ORNL/TM-2021/2284

Comparative Assessment of Data-driven Process Models in Health Information Technology

Hilda B. Klasky
Ozgur Ozmen
Femi A. Omitaomu
Mohammed Olama
Laura Pullum
Addy T. Malviya
Teja Kuruganti

MAY 2021

OAK RIDGE National Laboratory

ORNL IS MANAGED BY UT-BATTELLE LLC FOR THE US DEPARTMENT OF ENERGY

ORNL/TM-2021/2284

Computer Science and Engineering Division

COMPARATIVE ASSESSMENT OF DATA-DRIVEN PROCESS MODELS IN HEALTH INFORMATION TECHNOLOGY

Hilda B. Klasky
Ozgur Ozmen
Olufemi Omitaomu
Mohammed Olama
Laura Pullum
Addi T. Malviya
Teja Kuruganti

May 2021

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831-6283
managed by
UT-BATTELLE LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ACRONYMS

BPN	Business Process Notation
CDW	VHA's Corporate Data Warehouse
HIPAA	Health Insurance Portability and Accountability Act
HIT	Health Information Technology
OASIS	Organization for the Advancement of Structured Information Standard
ORNL	Oak Ridge National Laboratory
PM	Process mining
PM4CA	Process mining for conformance analysis
RMSD	Root mean square deviation error
UK	United Kingdom
VHA	US Veterans Health Administration

ABSTRACT

Background and Purpose

Process mining for conformance analysis consists of comparing a reference process model against a data-driven process model generated via log files from information technology systems. However, in the absence of a complete reference process model, we found no suggested approaches in the literature to address the need for evaluating process conformance among different healthcare facilities to assess standardization of care. Our goal is to find similarities and dissimilarities in data-driven process models among US Veterans Health Administration (VHA) facilities that can be indicative of patient safety issues. Our hypothesis was that the analysis would not produce statistically significant differences in outcome.

Methods

We present a unique implementation of conformance analysis in process mining that consists of combining process mining, process mapping and statistical metrics. We illustrate our approach by applying it to the analysis of two clinical radiology order process models generated from healthcare data provided by two similar facilities in the VHA.

Results

The comparative assessment showed that about 70% of the orders completed successfully and 30% were not completed due to policy and duplications. Our analysis found a good statistical correlation between both facilities, as the Spearman's correlation coefficient between facilities for the frequency of cases per total hours was 0.87879, for the frequency of cases by state transition was 0.79702 and for the throughput time per state transition was 0.63582. Additional statistical analyses using the Mann-Whitney U test and the root mean square error both produced values that were not significant.

Conclusions

The foregoing approach validated our hypothesis by demonstrating a good statistical correlation of data describing the flow of clinical radiology orders absent a credible reference model. Finding good agreement between both facilities was important in confirming that the clinical orders flow in a similar manner, suggesting standardization of care.

Keywords: Process mining, conformance analysis, health information systems, anomaly detection, health information technology hazards, data mining, data analytics, conformance assessment

1. INTRODUCTION

In healthcare organizations, Health Information Technology (HIT) processes [for example, the radiology order process within the VHA] often can be observed only on a case-by-case basis. This is a significant limitation because, while helpful on an individual basis, it does not produce a comprehensive understanding of the same process involving different providers, services, or facilities. Thus, large scale HIT processes within healthcare facilities have been essentially opaque to individuals/groups having

responsibility for assessing whether those processes conform to procedures/objectives of the subject organization and that there exists standardization of care among the different facilities.

1.1 CONFORMANCE ANALYSIS OF HIT PROCESSES

The motivation for performing such conformance analyses of clinical order processes is to identify possible anomalies that could cause patient harm from HIT, which we define as a hazard. It is also heightened by the recognition that hazards can be associated with deviations from an established process. The hazard we are investigating in this study is when the clinical service of a patient is not fully completed. Ideally, a clinical order should move along the process via an expected pathway. However, deviations can occur in the form of the process ends before the intended outcome, such that a clinical order is placed, yet never moves forward in the process. If the process is extended, there is potential that the intended clinical task is delayed or perhaps not completed. Unintended deviations from the expected clinical order flow may have implications for patient safety.

In the context of the current study, if we understand and identify what is normal and correct with the flow of clinical order processes, and what can precipitate unsuccessful flows, we will be able to identify when certain conditions deviate from normal or are completely incorrect. Then, we can create observation points for defined rules/metrics that serve as anomaly detectors and identify correct deviations such as purposeful discontinuations of clinical orders. The study presented herein is a continuation to our previous works in [1-3].

Our objectives are focused on HIT processes at the VHA and include the following:

- To discover the clinical order process for each of the four clinical domains: consults, radiology, laboratories and medications,
- To identify any observable rules between events in those processes,
- To identify deviations that may be impactful to the service completion process and thus may influence patient safety.

1.2 PROCESS MINING FOR CONFORMANCE ANALYSIS

Process mining (PM) provides the tools to better understand large scale processes to identify discrepancies between processes as envisioned and processes as performed. It can be used to verify conformance adherence between a model process and an observed process. We want to know how faithful the model is to our understanding of the true underlying data-driven process precipitated by actors interacting with the real world. We as humans are data generating agents and from a causal perspective, the latter process is how the world behaves.

Process mining for conformance analysis (PM4CA) consists of comparing a defined reference process model against a data-driven process model constructed via log files from information technology systems. That approach assumes the existence of a reference process model. However, our experience working with healthcare data is that, although there are many well documented high-level clinical processes, not all processes are documented; and those which are documented, may be incomplete or may become rapidly out of date due to fast evolving circumstances (personnel changes, evolving technology, new products, etc.). Moreover, some processes change so fast that segments of a process are either missing or not documented.

Although there are prescriptive/descriptive processes or workflows for healthcare services at the VHA, our studies revealed that each clinical order follows an individual, sometimes unique, path. Also, those clinical order flows are not documented in the Business Process Notation (BPN) flowcharts maintained by the VHA; what the BPN flowcharts did document was the clinical service. As a result, we observed the following discrepancies:

- Some observed behaviors in the logs were not captured in the VHA's BPN clinical service flowcharts; for example, holds and discontinuations.

• Some behaviors depicted in the VHA's BPN clinical service flowcharts were not included in the selection of events to create the logs (for example, reviewing prior exams, education materials given to the patient, patient consent, patient preparation, resulting imaging, and external imaging review).

Thus, our studies determined that a "reference model" was not established for clinical orders. Furthermore, it was perceived that a complex effort would be required to map VHA's BPN clinical service flowcharts to several database schemas in the VHA's Corporate Data Warehouse (CDW). Instead, this absence of a reference model to track the clinical order process for conformance was overcome in our work by using a unique implementation of conformance analysis in PM described in the following section.

1.3 A NOVEL APPROACH

The approach presented herein resulted from our need to address the absence of a reference model to track the clinical order process for conformance, as we had discovered when analyzing the clinical orders process through HIT systems, from creation to completion, to identify anomalies that could endanger patient safety. In this study, we present a unique implementation of conformance analysis in PM which consists of using two clinical order process models generated from healthcare data from similar facilities in the VHA.

For each facility, two process models were generated. One was from the raw data from CDW and another generated from the association of the events in the raw data to states using the OASIS [4] human task state transition diagram as reference model. Here, our focus is on PM for anomaly detection for patient safety. We describe some of the challenges that we encountered, and lessons learned.

PM4CA examines adherence of executed processes to model processes. With PM4CA, it is possible to detect process failures and breakdowns that normally would not be found or would only be detectable at a later stage with potentially more severe impact on healthcare delivery. As such, PM4CA provides valuable insights into non-adherence within a time frame that is unprecedented. In addition, since process failures can lead to serious harm to the patients, this analytical approach has the potential to improve patients' safety prior to harm, besides reaching large numbers of patients.

PM4CA required us to correlate as many steps as defined in the VHA workflows to CDW data. However, to perform a fair comparison, both processes must have analogous sets of activities and event sequences. The differences and deviations are documented and reported as part of the analysis.

1.4 CHALLENGES BEING TACKLED

In this study, we aim to answer the following research questions:

i. Can we apply PM4CA between process models of two different VHA facilities? If so, what do the process models of the two selected facilities look like?
ii. Are there deviations in data-driven process models from different facilities? And what are the exact differences?
iii. Are the observed deviations impactful to the clinical service completion that could be harmful to the patient?

We hypothesize that, in a broad comparison of two similar facilities, there should not be statistically significant differences between the two in terms of the PM of clinical order processes.

1.5 OUR CONTRIBUTIONS

Our scientific contributions are as follows:

1. We present a unique implementation of PM4CA, since this is typically used when a reference model is defined, and one is comparing the data-driven model to that reference model.

2. Our goal is to develop a custom approach using PM4CA to identify cases when the service of the patient is not fully completed, where we combine PM with process mapping and statistical metrics.
3. We performed the study with VHA healthcare data. We analyze process models designed by pathways that are discovered from the data and background knowledge of the mapping to the human task state transition (OASIS [4]). Then, we compare the two facilities' process models and statistics for deviation, rather than comparing a data-driven model to a reference model for deviation.
4. We illustrate our findings herein with a use case for radiology orders.

This paper is structured as follows:

- In the related work section, we present related PM4CA work.
- In the methods section, we present our approach.
- In the results section, we present the results of the comparative assessment.
- Finally, in the last sections, we provide discussion, limitations, future work and conclusion.

2. RELATED WORK

To our knowledge, PM was presented originally by Agrawal in [5] and has taken more prominence thanks to the ample work of Van Der Aalst [6-17], and his numerous collaborators. PM can be seen now as a relatively new research discipline that is set between data science, machine learning and data mining, on the one hand, and process modeling and analysis, on the other hand. PM adds the process perspective to data mining [18-21]. It comprises the application of several deterministic, statistical and 'intelligent' algorithms to discover association rules between steps, and from the association rules, PM discovers the process in log files from information systems [6, 7]. In addition to discovering the process, PM approaches generate metrics of frequency, performance and conformance [22].

The number of research articles related to the application of PM in healthcare follows a growth trend. In the published literature, we found several works related to literature review on PM specifically for healthcare [23-25]. All of them present evidence, at the time of publication, that the field has been growing in the previous decade. For example, in 2016, Rojas et al [23] presented a literature review in which he identified 74 articles concerned with PM in healthcare. In the same year, Ghasemi et al [25] identified 2371 publications related to PM, 168 of which were related to PM in healthcare. In 2018, Erdogan's systematic mapping of PM studies in healthcare analysis [24] found greater distribution of studies for healthcare processes than for clinical pathways. Both Rojas's and Erdogan's works report that most PM for healthcare studies is in the oncology medical field, followed by surgery. Erdogan's study found that most PM for healthcare works refers to services in departments, or clinical pathways and clinical services. Williams published a literature review of PM in primary care in [26]. Dos Santos's study in [27] is the most recent systematic mapping of general PM that states conformance analysis is the second most active research topic. Kusuma presented a literature review of PM in cardiology in [28].

Thus, PM has been applied to numerous studies in healthcare [18-21, 29]. Recent examples of such studies follow. In [30] Kusuma et al presented a novel application of PM focused on a feasibility study of disease trajectories. Gatta et al presented a framework for event log generation and knowledge representation for PM in healthcare [31].

Since inception, PM was meant to be used in decision support. For example, in [13] Van Der Aalst stated that "The outcome of PM is a better understanding of the process and accurate models that can safely be used for decision support because they reflect the reality". Another example is the work by Ying Liu in [32], where a systematic and generic business process simulation approach for operational decision support was presented, in which processes were modeled using graphs and nodes were events from workflow logs.

PM4CA had its origins in studies of graph similarity and process model similarity. Since the early 2000s, there have been studies on graph similarity and business process model similarity [33] measurements; for example, Agrawal presented a definition of conformance graph in [5] and process model similarity. Recent examples of graph matching algorithms for business process include the work of [33] by Dijkman et al, and a study in workflow simulation for operational decision support using event graph through PM by Liu et al in [32]. Graph comparison is a prominent field and there are many research articles in this topic. We would like to mention the excellent work of Wills et al who recently published their work on metrics for graph comparison as a guide for practitioners in [34].

PM4CA is well represented in [6], [35], [36] and [37]. Specifically, Rozinat [35] describes work on conformance checking of process based on monitoring real behavior. Munoz-Gama's book [36] examines conformance checking and diagnosis in PM by comparing observed and modeled processes. Carmona's book on Conformance Checking [37] focuses on relating processes and models and provides an extensive presentation on conformance techniques based on the principle of alignment or the shortest path through the process model; the latter also presents decomposition and heuristic checking techniques. An approach for conformance checking that uses a decomposition technique for large processes is presented in [38].

PM4CA has been implemented in several plugins in the open source PM toolkit ProM[1] [16], and some examples are the following: Rozinat's [35] conformance checker, Munoz-Gamma and Carmona's ETConformance checker [39], and in [40] Burattin and Carmona presented a framework for online conformance checking, to name a few. In the private arena, the companies Celonis[2] and Uipath[3] provide modules for conformance checking in their PM solutions.

PM4CA in healthcare has been applied largely to both the analysis of healthcare processes and clinical guidelines. Recent studies include the following: Rinner et al [41] presented a study on long running processes in the context of melanoma surveillance; Fernandez-Llatas et al [42] presented a study analyzing medical emergency processes; Badakhshan [43] et al presented a study of PM for process analysis, conformance and improvement for the process of pre-hospital emergency department; Tamburis et al [44] presented an approach to investigate conformance between a log file and a simulation tool's generated data while linking PM to discrete event simulation modeling; Kukreja et al [45] applied PM4CA to compare different PM approaches using a sepsis case study; Van Dongen et al [46] presented a conformance checking approach focused on mixed-paradigm process models; Asare et al [47] presented a conformance analysis between processes and the workflows on hospitals; Helm et al [48] introduced a modeling representation method based on multi-perspective declarative PM and a novel algorithm to trace and verify conformance; and Marazza et al [49] presented an approach to compare process models for patient populations and a case study in breast cancer care, by using cross-log conformance checking and standard graph similarity metrics.

Because of the abundance of the data and the complexity of the healthcare processes, new methodologies have recently surged in applications of PM to healthcare; for example, Pereira's work in [50] presents a case study in a tertiary hospital of PM project methodology in healthcare. In addition, recommendations have been published about the use, application, challenges and lessons learned; see for example [51].

Foregoing paragraphs present evidence of a growing trend in the number of studies applying PM for healthcare. However, the use of PM for clinical decision support to improve patient safety is still an area where more research needs to be performed. Our observations are consistent with the work of Williams et al in [52], whose findings imply a lack of work focused on primary care for patient safety. Williams et al [52] presented a study of PM in primary care in the UK, where they studied how to avoid adverse events due to hazardous prescribing. However, we were not able to find other references of the application of PM4CA in healthcare dedicated to identifying possible anomalies that could cause patient harm from HIT

[1] https://www.promtools.org/
[2] https://www.celonis.com/
[3] https://www.uipath.com/

systems. To our knowledge, at the time of writing this paper, Williams et al [52] and this paper are the first studies focused on the latter topic.

3. DATA AND METHODS

3.1 DATA

We focused our analysis on a selection of radiology orders from the VHA's CDW, which stores healthcare data. The orders were randomly selected for two different facilities. The observation window is from January 1st to December 30th, 2019. VHA has a way to classify their facilities based on their complexity [53]. The selected facilities' complex level is Type 1A, that is, these facilities have high complexity in terms of patient risk, high levels of teaching and research, high number and breadth of physician specialist; finally, they contain level 5 intensive care units. This study was performed by creating and analyzing event sequences of radiology orders in two enclave computer ecosystems: a) the Knowledge Discovery Infrastructure and b) the BlueRidge Collaborative Environment for Research Integration. Both of those computer facilities are located at Oak Ridge National Laboratory[4]. Researchers access and analyze HIT data within specifically augmented computer power and storage in these two secure enclaves, which follow Health Insurance Portability and Accountability Act (HIPAA) rules compliance and strict privacy and cybersecurity regulations to ensure the data stays secure. The latter is a prime requirement during the analysis of VHA's healthcare data.

3.2 METHODS

Figure 1 depicts each step of our approach. The steps of our approach are described in Table 1:

Table 1. Description of the comparative assessment approach presented herein.

Step No.	Step	Description
1.	Data identification and selection	Data identification and selection was performed by using the Entity Relationship diagrams of the VHA's CDW database and doing exploratory querying. That included identifying dates, activities, and status type columns of the different data domains for clinical orders.
2.	Data extraction	We wrote structured query language (sql) scripts to extract the data, preprocessed and stored in new database schemas. The creation of these smaller database schemas allowed us to work with reduced dataset selected from the CDW.
3.	Data preparation	Data preparation was performed within the schemas created in Step 2 within CDW. We wrote sql scripts to generate sequences of events for each case by following the PM methodology; thus, we appended a unique case id to each sequence of events. Then, we generated text files formatted with comma-separated values (log files).
4.	Process mining	The log files generated in Step 3 were imported to the Disco software tool [22], where the Fuzzy mining algorithm developed by Gunther [54] is implemented, to discover the data-driven process models of two VHA facilities. These process models were used in this study.

[4] https://www.ornl.gov/

5.	Process mapping	a) In order to simplify the very complex process model maps generated in Step 4, which had close to 50 different activities, we created state transition rules base on the OASIS Human Task state transition diagram in reference [4]. That state transition diagram describes the conditions when human tasks move to different states. This is important because we believe that it can be applied to many processes, in particular to our clinical orders processes. b) Next, we identified association rules between activities from the sequences of events in the log files to the OASIS task' states. This way, we mapped the OASIS Human Task states to clusters of activities based on transition rules which we defined specifically for the dataset as presented in [1] and [2]. c) The rules in b) were implemented in python scripts which generated a new set of log files. The new log files, with state transitions per case, were imported to Disco to create the state transitions process model maps. Those maps are shown in Section 4 as Figures 3 - 6.
6.	Comparative assessment	The assessment was performed by comparing, contrasting and analyzing descriptive statistics, activities statistics and frequency and performance process model maps against one another. Then, descriptive statistics were used to compute the Mann-Whitney U-test, the Spearman's correlation and the root mean square results to compare the two facilities, and to estimate correlation and error values.

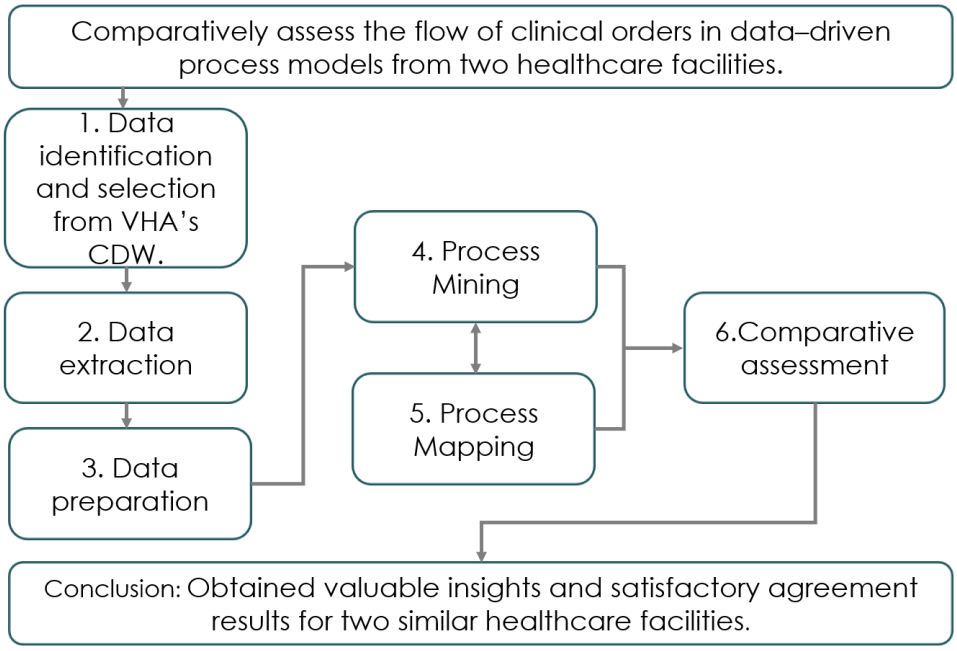

Figure 1. The six-step approach followed to perform this study.

4. RESULTS

Given the abundance of the data and the large size of the process models generated from the raw log files of activities, we limit this section to present mostly results of the mapped, simplified data from the OASIS state transition process model maps, as the latter are more generalizable.

4.1 DESCRIPTIVE STATISTICS

Table 2 Descriptive Statistics of Radiology Order Datasets.

	Facility A	**Facility B**
Events	61,122	111,911
Cases	3,137	5,482
Activities	39	47
Median duration	18.5 days	20.3 days
Mean duration	25.6 days	26.3 days

Table 2 presents the descriptive statistics of the datasets. Facility A had ~3K cases, and 39 different activities (each activity is a feature selected from the data set and resembles a step in the life cycle of the order). Observe how the median and the mean case durations are close for each facility, which shows that there are not too many outliers cases. For the same observation window, Facility B had almost twice the number of cases compared to Facility A, with ~5.5K cases. Facility B had 47 different activities.

4.2 ACTIVITIES

Figure 2 presents the clustered column chart for facilities A and B. On the left *y*-axis, we can see the frequency and on the right *y*-axis we can see the relative frequency. We can observe outstanding similarities among the frequencies of the activities for both facilities. Observe that there are three main groups or clusters of activities' frequencies. The first one to the left, is related to activities during the creating and termination stages. The second one, in the center, is related to the radiology domain when the service is scheduled, when the service takes place and when the results are recorded. Finally, the last cluster, to the right, mostly concerns activities related to rare radiology exams, which are less frequent activities or activities that almost never occurred.

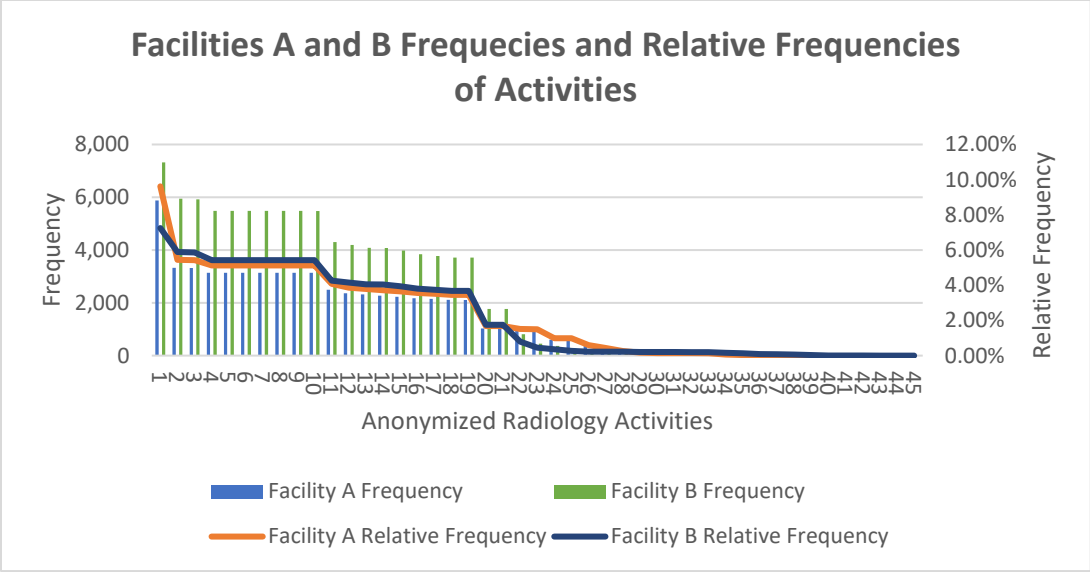

Figure 2. Facilities A and B Frequencies and Relative Frequencies of Radiology Activities.

4.3 PROCESS MODEL MAPS

Figure 3 and Figure 4 present the (OASIS) Task state transition for Facility A and Facility B, respectively. Only frequencies are shown in these maps. These charts present the most dominant connections between activities. Notice that these are still complex graphs; however, they only present the activities and paths from the most frequent process flows of radiology orders. Each box represents a cluster of activities, i.e., activities that were grouped based on times of occurrence and similarity. Each arrow or arc presents the process flow between two task's states. Observe here the different thickness of the arrows. The thickness of the arrow represents the frequency, thus, the thicker the arrow, the larger the number; and the darker the box, the larger the frequency of visiting that activity. Both charts provided us with a good understanding of the radiology's order process flow between facilities; the flows for the two facilities are the same. Both facilities present high levels of similarity. We observed that 10 different activities present the highest number of frequency visits; all of them are the same on the process maps of both facilities. The thick arrows show that most radiology orders move smoothly from Created to Ready, then from Ready to InProgress and finally, from InProgress to Complete. We can also observe that very few orders transition from Created to Error. Even more rare are cases that go from Complete to Failed. Most cases that are discontinued transition from the Ready, Reserved or InProgress states to the Exited state.

What about the time spent between the different activities of the radiology order process? Figure 5 and Figure 6 present the (OASIS) task state transitions for Facility A and Facility B, respectively, with throughput times. These process maps present the total duration (as the primary metric) and the mean duration (as the secondary metric – in smaller size font). Again, in these performance charts, the thickness of the arrows and the shadow of the boxes increase as the frequencies increase. Thus, thicker bright red arrows have higher performance values. Observe that bottlenecks are present in two main cases: a) in cases of Exited radiology orders for both facilities, and b) in some cases when the radiology order process transitions from Ready to Reserved. Further analysis of the latter case shows that the activity selected in this stage appears to schedule radiology orders in the future, when the radiology test appointment is desired but, of course, not guaranteed.

The generated process model maps provided information related to the successful and not successful termination states which are presented in Table 3. There, we can compare and contrast side by side termination states for both facilities. Our results presented in Table 3 show that, in both facilities, about 70% of the cases were completed successfully, and about 30% of the cases were incomplete, i.e., did not

complete successfully. We performed further classification regarding the incomplete cases for three different unsuccessful termination cases: Error, Failed, and Exited. We defined the unsuccessful termination cases as follows:

a) Error cases have a discontinuation activity after activities related to the Created state.
b) Failed cases have a discontinuation activity after activities in the Completed state.
c) Exited cases have a discontinuation activity after activities in the Ready or Reserved states.

Table 3 also shows that Failed and Error cases are very rare. Most unsuccessful cases are Exited cases. After performing further analysis on those cases, it was determined that those discontinuations are concerned mostly with policy and some duplicated records that arise due to imports from other subsystems. Consequently, no consequences of patient harm were identified from this analysis.

Table 3. OASIS Transition States Termination Frequencies and Relative Frequencies of Facilities A and B - Radiology Orders

Case Termination Classification		**Facility A**		**Facility B**	
		Total Frequency	Relative Frequency	Total Frequency	Relative Frequency
Total Cases		3137	100%	5482	100%
Complete		2111	67.29%	3715	67.76%
Incomplete	Failed	1	0.01%	1	0.01%
	Exited	1020	32.5%	1757	32%
	Error	7	0.2%	10	0.18%

The frequency and performance process maps (Figures 3 - 6) presented in this section provided a better understanding of the actual radiology order process flow at the VHA and answered our research questions:

- What do the process models of the two selected facilities look like?
- Are there deviations in data-driven process models from different facilities?

By observing the process model maps, we can state that they are very similar. In addition, the figures and tables in this section also answered our research question, i.e.,

- What are the of exact differences?

by providing metrics of the differences in frequencies. The analysis presented in Table 3 answered another research question, i.e.,

- Are the observed deviations impactful to the clinical service completion that could be harmful to the patient?

by providing evidence that cases with discontinuation activities were discontinued due policy and a few duplications due to imports from other sub-systems and did not impact any real patient.

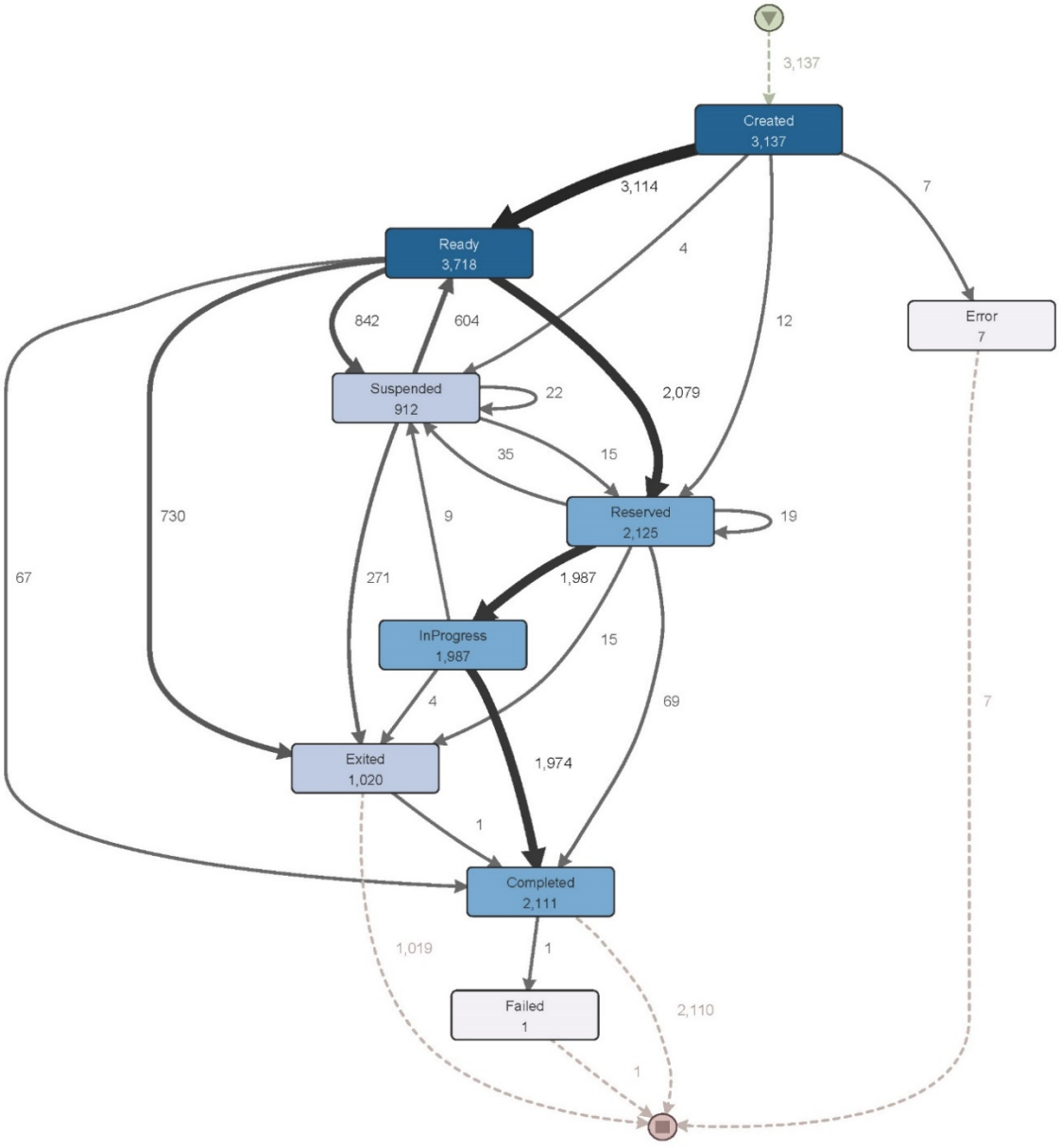

Figure 3. (OASIS) Task State Transitions for Facility A - Frequency Process Map for Radiology.

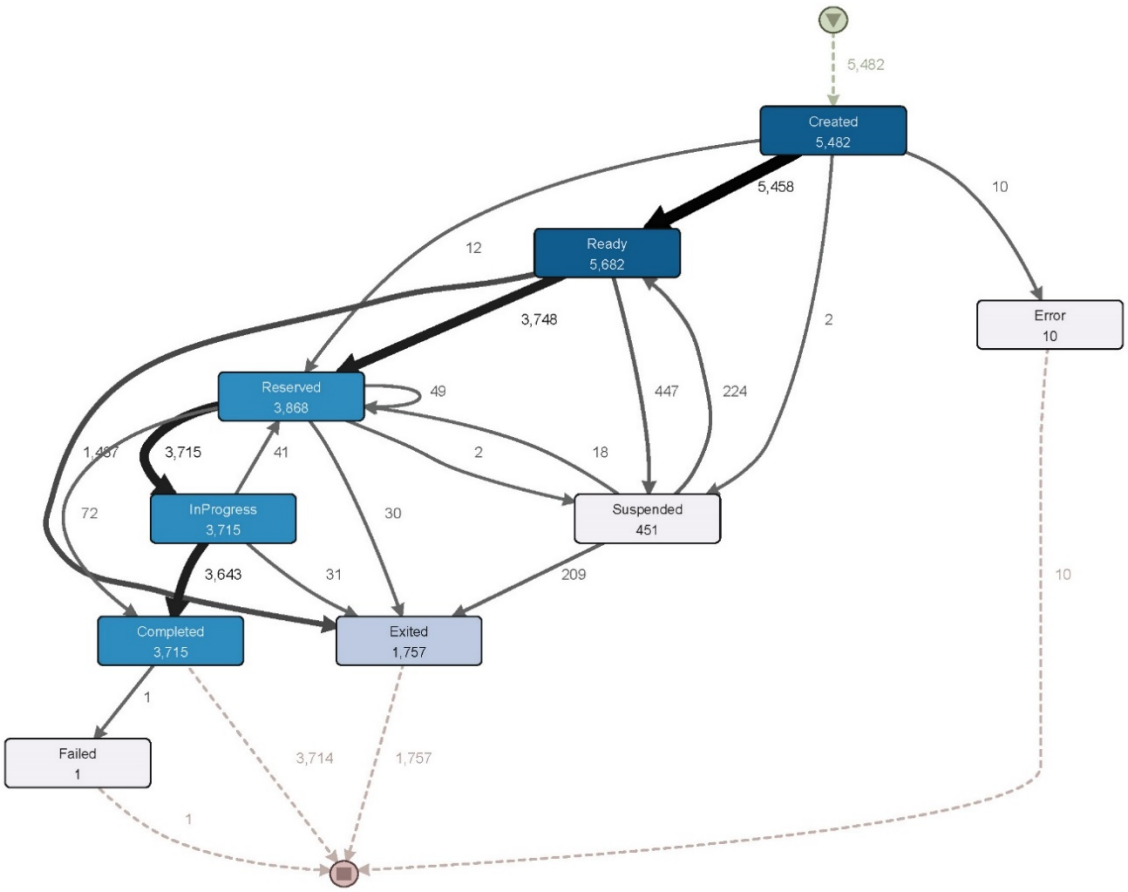

Figure 4. (OASIS) Task State Transitions for Facility B - Frequency Process Map for Radiology.

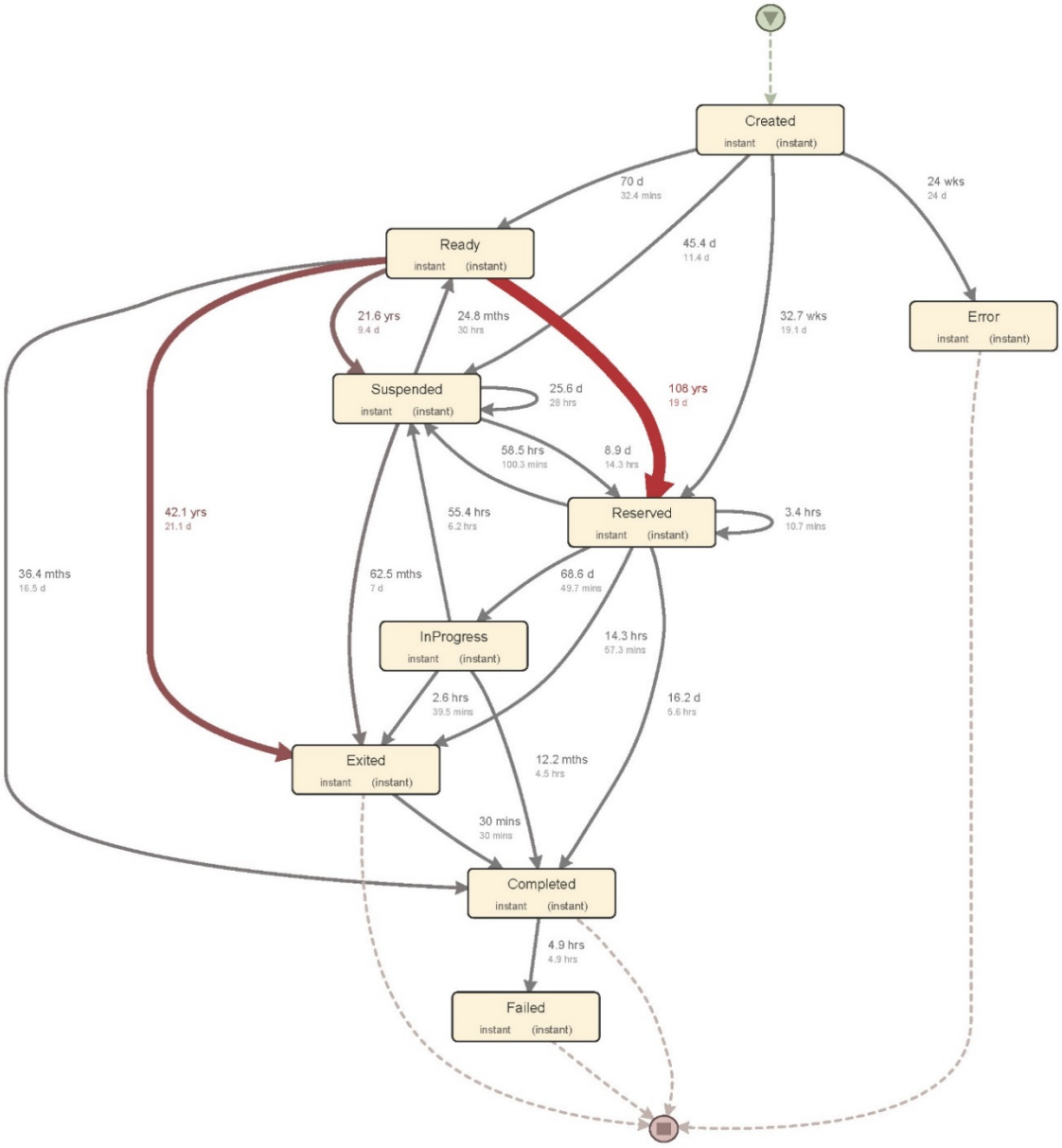

Figure 5. (OASIS) Task State Transitions for Facility A - Performance Process Map

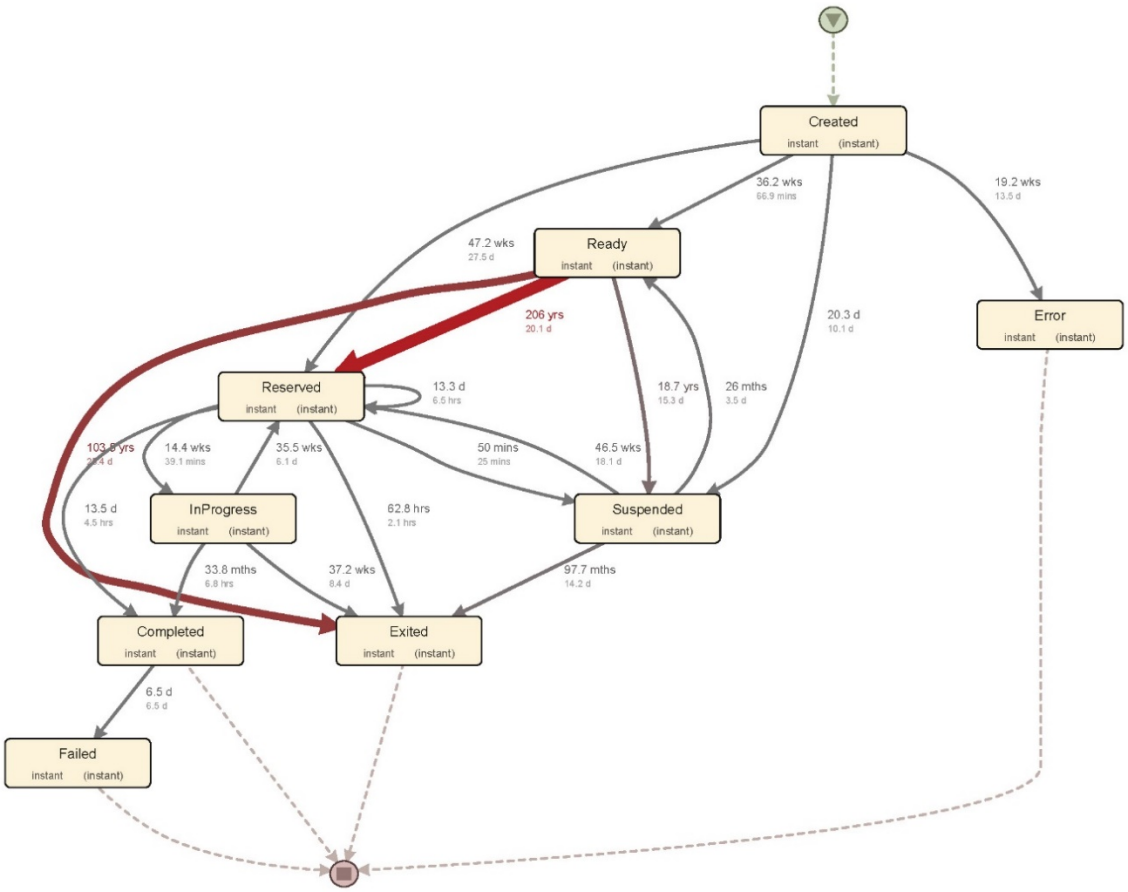

Figure 6. (OASIS) Task State Transitions for Facility B - Performance Process Map

4.4 STATISTICAL ANALYSIS

We performed a Mann-Whitney *U* test [55] to compare Facilities A and B for the cases presented in Table 4. Mann-Whitney *U* test is a nonparametric test that allows two data groups to be compared without assuming that values are normally distributed. The significance level was 0.05 and two-sided. For the case of *frequency of case per total hours from process start to end* the *U-value* is 24. For $p < 0.05$, the critical value of *U* is 23, thus, we conclude that the result is not significant. The *Z-Score* is -1.92762 and its *p-value* is 0.0536; therefore, the result is not significant at $p < 0.05$. For the case of *frequency of cases by state transition*, the value of *U* is 493.5, the Z-Score is 0.64764 and its *p-value* is 0.5157. Consequently, the result is not significant at $p < 0.05$. Finally, for the case of the *throughput time per state transition* the value of *U* is 474.5, the *Z-Score* is 0.8913 and its *p-value* is 0.3746. Thus, the result is not significant at $p < 0.05$.

Table 4. Mann-Whitney U Test for Facilities A and B.

	U-value	***Z-Score***	***Z-score's p-value***
1) Frequency of cases per total hours from process start to end	24	-1.92762	0.0536
2) Frequency of cases by state transition	493.5	0.64764	0.5157

3) Throughput time per state transition	474.5	0.8913	0.3746

In addition, we performed a correlation analysis using Spearman's correlation coefficient [56]. Spearman's Rho is a non-parametric test used to measure the strength of association between two variables, where the value of $r = 1$ means a perfect positive correlation and the value of $r = -1$ means a perfect negative correlation. The results in Table 5 show high correlation between Facilities A and B for the three cases analyzed. The Spearman's correlation coefficient between Facilities A and B for the *frequency of cases per total hours from process start to end* is 0.87879, for the *frequency of cases by state transition* is 0.79702 and for the *throughput time per state transition* is 0.63582. The significant value of this coefficient is very small at less than 0.05 for each case. Thus, we rejected the null hypothesis that there is no correlation, i.e. there is an association, between Facilities A and B for the cases outlined in Table 5. Note that no *p-values* were corrected for multiple testing.

Finally, the root mean square deviation (error) (*RMSD*) [57] is also presented in Table 5 for each case analyzed. The *RMSD* column shows that, in all cases, the *RMSD* is close to 0, i.e. almost a perfect fit of the data between the two facilities. The *RMSD* for the *frequency of cases per total hours from process start to end* is 0.026478, for the *frequency of cases by state transition* is 0.012718145 and for the *throughput time per state transition* is 0.039521815.

Table 5. Spearman's Correlation Coefficient (Rho) and RMSD Analysis for Facilities A and B.

	r_s	*p-value (2-sided)*	**RMSD**
1) Frequency of cases per total hours from process start to end	0.87879	<0.001	0.026478
2) Frequency of cases by state transition	0.79702	0	0.012718145
3) Throughput time per state transition	0.63582	<0.001	0.039521815

The data used in the generation of Figures 3-6 was also the source of Tables 4-5. Each row in Tables 4-5 corresponds, respectively, to Figure 7, the *relative frequency by duration from process start to end*; Figure 8, the *relative frequency by state transition*; and Figure 9, *the throughput time per state transition in days* for Facilities A and B. Notice that because the volume of data is so large, we included only a few cases in those figures to illustrate our point.

Figure 7 presents the relative frequency by duration from process start to end. We can observe the distribution of the frequencies. For Facility A, most cases had a duration between 574 and 1526 hours. For Facility B, most cases had a duration between 0-574 hours, and 1050-1526 hours.

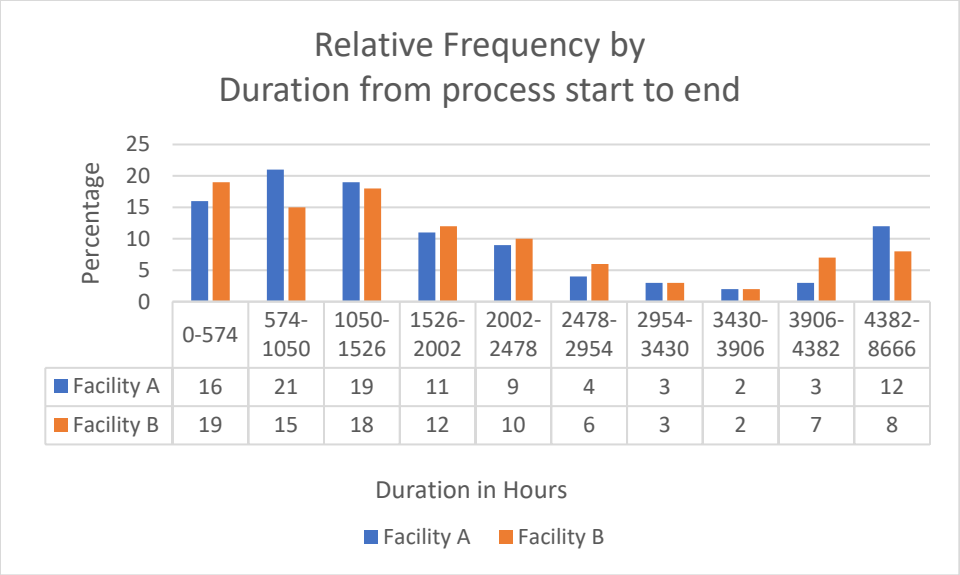

Figure 7. Relative Frequency by Duration from Process Start to End in Hours for Facilities A and B.

In Figure 8, which shows the six most common cases of relative frequency by state transition, it is observed that the two facilities had close relative frequencies in both the Created to Reserved and Ready to Reserved transitions. There was a notable difference between facilities in the Ready to Suspended, Suspended to Exited and Ready to Exited transitions.

In Figure 9, which shows the most common cases of throughput time per state transition, we can observe a notable difference between facilities in the Ready to Suspended, Suspended to Exited, and Suspended to Reserved transitions. This chart also presents close values in the transition between Created to Reserved throughput times.

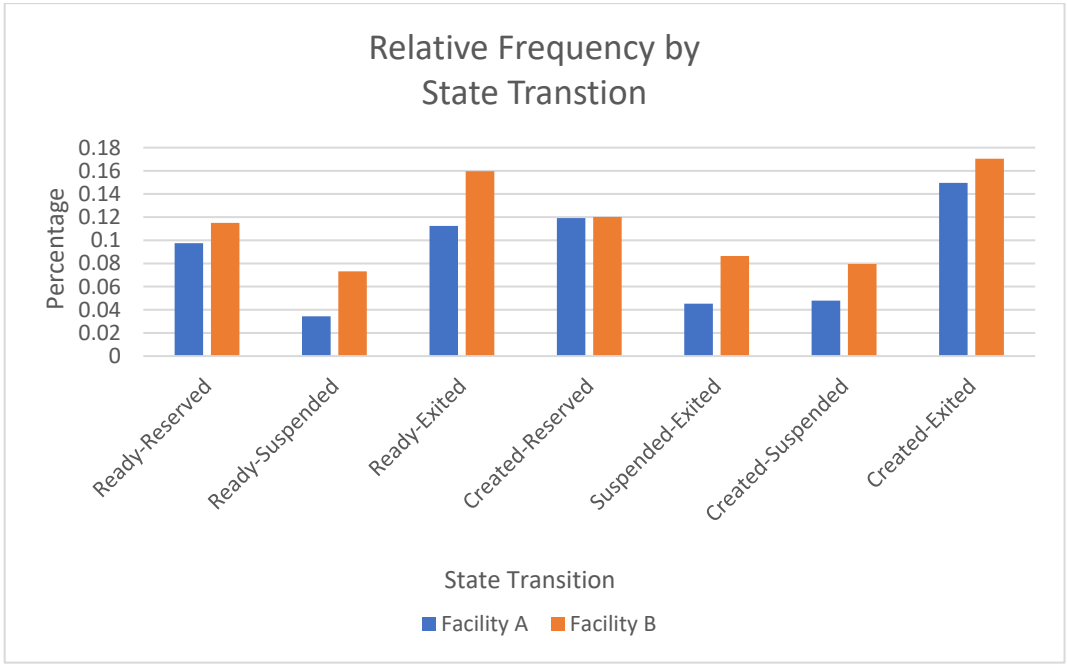

Figure 8. Frequencies by State Transition for Facilities A and B - most frequent cases

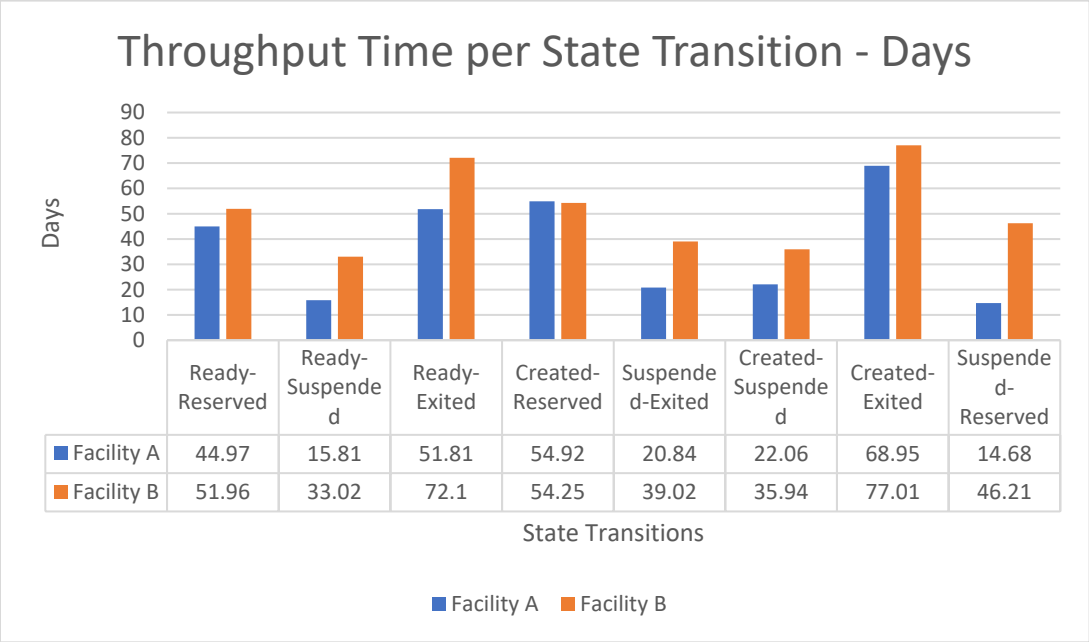

Figure 9. Throughput Time per State Transition in Days for Facilities A and B - most frequent transitions

The OASIS mapping of the data presented in Figures 3-6 and in Table 3, and the statistical analysis we conducted on those datasets using the Mann-Whitney U-test, the Spearman's correlation and RMSD in Tables 4-5 in essence make the case that our thesis is satisfied, i.e., there is no statistically significant difference between the two studied facilities.

5. DISCUSSION

5.1 GENERAL AGREEMENT IN RESULTS.

The hypothesis for standardized care in this study was that the application of a customized PM4CA methodology for comparison of data-driven process models from two similar facilities would not produce statistically significant differences in outcome. The results shown in the previous section supported that hypothesis by showing good agreement between the processes examined within the two facilities. The descriptive statistics and the process model maps show that most differences between the two facilities were expressed in terms of frequency.

Observe that the mean duration between facilities (i.e., Facility A 25.6 days and Facility B 26.3 days) is so close within the observation window. Our analyzes applying the Mann-Whitney U-test, the Spearman's correlation and the RMSD further confirm our findings. In terms of process, it was noted that most cases in the two facilities started and ended with the same event. The results also show that about one-third of the orders were discontinued. Those discontinuations are a 'normal' institutionalized process for dealing with cases in which the electronic clinical orders are incomplete. This particular outcome does not mean that the patient has never received the service. Rather, it was determined that those discontinuations are concerned mostly with policy and some duplicated records that arise due to imports from other subsystems.

5.2 IMPACT OF VARYING TEMPORAL CONDITIONS

A concern arising from the use of PM in this HIT environment is an assumption that all entities in the process have the same expected temporal conditions. In reality, even within a clinical domain of interest such as Imaging orders, there are conditions in which the process is based on clinical urgency such as the "stat" condition. In VHA's clinical care processes, the "stat" condition indicates an urgency. It is usually the highest priority for timeliness to do a task. Specifically, there is a need for those orders to flow through the process within minutes to hours. The "stat" condition is usually reserved for the Emergency department or Inpatient setting – it can mean that the Doctor/Nurse needs results in a matter of minutes because of the patient's deteriorating condition (or about to get worse). For example, a patient is having difficulty breathing and there is suspicion of pneumonia, so the doctor will order a "stat" chest X-ray so they can determine if there is pleural effusion and a need to start the patient on antibiotics. In another case, the doctor may order a routine chest X-ray to check for progress after, say, two weeks of antibiotics – that chest X-ray may be performed in 1-2 days from the time it was ordered. Or a chest X-ray can be ordered as part of a routine health physical. For a laboratory test, again some test results are needed in a matter of minutes (10 or less minutes) and other test results are simply for routine follow-up care.

On the other hand, the same type of Imaging order may be specified for a future date in which the temporal span of the process is acceptable within a range of time. Thus, analyzing a process must consider the acute clinical state as well as states that are related to follow-up care. For the foregoing reasons, PM studies in healthcare need to account for those expected differences.

5.3 ADDITIONAL CHALLENGES

There are several challenges to conducting a study on conformance analysis, beginning with the limitations presented by both the Business Process Models (BPM) that provide the "ideal" view, and the Electronic Health Record (EHR) data which provide the "real" view. In our opinion, both present an incomplete picture of the reality. BPM include tasks like reviewing the patient record and interviewing the patient, which are not captured in the EHR. The BPM are linear, present a single path for the process, and do not reflect the rework, interruptions, or multiple patients cared for simultaneously that we know are realistic aspects of healthcare delivery. The EHR data capture only a fraction of the processes outlined in the BPM, and do not include system data (log files) that could provide additional insights. Clearly, the EHR data source (the CDW) was not designed with the task in mind of tracking or validating BPM.

Another challenge is determining which deviations or to what degree a deviation from the process signals a potential problem. We know from subject matter experts that a process can deviate from the ideal path and still have a positive outcome. We also know that due to the nature of healthcare processes, new/different personnel can take over a process in progress and this is normal/expected. We also know that differences in care environments (inpatient versus outpatient) and the services being provided (an X-ray versus an MRI versus an ultrasound) can explain differences in time spent at different points along the process, and that these differences may be normal/expected. By comparing two facilities that are similar in size (patient beds), patient acuity, and population, the differences between the two facilities could at least be partially explained by local policies that influence how BPM are executed in real life. Unfortunately, we can never know nor measure all local policies that influence execution, or how institutional culture may influence that execution.

One way we could attempt to overcome these challenges is to analyze conformance with data stratified by care environment (inpatient or outpatient), type of service (radiology imaging type), urgency (stat or routine), and even diagnosis or underlying condition. The number of days out a procedure is scheduled could be categorized to provide additional stratification based on a clinically indicated/desired timeframe.

5.4 LIMITATIONS AND FUTURE WORK

The analysis presented herein has several limitations in addition to those mentioned in the above sections. Specifically, the study included a limited number of database schemas. Consequently, we may have excluded data that could provide more insights into the flow of the clinical orders. In addition, this study made conservative assumptions about working hours and holidays during the generation of the performance process model maps. Those assumptions may have resulted in slightly shorter mean durations between steps. Finally, we observed that when mapping activities to the OASIS Human Task State Transition diagram, the transitions between both the Created and Ready states and between the Reserved and InProgress states were minimal or non-existent in most cases. Thus, we could have represented both cases as one state, which implies the need for a customized state transition for healthcare data.

Finally, this study can be regarded as a pilot project for future work that applies PM4CA to assess a range of data-driven processes identified within healthcare facilities. It points out how the creation of BPM and data collection can be planned so that they are mutually supporting tasks crucial to the latter work scope. That future work needs to focus on tasks that include the following:

1. applying this approach to specific and well identified individual clinical processes,
2. performing comparative assessments across additional facilities,
3. developing a classification system for clinical processes to allow for processing class-wide analytical approaches that exceed individual processes, and
4. developing methods to measure process adherence that go beyond current data capturing.

Collectively, the foregoing items constitute a work scope that can provide an important analytical tool for confirming more general process conformance within healthcare facilities.

6. CONCLUSIONS

The study described herein focused on the application of an approach that included process mining, process mapping, and statistical metrics to establish similarity between data-driven process models from two healthcare facilities. That approach was successful in validating conformance in the flow of clinical radiology orders for the VHA when a reference model was absent or incomplete. Finding good agreement between the process models for both facilities was important to confirm that the clinical orders flow in a similar manner and that no apparent consequences of patient harm were identified from the study.

7. ACKNOWLEDGEMENTS

This work is sponsored by the US Department of Veterans Affairs. This manuscript has been authored by UT-Battelle LLC under contract DoE-VA IAA ORNL VICTOR Part A VA118-17-M-2015 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

We acknowledge the helpful discussions with Dr. Hong-Jun Yoon and Dr. Blair Christian about the Statistical Analysis Section.

8. REFERENCES

[1] H. B. Klasky, Ozmen, O., "Process Mining in Healthcare—A Case Study for the Corporate Data Warehouse of the Veterans Affairs Office - ORNL/TM-2019/1302," 2019.

[2] O. Ozmen, et al, "Feature Engineering and Process Mining to Enable Hazard Detection n Health Information Technology," *AMIA 2020 Informatics Summit,* 2020.

[3] O. A. Omitaomu *et al.*, "Real-Time Automated Hazard Detection Framework for Health Information Technology Systems," *Health Systems,* vol. 8, no. 3, pp. 190-202, 2019.

[4] OASIS, "Web Services Human Task (WS-Human Task) Specification Version 1.1 - Committee Specification Draft 12 / Public Review Draft 05," ed, 2012.

[5] R. Agrawal, D. Gunopulos, and F. Leymann, "Mining process models from workflow logs," in *International Conference on Extending Database Technology*, 1998: Springer, pp. 467-483.

[6] W. Van Der Aalst, *Process mining: discovery, conformance and enhancement of business processes*. Springer, 2011.

[7] W. Van der Aalst, "Data science in action," in *Process Mining*: Springer, 2016, pp. 3-23.

[8] W. Van Der Aalst, K. M. Van Hee, and K. van Hee, *Workflow management: Models, methods, and systems*. MIT press, 2004.

[9] W. Van der Aalst, T. Weijters, and L. Maruster, "Workflow mining: Discovering process models from event logs," *IEEE Transactions on Knowledge and Data Engineering,* vol. 16, no. 9, pp. 1128-1142, 2004.

[10] W. M. Van der Aalst, "The application of Petri nets to workflow management," *Journal of Circuits, Systems, and Computers,* vol. 8, no. 01, pp. 21-66, 1998.

[11] W. M. Van der Aalst, "Formalization and verification of event-driven process chains," *Information and Software Technology,* vol. 41, no. 10, pp. 639-650, 1999.

[12] W. M. Van Der Aalst, "Making work flow: On the application of petri nets to business process management," in *International Conference on Application and Theory of Petri Nets*, 2002: Springer, pp. 1-22.

[13] W. M. van der Aalst, "Decision support based on process mining," in *Handbook on Decision Support Systems 1*: Springer, 2008, pp. 637-657.

[14] W. M. van Der Aalst, "Workflow patterns," *Encyclopedia of Database Systems,* pp. 3557-3558, 2009.

[15] W. M. van der Aalst, "Object-centric process mining: Dealing with divergence and convergence in event data," in *International Conference on Software Engineering and Formal Methods*, 2019: Springer, pp. 3-25.

[16] B. F. Van Dongen, A. K. A. de Medeiros, H. Verbeek, A. Weijters, and W. M. Van Der Aalst, "The ProM framework: A new era in process mining tool support," in *International Conference on Application and Theory of Petri Nets*, 2005: Springer, pp. 444-454.

[17] M. L. van Eck, X. Lu, S. J. Leemans, and W. M. van der Aalst, "PM2: A Process Mining Project Methodology," in *International Conference on Advanced Information Systems Engineering*, 2015: Springer, pp. 297-313.

[18] R. Mans, M. Schonenberg, M. Song, W. M. van der Aalst, and P. J. Bakker, "Process mining in healthcare: A case study," in *conference; Healthinf 2008; 2008-01-28; 2008-01-31*, 2008: INSTICC Press, pp. 118-125.

[19] R. S. Mans, M. Schonenberg, M. Song, W. M. van der Aalst, and P. J. Bakker, "Application of process mining in healthcare–a case study in a dutch hospital," in *International Joint Conference on Biomedical Engineering Systems and Technologies*, 2008: Springer, pp. 425-438.

[20] R. S. Mans, W. M. Van der Aalst, and R. J. Vanwersch, *Process mining in healthcare: Evaluating and exploiting operational healthcare processes*. Springer, 2015.

[21] R. S. Mans, W. M. van der Aalst, R. J. Vanwersch, and A. J. Moleman, "Process mining in healthcare: Data challenges when answering frequently posed questions," in *Process Support and Knowledge Representation in Health Care*: Springer, 2012, pp. 140-153.

[22] A. Rozinat, Gunther, W. Christian, Niks, Rudi, "Process Mining in Practice," 2019. [Online]. Available: http://processminingbook.com.

[23] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, "Process mining in healthcare: A literature review," *Journal of Biomedical Informatics,* vol. 61, pp. 224-236, 2016.

[24] T. G. Erdogan and A. Tarhan, "Systematic mapping of process mining studies in healthcare," *IEEE Access,* vol. 6, pp. 24543-24567, 2018.

[25] M. Ghasemi and D. Amyot, "Process mining in healthcare: a systematised literature review," *International Journal of Electronic Healthcare,* vol. 9, no. 1, pp. 60-88, 2016.

[26] R. Williams, E. Rojas, N. Peek, and O. A. Johnson, "Process mining in primary care: A literature review," *Studies in health technology and informatics,* vol. 247, pp. 376-380, 2018.
[27] C. dos Santos Garcia *et al.*, "Process mining techniques and applications–A systematic mapping study," *Expert Systems with Applications,* vol. 133, pp. 260-295, 2019.
[28] G. P. Kusuma, M. Hall, C. P. Gale, and O. A. Johnson, "Process mining in cardiology: a literature review," *Int. J. Biosci. Biochem. Bioinform,* vol. 8, pp. 226-236, 2018.
[29] C. Fernandez-Llatas, "Interactive Process Mining in Healthcare," ed: Springer, 2020.
[30] G. P. Kusuma, S. Sykes, C. McInerney, and O. A. Johnson, "Process Mining of Disease Trajectories: A Feasibility Study," in *HEALTHINF*, 2020, pp. 705-712.
[31] R. Gatta *et al.*, "A framework for event log generation and knowledge representation for process mining in healthcare," in *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2018: IEEE, pp. 647-654.
[32] Y. Liu, H. Zhang, C. Li, and R. J. Jiao, "Workflow simulation for operational decision support using event graph through process mining," *Decision Support Systems,* vol. 52, no. 3, pp. 685-697, 2012.
[33] R. Dijkman, M. Dumas, and L. García-Bañuelos, "Graph matching algorithms for business process model similarity search," in *International Conference on Business Process Management*, 2009: Springer, pp. 48-63.
[34] P. Wills and F. G. Meyer, "Metrics for graph comparison: a practitioner's guide," *PloS one,* vol. 15, no. 2, p. e0228728, 2020.
[35] A. Rozinat and W. M. Van der Aalst, "Conformance checking of processes based on monitoring real behavior," *Information Systems,* vol. 33, no. 1, pp. 64-95, 2008.
[36] J. Munoz-Gama, *Conformance checking and diagnosis in process mining*. Springer, 2016.
[37] J. Carmona, B. van Dongen, A. Solti, and M. Weidlich, *Conformance checking*. Springer, 2018.
[38] J. Munoz-Gama, J. Carmona, and W. M. Van der Aalst, "Conformance checking in the large: Partitioning and topology," in *Business Process Management*: Springer, 2013, pp. 130-145.
[39] J. Munoz-Gama and J. Carmona, "A fresh look at precision in process conformance," in *International Conference on Business Process Management*, 2010: Springer, pp. 211-226.
[40] A. Burattin and J. Carmona, "A framework for online conformance checking," in *International Conference on Business Process Management*, 2017: Springer, pp. 165-177.
[41] C. Rinner, E. Helm, R. Dunkl, H. Kittler, and S. Rinderle-Ma, "Process mining and conformance checking of long running processes in the context of melanoma surveillance," *International Journal of Environmental Research and Public Health,* vol. 15, no. 12, p. 2809, 2018.
[42] C. Fernandez-Llatas *et al.*, "Analyzing medical emergency processes with process mining: the stroke case," in *International Conference on Business Process Management*, 2018: Springer, pp. 214-225.
[43] P. Badakhshan and A. Alibabaei, "Using process mining for process analysis improvement in pre-hospital emergency," in *ICT for an Inclusive World*: Springer, 2020, pp. 567-580.
[44] O. Tamburis, "Bridging the gap between process mining and des modeling in the healthcare domain," in *2019 E-Health and Bioengineering Conference (EHB)*, 2019: IEEE, pp. 1-4.
[45] G. Kukreja and S. Batra, "Analogize process mining techniques in healthcare: Sepsis case study," in *2017 4th International Conference on Signal Processing, Computing and Control (ISPCC)*, 2017: IEEE, pp. 482-487.
[46] B. F. van Dongen, J. De Smedt, C. Di Ciccio, and J. Mendling, "Conformance checking of mixed-paradigm process models," *Information Systems,* p. 101685, 2020.
[47] E. Asare, L. Wang, and X. Fang, "Conformance Checking: Workflow of Hospitals and Workflow of Open-Source EMRs," *IEEE Access,* vol. 8, pp. 139546-139566, 2020.
[48] E. Helm, A. M. Lin, D. Baumgartner, A. C. Lin, and J. Küng, "Towards the use of standardized terms in clinical case studies for process mining in healthcare," *International Journal of Environmental Research and Public Health,* vol. 17, no. 4, p. 1348, 2020.
[49] F. Marazza *et al.*, "Comparing Process Models for Patient Populations: Application in Breast Cancer Care," in *International Conference on Business Process Management*, 2019: Springer, pp. 496-507.
[50] G. B. Pereira, E. A. P. Santos, and M. M. C. Maceno, "Process mining project methodology in healthcare: a case study in a tertiary hospital," *Network Modeling Analysis in Health Informatics and Bioinformatics,* vol. 9, no. 1, pp. 1-14, 2020.
[51] N. Martin *et al.*, "Recommendations for enhancing the usability and understandability of process mining in healthcare," *Artificial Intelligence in Medicine,* vol. 109, p. 101962, 2020.

[52] R. Williams, D. M. Ashcroft, B. Brown, E. Rojas, N. Peek, and O. A. Johnson, "Process Mining in Primary Care: Avoiding Adverse Events Due to Hazardous Prescribing," in *MedInfo*, 2019, pp. 447-451.
[53] (2016). *VHA Directive 1232(1) Amended on 09/23/2016*.
[54] C. W. Günther and W. M. Van Der Aalst, "Fuzzy mining–adaptive process simplification based on multi-perspective metrics," in *International Conference on Business Process Management*, 2007: Springer, pp. 328-343.
[55] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The Annals of Mathematical Statistics,* pp. 50-60, 1947.
[56] C. Spearman, "Correlation calculated from faulty data," *British Journal of Psychology,* vol. 3, no. 3, p. 271, 1910.
[57] A. G. Barnston, "Correspondence among the correlation, RMSE, and Heidke forecast verification measures; refinement of the Heidke score," *Weather and Forecasting,* vol. 7, no. 4, pp. 699-709, 1992.