Identification of Physicians and EHR User Types

Yiqun Rong

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2018

Reading Committee:

Adam Wilcox, Chair

Gang Luo

Program Authorized to Offer Degree:

Biomedical and Health Informatics

University of Washington

Abstract

Identification of Physicians and EHR User Types

Yiqun Rong

Chair of the Supervisory Committee:
Professor Adam Wilcox
Biomedical Informatics and Medical Education

In recent years, the adoption and meaningful use of Electronic Health Records (EHRs) have undergone several promotions by governmental acts and incentive programs, with the promise of plenty of benefits. One of the most important objectives of EHR implementation was to make clinical providers to work more efficiently. However, previous studies have shown widespread dissatisfaction and low work efficiency among physicians due to rapid change of technologies and increasing requirements of EHR use. Based on a well-established model of EHR capabilities and physician EHR user, this study applied two strategies to identify different EHR user types among a large population of physicians from EpicCare Ambulatory Provider Efficiency Profile (PEP) dataset. The trends of physician work efficiency among three EHR user types identified –

basic users, strivers, and arrivers – were consistent with our hypothesis that basic users have the highest work efficiency; strivers have relatively low efficiency; and arrivers have medium work efficiency but probably gain other benefits from EHR use. This identification work of physicians and EHR user types is useful for future work focusing on facilitating healthcare providers to move forward to more efficient EHR user stages and deliver better care for patients.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

EHR	Electronic Health Record
PEP	Provider Efficiency Profile
UAL	User Action Log
HITECH	Health Information Technology for Economic and Clinical Health Act
ONC	Office of the National Coordinator
REC	Reginal Extension Centers
MU	Meaningful Use
HIT	Health Information Technology
HIE	Health Information Exchange
PHI	Personal Health Information
CQM	Clinical Quality Measures
DEA	Data Envelopment Analysis
CPOE	Computerized Physician Order Entry
TAM	Technology Acceptance Model
TTF	Task-Technology Fit
TSNE	T-distributed Stochastic Neighbor Embedding
GMM	Gaussian Mixture Model

Chapter 1: Executive Summary

This study was designed to identify physicians working with EpicCare systems and their Electronic Health Record (EHR) user types. Implications of this study should be of interest to researchers focusing on EHR usability, provider work efficiency, or clustering methods application, developers concerned about EHR function updates and EHR system design, and policymakers or managers making decisions to hasten EHR adoption and support healthcare providers. This study is deeply meaningful and highly impactful for future work of facilitating providers to move forward with EHR use, as well as contributed to the general knowledge about EHR user types and EHR benefits.

Background

Previous studies have shown widespread dissatisfaction and low work efficiency among physicians due to rapid change of technologies and increasing requirements of EHR use. Incentive governmental programs for EHR adoption, and the monoculture problem of EHR market also have produced some unintended negative consequences. To better help solo/small group physicians with practical information on EHR implementation and use, Miller and Sim identified in their analysis seven main EHR capabilities and five EHR user types[1]. Their work is of great importance to reflect actual EHR benefits and users' perspectives on EHRs. For EHR capabilities, *viewing* was a core EHR feature allowing physicians to view a variety of health records such as to-do lists, previous progress notes, medications, and allergies. Most physicians used *documenting* capability to electronically enter data during patient visits, including templates specific to some types of visits or diseases. *Ordering* was also known as electronic prescribing. *Messaging* capability was important to health information exchange(HIE) but always limited to interoffice communications. *Care management/follow-up* was designed to help clinicians longitudinally manage/follow patient conditions. *Analysis and reporting*, and *billing and scheduling* capabilities were an integral part of EHR use yet less often used. Five types of EHR users were identified *as viewers*, *basic users*, *strivers*, *arrivers*, and *system changers*. *Viewers* were physicians who almost only used the viewing capability of EHRs and obtained few benefits. *Basic users* were using capabilities directly benefiting physicians. *Strivers* were defined as physicians who began to use comprehensive EHRs and invest substantial additional time in customizing EHRs with hope of improved efficiency, but still obtain limited benefits. *Arrivers* were reaping more benefits from their investments and spending less time at work. *System changers* were physicians who were characterized by even more benefits and cost savings per patient, and advocating changes in EHRs for care quality improvement(QI).

Methodology

One month of User Action Log (UAL) Data from EpicCare Provider Efficiency Profile (PEP) Metrics was analyzed for 801 physician samples, each was given a meaningful efficiency score from PEP as well. Two strategies were formulated with different approaches to categorize physician samples into different groups in terms of their interaction with EHRs. First, automatic strategy was made to identify physician groups directly from existing data and categories by utilizing k-means clustering method and t-distributed stochastic neighbor embedding (TSNE) visualization algorithm. Then manual work was done on the dataset to further analyze physician EHR use patterns. Two approaches – Gaussian Mixture Model (GMM), and Jenks optimization method (i.e. Jenks natural breaks classification method) – were applied to re-classify physician EHR user groups, based on use types, as well as main EHR capabilities identified from Miller and Sim's theory.

Implications

This study had several meaningful implications. First, most physicians (>54%) working with EpicCare were identified as basic users, while arrivers account for the minority (< 6%). The trends of physician work efficiency among three EHR user types identified – basic users, strivers, and arrivers – were consistent with our hypothesis that basic users have the highest work efficiency; strivers have relatively low efficiency; and arrivers have medium work efficiency but probably gain other benefits from EHR use. Another interesting finding in this study is maximal use of EHR *documenting* function with multiple choices on it, and minimal use of other unusual EHR capabilities such us *care management* or *analysis and reporting*, probably more related to care quality improvement in EpicCare EHR systems.

In summary, this study showed the application of Miller-Sim model of EHR capabilities and user types to a large physician population, to quantitatively analyze different types of EHR users and their work efficiency with EpicCare systems. Although limitations remain in aspects of data source, topic scope and method design, this study can be seen as an initial and important contribution to future work of facilitating healthcare providers to move to more efficient EHR user stages and deliver better care for patients.

Chapter2: Background & Significance

Since the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009, the promotion of adoption and meaningful use of electronic health records (EHRs) has been one of the most important objectives of the Office of the National Coordinator (ONC) for Health Information Technology (HIT) to develop the Federal Health IT Strategic Plans[2]. Among various health IT products, EHRs are in the most wide-ranging use and of the greatest potential to benefit physicians in common clinical and administrative activities[3], in order to help improve quality of care for patients. Studies have shown that in some cases EHRs are positively associated with both clinical and societal outcomes. For example, computerized reminders have increased pressure ulcer prevention and treatment among hospitalized patients[4]; serious medication errors have been reduced a lot due to computerized physician order entry (CPOE) systems[5]; providers are able to charge patients more accurately and timely[6]; and more quantitative research can be conducted[7] to analyze problems and help improve healthcare quality eventually.

2.1 Supportive Programs and Problems

To better accelerate the adoption of EHRs by healthcare providers, health consumers, and the whole healthcare industry, HITECH established financial incentive programs for clinical practices and hospitals who are making “meaningful use” of EHRs[8] in 2010. Meaningful use (MU) defines minimum U.S. government standards for using EHRs. Once meet meaningful use criteria, also known as meaningful use measures, the healthcare provider is eligible to receive federal payments from the Center for Medicare and Medicaid Services (CMS). According to

Thurston's study[9], the criteria were created in purpose of promoting and supporting EHRs as well as maintaining the privacy and security. Improving quality, safety, efficiency while reducing health disparities; engaging patients and families in their conditions; improving care coordination; providing better population and public health; and guaranteeing privacy and security protection for personal health information (PHI) are five priorities of the MU program. To get an incentive fund, the healthcare provider must (1) use EHRs certified by the MU program, (2) meet the MU criteria while using EHRs in practice. Certification not only helps ensure technological capabilities of EHRs to provide patients with care in high quality, but also gives both providers and consumers more confidence that information in EHR systems is secure and confidential and can be shared with other systems or providers for better outcomes. MU emphasizes using EHRs in a meaningful way; using EHRs for health information exchange (HIE) to improve quality of care; and using EHRs to show clinical quality measures (CQM)[9]. MU is designed and conducted in three stages. From 2010 to 2012, Stage 1 required the adoption of EHRs. Stage 2, from 2012 to 2014, increased some thresholds of the meaningful use criteria and focused more on clinical decision support, care-coordination, and patient engagement[10]. By 2017, Stage 3 has detailed the meaningful use guidelines and stressed HIE across different systems and providers. The MU program makes it possible to provide patients safer and better care, and provide physicians a more efficient way to get resources, make communications, and practice medicine. This also helps to achieve the Quadruple Aim of healthcare[11] – patient experience improvement, population health enhancement, cost reduction, and better work life for healthcare providers.

However, although implementation of EHRs has promised a lot of benefits for patient care quality[12], Romano and Stafford[13] have provided important quantitative analysis about this gap between expectations and outcomes. Also, researchers have noticed that a part of physicians are inefficient while working with EHR systems, which is producing some unintended negative consequences.

With rapid change of technologies and increasing requirements of using EHRs, dissatisfaction among physicians is recently very common and growing. Problems such as impaired face-to face communication with patients[14], time-consuming documenting work, poor template-generated notes, unfulfilling performance of health information exchange between different EHR systems and indistinct work content[15] caused by disappointing EHR usability, are main factors resulting in physician dissatisfaction about their work. Sinsky et al.[12] pointed out that in ambulatory care, physicians might spend too much time on EHR products. Specifically, almost 2 additional hours is needed for EHR and desk work for every hour physicians spend on direct clinical time to their patients. Outside office time, physicians spend a mean of 1.5 hours of personal time per night using computers to do additional work, with nearly 60% of the time interacting with EHRs. Once physicians feel like they have become just data entry clerks or health information searchers, the goal of better healthcare quality for patients would never be achieved.

At Intermountain Health Care, a study[16] has demonstrated that although the productivity of physicians decreased immediately after EHR implementation, the average provider productivity increased during the whole evaluation period, and was no significantly different after half a year than the average level before EHR implementation. The work efficiency of different providers improved greatly due to EHRs. In the study, around ninety percent of the participant providers agreed with EHR benefits such as better communication within the office, reduction of transcriptions costs, improvement of patient care quality, and progress in information accessibility. A systematic review study[17] conducted by Chaudhry and colleagues showed the effectiveness of EHRs in improving practice efficiency and healthcare quality as well, but it stressed that how EHRs affect people and their workflow in most practice settings was still unclear. Some functionalities of EHRs were developed internally and available only to a limited number of users in specific contexts. The efficacy of these EHRs cannot be overvalued due to the lack of generalizability. Also, since EHRs themselves do not change anything about diseases or health status, the effectiveness and generalizability of EHR tools are of great importance with their responsibilities of providing better care for patients[17]. All of these are asking for the studies and analysis on how EHRs are used and how they make changes in different practices.

To measure the time of EHR interaction of EHR users, besides work sampling and interview/self-report methods, researchers also collected data with some time and motion approaches[18]. One of relevant studies[19] revealed the time requirements for EHR use among ophthalmologists at Oregon Health & Science University (OHSU) medical center. By manual time-motion observation and EHR audit log data analysis, the study concluded that the ophthalmologist participants had limited time to communicate with patients face-to-face in the offices, because EHR use required a substantial portion of patients' visit time. Another systematic review study[18], using the weighted average method to combine results from literature, showed that CPOE systems were inefficient and increasing physicians' work time. Some features like bedside terminals and central station desktops of EHR tools affected providers' workflow greatly. As addressed in the study, although these features slightly saved nurses time of entering data, they increased the documentation time of physicians a lot. The likelihood of EHRs to reduce documentation time was even rejected by the review.

By 2017, the majority of hospitals in the United States have reached Stage 5 or above, which means their EHR systems have equipped with capabilities of documentation with structured templates, task completion track/report, system intrusion prevention, data analysis, risk reporting, and external HIE, referring to the HIMSS Analytics EMR adoption model[20]. Doug Thompson pointed out that most EHRs were implemented in a "Big Bang" way, because hospitals and medical groups desired to move forward and meet the requirements of MU incentive program in the past few years[21]. Nevertheless, physicians and other clinical staff failed to keep pace with EHR implementation and change their workflow, which leaded to the opposite consequences of MU program intentions, including higher medical costs and heavier clinical workload. Other analysis on the reasons why hospitals have not achieved the expected outcomes included that EHRs were designed for some target audience, or they lack a holistic consideration of what the systems actually need[22]. Many health organization that have EHRs in place are facing the problems of user dissatisfaction and low work efficiency, and EHR optimization has been held up as a solution.

EHR optimization aims to refine EHRs to serve a user's own needs and emphasize clinical productivity and efficiency[23]. Since one widespread complaint of physician EHR users is too many clicks, EHR developers are expected to focus more on easier data entry and information retrieval while developing the systems. Moreover, according to the definition, EHR optimization should be different for each individual, and different clinic offices to maximize the efficiency benefits. This requires the customization function of EHR systems to let users optimize EHRs in accordance with their own work habits. There are some concerns about the conflicts between standardization and customization requirements of EHRs[24], as standardization has always been perceived as an effective way to improve efficiency and care quality. Pandhi et al.[24] suggested that one potential solution might be making EHR optimization plans with all relevant stakeholders to take as many as possible their EHR use, needs and expectation into consideration, to give them appropriate customization support.

2.2 Information Systems Theory

There has been a tremendous amount of studies dedicated to the acceptance of technology. The Technology Acceptance Model (TAM) is one of the most influential frameworks for predicting the perceptions of information system users[25][26]. The TAM focuses exclusively on factors that contributed to users' behavioral intentions toward accepting and using a new information technology, particularly, perceived usefulness and perceived ease of use[26]. A study discussed that perceived usefulness was highly correlated with physicians' attitude about EHRs and was an extremely strong predictor of their EHR use[25], as hypothesized by the TAM. While the TAM has been applied to make some predictions on EHR use successfully, its explanatory power has varied a lot in different settings. Other studies have suggested modifications to the TAM or combinations with other theories or frameworks to explain the usefulness and intentions more considerately, including the processes of social influence and cognitive instrument effect, also known as extended TAM in some cases[27][28][29].

However, use of technology including EHRs is not always voluntary, and increased use or acceptance does not always lead to better performance or increased efficiency. Therefore, in contrast with models explaining perceptions and technology acceptance like TAM, Task-Technology Fit (TTF) model was put forward in explaining user performance with the information systems[30]. On the premise of TTF, individual performance can be positively impacted when the functionalities of the technology meet the user's requirements and expectation[31]. This makes TTF in a good place to help explain the adoption and use situation of EHR systems. How EHRs can be used and potentially enhance the performance of healthcare providers is in an increasing demand[32]. A relevant study has confirmed eight positive dimensions of TTF and supported TTF as a good model for predicting performance influence in clinical settings[33]. Moreover, it has validated the explanatory power of TTF for ease of EHR use being important determinants of healthcare providers' performance.

2.3 Related work of EHR capabilities and EHR User types

Miller, Sim and Newman's studies have shown some important use and capabilities of EHRs as well as several lessons learned from EHR implementation including five types of physician EHR users they identified[3][33]. In the context that most active and practicing physicians were working in solo/small groups while little of their experience using EHRs had been published, Miller and colleagues were trying to understand EHR use pattern among solo/small group physicians, in order to help policymakers, coalitions and various funding agencies to make better decisions, and ultimately improve healthcare quality in solo/small clinical groups.

Qualitative data of physicians from 20 solo/small practices was collected from 2000 to 2002 in California. The investigators conducted semi-structured telephone interviews between one and four for each practice about physicians' EHR use. The researchers conducted additional face-to-face interviews when they felt necessary after observing physicians' interaction with EHRs. Selected physicians were from primary care, cardiology, urology, endocrinology, pulmonology, and family practices. The EHR products included Practice Partner, QD Clinical, Alteer, NextGen, A4 Healthmatics, AutoChart, Soapware and Cliniflow. Physicians had experience with EHRs between 6 months to ten years.

For data analysis, Miller et al. used QSR Nvivo (a qualitative research software) to do the transcribed interview text coding. They categorized the data into several concepts like "implementation activities" and "workflow changes" and analyzed interviewees' response. Qualitative research techniques like explanation building and pattern matching were also used in analysis.

2.3.1 Main EHR capabilities

First Miller and Sim identified that EHRs have seven main types of clinical capabilities that corresponding to seven sets of clinical activities in ambulatory practices. Specifically, although each EHR product had some different features than others, the eight EHR products involved in their research generally had the most-often-used capabilities of viewing, documenting, ordering, messaging, care management, analysis and reporting, and patient-directed. Below is the detailed description of each of these capabilities:

Viewing. The viewing capability represented a core feature of EHRs. Physician users from different practices may view a variety of health records but they generally included to-do lists, previous progress notes, medications, and allergies. Some physicians could also view lad results, consultant reports and other related clinical data if they were available in the EHR systems.

Documenting. The documenting capability allowed physician users to electronically enter data about progress notes, chief complaints, allergies, diagnosis, prescriptions and so on. Different templates were utilized across different physician individuals, practices, the types of visit, or the patient's health conditions. Electronic forms were expected to assist physicians to guide the exam process and communication with their patients during the visit, when most physicians typed to add information into the system. Physicians had the option to type in free text as well. Some combination of structured and unstructured entry saved EHR uses' time costs.

Ordering. The ordering capability was known as electronic prescribing as well. It enabled physicians to enter prescriptions (medication or non-medication orders) into electronic forms, which could be created and stored for later use, or selected from various existing ordering possibilities. Moreover, physicians could receive feedback(alerts) on their decisions about any potential drug-drug or drug-allergy interactions. Although initial prescription entry took a lot of time and effort, physicians reported that refilling orders did help save time, reduce error and improve efficiency to some extent.

Messaging. The messaging capability was designed to offer convenient inter-office or inter-provider communications. It could save time and effort when any information or request needed to be delivered or shared among different clinicians, but few of this kind of functionalities were integrated into existing EHR systems.

Care Management. The care-management capability included features related to disease prevention and management. It had some overlap with the documenting capability but can be considered more as follow-up. Physicians could customize their own templates to record important information that could remind them of health maintenance status, and health status improvement potentials for their patients. Users who had experience with the care-management capability in the research reported that they felt they had provided better care.

Analysis and Reporting. The analysis and reporting capabilities were for the purpose of specific patient population data query and review, in order to compare different care patterns for better clinical decision-making. However, these functions were very limited used in part due to their user-unfriendly and complicated manipulation steps.

Patient-directed. The patient-directed capability involved patients into EHR use. Patients could access the websites of their practices to review their diagnosis or have secure communications with their healthcare providers. Even though the capability was available, almost no patients directly interact with EHR systems. Physicians rarely used this function to response to their patients either.

This study also showed that some practices got EHRs with integrated billing and scheduling capabilities; others had interfaces for health information exchange between EHR products or EHR and management system software. Great integration might bring a lot of benefits such as reducing duplicate data entry. However, the research focused mainly on physician users, and many practices had none of these billing/scheduling or data-exchange capabilities at all, so these less often used EHR features somewhat for other clinical support staff were not included in the seven main EHR capabilities.

Miller, Sim and Newman then concluded three lessons they learned from their research on EHR implementation. The first lesson they learned was that initial EHR implementation costs were substantially high – ranged from $15000 to $50000 per physician, while financial benefits and patient care quality benefits varied greatly from physician reports. Though different types of EHR users might reap different benefits, it was found that decreased costs for transcriptionist, medical records, data entry and billing staff were common, but increased revenue for physicians were less common. Besides these benefits, their second lesson further demonstrated that technology differences among various EHR products could explain only some of the benefit variation. In other words, benefits and costs related to EHR using were determined by much more factors than just EHR software itself. How the EHRs were used is of great importance. Physician users might experience a wide range of benefits, satisfaction and time costs even if they use the same EHRs.

2.3.2 EHR User Types

The last lesson Miller and his team learned was not only valuable and influential, but also very critical to the frame of this thesis. On the basis of Cain and Mary's work of categorizing EHR users into innovators, early adopters, early majority, late majority and laggards[34], five different types of physician users within the spectrum of *innovators* and *early adopters* were created considering user characteristics, including their EHR use, benefits reaped and time spent on complementing the EHR systems. Figure 1 helps illustrate the relationships among these concepts.

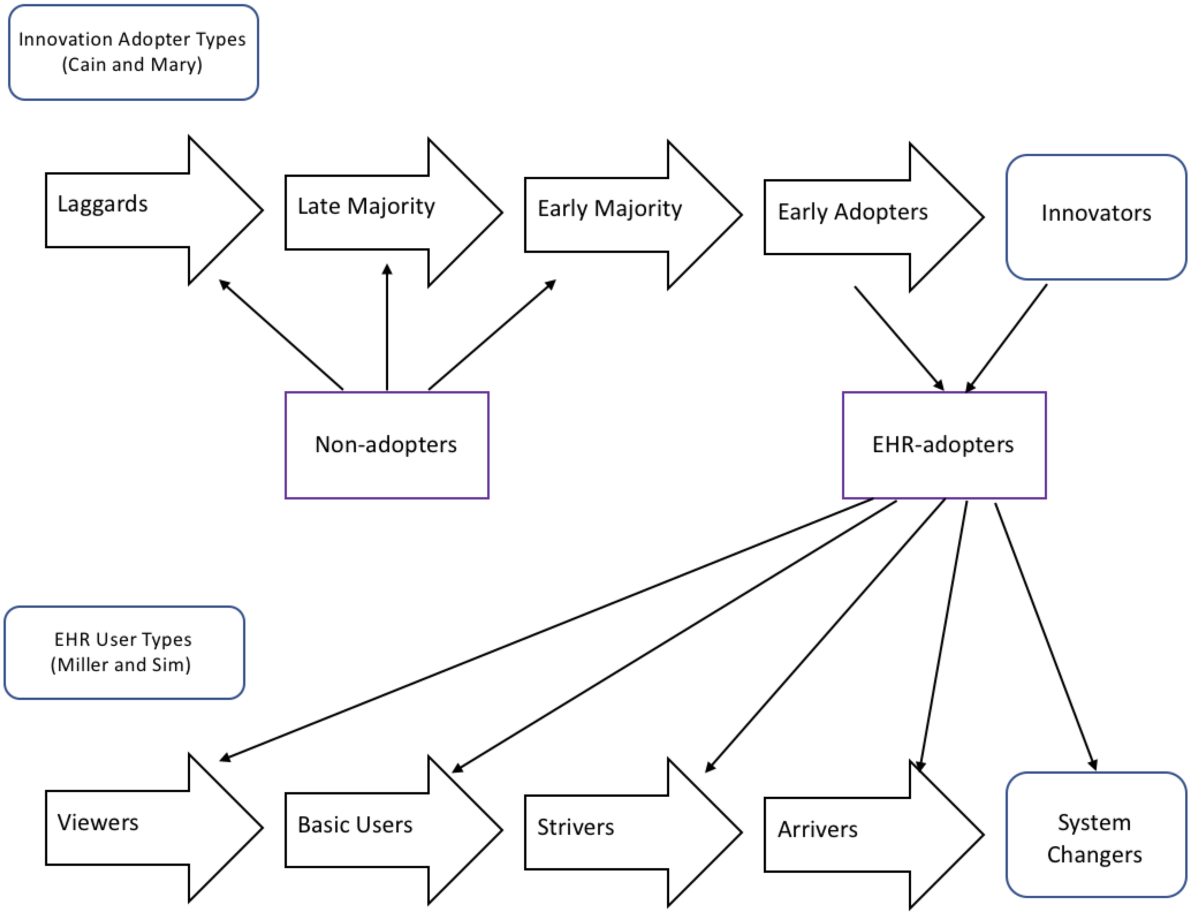


Figure 2. 1 Types of Innovation Adopters and EHR Users

Source: Electronic Medical Records: Lessons from Small Physician Practices

For better understanding each type of EHR users, it is essential to keep in mind the concept of complementary changes. Complementary changes played a crutial role in generating benefits for physician users' EHR experience. Activities like creating templates for convenient documenting, entering data from previous paper documents, customizing text shortcuts or phrases, giving solutions to fix technical problems, changing their wokflow correspongding to EHR use, and rearranging process in the whole office, are all included in complementary changes. These contributions would eventually save at least extra time costs because if implementing EHRs. Below the characteristers of each EHR user type and the relationship between each type and complementary changes to EHRs were described.

Viewers. Viewers spent the least extra time in making complementary changes. They interacted little with various EHR capabilities and obtain limited benefits. Other than viewing information, they used few else funtions within EHRs. Compared to electronically documenting, they did more diatation or hand-writing on progress notes and presctiptions. Viewers workflow was unchanged, and no saving on their time spent at work either. To make matters worse, viewers tended to attribute their mininal use of EHRs to the poor design of EHR systems negatively, which might avoid their movements to next stage – basic users.

Basic Users. Basic users still spent limited time in making changes to complement EHRs, but they used more EHR capabilities such as entering free text and ordering prescriptions besides viewing data. Some of them viewed disease-specific templates while they dictated visit notes, and then asked transcriptionist to fill out the templates. Basic users started to gain a little bit more benefits than viewers, but there was a concern among physicians: those basic users who had obtained more benefits from just investing mininal effort to make complementary changes, would kept their work at a low level of EHR use. This concern would influence their experience with EHRs and weaken their willingness to further embrace EHRs as a necessary part of their work, acting like strivers or arrivers.

Strivers. Strivers spent substantial extra time in making complementary changes to EHRs. Most of them were trained to use the EHR products more efficiently, for expamle, entering previous data about patients and conditions, customizing templates, fixing technicall problems of EHR softwares, and rearranging their workflow. Although some savings on time and finicial costs from reducing transcriptionists and data entry staff was a result, strivers themselves were still obtaining limited benefits. They spent additonal time at work with the hope of provide higher quality of care and improve work efficiency. Time was needed to enter progress notes in detail, communicate results with patients thoroughly, and sending messages to other clinicians from their perspectives. Strivers might also be physicians who were experienced with EHR use but recently switched to a new EHR product/system. More time was demanded to learn to use the product, create templates and change their workflow.

Arrivers. Arrivers had been strivers for some time and had made a lot of additional efforts in making complementary changes to EHRs. They had not only interacted with more EHR capablities, like developing messaging interfaces for better communication, and figuring out technical solutions for later use, but also documented a larger amount of patient history data, and created as well as utilized more useful templates. The most notable change was they reaped much more benefits from EHR use, in both time and financial aspects. Arrivers reported that they had spent much time in EHR use preparation while they were still considered as strivers, which made them be able to quickly generate key phrases, import lab data from just clicking mose buttons. Time saving during visits might lead to more chat with patients and higher care quality as a result. One more characteristic belonging to arrivers was workflow reorganization across the whole exam room and the office. Since EHRs brought many changes to physicians' work routine, better arrangement of tasks for each clinical staff member would help improve physicians' work efficiency as much as possible.

System Changers. System changers were those who obtained the most benefits from EHR use. They saved even more time for each patient visit than arrivers, and interacted with more capabilities more efficiently. Moreover, they had such activities related to the change of both internal and external environment as incentivizing EHR users staying in previous user type phases, negotiating with payers and independent practice associations to reward physicians, and encouraging policies to give practices bonuses for care quality improvement on account of EHR use. Most of the system changers were leaders in the office or locality, and they paid additional attention to quality improvement. System changers would be ideal examples of EHR champions in their practices or even larger groups.

After these lessons, researchers called on purchasers, public policymakers, and funding angencies to facilitate more EHR adoption and more meaningful EHR use by means of encouraging EHR exchange in communities, incenting physician users with more benefits, and increasing both qualitative and quantitative study on EHR users. As for suggentions for solo/small groups, identifying an EHR champion was recommended although it might not achieve any success. Most physicians were not willing to spend extra time to figure out how to document in the most efficient way, fix IT problems, change their workflow or their office's work pattern, and hasten health information exchange between themselves and outside providers, let along lead and help others. Also needed were comprehensive and multifacedted support services, and rewards from practices for physicians who generated more benefits from EHRs.

2.3 Challenges and Opportunities

Despite the contributions previous studies and programs have made to promoting EHRs, there still remain several challenges. On one hand, most previous influential conclusions are based on EHR user interviews and qualitative analysis. Although qualitative research can be examined in detail and in depth, which sometimes might be more comprehensive and compelling than quantitative research, the process is generally time-consuming and labor-intensive. The methods are hardly applicable to measures for a large population, and the findings can be difficult to verify and visualize. Researchers also mentioned that qualitative research quality may rely greatly on the individual skills and personal biases of the researchers themselves, and thus rigor is always hard to maintain, assess, and prove[35].

On the other hand, like what has been phrased earlier in this chapter, various benefits of meaningfully using EHRs are strong incentives for hospitals, as well as healthcare providers to adopt EHRs. As demonstrated that physicians are likely to continue and even increase their EHR use over time once they start to enter data into EHRs voluntarily[16], *basic users* of EHRs are pushed to be *strivers* easily, yet physicians who stay in the *strivers* stage are facing lack of support to help them move to *arrivers*. The Office of the National Coordinator (ONC) for Health Information Technology has funded many Reginal Extension Centers (REC) over the United States, attempting to slove the challenges of MU programs[36]. The REC program aims to support over 100,000 healthcare providers in small, rural, and underserved medical practices to accelerate EHR adoption and meaningdul use[37]. Furthermore, the ONC created the Nationwide Health Infoemation Network, a national infrastucture for HIE, to facilitate communication across different parties sharing information[38]. However, few studies or reports on the success of these projects were found. Supportive programs cannot be without their challenges; what they have brought to the healthcare industry, both positively and negatively, are of great significance but still unclear.

Before widespread use of EHRs, several studies applied data envelopment analysis (DEA), a linear-progeam based efficiency evaluation approach to measure physician practice patterns, and both technical and scale efficiency[39][40]. After EHR implementation, researchers have called much for improving physicians' EHR efficiency and satisfaction[41], but no general measures on physicians' work efficiency related to EHRs have been published to our knowledge. Most measures focused on the time users interact with EHR systems, yet time cannot reflect efficiency thoroughly, not to mention other benefits promised by EHR implementation.

A more recent study conducted by Koppel and Lehmann addressed a new challenge in the healthcare market[42]. Based upon the fact that EpicCare clinical systems have captured health information for more than 50% of the US population, Epic is taking the places of other EHR vendors in the market, building a "single-vendor landscape", which has raised concerns about Epic monoculture. Koppel and Lehmann elaborated that although one vendor's market has advantages like training needs reduction, and system development and maintenance costs decrease due to standardized data formats and user interfaces, the diversity of market suppliers will be compermised by this emerging phenomenon. Opportunities for transparency, user participation and patient interventions will also be negatively affected. What's worse, the monoculture is changing the motivation of system development or modification. The incentives for EHR vendors to build new features or update the systems are now primarily led by market forces rather than users' needs. Vendors' decision that selling EHR systems is much more important than helping users work more efficiently and deliver better care would be a big problem in the coming years.

Epic recently launched a program called Provider Efficiency Profile (PEP)[43] to focus on providers' workload and efficiency. Its purpose is to help managers identify providers who can utilize EHRs efficiently, as well as areas of systems where make changes and enhance providers' productivity in ambulatory settings. PEP can be a good support for stakeholders to prioritize continuous system improvement chances and make specific training plans for inefficient providers, in order to achieve long-term benefits from EHR implementation. Nevertheless, the direct users of PEP are managers and administrators, so the concrete metrics regarding how a provider's efficiency is defined, in other words, how a provider's *efficiency score* is calculated is unclear. Additionally, Epic does not have any framework for tracking how providers move from less efficient to more efficient, or in the other way. Though more work needs to be done, we believe the efficiency score from PEP is rational to some extent, and can be combined with other theories to further explain the providers' EHR use.

Therefore, recalling the EHR user stages identified by Miller and Sim, the identification of physicians and their EHR user types from quantitative analysis is in great need. This can contribute to help people get through the *strivers* stage, where EHR users struggle and complain the most, and facilitate providers to move to more efficient/advanced stage with better EHR use.

Chapter 3: Methods

This chapter describes the entire process of my research. I conducted two studies with different approaches to categorize physician samples into different groups in terms of their interaction with EHRs. First, I tried to automatically identify physician groups directly from existing data and categories by utilizing k-means clustering method and t-distributed stochastic neighbor embedding (TSNE) visualization algorithm. Since it didn't work well enough as expected, I decided to do some manual work on the dataset to analyze physician EHR use patterns. Then I applied two approaches: Gaussian Mixture Model (GMM), and Jenks optimization method (i.e. Jenks natural breaks classification method) to re-classify physician EHR user groups, based on use types, as well as main EHR capabilities identified from Miller, Sim and Newman's theory.

It is worth mentioning that this study was always trying to categorize physician samples into only three EHR user types: basic users, strivers, and arrivers, other than 5 types which also include viewers and system changers. In our rationale, viewers and basic users are both considered inefficient with EHRs. They may keep unchanged clinical workflow and still rely much on paper records or scanned images. On the contrary, both arrivers and system changers are advanced EHR users who benefit more than they invest. They make contributions to the EHR systems as well. Therefore, compared to the differences between viewers and basic users, or between arrivers and system changers, we focus more on how to distinguish strivers from the two ends of the EHR user spectrum. Strivers are staying in the stage that spending the most extra time at work in making changes but reaping little benefits, so they are those who always complain and need immediate help. Identifying basic users, strivers, and arrivers in this research can contribute to future studies on how to intervene in moving a striver to an arriver quickly.

I explain each step of my study in detail, to show the difficulties and significance of this type of work. This part also helps future researchers to have a better sense of how these methods might be applied with similar studies involving different EHR products.

3.1 Data Source

This analysis was performed with data from EpicCare Ambulatory Provider Efficiency Profile (PEP) of December 2017. Epic representatives collect more than 20 days of User Action Log data and produce different charts for each physician or non-physician user to show their work patterns. Stakeholders can use PEP data and dashboard to (1) reach out to individual EHR users who need additional help with the system use; (2) make policies or design training to support some specific providers; (3) modify the systems based on users' problems to improve their experience. Samples in this dataset are providers working with Epic EHR systems of different types (nurse, physician, medical assistant, coordinator, social worker et al), in different department specialties (urology, family medicine, cardiology, orthopedics, behavioral health, etc.), and from different service areas (UW Medical Center, Harborview Medical Center, Northwest Hospital and Medical Center and Seattle Cancer Care Alliance) in Washington State.

3.2 Data Cleaning

According to the aims of this research, I first did some data cleaning work. For samples, I kept EHR users who belong to the physician provider type and have meaningful efficiency scores (from 0 to 10), which built a dataset of 801 physician samples. The original data I got includes action log data of providers' EHR use for two months (November and December of 2017), and I used data recorded in 2017 December to implement the quantitative analysis. At last, I filtered out all sample descriptive variables such as user-mapped name, provide specialty, and primary provider department, as well as some other variables meaningless in this research like proficiency score and days out of contact; and finally kept a number of 95 numerical variables (including some binary variables with the value of 0 or 1) relevant to physicians' EHR usage for the following analysis work.

3.3 Study 1: Automatic Identification

With the purpose of identifying EHR user groups using existing data and categories from PEP, I chose K-means method to cluster 801 samples into 3 groups with all 95 variables, and TSNE method to deliver more understandable visualization when needed. Initial expectation of applying these approaches is (1) to divide samples into 3 separate groups; (2) to get 3 meaningful groups which can be interpreted by domain knowledge; (3) to further interpret 3 identified groups with work efficiency conditions (*efficiency score*) from PEP.

3.3.1 K-means

K-means is a widely used unsupervised machine learning method to cluster high dimensional dataset with given number of classes. The user first chooses the number of N centers, then the algorithm starts with arbitrary N center positions drawn from uniform distribution of data. The goal of the algorithm is to minimize the total distance of each point to its class center as shown. After the initialization of original centers, the next data point will be assigned to the class with the nearest center point to it, and the center of that class will be recalculated based on the average of the mass of all data points belonging to that class. This step will be continuously repeated until all data points are assigned. There are existing well-established machine learning libraries that provide convenient implementation of K-means clustering method. This algorithm is guaranteed to converge; however, the local minima problem may appear depending on random-picked initial centers. In order to prevent the algorithm from hitting local minima, multiple runs are performed with different initial centers, and all centers are assigned apart[44].

3.3.2 TSNE

T-distributed stochastic neighbor embedding (TSNE) is a commonly used machine learning algorithm to visualize high dimensional dataset. As a nonlinear dimensionality reduction technique, TSNE performs well in reserving the local structure of data. By converting high-dimensional dataset into pair-wise similarities, TSNE maps the dataset to two or three dimensions for visualization purposes[45].

In this analysis, to identify physicians into basic users, strivers, and arrivers, the number of N centers in k-means method was set to be three. All nighty-five meaningful numerical variables were taken into consideration while samples being classified by the K-means method. After grouping physician samples with 3 labels, TSNE was used to visualize the entire dataset by projecting all sample data points form high dimensional space to a 3D system and a 2D plane constituted by three and two principle components generated by the TSNE method respectively. Efficiency scores were reflected by color differences for visualization, both within each identified group and across the whole sample set.

The KMeans class and the TSNE class from the scikit-learn library[46] were used for implementation. A random seed was also fixed for reproducibility.

3.4 Study 2: Manual Identification

For the second study design, I manually selected variables reflecting physicians' EHR use, and calculated a use score for each physician sample regarding their average EHR use for each of their patients. With knowledge support from Miller and Sim in their analysis, expectation was that (1) physician samples can be grouped into three EHR user types by *use scores*; (2) identified groups can be interpreted with the combination of *efficiency score*s from PEP. In concrete implementation, two classification methods – Gaussian Mixture Model (GMM), and Jenks optimization method were applied under different assumptions of sample distribution.

3.4.1 EHR use measures

After initial analysis on the PEP dashboard raw data, I found several EHR capabilities identified by Miller and Sim, such as messaging, care-management, and billing, were still immature or not included at all in Epic EHR systems. For this reason, I created a set of EHR use-score metrics as close as possible to Miller and Sim's work to measure physicians' average EHR use for each patient during the report period.

Totally 30 out of 95 quantitative variables were taken into consideration while building physician use-score metrics. Other than the variable average appointments per user, the left 29 variables were categorized into four main EHR capabilities identified by Miller and Sim – viewing, documenting, ordering, and others, with 15%, 50%, 25% and 10% as their weights respectively.

Detailed weight allocations and descriptions for relevant PEP variables in each main EHR capability are following:

X: Average appointments per user
For provider records, the number of appointments during the report period. This variable is used as the denominator when needed to reflect the physician's average EHR use per appointment/patient.

15 % Viewing:

A. *% patient calls reviewed quickly*
   Comes from Pulse.
B. *% results message reviewed quickly*
   Comes from Pulse.
C. *Number of chart review quick filters created by user (5%)*
   Number of Quick Filters that list the provider as the owner.
D. *Are notes defaulted to sidebar? (20%)*
   Whether the provider has the default location for notes set to the sidebar.
E. *Visit diagnosis personalized? (20%)*
   Whether the user has speed buttons for visit diagnoses.
F. *Clinical review minutes: divided by X (50%)*
   Amount of time the provider spends in areas of the system designed for reviewing patient information, such as Chart Review, Patient Summary, SnapShot, Results Review, and the report viewer.
G. *Chart search use? (5%)*
   Whether the provider used Chart Search during the report period.

Measure of Viewing: ((F / X) * (A + B) / 2) * 50% + C * 5% + D * 20% + E * 20% + G * 5%

50 % Documenting:

A. *Number of SmartPhrases owned by user (12%)*
   The number of SmartPhases the provider owns.

B. *Number of SmartPhrases shared with user (8%)*
   The number of SmartPhases in which the provider is listed as a user.

C. *Number of note speed buttons (20%)*
   Number of note speed buttons the provider has in the Notes activity. Note speed buttons allow users to pull in note templates quickly.

D. *Notes/Letters minutes: divided by X (25%)*
   Amount of time the provider spends in notes and letters (either the Communication Management or Letters activity).

E. *Notes Entry: divided by X (35%)*
   Total number of characters in all notes entered by following ways:

   a. *All notes manual entry (20%)*
      Number of characters in all notes contributed by the note author that were typed.
   b. *All notes SmartTool entry (70%)*
      Number of characters in all notes contributed by the note author attributed to SmartLinks, SmartText, SmartPhases, or SmartLists.
   c. *All notes NoteWriter entry*
      Number of characters in all notes contributed by the note author attributed to NoteWrite (SmartBlocks, SmartBlocks from a macro, or SmartBlock templates).
   d. *All notes Copy/Paste entry (5%)*
      Number of characters in all notes contributed by the note author attributed to copy/paste, including PasteBoard and Copy Previous/Forward.
   e. *All notes voice recognition entry (5%)*
      *Number of characters in all notes contributed by the note author attributed to voice recognition such as Dragon or M*Modal.*
   f. *All notes transcription entry*
      Number of characters in all notes contributed by the note author attributed to transcription, including partial dictation.
   g. *All notes other entry*
      Number of characters in all notes contributed by the note author attributed to other methods, such as drag and drop, external document, system-generated etc.

Measure of E: a * 20% + (b + c) * 70% + d * 5% + (e + f + g) * 5%

Measure of Documenting: A * 12% + B * 8% + (C * 10) * 2% + (D / X) * 25% + (E / 100) * 35%

25% Ordering:

A. *Average number of medications per appointment*
   Number of medication orders placed during each outpatient visit during the report period.
B. *Average non-medication orders per appointment*
   Number of non-medication orders placed during each outpatient visit during the report period.

C. *% orders placed from preference list or smart sets*
   Comes from Pulse.

D. *Number of times QuickActions used: divided by X*
   Number of times a provider used a QuickAction during the report period.

E. *Order minutes: divided by X*
   The amount of time the provider spends writing and reviewing orders.

Measure of Ordering: (A + B) * C * 40% + D / X * 20% + (E / X) * 40%

10% Other:

- Patient-directed
    A. *% patient advice requests done promptly (40%)*
       Comes from Pulse.
- Scheduling
    B. *Schedule/Patient lists minutes: divided by X (25%)*
       Amount of time the provider spends in the Schedule or Patient Lists activity.
- Access/Search
    C. *LOS Personalized? (10%)*
       Whether the user has speed buttons for level of service.
- Other
    D. *Visit navigator minutes: divided by X (5%)*
       Amount of time the provider spends in Visit Navigator sections that are not already included in other metrics.
    E. *Other minutes: divided by X (20%)*
       Amount of time the provider spends in areas not listed in other metrics.

Measure of Other: A * 40% + B / X * 25% + C * 10% + C / X * 5% + E / X * 20%

Table 3. 1 Overview of PEP Dashboard Data and Sample Use-score Metrics

Before a final EHR use-score is calculated for each physician sample based on the measures stated above, all the measures are normalized with equation (3.1), where μ is the mean and σ is the standard deviation of the measured EHR capability. The normalization is performed to make sure that the scales of variation within each measure are similar, so that measures for each main EHR capability are able to account for the weighting percentages assigned to them to define the final use score.

$$x_{Normalized} = \frac{x - \mu}{\sigma} \tag{3.1}$$

Using *use score* as a feature, two classification approaches – GMM and Jenks natural breaks optimization – were implemented to group samples into three EHR user types. In the meantime, *efficiency scores* were compared with *use scores* across different groups to seek for potential implications.

3.4.2 GMM

Gaussian Mixture Model (GMM) is a parametric probability model that represents a weighted summation of multiple components of Gaussian distributions with unknown parameters[47]. GMM is commonly used for classifying data of different classes from mixture of normal distributions. Under the assumption that the use scores within three physician user types are following three normal distributions respectively, GMM was expected to produce an optimal classification of three classes depending on the *use-score* data.

In this analysis, GaussianMixture class from scikit-learn library[46] was used to implement the algorithm for GMM. The number of components was set to be three, and the algorithm converged shortly. Based on the estimated weights and parameters, all samples were finally assigned a label indicated the one of three distributions which it was most likely drawn from.

3.4.3 Jenks Natural Breaks Optimization

Jenks natural breaks optimization is another classification method for data clustering, which is used to classify data by maximizing average deviation of each class from other classes and minimizing deviation within classes[48]. Inappropriate hypotheses are always inapplicable to the actual data and influential to the rigor of research results greatly, yet as a nonparametric method, Jenks optimization does not require any assumption of data distribution. This method is an iterative searching process -- for each range combination, calculate sum of squared deviations for class means and find the smallest one.

For this part of work, Jenks optimization was realized utilizing an implementation from GitHub repository (https://github.com/perrygeo/jenks). Similarly, the number of classes was set to be three, and the computation converged shortly after started.

After the identification of physician user types, the efficiency score feature was combined, intending to further interpret similarities and differences across different EHR user types. Results from different identification studies were compared as well.

Chapter 4: Results

This chapter describes computational results of implementation described in the previous chapter including K-means clustering, calculation of use score, classification using GMM and Jenks optimization. The results contain visualization of classification results as well as descriptive statistics that will help with result interpretations.

4.1 Results of Automatic Identification

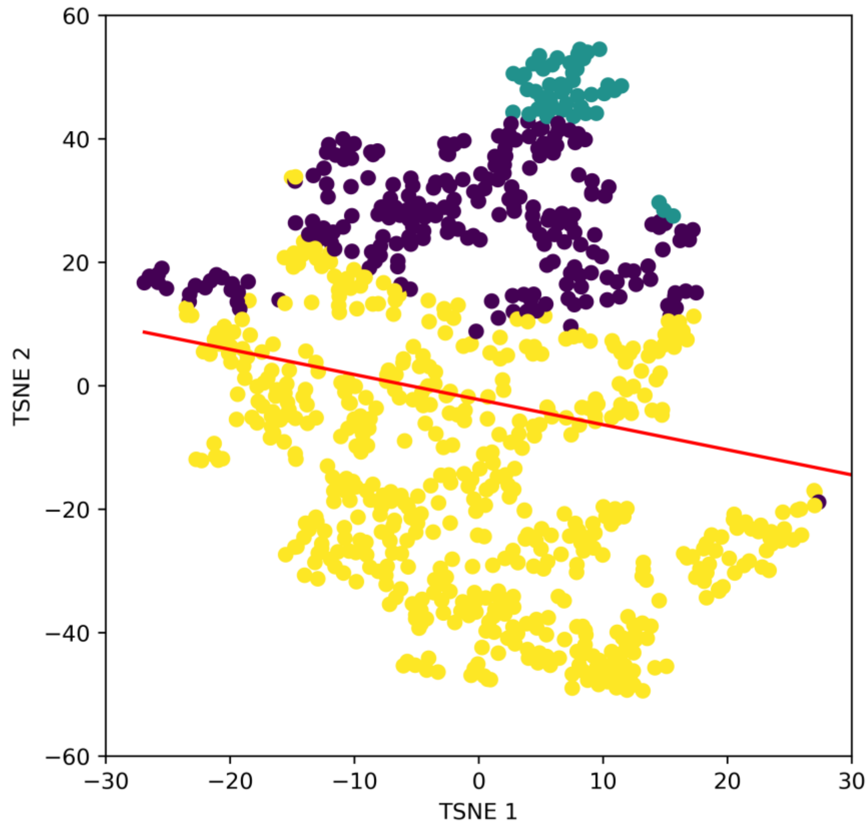


Figure 4. 1 Physician Identification Based on Automatic Work by K-means

Fig. 4.1 shows the results of K-means clustering. The data points are projected to a 2-dimensional plane constructed by main components generated by t-SNE. Purple, green, and yellow group of points indicate samples with class label 0, 1, and, 2. The regression line does not go across the three groups identified by K-means.

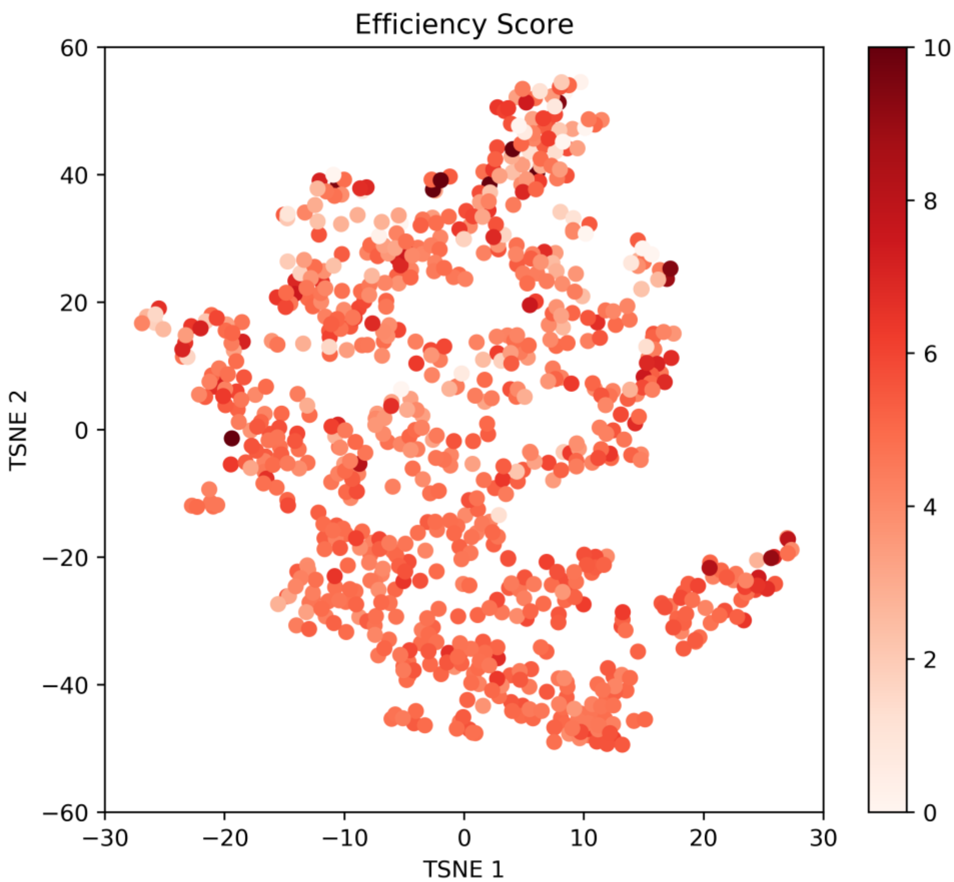


Figure 4. 2 Distribution of Efficiency Score through Automatic Physician Identification

Fig. 4.2 shows a visualization of the sample distribution with efficiency score. The data points are projected to a 2-dimensional plane constructed by main components generated by t-SNE.

The color depth, as shown in the side bar, indicates efficiency score value for each sample physician which varies from 0 to 10.

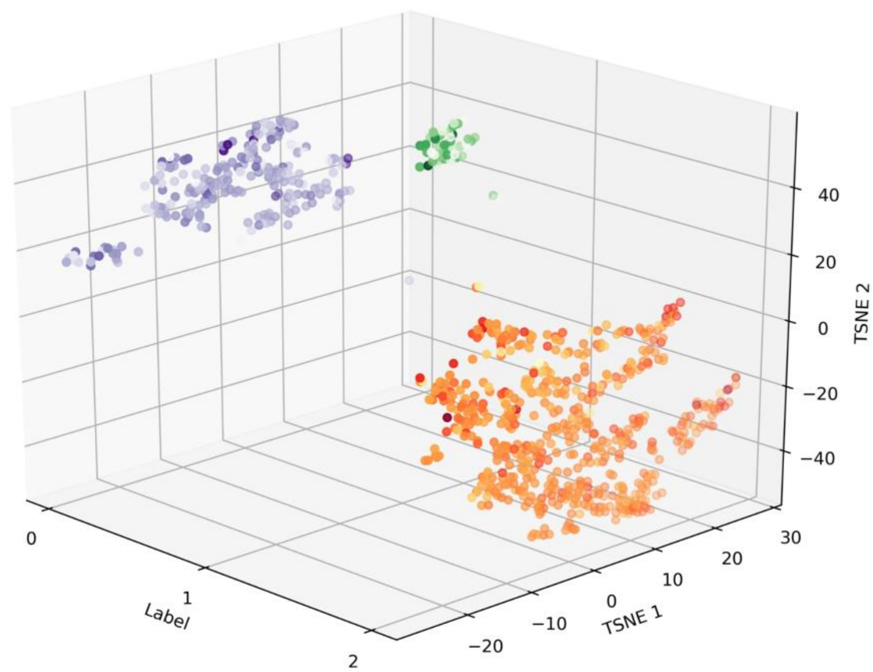


Figure 4. 3 Distributions of Efficiency Scores within Three Auto-Identified User Types

Fig. 4.3 is a visualization of efficiency score of different classes clustered by the K-means method. Same as shown in fig.1, purple, green, and orange indicate samples with class label 0, 1, and 2. Data points from three classes are project to 2-dimensional plane constructed by t-SNE main components and then separated to three planes according to their class label for visualization purpose. The color depths indicate values of efficiency score within each identified group, and darker color points show relative higher efficiency scores.

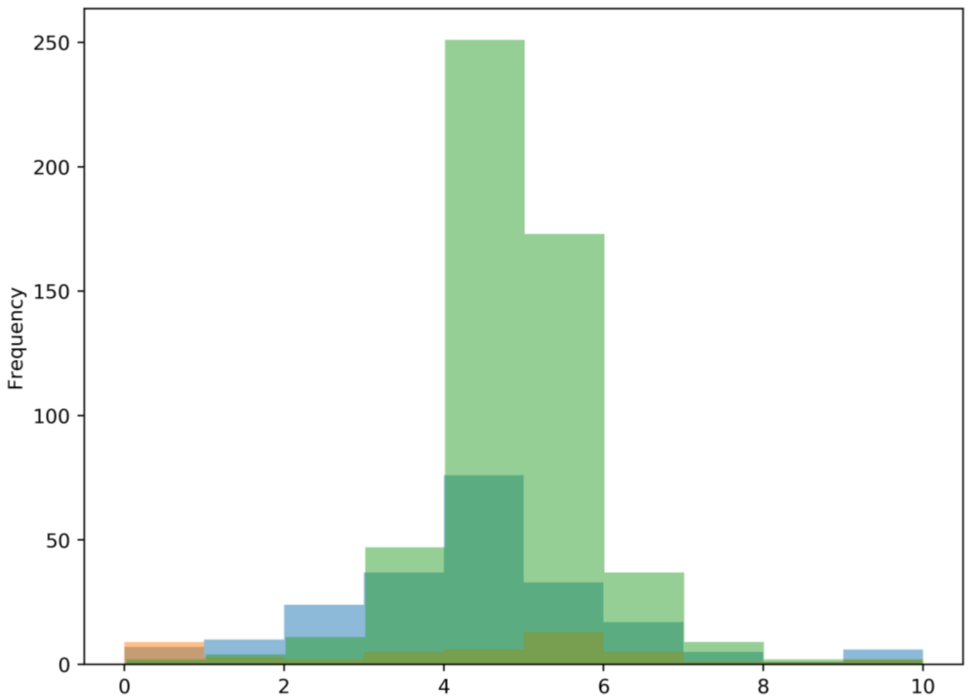


Figure 4. 4 Frequency Statistics of Efficiency Score of Three Auto-Identified User Types

Fig. 4.4 shows the frequency statistics of efficiency score of three classes. Class 2 has most samples. Counts of classes with label 0, 1, 2 are 216, 47, 538 respectively out of 801 samples. The distribution of three classes has different shapes as well.

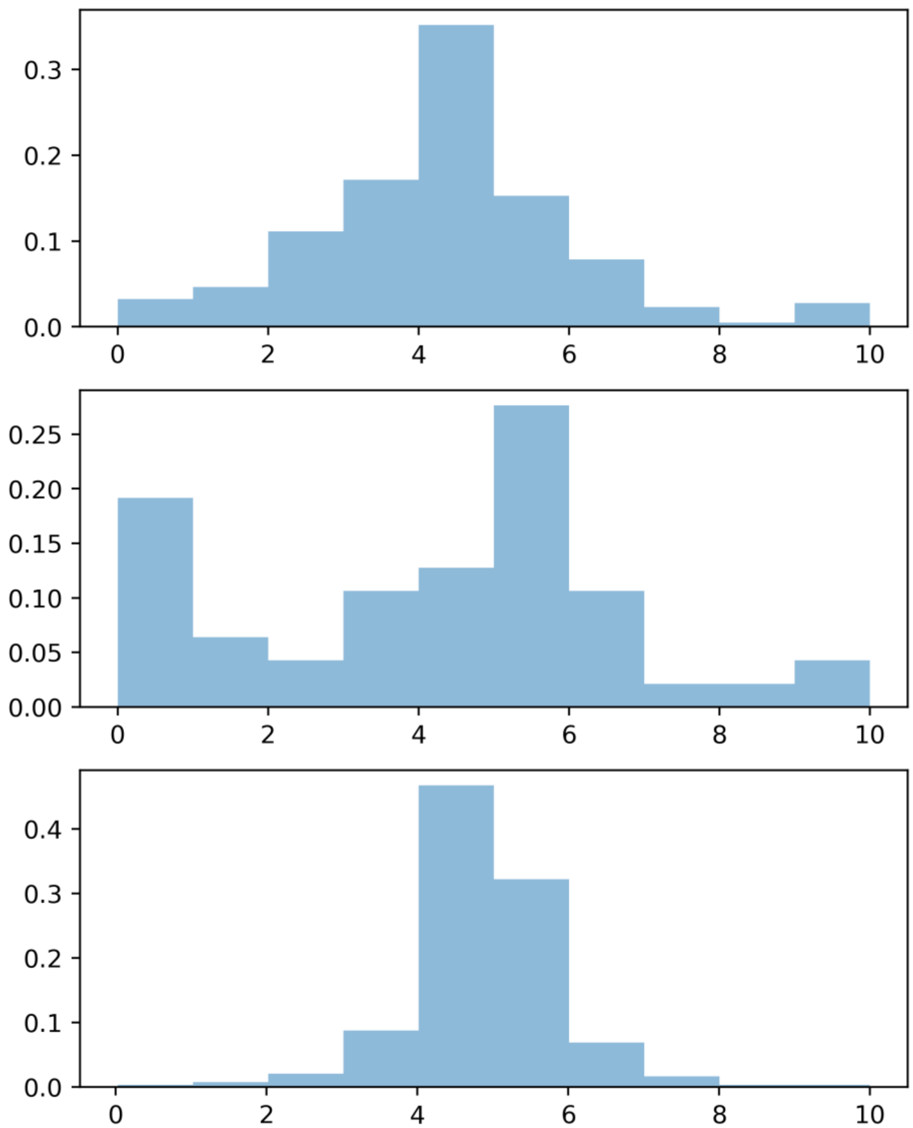


Figure 4. 5 Probability Densities of Efficiency Score of Three Auto-Identified User Types

Fig. 4.5 shows probability densities of three classes. From top to bottom are distributions for class 0, 1, and 2. The shape of sample distributions within each group is more clear from this figure.

4.2 Results of Manual Identification

This section contains three subsections describing properties of use score and classification of two methods, GMM and Jenks optimization respectively.

4.2.1 Efficiency score and use score

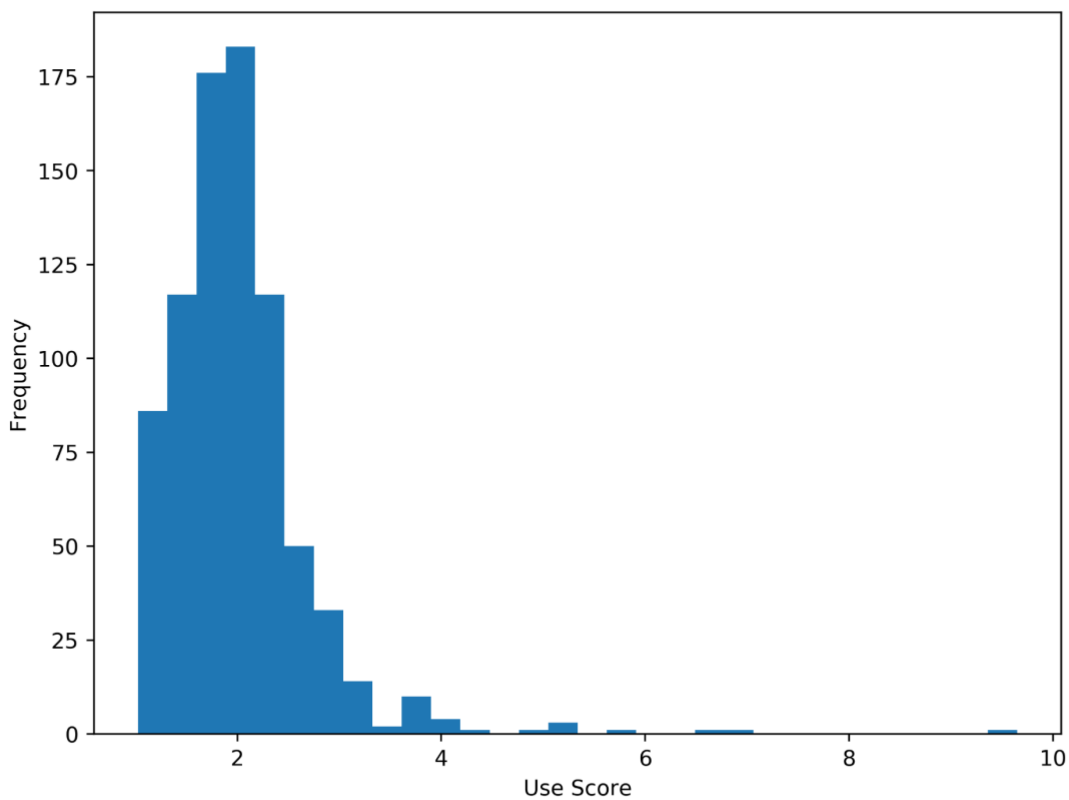


Figure 4. 6 Distribution of Use Score

Fig. 4.6 shows the distribution of use score of the entire dataset. The use scores vary from 0 to 10, with a mean of 2.000, a minimum of 1.025, and a maximum of 9.650. The use scores mainly concentrate at the lower region.

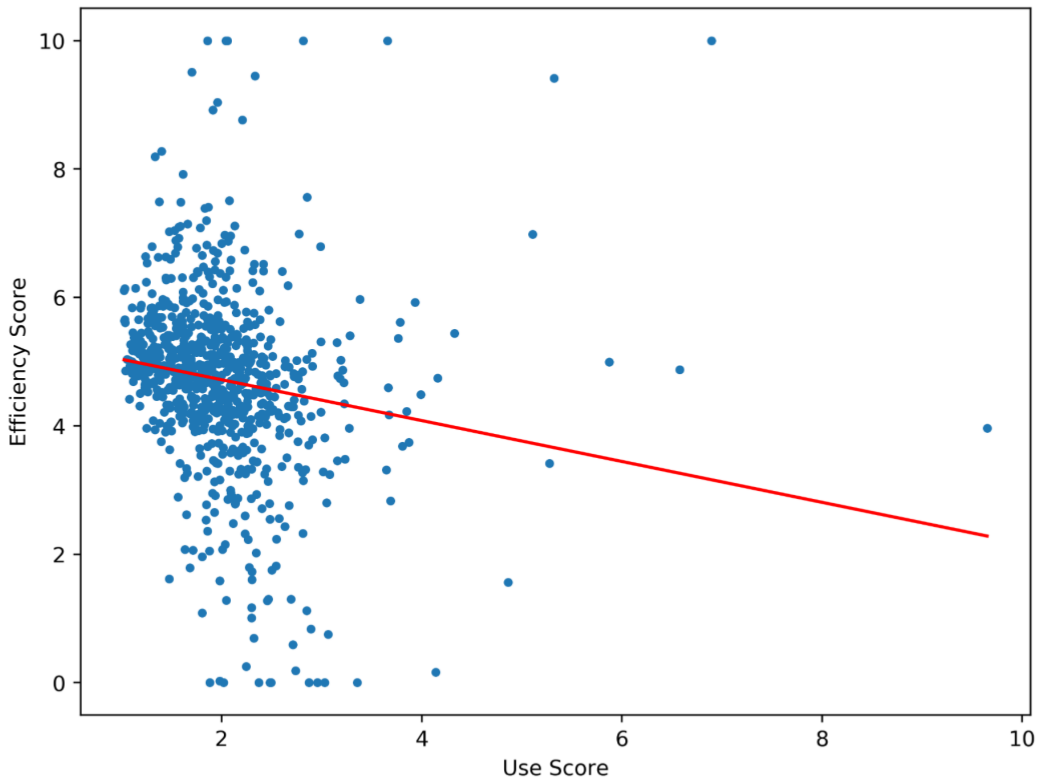


Figure 4. 7 Scatter Plot of Efficiency Score against Use Score

Fig. 4.7 shows a plot of use score against efficiency score. A red regression line is added on the plot displaying negative correlation between use scores and efficiency scores.

4.2.2 Identification with GMM

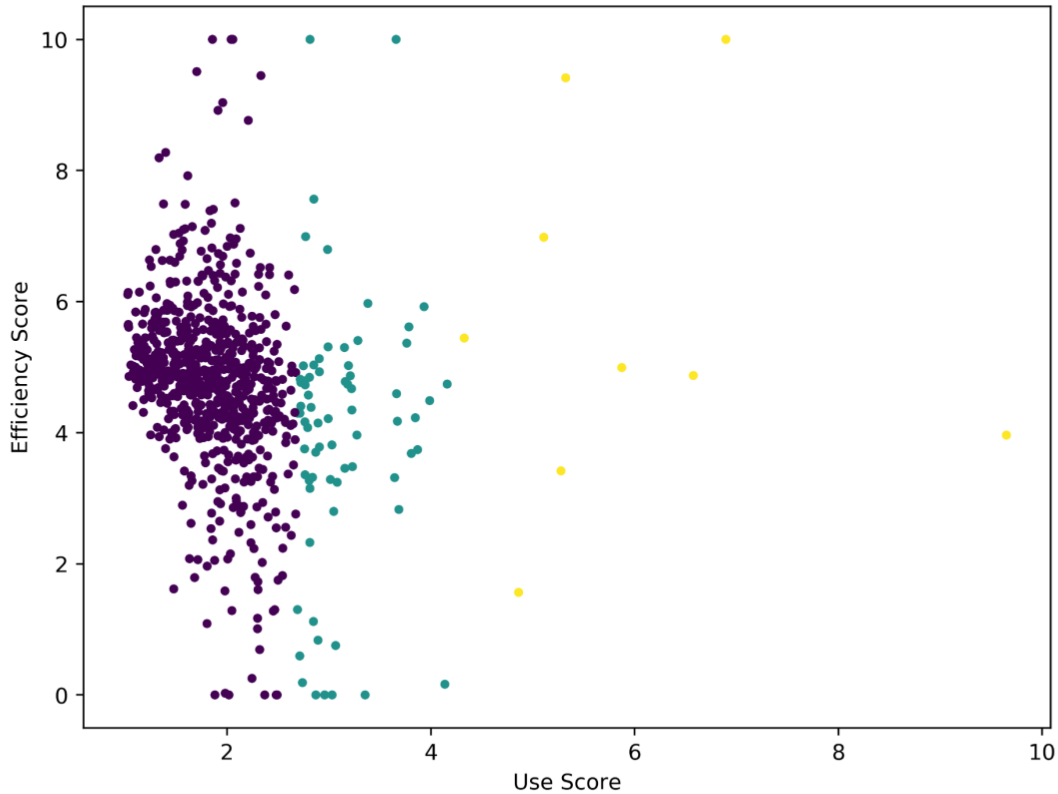


Figure 4. 8 Physician Identification Based on Use Score with GMM

Label	0	1	2
Mean	1.661	2.473	4.674
Standard deviation	0.294	0.300	1.418

Table 4. 1 Statistics of Use Score of Three EHR User Types Identified by GMM

Fig. 4.8 shows clustering results with GMM model. Purple, green, and yellow points indicate samples with class label 0, 1, and 2 respectively. Counts of classes with label 0, 1, 2 are 719, 73, 9 respectively out of totally 801 samples. Based on Miller-Sim model, classes with label 0, 1, and 2 represent basic users, strivers and arrivers.

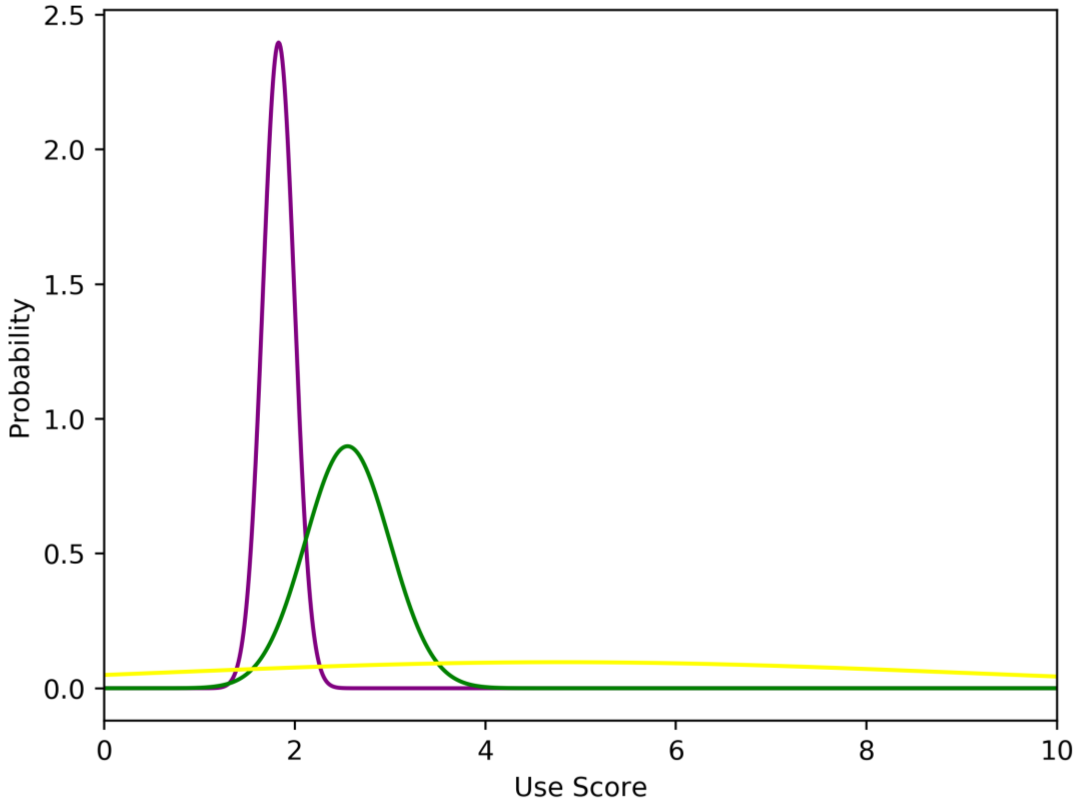


Figure 4. 9 Gaussian Distributions of Three EHR User Types Identified by GMM

Distribution	0	1	2
Mean	1.752	2.235	4.098
Standard deviation	0.138	0.287	3.066
Weight	0.415	0.041	0.544

Table 4. 2 Statistics of Use Score of Three Gaussian Distributions Identified by GMM

Fig. 4.9 is a visualization of three Gaussian distributions fitted by GMM model. Properties of three distributions estimate are shown in table 4.2.

4.2.3 Result of Jenks Natural Breaks Optimization

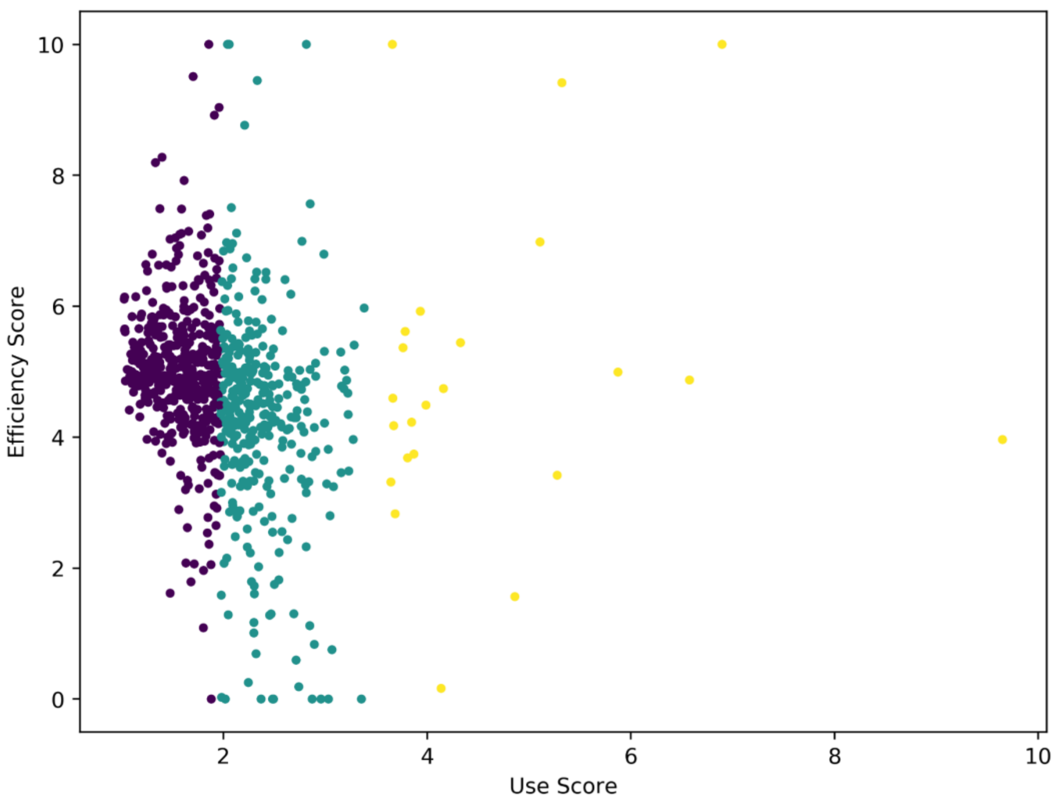


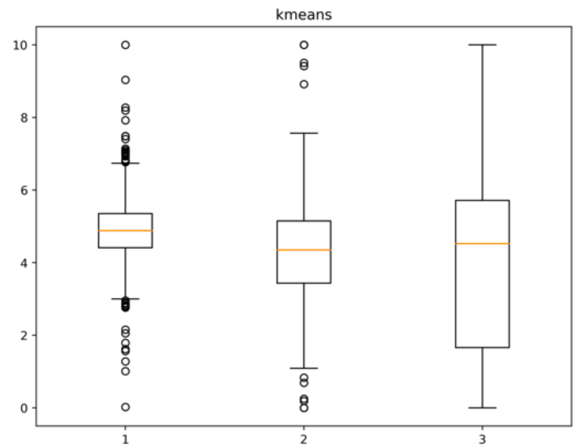
Figure 4. 10 Physician Identification Based on Use Score with Jenks Optimization

<table>
<tr><td>Label</td><td colspan="2">0</td><td colspan="2">1</td><td>2</td></tr>
<tr><td>Mean</td><td colspan="2">1.583</td><td colspan="2">2.359</td><td>4.674</td></tr>
<tr><td>Standard deviation</td><td colspan="2">0.260</td><td colspan="2">0.320</td><td>1.418</td></tr>
<tr><td>Cutoff</td><td colspan="3">1.97</td><td colspan="2">3.38</td></tr>
</table>

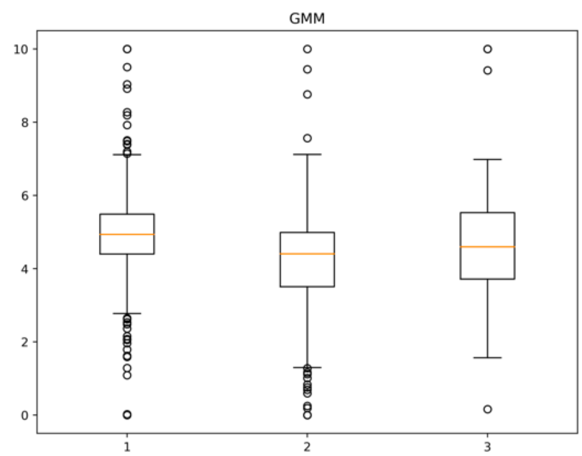
Table 4. 3 Statistics of Use Score of Three EHR User Types Identified by Jenks Optimization

Classification result using Jenks natural break optimization is shown in fig. 4.10. Purple, green and yellow points indicate samples with class label 0, 1, and 2 respectively. Similarly to GMM results, according to Miller-Sim model, classes with label 0, 1, and 2 represent basic users, strivers and arrivers. Counts of classes with label 0, 1, 2 are 439, 339, 23 respectively out of 801 samples. The cutoffs and statistics for three classes are displayed in table 4.3.

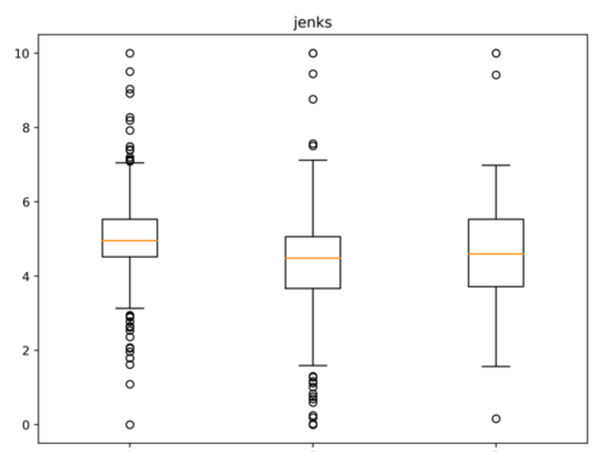
4.3 Comparison of Manual and Automatic Identification Results



(a)



(b)



(c)

Figure 4. 11 Boxplots of Efficiency Score of Three EHR User Types Identified by Each Method

method	k-means			GMM			Jenks		
label	0	1	2	0	1	2	0	1	2
mean	4.907	4.394	4.049	4.961	4.181	4.937	5.022	4.309	4.935
Standard deviation	0.983	1.769	2.610	1.188	1.555	2.335	1.088	1.569	2.343

Table 4. 4 Statistics of Efficiency Score of Three EHR User Types Identified by Each Method

In Figure 4.11 and Table 4.4, statistics of efficiency score for three methods are compared. Compared with K-means, classifications with manually engineered use score exhibit better performance in distinguishing efficiency score. Classifications with Jenks and GMM return similar results on arrivers, while for the other classes, the decision boundary of Jenks is lower than GMM.

Chapter 5: Discussion

The previous chapter has demonstrated that ambulatory physicians working with Epic EHR products have multiple choices regarding some certain identified EHR capabilities, but limited options for others. Both identification strategies – automatic identification based on all variables of interest, and manual identification based on physicians' interaction with EHR functions – of physicians' EHR user types show consistent results that physicians belong to the basic user stage have relatively high efficiency scores due to EpicCare PEP evaluation, while strivers have relatively low efficiency scores. Arrives have no higher efficiency scores than basic users, but show a little better performance than strivers on work efficiency through manual identification approaches.

In this chapter, I deliberate over some interesting findings in my study with exploration of external literature. Relevant limitations and future work are also discussed.

5.1 Implications

Implications of this study should be of interest to researchers focusing on EHR usability, provider work efficiency, or clustering method application, developers concerned about EHR functions and EHR system design, and managers making decisions on supporting providers move forward with better EHR use.

5.1.1 Applied Approaches

It is gratifying that each of the three clustering approaches from two identification strategies has successfully identified three physician groups from EpicCare PEP dataset, and identification results of each method has their meanings and applicability in different settings. Through results from all identification approaches, we find that among ambulatory physicians working with EpicCare EHR systems, basic users cover the majority of physicians (>54%), while arrivers only account for the minority (<6%). The K-means method for the automatic identification strategy did a great work for clustering samples with high-dimensional PEP dataset. It was the first method came into my mind when I accessed the original data and started thinking about grouping samples into different types using as much as possible information. Through the implementation of K-means method, all variables related to physicians' EHR use were taken into consideration while clustering, including "Patient Call/ Results/ Prescription Authorization Messages Received/ Incomplete in PEP Report Period", which were excluded in physicians' use score metrics in my second identification strategy. PEP was created much based on the time providers actually spent on EHRs versus time scheduled for them, so a large portion of PEP data was about providers' workload. How many what kinds of tasks are physicians assigned to complete is definitely a key component of measuring physicians' work performance and efficiency, but relevant variables like "Patient calls received" cannot directly reflect physicians' use conditions of EHRs. Also, it might be more reasonable to measure use-related variables with meaningful weights and in different ways with the purpose of better assessing physicians' EHR use. Therefore, a second study where identifying physician EHR user types based on their use of several main EHR capabilities– viewing, documenting, ordering and other functions – defined by Miller and Sim, was designed as a different attempt.

Other than automatic identification, which to some extent lacked support from previous research work on the meaning and interpretation of each physician type it presented, manual identification was able to clarify three physician types it identified as basic users, strivers, and arrivers with confidence of use score measurements and Miller and colleagues' model. Either one of the two common classification methods applied has its own advantages and implications for future similar work in different contexts. GMM was conducted under the assumption that each type of physician samples was following a normal distribution. This helps weaken the influence of noise points while clustering, so it would be a good method if the assumption is true. However, as shown in Figure 4.9, it is less likely that samples within each physician type completely follow a normal distribution in our case, after then Jenks natural breaks optimization was implemented to explore potential different identification results from GMM, with minimization the variance within each user types and maximization of the variance between each user type under no extra assumptions. Despite the strengths of natural breaks, it is important to recognize that Jenks optimization can be super time-consuming with a large dataset or the requirement of a larger number of classes, in case more detailed sub-types of physicians are expected to be identified in the future.

5.1.2 EHR Functionality

Use score metrics focus more on the smart and quick use of EHR comprehensive functions instead of just judging on the amount of time the physician spent in the systems. One interesting and meaningful finding in this study is maximal use of EHR documenting function and minimal use of other EHR capabilities in Epic EHR systems. From PEP Dashboard data, how notes were entered by healthcare providers was comprehensively measured. Electronic data including progress notes, patients' chief complaints, diagnoses, allergies, and prescriptions was recorded in user actions of manual entry, SmatTool entry, NoteWriter entry, Copy/Paste entry, voice recognition entry, transcription entry and other entry. In physicians use score metrics, we valued notes through SmartTool and NoteWrite entry the most, while notes through recognition and transcription entry the least. From Miller and Sim's model, the more advanced stage the physician stays, the more likely they would use smart tools of EHRs to help them save time and improve work efficiency. Features like the number of Note Speed Buttons, which allow users to pull in note templates quickly, also differentiate basic users, strivers, and arrivers to some extent. Although technology using habits might be one of impacting factors of physicians' choices, basic users are observed to tend to dictate notes while viewing templates, and train transcriptionists to fill out the templates as they transcribed. This is keeping adding costs of transcription, and tests or reports scanning. Interventions or trainings about how to document more efficiently by using smart tools and customized templates can be designed, with combination of more analysis on the documenting patterns of physicians belonging to the arriver stage.

In this study, *other* capability initially intended to include relevant data reflecting some unusual EHR functions such as messaging, patient-oriented, care management, analysis and reporting, billing, scheduling and access or information retrieval. As mentioned earlier, some of these capabilities were immature or not considered in Epic EHR systems, so there are only five variables from PEP dataset categorized in *other* EHR capability, and this section itself accounts for a small proportion in the whole use-score measure. Nevertheless, we find the use of these unusual EHR capabilities is closely related to the improvement of care quality. For example, the percentage of patient advice requests done promptly by physicians through EHR systems affects patient experience and satisfaction a lot, which may facilitate or hinder physician-patient communications, as well as healthcare quality improvement. Other studies have also shown that comprehensive EHR use with functions of health information exchange[49], acceptance of best practice alerts, viewing panel-level reports, and use of order sets[14] across delivery settings is associated with higher patient care quality. In addition, from Figure X in Appendix, we can see obvious differences between arrivers and the other two EHR use types on the *other* measure column. For arrivers, the use of viewing capability is positively correlated to the use of *other* capability, which indicates that physicians in relatively advanced EHR use stages may reach out to more EHR capabilities beneficial to patients, like care management or appointment/test scheduling while viewing visit notes.

5.1.3 Interpretation of Identification with Efficiency

From Figure 4.11 and Table 4.4, similar trends of efficiency score defined by EpicCare PEP Metrics among three physician EHR user types have been shown by three specific identification methods. Physicians belong to the basic user type have relatively high efficiency scores, while strivers have relatively low efficiency scores. Arrives have no higher efficiency scores than basic users, but show a little better performance than strivers on work efficiency through manual identification approaches based upon use score metrics. As mentioned before, *efficiency score* from PEP Metrics is based on the amount of time a provider spent in the systems versus the amount of time researchers predict in terms of the provider's workload. We can interpret from the results that basic users always spend fair time as they are supposed to on EHR use, while physicians who already move to the striver stage spend much more time in the systems, which explains the significant decrease in their work efficiency compared to basic users. This well matches Miller and Sim's finding that strivers started to invest substantial extra time on complementing EHRs with the hope of generating more benefits, while basic users remained satisfied with their low level of EHR use and benefits. The variance of efficiency score of the arriver physician type is very large, so we can hardly tell the difference between its mean and the other two types with much confidence. Nevertheless, we can still interpret the negative correlation between efficiency score and use score. Efficiency improvement is one of the most important objectives of EHR implementation, but it can neither be wholly reflected by individual user action log data nor represent all the benefits coming with EHR use. Although it seems like arrivers have no significant higher efficiency than the other two previous user types, it is worth to be aware that most arrivers have reorganized the workflows of their exam rooms or the entire offices. Their work efficiency should not be simply judged by their own EHR use from audit log data. For example, a physician of the arriver stage could ask nurses in her office for help entering patient data into EHRs while tracking down charts or viewing more history information by herself. This would reduce the time the physician need to document in the systems, yet actually improve work efficiency and even produce better care for patients. From this perspective, we also challenge the view of PEP that managers should focus on providers with scores of 0 and 1 before providers with scores of 3 and 4 when targeting providers for extra support. Efficiency score cannot interpret physicians' use of EHRs and their work conditions comprehensively. A provider with an efficiency score of 8 can be either a basic user remaining doing paper work and interacting little with EHRs, or an arriver who arranges EHR tasks to each stuff member in her office and spends more time communicating with patients and families during visits. Therefore, we suggest combination of more analysis on physicians' work patterns besides efficiency score from PEP before additional training design or policy making.

5.2 limitations and Future Work

There are specific limitations to this study from perspectives of data source and quality, question and aim scope, and study design.

Firstly, 801 samples in this study may not be representative of the entire physician population, let alone other types of EHR users such as nurses, pharmacists, and nutritionists. Analysis on EHR user status of other types of providers other than physicians may need to seek other frameworks before conducting studies with similar or different approaches applied in this study. PEP Dashboard data is structured and relatively well-organized. Specifically, most types of data in PEP are as descriptive as “clinical viewing minutes” and “orders minutes”, instead of rawer data like “15:00:00 – 15:03:27 clinical viewing” and “15:03:30 – 15:08:23 ordering”. This reduced a lot of work on data cleaning but as a tradeoff, missed much timeline information that could be useful to analyze concrete use patterns to identify “good” or “bad” EHR users. Also, EHR usability and work efficiency discrepancy among different user types might be inconsistent with the results in this study in other EHR products/systems.

Secondly, this study was initially aimed to analyze EHR usability and provider work efficiency, but only conducted in the way of identifying physicians and their EHR user types within limited time, although this would not affect main contributions of this study to the general knowledge. If more time allowed, detailed analysis can be made in a single office, exam room or among different specialties and service areas with current data. Comprehensively understanding EHR use of each member in the office, or the specialty can assist stakeholders better determine whom to help and how to support.

Lastly, there are few limiting factors in the process of study design and method application. One of the biggest drawbacks would be use score measures in the manual identification strategy. Use of EHR functions in reality is never less complicated than the model. Some variables in PEP dataset like “number of smart buttons” are of great significance in advanced EHR use but can be hardly categorized into any main EHR capabilities like “viewing” or “documenting” unless more details are informed. Additionally, this study simply considered variables directly related to EHR use, but other factors indirectly affect providers’ interaction with the systems should be valued as well. For example, information such as the percentage of new patients to the provider during the evaluation period, or demographics (e.g. gender, age) of patients would greatly influences the time a provider spend in the systems during visits. The problem of lacking ample literature support and objectivity on the use score metrics is expected to be better solved in the future.

Other method limitations have already been discussed and can be restated as (1) random initial center points and no weights for different EHR functions in K-means; (2) it is less likely that samples within each physician type obey the assumption of exhibiting normally distributed use scores for GMM; (3) exponential computation complexity of Jenks optimization algorithm will make it time-consuming when dealing with large datasets or complicated multi-class problems. Future work could attempt applications of semi-supervised methods to identify EHR user types to reduce negative effects of these limitations. One potential way is to first identify several samples as examples of each user type by manual data analysis, then use methods like K-means to group samples with well-determined sample point centers.

Chapter 6: Conclusion

In this study, I applied Miller-Sim model of EHR capabilities and user types to a large physician population, to quantitatively analyze different types of EHR users and their work efficiency with EpicCare systems. Several meaningful implications were found. First of all, most physicians (>54%) working with EpicCare were identified as basic users, while arrivers account for the minority (< 6%). The trends of physician work efficiency among three EHR user types identified – basic users, strivers, and arrivers – were consistent with our hypothesis that basic users have the highest work efficiency; strivers have relatively low efficiency; and arrivers have medium work efficiency but probably gain other benefits from EHR use. Another interesting finding in this study is maximal use of EHR documenting function with multiple choices on it, and minimal use of other unusual EHR capabilities such us *care management* or *analysis and reporting*, which are probably more related to care quality improvement (QI) in EpicCare EHR systems. I suggest implications in this study to be considered by researchers who are interested in EHR usability, EHR benefits and healthcare provider efficiency; policymakers in government or managers in hospitals who make decisions on which type of providers to give what kind of support regarding EHR use and benefits; and developers who are required on EHR function updates or relevant system design.

Here we propose some additional studies that could be carried out based on the findings in this study. Firstly, when more data come in, a study on EHR use analysis for the same population of physicians over time could generate meaningful results for EHR use training. As the observation continues, by evaluating EHR use of same physician over time, a provider's user type may change from basic user to striver, or striver to arriver, since she may start to learn how to use advanced functions of the systems or getting more familiar with some EHR capabilities so that more time and cost savings would be generated. By tracking these use pattern changes over time, associated with EHR user type evaluation, we will be able to identify key factors that prevent or facilitate users to move forward, thus provide better guidance for basic users and strivers to use EHRs better.

Another study related to physicians' satisfaction or experience working with EHR systems can also be combined. In this current study, UAL data from EpicCare PEP Dashboard was the main data source to characterize physician types. Objectivity is one of the metrics of UAL data leading to a meaningful identification of user types with a relatively large sample size. However, subjective perspectives from EHR users can hardly be reflected by UAL data, yet it can provide important insights to user behavior and is essential for EHR venders to design or improve their products in a user-centered and user-friendly way. Thus, qualitative studies taking the form of interviews or questionnaires are suggested to be carried out as follow-up work, targeting representative users of different EHR user types, to gather valuable first-hand information from their experience and comments on the systems.

Additionally, in [50], Li et al. combined a statewide survey data on physicians' adoption of EHRs indicating their EHR use, and claims data reflecting quality of care indicating quality of the services provided by physicians to investigate the relationship between EHR use and quality of care. Similar studies can be conducted by including patient outcome data or satisfaction data in addition to physician EHR use dataset in order to study the effect of EHR use on improving quality of care. By studying relationships between EHR use and healthcare quality, we may be able to identify types of physicians that provide different quality of services with different use of EHRs, as well as key features of EHR capabilities that help with QI. After all, improving quality of care for patients is one of the ultimate goals of EHR implementation.

I am incredibly pleased to have done this work on the identification of physicians and their EHR user types from the EpicCare Ambulatory Provider Efficiency Profile(PEP) dataset. My implications were deeply meaningful and highly impactful for future work of facilitating healthcare providers to move forward with EHR use, as well as contributed to the general knowledge about EHR user types and EHR benefits. Nowadays, technology is not to blame for restricted value achieved by hospitals, policies advocating EHR implementation and benefits should more consider about incentives leading to comprehensive EHR use.

Reference

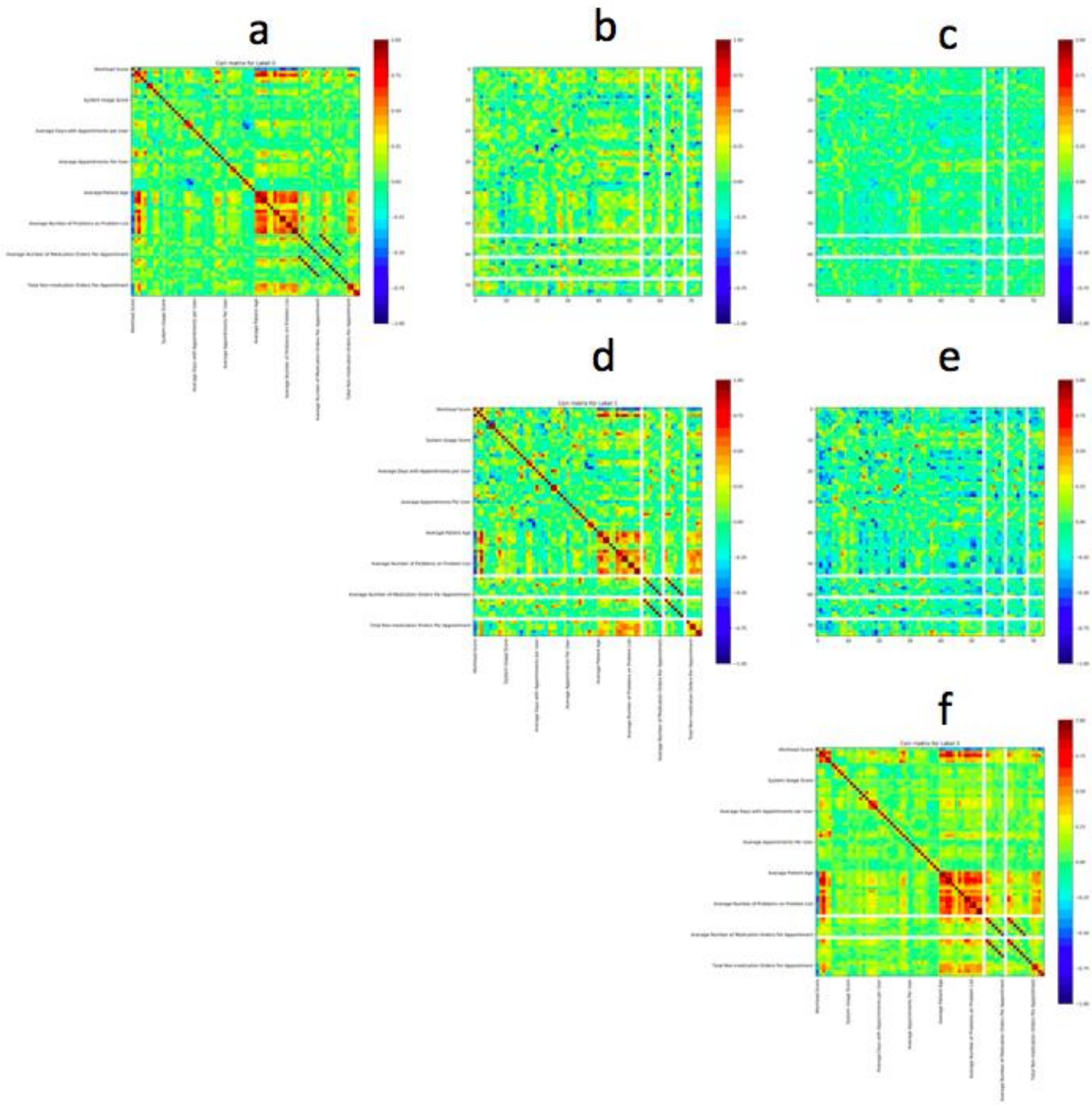
[1] R. H. Miller, I. Sim, and J. Newman, “Electronic medical records in solo/small groups: a qualitative study of physician user types,” *Stud Health Technol Inform*, vol. 107, no. Pt 1, pp. 658–662, 2004.
[2] J. Henry, Y. Pylypchuk, and T. Searcy, “Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015,” p. 11.
[3] R. H. Miller and I. Sim, “Physicians’ Use Of Electronic Medical Records: Barriers And Solutions,” *Health Affairs*, vol. 23, no. 2, pp. 116–126, Mar. 2004.
[4] D. Willson *et al.*, “Computerized support of pressure ulcer prevention and treatment protocols,” *Proc Annu Symp Comput Appl Med Care*, pp. 646–650, 1995.
[5] D. W. Bates *et al.*, “Effect of computerized physician order entry and a team intervention on prevention of serious medication errors,” *JAMA*, vol. 280, no. 15, pp. 1311–1316, Oct. 1998.
[6] K. F. Schmitt and D. A. Wofford, “Financial analysis projects clear returns from electronic medical records,” *Healthc Financ Manage*, vol. 56, no. 1, pp. 52–57, Jan. 2002.
[7] N. Menachemi and T. H. Collum, “Benefits and drawbacks of electronic health record systems,” *Risk Manag Healthc Policy*, vol. 4, pp. 47–55, 2011.
[8] S. Silow-Carroll, J. N. Edwards, and D. Rodin, “Using electronic health records to improve quality and efficiency: the experiences of leading hospitals,” *Issue Brief (Commonw Fund)*, vol. 17, pp. 1–40, Jul. 2012.
[9] J. Thurston, “Meaningful Use of Electronic Health Records,” *The Journal for Nurse Practitioners*, vol. 10, no. 7, pp. 510–513, Jul. 2014.
[10] M. Rouse, “What is meaningful use? - Definition from WhatIs.com,” *SearchHealthIT*. [Online]. Available: https://searchhealthit.techtarget.com/definition/meaningful-use. [Accessed: 05-May-2018].
[11] T. Bodenheimer and C. Sinsky, “From Triple to Quadruple Aim: Care of the Patient Requires Care of the Provider,” *The Annals of Family Medicine*, vol. 12, no. 6, pp. 573–576, Nov. 2014.
[12] C. Sinsky *et al.*, “Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties,” *Annals of Internal Medicine*, vol. 165, no. 11, p. 753, Dec. 2016.
[13] M. J. Romano and R. S. Stafford, “Electronic health records and clinical decision support systems: impact on national ambulatory care quality,” *Arch. Intern. Med.*, vol. 171, no. 10, pp. 897–903, May 2011.
[14] N. Ratanawongsa *et al.*, “Association Between Clinician Computer Use and Communication With Patients in Safety-Net Clinics,” *JAMA Intern Med*, vol. 176, no. 1, pp. 125–128, Jan. 2016.
[15] M. W. Friedberg *et al.*, “Factors Affecting Physician Professional Satisfaction and Their Implications for Patient Care, Health Systems, and Health Policy,” *Rand Health Q*, vol. 3, no. 4, p. 1, 2014.
[16] P. D. Clayton *et al.*, “Physician use of electronic medical records: issues and successes with direct data entry and physician productivity,” *AMIA Annu Symp Proc*, pp. 141–145, 2005.

[17] B. Chaudhry *et al.*, “Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care,” *Annals of Internal Medicine*, vol. 144, no. 10, p. 742, May 2006.
[18] L. Poissant, J. Pereira, R. Tamblyn, and Y. Kawasumi, “The impact of electronic health records on time efficiency of physicians and nurses: a systematic review,” *J Am Med Inform Assoc*, vol. 12, no. 5, pp. 505–516, Oct. 2005.
[19] S. Read-Brown *et al.*, “Time Requirements for Electronic Health Record Use in an Academic Ophthalmology Center,” *JAMA Ophthalmology*, vol. 135, no. 11, p. 1250, Nov. 2017.
[20] “Electronic Medical Record Adoption Model,” *EMRAM: A strategic roadmap for effective EMR adoption and maturity*, 24-Jan-2017. [Online]. Available: https://www.himssanalytics.org/emram. [Accessed: 06-May-2018].
[21] M. Hagland, “What Does EMR Optimization Actually Mean? So Much More Than Just Working Out the Kinks,” *Healthcare Informatics Magazine*, 29-Oct-2015. [Online]. Available: https://www.healthcare-informatics.com/blogs/mark-hagland/what-does-emr-optimization-actually-mean-so-much-more-just-working-out-kinks. [Accessed: 06-May-2018].
[22] K. McNamara, “EHR optimization,” *Electronic health records optimization: What is it and what’s in it for you?*, 17-Apr-2015. .
[23] K. Murphy, “What Is EHR Optimization, How Does It Start?,” *EHRIntelligence*, 15-Jun-2016. [Online]. Available: https://ehrintelligence.com/features/what-is-ehr-optimization-how-does-it-start. [Accessed: 07-May-2018].
[24] N. Pandhi *et al.*, “Approaches and challenges to optimising primary care teams’ electronic health record usage,” *Inform Prim Care*, vol. 21, no. 3, pp. 142–151, 2014.
[25] M. E. Morton and S. Wiedenbeck, “A framework for predicting EHR adoption attitudes: a physician survey,” *Perspect Health Inf Manag*, vol. 6, p. 1a, Sep. 2009.
[26] F. D. Davis, “Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology,” *MIS Quarterly*, vol. 13, no. 3, p. 319, Sep. 1989.
[27] V. Venkatesh and F. D. Davis, “A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies,” *Management Science*, vol. 46, no. 2, pp. 186–204, Feb. 2000.
[28] Venkatesh, Morris, Davis, and Davis, “User Acceptance of Information Technology: Toward a Unified View,” *MIS Quarterly*, vol. 27, no. 3, p. 425, 2003.
[29] C. S. Kruse, J. DeShazo, F. Kim, and L. Fulton, “Factors associated with adoption of health information technology: a conceptual model based on a systematic review,” *JMIR Med Inform*, vol. 2, no. 1, p. e9, May 2014.
[30] D. L. Goodhue, “Understanding User Evaluations of Information Systems,” *Management Science*, vol. 41, no. 12, pp. 1827–1844, Dec. 1995.
[31] S. Sarnikar, D. Bennett, and M. Gaynor, Eds., *Cases on Healthcare Information Technology for Patient Care Management:* IGI Global, 2013.
[32] Q. Gan, “Is the adoption of electronic health record system ‘contagious’?,” *Health Policy and Technology*, vol. 4, no. 2, pp. 107–112, Jun. 2015.
[33] O. F. E. Gayar, A. V. Deokar, and M. J. Wills, “Evaluating task-technology fit and user performance for an electronic health record system,” *International Journal of Healthcare Technology and Management*, vol. 11, no. 1/2, p. 50, 2010.

[34] M. Cain, R. Mittman, Institute for the Future, and California HealthCare Foundation, *Diffusion of innovation in health care*. Oakland, Calif.: California Healthcare Foundation, 2002.
[35] C. Anderson, “Presenting and evaluating qualitative research,” *Am J Pharm Educ*, vol. 74, no. 8, p. 141, Oct. 2010.
[36] M. B. Buntin, S. H. Jain, and D. Blumenthal, “Health Information Technology: Laying The Infrastructure For National Health Reform,” *Health Affairs*, vol. 29, no. 6, pp. 1214–1219, Jun. 2010.
[37] K. Lynch *et al.*, “The Health IT Regional Extension Center Program: Evolution and Lessons for Health Care Transformation,” *Health Services Research*, vol. 49, no. 1pt2, pp. 421–437, Feb. 2014.
[38] D. Blumenthal, “Launching HITECH,” *New England Journal of Medicine*, vol. 362, no. 5, pp. 382–385, Feb. 2010.
[39] J. A. Chilingerian and H. D. Sherman, “Managing Physician Efficiency and Effectiveness in Providing Hospital Services,” *Health Services Management Research*, vol. 3, no. 1, pp. 3–15, Mar. 1990.
[40] J. A. Chilingerian, “Evaluating physician efficiency in hospitals: A multivariate analysis of best practices,” *European Journal of Operational Research*, vol. 80, no. 3, pp. 548–574, Feb. 1995.
[41] M. T. Dastagir, H. L. Chin, M. McNamara, K. Poteraj, S. Battaglini, and L. Alstot, “Advanced proficiency EHR training: effect on physicians’ EHR efficiency, EHR satisfaction and job satisfaction,” *AMIA Annu Symp Proc*, vol. 2012, pp. 136–143, 2012.
[42] R. Koppel and C. U. Lehmann, “Implications of an emerging EHR monoculture for hospitals and healthcare systems,” *Journal of the American Medical Informatics Association*, Oct. 2014.
[43] “EpicCare Ambulatory Provider Efficiency Profile Metrics,” p. 31.
[44] D. Arthur and S. Vassilvitskii, “k-means++: The Advantages of Careful Seeding,” p. 11.
[45] L. van der Maaten and G. Hinton, “Visualizing data using t-SNE,” *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.
[46] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *MACHINE LEARNING IN PYTHON*, p. 6.
[47] D. Reynolds, “Gaussian Mixture Models,” in *Encyclopedia of Biometrics*, S. Z. Li and A. K. Jain, Eds. Boston, MA: Springer US, 2015, pp. 827–832.
[48] R. McMaster, “In Memoriam: George F. Jenks (1916-1996),” *Cartography and Geographic Information Systems*, vol. 24, no. 1, pp. 56–59, Jan. 1997.
[49] I. Graetz, M. Reed, S. M. Shortell, T. G. Rundall, J. Bellows, and J. Hsu, “The next step towards making use meaningful: electronic information exchange and care coordination across clinicians and delivery sites,” *Med Care*, vol. 52, no. 12, pp. 1037–1041, Dec. 2014.
[50] L. Zhou *et al.*, “The Relationship between Electronic Health Record Use and Quality of Care over Time,” *Journal of the American Medical Informatics Association*, vol. 16, no. 4, pp. 457–464, Jul. 2009.

Appendix

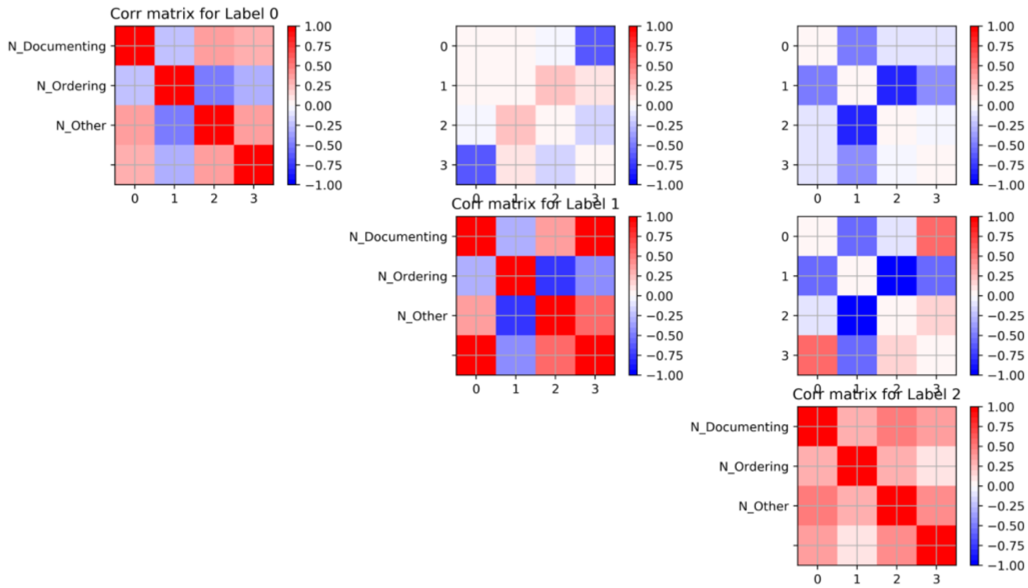
Fig 1. Correlation matrixes of variables across groups identified by K-means:



In Fig. 1, a, d, and f show correlation matrix of all numeric variables from each class, a for class 0, d for class 1, and f for class 2. b, c, and e show pointwise differences of correlation matrixes of two classes, b for class 0 and 1; c for class 0 and 2; and e for class 1 and 2. For correlation matrixes, color indicates values of correlation coefficients. For difference matrixes, color indicates the degrees of difference. From the difference matrixes, we can see that variable correlations between different classes are different, and class 2 has more differences with other two classes.

Fig 2. Correlation matrixes of EHR user capabilities across groups identified by:

(1) GMM:



(2) Jenks: