*Digital Healthcare Empowering Europeans*
*R. Cornet et al. (Eds.)*

*doi:10.3233/978-1-61499-512-8-692*

# Analysis of Care Pathway Variation Patterns in Patient Records

Xiang LI[a,1], Jing MEI[a], Haifeng LIU[a], Yiqin YU[a], Guotong XIE[a], Jianying HU[b], and Fei WANG[b]
[a] *IBM Research, Beijing, China*
[b] *IBM T. J. Watson Research Center, New York, USA*

**Abstract.** A care/clinical pathway (CP) is a standardized care process where temporal and data constraints of clinical activities are defined to ensure quality of care. In actual care practice, various situations of compliance and non-compliance with CPs can be observed. Analysis of these CP variation patterns (CPVPs) can help improve care quality and enhance decision support. In this paper, we propose an automatic method to detect CPVPs in electronic medical records (EMR), and statistically examine their correlation with patient outcomes. From each CP constraint, we first derive a CPVP tree, where each pattern is represented using first-order linear temporal logic and translated into a Büchi automaton for pattern detection. Then we identify the CPVPs that are evidently correlated with a patient outcome by examining the odds ratios. The method has been applied to a CP for congestive heart failure and real world EMR to demonstrate the effectiveness.

**Keywords.** Clinical Pathway, Electronic Medical Record, Variance Analysis, Logic, Quality Improvement

## Introduction

A care/clinical pathway (CP) is a complex intervention for the decision making and organization of care processes for a well-defined patient group during a well-defined period [1]. The aim of a CP is to enhance the quality of care by improving patient outcomes and promoting patient safety. A CP consists of various clinical activities such as diagnoses, medications and laboratory tests, as well as their constraints that describe temporal dependencies, data preconditions and contraindications.

Real world clinical scenarios, however, are more complex than those defined in CPs, and various situations of compliance and non-compliance can be observed in actual care practice. Variance analysis [2] is the process of collecting and analyzing these variations, aimed at improving the quality of future care. In variance analysis, it is inadequate to roughly detect the non-compliance with CP constraints. Detailed care pathway variation patterns (CPVPs), which represent the detailed execution situations of a constraint, should also be identified, especially for a complex constraint with multiple temporal and data relations. Moreover, the correlation between CPVPs and patient outcomes should be analyzed. The detection and analysis of CPVPs can help clinicians identify which detailed patterns lead to positive/negative outcomes, and can be used to provide additional decision support during future CP execution.

1 Corresponding Author. E-mail: lixiang@cn.ibm.com

It is a challenging problem, however, to identify and analyze detailed CPVPs in electronic medical records (EMR). Firstly, CPVPs should be defined in formal semantics that is not only detectable by machine but also understandable by clinicians. Manually formulating the patterns can be tedious and error-prone. Furthermore, for examining the correlation between CPVP and patient outcome, the relationships among the patterns must be explicitly defined. An existing tool [3] and our previous work [4] both focused on the detection of deviations from CPs. But these methods can neither detect the detailed variation patterns, nor analyze their correlation with outcomes.

In this paper, we address these issues by proposing an automatic method to detect and analyze CPVPs in EMR. The CPVPs are defined using first-order linear temporal logic (FO-LTL [5]), which can represent both temporal and data relations, and can be translated into Büchi automatons for the pattern detection in EMR. From each CP constraint, we first syntactically derive a CPVP tree where each node contains the FO-LTL formula of a variation pattern. Then we identify the patterns that are statistically correlated with a given patient outcome by examining the odds ratios in a patient group, based on the relationships defined in the CPVP tree. The analysis results can provide meaningful evidence for clinical practice improvement and further decision support.

## 1. Methods

Our method takes as input a set of patient traces from EMR and a set of constraints from a CP. A patient trace is a sequence of care events for a specific patient, sorted by time. Though there is no agreed standard for modeling a CP [3], temporal dependencies and data preconditions of activities should be defined in a CP format. In this paper, we use a general logic language FO-LTL [5] to represent a CP constraint, which has the capability to model both temporal and data relations. A FO-LTL formula is constructed by combining first-order formulas with temporal operators, which include the next operator **X**, the always operator **G**, the eventually operator **F**, the until operator **U** and the release operator **R**. For example, the following FO-LTL constraints

$$\forall k.\ (EL \wedge \neg(k > 5.0))\ \mathbf{R}\ \neg ACEI, \tag{1}$$

$$\forall k.\ \mathbf{G}\ (ACEI \rightarrow \mathbf{XG}\ ((EL \wedge k > 5.5) \rightarrow \mathbf{G}\ \neg PSD)), \tag{2}$$

$$\forall k.\ \mathbf{G}\ (ACEI \rightarrow \mathbf{XG}\ ((EL \wedge k > 6.0) \rightarrow \mathbf{G}\ \neg ACEI)), \tag{3}$$

mean that "(1) check electrolytes (*EL*) and ensure potassium (*k*) is not > 5.0 (mmol/L) before initiating angiotensin-converting enzyme inhibitors (*ACEI*); (2) after using *ACEI*, if *k* rises to > 5.5, check for use of potassium-sparing diuretics (*PSD*) and stop; (3) if *k* rises to > 6.0, stop *ACEI* immediately" [6]. Figure 1 shows these constraints in a CP based on the CMMN (case management model and notation) standard [7].

Our method consists of two phases: CPVP detection and outcome correlation analysis. The output of our method is the identified patterns and their outcome statistics.

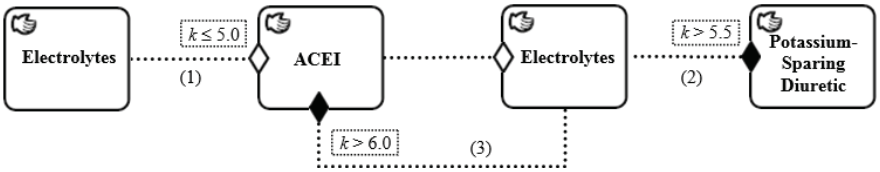


**Figure 1.** Example constraints from a care pathway

### *1.1. Variation Pattern Detection*

Essentially, a CPVP represents a specific state of a FO-LTL constraint, where some subformulas of the constraint are satisfied, some subformulas are violated, and the others are irrelevant. We first define irrelevant conditions of a constraint. Given a constraint $\varphi$ and a subformula $\psi$ of $\varphi$, the irrelevant condition of $\varphi$ with respect to $\psi$ is represented as $\varphi[\psi \leftarrow \bot]$, which is obtained from $\varphi$ by replacing $\psi$ with **true** or **false** [8], depending on the operators on $\psi$. Then a CPVP $\pi$ is defined as one of the Boolean conjunctions of the original constraint $\varphi$ and a set of irrelevant conditions.

Notice that not all CPVPs are unique, valid and meaningful. To address this problem, we propose a tree-based algorithm to generate and prune the patterns. First we build a binary subformula tree $T_{SF}$ for a constraint $\varphi$ by decomposing the FO-LTL formula according to the binary operators ($\vee$, $\wedge$, **U** and **R**), where each node represents a subformula of $\varphi$. Fig. 2(a) shows the subformula tree of the above constraint (1).

Then we build a binary CPVP tree $T_{VP}$, which is expanded in steps according to every inner nodes $\psi_1, \ldots, \psi_n$ in $T_{SF}$. In iteration $i$, an irrelevant condition $\omega$ is generated by replacing one subformula (child node in $T_{SF}$) of $\psi_i$ with $\bot$. Each formula generated in the previous iteration $\pi_{i-1}$ is decomposed into two formulas $\pi_{i-1} \wedge \omega$ and $\pi_{i-1} \wedge \neg\omega$, which will be added into $T_{VP}$ as the children of $\pi_{i-1}$. After the iterations, we combine each leaf node $\pi_L$ with $\varphi$, and add $\pi_L \wedge \varphi$ and $\pi_L \wedge \neg\varphi$ that are not equivalent to **false** (i.e., valid patterns) into the CPVP set. Since FO-LTL has clearly defined semantics, the patterns are understandable. For example, the CPVP tree of the constraint (1) is shown in Figure 2(b), where the pattern (1.I) means that "never use ACEI", (1.II) means that "initiate ACEI without checking electrolytes before", (1.III) means that "initiate ACEI after checking electrolytes and confirming that potassium is not > 5.0", and (1.IV) means that "initiate ACEI after checking electrolytes, but when potassium is > 5.0".

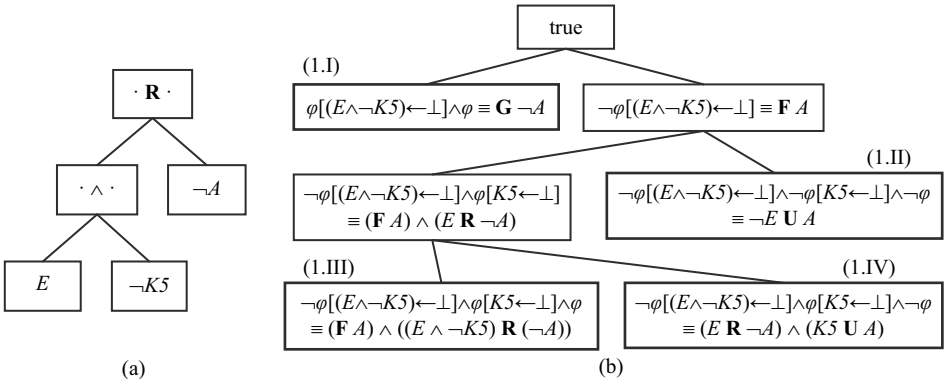


**Figure 2.** (a) Subformula tree; (b) CPVP tree: $\varphi \equiv (E \wedge \neg K5)$ **R** $(\neg A)$, $E \equiv EL$, $K5 \equiv \forall k.\ k > 5.0$, $A \equiv ACEI$

Given a CPVP $\pi$, we can detect whether $\pi$ is fulfilled in a patient trace $\sigma$. We first evaluate the truth value of each variable in $\pi$ at every time point in $\sigma$, yielding a sequence of truth value vectors $\alpha$. The FO-LTL formula $\pi$ can be automatically translated into a Büchi automaton $A_\pi$ [9], which is a non-deterministic finite automaton with acceptance conditions for input sequences. If $A_\pi$ accepts the sequence $\alpha$, then equivalently the pattern $\pi$ is identified in $\sigma$. Otherwise, $\pi$ is not fulfilled in $\sigma$. Figure 3 illustrates the Büchi automatons derived from the patterns in the above example. In this work, the above approach is implemented using Java based on the **ltl2ba** tool [9].

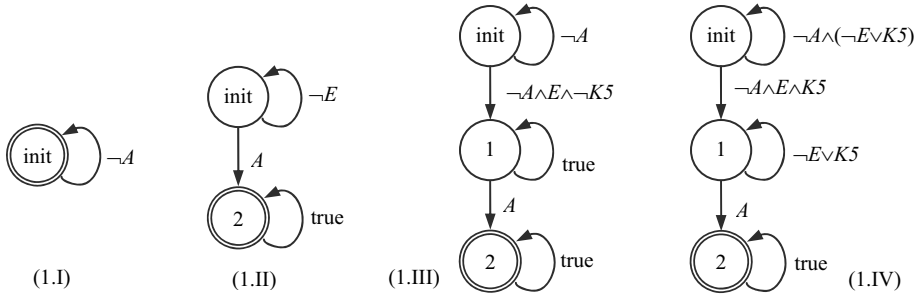


**Figure 3.** Büchi automatons of the example CPVPs in Figure 2

### *1.2. Outcome Correlation Analysis*

After the CPVP detection, we statistically examine the correlation between each pattern and a patient outcome in a group of patients. Let N($\pi$, $\gamma$) be the number of traces where the CPVP $\pi$ is identified and the outcome equals $\gamma$ ($\gamma$ = **true** for positive outcome, $\gamma$ = **false** otherwise), and Sib($\pi$) be the sibling of $\pi$ in the CPVP tree. Then we compute the odds ratio (OR) to quantify to what extent $\pi$ is associated with the outcome:

$$\mathrm{OR}(\pi) = (\mathrm{N}(\pi, \textbf{true}) / \mathrm{N}(\pi, \textbf{false})) / (\mathrm{N}(\mathrm{Sib}(\pi), \textbf{true}) / \mathrm{N}(\mathrm{Sib}(\pi), \textbf{false})).$$

$\mathrm{OR}(\pi) > 1$ means that $\pi$ is identified more in positive cases compared with its sibling pattern, and $\mathrm{OR}(\pi) < 1$ means that $\pi$ appears more in negative cases. This method can evaluate the contribution of every detailed condition of a CP constraint. We also use two-tailed Fisher's exact test to test the hypothesis of $\mathrm{OR}(\pi) = 1$. If p-value is $< 0.05$, then the correlation between the pattern $\pi$ and the outcome is statistically significant.

## 2. Results

We performed a case study using our method on a CP and real world EMR. The CP is derived from a clinical guideline for management of congestive heart failure (CHF) [6]. 22 constraints are defined in the CP to describe the temporal dependencies and contraindications of multiple pharmacological therapies, which include ACEIs (e.g., the constraints shown in Figure 1), angiotensin receptor blockers (ARBs) and diuretics. The EMR contains over 6 million care events from a cohort of 8193 CHF patients. The outcome associated with the patients is whether they are hospitalized or not after the CHF diagnosis. In this cohort, 4945 patients are negative ones whose patient traces are extracted beginning with their first diagnosis date of CHF to their first hospitalization date, while 3248 are positive patients who were not hospitalized after the diagnosis.

In the study, we generated 101 CPVPs from the constraints, detected every pattern in the EMR, and distinguished the negative patterns (OR < 1) from the non-negative ones. As shown in Table 1, most of the variation patterns (94%) can be identified in the EMR, and the majority of them (68%) tend to result in a negative outcome.

**Table 1.** Statistical results of the CPVP analysis

| Treatment | #Constraints | #CPVPs | #Undetected CPVPs | #Non-negative CPVPs | #Negative CPVPs |
|---|---|---|---|---|---|
| ACEIs | 6 | 27 | 1 | 8 | 18 |
| ARBs | 6 | 27 | 3 | 7 | 17 |
| Diuretics | 10 | 47 | 2 | 15 | 30 |

The quantitative analysis of the CPVPs provided meaningful evidence for the improvement of care practice. For example, in actual care, serum potassium is not always monitored on schedule for CHF patients and the related contraindications may also be ignored. However, as shown in Table 2, the variation patterns that violate the example constraints (1) and (3) are evidently prone to lead to a bad outcome, which reveals that potassium should be strictly monitored during ACEI therapy in the future.

The analysis results can also be used to provide additional decision support during future CP execution. For example, if a clinician tried to prescribe an ACEI to a CHF patient with historical records […, *ACEI*, *EL*($k$ = 6.5)], a suggestion for discouraging this action would be given because the negative CPVP (3.V) would be identified.

**Table 2.** Detailed results of the example CPVPs. ($A \equiv ACEI$, $E \equiv EL$, $K5 \equiv \forall k.\ k > 5.0$, $K6 \equiv \forall k.\ k > 6.0$)

| Constraint | CPVP | FO-LTL formula | Interpretation | OR |
|---|---|---|---|---|
| (1) | (1.II) | $(\neg E)\ \mathbf{U}\ A$ | initiate ACEI without checking electrolytes | 0.27 |
| | (1.IV) | $(E\ \mathbf{R}\ \neg A) \wedge (K5\ \mathbf{U}\ A)$ | initiate ACEI when potassium > 5.0 | 0.33 |
| (3) | (3.II) | $\mathbf{F}A \wedge \mathbf{G}(A \rightarrow \mathbf{G}\neg E)$ | not re-check electrolytes after using ACEI | 0.15 |
| | (3.V) | $\mathbf{F}(A \wedge \mathbf{XF}(E \wedge K6 \wedge \mathbf{F}A))$ | use ACEI despite potassium has risen to > 6.0 | <0.01 |

## 3. Discussion

In CP utilization, the analysis of detailed variation patterns, which can be positive or negative for patient outcomes, is critical for the quality improvement of care. In this paper, we proposed an automatic method to detect and analyze CPVPs in EMR. The results showed that our method can provide useful evidence for promoting care practice and enhancing decision support. Besides, the proposed approach has a potential to provide timely and evidence-based recommendation for future amendments of CPs.

One limitation of our current method is that the time interval conditions (e.g., "within 2 weeks") between activities are not defined in FO-LTL formulas. These conditions, however, are very common in CPs and should be taken into account during variance analysis. A potential solution is to model them using interval temporal logic, where the time intervals can be naturally defined.

## References

[1] Vanhaecht K, De Witte K, Sermeus W. The impact of clinical pathways on the organisation of care processes [dissertation]. Belgium: Katholieke Universiteit Leuven; 2007.
[2] Hyett KL, Podosky M, Santamaria N, Ham JC. Valuing variance: the importance of variance analysis in clinical pathways utilisation. Aust Health Rev. 2007 Nov; 31(4): 565-70.
[3] Ainsworth J, Buchan I. COCPIT: a tool for integrated care pathway variance analysis. Stud Health Technol Inform. 2012; 180: 995-9.
[4] Li X, Liu H, Zhang S, Mei J, Xie G, Yu Y, et al. Automatic variance analysis of multistage care pathways. Stud Health Technol Inform. 2014; 205: 715-9.
[5] Fitting M, Mendelsohn RL. First-order modal logic. Dordrecht: Kluwer Academic Publishers, 1998.
[6] Dickstein K, Cohen-Solal A, Filippatos G, McMurray JJ, Ponikowski P, et al. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur J Heart Fail. 2008 Oct; 10(10): 933-89.
[7] Object Management Group. Case management model and notation (CMMN). [released 2014 May; cited 2015 Feb 13]. Available from: http://www.omg.org/spec/CMMN/.
[8] Kupferman O, Vardi MY. Vacuity detection in temporal model checking. International Journal on Software Tools for Technology Transfer, 2003; 4(2): 224-33.
[9] Gastin P, Oddoux D. Fast ltl to büchi automata translation. In: Proceedings of the 13th International Conference on Computer Aided Verification; 2001 July; Paris, France. Berlin: Springer; 2001. p. 53-65.