Paper 14-2017

# Enhancement of Clinical Decision Making using Predictive Modeling

Archana Chinnaswamy, Oklahoma State University, OK, USA


## ABSTRACT

With the mandate of Electronic Health Records (EHR), there is an abundance of data available for every member in the healthcare supply chain. One big question that has remained is, "How effectively can we transform this data into a valuable resource for improving the quality and cost of healthcare? "The problem is not inadequacy of health data, but lack of a quantifiable health assessment tool, that will not only assess a patient's health status but also predict the patient's future visits to a hospital and potential future patient costs.

This paper aims to propose a quantifiable ranking tool based on predictive modeling to aid in improved clinical decision making. Data of over 300,000 AI/AN patients were analysed, cleaned, and transformed to predict future hospital visits and hospital costs. Using predicted numbers and segmentation techniques, patients were categorized into different risk levels. This information can help health-care providers provide quality healthcare. In addition, this invaluable data can be used to bring targeted health awareness programs specially to underserved communities. This data is also useful for hospital administrators, who can better plan for staffing and promotions targeted towards American Indian and Alaskan Native (AI/AN) patient wellness and insurance expansions.


## INTRODUCTION

Healthcare providers and patients in the United States have benefited from recent progresses in the fields of data collection, data management, and data sharing [1]. With the mandate of Electronic Health Records (EHR), there is an abundance of data available for every member in the healthcare supply chain. Nurses, practitioners, physicians, doctors, pharmacists, family care providers, insurance agencies, and several others have access to this vast range of data. While data sharing and exchange have helped patient care and health care providers, one big question that has remained is, "How effectively can we transform this data into a valuable resource for improving the quality and cost of healthcare? ". This has given birth to the field of healthcare analytics. Use of predictive models and advanced data science tools to provide a thorough analysis with direction, has established the importance of this field. As opposed to only looking at facts and figures, it is imperative that historical data, be used to forecast useful information that can help hospitals; health systems and most importantly patients [2]. Considerable mortality disparities have been reported in recent years, especially in the American Indian and Alaskan Native (AI/AN) communities. AI/AN population is reported to have a life expectancy - that is 4.4 years less than all other races in the U.S. population (ages 73.7 vs. 78.1) [3].

Health care experts, tribal leaders, and policy makers have been looking into adequacy of funding, health awareness initiatives, and outreach methodologies, special service eligibilities, to improve this status. Several Organizations [such as the Indian Health Service (IHS)] that provides care for 566 AI/AN tribes representing about 2.2 million AI/AN natives, have been working on this issue. The problem is not inadequacy of health data, but lack of a quantifiable health assessment tool, that will not only assess a patient's health status but also forecast the patient's future visits to a hospital and potential future patient charges. This type of information, in addition to aiding future administrative plans will also help doctors provide better healthcare advice, enhancing decision making. Pharmacy and physician visits also could be cut down. All of this will have a cumulative impact on patient well-being and efficiency of the healthcare system. Availability of such data and a ranking tool as a result, can also help target wellness programs for under-served communities (such as AI/AN). Such a tool can also provide logical bases to implement effective methodologies towards AI/AN health awareness.

One fourth of all healthcare budget expenses go towards administrative costs [4]. This is a proof that, there is a room for significant improvement to cut down costs and improve operational efficiency. Most hospital administrations schedule their operational capabilities based on historical trends in patient volume and acuity. Recent advances in healthcare analytics however, have helped make better administrative decisions improving efficient and cutting down on overhead. For example, Yale-New Haven Health System (YNHHS) has reported 150 million dollars in savings, with the help of predictive analytics [5]. Other quantifiable data have been reported signifying the benefit of predictive analytics [6]. Innovations in data discovery methods and analytical tools have aided such tremendous progress. Healthcare analytics industry size is estimated at 5 billion dollars (2016) and is projected to grow to 19 billion dollars over the next 3 years. This includes key players' such as Cerner, McKesson, IBM, Allscripts, and Optum [7]. This is almost 50% of the size of medical devices industry. This is evidence, corroborating the importance of healthcare analytics and the data-driven culture adapted by the healthcare industry. One problem that industry has been addressing and continually trying to improve is, "the effective use the analytical tools to enhance clinical decision-making and cut down administrative costs". Over 80% of the collected data has been reported as "unstructured" and the number is growing continuously. Reported literature has alluded to a multitude of areas where analytics can help improve healthcare costs and quality.

1. Making use of historically available clinical data and factoring them to make informed clinical decisions.
2. Provision of a more complete view and valuable insights into patient care coordination to identify and engage high-risk patients into tailored awareness and assessment programs.
3. Reduction of healthcare costs: cut-down avoidable overuse from an administrative and patient's standpoint [7].

## METHODOLOGY

Data for over 254,900 AI/AN patients were provided in a spreadsheet format. De-identified patient data including – hospital information, visit information, diagnosis information and drug information were also provided.

**Method 1: Predictive modeling using "JMP" tool to identify high risk patients based on number of hospital visits.**

A predictive model was developed to forecast the number of future visits for existing patients. As a first step, SAS Enterprise Guide was used to clean and transform the provided data. JMP tool was used to execute the model. The provided data was grouped based on patient identity numbers for the years 20002015. A multiple regression model was used to predict the number of future visits. Based on a detailed literature review and examination, the most relevant variables were selected to perform this analysis. The selected variables include - age, total charges, diagnostic information, drug information, Blood pressure, resp-rate, bun-ratio, corpuscular hemoglobin, platelet-volume, glucose, hematocrit, platelet-count, RBC-count, WBC-count. These are regularly collected values for inpatient in many hospitals.

**Method 2: Predictive modeling using "SAS Enterprise Miner" for assessment of a patient's risk level based on patient charges.**

SAS Enterprise Miner was used to sort, segment and transform the provided data. The cleaned data was then used to build a predictive model using decision tree technique. The data was segregated sequentially. Five different buckets were created to segregate patient charges and visits. These were categorized based on the required level of medical attention and potential clinical wellness recommendation.

**Table 1 Proposed Clinical Interpretation Terms**

| Defined level | Predicted total patient charges | No. of predicted patient visits | Medical attention required | Clinical wellness interpretation |
|---|---|---|---|---|
| 1 | < $3,000 | < 3 | Very low | Limited requirements |
| 2 | $3,000 - $10,000 | 3-5 | Low | Minimal observations |
| 3 | $10,000 - $20,000 | 5-10 | Moderate | Candidate for wellness programs |
| 4 | $20,000 - $60,000 | 10-20 | High | Candidate for disease management program |
| 5 | > $60,000 | > 20 | Very high | Candidate for case management program that require continuous observation. |

Based on the results from the predictive model, both accuracy and penalty errors are reported. Both these models and algorithms can be integrated in the backend of existing electronic dashboards currently available to providers today.

## RESULTS

The two methods in previous section have been proposed to predict:

1. Number of future hospital visits for a patient
2. Total future charges from a patient
3. Combining a) and b), a patient's risk level, recommend appropriate health programs.

This can be done with a help of a web interface. This interface will consist of a series of patient information and their health history. As soon as the health administrator enters the required information, it populates the entered information as an input for the proposed model. The model then produces the required output, which consists of 3 key parameters.

- Risk level of the patient (very low – very high), assigning a rank
- Rank based health program recommendations
- Forecasted number of future visits
- Total future charges from the patient

An example snapshot, of the web interface (input) and the front-end (results) that a healthcare administrator can view is presented in the figure1.

**Figure 1 Web Interface for Patient Health Risk Assessment**

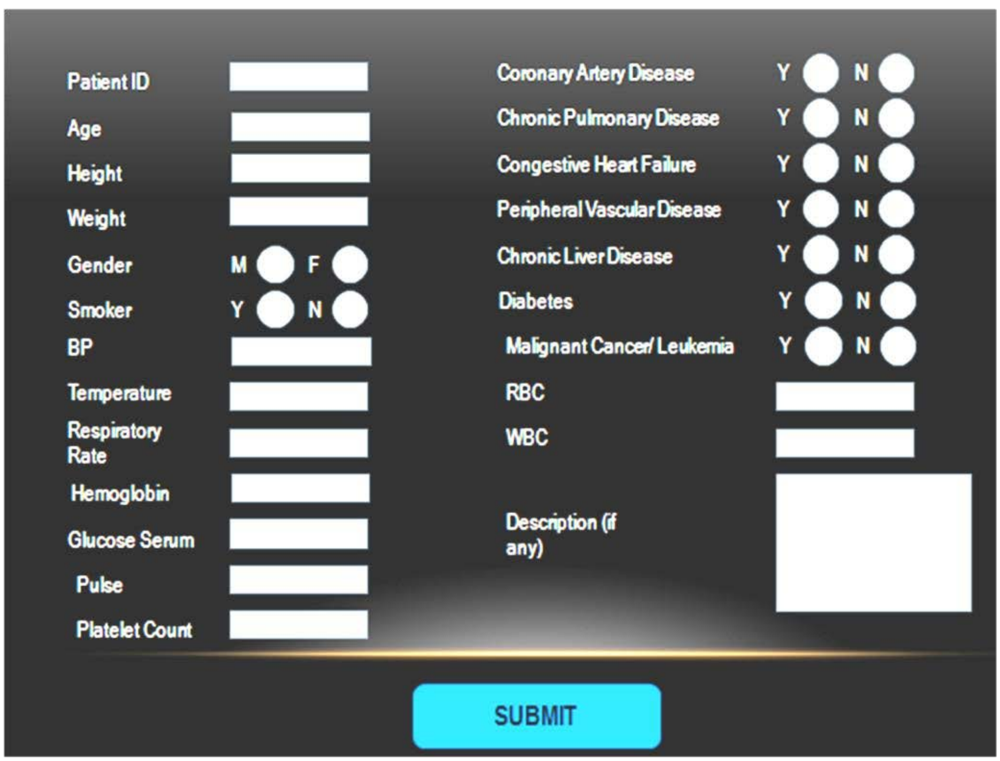
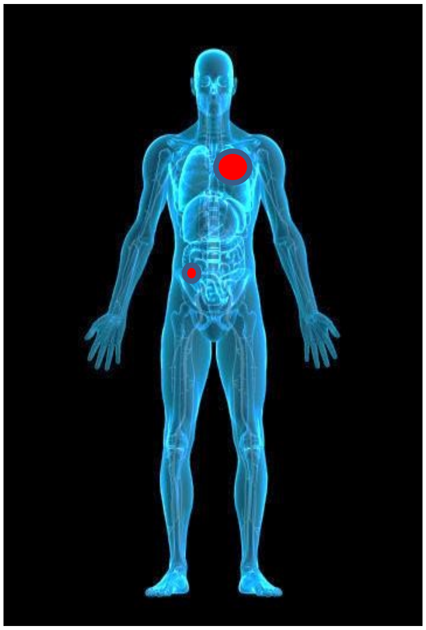


**Figure 2 Health Risk Assessment Output**



| Risk | Moderate |
|---|---|
| Projected hospital visits | 10 |
| Projected patient charges | $15,000-$20,000 |
| Recommended health programs | Wellness Programs |

The web interface was built using Python and Flask. The inputs from the form is passed to the model built and the risk levels are identified. Results from the predictive model have an accuracy of 66% (moderate strength). With additional data and availability of more independent variables, the accuracy of this model can be improved.

A penalty matrix was developed in assessing the patient's health rank levels and cost buckets. The proposed model reduced the penalty error from 88% (baseline) to 72%. Note that the baseline accuracies were calculated by adopting the proposed analytical model and the presented data (2000-2015).

**Figure 3 Output from Decision Tree**

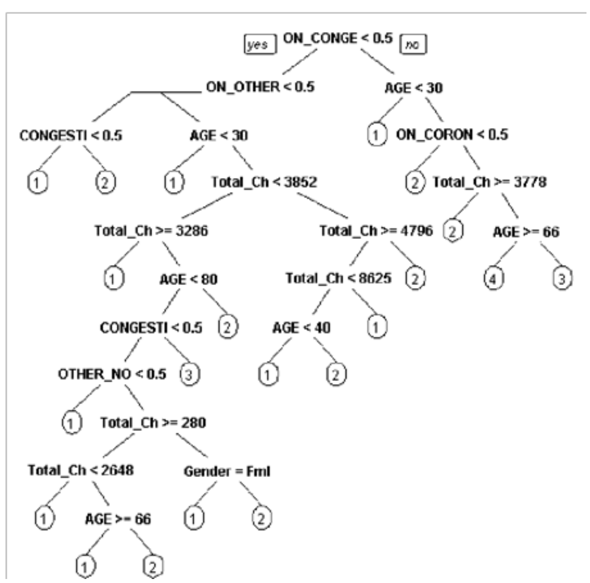


Figure 3 and 4 shows the variables selected using decision tree model and stepwise regression model. We can see that cost, age, gender, and chronic illness plays a significant role in predicting the number of hospital visits.

**Figure 4 Output from Stepwise Regression**

Step History

| Step | Parameter | Action | "Sig Prob" | Seq SS | RSquare | Cp | p | AICc | BIC | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | RBC_COUNT1 | Entered | 0.0000 | 132537.7 | 0.1319 | 14439 | 2 | 325517 | 325544 | ○ |
| 2 | BP_SYS | Entered | 0.0000 | 43779.43 | 0.1754 | 10766 | 3 | 322492 | 322528 | ○ |
| 3 | PERIPHERAL_VASCULARDISEASE | Entered | 0.0000 | 34016.96 | 0.2093 | 7912.7 | 4 | 320029 | 320074 | ○ |
| 4 | CHRONIC_PULMONARYDISEASE | Entered | 0.0000 | 19304.33 | 0.2285 | 6294.3 | 5 | 318585 | 318639 | ○ |
| 5 | CONGESTIVE_HEARTFAILURE | Entered | 0.0000 | 14985.9 | 0.2434 | 5038.4 | 6 | 317440 | 317503 | ○ |
| 6 | MALIGNANT_CANCERLEUKEMIA | Entered | 0.0000 | 12241.84 | 0.2556 | 4012.9 | 7 | 316488 | 316559 | ○ |
| 7 | ON_MALIGNANT_CANCER_LEUKEMIADRUG | Entered | 0.0000 | 7920.941 | 0.2634 | 3350 | 8 | 315864 | 315945 | ○ |
| 8 | RENALFAILURE | Entered | 0.0000 | 7445.928 | 0.2708 | 2727 | 9 | 315271 | 315361 | ○ |
| 9 | SEVERE_CHRONICLIVER_DISEASE | Entered | 0.0000 | 5383.536 | 0.2762 | 2277.2 | 10 | 314840 | 314939 | ○ |
| 10 | CORONARY_ARTERYDISEASE | Entered | 0.0000 | 4102.27 | 0.2803 | 1934.8 | 11 | 314509 | 314617 | ○ |
| 11 | GENDER{Unknown/Invalid&Male-Female} | Entered | 0.0000 | 3879.416 | 0.2841 | 1611.2 | 12 | 314195 | 314312 | ○ |
| 12 | ON_SEVER_CHRONIC_LIVER_DISEASEDRUG | Entered | 0.0000 | 2832.803 | 0.2870 | 1375.4 | 13 | 313965 | 314091 | ○ |
| 13 | GLUCOSE | Entered | 0.0000 | 2443.713 | 0.2894 | 1172.3 | 14 | 313766 | 313901 | ○ |
| 14 | DIABETES_W_END_ORGANDAMAGE | Entered | 0.0000 | 1758.181 | 0.2911 | 1026.7 | 15 | 313623 | 313767 | ○ |
| 15 | BUN_RATIO1 | Entered | 0.0000 | 1724.632 | 0.2929 | 883.96 | 16 | 313483 | 313635 | ○ |
| 16 | ON_OTHERDRUG | Entered | 0.0000 | 1594.153 | 0.2944 | 752.15 | 17 | 313353 | 313514 | ○ |
| 17 | RESP | Entered | 0.0000 | 1793.158 | 0.2962 | 603.63 | 18 | 313206 | 313376 | ○ |
| 18 | ON_CHRONIC_PULMONARY_DISEASEDRUG | Entered | 0.0000 | 1527.436 | 0.2977 | 477.43 | 19 | 313081 | 313260 | ○ |
| 19 | WBC | Entered | 0.0000 | 1059.642 | 0.2988 | 390.48 | 20 | 312994 | 313183 | ○ |
| 20 | BP_DIAS | Entered | 0.0000 | 1031.821 | 0.2998 | 305.87 | 21 | 312910 | 313108 | ○ |
| 21 | BP_SYS | Removed | 0.9418 | 0.063896 | 0.2998 | 303.88 | 20 | 312908 | 313097 | ○ |
| 22 | ON_RENAL_FAILUREDRUG | Entered | 0.0000 | 408.2492 | 0.3002 | 271.61 | 21 | 312876 | 313074 | ○ |
| 23 | ON_DIABETES_W_END_ORGAN_DAMAGEDRUG | Entered | 0.0000 | 353.5776 | 0.3006 | 243.93 | 22 | 312848 | 313055 | ○ |
| 24 | WEIGHT1 | Entered | 0.0000 | 310.814 | 0.3009 | 219.84 | 23 | 312824 | 313040 | ○ |
| 25 | DEMENTIA1 | Entered | 0.0000 | 301.9718 | 0.3012 | 196.5 | 24 | 312801 | 313026 | ○ |
| 26 | RBC_DISTR | Entered | 0.0000 | 226.5104 | 0.3014 | 179.48 | 25 | 312784 | 313018 | ○ |
| 27 | CORPUSCULAR_HEMOGLOBIN1 | Entered | 0.0000 | 891.9066 | 0.3023 | 106.62 | 26 | 312711 | 312954 | ○ |
| 28 | HEIGHT1 | Entered | 0.0000 | 257.47 | 0.3026 | 87.009 | 27 | 312692 | 312943 | ○ |
| 29 | PLATELET_VOLUME1 | Entered | 0.0000 | 201.553 | 0.3028 | 72.091 | 28 | 312677 | 312937 | ○ |
| 30 | HEMATOCRIT1 | Entered | 0.0000 | 221.0683 | 0.3030 | 55.535 | 29 | 312660 | 312930 | ○ |
| 31 | TOT_CHARGES | Entered | 0.0002 | 171.1163 | 0.3032 | 43.172 | 30 | 312648 | 312926 | ○ |
| 32 | PLATELET_COUNT1 | Entered | 0.0192 | 65.34261 | 0.3032 | 39.687 | 31 | 312645 | 312932 | ○ |
| 33 | ON_DEMENTIADRUG | Entered | 0.0507 | 45.5082 | 0.3033 | 37.867 | 32 | 312643 | 312939 | ○ |
| 34 | OTHER_NONCHRONICDISEASE | Entered | 0.0675 | 39.83167 | 0.3033 | 36.524 | 33 | 312641 | 312947 | ○ |
| 35 | ON_CONGESTIVE_HEART_FAILUREDRUG | Entered | 0.1392 | 26.058 | 0.3033 | 36.337 | 34 | 312641 | 312955 | ○ |
| 36 | ON_CORONARY_ARTERY_DISEASEDRUG | Entered | 0.0358 | 52.47882 | 0.3034 | 33.932 | 35 | 312639 | 312962 | ○ |
| 37 | ON_PERIPHERAL_VASCULAR_DISEASEDRUG | Entered | 0.1060 | 31.12258 | 0.3034 | 33.32 | 36 | 312638 | 312970 | ◉ |

## DISCUSSION

The major aim of this work is to predict the future visits for a new patient based on current diagnostics and vital signs. This has been achieved by using a predictive model built on historical data. The availability of this information will provide valuable insights to hospitals and the physicians treating the patient. Decisions like referring to a specialist, ordering lab tests, and scheduling follow up appointments can be made by physicians based on the insights provided by this model. The hospital can also use this information to plan their resources and prepare of the patient's future visits.

For existing patients, the model will predict the cost, which a patient can expect to incur in the coming year. Several factors like existing ailments, number of visits, age, existing costs, and medications are used in determining the future expenditure. The estimated future cost will help set financial expectations for the patient and will aid in financial planning. Insurance companies can minimize claims by collaborating with hospitals for preventive health plans such as wellness programs, specialist check-ups, disease management, etc. These could also serve as marketing campaigns for the insurance providers.

This new method to predict future cost and number of visits will be useful to healthcare decision makers - ultimately resulting in efficient patient care and wellbeing.

Most of the tools used in this work are "open source" tools such as Python and are easily available (at no cost). It can be easily implemented with a web interface and algorithms. The Python web application can also be deployed freely using Dokku, Django, etc. Future research is required with higher variance in data and more variables to increase the accuracy of the model.

## LIMITATIONS

1. The lab values are averaged across all values for the same lab test performed in that particular encounter which could affect the significance of the results.
2. The diagnoses were grouped into categories based on the Dartmouth Atlas [9]. This is one classification system for reducing the number of levels in diagnoses codes. Other systems that are more popularly referenced in literature include the CCS and Charleson comorbidity indexes.
3. The drugs were similarly grouped based on a classification system provided by drugs.com. It is possible that the same drug could show up more than once in each of the different categories because a drug can have multiple indications. Without additional knowledge, we are not able to clearly identify the reason why a drug was prescribed.
4. The variable 'total_charges' is not a true representation of the actual cost of hospital stay. The total charges were null in most cases because the hospital charges might be covered under some agreement between the hospital and the payer.

## CONCLUSION

1. Using the data provided on AI/AN patients, a new predictive model was developed to forecast the number of future patient visits and projected patient charges. Risk levels were created based on patients' spending history and health statuses.
2. A web interface and a resulting visual of an individual patient's health assessment are proposed.
3. Health-care providers can bring targeted health awareness programs.
4. Programs such as the Indian Health Service (The Federal Health Program for American Indian and Alaskan Natives) can benefit by providing targeted wellness programs to AI/AN patients and communities.
5. Hospital administrators can better plan for staffing and promotions targeted towards AI/AN patient wellness and insurance expansions.

## REFERENCES


[1] Health Catalyst "*Healthcare Analytics Adoption Model*" https://www.healthcatalyst.com

[2] Kelsey Brimmer, 2013 Healthcare IT news http://www.healthcareitnews.com

[3] Indian Health Services, "Mortality Discrepancy Rates Fact Sheet March 2016". https://www.ihs.gov/newsroom/factsheets/disparities/

[4] Kelsey Brimmer"2013 Healthcare IT news" [accessed Oct 01 2016] http://www.healthcareitnews.com

[5] Daily Briefing May 2016 Advisory Board "*How predictive analytics helped Yale-New Haven save $150 million*" https://www.advisory.com/daily-briefing/

[6] Carol McDonald "How Big Data is Reducing Costs and Improving Outcomes in Health Care" https://www.mapr.com/blog/reduce-costs-and-improve-health-care-with-big-data

[7] Markets and markets "*Healthcare Analytics Market worth $18.7 Billion by 2020*" *http://www.marketsandmarkets.com/PressReleases*

[8] Linda A. Winters-Miner "Seven ways predictive analytics can improve healthcare" [6 October 2014]

[9] Dartmouth Atlas of Health Care, 2014 *http://www.dartmouthatlas.org/data/topic/topic.aspx?cat=21*


## ACKNOWLEDGEMENT


I thank Dr. William Paiva and Ms. Elvena Fong of Center for Health Systems Innovation at Oklahoma State University for granting me permission to publish my findings. Special thanks to Dr. Goutam Chakraborty and Dr. Miriam McGaugh of Business Analytics department at Oklahoma State University for their continuous support and guidance.


## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:


Archana Chinnaswamy
MS in Business Analytics
Oklahoma State University
Phone: 405-338-8705
Email: archana.chinnaswamy@okstate.edu
LinkedIn: www.linkedin.com/in/archanachinnaswamy/


Archana Chinnaswamy is a graduate student enrolled in Business Analytics at the Spears School of Business, Oklahoma State University. She has been working on healthcare analytics projects with the Center for Health Systems and Innovation, as a Graduate Research Assistant since Jan 2017.

This work was conducted with data from the Cerner Corporation's Health Facts database of electronic medical records provided by the Oklahoma State University Center for Health Systems Innovation (CHSI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and not necessarily reflect the views of the CHSI.