., the length of hospitalization, could be used since long-term inpatients rapidly increase the diversity of clinical orders and short-term inpatients reduce the importance of typical orders. To eliminate outliers in the length of hospitalization, statistical methods such as the interquartile method, which uses a boxplot, and Z-score, which uses a standard normal distribution could be used [37].

In addition to outlier elimination, event data aggregation is required for CP feasibility evaluation. Event data aggregation is the process of combining order events with low-level abstraction into high-level abstraction [38]. Clinical orders show low-level granularity because they contain detailed instructions from medical practitioners to others such as the medical staff, pharmacy, and laboratory [31,30]. For example, a medication order contains the drug name, start time, delivery route, dosage to be provided each time, and frequency of dosage per day [23]. However, high level information, such as the ingredients or purpose of the drug, is sufficient to test CP feasibility. A test order also contains detailed information. For example, blood tests include several tests, such as human immunodeficiency virus screening tests and hepatitis B blood tests, and each test is directed by different clinical orders. However, the representative of them, i.e., the blood test, is sufficient to test the CP feasibility.

Using the anatomical therapeutic chemical (ATC) classification system[2], the medication orders could be aggregated. The ATC code classifies drugs based on the organs or systems they act upon and their chemical properties. The first digit of the ATC code represents the anatomical group acted on by a drug. The second and third digits represent the pharmacological or therapeutic properties. Therefore, if the first three digits of the ATC code match, the medication orders can be classified as a drug used for the same purpose.

3.2. Activity & Trace homogeneity evaluations

3.2.1. Activity homogeneity evaluation

The purpose of activity homogeneity evaluation is to investigate the generality of clinical orders that compose the event log. If the orders applied to each patient were different, i.e., the orders were not general, standardization would be difficult, then the target group would not be CP-feasible. To assess the generality of the orders, three metrics were developed: *area under the order set utilization rate curve (AUO)*, *variation of timing (VOT)*, and *variation of frequency (VOF)*.

Area under the order set utilization rate curve The AUO evaluates the diversity of orders. Intuitively, the diversity of orders can be evaluated using 1) the percentage (w.r.t. all orders) of the orders that appear in most of the traces or 2) the average utilization of orders. However, it was found that these two metrics were insufficient for evaluating the diversity of orders in real-world care processes. Even when the target was CP-feasible, the ratio of common orders was substantially small, making the distinction of CP-infeasible and CP-feasible cases difficult. The average order utilization also could not provide a comprehensive assessment of the generality. For example, it was assumed that there were ten patients and two orders o_1 and o_2, each of which was used by eight patients. Then, the utilization rate of each order was 80% and when the utilization of two orders with the average utilization rate was assessed, their generality was determined to be approximately 80%. However, even if each order was used by eight patients, there could be only six patients who utilized both of the orders; in the instance where two patients used o_1, two patients used o_2, and six patients used both. That is, in this case, generality calculated by the average order utilization was 80%, but the actual generality of the two orders was 60%. Therefore, the AUO metric to evaluate the generality of order sets was developed.

The AUO denotes the area under the order set utilization rate (OU) curve (Fig. 2). The OU curve plots the utilization rate of the order set as the elements of the order set change. Starting with the entire order set, the least used order was excluded. The importance of the excluded order (x-axis in Fig. 2) and the utilization rate of the order set after exclusion (y-axis in Fig. 2) were then calculated and plotted to the graph. Thereafter, the next least used order continued to be excluded, calculating the x and y values, and plotting the graph; this step was repeated until the entire order set became empty.

The importance of the order was calculated as follows:

$$I\left(o_k\right) = r\left(\left\{o_k\right\}\right) \Big/ \sum_{i=1}^{m} r\left(\left\{o_i\right\}\right), \tag{1}$$

where $O = \{o_1, o_2, \cdots, o_m\}$ be the entire set of orders from the clinical order log, $I(o_k)$ be the importance of order k, and $r(S)$ be the proportion

[2] https://www.whocc.no/atc/structure_and_principles/

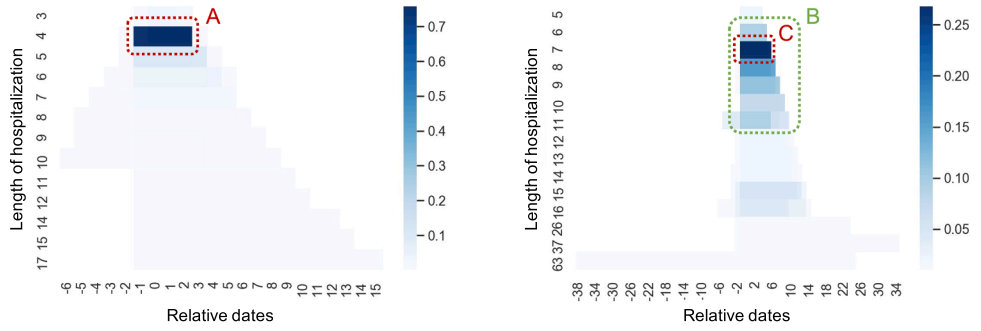

Fig. 3. Examples of the TA heatmap. (a): very homogeneous; the darkest area A: > 70% of patients, no other dark area but A. (b): not so homogeneous; the darkest area C: ≈ 25% of patients, dark area B ($C \subset B$).

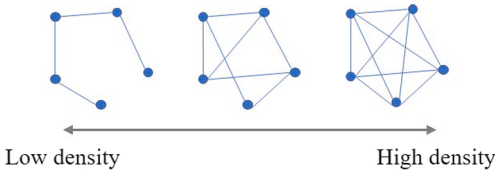

Fig. 4. Examples of trace network of five patients.

of patients who use order subset $S \subset O$.

After the importance of the orders was obtained, they were sorted in descending order. Then, the AUO was calculated as

$$AUO = \sum_{i=0}^{m-1} I\left(o'_{m-i}\right) r\left(\bigcup_{j=1}^{m-i} o'_j\right), \tag{2}$$

where $O' = \{o'_1, o'_2, \cdots, o'_m\}$ is the sorted order set. The AUO value increased as the ratio of important orders and the ratio of patients using the orders increase. In Fig. 2, the AUO of graph A was higher than that of graph B, since the ratio of important orders was higher and the ratio of patients using the orders increased at a faster rate in graph A.

Variation of timing The VOT is a metric used to evaluate the diversity of times at which orders are used, which was not considered in the AUO. This metric was required because the CP standardized the tasks along the timeline. The VOT was defined as follows:

$$VOT = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{l}\sum_{j=1}^{l}\left(\frac{2}{n}x_{ij} - 1\right)\cdot 1_{x \geqslant \frac{n}{2}} + \left(-\frac{2}{n}x_{ij} + 1\right)\cdot 1_{x < \frac{n}{2}}, \tag{3}$$

where $O = \{o_1, o_2, \cdots, o_m\}$ is a set of orders and $D = \{d_1, d_2, \cdots, d_l\}$ is a set of relative dates, x_{ij} is the number of patients who used order o_i on relative date d_j, and n is the total number of patients. When x_{ij} is close to n or 0, the VOT of an order is close to 1. When x_{ij} is neither large nor small, i.e., close to $n/2$, the VOT of the order becomes nearly 0. The VOT of the entire order is the average of the VOT of each order.

For example, when there is a log $\mathscr{L} = \{\langle(o_2,-1),(o_1,0),(o_2,0)\rangle^3, \langle(o_1,0),(o_2,0),(o_2,1)\rangle^3, \langle(o_1,0),(o_2,0),(o_2,2)\rangle^3, \langle(o_1,0),(o_2,0)\rangle\}$, VOT of order $o_1 = \frac{1}{4}\left(\left(-\frac{2}{10}\cdot 0+1\right)+\left(\frac{2}{10}\cdot 10-1\right)+\left(-\frac{2}{10}\cdot 0+1\right)+\left(-\frac{2}{10}\cdot 0+1\right)\right) = 1$ and VOT of order $o_2 = \frac{1}{4}\left(\left(-\frac{2}{10}\cdot 3+1\right)+\left(\frac{2}{10}\cdot 9-1\right)+\left(-\frac{2}{10}\cdot 3+1\right)+\left(-\frac{2}{10}\cdot 4+1\right)\right) = 0.45$, so VOT of log $\mathscr{L}$ becomes 0.725.

Variation of frequency The VOF is a metric used to evaluate the homogeneity of the number of times an order has been used. Using the same order at a different frequency for each patient can also be an obstacle for standardization; hence, the VOF should be evaluated. The VOF was defined as follows:

$$VOF = \frac{1}{m}\sum_{i=1}^{m} max_{j\in\{1,\cdots,l\}}\left(\frac{1}{n}\sum_{k=1}^{n}\left(y_{ij}^{k} - \overline{y_{ij}}\right)^2\right), \tag{4}$$

where y_{ij}^{k} is the number of times patient p_k used order o_i on day d_j. The VOF is the average of the largest standard deviation of each order; because a frequency deviation in one day affects the standardization of the process. Contrary to other metrics, a small VOF indicates a high feasibility of CP development.

For example, when there is a log $\mathscr{L} = \{\langle(o_1,-1),(o_2,-1),(o_1,0),(o_2,0)^3,(o_1,1),(o_2,1)\rangle, \langle(o_1,-1)^2,(o_2,-1)^2,(o_1,0),(o_2,0)^7\rangle, \langle(o_1,-1),(o_2,-1)\rangle, \langle(o_1,-1),(o_2,-1),(o_2,0)^5\rangle, \langle(o_1,-1),(o_2,-1),(o_2,0)^2\rangle\}$, VOF of order $o_1 = max(std([1,2,1,1,1]), std([1,1,0,0,0]), std([1,0,0,0,0])) = 0.490$ and VOF of order $o_2 = max(std([1,2,1,1,1]), std([3,7,0,5,2]), std([1,0,0,0,0])) = 2.417$, so VOF of log $\mathscr{L}$ becomes 1.453.

3.2.2. Trace homogeneity evaluation

The purpose of trace homogeneity evaluation is to examine the variation in the processes experienced by target patients (i.e., the variety of the traces). Herein, a visualization method was proposed, *trace abstraction heatmap (TA heatmap)*, and a metric, *trace network density (TND)*.

Trace abstraction heatmap The TA heatmap (Fig. 3) is a visualization method for evaluating the similarity of the trace length (i.e., the length of hospitalization) and the similarity of the occurrence time of major events (i.e., the time interval between admission and surgery). To roughly grasp the similarity of the traces, the TA heatmap is effective.

The heatmap was drawn by coloring the cells corresponding to each date (x-axis) during hospitalization and the length of hospitalization (y-axis) for each patient. The dates during the hospitalization were relative dates based on the surgery date. For example, if a patient was admitted to a hospital on January 19, underwent surgery on January 20, and was discharged on January 22, the relative dates of admission and discharge would be −1 and 2, respectively, and the length of hospitalization would be 4. Then, the cells of $(-1,4),(0,4),(1,4)$, and $(2,4)$ were colored. The traces of patients who showed the same time interval between admission and surgery and the same length of hospitalization were colored on the same cells. Consequently, the homogeneity of the trace based on the intensity of the color could be assessed.

In Fig. 3, for example, the graph (a) was considerably homogeneous because only a narrow area (A) was dark and > 70% of patients were in this area. However, the graph (b) was not homogeneous because the dark area (B) was wide and only ≈ 25% of patients were in the darkest area (C).

Trace network density While the heatmap roughly evaluates trace homogeneity using only the trace length and time of the major event, the TND evaluates homogeneity by directly comparing traces between

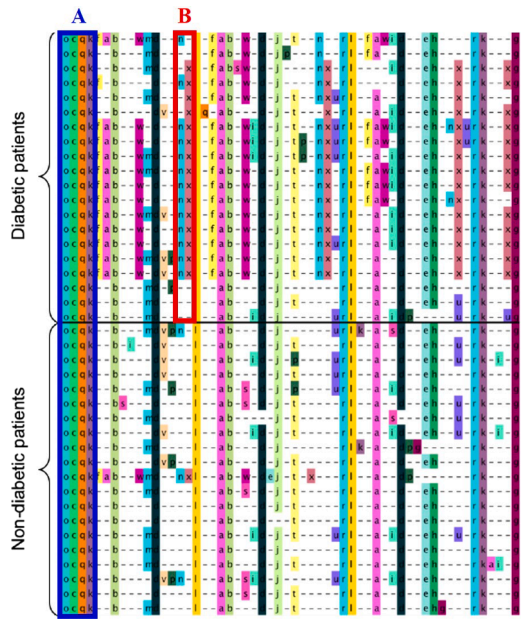

Fig. 5. Example of trace alignment in an analysis of the process in the medical field: Trace alignment results of diabetic patients and non-diabetic patients [43]. A: orders common in both diabetic and non-diabetic patients, B: orders common only in di.abetic patients.

patients. A network graph was used to compare all traces. A node of the graph represents a trace of a patient, and two nodes were connected when the similarity between two order traces exceeded a certain threshold. Then, the homogeneity of traces was quantified using network density that was defined as the number of connections divided by the total possible connections. The number of connections increased as the pair of patients with similar traces increased, and the TND increased as the number of links increased (Fig. 4).

The similarity of traces were evaluated from various viewpoints, such as the activity and originator [39]. From the activity perspective, the similarity between order vectors was calculated using the jaccard distance [40]. The order vector represents the order occurrences in each patient's trace among all orders in the event log. For example, if the order set in the event log was $O = \{o_1, o_2, o_3, o_4, o_5\}$ and only o_2 and o_5 appeared in a patient's trace, the patient's order vector was expressed as $v = (0,1,0,0,1)$.

The TND varies depending on the criterion (i.e., threshold) used to determine whether to connect the nodes. Selecting an appropriate threshold requires a preliminary analysis based on a comparison between CP-feasible and CP-infeasible data. In the experiment, it was found that acceptable results were obtained when the threshold was around 70%.

3.3. Process inspection using process mining

In this step, the process was verified using process mining techniques for the target determined in the previous step to ensure a high homogeneity of the activity and trace. The previous steps enable to reduce the

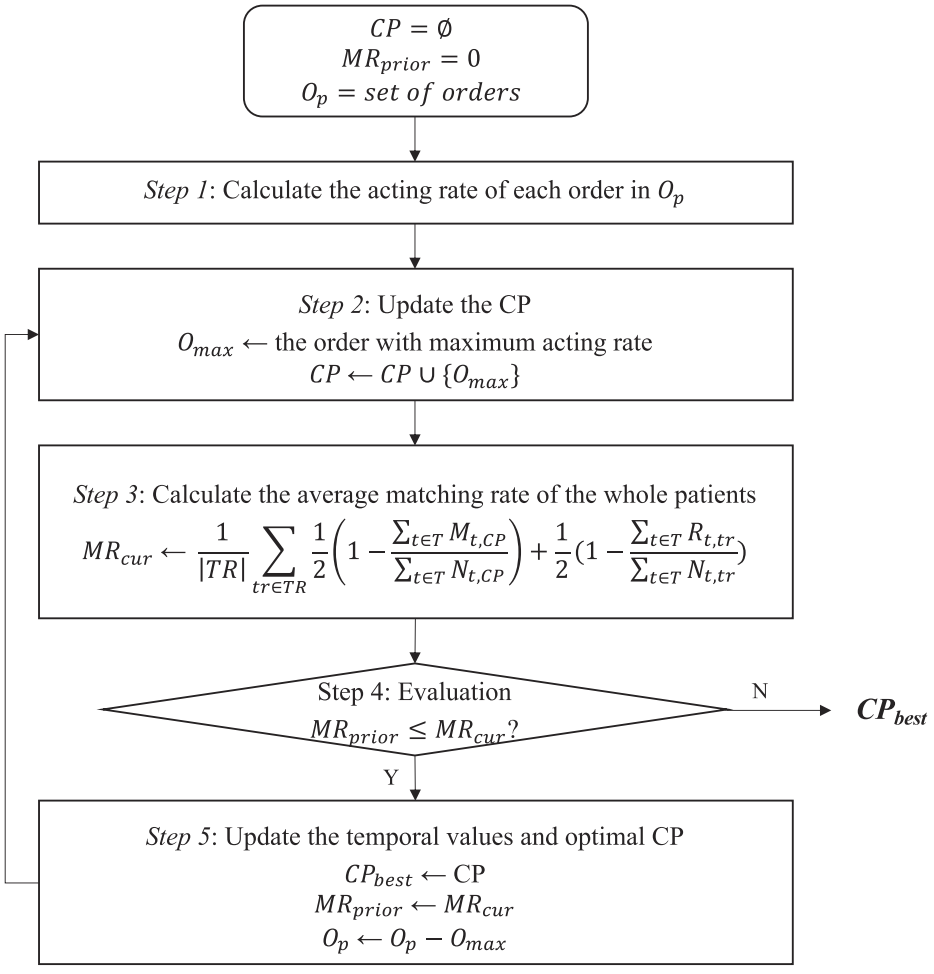

Fig. 6. The flow diagram of MRCPM algorithm.

Table 1
Summary of data.

Target group	GS51	OL21	OG3	JRC15	OG4
Patients: total	135	220	1030	522	646
Orders	521	347	347	323	295
Patients: no CP	2 (1.48%)	7 (3.18%)	0	2 (0.38%)	1 (0.15%)
Patients: stop CP	69 (51.11%)	74 (33.64%)	147 (14.27%)	18 (3.45%)	20 (3.10%)
Patients: complete CP	64 (47.41%)	139 (63.18%)	767 (85.73%)	500 (96.17%)	542 (96.75%)
CP-feasible level	poor	fair	fair	good	good

Table 2
Summary of prepared data.

Target group	GS51	OL21	OG3	JRC15	OG4
Patients: total	127	204	781	509	542
Orders	94	80	78	83	67
Patients: no CP	2 (1.57%)	4 (1.96%)	-	-	-
Patients: stop CP	61 (48.03%)	61 (29.90%)	14 (1.79%)	9 (1.77%)	-
Patients: complete CP	64 (50.39%)	139 (68.14%)	767 (98.21%)	500 (98.23%)	542 (100%)
CP-feasible level	poor	fair	good	good	good

number of candidate processes by calculating the metrics. The aim of this step is providing useful (visual) information to medical staffs to help them determining the CP-feasibility.

Using trace alignment and process model discovery, the similarity between traces and the expected process model was visualized. The CP development used matching-rate-based CP mining (MRCPM) [24]. The CP feasibility was examined more in detail based on the validation of medical experts on the CP derived using the MRCPM algorithm, and the derived CP was used as the reference model. Even if the CP development is sufficiently assessed to be possible in the previous step, if the visualization result was complex and the CP derived using the MRCPM algorithm was invalid, the target was determined to be CP-infeasible.

3.3.1. *Visualization using trace alignment and process map discovery*

Trace alignment aligns traces while minimizing the substitution, insertion, and deletion of activities [41]. Therefore, trace alignment can be used to visually evaluate the information [42] that is difficult to grasp using only the TND value, such as identifying the commonly used and uncommonly used orders (Fig. 5). As the number of common orders decreased and the number of individual orders increased. The feasibility of CP development is low when the total aligned length is long.

A process model can also be used to determine the feasibility of CP development. Complex and varying traces yielded spaghetti-like models that were difficult to understand. Therefore, if the derived process model was excessively complex to comprehend, it could be concluded that the feasibility of the CP development was low.

3.3.2. *CP development using matching-rate-based clinical pathway mining*

The MRCPM algorithm discovers the most suitable CP from the order traces of patients using the matching rate [44]. The matching rate is a metric to represent the conformance between CP and an order trace; this

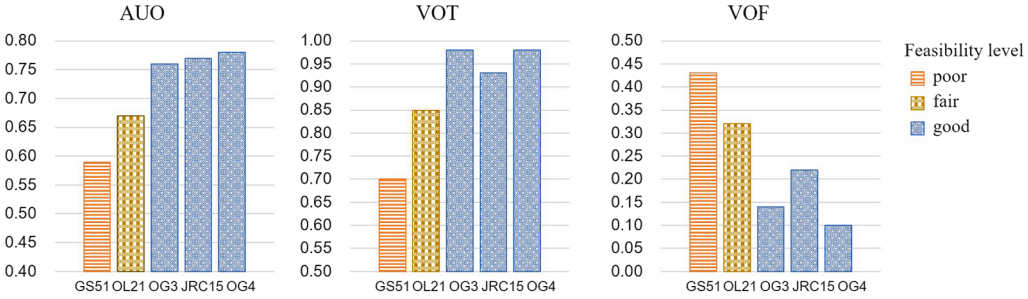

Fig. 7. AUO, VOT, and VOF values of GS51, OL21, OG3, JRC15, and OG4.

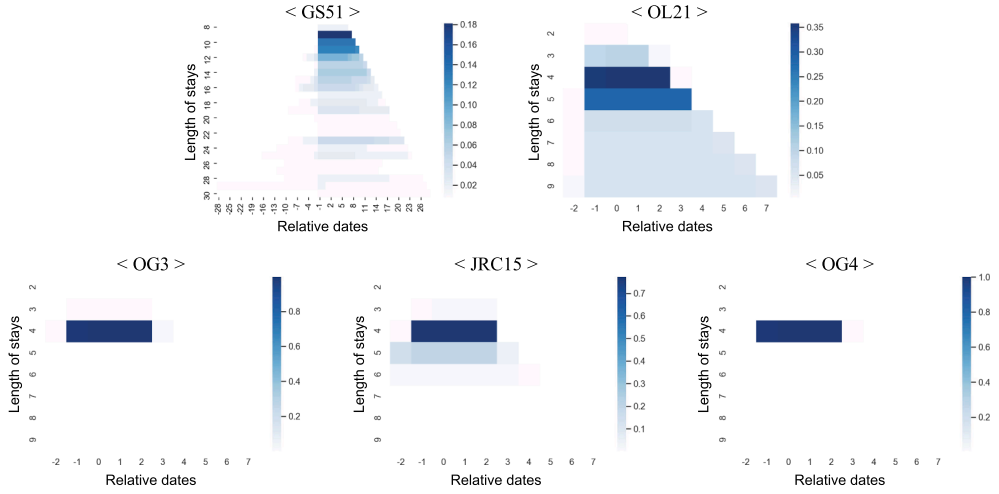

Fig. 8. TA heatmaps of GS51, OL21, OG3, JRC15, and OG4. Darkness of color: the relative magnitude of patient proportion.

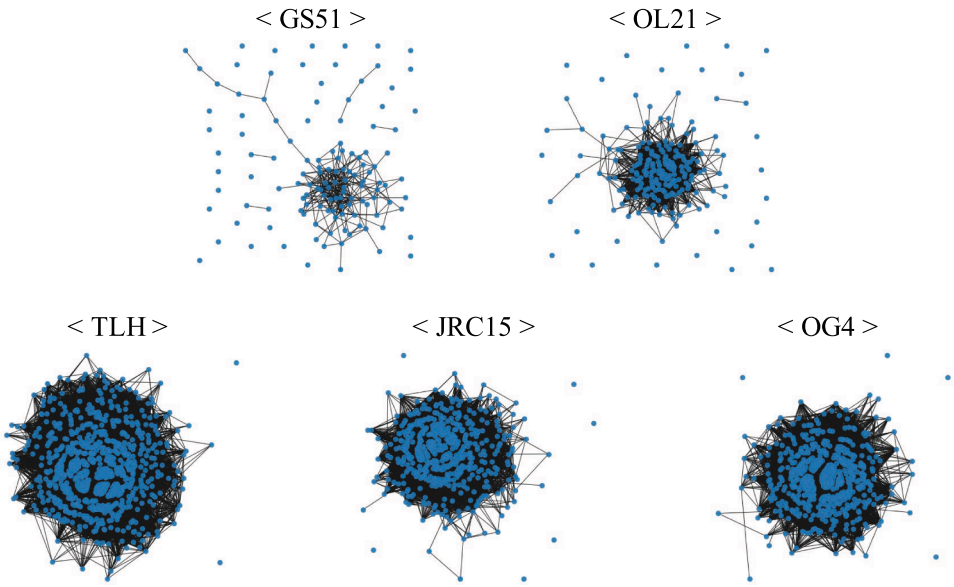

Fig. 9. Trace network graphs of GS51, OL21, OG3, JRC15, and OG4. Edges were connected when the trace similarity between two nodes is more than 75%.

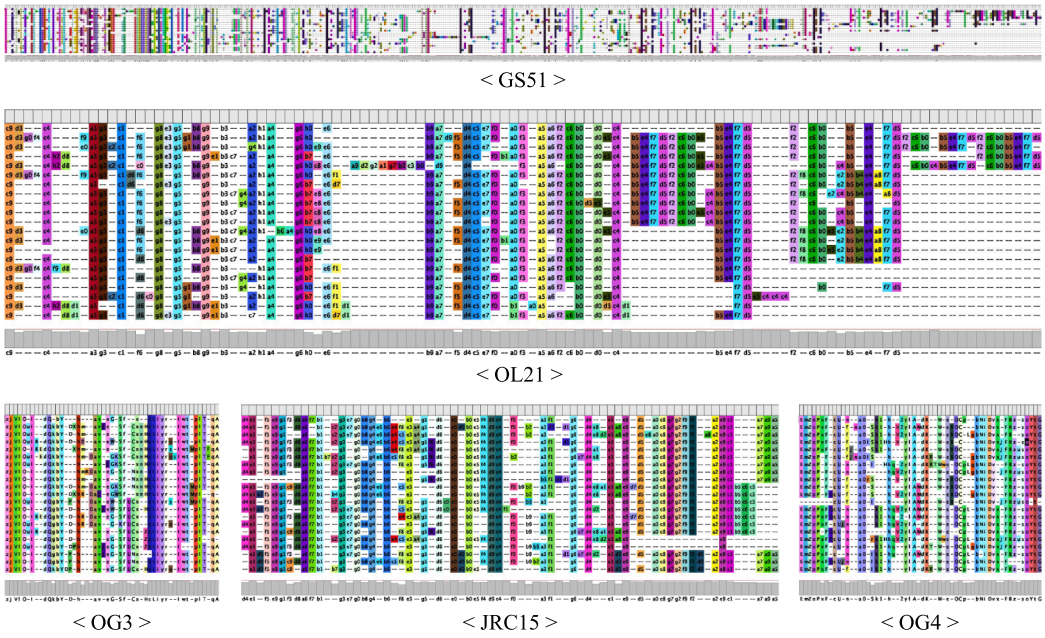

Fig. 10. Trace alignment results of GS51, OL21, OG3, JRC15, and OG4 from randomly sampled 20 patients' traces.

metric is used as the fitness measure in conformance checking [45]. The CP derived using the MRCPM algorithm can be used as a reference for CP development after the feasibility evaluation of an expert. The MRCPM algorithm derives an optimal CP in a greedy way choosing the best order in each step [24]. The optimal CP started with an empty set; at this time, the average matching rate was zero. To compose the optimal set, the best action was choosing the most utilized orders. Thus, the most utilized order was continuously added to the optimal set at each step, and the average matching rate kept increasing at some point. However, the average matching rate started to decrease after the point because an order that is not used by the majority of the patients starts to be included in the optimal set. Orders stopped being added to the set after that, and the set became the optimal CP. The detailed procedure of the MRCPM algorithm [24] is shown in Fig. 6.

4. Evaluation

Several experiments were performed to validate the proposed method. First,the proposed method was applied to real-world data with different levels of CP feasibility and compared the results. Second, an experiment was conducted using synthetic data to verify the feasibility representation ability of metrics in the proposed method. The degree of validity was changed ten times and the changes in the metric values were observed. Finally, experiments were performed to determine the appropriate number of sample data to apply the presented metrics.

4.1. Data

The proposed method was assessed using patients' order data from the HIS in a tertiary hospital in South Korea. After discussion with

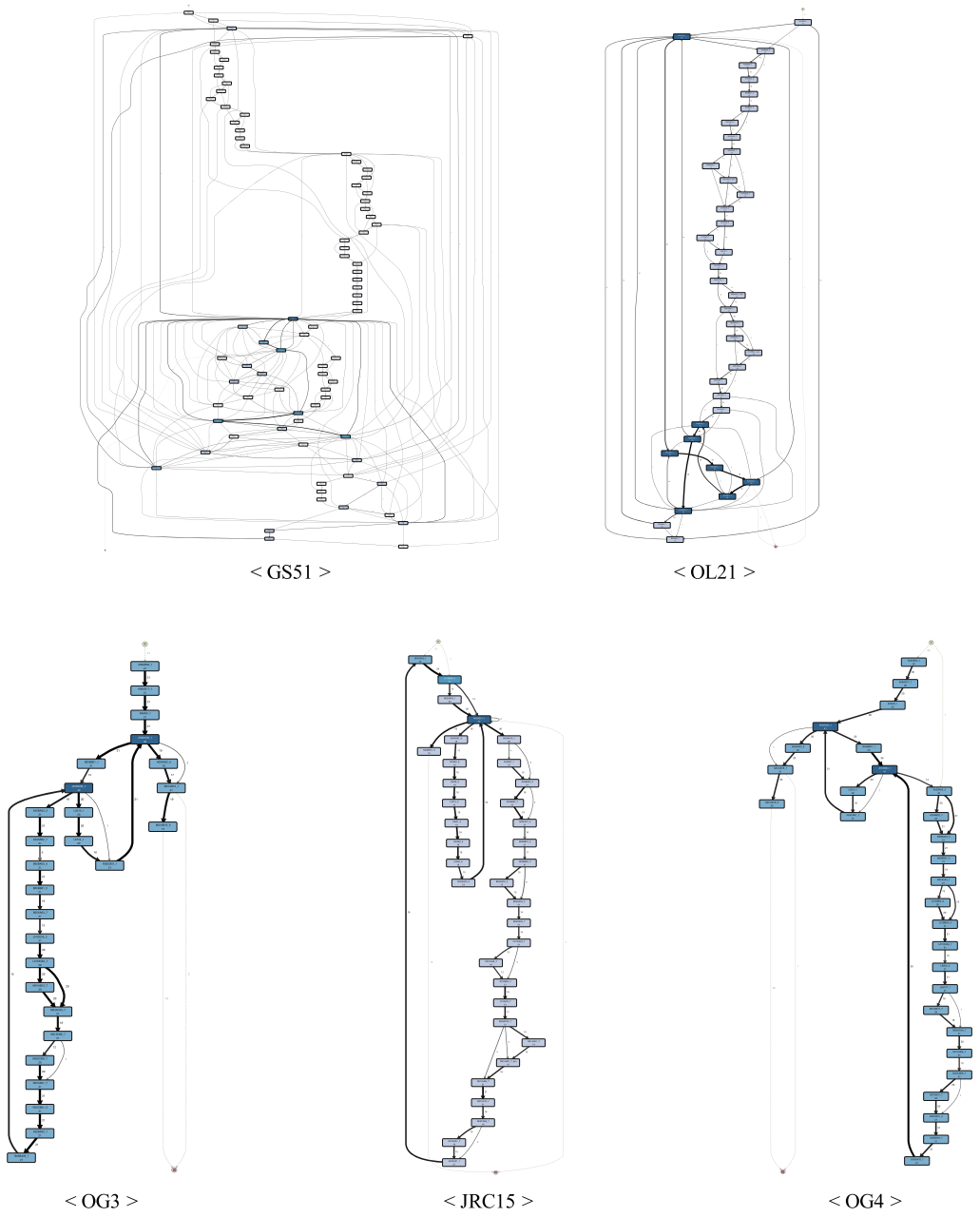

Fig. 11. Process models of GS51, OL21, OG3, JRC15, and OG4 discovered by Disco (activities threshold: 50%, paths threshold: 50%).

medical experts in the hospital, five surgical CPs were selected, named GS51, OL21, OG3, JRC15, and OG4. The selection was based on interruption rates. The interruption rate referred to the percentage of patients who stopped following CP. The interruption occurred since the CP did not provide proper orders for a patient. It is known that clinical processes with high interruption rates have more variations and are difficult to standardize.

GS51 is a nine-day CP for patients receiving pancreaticoduodenectomy, known as the Whipple procedure. The operation removes the head of the pancreas and the first part of the small intestine (duodenum), including the gallbladder and bile duct. OL21 is a five-day CP for patients receiving minor head and neck surgery, primarily for parotidectomy, submandibular gland resection, lipoma excision, and glossectomy, which the otolaryngology department categorized as relatively simple. OG3 is a four-day CP made by obstetrics and gynecology. It is applied to patients receiving laparoscopic surgery when a patient undergoes total laparoscopic hysterectomy (TLH) operation for uterine myoma. JRC15 CP is a four-day CP for patients with shoulder injuries requiring rotator cuff repair or superior labrum anterior to posterior (SLAP) repair. OG4 is similar to OG3; it is a four-day CP for patients receiving laparoscopic surgery. However, OG4 CP is applied to pelviscopy procedures for removing benign neoplasms. Hereafter, each group of patients was called as the name of the target CP.

For GS51, OL21, JRC15, and OG4, the clinical records between December 2018 and December 2019 were extracted; for OG3, the clinical records between January 2012 and April 2016 were extracted. Table 1 shows the summary statistics. The number of patients for GS51, OL21, OG3, JRC15, and OG4 were 135, 220, 1030, 522, and 646, respectively. The number of orders used for GS51, OL21, OG3, JRC15, and OG4 were 521, 347, 347, 323, and 295, respectively; here, orders that were used for only one patient were excluded. The interruption rates for GS51, OL21, OG3, JRC15, and OG4 were 51.11%, 33.64%, 14.27%, 3.45%, and 3.10%, respectively. According to the interruption rates, five sets were rated into three levels in the CP feasibility: GS51 as

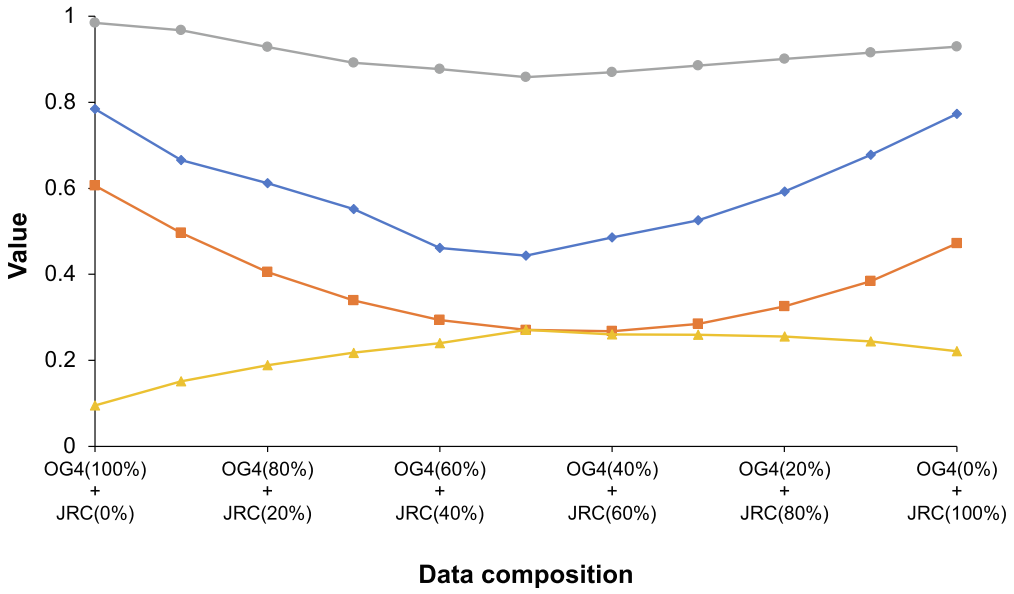

Fig. 12. AUO, VOT, VOF, and TND vs. data composition ((100-x)% of OG4 + x% of JRC15). The x increases by 10. The values were derived by averaging 100 results from random sampling. ◆: AUO, ●: VOT, ▲: VOF, ■: TND.

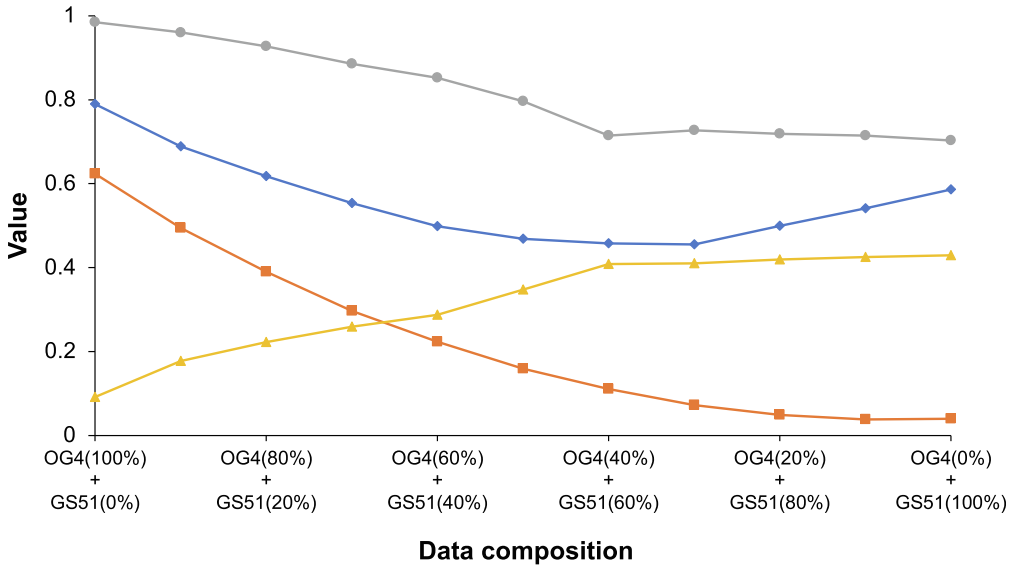

Fig. 13. AUO, VOT, VOF, and TND vs. data composition ((100-x)% of OG4 + x% of GS51.). The x increases by 10. The values were derived by averaging 100 results from random sampling. ◆: AUO, ●: VOT, ▲: VOF, ■: TND.

poor, OL21 and OG3 as fair, and JRC15 and OG4 as good.

Note that this study was approved (IRB No. B-1609–361-105) by the Institutional Review Board of the Seoul National University Bundang Hospital, which waived patients' informed consent. All EHR data were deidentified and then provided to the researchers for this study.

4.2. Method validation using real-world data

We applied the proposed method to the five surgical groups and examined whether the results are consistent with the feasibility level.

4.2.1. Data preparation

The extracted data were prepared for CP feasibility assessment as described in Section 3.1. Table 2 shows the summary statistics of data after preparation. The number of patients for GS51, OL21, OG3, JRC15, and OG4 were 135, 220, 1030, 522, and 646. The number of orders used for GS51, OL21, OG3, JRC15, and OG4 were 94, 80, 78, 83, and 67; here, orders that were used for only one patient were excluded. The interruption rates for GS51, OL21, OG3, JRC15, and OG4 were 49.03%, 29.90%, 1.79%, 1.77%, and 100%; therefore, we classifiedGS51 as poor, OL21 as fair, and OG3, JRC15, and OG4 as good to implement a CP.

OG3 became 'good' from 'fair' after the data preparation. This is because OG3 contained several outliers that could not be standardized as a CP. If the outliers are removed by preprocessing, a CP can be developed from the remaining patients. In the field, CP development is conducted targeting only about 70–80% of patients since there are many outliers in most medical processes. Thus, the ground truth level of each dataset was set after the data preparation; i.e., GS51 as 'poor,' OL21 as 'fair,' and OG3, JRC15, and OG4 as 'good.'.

4.2.2. Activity & Trace homogeneity evaluation

The feasibility of each target in terms of activity and trace homo-

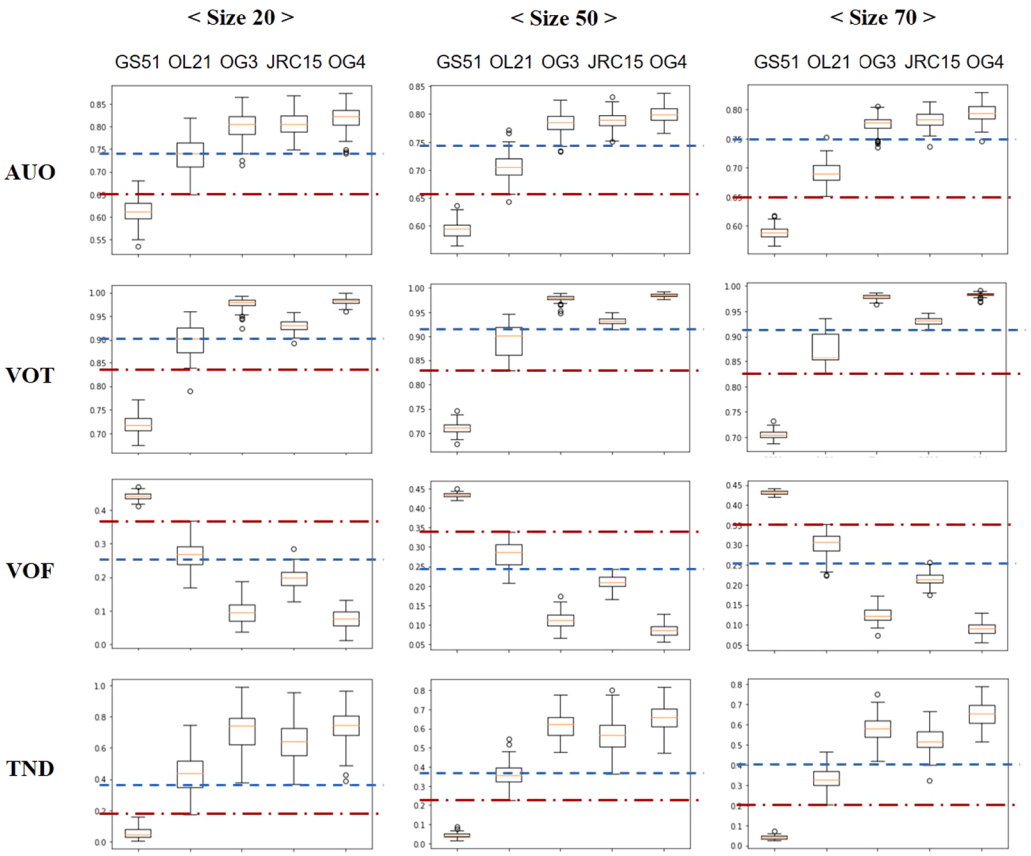

Fig. 14. Box plots of AUO, VOT, VOF, and TND of each CP group when the sample size was 20, 50, and 70 (n = 100). – –: The smallest whisker of AUO, VOT, and TND or the largest whisker of VOF in good feasibility groups, – • –: The smallest whisker of AUO, VOT, and TND or the largest whisker of VOF in the fair feasibility group.

geneity was evaluated. First, to assess the activity homogeneity, AUO, VOT, and VOF were calculated and the results were compared between GS51, OL21, TLH, JRC15, and OG4. The AUO and VOT increased, and VOF decreased as the feasibility increased (Fig. 7). The AUO values of GS51, OL21, OG3, JRC15, and OG4 were 0.59, 0.67, 0.76, 0.77, and 0.78, respectively (GS51 < OL21 < OG3 < JRC15 < OG4). The VOT values of GS51, OL21, OG3, JRC15, and OG4 were 0.70, 0.85, 0.98, 0.93, and 0.98, respectively (GS51 < OL21 < JRC15 < OG3 = OG4). The corresponding VOF values were 0.43, 0.32, 0.14, 0.22, and 0.10 (GS51 > OL21 > JRC15 > OG3 > OG4).

Second, using the TA heatmap and TND metric, the trace homogeneity of GS51, OL21, OG3, JRC15, and OG4 were compared. The similarity of the length of stay (LOS) and time interval (TI) between admission and surgery increased as the CP feasibility increased. In Fig. 8, GS51 has the broadest range of dark areas, and OL21 has a narrower range of dark regions than GS51, and OG3, JRC15, and OG4 have the densest dark area, which indicated that more patients of OG3, JRC15, and OG4 had similar LOS and TI between admission and surgery. Patient proportions of the darkest areas were also discriminative: GS51 was 0.181, OL21 was 0.358, OG3 was 0.997, JRC15 was 0.772, and OG4 was 1.0.

The TND increased as the feasibility increased: 0.04 for GS51, 0.30 for OL21, 0.53 for OG3, 0.47 for JRC15, and 0.60 for OG4. The trace network graph intuitively showed the difference (Fig. 9). GS51 and OL21 have many disconnected nodes, which were sparsely connected, whereas OG3, JRC15, and OG4 have few unconnected nodes, which were tightly connected.

4.2.3. Process inspection using process mining

The complexity of the actual process was visualized using process mining. First, trace alignment was used to align the traces in the event log and compared how similar the traces were. A total of 20 patients were randomly selected for each CP and aligned traces composed of orders whose utility rate was over 10% (Fig. 10). The traces of GS51 were poorly aligned; the aligned full length was 481. It was much more than the number of orders each patient used during the hospitalization; average: 170.0, min: 119, and max: 221. OL21 was aligned better than GS51. The aligned full length was 111, while the average number of orders each patient used was 52.0 (min: 36, max: 73). The aligned full lengths of OG3, JRC15, and OG4 were much shorter than GS51 and OL21. The aligned lengths were 51 for OG3, 72 for JRC15, and 58 for OG3, though the average number of orders (min, max) in each group was 37.3 (34, 42), 36.0 (33, 53), and 41.7 (30, 49).

Here, we presented the trace alignment results for 20 patients due to space constraints. Still, it was similar to the trace alignment results for all patients, indicating that the more feasible the CP, the better the trace alignment.

Second, process models were derived using the same data in the trace alignment (Fig. 11). Disco [46], a popular tool for process discovery was used. The threshold setting used for process discovery was 50% of activities and 50% of paths. The complexity of the process map was the same as the result of the trace alignment. The map of GS51 was most complicated, the map of OL21 was less complicated than GS51, and the maps of OG3, JRC15, and OG4 were much more straightforward than GS51 and OL21, although the process map was derived on the same setting.

Finally, using MRCPM algorithm, CPs were discovered and the feasibility level was compared with the expected matching rate from the derived CPs. The average matching rate increased as the CP feasibility increased: 65.1 for GS51, 73.8 for OL21, 83.4 for OG3, 81.0 for JRC15,

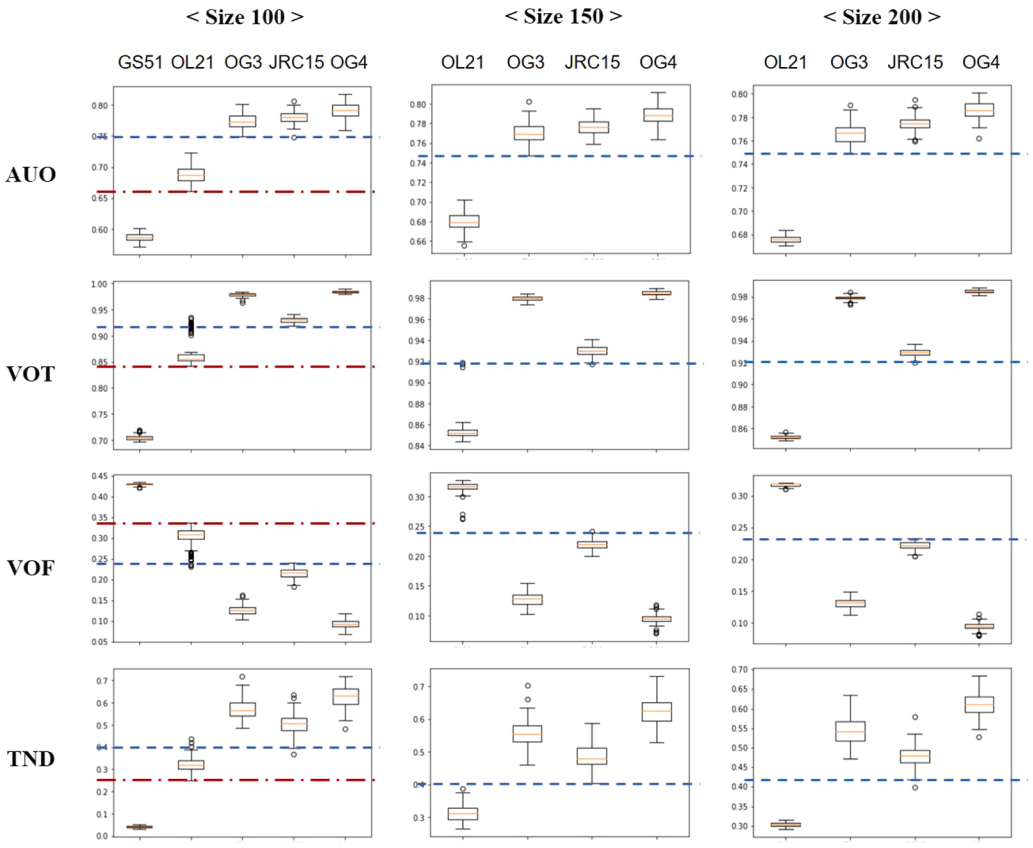

Fig. 15. Box plots of AUO, VOT, VOF, and TND of each CP group when the sample size was 100, 150, and 200 (n = 100). – –: The smallest whisker of AUS, SPD, and VOT or the largest whisker of VOF in good feasibility groups, – • –: The smallest whisker of AUS, SPD, and VOT or the largest whisker of VOF in the fair feasibility group.

and 85.1 for OG4.

4.3. Metric validation using synthetic data

In addition to experiments with real-world data, metric validation was performed using synthetic data to verify the feasibility representation ability of metrics. Previously, the results of each step of the method was compared according to the level of feasibility by dividing the five real-world datasets into three feasibility levels, whereas, here, 11 datasets were synthesized with different activity and trace homogeneity and the values of AUO, VOT, VOF, and TND were compared.

Two experiments were conducted. First, two homogeneous groups, OG4 and JRC15, were synthesized. To generate 11 datasets, the order logs of two patient types were combined: $(100-x)\%$ from OG4 + $x\%$ from JRC15 ($x = 0$ to 100 by 10) summing up the total 300 traces. Second, homogeneous group, OG4, and heterogeneous group, GS51, were synthesized. The datasets were synthesized in the same way as synthesis of OG4 and JRC15, aside from the size of each dataset. GS51 had only 127 traces, which was not enough to generate 300 traces. Therefore, in the second experiment with synthetic data, each dataset was synthesized to have 120 traces. The total traces were randomly selected 100 times and the metric values were derived by averaging 100 results.

In the first experiment, AUO, VOT, and TND increased, and VOF decreased as the graph went to both ends, which were homogeneous since the dataset comprised only order log of OG4 or JRC15 (Fig. 12).

In the second experiment, the values of AUO, VOT, and TND tended to decrease, and VOF tended to increase, as the percentage of GS51 increased (Fig. 13). After the percentage of GS51 was over 50%, the increasing or decreasing size became smaller and AUO increased as the percentage of GS51 increased.

4.4. Sample size test for valid evaluation

Lastly, a sample size test was conducted to inspect the effect of data size on suggested metrics. Samples of size 20, 50, 70, 100, 150, and 200 traces were randomly selected from each group of GS51, OL21, OG3, JRC15, and OG4 100 times, and AUO, VOT, VOF, and TND were calculated, and then they were compared. Note that GS51 was not tested with a data size of 150 and 200, since the total traces of GS51 were only 127. During sampling, the ratio of patients who did not use the CP, who stopped following the CP partway through, and who completed a CP remained the same.

The proposed metrics were reliable when the sample size was over 100, i.e., the traces in the order log was over 100. In other words, AUO, VOT, VOF, and TND can distinguish groups with different feasibility levels (i.e., poor, fair, and good level) when the sample size was over 100. In Fig. 14, the whiskers of the poor, fair, and good groups overlapped when the sample size was less than 100; however, in Fig. 15, the whiskers did not overlap when the sample size was over 100.

4.5. Comparison with other log complexity metrics

A comparative evaluation of the proposed metrics was conducted with the 19 metrics of Back et al. [29]. As a result, three metrics, global block entropy (GB), difference-Based k-block entropy rate with the third k estimator (DER), and Lempel–Ziv entropy rate (LZ), among 19 metrics could sort three levels of CP feasibility (Table 3). However, it was concluded that our metrics were more appropriate to the assessment of CP feasibility because the three metrics showed a small gap between different levels (Fig. 16 a-c) compared to that our metrics which showed a large gap between different levels (Fig. 7 and Fig. 16 d). In the result of applying GB, the gap between 'fair' and 'good' was small compared to

Table 3
AUO, VOT, VOF, TND, and Back et al.'s 19 metrics values.

	GS51	OL21	OG3	JRC15	OG4	Fail cases
AUO	0.59	0.67	0.76	0.77	0.78	–
VOT	0.7	0.85	0.98	0.93	0.98	–
VOF	0.43	0.32	0.14	0.22	0.1	–
TND	0.04	0.3	0.53	0.47	0.6	–
Global block entropy	22.78	18.51	16.9	17.8	16.52	–
Lempel–Ziv entropy	2.55	2.5	1.74	1.87	1.64	–
Difference-based k-block entropy rate (estimator 3)	1.74	1.72	1.46	1.58	1.33	-
Trace entropy	7	7.68	9.61	8.99	9.08	GS51 < OL21 < OG3, JRC15, OG4
Prefix based entropy	15.41	13.58	12.96	14	12.65	OL21 < JRC15
Kozachenko-Leonenko entropy	5.25	4.9	5.71	5.46	5.12	GS51 < OG3, JRC15
K-nearest neighbor entropy (k = 1)	5.25	4.9	5.71	5.46	5.12	GS51 < OG3, JRC15
K-nearest neighbor entropy (k = 2)	4.3	4.05	4.85	4.58	4.28	GS51 < OG3, JRC15
K-nearest neighbor entropy (k = 3)	3.83	3.62	4.42	4.13	3.86	GS51 < OG3, JRC15, OG4
K-nearest neighbor entropy (k = 4)	3.51	3.33	4.13	3.83	3.58	GS51 < OG3, JRC15, OG4
Difference-based k-block entropy rate (estimator 1)	0	0	−0.01	0	−0.01	GS51 = OL21 = JRC15
Difference-based k-block entropy rate (estimator 2)	0.12	0.24	0.5	0.46	0.49	GS51 < OL21 < OG3, JRC15, OG4
Difference-based k-block entropy rate (estimator 4)	1.74	1.72	4.33	1.58	1.33	GS51 < OG3
Difference-based k-block entropy rate (estimator 5)	1.08	1.06	1.05	1.14	0.94	OL21 < JRC15
Ratio-based k-block entropy rate (estimator 1)	2.27	2.56	2.89	2.3	2.12	GS51 < OL21 < OG3
Ratio-based k-block entropy rate (estimator 2)	0.35	0.55	0.93	0.82	0.85	GS51 < OL21 < OG3, JRC15, OG4
Ratio-based k-block entropy rate (estimator 3)	4.51	4.62	4.33	4.17	4.08	GS51 < OL21
Ratio-based k-block entropy rate (estimator 4)	4.51	4.62	4.33	4.17	4.08	GS51 < OL21
Ratio-based k-block entropy rate (estimator 5)	3.13	4.62	4.33	4.17	4.08	GS51 < OL21

the gap between 'poor' and 'fair,' so GB was inappropriate to distinguish between 'fair' and 'good.' In the result of applying DER and LZ, the gap between 'poor' and 'fair' was small compared to the gap between 'fair' and 'good,' so DER and LZ were inappropriate to distinguish between 'poor' and 'fair.' Additionally, GB had a limitation in computational complexity; our metrics, DER, and LZ took less than one minute to calculate, but GB took up to 20 min.

5. Discussion

In the case study, it was confirmed that the homogeneity of clinical orders, the uniformity of order usage, the diversity in the hospitalization period and surgery timing, and the similarity of traces were properties that can identify a CP-feasible group. The experiments showed that the metrics (AUO, VOT, VOF, TND), the visualization method (TA heatmap), and process mining techniques (trace alignment, process discovery, MRCPM) were reasonable to assess these characteristics.

As the homogeneity increased, AUO, VOT, and TND increased, and VOF decreased; thus, the CP-feasible groups could be distinguished using these metrics. The TA heatmap illustrated the diversity in the hospitalization period and surgery timing well, so the clinical experts evaluated that the heatmap was useful to grasp the variety at once. Furthermore, using process mining techniques, the group of patients with higher feasibility demonstrated to have a more structured process with higher matching rates. They had a short aligned result in trace alignment and had a more straightforward process model in process discovery.

The patient groups with high CP feasibility, i.e., OG3, JRC15, and OG4, were characterized by shorter hospitalizations than OL21 and GS51. In other words, the shorter the hospitalization period, the higher the homogeneity of activity and trace. On the other hand, longer hospital stays showed lower uniformity of orders and had more variations. The AUO and TND values, which were metrics that did not consider the frequency and time of orders, decreased as the hospitalization period increased. It indicated that the various kinds of orders were used as the hospitalization period increases. Therefore, for the surgeries with longer hospital stays, it may be necessary to develop CPs for only 2–3 days before and after the surgery when there exists higher homogeneity of orders.

In the two experiments with synthetic data, the metrics were confirmed to express homogeneity well. In the first experiment, which mixed two homogeneous data (OG4 and JRC15), the lowest homogeneity resulted when the two data were half-mixed. This was a result showing that the metrics could express homogeneity well. In the second experiment, which mixed homogeneous data (OG4) and heterogeneous data (GS51), the homogeneity tended to decrease as the proportion of heterogeneous data increased. When the ratio of GS51 exceeded half, the change in metrics decreased, and the value AUO rather increased (an increase of AUO shows that homogeneity increased). This result was caused since the heterogeneity from combining two different data worked with the heterogeneity of GS51 itself. Thus, it could be concluded that the metrics could represent the homogeneity well.

In the analysis for estimating the suitable sample size, the proposed metric worked well in a group of more than 100 patients. The metrics assessed the diversity in orders and traces, so when the number n of patients was small, the diversity decreased; consequently, the metrics evaluate the feasibility optimistically. When the test with different n from 20 to 200 was performed, the worst metric values, i.e., the case considered the least feasible, were similar regardless of data size; however, the best metric values, i.e., the case considered the most feasible, increased as data size decreased. Nevertheless, when $n > 100$, the best value did not interfere with classifying the level of CP feasibility, and it could be concluded that the method should be used when patients in a group were more than 100 ($n > 100$).

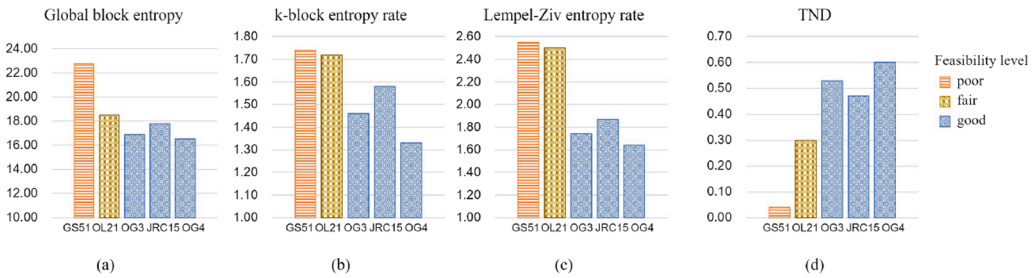

Fig. 16. (a): Global block entropy, (b): k-block entropy rate, (c): Lempel–Ziv entropy rate, and (d): TND values of GS51, OL21, OG3, JRC15, and OG4.

6. Conclusion

This paper proposed a method for assessing CP feasibility regarding clinical complexity using clinical order logs to identify the group of patients whose care process could be managed using CP. Using the proposed method, hospitals could reduce the cost of developing and maintaining CP since the proposed method helps to find appropriate target processes that can be standardized as CPs. Standardization of treatment processes allowed hospitals to take all the best practices and combine them into a single, well-defined process. In addition, a standardized process could be a good treatment guideline for medical staff. Thus, hospitals can gain more patients' satisfaction and improve the quality of care for diseases. The utilization of standard processes can also reduce the operating cost of hospitals.

The proposed method consisted of data preparation, activity & trace homogeneity evaluation, and process inspection using process mining. For evaluating activity homogeneity, three metrics were developed, namely AUO, VOT, and VOF, to evaluate the generality of orders in the log. For trace homogeneity evaluation, TA heatmap was suggested as a visualization method to help analyze the homogeneity of surgery timings and hospitalization periods and a metric, TND, to measure the similarity of traces. Finally, in the process inspection step, a trace alignment, process discovery, and MRCPM were proposed to inspect the process of CP-feasible target. In the experiments, the method was validated and the minimum sample size required to apply the method was estimated. Using the proposed method, the CP-feasible group could be identified in advance and the success rate of data-driven CP development was increased.

This study has some limitations. First, the proposed metrics was validated with a case study using five real-world data sets. However, the reference values were not provided to classify the CP-feasible levels. Thus, more experiments are needed to investigate the reference values. Second, the proposed method focused on the surgical processes. However, there are also other processes requiring CP, such as acute and chronic conditions [47]. The surgical processes are relatively homogeneous. However, the care process for an acute or chronic condition varies in length and interval between visits based on patients. Last, all the orders were treated as entirely distinct. However, some orders were quite similar even though they had different order codes. Considering the similarity among orders, more standardized CPs could be generated.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (No. NRF-2021R1A2C301229211).

References

[1] P.A. Hofmann, Critical path method: an important tool for coordinating clinical care, Joint Comm. J. Qual. Improve. 19 (7) (1993) 235–246.
[2] T. Rotter, R.B. de Jong, S.E. Lacko, U. Ronellenfitsch, L. Kinsman, Clinical pathways as a quality strategy, Improv. Healthcare Qual. Europe (2019) 309.
[3] H. Campbell, R. Hotchkiss, N. Bradshaw, M. Porteous, Integrated care pathways, Bmj 316 (7125) (1998) 133–137.
[4] F. Fox, V.R. Aggarwal, H. Whelton, O. Johnson, A data quality framework for process mining of electronic health record data, in: 2018 IEEE International Conference on Healthcare Informatics (ICHI), IEEE, 2018, pp. 12–21.
[5] R.J. Luttman, G.L. Laffel, S.D. Pearson, Using pert/cpm (program evaluation and review technique/critical path method) to design and improve clinical processes, Qual. Manage. Health Care 3 (2) (1995) 1–13.
[6] S. Ovaere, I. Boscart, I. Parmentier, P.J. Steelant, T. Gabriel, J. Allewaert, H. Pottel, F. Vansteenkiste, M. D'Hondt, The effectiveness of a clinical pathway in liver surgery: a case-control study, J. Gastrointest. Surg. 22 (4) (2018) 684–694.
[7] G.A. Porter, P.W. Pisters, C. Mansyur, A. Bisanz, K. Reyna, P. Stanford, J.E. Lee, D. B. Evans, Cost and utilization impact of a clinical pathway for patients undergoing pancreaticoduodenectomy, Ann. Surg. Oncol. 7 (7) (2000) 484–489.
[8] S. Saint, T.P. Hofer, J.S. Rose, S.R. Kaufman, L.F. McMahon, Use of critical pathways to improve efficiency: a cautionary tale, Am. J. Managed Care 9 (11) (2003) 758–766.
[9] R. van Zelm, E. Coeckelberghs, D. Aeyels, W. Sermeus, A. Wolthuis, M. Panella, K. Vanhaecht, Qualitative evaluation of the implementation of a care pathway for colorectal cancer surgery, Qualitative Health Research 31 (2) (2021) 241–253.
[10] C. Hwang, K. Issokson, C. Giguere-Rich, S. Reddy, A. Tinsley, W.K. Van Deen, H. Bray, D. Lum, H. Aguilar, T.L. Zisman, et al., Development and pilot testing of the inflammatory bowel disease nutrition care pathway, Clin. Gastroenterol. Hepatol. 18 (12) (2020) 2645–2649.
[11] M. Proietti, G.F. Romiti, B. Olshansky, D.A. Lane, G.Y. Lip, Improved outcomes by integrated care of anticoagulated patients with atrial fibrillation using the simple abc (atrial fibrillation better care) pathway, Am. J. Med. 131 (11) (2018) 1359–1366.
[12] Z. Huang, X. Lu, H. Duan, W. Fan, Summarizing clinical pathways from event logs, Journal of biomedical informatics 46 (1) (2013) 111–127.
[13] M. Lang, T. Bürkle, S. Laumann, H.-U. Prokosch, Process mining for clinical workflows: challenges and current limitations, in: MIE, Vol. 136, 2008, pp. 229–234.
[14] A. Perer, F. Wang, J. Hu, Mining and exploring care pathways from electronic medical records with visual analytics, Journal of biomedical informatics 56 (2015) 369–378.
[15] W. Van Der Aalst, Data science in action, in: Process mining, Springer, 2016, pp. 3–23.
[16] R. Mans, H. Schonenberg, G. Leonardi, S. Panzarasa, A. Cavallini, S. Quaglini, W. Van Der Aalst, Process mining techniques: an application to stroke care, in: MIE, Vol. 136, 2008, pp. 573–578.
[17] X. Xu, T. Jin, J. Wang, Summarizing patient daily activities for clinical pathway mining, in: 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), IEEE, 2016, pp. 1–6.
[18] X. Xu, T. Jin, Z. Wei, C. Lv, J. Wang, Tcpm: topic-based clinical pathway mining, in: 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), IEEE, 2016, pp. 292–301.
[19] X. Xu, T. Jin, Z. Wei, J. Wang, Incorporating topic assignment constraint and topic correlation limitation into clinical goal discovering for clinical pathway mining, Journal of healthcare engineering 2017 (2017).
[20] Z. Huang, X. Lu, H. Duan, On mining clinical pathway patterns from medical behaviors, Artificial intelligence in medicine 56 (1) (2012) 35–50.
[21] Z. Huang, W. Dong, L. Ji, C. Gan, X. Lu, H. Duan, Discovery of clinical pathway patterns from event logs using probabilistic topic models, Journal of biomedical informatics 47 (2014) 39–57.
[22] Z. Huang, W. Dong, L. Ji, C. He, H. Duan, Incorporating comorbidities into latent treatment pattern mining for clinical pathways, J. Biomed. Inform. 59 (2016) 227–239, https://doi.org/10.1016/j.jbi.2015.12.012, https://www.sciencedirect.com/science/article/pii/S1532046415002920.
[23] J. Chen, L. Sun, C. Guo, W. Wei, Y. Xie, A data-driven framework of typical treatment process extraction and evaluation, Journal of biomedical informatics 83 (2018) 178–195.

[24] M. Cho, K. Kim, J. Lim, H. Baek, S. Kim, H. Hwang, M. Song, S. Yoo, Developing data-driven clinical pathways using electronic health records: The cases of total laparoscopic hysterectomy and rotator cuff tears, International journal of medical informatics 133 (2020) 104015.
[25] W. Yang, Q. Su, Process mining for clinical pathway: Literature review and future directions, in: 2014 11th International Conference on Service Systems and Service Management (ICSSSM), IEEE, 2014, pp. 1–5.
[26] C.W. Günther, Process mining in flexible environments, Ph.D. thesis, Technische Universiteit Eindhoven (2009).
[27] M. Benner-Wickner, M. Book, T. Brückmann, V. Gruhn, Examining case management demand using event log complexity metrics, in: 2014 IEEE 18th International Enterprise Distributed Object Computing Conference Workshops and Demonstrations, IEEE, 2014, pp. 108–115.
[28] M.O. Kherbouche, N. Laga, P.-A. Masse, Towards a better assessment of event logs quality, in: 2016 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2016, pp. 1–8.
[29] C.O. Back, S. Debois, T. Slaats, Entropy as a measure of log variability, Journal on Data Semantics 8 (2) (2019) 129–156.
[30] R. Khajouei, M. Jaspers, The impact of cpoe medication systems' design aspects on usability, workflow and medication orders, Methods of information in medicine 49 (01) (2010) 03–19.
[31] M. Tschopp, M. Despond, D. Grauser, J.-C. Staub, C. Lovis, Computer-based physician order entry: implementation of clinical pathways, in: MIE, 2009, pp. 673–677.
[32] D.F. Sittig, A. Wright, J.A. Osheroff, B. Middleton, J.M. Teich, J.S. Ash, E. Campbell, D.W. Bates, Grand challenges in clinical decision support, Journal of biomedical informatics 41 (2) (2008) 387–392.
[33] M. van der Kolk, M. van den Boogaard, C. ter Brugge-Speelman, J. Hol, L. Noyez, K. van Laarhoven, H. van der Hoeven, P. Pickkers, Development and implementation of a clinical pathway for cardiac surgery in the intensive care unit: effects on protocol adherence, Journal of evaluation in clinical practice 23 (6) (2017) 1289–1298.
[34] M. Jabbour, A.S. Newton, D. Johnson, J.A. Curran, Defining barriers and enablers for clinical pathway implementation in complex clinical settings, Implementation Science 13 (1) (2018) 1–13.
[35] B. Cardoen, E. Demeulemeester, Capacity of clinical pathways–a strategic multi-level evaluation tool, Journal of medical systems 32 (6) (2008) 443–452.
[36] G.M. Veiga, D.R. Ferreira, Understanding spaghetti models with sequence clustering for prom, in: International conference on business process management, Springer, 2009, pp. 92–103.
[37] C.M. Salgado, C. Azevedo, H. Proença, S.M. Vieira, Noise Versus Outliers, Springer International Publishing, Cham, 2016, Ch. 14, pp. 163–183. doi:10.1007/978-3-319-43742-2_14. URL https://doi.org/10.1007/978-3-319-43742-2_14.
[38] W. Van Der Aalst, A. Adriansyah, A.K.A. De Medeiros, F. Arcieri, T. Baier, T. Blickle, J.C. Bose, P. Van Den Brand, R. Brandtjen, J. Buijs, et al., Process mining manifesto, in: International conference on business process management, Springer, 2011, pp. 169–194.
[39] M. Song, C.W. Günther, W.M. Van der Aalst, Trace clustering in process mining, in: International conference on business process management, Springer, 2008, pp. 109–120.
[40] P. Jaccard, Nouvelles recherches sur la distribution florale, Bull. Soc. Vaud. Sci. Nat. 44 (1908) 223–270.
[41] R.J.C. Bose, W.M. van der Aalst, Process diagnostics using trace alignment: opportunities, issues, and challenges, Inform. Syst. 37 (2) (2012) 117–141.
[42] P.M. Dixit, H.G. Caballero, A. Corvo, B. Hompes, J.C. Buijs, W.M. van der Aalst, Enabling interactive process analysis with process mining and visual analytics., in: HEALTHINF, 2017, pp. 573–584.
[43] J. Lim, K. Kim, M. Cho, H. Baek, S. Kim, H. Hwang, S. Yoo, M. Song, Deriving a sophisticated clinical pathway based on patient conditions from electronic health record data, in: S. Leemans, H. Leopold (Eds.), Process Mining Workshops, Vol. 406, Springer International Publishing, 2021, pp. 356–367.
[44] S. Yoo, M. Cho, S. Kim, E. Kim, S.M. Park, K. Kim, H. Hwang, M. Song, Conformance analysis of clinical pathway using electronic health record data, Healthcare Inform. Res. 21 (3) (2015) 161.
[45] W.M.P. van der Aalst, Process Mining: Data Science in Action, 2nd Edition,, Springer, Heidelberg, 2016, https://doi.org/10.1007/978-3-662-49851-4.
[46] C.W. Günther, A. Rozinat, Disco: Discover your processes, BPM (Demos) 940 (2012) 40–44.
[47] E. Aspland, D. Gartner, P. Harper, Clinical pathway modelling: a literature review, Health Syst. 10 (1) (2021) 1–23.