Reduction of Clinic Office Visit Costs By Grouping Patients Into 6 to 8 Small Groups and Using One 45-60 Minute Office Visit Slot Instead of 6 to 8 Individual Visit Slots

Executive Summary

Rising costs of health care in the United States has contributed to 19.7% of GDP (Gross Domestic Product) in 2021 resulting in $12,530 per person in medical expenses per CMS (2022). The highest costs of all other countries. Between 2015 and 2016 clinic needs rose 51% according to Kacik 92016), which means we have to find better solutions. The proposal is a threefold effect with the primary benefit being reduction of overall clinic appointment costs. 1)Reduce overall clinic appointment cost 2) Increase clinic hour availability of providers 3) Provide better health outcomes. Using CMS data for office costs, using coronary heart disease data set from Kaggle in which there will be randomly added social determinant answers as well as patient ages (if not present) to a coronary heart disease patient with the intention of grouping programmatically, 6-8 patients into a SMA appointments. Currently there have been research papers addressing utilization of SMA's to address cost or patient outcomes, however, there has not been a paper that combines cost, patient outcome, availability of provider with commonalities in diagnosis and (social determinants to include mocked up answers randomly applied to a set of coronary disease patients for financial resource strain, transportation needs, food insecurity, housing stability, stress and depression and social connections). This mocked up data set will hopefully provide a catalyst for the healthcare community to build models that addresses the three-pronged question of delivering healthcare in a conscientious effort to reduce cost all around, increase provider availability for other patients, and provide positive health outcomes to the patients grouped together for the SMA's.

By asking ourselves if there is a significant advantage by grouping patients together in groups of 6 to 8 for clinic visits in regards to cost we can reduce the cost and free up slots of clinic hours for other patients to receive care. We can also significantly impact patient care based on Social Determinants of Health in that patients will be like in needs of care. One other impact that we can make in significance is facilitating patient bonding that can go beyond the small group medical visit into the social connections in their home areas.

To conduct our proof of concept in grouping patients together we used the Coronary Heart Disease Patient Data set available on Kaggle.com as well as CMS General Office Appointment Cost Dataset for general office appointment costs. And we used the ICD-10-CM/PCS MS-DRG v37.0 Definitions manual for Diseases & Disorders of the Cicurlatory System from CMS.gov to randomly assign a DRG to the 70,000 cardio patient data set from Kaggle.com.

The DRG's we will randomly assign the 70,000 patients are:

Acute Myocardial Infarction	280
Acute Myocardial Infarction	281
Acute Myocardial Infarction	282
Hypertension	304
Coronary Bypass	231
Coronary Bypass	232
Atherosclerosis	302

The available datasets were combined and/or randomly assigned Social Determinants of Health based on the Virginia Commonwealth University Health System's Social Needs Assessment Form (Virginia Commonwealth University Health System, 2017) as a template and the AAFP (American Academy of Family Physicians, 2019) Guide to social needs screening to help formulate questions.

**Tobacco - Former smoker, how many years?

Tobacco - smokeless

Tobacco - e-cig/vape

**Demographics - divorced, separated, life partner, married, significant other, sinle, unknown, widowed

**Children- number of children

**Education- years of education

**Financial Resource Strain - How hard is it for you to pay for the very basics like food, housing, medical care, heating?

Very Hard, Hard, Somewhat hard, not very hard, not hard at all

**Transportation Needs - do you need help getting to medical appointments?

Stress

**Housing Stability - do you need help paying your mortgage

Intimate Partner Violence

Alcholhol Use

**Food Insecurity - Last 12 months have you worried that your food would run out before you got more money to buy more?

**Physical Activity --On average how many days per week do you engage in moderate to strenuous exercise a week?

On average, how many minutes do you engage in exercise at this level?

**Depression - never, sometimes, often, always

**Social Connections - In a typical week how many times do you talk in the phone with family, friends, or neighbors?

How often do you get together with friends or relatives?

How often do you attend church or religious services? Never, 1 to 4 a year, more than 4 a year,

Do you belong to any clubs or organizations? Yes, No

How often do you attend meetings of clubs or organisations? Never, 1 to 4 more than 4 a year

Are you married, widowed, divorced, separated, single or living with someone

These culminated in the following categories of assignments:

Tobacco Use, Marital Status, Housing Stability, Food Stability, Depression Status and Transportation Needs which will aide in deciding what future small group medical visits can be appropriated to.

The data included 24 variables and 70,000 rows. Splitting training and testing data was into an 80 training and 20% testing of the data set used for model building and a

validation set used for model validation. The set of 24 independent variables included continuous, discrete and nominal categorical types related to the randomly assigned data for patients. Patient attributes included, age, height, weight, blood pressure, cholesterol, glucose, tobacco use, alcohol use, physical activity, randomly assigned drg id's, randomly assigned SDOH's, insurance cost and patient co-pay cost. The target variable was the patient attribute cardio which was a binary true/false variable.

Primary Component Analysis (PCA) which was further validated by using CHI testing on contingencies was used to identify the statistically significant independent variables and interaction effects, build the model, and score new observations. After the PCA was performed it was determined that tobacco and gender were not giving us enough evidence to reject the null hypothesis at the $\alpha=0.05$ level. The data suggest that the two categorical variables are not independent.

After the PCA and CHI testing was performed a KNN (K-Nearest Neighbor) model to cluster the data was used with a parameter of 6. However, this resulted in a 58% accuracy in the model. Standard Scalar was then applied to the model to improve the accuracy of the model however, it too only produced a 58% accuracy. To which then we asked the KNN model to be assessed by machine language programming to suggest what parameter fit better. The result was also 58% accuracy. So all 3 models of the data resulted in 58% accuracy.

The limitations of the data is first and foremost that it was a proof of concept: Can we significantly impact patient care, reduce costs and free up visit slots for providers using the data we found? The dataset does not contain nor represent all 40,600 outpatient clinics entire population according to Modern Healthcare (2016). Since the data is six years old, there are outpatient clinics not in this total. It only contains information from the Kaggle derived data set of 70,000 rows to which 35,021 of those were declared cardio only patients leaving 34,979 non-cardio patients. In addition, clinics typically treat many other patients with many different types of disease and illnesses and who may have more than one

medical issue, this study only looks at eight DRG (ICD10) diagnosis codes for cardiac diagnosis.

There are 45,330 males and 24,670 females in the resulting patient list. From a PCA (not normalized) we see that males median age is in the upper fifties and females median age is in the forties.

Some suggestions for future research include but are not limited to the following:

- using actual patient data gathered from our clinics
- cleaning our data
- using different clustering techniques

The most important benefit of this analysis is that it provides a way of automatic identification of patients that could be grouped together for small medical clinic visits. This would not only allow patients to receive targeted care, but targeted healthcare in small groups of patients that have similar healthcare needs, social determinant needs, and can foster out of clinic relationships between small grouped patients. Healthcare organizations collect large amounts of data about their patients through EMR/EHR applications. These data allows us the opportunity to meet the purpose of this proof of concept and its three pronged proposal with the primary benefit being reduction of overall clinic appointment costs. 1)Reduce overall clinic appointment cost 2) Increase clinic hour availability of providers 3) Provide better health outcomes

The findings of this study have the potential to begin working on real patient data to drive the initiatives our healthcare community needs and invites to conversation to approach negotiating with healthcare insurance companies better rates for Small Medical Visit Group appointments and smaller-co-pays for our patients.

References:

AAFP (2019). Social determinants of health: A guide to social needs screening. American Academy of Family Physicians

https://www.aafp.org/dam/AAFP/documents/patient_care/everyone_project/hops19-physician-guide-sdoh.pdf

CMS (2022). National Health Expenditure Data. Historical.

https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical

Kacik (2016).Number of outpatient facilities surges as industry values more convenient, affordable care. Modern Healthcare.
https://www.modernhealthcare.com/article/20181220/NEWS/181229992/number-of-outpatient-facilities-surges-as-industry-values-more-convenient-affordable-care

Virginia Commonwealth University Health System (2017).. VCU-Health-Social-Needs-Assessment_102517.pdf. Virginia Commonwealtth University Health System.

https://www.chcs.org/media/VCU-Health-Social-Needs-Assessment_102517.pdf

DATA ACQUISITION:

Kaggle Coronary Heart Disease Dataset:
https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset

CMS General Office Appointment Cost Dataset:

https://data.cms.gov/provider-data/dataset/2e55-8767

CMS.gov ICD-10/PCS MS-DRG v37.0 Definitions manual for Diseases & Disorders of the Circurlatory System Dataset:

https://www.cms.gov/icd10m/version37-fullcode-cms/fullcode_cms/P0001.html

Panapto recording:https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=f415b825-b0d5-45c4-98f4-af2201682662