Inferring Clinical Workflow Efficiency via Electronic Medical Record Utilization

You Chen, PhD[1], Wei Xie, MS[1], Carl A Gunter, PhD[2], David Liebovitz, MD[3], Sanjay Mehrotra, PhD[3], He Zhang, PhD[4], and Bradley Malin, PhD[1]
[1]Vanderbilt University, Nashville, TN; [2]University of Illinois, Urbana, IL; [3]Northwestern University, Chicago, IL; [4]University of South Florida, Tampa, FL

Abstract

Complexity in clinical workflows can lead to inefficiency in making diagnoses, ineffectiveness of treatment plans and uninformed management of healthcare organizations (HCOs). Traditional strategies to manage workflow complexity are based on measuring the gaps between workflows defined by HCO administrators and the actual processes followed by staff in the clinic. However, existing methods tend to neglect the influences of EMR systems on the utilization of workflows, which could be leveraged to optimize workflows facilitated through the EMR. In this paper, we introduce a framework to infer clinical workflows through the utilization of an EMR and show how such workflows roughly partition into four types according to their efficiency. Our framework infers workflows at several levels of granularity through data mining technologies. We study four months of EMR event logs from a large medical center, including 16,569 inpatient stays, and illustrate that over approximately 95% of workflows are efficient and that 80% of patients are on such workflows. At the same time, we show that the remaining 5% of workflows may be inefficient due to a variety of factors, such as complex patients.

Introduction

The success of an electronic medical record (EMR) system implementation and its subsequent adoption by users is contingent upon the initial design and refinement of workflows in the healthcare organization (HCO)[1,2]. The appropriate design and management of workflows can significantly improve efficiency of clinical treatment[3,4], care quality[4], patient safety[5], and care decisions[6]. Yet, despite their potential, workflows can be quite complex[7], creating barriers to EMR system utilization[2,3]. This can ultimately lead to inefficiency in diagnoses, ineffectiveness of treatment plans, and uninformed management of an HCO. To mitigate workflow complexity in EMR systems, it has been suggested that HCOs design workflows to optimize business processes or manage the complexity of current workflows, rather than rely upon of current workflows[1,7,8,9].

To enable such strategies, various approaches have been developed to measure the gap between the workflows defined by an HCO and the actual processes followed by individuals in a clinical setting. In general, for the gap analysis approaches, the HCO workflows are compared to the expectation of experts in the organization, as learned through surveys, interviews or observational data collected in the physical healthcare setting[7,8,12,14,15,16]. Application of these methods requires a substantial exertion of manual effort because it requires invasive interviews and patience while observing the interactions between care providers and patients. As a result, this type of approach is often limited to specific areas of clinical care. Furthermore, these methods only measure the gap between the organization's workflows and the expectation of care providers. This neglects the influences of EMR systems on the utilization of workflows, which could be leveraged to optimize workflows facilitated through the EMR systems.

By contrast, most recent approaches model workflows by mining data recorded in the EMR system. This type of strategy considers the influences of EMR systems on the utilization of workflows, but it only models the patterns of care paths, and neglects the efficiency management of workflows [9-11].

Good management of workflow efficiency can improve quality of clinical care and reduce costs of patients[1,3]. By providing HCOs with such knowledge of workflows, we anticipate healthcare administrators will be able to optimize the efficiency, as well as minimize the complexity of workflows in a more productive manner. For instance, imagine that an HCO administrator learns a particular clinical process tends to require a long duration in time, but that the process has a large variance in its duration across the enterprise. Based on this knowledge, they can investigate the reasons behind such long waiting times and variation. In this paper, we introduce a framework to model clinical workflows at two levels of granularity and categorize these workflows according to their efficiency. To the best of our knowledge, this is the first approach to automatically learn and categorize workflows according to their duration.

Background

The past several years have witnessed a number of investigations into modeling and characterizing the workflows associated with clinical practices. As alluded to earlier, we characterize these investigations, and the approaches upon which they are based, into two classes. These classes are dependent upon the methods of data acquisition and analysis that are invoked: i) physical observation studies (in the clinic) and ii) virtual observation studies (data in the EMR). Here, we take a moment to review representative research from both categories and illustrate the relationship with our own approach.

Workflows Based on Physical Observations

Observation-driven studies often rely on manual data collection approaches, such as observations and interviews. One such example was presented by Unertl and colleagues[8], which analyzed direct observations and interviews in hospitals to understand workflow and information flow in the care of chronic diseases. Similarly, Ramaiah and colleagues[7] designed surveys consisting of questions, interviews of care providers and patients to discover workflows associated with time delays in the HCO. In another setting[15], comparative data were collected from the operating room and statistical analysis was performed with respect to gains in efficiency. The ultimate goal was to justify the need for reorganizing clinical workflow to increase throughput in the operating room. In other work[16], ethnographic observation and interview data were applied to study the evolution and management of medical errors.

Workflows Based on Virtual Observations

Modern studies[10,13,20,21] are increasingly turning to EMR-related data because they can enable large scale analytics at a low acquisition cost. While more comprehensive surveys[10,13] exist about workflow mining in the EMR, we briefly examine investigations relevant to our approach.

In one study[9], Zhang and colleagues utilized EMR usage logs to model patients' flow in the healthcare system. After learning patterns of patient record usage, deviations from the average workflows were detected and promoted for investigation as either undocumented policies or misuse of the EMR system. Similarly, another study[12] employed sequence alignment methods to derive a consensus workflow and automatically detect outliers from surgical activity logs. Other studies have focused on contrasting treatment differences for certain diseases. For instance, Mans and colleagues[14] studied stroke care by applying process mining to clinical data. They additionally compared the pathways from disparate healthcare systems and various types of patients. Most recently, Partington and colleagues[17] focused on cross-hospital process mining and performed a comparative analysis by leveraging a combination of administrative and clinical data. This investigation yielded detailed insights into patient care and hospital budget pressures.

Methods

We provide a framework to 1) learn workflows and 2) categorize the workflows into four general types according to the length and variation in their temporal duration. For reference purposes, we summarize the common terms and notation used in this paper in Table 1.

Event. An event corresponds to the smallest granularity associated with a workflow. This corresponds to an action invoked by an EMR user over a patient's medical record at a certain time. For instance, a user, acting as a pathologist, can initiate an event by accessing a patient record to request a lab test. Alternatively, another user, acting as a primary care provider, can initiate an event by accessing a patient record to approve a request to refill a medication.

Sequence. A patient sequence consists of a series of ordered events that represents an episode of a patient process. For example, an ordered series of events could be: *a user, acting as a physician requested a lab test for a patient* → *a laboratory user uploaded a lab test result for the patient* → *the lab test results were returned to a care provider in a physician office* → *a registered nurse provided customer service support to the patient in response to an inquiry about the lab test results*. We assume we are provided with m events that are classified into n patient sequences.

Block. A block is an ordered series of events that have strongly ordered relations with each other. For example, if the relation between an event *Rehab Service Clinician* and another event *Rehab Quality Audit* is strong in a way we will define precisely below, then *Rehab Service Clinician* → *Rehab Quality Audit* belongs to a block. A block represents a specific stage of a patient process, and a patient sequence consisting of events can be represented using the corresponding blocks. For instance, imagine there are two blocks [*Primary Care Physician* → *Laboratory Tester*] and [*Physician Office Care Provider* → *Primary Care Staff Nurse*]. Then, the earlier patient sequence example (with four events) can be represented using these blocks to express a specific stage of a workflow corresponding to the four events.

Topic. A topic consists of a set of blocks, which together represent the main processes associated with a patient type. For example, imagine there is a group of patients with the conditions *Urinary Tract Infection, Other Specified Retention of Urine,* and *Unspecified Essential Hypertension*. Then a topic consisting of the following blocks may characterize this group: [*Advanced Practice Clinician – CPOE → Physician Office*], [*SN-OR RN SC → Patient Care Assistive Staff*], [*Physician Office → Advanced Practice Clinician – CPOE*], [*OR RN SC-Primary → Primary Assistive Staff*], [*Unit Secretary → Rehab PT, Respiratory → SN-RN/Customer Service*], and [*SN-RN/Customer Service → NMH Physician Hospitalist-CPOE*].

Table 1. Common notation used in this study with their corresponding definitions.

Notation	Description
$\mathrm{E} = \{e_1, e_2, \cdots, e_i, \cdots e_m\}$	A set of events
$\mathrm{S} = \{s_1, s_2, \cdots, s_i, \cdots s_n\}$	A set of sequences
$s_i = [\cdots \to e_i \to e_j \to \cdots]$	A sequence, which consists of a series of events in order.
$\mathrm{B} = \{b_1, b_2, \cdots, b_i, \cdots b_l\}$	A set of blocks
$b_i = [\cdots \to e_k \to e_l \to \cdots]$	A block, which is a series of events in order.
$s'_i = \{\cdots, b_i, b_j, \cdots\}$	A revised sequence, which consists of blocks.
$\mathrm{S}' = \{s'_1, s'_2, \cdots, s'_i, \cdots s'_n\}$	A set of revised sequences
$\mathrm{T} = \{t_1, t_2, \cdots, t_i, \cdots t_k\}$	A set of topics
R_E	An asymmetric matrix representing the relations between events
R_B	A symmetric matrix representing the relations between blocks
$R_{S' \times B}$	A matrix representing the relation between sequences in S' and blocks in B

We learn workflows at a fine granularity, in the form of blocks of events, and a coarse granularity, in terms of topics of blocks. Representation at the block-level characterizes the stage of a patient process. By contrast, representation at the coarse-grain characterizes different types of patient processes. Fig.1 shows an example consisting of 4 models to learn workflows at the block- and topic-level. These are formalized in the Workflow Mining Algorithm (WMA) depicted in Fig. 2.

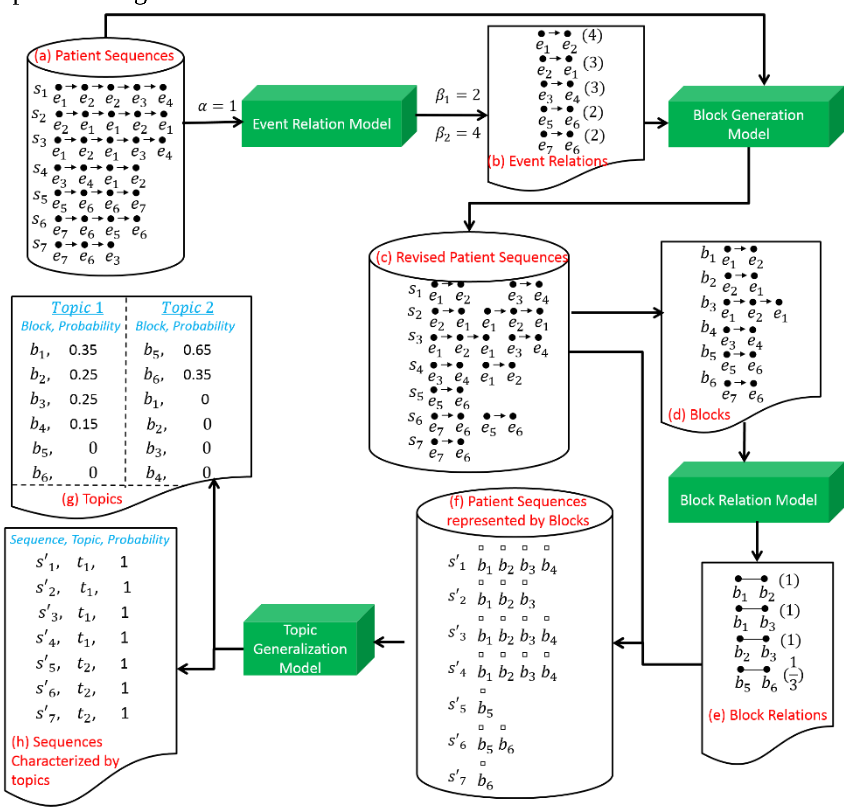

Figure 1. Generation of fine-gained (blocks) and coarse-gained (topics) workflows from patient sequences.

1. Event Relation Model (ERM): Measures relations between events according to the patient sequences. The details of relation measurement are depicted in steps 1 through 5 of WMA.

2. Block Generation Model (BGM): Generates blocks of events according to the relations of events and patient sequences. The details of block generation are described in steps 6 through 19 of WMA.

3. Block Relation Model (BRM): Measures the relations between blocks according to the common events they contain. The details about measurement are in steps 20 through 22 of WMA.

4. Topic Generation Model (TGM): Generates topics of blocks to represent a similar type of patient processes. The details are depicted in steps 23 to 26 of WMA.

We introduce these models in detail in the following sections on workflow generation at the block- and topic-level.

Block-level Workflows

First, we generate event blocks (i.e., workflows at the block-level) according to steps 1 through 19 of WMA. A workflow at the block-level aims to characterize the efficiency of a stage of a patient process. In this work, we aim to learn four types of blocks:

(1) **Stable Efficient Blocks (SEB):** Have short average duration with small variance across the patient population.

(2) **Unstable Efficient Blocks (UEB):** Have short average duration with large variance across the patient population.

(3) **Stable Inefficient Blocks (SIB):** Have long average duration with small variance across the patient population.

(4) **Unstable Inefficient Blocks (UIB):** Have long average duration with large variance across the patient population.

We anticipate that the categorization of blocks into these types can assist HCOs in speeding up the discovery of (in)efficient blocks and refine their policies accordingly. For instance, imagine the block [*Radiology Mgr/RC* →*Attending Physician/Provider*] exhibits a large variance in its duration, such that it requires less than 1 hour for one patient, but 240 hours for another patient. Our model could promote this block to administrators for investigation.

We infer blocks from the event logs generated by EMR systems. This is because EMR-facilitated workflows, and the utilization logs in particular, do not necessarily follow the exact order of events in the physical world. For instance, in the real world, the order of events may be $e_i \rightarrow e_j$, but in some cases, the order recorded by an EMR may be $e_i \rightarrow e_k \rightarrow e_j$. To relax the order relations of events, we consider relations within a sliding time window. Specifically, we assume that if an event and its following events are within a window of size α, the order relations holds, but the strength of the relation is proportional to their distance. We measure the order relation between a pair of events within a sequence as:

$$Event_{relation}(e_i, e_j) = \begin{cases} \frac{1}{(p(e_j)-p(e_i))^2}, & (0 < p(e_j) - p(e_i) \leq \alpha) \\ 0, & otherwise \end{cases} \tag{1}$$

Where $p(e_i)$ is the position of the event e_i in a sequence. The position of the first event of a sequence is set as 1, and the position of the last event of a sequence is set as the length of a sequence. The relations of events are incrementally measured over all of the sequences (as depicted in steps 1 through 5 of WMA). As a consequence, this type of relation considers the ordinal distance between the events and their frequency in the patient population. When we set α to 1, the relation between $e_1 \rightarrow e_2$ is 4 (1+1+1+1) accumulated from sequence s_1, s_2, s_3,and s_4 respectively (as shown in Fig.1(b)).

To represent each stage of a patient process via temporal patterns, we infer event blocks through relations of events. The generation of such blocks is shown in steps 7 through 19 of WMA. We assume that, for a sequence of events, if an event has a strong relation with the following events, then they should be grouped into a block. Since the relations of events are already considered in the sliding time window, we only consider the immediately following neighbor in block generation. As shown in steps 10 through 13 of WMA, if the relation of every neighboring event is within a range of $[\beta_1, \beta_2]$, these events are included in a block. The lower bound for the range is applied to filter out weak relations, whereas the upper bound filters highly frequent relations (like stop words, "a", "of", or "the" in natural language processing). The end of a block is realized when the relation between an event and its neighbor is outside of the range. For instance, when $\beta_1 = 2$ and $\beta_2 = 4$, sequence $s_1 = e_1 \rightarrow e_2 \rightarrow e_2 \rightarrow e_3 \rightarrow e_4$ (shown in Fig.1(a)) generates two blocks $e_1 \rightarrow e_2$ and $e_3 \rightarrow e_4$ (shown in Fig.1.(c)) . This is because the event relations of $e_2 \rightarrow e_2$ and $e_2 \rightarrow e_3$ occur only once, which is below the lower bound threshold of 2.

To evaluate and characterize the time efficiency of blocks, we need contextual information, such as the duration of a block or the reasons for each event in a block. Note, a block can appear multiple times in different patient sequences, such as when $b_1 = e_1 \rightarrow e_2$ exists in s_1, s_2, s_3,and s_4. As such, an event block can have multiple reasons and time durations. For example, if $e_1$happens at 9:00am, and the following event $e_2$happens at 10:00am in sequences_1, then the time duration for $b_1$in $s_1$is 1 hour.

We summarize the efficiency of a block using several basic statistics. Specially, we compute the duration and variance for a block as the average and standard deviation over all occurrences in the patient population. For example, if the time duration of block $b_1$in the four sequences is 1, 2, 1, and 1.5 hours, respectively, then the average time duration is 1.375 hours and the variance (standard deviation) is 0.4787 hours.

Workflow Modeling Algorithm (WMA)

Input: S: a set of patient sequences; **E**: a set of events; and **β**: a threshold for the event relation.
Output: B: a set of blocks and **T:** a set of topics
Steps:
1: **for** each sequence $\boldsymbol{s_i} \in \boldsymbol{S}$ **do** // Process each sequence in the sequence set
2: **for** each pair of events $\boldsymbol{e_{i,j}} \in \boldsymbol{s_i}$ and $\boldsymbol{e_{i,q}} \in \boldsymbol{s_i}$ **do** // Process each pair of events in a sequence
3: $\boldsymbol{R_E(e_{i,j}, e_{i,q}) = R_E(e_{i,j}, e_{i,q}) + Event_{relation}(e_{i,j}, e_{i,q})}$ // Measure the relation between events
4: **end for**
5: **end for**
6: **for** each sequence $\boldsymbol{s_i} \in \boldsymbol{S}$ **do** // Generate workflows at block-level
7: j ← 1; // First event of a sequence
8: **while** ($\mathbf{j} < \boldsymbol{i_m}$) **do** // Iterate over each event in a sequence
9: $\mathbf{blockStart = j}$; $\mathbf{blockEnd = j}$; // Initialize the start and end position of a block
10: **while**(($\mathbf{j} < \boldsymbol{i_m}$) and ($\boldsymbol{\beta_1 < R_E(e_{i,j}, e_{i,j+1}) < \beta_2}$)
11: j ← j+1; // If the relation of neighboring events is within a range keep the order of events in a block
12: $\mathbf{blockEnd} \leftarrow \mathbf{j}$; // Update the end position of a block
13: **end while**
14: $\boldsymbol{B \leftarrow union(B, e_{i,blockStart} \rightarrow e_{i,blockEnd})}$ // Add new block into the block set
15: $\boldsymbol{s'_i \leftarrow Append(s'_i, e_{i,blockStart} \rightarrow e_{i,blockEnd}})$ // Append the generated block to the new sequence
16: j ← j+1; // Generate the next block
17: **end while** // End of processing a sequence
18: $\boldsymbol{S' \leftarrow union(S', s'_i)}$ // Add new sequence into the revised sequence set
19: **end for**
20: **for** each pair of blocks $\boldsymbol{b_i} \in \mathbf{B}$ and $\boldsymbol{b_j} \in \mathbf{B}$ **do** // Measure relations of pairs of blocks
21: $\boldsymbol{R_B(b_i, b_j) \leftarrow Block_{relation}(b_i, b_j)}$ // Measure the relation between blocks
22: **end for**
23: **for** $\boldsymbol{s'_i} \in \mathbf{S'}$ and $\boldsymbol{b_j} \in \mathbf{B}$ **do**
24: $\boldsymbol{R_{S' \times B}(s'_i, b_j) \leftarrow SB_{relation}(s'_i, b_j)}$ // Measure the relation between a sequence and a block
25: **end for**
26: **T ← LDA(**$\boldsymbol{R_{S' \times B}}$) // Generate workflows at the topic level
27: **Return B, T** // Return workflows at block and topic level

Figure 2. Pseudocode for the algorithm to generate workflows at the block- and topic-level.

Topic-level Workflows

A workflow at the block-level only provides a description for a particular stage of a patient process. However, it neglects the relations between blocks within a patient process. To do so, we summarize collections of blocks using a topic modeling strategy (e.g., latent Dirichlet allocation (LDA))[18,19], a popular approach to learn latent concepts from a corpus of documents. In our setting, this corresponds to learning a set of latent workflow patterns to represent patient sequences. Conceptually, patient sequences can be thought of as documents, where the event blocks constitute a vocabulary and the specific event blocks assigned to a patient's sequence are the semantic ideas derived from the vocabulary.

To learn topics, we begin by generating the set of unique blocks. We then represent each original patient sequence s_i as a new sequence s'_i, consisting of blocks instead of events (step 15 of WMA). For instance, Fig.1(d) depicts 6 unique blocks generated from Fig.1(c). Now, to infer topics of terms (i.e., blocks), we need to prepare the documents (i.e., patient sequences). Here, a term is a block and a document is a patient sequence $\mathbf{s'}_i$ that consists of a series of blocks. While the events within a block are ordered, the order is not necessarily consistent with the real world. As such, we group the blocks associated with similar sets of events. In many respects, this can be thought of as creating a set of synonyms in a vocabulary.

To discover synonyms of blocks, we measure the similarity using the Jaccard coefficient between the blocks (as invoked in step 21 of WMA):

$$\boldsymbol{Block_{relation}}(\mathrm{b}_i, \mathrm{b_j}) = \frac{set(b_i) \cap set(b_j)}{set(b_i) \cup set(b_j)} \qquad (2)$$

where $set(b_i)$ corresponds to the set of unique events in a block. Note that this function ranges from 0 (no relation) to 1 (perfect relation). For instance, the relation between block b_1 and b_2 (as shown in Fig.1(d)) is 1 (as shown in Fig.1 (e)).

As described earlier, we want each patient sequence, including the synonyms of its blocks, to improve the quality of learned topics through the LDA model. We use Equation 3 to represent a patient sequence using its blocks, along with their synonyms, as:

$$\boldsymbol{SB_{relation}}(\mathrm{s'_i}, \mathrm{b_j}) \begin{cases} 1 & b_j \in s'_i \\ 1 & b_j \notin s'_i \ (\exists b_q \in s'_i, B_{relation}(b_q, b_j) = 1) \\ 0 & other\ wise \end{cases} \tag{3}$$

We generate a matrix $R_{S'\times B}$ to represent the relation between a sequence $\mathrm{s'}_i$ and a block b_j (steps 23 through 25 of WMA). For instance, s'_1 has block b_1 and b_4 (as shown in Fig.1(c)), but b_1 has synonyms b_2 and b_3 according to Equation 3. As such, s'_1 is represented by blocks b_1, b_2, b_3, and b_4 (as shown in Fig.1(f)). The matrix $R_{S'\times B}$ serves as the input to the LDA learning process (step 26 of WMA).

A topic consists of a probability distribution over a set of blocks as shown in Fig.1(g). The larger the probability of a block, the more this block is representative of the topic. At the same time, the patient process can be characterized by inferred topics as shown in Fig.1(h).

Results

We evaluate our framework on four months of inpatient event logs generated by the Northwestern Memorial Hospital (NMH). In this dataset, an event corresponds to a chart access, each of which is associated with the user and the user-designated reason for the access. It should be noted that the initial reasons selected by a chart user during the hospitalization of a patient persists throughout the hospitalization. There are 1,138,317 total events distributed over 16,569 patient processes. These events were generated by 144 user roles with access to 142 reasons. Additionally, each patient is associated with a set of ICD-9 codes assigned after discharge. The total number of unique ICD-9 codes for this set of patients is 4,543.

To apply WMA, we need to set three parameters: i) the sliding window size α, ii) the event relation thresholds β_1 and β_2, and iii) the number of topics for the LDA model. We set α to the average number of events that transpired during a 24 hour window. This corresponded to 3 events. β_1 was set to 50 because smaller values led to an extremely large number of blocks (over 100,000), which suggested a substantial amount of noise. β_2 was set to 500 because, at this point, the only events in the resulting blocks corresponded to either : 1) *Physician-CPOE*, 2) *Residence*, 3) *Patient Assistive Staff*, 4) *Patient Care Staff Nurse*, 5) *Respiratory*, and 6) *Unit Secretary*. These extremely frequent blocks limited the generation of meaningful blocks and, thus, we removed these from further consideration. For the LDA model, we set the number of topics according to topic similarity instead of perplexity, based on a previous study[19]. In doing so, we searched for a model that minimizes the workflow topic similarity, measured as the cosine similarity. We set the number of topics as 15, 20, 25, 30, and 35 respectively, and calculate the corresponding topic similarity in each setting. The topic similarity was minimized (0.0033) when the number of topics was set to 25. Given these parameters, WMA generated 22,442 event blocks and 25 topics of blocks.

In the remainder of this section, we first show the block- and topic-level efficiency results. To illustrate the results at each level, we then provide a case study with respect to the learned workflows.

Block Types

For each block, we record the average and variance in time duration. Figure 3 depicts the 22,442 blocks as a function of these concepts. We partition this space, based on their length and stability, into the four types mentioned above. To do so, we partitioned the system using thresholds of an average length of 100 hours and a variance of 100 hours. According to this split, 94.7% of the blocks are in the SEB area (i.e., short duration with low variance). This suggests that the HCO mainly manages inpatients associated with short processes.

At the same time, this finding implies that approximately 5% of the blocks are of potential concern. Among the remaining blocks, 70% correspond to SIB (i.e., long duration with low variance), 22% correspond to UIB (i.e., long duration and long variance), and 8% correspond to UEB (i.e., short duration and large variance). It is possible, however, that these blocks are artifacts of an insufficient amount of evidence to draw meaningful conclusions. To assess this issue, Figure 4 provides a frequency analysis for all of the blocks as a function of their average duration and variance. Clearly, the frequency of the efficient blocks (i.e., SEB, UEB) is substantially larger than the inefficient blocks (i.e., SIB, UIB). In most cases, the frequency of SIB and UIB blocks is small (around 2~3), which indicates they are not popular in the management of inpatients.

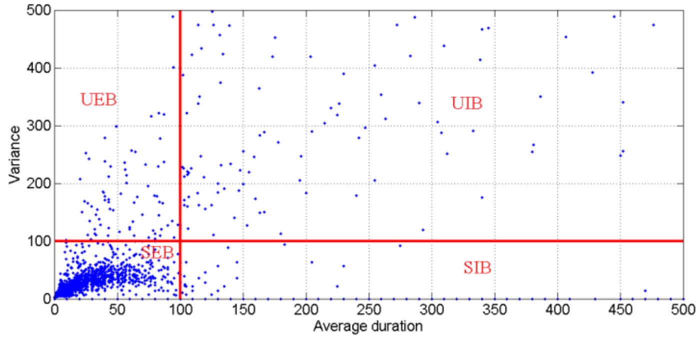

Figure 3. Blocks (n = 22,442) as a function of their average duration and variance. The blocks are partitioned into four types: i) SEB: stable efficient blocks, ii) SIB: stable inefficient blocks, iii) SEB: unstable efficient blocks, and iv) UIB: unstable inefficient blocks.

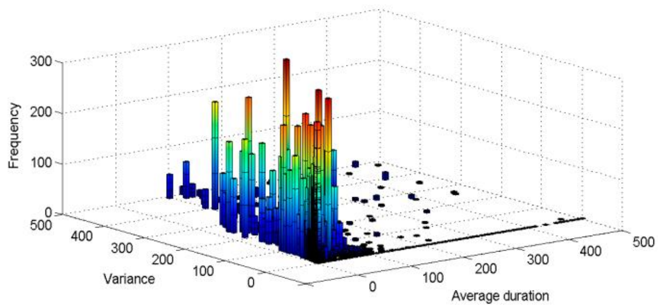

Figure 4. Blocks as a function of their average duration, variance in duration, and frequency in the patient population.

Block Type Case Studies

To gain a deeper appreciation of the different block types, we provide examples of UEB and UIB in Table 3. Block B1 belongs to UEB and has 6 unique reasons associated with it. Its reason R1 is associated with two significantly different durations. The first is less than 1 hour, while the other is 240 hours. The same phenomenon occurs for the reason R4, which has a duration of less than 1 hour and another of 160 hours. These phenomena illustrate how a block with the same chart access reason, can exhibit significantly different durations. They indicate that, though a block encompasses the same transitions between reasons, the time allotted for doing so may be significantly different. This may stem from a number of complications, such as varying patient symptoms and purposes for the hospitalization, the urgency of imaging needs, the ability of some attending physicians to rely on residents in training to access charts for them and provide relevant updates, the resource allocation strategies of HCOs, or the workflow timing of chart access by care providers.

Table 3. Examples of UEB and UIB block types.

Block		Duration (hours)		
		Average	Variance	Frequency
B1	[*RAD - Mgr/RC → NMH Physician-CPOE*]	25.6	112.4	89
B2	[*NMPG MD – CPOE → NMPG APRN → NMPG MD – CPOE*]	203	418	18

Table 4 shows two patients associated with Block B1. The condition for one patient is related with polyneuritis, while the condition for the other patient is related with septicemia, which is more complex. This may be the reason why this block exhibits high variance in time duration, even for the same chart access reason.

Block B2 belongs to UIB, and its reason R2 has long duration and large variance. As can be seen, the patients associated with this block are related to obstetric care. Furthermore, advanced practice nurses function as care providers, which includes the ability to create and sign chart orders, lessening the need for rapid access by a supervising attending.

Topic Workflow Types

To characterize the categories of workflows at the topic-level, we inferred 25 topics through distribution of blocks on patients, and then categorize the 25 topics of blocks into four groups using thresholds of a duration of 20 hours and a variance of 20 hours, as shown in Figure 5. Each topic is composed of top 10 blocks with highest probabilities. The duration and variance of a topic were calculated based on average durations of top 10 blocks. Topics 1, 7, 19, 20 were classified as Unstable Inefficient Topics (UITs), topic 8 was a Stable Inefficient Topic (SIT), topic 4 and 11 were Unstable Efficient Topics (UETs) and the remaining 18 topics were Stable Efficient Topics (SETs).

In Figure 5, it can be seen that topic 20 exhibits the longest time duration and corresponding variance. This is because one of the top 10 blocks of topic 20 belongs to UIB. This block corresponds to [*NMH Physician Office – CPOE → SN-RN/Customer Service → NMH Physician Office – CPOE → SN-RN/Customer Service*], which has 8 different reasons. The duration and variance of this block is 300 and 170 hours, respectively.

Table 4. Examples of patients associated with B1 and B2.

Block	ICD-9 Codes	Description of ICD-9 Codes	Reasons
B1	Patient 1: 340, 35781	Multiple sclerosis; Chronic inflammatory demyelinating polyneuritis	R1: Radiology Mgr/RC (a non-physician manager within the radiology department) → Attending Phys/Prov (the primary physician responsible for an inpatient's care) [appeared twice: one time for less than 1hour, and one time for 240 hours] R2: Radiology Mgr/RC → Resident- Inpatient Consulting Service [low duration, low variance] R3: Radiology Mgr/RC → Approved Quality or Peer Review Process [low duration, low variance] R4: Radiology Mgr/RC → Patient Care (associated with nursing roles) [appeared twice: one time for less than 1 hour, and one time for 160 hours] R5: Radiology Mgr/RC → Resident-Inpatient Covering Service [low duration, low variance] R6: Radiology Mgr/RC → Resident-Inpatient Primary Service [low duration, low variance]
	Patient 2: 78552, 7907, 99592, 0389, 0417, 2760, 2762, 2875, 5070, 51881, 5849, 6826, 68601, 70705, 70719	Septic shock; Bacteremia; Severe sepsis; Unspecified septicemia; Pseudomonas infection in conditions classified elsewhere and of unspecified site; Hyperosmolality and/or hypernatremia; Acidosis, Thrombocytopenia, unspecified; Pneumonitis due to inhalation of food or vomitus; Acute respiratory failure; Acute kidney failure, unspecified; Cellulitis and abscess of leg, except foot; Pyoderma gangrenosum; Pressure ulcer, buttock; Ulcer of other part of lower limb	
B2	Patient 3: 65971, 66411, V270	Abnormality in fetal heart rate or rhythm, delivered, with or without mention of antepartum condition; Second-degree perineal laceration, delivered, with or without mention of antepartum condition	R1: Other Phys/Prov → NMPG APRN → Other Phys/Prov [long duration, low variance] R2: Attending Phys/Prov → NMPG APRN (an advanced practice nurse) → Attending Phys/Prov [long duration, large variance]
	Patient 4: V270, 64891, 66541, V0251	Outcome of delivery, single liveborn; Other current conditions classifiable elsewhere of mother, delivered, with or without mention of antepartum condition; High vaginal laceration, delivered, with or without mention of antepartum condition; Carrier or suspected carrier of group B streptococcus	

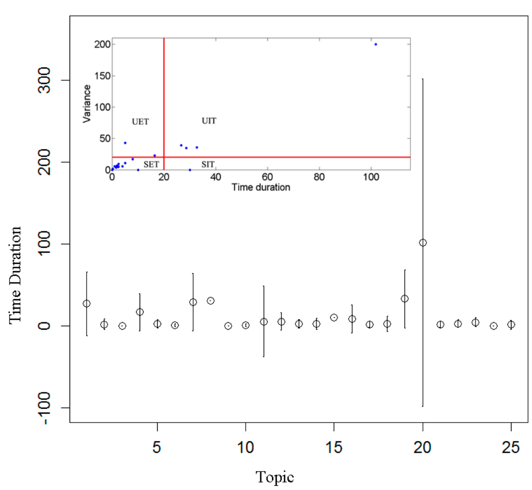

Figure 5. Average duration and variance of each of the 25 topics.

Case Study of a Complex Patient

The long duration and variance of topics may be due to the complexity of patients. To illustrate we conducted a case study of a patient who is characterized by seven topics.

Figure 6 shows the distribution of the number of topics needed to characterize a patient. It can be seen that most of the patients are characterized by 1 or 2 topics. We say patients who are associated with multiple topics are complex patients. Different topics characterize different types of patients. If a patient is associated with multiple types of topics, his condition should be quite complex.

To illustrate, let us consider one patient who was characterized by 5 topics and is associated with the conditions of *multiple myeloma*, *diabetes mellitus*, *esophageal reflux*, *urinary tract infection*, and *personal history of malignant neoplasm of breast*. This patient is associated with multiple blocks coming from different topics, of which we show two examples in Table 5. The first block is B3: [*Rehab Speech* → *Rehab Speech*] which has a long duration and large variance, while the other is B4: [*Medical Records – Scanner* → *NMH Physician-CPOE*], which has a short duration and variance.

Block B3 is associated with 3 reasons, and block B4 is associated with 5 reasons as shown in Table 5. The difference between reason R2 and R3 of the block B3 is *Med Rec Release of Info* and *Med Rec Quality*. The reason R3 associated with *Med Rec Quality* requires more time. If the reason following *Med Rec Quality* is *Attending Phys/Prov* (R1) instead of *Patient Care* (R3), then the time duration will increase. For block B4, the associated five reasons require a shorter duration, which suggests the block is stable and efficient.

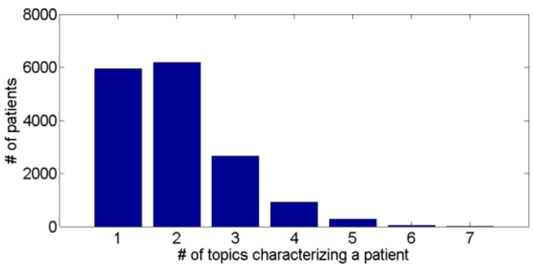

Figure 6. The number of topics necessary to characterize each patient.

Discussion

Clinical workflow modeling can be a challenging endeavor because of the complexity of patients, variability in the healthcare environment, and changeover in staff. Poor documentation of workflow can limit the adoption, or successful implementation, of EHR systems. Our work provides a framework to generate such workflows at multiple levels of granularity via data mining. Our evaluation of the duration of such workflows enables the clear partitioning of existing workflows into four types, depending on their length and variability. We believe that such a characterization of the system can enable an HCO to investigate and refine inefficient and unstable workflows.

We note that this is a pilot study on workflow extraction and modeling, but that the findings are promising as a roadmap towards the future. Specially, we believe that this type of investigation can be enhanced through several lines of research.

First, the reasons behind inefficient workflows clearly need to be studied in a more refined manner. Our work characterizes inefficiencies as a function of complexity in the patients and the surrounding workflows. For instance, one of our examples illustrates that residents assisting attending physicians (via access charts) may be a possible cause of inefficient workflows. Yet this explanation is limited in its explanatory power. Alternatively, or perhaps additionally, the reasons for inefficient workflows may include varying clinical urgency, variation in resource availability, varied clinical and system experiences of users, or even the design of EHR system itself. Further investigation of these factors will likely yield additional workflow representation optimization opportunities.

Table 5. Two typical blocks and their corresponding reasons and duration.

Block	Reason	Duration
B3	R1: Med Rec Quality → Attending Phys/Prov	440
	(the primary physician responsible for an inpatient's care)	450
	R2: Med Rec Release of Info → Patient Care (associated with nursing roles)	0 0
	R3: Med Rec Quality → Patient Care	460 10
B4	R1: Rehab Services Clinician → Rehab Quality Audit	0
	R2: Rehab Assigned Therapist → Rehab Quality Audit	0
	R3: Rehab Quality Audit → Rehab Services Clinician	0
	R4: Rehab Quality Audit → Rehab Assigned Therapist	0
	R:5 Rehab Assigned Therapist → Rehab Services Clinician	0

Second, the learned 25 topics need to be confirmed by clinicians. To do so, we will need to provide more nuanced contextual information about the patients associated with these workflows. Currently, the context of these patients is limited to the ICD-9 codes billed to insurance companies, which certainly does not cover the detailed conditions of a patient.

Third, the workflows our technique inferred are not associated with specific diseases. This will make it difficult for HCOs to determine where to invest in workflow optimization. We plan to construct the association study between workflows and diseases through data mining and machine learning technologies as a next step.

Conclusion

Modeling workflows for healthcare is challenging due to the complexity of clinical processes. In this work, we introduced a framework to model workflows at multiple levels of granularity. We illustrated that this framework can enable the categorization of workflows into four classes based on their duration: i) stable efficient, ii) unstable efficient, iii) stable inefficient and iv) unstable inefficient. We performed an extensive evaluation on Northwestern EHR event logs in the inpatient setting, where the results showed almost 95% of blocks are stable as well as efficient, and that over 80% patients are associated with efficient workflows. We further provided several illustrations of the reasons for inefficiency in a workflow and posited the main reason may derive from complexity of patients and the fact that Northwestern is a teaching hospital where residents are trained. Nonetheless, the reasons for inefficiency of workflows are diverse and will require additional contextual information (e.g., nuanced clinical data on patients, resource allocations of HCOs, and experiences of care providers) to further investigate and optimize workflows accordingly.

List of Abbreviations Used in this Paper: Rehab PT: Rehabilitation Physical Therapist; NMPG MD: A physician who belongs to the Northwestern Memorial Physicians Group; CPOE: Computerized Physician Order Entry; APRN: Advanced Practice Nurse Provider; SN-OR RN SC: Surgical Nurse Operating Room Service Coordinator (includes scheduling and patient experience issues); Radiology Mgr/RC: Radiology Manager Resource Coordinator (includes scheduling and patient experience issues); SN-RN/Customer Service: Surgical Nurse also Customer Service (includes registration and scheduling issues); RAD - Mgr/RC: Radiology Manager and Resource Coordinator (includes scheduling and patient experience issues).

Acknowledgements: This work was supported in part by NIH grants K99LM011933 and R01LM010207, and NSF CNS grants 09-64392 and CNS 13-30491.

References

1. Ash JS, Bates DW. Factors and forces affecting EHR system adoption: report of a 2004 ACMI discussion. *J Am Med Inform Assoc* 2005; 12: 8–12.
2. Bowens FM, Frye PA, Jones WA. Health information technology: integration of clinical workflow into meaningful use of electronic health records. *Perspect Health Inf Manag*. 2010; 7: 1d.
3. Unertl KM, Novak LL, Johnson KB, Lorenzi NM. Traversing the many paths of workflow research: developing a conceptual framework of workflow terminology through a systematic literature review. *J Am Med Inform Assoc* 2010; 17(3): 265–73
4. Cain C, Haque S. Patient safety and quality: an evidence-based handbook for nurses. Hughes RG, ed. Rockville (MD): Agency for Healthcare Research and Quality (US). 2008.
5. Egan M. Clinical dashboards: impact on workflow, care quality, and patient safety. *Crit Care Nurs Q*. 2006; 29(4): 354–61.
6. Mensah N, Sukums F, Awine T, et al. Impact of an electronic clinical decision support system on workflow in antenatal care: the QUALMAT eCDSS in rural health care facilities in Ghana and Tanzania. *Glob Health Action* 2015; 8: 25756.
7. Ramaiah M, Subrahmanian E, Sriram RD, Lide BB. Workflow and electronic health records in small medical practices. *Perspect Health Inf Manag* 2012; 9: 1d.
8. Unertl KM, Weinger MB, Johnson KB, Lorenzi NM. Describing and modeling workflow and information flow in chronic disease care. *J Am Med Inform Assoc* 2009; 16(6): 826–36
9. Zhang H, Mehotra S, Liebovitz D, Gunter CA, Malin B. Mining deviations from patient care pathways via electronic medical record system audits. *ACM Transactions on Management Information Systems*. 2013; 4(4): 17.
10. Ghattas J, Peleg M, Soffer P, Denekamp Y. Learning the context of a clinical process. *Proc Business Process Management Workshops* 2010; Springer LNBIP 43: 545–56.
11. Webster C. EHR business process management: from process mining to process improvement to process usability. *Proc Healthcare Systems Process Improvement Conference* 2012.
12. Bouarfa L, Dankelman J. Workflow mining and outlier detection from clinical activity logs. J *Biomed Inform* 2012; 45(6): 1185-90.
13. Mans RS, van der Aalst WMP, Vanwersch R, Moleman AJ. Process mining in healthcare: data challenges when answering frequently posed questions. *Proc Process Support and Knowledge Representation in Health Care Workshop* 2013; Springer LNCS. 140-53.
14. Mans RS, Schonenberg H, Leonardi G, et al. Process mining techniques: an application to stroke care. *Stud Health Technol Inform* 2008; 136: 573-8.
15. Stahl, JE, Sandberg WS, Daily B, et al. Reorganizing patient care and workflow in the operating room: a cost-effectiveness study. *Surgery* 2006; 139(6): 717-28.
16. Malhotra S, Jordan D, Shortliffe E, Patel VL. Workflow modeling in critical care: piecing together your own puzzle. *J Biomed Inform* 2007; 40(2): 81-92.
17. Partington A, Wynn M, Suriadi S, Ouyang C, Karnon J. Process mining for clinical processes: a comparative analysis of four Australian hospitals. *ACM Transactions on Management Information Systems* 2015; 5(4): 19.
18. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. Journal of Machine Learning Research. 2003; 3: 993-1022.
19. Chen Y, Ghosh J, Bejan CA, et al. Building bridges across electronic health record systems through inferred phenotypic topics. *J Biomed Inform* 2015; 55: 482-493.
20. Chen Y, Nyemba S, Zhang W, Malin B. Specializing network analysis to detect anomalous insider actions. *Secur Inform* 2012; 1: 5.
21. Chen Y, Nyemba S, Malin B. Auditing medical records accesses via healthcare interaction networks. *AMIA Annual Symp* 2012: 93-102.