Journal of Biomedical Informatics 58 (2015) 186–197

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Paving the COWpath: Learning and visualizing clinical pathways from electronic health record data

Yiye Zhang [a,*], Rema Padman [b], Nirav Patel [c]

[a] School of Information Systems Management, The H. John Heinz III College, Carnegie Mellon University, Pittsburgh, PA, United States
[b] The H. John Heinz III College, Carnegie Mellon University, Pittsburgh, PA, United States
[c] Teredesai, McCann & Associates, P.C., Pittsburgh, PA, United States

ARTICLE INFO

Article history:
Received 1 April 2015
Revised 18 June 2015
Accepted 10 September 2015
Available online 28 September 2015

Keywords:
Clinical pathway
Clinical practice guideline
Visualization
Chronic kidney disease

ABSTRACT

Objective: Clinical pathways translate best available evidence into practice, indicating the most widely applicable order of treatment interventions for particular treatment goals. We propose a practice-based clinical pathway development process and a data-driven methodology for extracting common clinical pathways from electronic health record (EHR) data that is patient-centered, consistent with clinical workflow, and facilitates evidence-based care.

Materials and methods: Visit data of 1,576 chronic kidney disease (CKD) patients who developed acute kidney injury (AKI) from 2009 to 2013 are extracted from the EHR. We model each patient's multi-dimensional clinical records into one-dimensional sequences using novel constructs designed to capture information on each visit's purpose, procedures, medications and diagnoses. Analysis and clustering on visit sequences identify distinct types of patient subgroups. Characterizing visit sequences as Markov chains, significant transitions are extracted and visualized into clinical pathways across subgroups.

Results: We identified 31 patient subgroups whose extracted clinical pathways provide insights on how patients' conditions and medication prescriptions may progress over time. We identify pathways that show typical disease progression, practices that are consistent with guidelines, and sustainable improvements in patients' health conditions. Visualization of pathways depicts the likelihood and direction of disease progression under varied contexts.

Discussion and conclusions: Accuracy of EHR data and diversity in patients' conditions and practice patterns are critical challenges in learning insightful practice-based clinical pathways. Learning and visualizing clinical pathways from actual practice data captured in the EHR may facilitate efficient practice review by healthcare providers and support patient engagement in shared decision making.

1. Introduction

Evidence-based medicine is widely acknowledged as a systematic approach for delivering consistent, credible, and safe healthcare [1]. It is a critical assumption in the Affordable Care Act of 2010 to achieve coordinated, patient-centered, and effective healthcare [2]. However, validated models, methods and tools required to apply evidence-based medicine at the point of care, particularly those that accommodate insights from current practice, are lacking. An important source for medical evidence is clinical pathways, indicating the most widely applicable order of treatment interventions for particular patient groups [3]. Clinical pathways translate best available evidence into practice, are known to reduce in-hospital complications, and reduce length of stay and medical expenses [3]. In the US, more than 80% of the hospitals use clinical pathways for at least one intervention [4]. Clinical pathways are developed based on clinical practice guidelines (CPGs), which list recommendations for various treatments based on evidence from randomized clinical trials (RCTs). However, the strength of such evidence differs by clinical areas. For example, while recommendations in CPGs for conditions such as hypertension are frequently based on consistent and good quality patient-oriented evidence [5], those for chronic kidney disease (CKD) are known to be mostly consensus-based due to the difficulties in patient recruitment and high cost of RCTs [6]. Hence, CPGs often face concerns about oversimplification, inadequate applicability to the average patient, and resistance from practicing physicians [7,8].

* Corresponding author.
E-mail addresses: yiyez@andrew.cmu.edu (Y. Zhang), rpadman@andrew.cmu.edu (R. Padman), ndpatel219@gmail.com (N. Patel).

http://dx.doi.org/10.1016/j.jbi.2015.09.009
1532-0464/

As health information technology (IT), such as electronic health records (EHRs), gain widespread adoption and use in healthcare industry, thereby accumulating vast amounts of real-time patient care data, there is tremendous opportunity to develop data-driven models, methods and tools to facilitate review of practice workflows and improve evidence-based care delivery by learning practice-based pathways of care [7]. In this study, we aim to leverage this opportunity using a 'paving the COWpaths' approach [9], where 'COWpaths' refer to the most common clinical 'paths' or treatment patterns followed by patients, and oftentimes recorded using 'computer on wheels' (COW), and identified via retrospective analysis of patients' EHR data. They reflect the multitude of day-to-day clinical decisions made by many clinicians across many patients over a significant duration, and is an important and effective first step in managing change resulting from technological innovation. Upon thorough evaluation and validation by medical experts, we anticipate that these 'COWpaths' may be utilized as 'practice-based clinical pathways', to overcome the challenges faced by current CPGs, and also serve as individualized treatment guidelines.

In this paper, we aim to learn practice-based clinical pathways for chronic kidney disease (CKD), a chronic condition where patients gradually lose their kidney's functions, progressing from stage 1 to stage 5, and end stage renal disease (ESRD) [10]. It is a costly, complex and high mortality health condition affecting 26 million US adults, with another 73 million at increased risk for the disease [11]. In 2011, CKD patients above 65 years old made up 9.2% of US Medicare population (2.3 million), but incurred 18.2% of Medicare costs ($45.5 billion) [12]. Typically, the management of CKD is focused on delaying the progression of the condition, such as maintaining patients in their current disease stage and delaying the progression from stage 5 to dialysis [13]. Two widely known CPGs for CKD are Kidney Disease Outcomes Quality Initiative (KDOQI) and Kidney Disease: Improving Global Outcomes (KDIGO), which summarize global CPGs for various complications of kidney disease [14,15]. Yet, many recommendations are consensus-based, or based on evidence from small-scale randomized clinical trials (RCTs) [14,15]. Recent studies suggest that care delivery changes in CKD management may improve clinical outcomes, enhance quality of patient experience, and reduce annual total per capita health spending [16]. This is of critical importance in the special case of CKD patients who have developed acute kidney injury (AKI) whose consequences are serious and even fatal at times. Yet, few standard preventive and therapeutic options exist for AKI, and there is a growing movement among nephrologists to develop a CPG for AKI [6]. In this study, we specifically focus on AKI as a likely condition that can benefit from insights provided by practice-based clinical pathways that may be learned from EHR data.

We propose a practice-based clinical pathway development process, shown in Fig. 1, that integrates health IT and domain knowledge, including representation of multidimensional and longitudinal EHR data, identification of distinct patient subgroups, and extraction of common treatment patterns as candidate clinical pathways that constitute the focus of this paper. Subsequently, medical experts need to evaluate candidate clinical pathways and their outcomes, and make modifications and redesign, when necessary, to complete the process. Furthermore, we provide visualizations of the learned pathways that can be used by healthcare providers for practice review and decision support, and by patients to engage in shared-decision making, communication, and education. This paper is organized as follows. In Section 2 we describe the data representation, clinical pathway extraction methods, as well as our study data. Analytical results and evaluation are presented in Section 3. We discuss limitations and future work in Section 4 and Section 5 concludes the paper.

2. Materials and methods

2.1. Related work

Previous studies have identified clinical pathways and care patterns from EHR using statistical models, process mining, and machine learning [17–21]. Among studies that used statistical models, Lin et al. and Poelmans et al. modeled clinical pathways as a hidden Markov model (HMM), a stochastic model for randomly changing processes, that includes sequences of hidden states and observations [22,23]. Hidden states in the HMM are assumed to be states in the clinical pathway, and observations are actual interventions that occurred in practice. Generally, a large amount of data is necessary to train a HMM. On the other hand, process mining approaches take process logs as inputs, and in the case of clinical pathway mining, clinical activities recorded in EHR are used as process logs [20]. Many of the process mining approaches, such as Heuristic Miner and Fuzzy Miner, assume that event logs contain sufficient information and minor noise [24,25]. This assumption leads to spaghetti-like workflow models that are hard to interpret, as diversity is innate to most of the health data [19]. Recently, while still a process mining approach, Huang et al. presented sequence mining algorithms for clinical pathway patterns that did not look for an entire clinical pathway from start to end, but rather patterns of clinical pathways including time differences between events [21]. Also, machine learning techniques provide a potential solution to spaghetti-like workflow models by segmenting patients into relatively homogeneous groups before learning clinical pathways [18,26,27]. For example, Greco et al. used hierarchical clustering to cluster patients' sequences of visits [27], and Lakshmanan et al. segmented patients by their outcomes, followed by further clustering using DBScan and frequent pattern mining using SPAM [18].

Models from these studies have primarily been applied to clinical processes where each point in the pathway is a clinical event that bears, or is assumed to bear, clear temporal relationship with another [19,21]. However, in the outpatient setting especially, time stamps associated with clinical activities are typically *per day* [18]. Such lack of accurate temporal ordering presents a mining challenge for previous methods to accurately learn the co-progression of interventions and outcomes. For example, CKD patients visit clinics and hospitals every few months depending on the severity of their conditions. During each visit, they

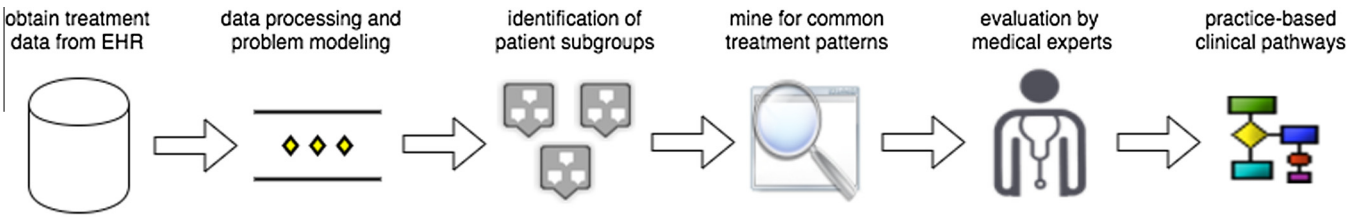

Fig. 1. Practice-based clinical pathway development process.