Applying Data Mining to Healthcare: A Study of Social Network of Physicians and Patient Journeys

Shruti Kaushik(✉), Abhinav Choudhury, Kaustubh Mallik,
Anzer Moid, and Varun Dutt

Applied Cognitive Science Laboratory, Indian Institute of Technology Mandi, Mandi 175005,
Himachal Pradesh, India
{shruti_kaushik,abhinav_choudhury,
kaustubh_priya,md_anzer}@students.iitmandi.ac.in,
varun@iitmandi.ac.in
http://www.acslab.org/index.html

Abstract. In 2004, the US President launched an initiative to make healthcare medical records available electronically [27]. This initiative gives researchers an opportunity to study and mine healthcare data across hospitals, pharmacies, and physicians in order to improve the quality of care. Physicians can make better informed decisions regarding care of patients if physicians have proper understanding of patient journeys. In addition, physician healthcare decisions are influenced by their social networks. In this paper, we find patterns among patient journeys for pain medications from sickness to recovery or death. Next, we combine social network analysis and diffusion of innovation theory to analyze the diffusion patterns among physicians prescribing pain medications. Finally, we suggest an interactive visualization interface for visualizing demographic distribution of patients. The main implication of this research is a better understanding of patient journeys via data-mining and visualizations; and, improved decision-making by physicians in treating patients.

Keywords: Diffusion of innovation · Patient journey · Social network analysis · Physicians · Visualization · Pain medications

1 Introduction

Modern healthcare has started using a patient-centered approach by building and evaluating patient journeys through their sickness and recovery. The study of patient journeys is relatively recent innovation in the healthcare quality improvement process [1]. A patient's journey involves a sequence of events that a patient proceeds through from the point of entry into the healthcare system (triggered by sickness) until the complete recovery or death. Thus, patient journeys include filling prescriptions at a pharmacy, visiting a doctor, being admitted to the hospital, undergoing lab tests, getting treatment, and recovering from sicknesses. Understanding the whole journey from patient's point of view is important as the patient is the only person who experiences the whole journey [2]. Patient journeys highlight bottlenecks in the healthcare

P. Perner (Ed.): MLDM 2016, LNAI 9729, pp. 599–613, 2016.
DOI: 10.1007/978-3-319-41920-6_47

system, which helps providers improve the system. Hence, a comprehensive approach to patient journeys can help impart provocative vision that may lead physicians to revise their treatment plan.

During a patient's journey, the decisions of the physicians may be influenced by their social network. Most importantly, the medication that a physician prescribes or any new innovation that he adopts is highly influenced by the interpersonal communication of the physician with the members of his personal network [5]. Here, social network analysis investigates various direct and indirect interpersonal communications and interaction patterns between the members of a social system.

Since patients are the end-users of medications, the understanding of patient journeys will help us specify how and when patients consume pharmaceutical products. Proper evaluation of patient journey will help a physician know the unspoken needs of patients. Thus, it is important to investigate factors involved in creating good patient journeys. Up to now, researchers have created these journeys using process mapping and visualization tools by conducting surveys for a few hundred patients [1]. In this paper, however, we have adopted a bottom-up approach to understand patient journeys. We focus on pain medications being used by patients and prescribed by physicians in the United States of America. We build patient journeys by mining through billions of patient records and creating a social network of the physicians using their prescribing histories, specialty, and diffusion of innovation theory [7]. Furthermore, using social network analysis, we pinpoint key-opinion leaders in the physician's social network. These opinion leaders have high influencing power and can bring about behavioral change in ways that medications are prescribed [7]. Moreover, in order to visualize multidimensional patients' data we have implemented visualization techniques for exploratory data analyses. Starting in the next section, we discuss background work in the area of patient journeys, social network analysis and visualization. Then, we divide our analysis into three sections: The first section deals with patient journeys; second section deals with diffusion of innovation of the medication and social network analysis of the physicians' network; and, the last section shows the visualization of multivariate dataset. We close this paper by discussing the significance of our results for improving existing healthcare system.

2 Background

Patient journey motivates us to observe and examine the various paths a patient goes through and the decisions that are taken by various stakeholders through the amelioration of a disease and its treatment. Process mapping is a framework used for building patient journeys through patient's perspective. It allows us to figure out a series of sequential-events linked with the patient's experience [3]. It aims to boost the quality of clinical role, eliminate unnecessary activities from the care, and finally focus on more valuable activities [4]. Though process-mapping technique successfully maps patient's experiences; however, in order to visualize each stakeholder's role we need graphical communication tools. These graphical communication tools act as communication medium which encourages communication between various participants and helps to improve the journey experiences of the patient. Furthermore, the communication tool encourages stakeholder involvement and highlights the importance of relevant variables that contribute to the whole experience of the patient. However, less attention has been paid to the analysis of patient journeys using bottom-up approach, i.e., from low-level data. In this paper, we use data from US for pain medications and build patient journeys bottom-up. We believe that building such journeys will provide insights into improving patient experience across their journeys.

Furthermore, the rate of adoption of a new medication/innovation varies from physician-to-physician [5]. While some physicians adopt a new medication early, most of them tend to test the waters first. Thus, most physicians wait for others in his social system to have tried the innovation first. The diffusion of innovation is the process in which an initial few people adopt an innovation first and, through their social network, the innovation diffuses to others in the network. As time goes on the rate of adoption increases and all or most members of the social system start adopting the innovation [6-8]. A social network/system is the pattern of friendship, advice, communication or support which exists among the members of a social system [9-12]. Such networks can be used to find key-opinion leaders inside a social system. Thus, a major contribution of social networks to diffusion research has been the categorization of adopters based on innovativeness as measured by the time-of-adoption [13]. Here, innovativeness is the degree to which an individual is relatively early in adopting new ideas compared to other members of a social system [15]. As per theory of innovation of diffusion, there are five adopter categories among members of a social system on the basis of their innovativeness: 1) Innovators, 2) Early adopters, 3) Early majority, 4) Late majority, and 4) Laggards [6, 14, 7]. According to theory of diffusion of innovation given by Everett Roger [7, 17] adopter distribution takes the form of a bell-shaped curve. Using two basic statistical parameters of the normal adopter distribution-mean time of adoption (t) and its standard deviation (σ) we obtain the five adopter categories [17] (Table 1). In this paper, we use this categorization and apply it to physicians prescribing pain medications in the US. As suggested by theory of innovation-diffusion, physicians' relative location in the social network with other physicians affect their decisions concerning the adoption of new innovation [16]. Thus, categorization based upon theory of innovation-diffusion helps us understand how pain medications diffuse over a social network of physicians and allows us to measure the rate of diffusion of innovation over time.

Table 1. Adopter categories based on Innovativeness [17]

Adopter Categories	% adopters	Area covered under curve
Innovators	2.5	Between t - 2σ
Early adopters	13.5	Between $t - \sigma$ *and* $t - 2\sigma$
Early majority	34	Between t *and* $t - \sigma$
Late Majority	34	Between t *and* $t + \sigma$
Laggards	16	Between $t + \sigma$

Furthermore, huge amounts of multivariate data are being generated on a daily basis nowadays about patient journeys and physician prescribing histories. In order to find patterns in the growing data, we need to be able to visualize it efficiently. Conventional visualization methods like 2-d plots, scatter grams, histograms are limited in the sense that they can only depict 2-dimensions at one time. In contrast, Parallel Coordinates system [24] has been recently proposed and it allows end-users to visualize entire data together at one time. Furthermore, various functionalities like brushing, scattering and distribution can be attained using Parallel Coordinates. Currently, healthcare industry lacks a generalized, interactive and easy-to-use visualization interface which incorporates features like brushing, distribution, visualization of selective dimensions, and correlation among dimensions. Parallel Coordinates helps us overcome this necessity. In this paper, we build a tool based upon Parallel Coordinates that accepts a CSV dataset of any size and generates the visualization plot, correlation matrix, and distribution lists from this data.

3 Method

The patient journeys, social networks, and visualizations were created for patients and physicians residing in the US. We used a large medical-prescriptions dataset[1] in order to build patient journeys, social network of the physicians, and visualizations. The dataset, containing patients and physicians, was provided by a pharmaceutical company.

3.1 Patient Journey

We have focused our analyses on outpatient refill data and inpatient hospital-visitation data. The data is Big in nature as it has more than 100 million records between years 2008 and 2014. Inpatients are those who consume pain medications and are admitted to hospitals. Outpatients are those patients who consume pain medications but were never admitted to hospitals. We used a Big-Data architecture consisting of q-programming language to query a kdb+ database (from Kx systems) in order to find patterns among inpatients and outpatients [18]. In our patient journeys, each activity of patients' is coded using a letter code. For example, *H* represents that a patient has been admitted to the hospital, *D* represents discharge from hospital. Furthermore, *px* and *dy* represent procedure and diagnostic-test codes corresponding to procedure *x* and diagnostic-test *y* performed on patients, respectively. Medicine consumption is coded as amount of potency consumed by the patient (e.g., 5 mcg/hr, 10 mcg/hr, 15 mcg/hr, and 20 mcg/hr). After building the long chain of sequences for each patient who consumed the pain medications, we applied Apriori algorithm [19] to find out strong association rules across journeys. The Apriori algorithm helps us to find the frequently appearing item sets in a large database. The frequent item sets are used to determine the association rules that highlight patterns in data. We have performed a demographic analysis based on sex and age-group of patients to see how

[1] Data shown has been altered to protect privacy.

these characteristics affect the refill behavior among patients. Furthermore, we analyzed the switch behavior to see how patients switched from one medicine level to another during their journeys.

3.2 Social Network Analysis

The dataset consisted of billions of physicians prescribing medicines between years 2011 and 2015. The nature of data was prescription records of each physician during this period. Using social-network analysis, we created a social network for 50,000+ physicians, who prescribed pain medications and satisfied the constraints discussed below.

The following assumptions were used to create a social network:

1. Physicians living in close vicinity are likely to be in contact with each other and influence each other
2. Physicians having the same specialty are likely to be in contact with each other and influence each other

The distance d (i, j) between the physician i and j was computed using the haversine distance formula [20]

$$d(i,j) = 2\text{rarcsin}(\sqrt{sin^2\left(\frac{\varphi_j - \varphi_i}{2}\right) + \cos(\varphi_i)\cos(\varphi_j)sin^2\left(\frac{\lambda_j - \lambda_i}{2}\right)} \tag{1}$$

where φ_i, φ_j and λ_i, λ_j are the latitude and longitude respectively, of prescribers i and j. The distances calculated are in miles. Thus, all physicians prescribing pain medications, having the same specialty, and within a certain threshold distance of each other were in each other's social network. In this paper, the threshold distances selected were the 12.5th, 25th and 50th percentiles of all distances calculated between each physician (12.5th percentile: 323 miles; 25th percentile: 527 miles; 50th percentile: 919 miles). Once the social network was created, each physician was given a diffuser level based on when he/she first prescribed pain medications with respect to his/her personal network. The algorithm for assigning a diffuser level is as follows:

1. Prescribing physicians are given an integer diffuser level in the set {1, 2…n}
2. A physician *i* is given diffuser level (n) = 1 if he/she prescribed in the first month or is the first person in his/her personal network to prescribe the pain medications, i.e., all members of his/her personal network had a time of adoption later than *i*
3. The diffuser level assigned to a physician is n+1 where n is the diffuser level of physician i who is in the personal network of j but has prescribed the pain medications earlier than j.

3.2.1 Algorithm for Social Network Analysis

The algorithm used for social network analysis can be summarized as follows:

1. Calculate distances between all prescribers of pain medications using *haversine* [20] distance formula
2. Set a distance threshold for creating the social network

3. Create a social network by connecting physicians satisfying the vicinity and specialty assumptions
4. Find the key-opinion leaders i.e., prescribers with the highest number of connections
5. Assign diffuser level to each prescriber
6. Apply the diffusion of innovation theory to the social network to categorize the physicians into the five adopter categories

Based upon the algorithm above, key-opinion leaders were those physicians who were categorized as early adopters and whose personal network had the highest number of prescribers of the medication.

3.3 Visualization

Parallel Coordinates are becoming popular methods for data visualization, especially for multivariate data. The technique was proposed by Inselberg for analysis of hyper-dimensional geometry [21]. To show a set of points in an n-dimensional space, a backdrop is drawn consisting of parallel lines, typically vertically and equally spaced. A point in an n-dimensional space is represented as a polyline with vertices on the parallel axes; the position of the vertex on the i-th axis corresponds to the i-th coordinate of the point. In order to make the interface more interactive various functionalities were incorporated by making use of the d3 parallel coordinates library of JavaScript and enabling brushing experience for the user [22, 23]. This library follows an object-oriented design and consists of core functionalities implemented in JavaScript. It provides APIs that are used for further development of new features [23]. For testing our implementation, we used Parallel Coordinates on a large medication refill dataset (5000+ patients).

4 Results

4.1 Patient Journey

The Apriori algorithm [19] measures the quality of association rules using confidence of the rule. The confidence of a rule is the number of cases in which the rule is correct related to the number of cases in which it is applicable. Based upon the Apriori algorithm, we found the following three rules with 100% confidence among patients:

1. *If patients suffer from morbid obesity, then they go for gastric bypass and gastric restrictive procedures and consume pain medications*
2. *If patients are females and they experience infections related to giving birth, then they go for spine-related surgeries and consume pain medications*
3. *If patients have severe knee conditions (e.g. osteoarthritis), then they go for total knee replacements and consume pain medications*

As the above rules had 100% confidence, whenever the “if” condition occurred in data, then the “then” part of the rule occurred with 100% probability. These rules indicated that the same pain medications were used to treat a number of post-operative patient conditions.

Furthermore, results showed that more females went for refills as compared to males among both inpatients and outpatients (Fig. 1a and 1b). Total number of males and females going for refill were 4,800 and 11,057, respectively, for inpatients (Fig. 1a) and 8,543 and 14,762, respectively, for outpatients (Fig. 1b). On average both males and females went for 4 refills among inpatients (Fig. 1a) and outpatients (Fig. 1b) (shown as the vertical dotted line). In addition, the average number of refills and the distribution of refills were the same among both inpatients and outpatients.

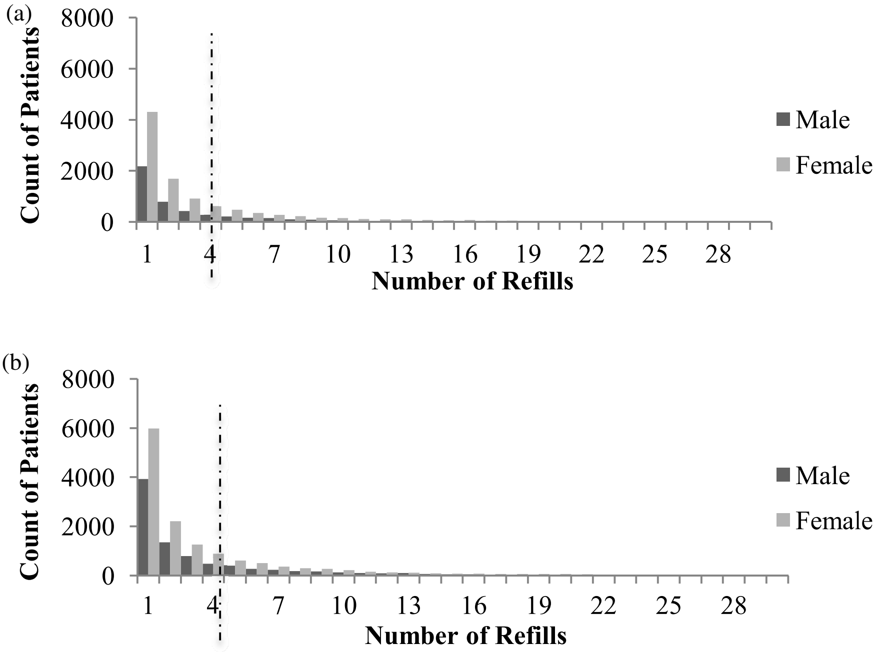

Fig. 1. Refill distributions among males and females for inpatients (a) and outpatients (b). The x-axis shows the number of refills and y-axis shows the number of patients going for refills. Vertical dotted line shows the average number of refills.

Next, we analyzed the distribution of refills with respect to different age groups among inpatients and outpatients (Fig. 2a and 2b). As shown in Fig. 2a, the total number of inpatients belonging to the age-group 0-17, 18-34, 35-44, 45-54, and 55-64 were 35, 2,468, 3,164, 4,705, and 4,664, respectively. Similarly, the total number of outpatients belonging to the age-group 0-17, 18-34, 35-44, 45-54, and 55-64 were 44, 3,964, 5,643, 7,247, and 5,449, respectively. Overall, there was more number of patients belonging to 45-54 and 55-64 age-groups refilling pain medications compared to other age groups. Interestingly, the average number of refills were same (~ 4) for both inpatients and outpatients.

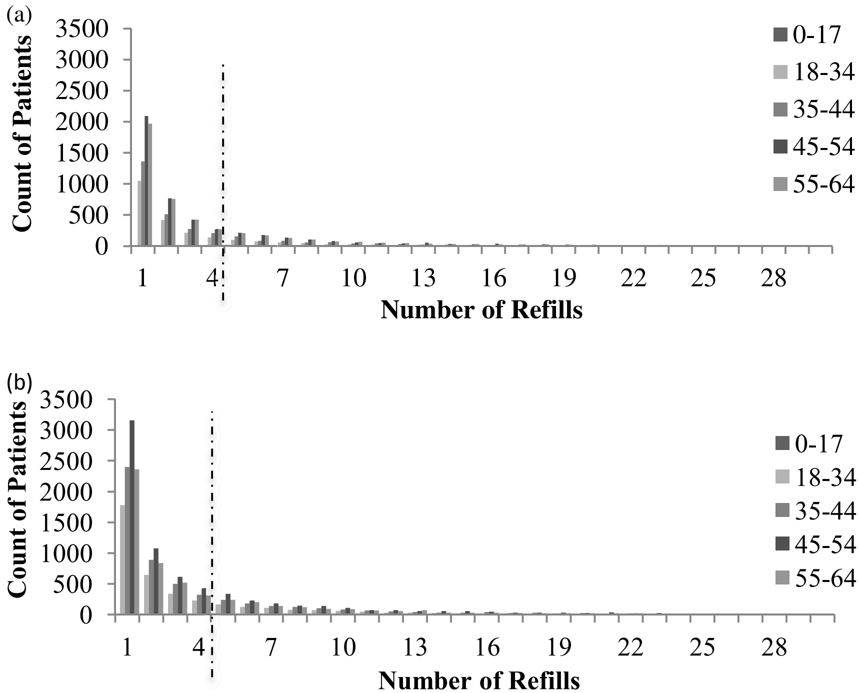

Fig. 2. Refill distributions among different age-groups for inpatients (a) and outpatients (b). The x-axis shows the number of refills and y-axis shows the number of patients going for refills. Vertical dotted line shows the average number of refills.

Lastly, we analyzed how patients switched between different potency of pain medications across their journeys. For this purpose, we counted whether a patient increased or decreased medication potency from one refill to the next (a switch). If the patient consumed the same potency across two consecutive refills, then she did not switch. Next, using this switch data, we computed the switch ratio as:

$$\text{Switch Ratio} = \frac{m}{n} \tag{2}$$

Where, m is the total number of times a patient switched medication potency across her journey and n is the total number of possible switches (= number of refills – 1). Fig. 3 shows the distribution of potency switching among patients using the switch ratios. As shown in Fig. 3, the distribution of switch ratios had a bimodal distribution with two peaks at 0.0-0.1 switch ratio and 0.9-1.0 switch ratio. Thus, there existed two kinds of switch-ratio behaviors: Patients who switched very little between different potencies and patients who switched a lot between different potencies.

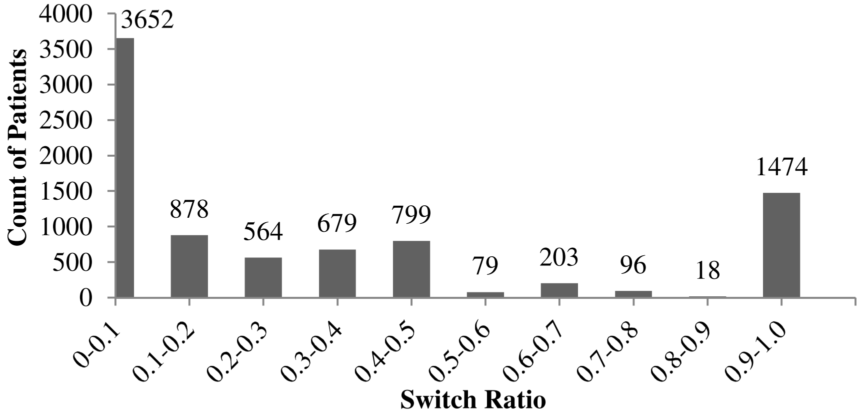

Fig. 3. Distribution of potency switching among patients. The x-axis shows the switch ratio and the y-axis shows the number of patients showing a particular switch ratio.

4.2 Social Network Analysis

In this section, we look into the results of how physicians are connected with other physicians in their network. Using the algorithm described above, first, we constructed a list of top-500 key-opinion leaders. Among this list, we found that most opinion leaders were from mid-western and southern regions of USA[2] for different percentile distances (Fig. 4). Among physicians, the specialties with the highest number of prescribers were the following (in decreasing order): Family Medicine, Internal Medicine, and Nurse practitioners.

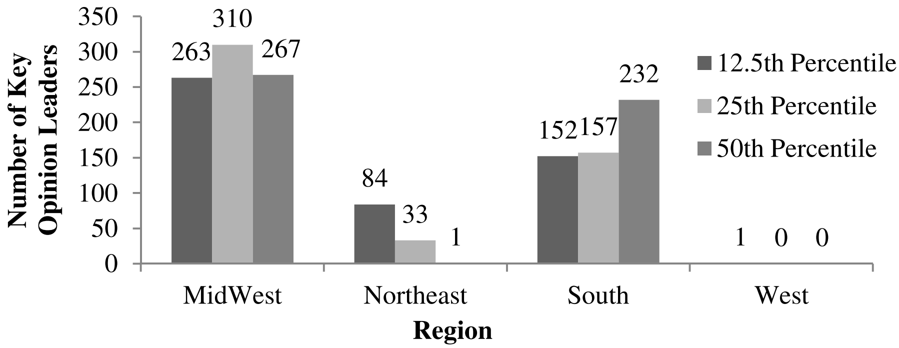

Fig. 4. Region-wise key-opinion leaders for 12.5th, 25th and 50th percentile.

[2] The regions were divided as per U.S. Census Bureau Regions and Divisions.

Next, we analyzed the count of physicians under different diffuser levels (as explained above a diffuser level was assigned to each physician). As shown in Fig. 5, we found that there were slightly more number of 3rd and 4th level diffusers in the 12.5th percentile compared to the 25th and 50th percentiles. We also observed that most physicians were 2nd level diffusers across all three percentiles and only the 50th percentile had 6th level diffusers.

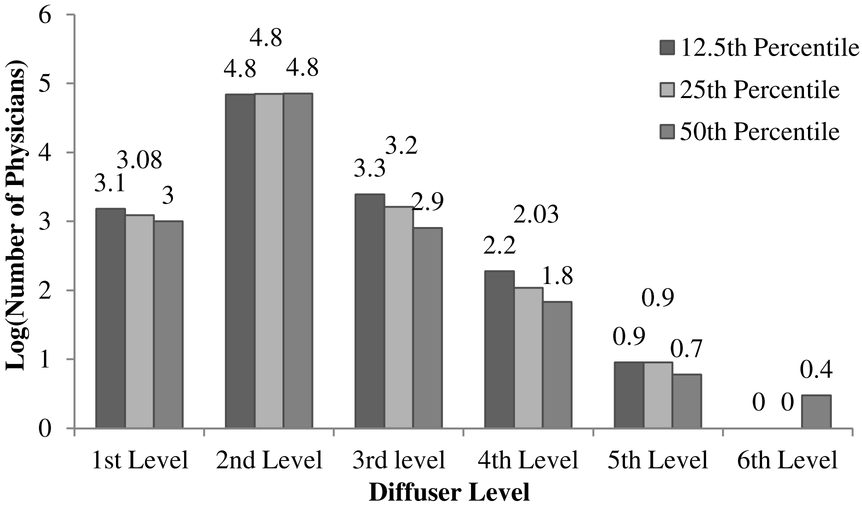

Fig. 5. Log (Number of Physicians) of Diffuser Level at 12.5th, 25th, and 50th percentile.

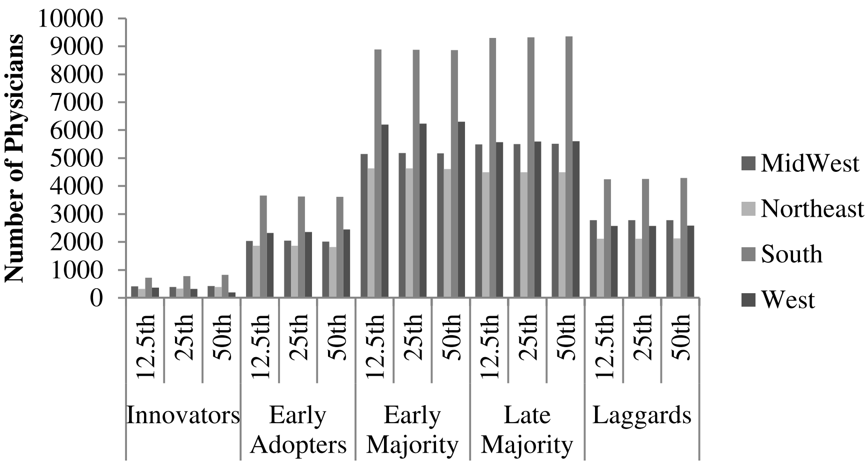

Fig. 6. Region-wise distribution of adopter categories across 12.5th, 25th and 50th percentile.

Next, using the theory of diffusion of innovation [7], we calculated the number of physicians in the different diffuser categories as well as distance percentiles (Fig. 6). As shown in Fig. 6, majority of physicians were from the southern region across all diffuser categories and different percentiles. The second highest numbers of physicians were from mid-west for innovator and laggards categories and from west for early-adopters, early-majority, and late-majority categories, respectively.

4.3 Visualization

In this section, we describe results of visualizing data of patients using parallel coordinates.

4.3.1 Distribution of Data in Different Dimensions by Brushing a Particular Dimension

It is important to visualize how a selected (brushed) dimension is distributed among other dimensions. For example, how a patient age-group 35 to 44 years is distributed among their healthcare plan type (Fig. 7). For this functionality, a user first needs to select a subset of data via brushing on a dimension (e.g., AGEGRP in Fig. 7). Then, the user needs to click on a different dimension's header (PLANTYP) to display a list showing the count and percentages of various values within the clicked dimension (shown in the box on the right in Fig. 7).

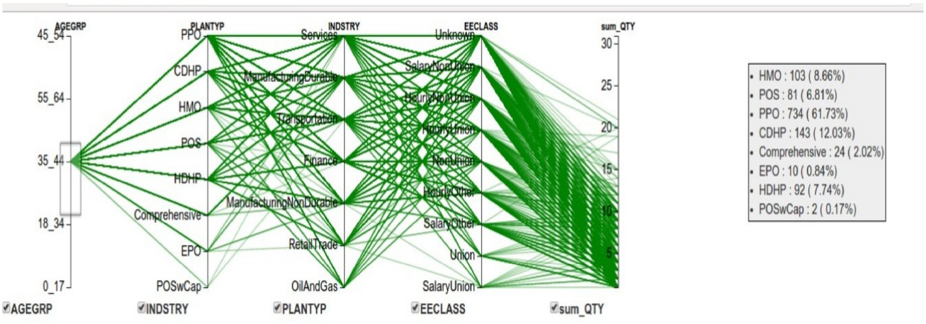

Fig. 7. Figure shows different dimensions (AGEGRP, PLANTYP, INDSTRY etc.) in parallel coordinate system. AGEGRP is brushed at the point 35_44 and PLANTYP header is clicked. The table on the right shows the distribution of different PLANTYP with values and percentages.

4.3.2 Correlation Between Dimensions

It is important to visualize if there is interdependency (correlation) among the different dimensions of a multivariate dataset. In order to find correlation among categorical dimensions (e.g., gender, region, etc.) integer codes (starting from 1 and above) were assigned in lexicographical order of the category. For numerical categories values were directly used to calculate correlation coefficients. Fig. 8 shows an example where correlation coefficients have been computed between different dimensions (some numerical and some categorical). The correlation matrix is dynamic in nature and changes whenever a user brushes a subset of points in a certain dimension (e.g., AGEGRP is brushed between points 18_34 and 35_44 and correlations are computed only for this data range).

Fig. 8. Table on the bottom-right shows correlation coefficients computed between different dimensions. Also, shown is the data in parallel-coordinates format (top panel) and tabular format (bottom panel).

5 Discussion and Conclusions

5.1 Patient Journey

The initiative launched by the US government in 2004 to make anonymized health records public has given data-mining researchers an opportunity to find patterns in patient journeys to identify bottlenecks in the healthcare system, boost the quality of clinical role, and eliminate unnecessary activities.

First, we found that certain pain medications were used to treat a number of post-operative patient conditions. This result shows that pain medications are being used by pharmaceutical industry across a wide spectrum of ailments among patients. Thus, it would be best for start-up companies to align their drug production strategies to those medications that are robust against a number of illnesses; rather, being focused on only a few ailments. Second, results showed that more females went for refills as compared to males among both inpatients and outpatients. One reason for this result could be the fact that females go through a number of medical procedures (like child-birth and spinal-cord injuries), which make them consume more pain medications compared to males. Third, results showed that there were more number of patients belonging to 45-54 and 55-64 age-groups refilling pain medications compared to other age groups. As per evidence from literature [26], knee problems are one of the most common health problems among middle-aged patients. Lastly, we found there existed two kinds of switch-ratio behaviors among patients' potency-consumption patterns. Thus, it seems that physicians are good at prescribing the right quantity of medication to majority of people. However, there also exists a large population of patients, where the pain medications require trial-and-error adjustments before the right potency is prescribed.

5.2 Social Network Analysis

As per theory of diffusion of innovation [7], innovators and early-adopters are known for their innovativeness and ability to take risks. Innovators are global leaders while early adopters are local leaders. As such, early-adopters have a higher influence as compared to innovators. In this paper, we classify physicians as key-opinion leader who are early-adopters, and have the highest number of prescribers in their personal network that they have influenced in diffusing pain medications.

First, we found that even though the innovation was pain medications, the highest number of prescribers were from the specialty family medicine rather than pain medicine. This finding seems counter-intuitive; however, it could simply be explained by the fact that family-medicine doctors are general physicians available in larger numbers to whom patients go to at the onset of sickness. In contrast, pain medicine is a specialized field where doctors would be fewer in number and only referred to by family-medicine doctors as a second step in patient journeys. Second, results show that most physicians are second level diffusers i.e. most physicians wait for other physicians in their personal network to prescribe before they themselves prescribe. This is also consistent with the diffusion of innovation theory [7], which shows fence-sitting effects, as bulk percentage of adopters are early-majority (34%) and late-majority (34%) rather than innovators or early-adopters. Third, our results show that there were slightly more number of 3rd and 4th level diffusers in the 12.5th percentile compared to the 25th and 50th percentiles. This is because in a smaller geographical area (12.5th percentile), people are likely to know each other and diffuse innovation across their small community; however, in a larger geographical area, people are less likely to know each other due to distance and would be less likely to diffuse innovation. Lastly, we found that southern region of US shows the highest concentration of physician prescribing pain medications across all the adopter categories. This peak for southern region was followed by western region; with a slight deviation in the case of innovators and laggards, where mid-western region was higher. This may be due to the fact that southern region has warmer temperatures and denser population followed by west and mid-western region [25].

5.3 Visualization

It became quite evident that scatter-grams and conventional methods are not sufficient enough for visualization of Big-Data in healthcare and other domains. The tool created on parallel coordinates technologies is very interactive and helps data analysts to mine data more efficiently and find meaningful patterns.

6 Future Scope

Patient journeys can provide a roadmap to create a better healthcare plan. In the future we will enhance patient journeys to highlight reasons that made patients or physicians to change their medication/treatment. We will also interleave diagnosis and procedures in the patient journeys which will explain reasons behind prescribing a particular medication.

Social network analysis can be an efficient and powerful tool to find key opinion leaders. In the future we will consider referral patterns and nomination studies to strengthen the reliability of the connections between the physicians and also use the centrality measures (e.g., eigen-vector centrality) as a metric for finding key-opinion leaders.

The visualization tools could be extended to incorporate features like pie-chart for distribution within a dimension, manual assignment of codes for non-numeric attributes for calculation of correlation, and showing only significant correlation among dimensions using t-distributions and p-values. The tools can also include linear-regression models between dimensions using correlations found between dimensions.

These and other ideas are some of the immediate next steps that we plan to take in this ongoing project.

Acknowledgement. The project was supported from grants (awards: #IITM/CONS/PPLP/VD/03 and #IITM/CONS/PPLP/VD/05) to Varun Dutt.

References

1. Curry, J., McGregor, C., Tracy, S.: A communication tool to improve the patient journey modeling process. In: Conference Proceedings of the International Conference of IEEE Engineering in Medicine and Biology Society, pp. 4726–4730 (2006)
2. Ben-Tovim, D.I., et al.: Patient journeys: the process of clinical redesign. Medical Journal of Australia **188**(6), S14 (2008)
3. Trebble, T.M., et al.: Process mapping the patient journey through health care: an introduction. BMJ **341**(7769), 394–397 (2010)
4. NHS Modernisation Agency. Process mapping, analysis and redesign. Department of Health, London, pp. 1–40 (2005)
5. Valente, T.W.: Social network thresholds in the diffusion of innovations. Social networks **18**(1), 69–89 (1996)
6. Ryan, B., Gross, N.C.: The diffusion of hybrid seed corn in two Iowa communities. Rural Sociology **8**, 15–24 (1943)
7. Rogers, E.M.: Diffusion of Innovation, 3rd edn. Free Press, New York (1983)
8. Valente, T.W.: Diffusion of Innovation and policy decision-making. Journal of Communications **43**, 30–45 (1993)
9. Knoke, D., Kuklinski, J.H.: Network analysis. Sage, Newbury Park (1982)
10. Burt, R.S., Minor, M.J.: Applied Network Analysis. Sage, Newbury Park (1983)
11. Wellman, B.: Structural analysis: From method and metaphor to theory and substance. Contemporary Studies in Sociology **15**, 19–61 (1997)
12. Scott, J.: Network Analysis: A Handbook. Sage, Newbury Park (1991)

13. Rogers, E.M.: Categorizing the adopters of agricultural practices. Rural Sociology **23**, 345–354 (1958)
14. Beal, G.M., Bohlen, J.M.: How Farm People Accept New Ideas. Cooperative Extension Service Report 15, Ames, IA (1955)
15. Valente, T.W., Rogers, E.M.: The origins and development of the diffusion of innovations paradigm as an example of scientific growth. Science communication **16**(3), 242–273 (1995)
16. Anderson, J.G., Jay, S.J.: The diffusion of medical technology: Social network analysis and policy research. The Sociological Quarterly **26**(1), 49–64 (1985)
17. Mahajan, V., et al.: Determination of adopter categories by using innovation diffusion models. Journal of Marketing Research, 37–50 (1990)
18. Big Data Analytics, Time-Series Database, Kx Systems. http://www.kx.com
19. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, vol. 1215 (1994)
20. Don Josef de Mendoza y Rios, F.R.S.: Recherches sur les principaux problemes de l'astronomie nautique. In: Proceedings of the Royal Society (1796)
21. Johansson, J.: On the Usability of 3D Display in Parallel Coordinates: Evaluating the Efficiency of Identifying 2D Relationships (2013)
22. Bostock, M.: Parallel Coordinates, November 5, 2011
23. Parallel coordinates (0.7.0) A visual toolkit for multidimensional detectives
24. Nutrient Contents - Parallel Coordinates. http://www.exposedata.com
25. U.S. Population By Region (1990–2010). http://www.infoplease.com
26. Heidari, B.: Knee osteoarthritis prevalence, risk factors, pathogenesis and features: Part I. Caspian Journal of Internal Medicine **2**(2), 205–212 (2011)
27. The White House, authors. Transforming Health Care: The President's Health Information Technology Plan, December 16, 2004. http://www.whitehouse.gov/infocus/technology/economic_policy200404/chap3.html