An Innovative Decision Support Service for Improving Pharmaceutical Acquisition Capabilities

Phattharat Songthung, Kunwadee Sripanidkulchai
National Electronics and Computer Technology Center
Pathumthani, Thailand

Rungpetch C. Sakulbumrungsil
Chulalongkorn, University
Bangkok, Thailand

Nusaraporn Kessomboon
Khon Kaen University
Khonkaen, Thailand

Paithip Luangruangrong
DMSIC, Ministry of Public Health
Bangkok, Thailand

Siripa Udomaksorn
Prince of Songkla University
Songkla, Thailand

Inthira Kanchanaphibool
Silpakorn University
Nakhonpathom, Thailand

Abstract—The cost of pharmaceutical products is one of the largest contributors of operating costs in providing healthcare services in Thailand. As drug prices change according to shifting market forces, the distribution of purchase prices for each drug varies according to the type of medication and the purchasing power of the healthcare provider. These changes can have significant impact on how well healthcare providers can effectively provide services. PAC-DSS (Pharmaceutical Acquisition Capability Decision Support Service) is an innovative decision support service that enables hospitals to pool together and share information to better understand current market prices and run analytics on drug prices in order to improve their operating costs. Our service allows users to interact with real pricing data to dissect various factors that can contribute to acquisition capabilities of individual drugs such as specific brand names or groups of drugs such as all brands of a given generic drug, without sacrificing individual providers' privacy as we do not disclose individual purchase prices. In developing PAC-DSS, we have had to address a range of technical challenges such as data privacy and alignment of disparate drug ontologies. In this paper, we describe PAC-DSS's service architecture, analytic services, and the benefit of PAC-DSS on improving healthcare services by lowering operating cost without sacrificing service quality. We also discuss the initial benefits from deployment of the service currently hosted by the Ministry of Public Health.

Keywords—decision support system; innovative service; healthcare IT

I. INTRODUCTION

In many countries including Thailand, the rising cost of healthcare is a national concern that impacts all citizens and their respective governments. The primary provider of healthcare services in Thailand is the government with more than 1,000 hospitals throughout the country. Previous work has shown that the majority of cost for providing services in these hospitals is the cost of pharmaceutical products, and half of the total pharmaceutical product consumption in the country in 2001 was through these hospitals [1]. Thus, being able to effectively manage the cost of pharmaceutical products is a highly effective mechanism to control the total cost of providing healthcare services in order to continue to make services accessible to all citizens.

There have been several approaches to help government healthcare providers manage the cost of pharmaceutical products in Thailand. For example, the Ministry of Public Health (MoPH) has an effort to increase the efficiency of procurement through the Drug and Medical Supply Information Center (DMSIC) whose role is to gather information about drugs purchased by hospitals operated by the MoPH. The approach taken by DMSIC is to identify and publish suitable reference purchase prices that hospitals can use as a ceiling. However, even with established reference prices, different hospitals still purchase at different prices even when buying the exact same product from the same manufacturer and distributor. The variation in price can be caused by a number of factors such as volume discounts, bargaining power, differences in distribution costs, etc.

In this paper, we develop a decision support service called PAC-DSS (Pharmaceutical Acquisition Capability Decision Support Services) that healthcare providers can use to better understand their own bargaining power in order to negotiate for appropriate prices. Whether providers negotiate with suppliers individually, or pool together with other providers to negotiate for group discounts, they can leverage the data and analytics provided by PAC-DSS to reason about the behavior of the market for individual drugs...
at the granularity of generics and brands in order to improve and control cost.

The main idea behind PAC-DSS is to enable healthcare service providers to share and run analytics on pricing data that is commonly considered sensitive and private without needing to disclose their prices to other healthcare providers, thus keeping each provider's data private. Users run analytics in real-time by decomposing various key factors that contribute to overall drug prices such as size and type of hospital, type of drug (imported vs. domestic), etc. Based on the results of these analytics, healthcare service providers can better determine the target prices for each drug and develop an acquisition plan that results in lowering their overall operating cost. In order to support these analytics, we use domain-specific classification of pharmaceutical products developed by aligning existing ontologies in order to support products that are distributed in Thailand. Our work extends existing national and international drug ontologies.

This paper is organized as follows. In Section II, we provide a brief overview of PAC-DSS. Section III discusses the background required to understand our work. In Sections IV and V, we present the implementation of PAC-DSS and the initial benefits from piloting PAC-DSS with more than fifty healthcare providers. Section VI summarizes our work.

II. PAC-DSS OVERVIEW

In this section, we provide an overview of data collection and analytics in PAC-DSS.

A. Data collection

The current practice in drug price data collection is that government hospitals submit drug purchase data on a voluntary basis to DMSIC under the Ministry of Public Health. Each hospital individually selects the set of drugs they wish to report and the frequency of the reports such as daily, monthly or annually, as depicted in the top half of Figure 1. DMSIC then sifts through this massive amount of data (there are more than 40,000 distinct drugs in the national market, more than 1,000 hospitals) to perform data cleansing and traditional statistical analysis (mean, median, etc.). The excruciating data cleansing process is performed by hand and is very time-consuming. In Section IV-C, we discuss about our future work to support automated data cleansing. The most requested information from this data prior to PAC-DSS is the minimum, average, and maximum price and quantity of each drug in order to help hospitals make pricing decisions for their next orders.

B. Data analytics

PAC-DSS improves upon DMSIC's statistical calculations by providing a richer, more interactive way of analyzing the exact same data reported into DMSIC. PAC-DSS is implemented as a Web-based application that hospitals can log in to via the Internet and select various parameters to run analytics. For example, users can select the drugs they are interested in and who (which group of drugs/providers) they want to compare their acquisition capabilities against as depicted in the bottom half of Figure 1.
• Compare my capability for the given drug group with hospitals of similar size (and demand),
• Compare my capability for the given drug group with all hospitals.

We call the selected drug grouping and hospital grouping the selected “drug market”.

PAC-DSS will return analytic results in useful presentation formats such as graphs and summary tables. Because pricing behavior changes extremely quickly due to market forces, the analytics is performed in real-time based on the latest reported data from hospitals. The results of the analytics enables hospitals to make a decision on the target purchase price for each drug based on the pricing behavior of similar drugs and other hospitals. In addition, the MoPH can also use this information to increase operational efficiency, create policies for acquisition based on market prices, and implement medium-term and long-term policies that will lead to reasonable price control on drugs in the market.

III. BACKGROUND

In this section, we present a review of the background work upon which PAC-DSS is built. There are three concepts we leverage: Decision Support System (DSS), Pharmaceutical Acquisition Capabilities Index (PAC Index), and the Gini Coefficient which is often used as an inequality index. We will briefly describe each concept next.

A. Decision Support System (DSS)

A DSS is an interactive computer-based system that organizes concepts and data to help users find answers in a simple and rapid way. The data in a DSS may be in the form of semi-structured or unstructured. Often, a DSS is not used for the purpose of making a final decision but for the purpose of presenting important and interesting information to support the users decision-making process. Users must make decisions using their own reason, experience, wisdom, and creativity. The quality of data used in a DSS is of critical importance. Low quality data will likely be insufficient for making decisions or may lead to poor decisions. Therefore, data cleansing is required prior to use.

PAC-DSS is built on the DSS concept as we do not provide final decisions about target drug prices but rather data to support decision making. The presentation of data which we will discuss in Section IV is based on line graphs and other statistical means such as box charts.

B. Pharmaceutical Acquisition Capability Index (PAC Index)

PAC Index [2] is an analytical tool we use in PAC-DSS that provides meaningful comparisons between hospitals without needing to disclose individual hospitals' purchase prices to other hospitals which may violate privacy of purchase price information. By using the PAC Index, we can measure the ability of purchasing power of buyers for a particular drug. PAC transforms absolute price and purchased quantity information into a relative combined price and quantity index as described below:

\[
PAC_i = \frac{-\ln(P_i / P_{\text{min}})}{(Q_i / Q_{\text{min}})}, \quad \text{for every } P_i > P_{\text{min}}
\]

Where \(PAC_i\) = Pharmaceutical Acquisition Capability (PAC) of buyer \(i^{th}\) who bought the product

\[
P_i = \text{Price of buyer } i^{th} \text{ who bought the product}
\]

\[
P_{\text{max}} = \max\{P_1, P_2, P_3, \ldots, P_n\}
\]

\[
P_{\text{min}} = \min\{P_1, P_2, P_3, \ldots, P_n\}
\]

\[
Q_i = \text{Quantity of buyer } i^{th} \text{ who bought the product}
\]

\[
PAC_{\text{max}} = \max\{PAC_1, PAC_2, PAC_3, \ldots PAC_n\}
\]

If prices vary linearly with purchased quantity, then the value of PAC is a function of \(\ln(x)/x\). Otherwise, there are other factors that affect the behavior of the price difference such as individual hospitals' bargaining power. By using only PAC in the analytics, users will not know about other hospitals' purchase prices. A hospital with a high PAC for a given drug market has high bargaining power for that drug and vice versa. For example, a high PAC value can occur when a hospital pays a very low price for a small purchase quantity relative to other hospitals. We allow a hospital to see its own PAC for each of its purchased drugs but do not disclose other hospitals' PACs in PAC-DSS. Instead, we use the average PAC to represent the distribution of purchasing capabilities for the given drug market. This provides purchase price confidentiality.

C. The Inequality Index Gini Coefficient

An additional analytics tool that we use to characterize the market price of drugs is the Gini Coefficient [3] which is commonly used to measure the inequality of income or wealth. PAC-DSS applies the Gini Coefficient to capture the inequality of PAC. There are three general cases to interpret the value of the Gini Coefficient: a value close to 1, a value close to 0, and a value higher than 0.5. In case the Gini Coefficient has a value close to 1, most products on the market were purchased at a price higher than should be. A Gini Coefficient with a value close to 0 means that most products on the market were purchased at a similar price. A Gini Coefficient higher than 0.5 indicates high inequality.

\[
\text{Figure 2. Gini Coefficient}
\]
Figure 2 illustrates how the Gini Coefficient is calculated as the ratio of the area under the Cumulative Distribution Function (CDF) of the value to be measured (in this case the value of PAC) and the area under the Line of Equality or the CDF of the Uniform Distribution using the following formula.

\[ G = \frac{A}{A+B} \]

Where \( G \) = Gini Coefficient

\( A \) = Area under the graph between Line of Equality and CDF of the value to be measured (value of PAC)

\( B \) = Area under the graph CDF of the value to be measured (value of PAC)

In Section V, we report the results from using the Gini Coefficient to characterize the pricing behavior of different drugs.

IV. IMPLEMENTATION

PAC-DSS has three core components: the client-side user interface, server-side data services, and raw data transformation services as depicted in Figure 3. PAC-DSS is a web-based application designed and developed on top of Joomla, a well-known Content Management System (CMS). We add additional content as PHP pages and scripts to process data for PAC-DSS. The server-side data services consist of a MySQL database serving as a repository for raw price reports and for the CMS content. The raw data transformation services is by far the most complex part of the system that implements the interactive data analytics bridging between the server-side data services and the client-side UI. The analytics and the graphs are implemented using PHP, AJAX, Dojo and Google Charts. We first present the data analytics and the corresponding UI in PAC-DSS, then challenges and solutions to support grouping of drugs for the analytics, and lastly data cleansing challenges.

A. User Interface and Analytics

The PAC-DSS UI has two main subsystems “Graph” and “DSS” as shown in Figure 4. The PAC, Gini Coefficient and other statistical computations discussed in Section III are part of the “Graph” menu depicted on the top of Figure 4. The decision support services to help users decide on target purchase prices and quantities are under the “DSS” menu depicted on the bottom of Figure 4.

Users selecting either of these menus will be presented with an option to select a subset of data from the PAC-DSS repository for the analytics such as the group of drug that is of interest (i.e., a brand name and strength), the group of hospitals to compare against, and the time frame (i.e., in the last 12 months) which we call the selected “drug market”.

For the selected drug market, if the “Graph” menu is selected, the user will have a choice of two types of analytics as depicted at the top of Figure 5. On the top left hand corner is the PAC line graph that shows the calculated PAC for the hospital and the representative PAC for the selected subset of data. Interpreting this graph is rather straightforward – if the “x-mark” (the hospital's PAC) is above the average PAC curve, the hospital’s price is higher than it ought to for the given drug. If the “x-mark” is below the average PAC curve, the hospital has a better/higher PAC than the rest of the hospitals and is already purchasing at
Figure 5. PAC-DSS Analytics
a competitive price. The chart on the top right hand corner is the statistical distribution shown as Box plots of prices, quantities purchased and PAC for the selected drug. Gini values for the selected drug market are listed in the table under the charts.

If the “DSS” menu is selected, the user will have a choice of two types of analytics as depicted at the bottom of Figure 5. On the left hand-side, users can input combinations of prices and quantities to purchase to see how well those combinations fair against the market. If the “x mark” (based on the user’s inputted values) is below the average PAC curve, then that combination is considered to be doing better than the average in the market. On the right-hand side of Figure 5, users can input the total quantity they wish to purchase and the system will return a recommendation for an appropriate price. The price will fall on the curve which represents the average pricing behavior in the market. Cost savings that may result if hospitals purchase based on these prices instead of their current purchase prices are also displayed.

B. Grouping Drugs for the Analytics

PAC-DSS supports analytics based on grouping of drugs that serve similar purposes such as drugs that are “painkillers” using the semantic grouping and classification information from the ontologies we discuss next. In short, we would like to group Thai drugs in TDC based on the groupings defined by the World Health Organization’s ATC classification but there is no direct way to map between the two ontologies. They use completely different conventions. The only possible linkage is the chemical substance names in ATC and the generic drug names in TDC which sometimes may be the same, but sometimes may not be as they are named differently. To solve this problem, we align and match these two disparate drug ontologies, the international Anatomical Therapeutic Classification (ATC) ontology and the national Thai Drug Code (TDC) ontology using two helper ontologies.

![Anatomical Main Group “A”](image)

Anatomical Therapeutic Classification (ATC)

Drugs that serve similar purposes can be grouped based on the World Health Organization’s (WHO) Anatomical Therapeutic Classification (ATC) [4]. The classification of drugs is based on the groupings of active substances in the drugs, the 14 organs or systems on which they act and their 5 levels of therapeutic, pharmacological and chemical properties, as depicted in Figure 6. The fifth level in the ATC is generally comparable to the generic name of the drug. There are approximately 5,000 generic names of drugs in the fifth level of ATC that we would like to align with the drugs in Thailand.

Thai Drug Code (TDC)

The Ministry of Public Health has developed a standard 24-digit code TDC to identify pharmaceutical products used in Thailand [5]. It is a reference code that is used in all Electronic Health Records nation-wide in order to refer to the same specific drug down to the level of more than 40,000 brand names, dosage forms and strengths. The definition of the 24 digits is described in Table 1.

<table>
<thead>
<tr>
<th>Code Description</th>
<th>Digit</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Types of medicines</td>
<td>1</td>
<td>1 – single-ingredient drug 2 – in combination preparations drug 3 – medicines manufactured by each hospital 4 – herbal medicines 5.9 – reserve for future use</td>
</tr>
<tr>
<td>The drug</td>
<td>10</td>
<td>The single-ingredient drug according to Iowa Food and Drug Administration, there are 16-digit but we discarded the first 6-digit and used only the last 10-digit by separated into three groups as follow: 7-11th digit 12-14th digit 15-16th digit</td>
</tr>
<tr>
<td>Main ingredient drug</td>
<td>Sodium Chloride</td>
<td>Synonym</td>
</tr>
<tr>
<td>Main ingredient drug</td>
<td>Sodium Chloride</td>
<td>Synonym</td>
</tr>
</tbody>
</table>

In combination preparations drug according to ATC system is used by Food and Drug Administration, there is 5-digit composed of English alphabets with the numbers. They adapted this ATC code to be 7-digit number follow by 3-digit sequence number for all drugs in the same ATC group as shown in table below.

<table>
<thead>
<tr>
<th>7-digit</th>
<th>3-digit</th>
</tr>
</thead>
<tbody>
<tr>
<td>The transformation of ATC group code</td>
<td>The sequence number for group</td>
</tr>
<tr>
<td>The strength of drug</td>
<td>5</td>
</tr>
<tr>
<td>The form of drug</td>
<td>3</td>
</tr>
<tr>
<td>Manufacturer/ Importer</td>
<td>5</td>
</tr>
</tbody>
</table>

(a) Types of medicines (1-digit)
(b) The drug (10-digit)
(c) The strength of drug (5-digit)
(d) The form of drug (3-digit)
(e) Manufacturer/ Importer (5-digit)

Alignment of ATC and TDC

Automated ontology mapping is a research challenge [6, 7]. The purpose of the mapping is to determine the relationship between concepts in two or more different ontologies. TDC is specific to Thailand and encapsulates more complex information than the 7-digit ATC code. In order to use the drug classification from ATC with drugs in the Thai market, we need to align the two ontologies. The use of this alignment in PAC-DSS is to better support decision-making by allowing users to compare drugs that serve similar purposes such as comparisons against all drugs.
that treat diabetes which is a much larger grouping than based on generics or brand names alone.

Figure 7. Ontology for mapping ATC ontology and TDC ontology in PAC-DSS system

Figure 7 depicts our alignment approach. We take the two existing ontologies ATC and TDC, and use a third ontology RxNorm [8] from the United States and our own “Dosage Form Ontology” to map between ATC and TDC. The resulting alignment is used in PAC-DSS and is called the PAC-DSS’s ATC-TDC Mapping Ontology. We use semi-automatic mapping heuristics to link ATC and TDC ontologies together. We estimate that our approach saves more than 1000 hours compared to hand-created linkages.

The approach is described in Figure 8 in pseudo-code. We analyze each generic name to identify the same drug using 3 heuristics.

1. Exact match: If the generic names/chemical substances and the dosage forms are the same between ATC and TDC, it is considered to match (labeled A in Figure 8).

2. Semantic-based heuristics (labeled B): a “synonym” of a drug which is often a name of chemicals or main elements derived from information from ontology of drugs in the public domain such as RxNorm [8].

3. Name-based heuristics (labeled C): calculate the similarity of drug names using Levenshtein Distance [9] between string 1 and string 2 which counts the number of characters replaced, inserted, or deleted in order to transform string 1 into string 2. The smaller the distance of the two strings, the closer those two strings are to each other.

The result of using these heuristics to align TDC and ATC is depicted in Figure 9. Overall, we were able to match approximately 96% of drugs in TDC to ATC. Going clockwise, of the 96% matched, 72% was matched based on exact names, 10.7% was matched via semantic-based mapping, and 11.1% was matched using the minimum string distance of drug name. Around 6% of drugs in TDC did not match. One important factor that contributes to this 6% is that some drugs used in Thailand such as herbal remedies are not used abroad, thus there are no ATC codes that could possibly match.

C. Data cleansing

There are many factors that cause poor quality data. We have found that for the most part, human error played a significant role in all data that needed cleansing. For example, typos of drug names, 24-digit codes, dosage forms, strengths, and units were commonly misspelled. In addition, the mixed use of Thai and English spellings further
obfuscates the data. Secondly process problems also existed. Hospitals sometimes do not maintain the use of 24-digit codes. For example, a hospital may have recently changed suppliers resulting in a different brand name for the same generic drug. However, they continue to use the 24-digit code for the previous brand instead of updating the 24-digit code to the current latest brand they use. Reporting of such codes into DMSIC presents a problem because DMSIC cannot know that the drug was changed at the hospital and will perform analytics on wrong data. Third, we have deployed a best practice exclusion criteria for ignoring certain reports from our analysis. For example, for privacy reasons and to support reasonable interpretation of Gini Coefficients, any drug that was purchased by fewer than four hospitals are not included in the analytics. In addition, any drug that was reported to be purchased only once by a hospital within a year is considered incomplete data because it is rarely the case that a drug is only purchased once a year.

Automated data cleansing is a challenging problem that we plan to address as part of future work. We are exploring data mining techniques to help differentiate inconsistent and poor quality data for cleansing. We believe that we can reduce the overwhelming amount of time currently spent on manual cleansing.

Figure 9. Results from mapping ATC and TDC Ontologies

V. DEPLOYMENT RESULTS

PAC-DSS is currently hosted by the Ministry of Public Health in Thailand and made available to all healthcare service providers that choose to contribute data to utilize the service. More than fifty hospitals have used the system to run analytics on the data.

PAC-DSS provides information about potential cost savings as one of the major advantage of our system. Users can use PAC-DSS to make the right decision to negotiate with vendors. Figure 10 above shows the percentage of all hospitals with cost savings if hospitals were to purchase generic drug according to our recommended price. The y-axis is the percentage of cost savings and the x-axis is experiment number sorted by percentage of savings. An experiment is a particular combination of generic drug and strength across all hospitals which we call a drug market. As you can see from the figure, there are 25 experiments out of 31 drugs that if using our system to set the target price could result in cost savings. The drug with the maximum cost savings is results in a 32% discount. On the left side of the chart, experiment numbers 1-6 have no cost saving which means that presently hospitals are already purchasing at or less than the recommended price. These potential cost savings can help users decide how best to approach their next procurement cycle.

Figure 10. Cost Savings of Generic Drugs for all Hospitals

Figure 11 shows a bar chart of the Gini Coefficient of generic drugs for all hospitals in PAC-DSS. The y-axis is the Gini Coefficient and the x-axis is experiment number where an experiment is a particular combination of generic drug and strength across all hospitals which we call a drug market. There are a total of 31 experiments depicted in Figure 11. Using a threshold of 0.5 (horizontal line) where the Gini Coefficient is less than 0.5 and Gini Coefficient is more than 0.5, we see that we can segregate the drugs into two groups. Drugs 1-8 have a Gini Coefficient of less than 0.5 meaning that the market has little inequality. However, drugs 9-31 have a Gini Coefficient larger than 0.5 which indicates that hospitals purchasing these drugs are doing so
with disparity. Some are purchasing at much lower capabilities than others. The Gini Coefficient is useful for indicating the market behavior of each drug.

VI. SUMMARY AND FUTURE WORK

We developed and deployed PAC-DSS as a decision support system for hospitals to determine their target purchase price and cost savings for drugs. PAC-DSS ensures that actual prices and quantities of drugs purchased by individual providers are not shared or disclosed to other healthcare providers. Instead, we use capability indices to measure each provider's capabilities for a given drug or class of drugs. The result is useful and intuitive for cross-provider comparisons but powerfully abstracts away the individual's price and purchased quantity information so privacy concerns are minimum. PAC-DSS has three key features. First, the service supports useful analytics of drug cost data. Second, our PAC-DSS service allows users to run analytics in real-time by decomposing various key factors that contribute to overall drug prices such as size and type of hospital, type of drug (imported vs. domestic), etc. Users can run their own custom analytics in an interactive manner. Based on the results of these analytics, healthcare service providers can better determine the target prices for each drug and develop an acquisition plan that results in lowering their overall operating cost. In order to support these analytics, we use domain-specific classification of pharmaceutical products based on an alignment of ontologies that we have built to support classification of products that are distributed in Thailand that extend existing national and international drug ontologies. Third, we have successfully convinced more than fifty healthcare providers to participate in using our service to increase their acquisition capabilities. The service is currently hosted by the Ministry of Public Health in Thailand and made available to all healthcare service providers that choose to contribute data to utilize the service. The benefit of using PAC-DSS results in improvements in healthcare services by lowering operating cost without sacrificing service quality.

As future work, we will continue to improve upon our research in ontology alignment for Thai Drug Code with other codes. The resulting ontology can be leveraged for other uses. Another area for improvement is to explore different reference PAC values for a given drug market that does not necessarily need to be based on the average PAC such as the median PAC, 60th-percentile PAC, 75th-percentile PAC, etc. Users can adjust and interact with different reference PAC values for their custom planning and decision-support. In addition, time sensitivity of data is another interesting problem since prices of different drugs fluctuate differently, some drugs may need to be analyzed over the course of a year while other drugs may need to be analyzed at shorter time scales. Lastly and most importantly, we are looking at leveraging classic data mining techniques such as clustering to help us classify drug markets that have similar behavior, and text mining techniques to enable automated data cleansing to solve the problem of typos in drug names both in Thai and English. For example the use case we commonly faced in our data set was the spelling of “PARACETAMOL”, which varied significantly across hospitals and very difficult to identify all semantically similar spellings: “Paracetamol”, “para”, “พาราเซตามอล”-in Thai, “พารา”-in Thai, etc. All these future works will be useful not only for PAC-DSS system, they can be used to improve the other systems as well.

REFERENCES