

# Profiling Patterns in Healthcare System: A Preliminary Study

Nicholas Khin-Whai Chan<sup>1</sup>, Angela Siew-Hoong Lee<sup>2</sup>  
Department of Computing and Information Systems  
Sunway University, Kuala Lumpur  
Malaysia

Zuraini Zainol<sup>3</sup>  
Department of Computer Science  
Universiti Pertahanan Nasional Malaysia  
Sungai Besi Camp 57000 Kuala Lumpur, Malaysia

**Abstract**—In the 21<sup>st</sup> century, our planet revolves around data and is known as a digital earth. The astonishing growth in data has resulted in an increase in interest of Big Data Analytics to capture, store, process, analyze and visualize unprecedented amount of information. Big data has undoubtedly and will continue to shape modern information driven society where behind all the available data, there is a hidden potential to discover meaningful insights and patterns which may impact businesses in unexpected measures. The exponential growth of data is also present in the healthcare sector. In Malaysia, most employees are provided with medical benefits which includes general medical costs to hospitalization benefits and insurance coverages. With the healthcare data and information stored with the Human Resource (HR), employers could potentially analyze and identify patterns in the historical medical claims which could then help in making specific decisions to understand their employee population health and the usage of the premium coverage. Therefore, the aim of this research is to better understand the patterns presented in the employees' healthcare data. Through the analysis and understanding of the patterns in past medical claim history, potential strategies can be proposed to allow employers to provide proactive and reactive measures to potentially help sustain medical expenditure.

**Keywords**—Big data analytics; data mining; descriptive analysis; healthcare; pattern profiling

## I. INTRODUCTION

Today, in the 21<sup>st</sup> century, our planet is considered as a digital earth [1]. Big Data would be the biggest revolution in the 21<sup>st</sup> century [2]. The massive growth of data in the world of Big Data Analytics has triggered every industry to try and understand the Big Data revolution [3]. This exponential growth of data is also presented in the healthcare sector [4]. More importantly, there are potential in using Big Data to support healthcare in medical including clinical decision support, disease prediction and population health [4]. Information, knowledge and data in healthcare is growing continuously on a daily basis [5].

In Malaysia, employers usually offer their staff health benefits [6]. This includes general medical costs to hospitalization benefits and surgical insurance coverage [6]. With the healthcare data and information stored with the HR (Human Resource), employers could potentially analyze and identify patterns in the historical medical claims which could then assist them in making specific decisions such as to understand their employee population health and the usage of

the premium coverage. Employers will then be able to better understand the population health of their employees through analysis, and potentially prepare proactive measures instead of reactive measures. Moreover, employers are consistently playing premium coverage to provide employee with the medical benefits [7]. Unfortunately, due to the ever-increasing medical costs, employers are facing a major issue where more resources are needed to sustain the premium coverage [7]. What is more crucial is that a large part of an employee's satisfaction is influenced by the benefits provided by the employer [8]. Therefore, how do employers potentially reduce the premium coverage in medical benefits? In Malaysia, there have been minimal research performed in this area of interest, hence, it could provide a breakthrough in providing crucial insights which could benefit the employers.

## II. LITERATURE REVIEW

Big Data has shaped Modern Technology [3] and is considered as a major revolution in the 21<sup>st</sup> century and the scale of growth is happening in a tremendous rate while the changes in modern technology is dramatic [2]. The intensity and speed of growth has triggered every industry to discuss about the evolution of Big Data [3]. Behind all the data generated, there is a hidden potential to be collect, share, process and analyze varied data [3]. The impact of such insights and hidden patterns could have major capability in enhancing processes, making business operations more efficient, creating more strategic business opportunities while reducing risk and resources [3]. Big Data is usually categorized by 3V's, however, as time passes on, more variations to describe the characteristics of Big Data began to surface as more industries began to explore the potential. The 3V's describing Big Data are Volume, Variety and Velocity [9]. Volume describes the ginormous amount of data while Variety describes the varying types of data generated (structured, unstructured or semi-structured) and the last V would be Velocity which describes the varying speed at which data is generated and processed (real-time, batch, periodic) [9].

Healthcare has become a thriving sector in many developed and developing countries [10]. Because of this growth, there comes other difficulties which follows along such as rising healthcare and medical costs, inefficiencies, poor management and quality and even an increase in complexity [10]. Hence, this leads to thoughts being put into making better decisions based on the available data and information that could potentially mitigate such difficulties

and challenges [10]. Big Data Analytics has opened new opportunities to improve service delivery, reduce cost and solve problems in the healthcare industry and enable timely decision making [11] [12]. Data in the healthcare industry is not just growing immensely because of the sheer volume of healthcare data but the diversity as well as the need for managing the data at varying speeds [13]. Most healthcare related personnel have recognized and understand the importance of Big Data and how it opens the door to new possibilities through the development of predictive models, pattern discovery, potential reduction in cost as well as to improve services, real-time analysis and decision making [11]. The increasingly promising outlook of how Big Data Analytics could shape the healthcare industry has led to an increase in interest from both academic and professional forces [12]. Hence, with the emergence and continuous growth in popularity in Big Data Analytics, the healthcare industry should leverage on the potential of Big Data technology [12].

Data analytics is a process of extracting knowledge from data and explored to identify meaningful insights and to obtain answers to specific questions. On one hand, data analytics is related to business intelligence and business analytics while Data Mining is related more towards a science and mathematical approach [14]. In data analytics, it can be segregated into three categories which include Descriptive analytics, Predictive analytics and Prescriptive analytics - the most common of the three are Descriptive and Predictive analytics [15]. Descriptive analytics aims to provide answer to the "What has happened?" question, which uses information from the past to easily explain and present those information using data visualization such as bar charts, pie charts, line graphs, etc. to provide an insight to those data [15]. Predictive analytics aims to provide answer to the "What can be predicted?" question, where the available and current data is used to define what to be expected in the future outcomes - predictive analytics usually uses mathematical methods and algorithms to discover relationships, patterns and insights in data which can be impactful for an organization [15]. Predictive analysis usually uses Data Mining techniques to obtain the expected outcomes. Prescriptive analysis aims to answer the "What should I do and What action should be taken?" question - with regards to the data and outcomes obtained, this would be to identify the opportunities or most viable solution to solve existing issues and problems; meaning, prescriptive analytics would equip organizations with the tools to achieve their objectives [15].

Medical expenditure will continue to rise in a tremendous rate and it has become a cause of concern for employers who are consistently paying a premium for the employee medical insurance. That is why the need to explore and identify the current employee medical claim trend to better understand and to propose potential recommendations to employers to help sustain the premium insurance.

However, employers are experiencing a major issue in health and benefits programme as medical costs is continuously increasing [16]. Medical costs will continuously increase in a relatively flat rate in 2019 [16]. The growth is

between the estimated trend of 5.5% to 7% over the past five years - the expected growth in 2019 is 6% which is a welcome change from the double-digit spikes in the 2000s [16]. But according to PricewaterhouseCoopers (PwC) the higher costs have not improved in terms of gains in consumer health and productivity [16]. Employers are facing a steep increase in medical costs and the benefits. As mentioned, for example, the costs of consultations with General Practitioners (GP) consistent increase in medical costs [17].

### III. PROBLEM STATEMENT AND PROJECT OBJECTIVES

Companies and organizations are paying a premium for healthcare benefits for their employees. And because of this, the cost has been consistently increasing every year as medical costs are ever increasing. Hence, companies and organizations are spending more resources on these premium healthcare coverages. More importantly with companies and organizations paying the premium healthcare coverage, is the healthcare benefit fully maximized or is it going to waste? Also, companies are unaware of the current employee health profiles. Through the analysis of the current employee healthcare trend and claims, the discovery of the hidden patterns of employee healthcare trends would be identified. This project will enable companies to better understand the underlying patterns in employee medical claims. The aim of the research is to better understand the patterns present in employee medical claims data, to better understand the claim behavior of employees and whether employees are fully utilizing or underutilizing the medical coverages provided. To potentially help optimize the medical expenditure for a company based on the claim behavior and pattern analysis. And finally, to provide strategies and recommendations through the analysis.

### IV. RESEARCH METHODOLOGY

#### A. Data Collection

Data collection (see Fig. 1) is the phase where data to be used in this project analysis will be collected. Data collection can occur in many different scenarios such as through surveys, questionnaire, databases, text, etc. - depending on the field of research. In this case, we will be using fictional data of human resource to demonstrate the use case on how one can conduct profiling patterns on the data.

#### B. Data Cleaning and Data Preparation

Data preparation and data cleaning (see Fig. 1) will involve the ETL process which is to extract the data, transform it into the manner which analysis can be performed and then loading it for analysis to be performed as well as data cleaning to remove any data deemed as unnecessary such as outliers, duplicates, wrong spelling or noisy data. This process is to prepare the data necessary and required to build the predictive model. Data transformation techniques will be applied such as replacement, binning, imputation as well as merging of different datasets before further analysis can be performed. Data preparation has to be performed prior to performing the analysis as it helps to prepare the data to achieve the project objectives.



Fig. 1. Flow of Work.

### C. Data Understanding

The next phase would be to have a comprehensive understanding of the data. Data understanding (see Fig. 1) is one of the most important steps when performing an analysis because it provides a clear and complete overview of the collected data and which variables might be important and which to be included or excluded for the analysis. Data understanding would also allow us to have an overview of the available variables which would allow for the predictive models to be built.

In this phase, the types of data (nominal, ordinal and continuous), correlation and relationships in data, basic descriptive analysis such as bar charts, pie charts, etc. will be created to enhance the understanding of the data which will be obtained. Data understanding would also provide us with knowledge on what variables and types of transformation techniques which is required for the dataset.

### D. Data Analysis

Data analysis (see Fig. 1) phase in this project would involve graphical representations such as bar charts, histograms, pie charts, line graphs as well as clustering analysis to discover segmented profiles and hidden patterns. Furthermore, segmentation will be performed as well where Clustering will be used to segment individuals into distinct groups with similar characteristics.

## V. ANALYSIS AND DISCUSSION

The medical claims were split into various categories such as General Practitioner (GP), Specialists (SP) and In-Patient (IP). Each of the category has its own meaning, GP would represent the common outpatient and normal visits to the hospitals or clinics, while SP would represent the trip to a specialist and finally, IP would represent patients who have been admitted into the hospital.

As shown in Fig. 2 there has been a gradual increase in the overall amount incurred throughout the period of 2016 to 2018. For GP it was 4.69mil in 2016, 5.1mil in 2017 and increased up to 5.64mil in 2018. For SP, it was 2.46mil in 2016, 2.76mil in 2017 and 2.85mil in 2018. And lastly, for IP, in 2016 it was 10.86mil, 2017 it was 12.2 and in 2018 it increased another 800,000 to 13mil incurred in in-patient cases. This shows that the amount incurred has been gradually increasing every year due to inflation as well as more medical claim cases as well which is driving companies to better understand the claim patterns.

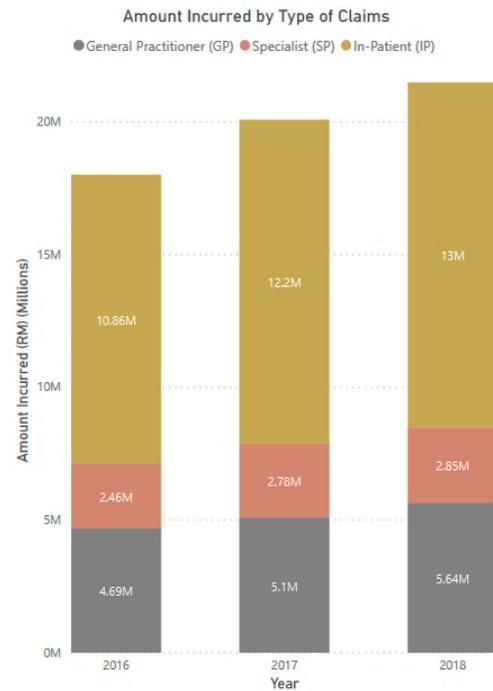


Fig. 2. Amount Incurred by Type of Claims.

Fig. 3 shows the increase in the number of employees between year 2016 and 2018. As shown on the bar chart, there is a gradual increase over the period of 3 years from 8347 total employees in 2016, to 8935 in 2017 and 9186 in 2018. In addition, there is a correlation between the increase in employees as well as the increase in the number of claims made over the 3 year period. In 2016, 51,000 claims were made, while in 2017, 53,000 claims were made and in the most recent year, 2018, 55,000 claims were made by the employees.

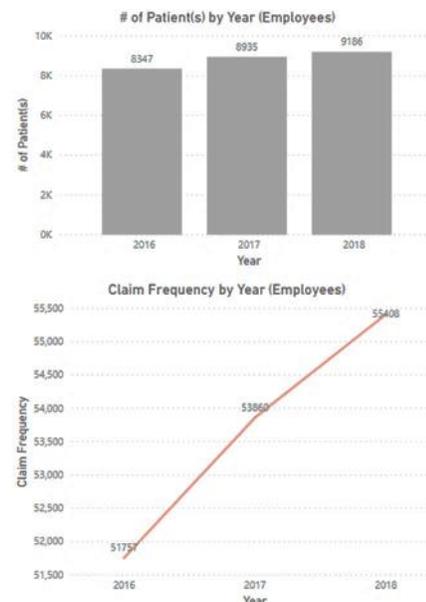


Fig. 3. Claim Frequency and # of Employees (GP) 2018.

Above in Fig. 4 shows the demographics of the company's employees, spouses and children who are under the medical insurance coverage. The demographics shown are for the year 2018 under General Practitioner (GP) claims history. As shown, there are a total of approximately 16,600 patients who are medically covered - out of the 16,600; 9190 are employees, 4800 are spouses and 2600 are children.

Moving on to Fig. 5, an additional visualization was done to identify the most common diagnosis found among the medical claims. Based on the findings, Acute Upper Respiratory Infections (8516) and Fever (6506) were the most commonly diagnosed illnesses among the claims. Besides, there were 2 chronic diseases which would be a cause of concern - Low Back Pain (2942) and Hypertension (1231). Upper Respiratory Infection also translates to sore throat in lay man terms, furthermore, there seems to be a correlation between the top 2 common diagnosis of Upper Respiratory Infection and Fever as usually when patients get a sore throat, there would be fever present as well due to the inflammation in the throat. Hence, in this case, the diagnosis has proven this assumption.

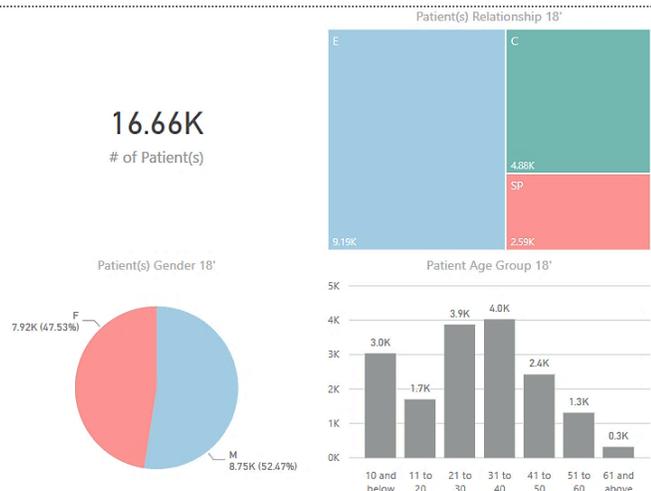


Fig. 4. Demographics of the Medical Coverage Patients.

Top Diagnosis with > 500 Claim Frequency 18'

Diagnosis	Count of Claim/Frequency
Acute Upper Respiratory Infections Of Multiple And Unspecified Sites	8516
Fever, Unspecified	6506
Upper Respiratory Tract Infection	4498
Diarrhoea And Gastroenteritis Of Presumed Infectious Origin(Acute Gastroenteritis /Dysentery)	3574
Low Back Pain	2942
Acute Pharyngitis	2365
Atopic Dermatitis	1558
Gastritis, Unspecified	1512
Headache	1509
Essential (Primary) Hypertension	1231
Conjunctivitis	1064
Acute Tonsillitis	1020
Acute Bronchitis	775
Urinary Tract Infection, Site Not Specified	714
Cough	677
Migraine	645
Asthma	592
Abdominal And Pelvic Pain	574
Acute Nasopharyngitis (Common Cold)	535
Acute Pain	529
<b>Total</b>	<b>41336</b>

Fig. 5. Diagnosis (GP) 2018.

Fig. 6 presents the demographics for patients under SP in 2018. Here, the patients have been filtered out to include only employees. There were 2107 patients who made 5791 claims in 2018 under specialists claim category. Out of the 2107 patients, 1046 were males and 1061 were females, it is almost a 50-50 split between males and females. RM2.1 mil was incurred in 2018 while RM2mil were insured. Looking at the age group distribution, the most common age group who visited specialists are between 31 and 50, this category comprises of over 50% of the patients.

From Fig. 7 the diagnosis section, under the specialists claims, the most common diagnosis was Hypertension at 64 claims, and Low Back Pain had 60 claims. The 3<sup>rd</sup> most common would be Coronary Artery Disease which is another chronic condition but there were only 49 of such claims. These results would show that there is a Hypertension and Low Back Pain issue within the company as it was present in the GP claims as well. And in comparison to the GP claims, under GP there were thousands of claims with Hypertension and Low Back Pain issue. This would draw the attention of the HR to present employees with short term fixes to try and fix this ongoing and growing issue.

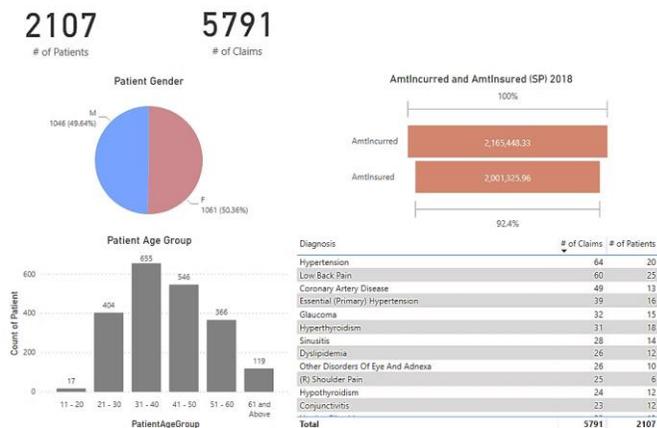


Fig. 6. Demographics for Employees (SP) 2018.

Diagnosis	# of Claims	# of Patients
Hypertension	64	20
Low Back Pain	60	25
Coronary Artery Disease	49	13
Essential (Primary) Hypertension	39	16
Glaucoma	32	15
Hyperthyroidism	31	18
Sinusitis	28	14
Dyslipidemia	26	12
Other Disorders Of Eye And Adnexa	26	10
(R) Shoulder Pain	25	6
Hypothyroidism	24	12
Conjunctivitis	23	12
<b>Total</b>	<b>5791</b>	<b>2107</b>

Fig. 7. Diagnosis (SP) 2018.

Fig. 8 shows the demographics for employees under IP in 2018. Again, the patients have been filtered to only include employees. There were 1034 patients who were admitted into the hospital in year 2018. 3839 claims were made under the IP claim category. Out of the 1034 patients, 434 were females while 600 of them were males which would translate to 42% were females and 58 were males. The age group distribution differed slightly from the SP claims, with the age group between 21 and 50 as the most common age groups who were admitted.

Out of the 3839 claims under IP (see Fig. 9), the most common diagnosis for patients to be admitted in 2018 was due to Gastritis (172), Intervertebral Disc Disorder (105), Dengue Fever (89), Acute Sinusitis (88) and Tear of Meniscus (80). These were the common diagnosis among the hospitalized patients.

Above in Fig. 10 is a demographics which was churned out based on the IP dataset. Based on the analysis, there was an identifier which showed the TypeofClaims of a patient. We identified that whenever a patient was admitted for the first incident, there would be a record which showed GHS (General Hospitalization). However, there is also another label which shows Post GHS (Post General Hospitalization), this is to signify post hospitalization checkups which would suggest that there were not so many claims on in-patient cases. Every in-patient patient would have a GHS row and multiple PostGHS row, counting the rows would not provide the count of a specific encounter. For example, if a patient was admitted for "appendicitis", there would be one row signifying GHS. After the surgery, the patient would have follow-up check up by the doctor, the rows are recorded as PostGHS. If we were to count the number of claims, it would be multiple claims, maybe 3-4 or 5-6 claims. However, this is not the actual number of encounters by a patient. This triggered us to unique identify each encounter by segregating them accordingly.

As shown in the illustration above Fig. 10, there were 950 patients who had over 1183 in-patient cases instead of the previous 3839 cases. Out of the 950 patients, 394 were females and 556 were males. The age group distribution still mimics the previous analysis with the most common age group between 21 to 50 years.

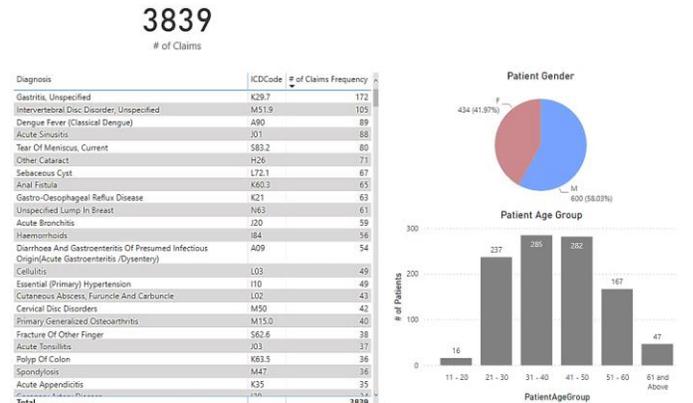


Fig. 9. Diagnosis (IP) 2018.

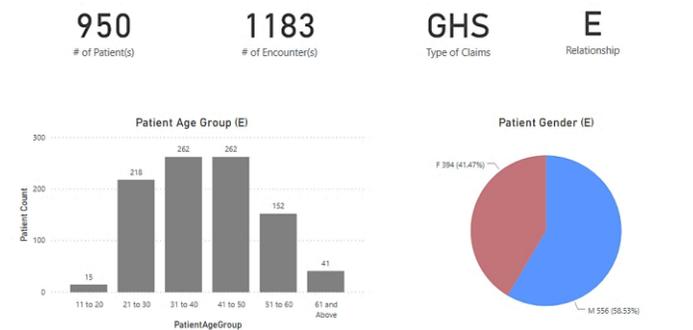


Fig. 10. IP Encounters 2018.

Here, we listed the common diagnosis such as Upper Respiratory issues, Low Back Pain and Hypertension. Out of the 3 diagnosis, 2 are chronic conditions such as Low Back Pain and Hypertension. The top four industries with Upper Respiratory issues (see Table I) based on claims in descending order are Learning Centre A (1536), Hotels (1263), Construction Company (1232) and Retailers (870). Learning Centre A Referring to (see Table II) has approximately 47.66% of employees who made claims in 2018, Hotels has more than 65.77% of employees, Construction Company only had 35.18% of employees and Retailers had around 49.24% of employees, over their total industry employee headcount.

For Low Back Pain issues (see Table I), Construction Company (484) had the highest number of claims with the most employees who made claims as well, followed by Hotels (442), Learning Centre A (250) and then Security Company (234). (see Table II) 298 employees out of 1697 (17.56%) from Construction Company made claims in 2018, while 250 out of 815 (30.67%) employees from Hotels, 194 out of 1563 (12.41%) employees from Learning Centre A and 139 out of 717 (19.39%) employees from Security Company.

Hypertension (see Table I) is another common diagnosis recorded among the employees. Learning Centre A (413) has the highest number of claims made for this diagnosis, Construction Company is the 2nd highest at (396) claims, then Hotels (349) and Learning Centre B (161) in 4th. Out of 1563 employees in Learning Centre, 100 of them (approx. 6.40%) (see Table II) made claims for hypertension diagnosis, while in Construction Company 103 employees out of 1697 (approx. 6%) made claims, 81 out of 815 (approx. 4.77%) for Hotels

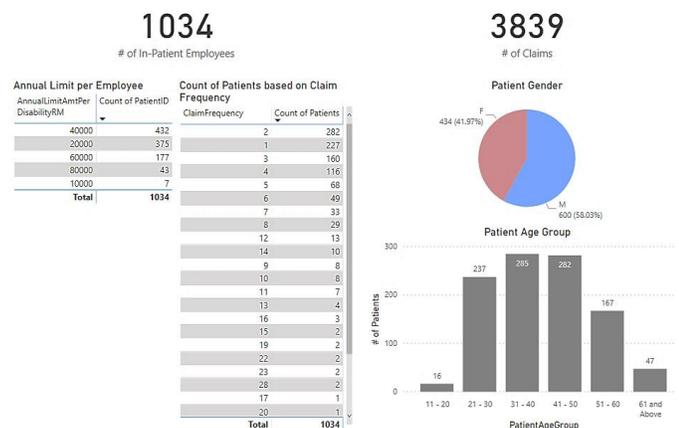


Fig. 8. Demographics for Employees (IP) 2018.

and 49 (approx. 5.58%) from Learning Centre B out of 878 employees.

TABLE I. COMMON DIAGNOSIS WITH # OF PATIENTS AND CLAIMS

Diagnosis	Business Industry	Industry Headcount	# of Patients	# of Claims
Upper Respiratory	Learning Centre A	1563	745	1536
	Hotels	815	536	1263
	Construction Company	1697	597	1232
	Retailers	725	357	870
Low Back Pain	Construction Company	1697	298	484
	Hotels	815	250	442
	Learning Centre A	1563	194	260
	Security Company	717	139	219
Hypertension	Learning Centre A	1563	100	413
	Construction Company	1697	103	396
	Hotels	815	81	349
	Learning Centre B	878	49	161

TABLE II. COMMON DIAGNOSIS WITH % OF PATIENTS AGAINST HEADCOUNT

Diagnosis	Business Industry	Industry Headcount	# of Patients	% of Patients
Upper Respiratory	Learning Centre A	1563	745	47.66
	Hotels	815	536	65.77
	Construction Company	1697	597	35.18
	Retailers	725	357	49.24
Low Back Pain	Construction Company	1697	298	17.56
	Hotels	815	250	30.67
	Learning Centre A	1563	194	12.41
	Security Company	717	139	19.39
Hypertension	Learning Centre A	1563	100	6.40
	Construction Company	1697	103	6.07
	Hotels	815	81	4.77
	Learning Centre B	878	49	5.58

As presented in Table III, we split the total number of patients into gender based with the value and the percentage of split as per gender, male or female. Let us look at Upper Respiratory first, under Learning Centre A, out of 745 patients, 228 (30.60%) were males and 517 (69.40%) were females. Hotels, out of 536 patients, 324 (60.45%) were males while 212 (39.55%) were females. Construction Company had 597 patients, 453 (75.88%) were males and 144 (24.12) were females. And finally, Retailers had 204 (57.14%) males and 153 (42.86%) females, in total 357 patients. Looking at Low Back Pain, under Construction Company, a high percentage

were males at 249 (83.56%) out of 298 patients and 49 (16.44%) out of 298 patients. Hotels had 250 patients, 177 (70.80%) out of 250 were males and 74 (29.60%) were females. Learning Centre A had 194 patients with 58 (29.90%) were males 136 (70.10%) were females. For Security Company, 123 (88.49%) were males and only a minimal 16 (11.51%) were females. For Hypertension, Learning Centre A has almost a 50-50 split, with 46 (46%) being males and 54 (54%) being females. Construction Company skewed towards males 92 (89.32%) and females only 11 (10.68%). Hotels had 47 (58%) males and 34 (41.98%) females. Learning Centre B had 49 patients in total, 33 being males and 16 being females.

TABLE III. % OF PATIENTS AGAINST PATIENT GENDER

Diagnosis	Business Industry	# of Patients	Gender (M)	Gender (F)
Upper Respiratory	Learning Centre A	745	228 - 30.60%	517 - 69.40%
	Hotels	536	324 - 60.45%	212 - 39.55%
	Construction Company	597	453 - 75.88%	144 - 24.12%
	Retailers	357	204 - 57.14%	153 - 42.88%
Low Back Pain	Construction Company	298	249 - 83.56%	49 - 16.44%
	Hotels	250	177 - 70.80%	74 - 29.60%
	Learning Centre A	194	58 - 29.90%	136 - 70.10%
	Security Company	139	123 - 88.49%	16 - 11.51%
Hypertension	Learning Centre A	100	46 - 46%	54 - 54%
	Construction Company	103	92 - 89.32%	11 - 10.68%
	Hotels	81	33 - 58.02%	34 - 41.98%
	Learning Centre B	878	33 - 67.35%	16 - 32.65%

## VI. CONCLUSION

Based on the findings and analysis which were performed, we managed to discover meaningful insights within the medical claims data among the employees. One of the key discoveries would be the most common diagnosis among the employees which included, Upper Respiratory Infection, Low Back Pain and Hypertension. Out of the three diagnosis, there are two which are chronic conditions which is a cause of concern for the Human Resource department. As per the analysis performed, we have looked at the various claim categories, GP, SP and IP, identified the common diagnosis and the age groups associated to the diagnosis, identified the pattern of claims based on dates and looked at the total amount incurred throughout the 3 year period. These preliminary analysis will allow us to perform drill down analysis in the different areas as it gave us a better overview of the current employee health population and state.

Some assumptions made as to why Hotels result are more to upper respiratory issues and it is common due to employees are consistently expose to A/C in the working environment which is drying and the constant talking to the guests may be a cause of the high number of diagnosis among this BU. Construction Company could be due to the constant exposure to the sun and outdoors with the lack of consumption of water which led to upper respiratory infections. As per our discussion with the Human Resource department, Low Back Pain issues are common among business industries where their job require them to do a lot of walking and long periods of standing. As shown in the analysis tabulated in Table I, all the top 4 business industries would require the employees to do a lot of walking such as Construction Company, Hotels as well as Security Company, while Learning Centre employees are consistently standing to teach. Finally, for Hypertension issues, an assumption made was reflected on the working environment which contributed to this diagnosis.

To sum up the above explanation as shown in Table I, the numbers shown do not reflect major concern as the percentage of chronic condition patients are below the 30% mark out of each respective industry headcount. The only business unit which had a 30% and above mark would be Hotels who had 250 patients out of 815 employees. Looking at the percentage of Hypertension patients, they are all below 10% of their respective business industry. This shows that there is not a major cause of concern once you compare the total number of patients against the total headcount of each of the business industry.

As shown from the gender comparison chart in Table III, Upper Respiratory issues are generally more common among M as compared to F; except for Learning Centre A, where F had 69.40% as compared to M who had 30.60%. For other industries, as mentioned it is more common among M - Hotels (60%), Construction Company (75%) and Retailers (57%). Low Back Pain is also more common among M, with only Learning Centre A being the opposite; where F had 70.10% and M were only 29.90%. Hotels had 70% who were M and 30% who were F while both Construction Company and Security Company had over 80% cases of Low Back Pain among M. There is a similar pattern as well for Hypertension issues where Learning Centre A had more F (54%) as compared to M (46%). This shows that for all three diagnosis, Learning Centre A had more F who were diagnose with the issues as compared to M.

An additional pattern analysis performed was the distribution of patients based on patient age group, this will allow us to see the pattern among the different age groups. Within the Upper Respiratory diagnosis, Learning Centre A and Hotels, had the highest number of patients found in age group 31 – 40, at 355 and 192 patients respectively while Construction Company and Retailers recorded the highest number of patients within the age group of 22 – 30, at 237 and 139 patients, respectively. Under Low Back Pain, Construction Company and Security Company recorded the highest number of patients within the 22 – 30 age group at 100 and 45 patients respectively. Hotels and Learning Centre A both had the highest number of patients within the 31 – 40 age group, 110 and 92 patients. Hypertension diagnosis was most

commonly found among the age group of 41 – 50 across the industries including Learning Centre A, Construction Company, Hotels and Learning Centre B.

Apart from the common diagnosis, looking at the excess claim amount made by taking the coverage amount deducting the total amount insured for each employee throughout the year allowed us to identify the excess claim amount by each employee. This would show if an employee has exceeded the coverage amount and if there is a consistent pattern of chipping in to pay for the excess amount claimed. However, there is no such pattern and it is fair to say that there is a minimal excess claim amount made by the employees throughout the 3 year period of 2016 to 2018. This would suggest that the medical coverage currently is more than sufficient to cover for every employee in the company. The number of patients who has spent an excess of the insurance coverage are minimal and the ratio is comparison is approximately 90 - 10; 90% of them being those who are within the medical coverage and 10% who had spent an excess of the coverages. Which draws another conclusion that the current insurance package which the company is paying would be more than sufficient to cover for the current employees based on the population health. The company would have an option to choose to remain with the current insurance package without the need to increase on the medical premium as per suggested by the insurance companies.

To conclude the analysis, we have achieved the objectives first stipulated by understanding the pattern of medical claims among the employees, and identifying the most common diagnosis throughout the year; to potentially help optimize medical expenditure by focusing on the common diagnosis segment as this segment of individuals are one of the largest group who drives medical claims; and lastly, to provide recommendations in the following section. This analysis has allowed us to better understand the employees' medical claims and to have an overview of the current employee health population. With reference to the findings, we will provide a better overview on the recommendations which will be proposed in the next section.

## VII. RECOMMENDATIONS AND FUTURE WORK

### 1) Upper Respiratory Infection

- Human Resource department could install air purification systems in every department to ensure the air within the department would be purified as most employees spend most of their time in the office spaces. So, there is a need to have clean and fresh air.
- Human Resource department could potentially provide the necessary vaccination to the specific group of target segment which has highest volume of medical claims within the business unit.
- With the specified recommendation, the Human Resource department could monitor the changes within the next 3 months to observe if there are any changes within the claim pattern.

## 2) Low Back Pain

- Human Resource department could start by targeting the business units with the highest medical claims and try to observe the day-today operations within the business units to better understand why the business units are experience such an issue.
- Human Resource department could provide “Back Pain Relief Lumbar Support Cushion Pillow” to help employees with their posture and comfort levels. As consistently sitting on a chair without proper back support could affect the lower back.
- Human Resource department could provide encourage and simple exercises which employees could do while at the offices such as simple stretching exercises to help loosen the muscles.

## 3) Hypertension

- Human Resource department could start by targeting the business units with the highest medical claims and try to observe the day-today operations within the business units to better understand why the business units are experience such an issue.
- Human Resource department could provide simple exercises and stress relief techniques to help employees relax during their day-to-day operations.

## 4) Future Works

- We will explore the possibility of implementing a new algorithm to enhance the predictive outcomes while also providing more innovative results as this is only a preliminary study. The recommendation should be further confirming and verification of the causes before instituting measures. However general measure to promote healthcare can be put in place. The outcome of this studies shows how one could segregate the data and conduct profiling patterns.

### ACKNOWLEDGMENT

The authors would like to thank all the relevant parties in supporting this research works.

### REFERENCES

- [1] Gore, "The Digital Earth: Understanding Our Planet in the 21st Century," *The Australian Surveyor*, vol. 43, no. 2, pp. 89-91, 2012.

- [2] E. Kar, "The Evolution of Big Data and its implications," *Happiest Minds Blogs*, 15 May 2015. [Online]. Available: <http://www.happiestminds.com/blogs/the-evolution-of-big-data-and-its-implications/>. [Accessed 9 October 2017].
- [3] E. Rahm, "Big Data Analytics," *IT - Information Technology*, vol. 4, no. 58, pp. 155-156, 2016.
- [4] R. Sippe, "The Relation between Big Data and Informational Privacy in the Context of the Healthcare," *Delft University of Technology*, Delft, 2015.
- [5] A. G. Eapen, "Application of Data Mining in Medical Applications," *University of Waterloo*, Ontario, 2004.
- [6] B. Beh, "Avoid Overdependence on Employee Health Benefits," *Focus Malaysia*, 22 February 2019. [Online]. Available: <http://webcache.googleusercontent.com/search?q=cache:bJtd5GUvYJJ:www.focusmalaysia.my/Income/avoid-overdependence-on-employee-health-benefits+&cd=1&hl=en&ct=clnk&gl=my>. [Accessed 26 February 2019].
- [7] R. Abelson, "Employer Health Insurance is Increasingly Unaffordable, Study Finds," *The New York Times*, 25 September 2019. [Online]. Available: [nytimes.com/2019/09/25/health/employer-health-insurance-cost.html](http://nytimes.com/2019/09/25/health/employer-health-insurance-cost.html). [Accessed 5 March 2020].
- [8] Malaysian Reserve, "Reining in Medical Benefits Cost," *Malaysian Reserve*, 3 April 2017. [Online]. Available: <https://themalaysianreserve.com/2017/04/03/reining-in-medical-benefits-cost/>. [Accessed 26 February 2019].
- [9] M. Rouse, "Big Data," *TechTarget*, 7 July 2014. [Online]. Available: <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data>. [Accessed 15 April 2017].
- [10] M. S. Islam, M. M. Hasan, X. Wang, H. D. Germack and M. Noor-E-Alam, "A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining," *Healthcare Basel MDPI*, vol. 6, no. 2, pp. 1-43, 2018.
- [11] S. G. Alonso, I. d. I. T. Díez, J. J. P. C. Rodrigues, S. Hamrioui and M. López-Coronado, "A Systematic Review of Techniques and Sources of Big Data in the Healthcare Sector," *Journal of Medical Systems*, vol. 41, no. 1, pp. 1-9, 2017.
- [12] N. Mehta and A. Pandit, "Concurrence of Big Data Analytics and Healthcare: A Systematic Review," *International Journal of Medical Informatics*, vol. 114, no. 1, pp. 57-65, 2018.
- [13] W. Raghupathi and V. Raghupathi, "Big Data Analytics in Healthcare: Promise and Potential," *Health Information Science and Systems*, vol. 2, no. 3, pp. 1-10, 2014.
- [14] R. A. Soeini and K. V. Rodpysh, "Applying Data Mining to Insurance Customer Churn Management," *2012 IACSIT Hong Kong Conferences*, vol. 30, pp. 82-92, 2012.
- [15] VRajaraman, "Big Data Analytics," *Resonance*, pp. 695-716, 2016.
- [16] A. Kacik, "Healthcare Costs Increasing at Unsustainable Pace," *Modern Healthcare*, 13 June 2018. [Online]. Available: <https://www.modernhealthcare.com/article/20180613/NEWS/180619961/healthcare-costs-increasing-at-unsustainable-pace>. [Accessed 11 March 2019].
- [17] Human Resources Online, "A Quick Fix to Employee Benefit Costs in Malaysia," *Human Resource*, 21 October 2015. [Online]. Available: <https://www.humanresourcesonline.net/quick-fix-employee-benefit-costs-malaysia/>. [Accessed 11 March 2019].