Exploring Variations from Electronic Medical Records (EMRs): A Case Study of Chronic Disease

Praowpan Tansitpong¹, Wanpracha Chaovalitwongse², and Apirak Hoonlor³

¹ NIDA Business School, Bangkok, Thailand
² Department of Industrial Engineering and Institute for Advanced Data Analytics, University of Arkansas, Fayetteville, USA
³ Faculty of Information and Communication Technology, Mahidol University, Nakon Pathom, Thailand
Email: praowpan.tan@nida.ac.th, artchao@uark.edu, apirak.hoo@mahidol.edu

Abstract—Health policy reform in major developing countries has changed how patients are treated for chronic diseases due to high cost and government regulations. In order to understand the overall patterns of the treatments, this study explores methods to collect Electronic Medical Records (EMRs) and analyzes patterns in prescription plan and treatment cost in health coverage programs from local data source. Focusing on admitted patients with conditions related to diabetes and hypertension, the results suggested high variations of treatment in patients with hypertension, but very limited variations among diabetes patients. The results also summarized prescription lists and treatment cost among health benefit programs.

Index Terms—electronic medical records, healthcare operations, prescriptions, pattern identification

I. INTRODUCTION

Managing treatments for chronic diseases is critical for developing countries due to long term nature of frequent medical needs and continuous monitoring process [1]. While modern devices and sensor technologies allow physicians and medical staffs to create treatment individualization and real-time information for medical service delivery, these technologies may not be available in most countries of the world [2]. Treatments for chronic conditions are focusing mainly on monitoring symptoms such as blood pressure control at home, however, electronic medical process of acute treatments for hypertensive- and diabetic-related admissions has not been vastly studied [3]-[5]. Treatment for chronic diseases became new challenges in healthcare management due to increasing number of aging patients and high cost of long-term treatments. Majority of patients with chronic diseases, such as diabetes and hypertension, are facing the increasing cost and diversities of prescription drugs. Conditions and costs are also intertwined; combined conditions among diseases could add up into cost of treatment. Two third of diabetes patients are diagnosed with high blood pressure according to the Centers for Disease Control and Prevention [6]. Number of patients with conditions of hypertension and high blood pressure are accounted for nearly half of US population. In western countries, patients are estimated to spend an estimate of $6,000 per year on expenses related to chronic diseases; the cost could be raised to $13,000 per year for diabetes.

Physicians make various decisions in treating patients due to market competition between drug choices [7]-[9]. According to European market, high fixed cost for higher quality in active ingredients forced generic drug manufacturers to be inflexible in price competition; brand name drugs gained benefit to lower the price and became popular choice in European copayment system [10]. In a study of health systems in Italy, frequency of prescriptions is made due to information spillovers and word-of-mouth [11]. Repetitive behavior of physicians is varied due to common practice, observed quality, or new learning experience from previous treatment or medical society. As approximately 5 percent of patients are accounted for 50 percent of health care spending in the US [12], prescription patterns reflect similar characteristics of overall healthcare spending where majority of drugs that are highly used and prescribed and accounted for majority of the cost. In general cases, physicians make decisions on behalf of patients to optimize treatment provide the best benefit for patients, however, some influences such as health administered organization may shape drug choice decision. Drug choices can be decided base on factors such as drug options and types, repetitive physician’s decisions, knowledge on drugs, and prior experience. Suggested by literature, physicians tended to prescribe more expensive drugs or higher dosage when prescriptions are co-paid [13], [14]. Health administered programs may play role in determining drug options due to criteria in quality of treatment [15]-[17]. For chronic disease treatments, more than 500 preferred brand names and generic brands to treat chronic conditions are available in the market. The list is extended triple for non-preferred drugs [18] which incurs researchers to define pattern of choices to prevent over-prescriptions and overdiagnosis [19], [20]. Most

Manuscript received November 7, 2019; revised March 15, 2020.
importantly, cost and profit made in compliance with drug choices may be made specific guidelines of health administered programs. It is important for researchers to understand how guidelines designate orders for prescriptions on multiple benefit plans. Hence, the goal of this study is to utilize data mining techniques to identify prescription patterns in health administered programs, and to clarify whether health insurance and benefit plans play role in changing the decision of drug choices.

II. LITERATURE REVIEW

Decisions made by physicians in treatment process are critical; one physician may have different decision from others. Due to rising prescription drug cost, literature provided supports that physicians and health benefit programs plays significant role in prescriptions [21]-[23]. A survey evidence suggested that cost, quality, efficacy of drugs, along with physicians’ discretions are important in supporting treatment decision. Electronic Medical Records (EMRs) could facilitate modern medical workers in understanding health process, the records also provided opportunities for medical society to observe medical process and improve efficiency and utilization [24]-[28]. Researchers utilized EMRs to characterize patient segments [29]-[34]. Although participation with patients do share influences in medical treatment decision making process [35]-[38], these decisions are varied. In the review paper from Bates et al. [39], the study suggested opportunities in using data analysis to prioritize patients from causes such as readmissions, triage, decompensation, and adverse events, and highlighted difficulties for healthcare system to optimize treatment complexities for patients.

Data analytics in health applications is widely studied in previous literature. Research community studied emerging issues of chronic disease by observing bioinfomatics and assessing risk factor for individuals in health process [40]-[42]. Data mining in health process offers an understanding and exploration of patient-related data to help reduced cost of treatments, facilitated with compliances, and improved performance of health providers [43]-[45]. Medical electronic records created opportunities for researchers to access clinical practices from coded diagnostic process, patient demographics, visits history, and prescribed medications related to each visit. However, while most of the studies in the past were focusing on identifying phenotypes, seeking algorithm for comorbidity, and exploring personalized prescription for individuals [46]-[49], less attention was paid to understand patterns in prescription behavior and equity of care. As chronic disease become major global concerns in modern health systems, monitoring standard of treatment requires cooperative database sharing among national public health administers, healthcare providers, insurance companies, and pharmacies. This study focuses on a full spectrum of government funded programs by gathering data at hospital level. The purposes of this study are to investigate variations of prescriptions related to benefit plans and categorizes frequency and variety of prescriptions under government guidelines. This study aims to identify significant similarities and differences in prescribing medications and summarize cost of treatments in multiple benefit groups and summarize decision for prescriptions.

III. BACKGROUND

A. Health Benefit Plan in Thailand

Thailand has introduced Universal Coverage Scheme (UCS) in 2001 which provided access to healthcare to approximate 48 million or 75% of population [50], complementing the existed Civil Servant Medical Benefit Scheme (CSMBS) which includes approximately 5 million government and state enterprise employees, and Social Security Scheme (SSS) which is mandatory health insurance for 11 million employees in private sectors. Under an establishment of UCS, CSMBS, and SSS, 99.5% of Thai population has health protection coverage. While UCS and CSMBS are administered by the government and funded from tax revenue, SSS is funded by contributions from employee payrolls and partly through government budget. In 2017, Thai government spent 12.6% of healthcare GDP and the three programs covered 64 members in the systems. According to WHO [50], healthcare system of Thailand received financial flow from general tax, payroll contribution from SSS, and premiums from private insurance. A summary of health financial flow is shown in Fig. 1.

![Figure 1. Financial flow in Thailand’s health system (WHO, 2017).](image)

CSMBS allows three physicians to co-endorse use of drug outside the list if required for treatments. This regulation reveals differentiation of prescriptions to population of Thailand. A unique exception of CSMBS prescriptions is to allow physicians to choose the prescriptions under discretions of physicians. According to Suraratdecha et al. [51], benefits included in the three benefit schemes, the highest spending are found in the CSMBS which includes all treatments and interventions, all prescriptions within and beyond government suggested list. The lowest spending is UCS. SSS spending is found to be in between the other two programs [52], [53]. In Thailand, the hospitals collected information of treatments costs and patients diagnosis under subsidized coverage programs (Security Scheme (SSS), Civil Servant Medical Benefit Scheme (CSMBS),
Universal Coverage Scheme (UCS)), and two others (private insurance holder and out-of-pocket payers).

IV. DATA COLLECTION

The datasets were transferred from locally store database that includes 55,601,628 rows from 18 tables in Structured Query Language (SQL) format. The data were primarily stored separately for departmental use. Data conversion process included identification of Primary Key (PK) and Foreign Key (FK) in all tables to access information in diagnosis and health benefit details. The overview of data migrating process is summarized in Fig. 2. The script formats were transformed from SQL to JavaScript Object Notation (JSON). Data cleaning process includes dropping missing values and null values. Some translation from string scripts of prescription dosage and units are required. The cleaning process also includes separation of number and text units. Translation requires open-source software (e.g., MongoDB, Hortonworks, CouchDB, and Cloudera) to convert files in Comma-separated Values (.csv) format to JSON. Under high security mask to transform data into encrypted scripts, Cloud service of Amazon Web Service (AWS) is chosen for deploying data and managing this data ecosystem.

The electronic records were transferred from locally store database to an assigned Internet Protocol (IP) address. In data cleaning process, tables from patient identification number, the diagnosis codes, treatments and prescriptions, test results, and other related health information contain 62,050 patient records under three main government-subsidized coverage programs (Security Scheme (SSS), Civil Servant Medical Benefit Scheme (CSMBS), Universal Coverage Scheme (UCS)), and two others (private insurance holder and out-of-pocket payers). Relational database was created to cross-list chronic disease patients with multiple arrays of International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes for the selected chronic diseases (hypertension and diabetes). The final step was to generate a list of patients with the eligible benefit plans in numerical orders.

Figure 2. Structure of data process.

V. METHODOLOGY

This study extracted a data set of prescriptions and diagnosis codes. The dataset contains prescription decision on dosage, drug brand, and manufacturer identifier, diagnosis patterns in a structured ICD-10 format, reimbursement plans, admission date, patient demographics, cost and sale price of each prescription. The dataset contains 160 GB of 3-year EMRs that includes (1) Encrypted Hospital Identification Number (HN), date, health information of the patients (2) prescription data (including retrieved medicine orders, recommended usage, and dosage), and (3) Treatment details and charges. A summary of prescription frequency is analyzed using MATLAB software. MATLAB frequency plots provide a visualization of distributions for prescription frequency between diabetic and hypertensive patients among health programs (UCS, SSS, CSMBS).

A relational diagram provides linkage of patient identification to diagnosis and prescriptions. For observations in terms of prescriptions, 62,050 in-patient visits were extracted for all diagnosis codes. Out of 62,050 prescription records, a 12,886 prescription records were 553 patients with diabetes (E.08x - E11x) and 2,193 prescription records were issued for 176 patients diagnosed with essential primary hypertension (R73x, I.10x, I.11x, and E78x), 13 patients were shared diagnosis between health programs are displayed in two frequency plots. A summary of prescriptions to the three health benefit schemes will be segmented in percentages. The most frequent prescriptions will be extracted into a list from highest to lowest percentage of use.

VI. RESULTS

Prescriptions for patients with diabetes under UCS are clustered to the right-hand side of proportion of prescription (Shown in Fig. 3.). The first 100 most frequent items with percentage of prescription greater than 6% (listed on y-axis) are located on the right-hand side of the chart, and prescriptions under SSS and CSMBS are clusters within the first 200 items. In comparison to diabetes, prescription variations in hypertension patients in three schemes (CSMBS) distributed with greater range to the overall list of 597 prescribed items. High percentage of prescriptions was found in CSMBS than other schemes with average percentage of distribution of 0.29% compared to 0.25% in SSS, and 0.24% in UCS. High percentage of prescribed items were found for hypertension patients under SSS which in the last 200 items on the list are also clustered onto the right side of Fig. 3. The distributions show a long-tailed distribution of repeating prescriptions in diabetic patients with CSMBS with nearly 0% of prescription. Most of the prescriptions was being selected with less than 1% of total options in the prescription list. Among 596 prescriptions, 467 unique items were prescribed 20 times or less out of 12,886 total prescriptions (0.15%). Out of 596 brands, only 29 items from the list were uniquely prescribed to either diabetes or hypertension patients, the other 569 prescriptions were selected in treatments of both diabetes and hypertension patients under UCS, SSS, or CSMBS. Although
variations of brand selection co-existed between the two treatments, frequencies of overall prescriptions are found to be dispersed. The distributions of variation among three benefit plans (UCS, SSS, CSBMS) are shown in Fig. 3.

The dataset contains prescription decision on dosage, drug brands, and manufacturer identifiers, reimbursement plans, and charges for in-patient diabetes and hypertension prescriptions. The most frequent prescriptions were found to be 47 repeats for Norvasc (a long-acting calcium channel blocker) in patients with hypertension, and 167 prescription orders for Zorcor (high cholesterols treatment) in diabetes patients under UCS. These two prescriptions also shared most frequent prescribed drugs as for all Hypertension and Diabetes patients. Top ten prescriptions on each diagnosis are listed in Table II. The most frequent uses for Hypertension are Norvasc, Sodium Chloride 5ML (Saline), Paracetamol. Top Diabetes prescriptions are Saline, Zocor, Mulcinil, and Both diagnoses shared same similarity, for example, Olmetec (high blood pressure), Paracetamol (pain relief), and Saline was listed with frequent usage in both groups.

Table II provides a summary of prescriptions to the three schemes. A total of 1,399 diabetes and hypertension prescription list under government guideline, 404 brands that can be prescribed to UCS patients, or equivalent to 28%; 497 brands are suggested prescriptions for SSS, or 35%, and 498 brands (or 35%) on the list are applicable for CSBMS patients. This suggested the guideline provides similar list of prescriptions in three programs (or CSBMS = SSS = UCS). Average cost of prescription in each group was calculated by cost of prescriptions multiply by identification. The outcome suggested that the average costs of treatment are similar ($15.64, $16.58, and $15.32 for UCS, SSS, and CSBMS) among the three health benefit programs. By comparing number listed medications under government guideline to the actual number of prescriptions, total percentage of prescriptions prescribed to the UCS was highest (5,825 items, equivalent to 45%), while prescriptions prescribed to SSS and CSBMS were 37% and 18% accordingly (or CSBMS < SSS < UCS).

VII. CONCLUSION AND RECOMMENDATIONS

This study investigates an outcome of health policy reform in Thailand in terms of treatments and prescription plans. The results summarized percentages and lists of prescriptions that patients received from different health benefit programs. By using local data sources, this study showcases patterns in Electronic Medical Records (EMRs) in terms of prescriptions and treatment cost from in-patient data with identification diagnoses for hypertension and diabetes. According to this investigation, fundamental similarities in prescriptions for diabetes and hypertension are found in the dataset, however, pre-assigned benefit schemes (such as UCS, SSS, and CSBMS) are different. Prescriptions among groups related to CSBMS are hypothesized to be higher than SSS and UCS (CSBMS > SSS = UCS), this study found that prescription costs were lowest for UCS, moderate for SSS, and highest in CSBMS patients accordingly. The results also suggested high variations of treatments in patients with hypertension, but very limited variations among diabetes patients. As variety of prescriptions are similar in the three programs, the amount of prescriptions was significantly different. Frequency of proportion of prescriptions assigned to each group of patients indicated decision patterns, and
variations among patient groups, and suggested average cost of treatments.

The outcome of this study should be beneficial to all stakeholders in healthcare industry, especially those in Thailand. By utilizing electronic records, this study depicts emerging issues of treatment and prescription equity by comparing the outcomes of cost and treatment procedures from major government administered health benefit plans in Thailand. Patients with chronic disease should be able to compare the cost-benefit from multiple health programs, to know their rights, and assess their expected future reimbursed costs and outcomes. This study aims to build an awareness to stakeholders such as physicians, pharmacists, and healthcare institutions, in determining the process of treatment and its influences inherited by characteristics of each program. The implications of this electronic data investigation method can be applied to monitor the performance of government administered medical procedures and prescription programs, and to provide better understanding in inherited mechanisms of the healthcare system for other countries.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

In regards to this paper contribution, Dr. Tansitpong was responsible for theory building, writing, and data analysis; Dr. Chaovalinwongse and Dr. Hoonlor were responsible for supervising and editing; all authors had approved the final version.

ACKNOWLEDGMENT

The authors would like to thank Dr. Pongtorn Kietdumrongwong and Mr. Tananat Kaesaprasit for their contribution in facilitating with medical data transfer, extraction, and data collection. Funding This work was supported by Mahidol University International College, [O10/2015]. Ethics approval This project has been approved by The Committee for Research (Social Sciences) of Mahidol University with Ethics certificate of exemption no. 2015/033.1108, MU-SSIRB no. 2015/314 (B2).

REFERENCES


Prawpan Tansitpong is an Lecturer of Operations Management at NIDA Business School in Bangkok, Thailand. She also serves as a Director of the Accelerated MBA Program. She received a Ph.D. in Operations Management from Lally School of Management and Technology, Rensselaer Polytechnic Institute, a Master of Science in Innovation Management and Entrepreneurship from Brown University, and a bachelor degree in Engineering from Chulalongkorn University in Thailand. She is a former US-ASEAN Fulbright Visiting Scholar at the School of Medicine, University of Washington. Her research interests are in the areas of healthcare operations, health analytics, and product and service design in operations/marketing interface.

Wanpracha Chaovilaitwongse is a Professor of Industrial Engineering at Mahidol University. He received a Ph. D. in Computer Science from Rensselaer Polytechnic Institute. His current research interests are in data mining (graph mining and text mining), data science, decision support system, and AI applications. He also works with several organizations from both the public and private sectors to design and create their data analytics pipeline and/or platform.

Apirak Hoonlor is a lecturer at the Faculty of Information and Communication Technology, Mahidol University. He received a Ph. D. in Computer Science from Rensselaer Polytechnic Institute. His current research interests are in data mining (graph mining and text mining), data science, decision support system, and AI applications. He also works with several organizations from both the public and private sectors to design and create their data analytics pipeline and/or platform.