Predicting Incidents from Residential Aged Care Homes

Author :
Gayatri Aniruddha
Student ID: 30945305

Primary Supervisor :
Dr. Anuradha Madugalla

Co-Supervisors:
Prof. Jennifer McIntosh and Dr. Humphrey Obie

MASTER THESIS REPORT

Monash University
Faculty of Information Technology
FIT5210 - Masters Thesis
June 2021

Word Count : 6808

Acknowledgements

Throughout the writing of this thesis, I have received a great deal of support, assistance and guidance.

I would like to take this opportunity, first to thank my primary supervisor Dr. Anuradha Madugalla, and my co-supervisors, . Jenifer McIntosh and Dr. Humphrey Obie, for their support, guidance and feedback throughout the project.

I would also like to thank my Thesis Coordinator, Dr. Reuben Kirkham and Assistant Coordinator Tian Goh for valuable guidance throughout my thesis.

In addition, I would like to thank my parents, my baby sister and my grandparents for their constant motivation and strength throughout.

This thesis report could not have been completed without the support of my fellow student researchers, with whom I was able to have discussions regarding the reports and formatting. In addition, I would like to thank my housemates who provided me with happy breaks and game sessions to rest my mind, outside of my research.

Finally, I would like to thank and acknowledge the Faculty of IT, Monash University for giving me this opportunity to conduct this valuable research.

TABLE OF CONTENTS

Part 1: General Literature Review **4**

1.1 Introduction 4

1.2 Substantive Literature Review 5

1.3 Summary of the State of the Art 10

1.4 Plan for your research project 11

1.5 Conclusion 12

Part 2: The Research Paper **14**

2.1 Introduction 14

2.1.1 Contributions of the Research Project 14

2.1.2 Thesis Report Organisation 15

2.2 Background 16

2.3.1 Overview of the Elderly Population in Healthcare 16

2.3.2 Need and motive of the research work 16

2.3 Methodology 17

2.3.1 Aim 17

2.3.2 Design 18

2.3.3 Tools Used and Steps Followed 18

2.3.4 Data Source and Study Variables 19

2.3.5 Model Development 19

2.3.6 Statistical Explorations, Observations and Analysis 20

2.4 Results and Discussions 24

2.4.1 Results 25

2.4.2 Discussions 26

2.5 Limitations and Future Work 27

2.6 Conclusion 29

2.7 References 30

Part 3: Appendices **34**

Part 1: General Literature Review

1.1 Introduction

This project revolves around Predicting Hospital Admissions of Residents from Australian Aged Care Homes, in collaboration with Health Metrics. Health Metrics was established in 2008. Health Metrics's software - eCase helps in supporting the functionality of Residential Aged Care, Home Care, Disability Care and Primary Health Care. With each passing day, eCase is becoming one of the fastest growing choices of leading healthcare providers in the markets of Australia and New Zealand. eCase provides software solutions and supports the clinical, financial and operational aspects of Residential Aged Care, Retirement Living, Disability Support and Community Living.

There are around three main sectors of Health Care provided which includes Homecare, Residential Care and Disability Care. Home Care is also commonly known as in-home care, home help, home support and in-home nursing care. In Home Care, staff are sent at home to provide the required care and support so that older people can live in the comfort of their own homes. Home Care allows the recipients to experience independence and choice with respect to the services they recieve. Most older people prefer these services over Residential Aged Care. Some of the most common services provided by Home Care include bathing, dressing, laundry, gardening, preparing meals, nursing, maintenance of home, transportation and therapy. These services can be classified into basic level, low level, intermediate level and high level needs and receive funding from the government.

Residential Care refers to the services and care provided when you are living in a care facility, which includes Aged Care Homes. Residential Care is for older people who can no longer live at home due to various reasons such as illness, disability or because they can no longer live at home without any help. This includes Nursing Homes and Retirement Villages. Nursing Homes provide different types of personal and nursing care on a temporary and permanent basis. Residential Aged Care provides help with day to day tasks, personal care, provides access to health services, clinics and therapies and also provides other services such as emotional support, social life and entertainment. The Australian Government funds these Residential Aged Care Homes to make them more accessible and affordable for everyone. Disability Care is a combination of the two where workers provide support for older people with disabilities in their home or in Residential Care centers.

The elders present in these Aged Care Homes are definitely at a high risk for Hospital Admissions. These Hospital Admissions are due to various risk factors such as age, lifestyle, meteorological factors, history of diseases, multimorbidity etc and can also take place due to various incidents, the most common incident being falls. This first step of my project revolves around researching, identifying and analysing the various factors that are responsible for causing these Hospital Admissions. (O'Caoimh, Cornally, Weathers, O'Sullivan, Fitzgerald, Orfila & Molloy, 2015). These factors were identified after going through 30 Research Papers which gave me an insight into the different research work which was already done in this field. Based on my reading and analysis, the various common factors that lead to hospital admissions are age, lifestyle, prior cases of hospitalisations, existing medical conditions, multimorbidity, financial conditions, social and/or family support and meteorological factors such as pollution. (Cornette, D'Hoore, Malhomme, Van Pee, Meert & Swine, 2005)

After identifying these factors, the second part of my project involves looking into the various models which have already been developed in order to predict these admissions. In order to find out the importance of these factors while predicting hospital admissions, a common method included p value analysis and c statistics analysis. (Zhang, Kim, Patzer, Pitts, Patzer & Schrager, 2017). After going through the different Research work, some of the common models implemented included Stepwise Linear Regression and Multivariable Logistic Regression. (Hong, Haimovich & Taylor, 2018)

Thus, the objective of the Literature Review was to get a better understanding of the various contributing factors, importance of these factors and models implemented to predict the Hospital Admissions based on these factors. Furthermore, after the identification of the factors and models implemented, this report revolves around the various limitations and gaps present in the existing methodologies and surveys provided. Finally, based on the gaps identified, this report finally throws light on the various tasks needed to be implemented in this project and the proposed timeline of completion.

1.2 Substantive Literature Review

This section revolves around the various papers which have been explored in detail and thoroughly analysed in order to find out the range of the research which was conducted in the research area and to find out what the existing research conveyed. Since my project aims at predicting the hospital admissions, there were two key points to look at from the papers, namely the factors responsible for hospital admissions and the models used to predict the admissions. The range of the research conducted in this area revolved around the data used to determine the factors responsible for hospitalisation, the target audience and age group of the patients involved in the research, the location of the research, time frame of the research conducted and finally the number of patients involved in the study. The findings of the methodology followed, i.e the models developed, results obtained from the study, and the subsequent discussion and analysis.

There has been a diverse range of data collected for the research work. In order to analyse the factors contributing to the hospital admissions, studies have been conducted on a wide range of audience, ranging from the entire population to specific age groups such as above 60+ age group, 80+ age group etc. Age was one of the most common factors which determined the hospital admissions. (Lyon, Lancaster, Taylor, Dowrick & Chellaswamy, 2007).

Another distinguishable factor was the air pollution levels which are a major cause of Respiratory Diseases, which in turn have chronic effects on the health of older people and lead to admissions, generally in the emergency department. In order to understand the relationship between the air pollution levels and the corresponding hospital admissions, research was conducted on the entire population of Greece, over a period of eight years from 1992-2000. The various relationships were explored and studied using Multivariate Stepwise Analysis and Artificial Neural Networks only to reveal that meteorological conditions such as southern and western winds, accompanied with periods of high humidity and high temperatures, contribute to the accumulation of air pollutants which lead to an increased amount of air pollution which further leads to an increase in the number of hospital admissions. (Kassomenos, Papaloukas, Petrakis, & Karakitsios, 2008)

Similarly, another factor contributing to hospital admissions were the number of GP Visits in the preceding twelve month and the level of use of assistive technology used by the residents. To study the effects of these factors, research was conducted on the 80+ year old residents residing in the Linkoping Municipality of Sweden during the year 2010 using p value analysis. From this analysis, it was understood that patients which reported and required more than two visits from a GP in the preceding twelve months were significantly at a greater risk of hospitalisations. Also, increased use of assistive devices and technology and community assistance indicated an increase in the number of hospitalisations. (Nägga, Dong, Marcusson, Skoglund, & Wressle, 2012)

Another interesting determining factor was the arrival mode of patients. This was researched with an intense study that was conducted on around 317,581 Emergency Department Patients, in 2007 and 2008 over a period of 3 months in the departments of National HealthCare Group and Emergency Department, Tan Tock Seng Hospital, Singapore using p value analysis, c statistic analysis and Logistic Regression in order to develop the prediction model. This analysis developed a model which was represented by an equation along with the different factors which were indicated by the various coefficients. Thus, for every resident, a risk factor is calculated which determines whether the patient needs hospital admission or not. This model was used to predict early Emergency Admissions in order to avoid overcrowding and have better resource utilization. (Sun, Heng, Tay & Seow, 2011)

The resident's existing medical readings also contributed to predicting a possible hospital admission. These medical readings involved hemoglobin and sodium levels at discharge. In order to understand and analyse the importance of these readings, a detailed study was conducted on all patients discharged, around 10731 patients from any/all medical services between July 1, 2009 and June 30, 2010 in the Academic Medical Center in Boston, Massachusetts using Multivariable Logistic Regression. It was understood and reported from the study that low levels of hemoglobin and sodium at discharge resulted in a greater risk of a possible readmission for the residents. (Donzé, Aujesky, Williams & Schnipper, 2013)

Another contributing factor was the resident's frailty, morbidity and multi-morbidity. The relationship between healthcare and hospital admissions of the aging population depends upon these factors and was thoroughly studied by researching on the healthcare data of 40,728 people in the age group of 75-109 years of age in the county of Ostergotland, which is located in the south-east of Sweden using Multivariable Logistic Regression. After the study and development of the predictive model, it was understood that an increased level of multi-morbidity resulted in a greater number of Emergency Room visits which resulted in increased likelihood of hospitalisations. (Marcusson, Nord, Dong & Lyth, 2020)

Furthemore, other contributing factors include chronic medical problems, functional disability, and poor self-reported health. There were also less powerful factors such as low income and lack of a spouse. This study was conducted on all US civilians above 70 years of age in 1984 and the effect of these factors was again studied using p value analysis and Multivariable Logistic Regression. It was revealed that elders who had frequent health problems and chronic medical problems needed more hospital resources and this resulted in an increased number of hospitalisation. Also, residents who were not able to perform their daily activities and needed caregivers on a daily basis were at a higher risk of hospitalisation. (Boult, Dowd, McCaffrey, Boult, Hernandez & Krulewitch, 1993)

Another study was conducted on elderly Emergency Admissions data from National Healthcare Service Data over the four years from April 2007 to March 2011. The tests were done using Multivariate Linear Regression to test for multicollinearity only to again reveal that old age again played a substantial role in determining hospital admissions as more than half the number of emergency admissions were for people aged over 65 years. (Chenore, Gray, Forrer, Wright & Evans, 2013)

Similarly, studies were also conducted on around 45,000 residents aged 65 years of age and above from Group Health Cooperative of Puget South, which happens to be a Health Maintenance Organization located in Western Washington State in 1997. Studies were performed using Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC). This again revealed that older patients who had a history of diabetes and heart diseases and needed increased caregiving were at a greater risk of hospitalisation. (Coleman, Wagner, Grothaus, Hecht, Savarino & Buchner, 1998)

Similarly, studies were also conducted on seven hundred thirteen individuals who lived in community dwelling aged above 65 years of age in Germany, United Kingdom and Switzerland. Analysis was again done using Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC). Factors mentioned above such as increased age, need for caregiving, poor self-rated health, greater number of GP Visits in the previous year contributed towards increased hospitalisation. (Wagner, Bachmann, Boult, Harari, Von Renteln-Kruse, Egger, & Stuck, 2006)

Similar studies were conducted on patients aged 75 years and above present in the University of North Carolina Center for Aging and Health from January 1, 2007 to December 31, 2007 and June 1, 2008 to December 31, 2008. Logistic Regression and p value analysis was performed in order to understand the relationship between above mentioned factors such as age and resident's existing health readings such as heart rate and blood pressure.
It was understood from the studies performed that high blood pressures and need for facilities such as tube insertion for nutrition resulted in greater number of GP visits which further lead to increasing hospital admissions. This analysis was done to identify patients which were at a greater risk of hospitalisations in order to avoid unnecessary delay in admissions and have increased and improved resource utilization. (LaMantia, Platts-Mills, Biese, Khandelwal, Forbach, Cairns & Kizer, 2010)

Similarly, factors such as distance to hospital, rurality, access to primary care and lifestyle patterns also contribute towards hospital admissions which can be avoidable. Also, factors such as self management and homecare, exercise and provision of telemedicine services also contribute towards predicting and preventing these avoidable hospital admissions.
(Aminzadeh & Dalziel, 2002) However there is uncertainty over which factors are avoidable and unavoidable. The contribution and effect of these factors was analysed using a study conducted between 2004 to 2005 on the Emergency Elderly Hospital Admissions by the Nuffield Trust in the UK. Based on the studies, it was observed that age again played an important role and an aging population contributed towards more than half of the emergency admissions. (McCusker, Karp, Cardin, Durand & Morin, 2003)

It was also found that people who lived in deprivation were also more likely to be admitted to hospitals. (Salvi, Morichi, Grilli, Giorgi, De Tommaso & Dessi-Fulgheri, 2007)
Lower education also contributed towards an increased number of hospital admissions. It was also noted that people living in Urban Areas had higher hospital admission rates than people living in rural areas. (McCusker, Bellavance, Cardin, Trepanier, Verdon & Ardman, 1999)

Living closer to an ED, somehow also increased the risk of admission. Lifestyle habits such as smoking, also saw an increase in the level of hospitalisations. (Miller & Weissert, 2000)
Also, improvement in telemedicine services and homecare showed a decrease in hospitalisations. (Purdey & Huntley, 2013)

Furthermore, studies were also conducted on older Thai adults from local hospitals from more than 66 provinces of Thailand during June 1999 to August 1999. Univariate, multivariate and p value analysis were also performed to understand the importance of various factors collected from the Thai dataset associated with hospital admissions. The studies showed that poor financial status resulted in increased hospital admissions. Similar to the above findings from the Nuffield Trust in the UK, it was clear that current and former smokers were at a greater risk of hospitalisation. This was also true in the case of alcohol consumption. Age again was a major contributing factor with elders aged over 65 years of age were at a higher risk of hospitalisation. Other factors such as history of chronic lung diseases, repeated falls, poor self health, consuming calcium tablets and fatty foods, low body mass index and hypertension lead to increasing hospital admissions. Also, limited physical independence is also seen as an important predictor for hospital admissions.
(Assantachai & Maranetra, 2005)

Another important factor for predicting hospital admissions include having two or more comorbidities and taking more prescription medications. Studies were performed in order on around 411 patients aged 65 years of age and above who took part as control patients in the Generalist Physician Initiative at the Carle Clinic site, Urbana, Illinois between May 1993 and May 1994. Logistic Regression was used to understand and interpret the influence of these characteristics. A scoring system was then developed using these predictors to divide the residents into groups of low-risk and high-risk categories. It was clearly observed that patients with more comorbidities such as cancer, heart failure, strokes, diabetes, restricted activity in bed and taking multiple medications lead to increased GP visits which further lead to increased hospitalisations. (Shelton, Sager & Schraeder, 2000)

Similar to the above studies, it can be seen that factors such as chronic heart and respiratory diseases, diabetes and presence of palliative care also contribute towards predicting early hospital admissions. In order to further understand the importance and effect of these factors, studies were conducted in the Comprehensive Healthcare Centre of Burjassot in Valencia, Spain on around 6905 people aged over 65 years of age in 2013. Logistic regression was performed to assess the correlation between these variables. Thus, based on the analysis and the model developed, they were able to successfully identify patients which were at a high risk of admission. Presence of heart and respiratory diseases lead to an increased risk of hospital admission. It was also noted and observed that the presence of palliative care had a strong effect on the prediction of hospital admissions. Furthermore, other factors such as lack of social/family support and economic problems also contributed to increased hospitalisation. (Doñate-Martínez, Ródenas-Rigla & Garcés-Ferrer, 2017)

We can see that commonly occurring factors for model development and hence prediction include age, previous records of hospitalisations, existing medical conditions, presence of multimorbidity, prescription of multiple medicines and social/family support. From the above range of research work conducted, we can see that the models which were developed for predicting the hospital admissions were mostly based on Multivariate Logistic Regression Analysis. Based on a systematic review of the various risk models developed in the United States, United Kingdom, Italy, Spain and Canada for adults aged above 18 years of age subject to emergency hospital admissions in 2014, it was observed that models which were developed using administrative and clinical data and which was developed on larger datasets had an higher predictive ability than models based on self-report questionnaires. Models which included family history and GP data also improved the predictive accuracy. (Wallace, Stuart, Vaughan, Bennett, Fahey & Smith, 2014)

Similarly, studies were also conducted among around 185 older adults aged 65 years and above in Ramsey County, Minnesota in June 1992. This study was carried out in order to understand the factors which the older population thought was responsible for their hospitalisations. Similar to the above findings, factors such as age, sex, self-rated health, diabetes, presence of heart diseases and recent hospital admissions played an important role in predicting these hospital admissions. Logistic regression was then used to understand the relationship between these factors identified. Similar to the above findings, it was observed that residents above 85 years of age, poor self-rated health and presence of diabetes were involved in greater numbers of GP visits and hence were at a higher risk of hospitalisation. Similarly, residents who spent a larger amount of monetary resources in medicines and pharmacy were more likely to have a larger number of GP visits and consequently be at higher risk for hospitalisation. (Pacala, Boult & Boult, 1995)

Similarly, studies have also been conducted on around 1,122 Emergency Department patients aged 65 years and more who visited four urban, university affiliated hospitals in Montreal, Quebec, Canada over a period of three months in 1996. Logistic Regression was performed and p analysis was also conducted in order to understand the importance and influence of factors such as previous history of hospitalisations, heart diseases, marital status, drinking of alcohol and lack of social/family support. It was observed from the studies that, similar to the above mentioned findings, chronic medical conditions which result in frequent relapse, poor care for the different chronic conditions, social isolation, psychological conditions of the patients and bad quality of the services and care during the initial visits also result in increased hospitalisations. It was also found that patients who had a history of heart diseases were more likely and prone to earlier hospitalisations.

Similarly, patients with a history of diabetes were more prone and likely to frequent hospitalisations. (Hasan, Meltzer, Shaykevich, Bell, Kaboli, Auerbach & Schnipper, 2010)

Based on the findings from various other studies, including the present one, it was once again established that recent history and cases of hospitalisations were an crucial indicator of seriousness and illness and predicted both frequent visits and early visits. (Donzé, Lipsitz, Bates & Schnipper, 2013)

Furthermore, patients who were diagnosed with digestive and respiratory problems were also more prone to frequent cases of hospitalisations. It was also found out that males, who were living independently and alone also had frequent cases of hospitalisations. Depression was also one of the deciding factors for hospitalisation. It was observed that people who were diagnosed with depression and psychiatric problems, were more likely to use more hospital services and were at an increased risk for hospitalisation. Marital status and alcohol consumption were also key factors while determining admissions. (Hu, Gonsahn & Nerenz, 2014)

It was observed that residents who were not married before were less likely to have a return visit to the hospital when compared to those who were married or widowed. This may be because of the fact that individuals who have not been married before are more likely to be self-sufficient and independent than those who have been married. Elders who are married may have partners who themselves have existing problems and elders who are divorced or separated, may lack family/social support. It was also seen that those who drank alcohol daily were less likely to have a return visit to the hospital than those who did not drink alcohol daily. This can be explained because of assumptions such as the ability to drink alcohol daily may be considered as a good indicator of health as elders with chronic health problems may be advised to reduce/stop their intake of alcohol. Finally, lack of support remains to be one of the strong predictors of hospital admissions and readmissions. Lack of family and social support for the elderly has been connected with increased use of hospital resources and hence increased hospital admissions. (McCusker, Cardin, Bellavance & Belzile, 2000)

Thus, we have explored in depth around 30 papers for our substantive literature review. The literature review has carried research across all age groups, mostly targeting the elderly population aged 65 years of age and above. It covers research conducted over a period of twenty years from the 1980s to 2000s.It also covers research and studies conducted around the world in places including the United States, Canada, United Kingdom, Italy, Germany, Spain, Thailand, Singapore and Switzerland. The various sections presented on top discuss the different topics being covered. The review has given a thorough analysis of the range and findings of the existing literature which can be used to predict the various hospital admissions.It covers, what/where/who/when and how of the research. Based on the diverse range of research conducted, age, meteorological conditions, recent cases of hospital admissions, use of supportive and assistive technology, existing chronic medical conditions such diabetes, heart, respiratory and digestive diseases, presence of multi-morbidity and comorbidity, poor functional ability, increased need for care, provision and availability of telemedicine services, educational background, lifestyle habits such as smoking and consumption of alcohol, proximity to an ED, urban habitats, economic conditions, intake of multiple medicines, recent cases of strokes and falls, social isolation, marital status and finally a lack of family/social support were identified as some of the most important factors which contributed to increase use of medical services and henceforth lead to early and frequent hospitalisations. In conclusion of the literature review, the most commonly used tool for analysing these factors has been p values analysis and the most commonly used model to carry out these hospital admissions has been Multivariate Logistic Regression.

1.3 Summary of the State of the Art

This section presents the limitations and the gaps which were observed in the literature review conducted above with respect to my project proposal in mind. Now, after going through 30 review papers, we have now been able to narrow down our hospital admissions to the aforementioned factors. However, there are many challenges involved while analysing these factors too. To start with the meteorological factors affecting health and consequently hospital admissions, there is a problem with respect to measured air pollution levels and the corresponding population it covers. The concentrations which are measured in the various sites monitored don't really represent the entire population. Similarly, while analysing the effects of GP visits and use of assistive technologies, in the study conducted, all participants present belonged to the same municipality. Also, individuals who were present in shared and sheltered accommodation were under more closer surveillance than individuals residing by themselves. Hence, the data collected would be biased and incorrect. With respect to the ED data collected from Singapore in order to perform logistic regression, the gap in the study was that only routine ED data which was collected at the time of triage was used for developing the predictive model. Other factors such as underlying symptoms, vital signs of the patients and socioeconomic status were not incorporated in the development model.

Furthermore, while analysing factors such as morbidity and multi-morbidity, the study excluded inpatient deaths and post-discharge deaths were not considered in the outcome analysis. Also, these models did not include the effect of factors such as functional status, literacy levels with respect to health, social/family support and previous conditions of medications provided. Furthermore, while analysing the factors like hemoglobin and sodium levels, it becomes a problem because these values become available only at the time of discharge and this becomes difficult to incorporate into the model. Also, these studies only have internal model validations. They have been implemented only in smaller hospitals at a small scale and need to have a more widespread implementation.

Similarly, while analysing factors such as functional disability and chronic medical problems, the study was based strictly in the US and was not able to include data from other counties and countries. This study and model too was not able to incorporate the effect of socio-economic and demographic data and data which was not available in the administrative health care department.

In a similar way, while analysing the effect of history of diabetes using ROC and AUC statistics, a major limitation was that the administrative approach is heavily dependent on information which may not be available at the time of enrolment. What we could have is a screening hybrid strategy which incorporates a new technique. Here, the newer population would be screened at their times of enrolment and for the population which has already been enrolled, the administrative approach can be used.

Thus, there are many limitations and problems in the existing factors and models which have been implemented and my research proposal aims at bridging these gaps in an effective manner.

1.4 Plan for your research project

This section revolves around a feasible and focussed plan which explains what research I intend to do in order to resolve the various issues which I identified in the literature review.
This project revolves around successfully developing a model which can be used to Predict the Hospital Admissions from Australian Aged Care Homes in collaboration with the Health Metrics Data. This project proposal has been divided into three sections as per the requirements provided.The first section of the project proposal will cover the various tasks I intend on completing. The second section will throw light on the time which will be needed to complete every sub-task. The final section will discuss the various existing limitations which were identified in the literature review and the different ways in which I plan on addressing these issues and gaps.

The data provided by Health Metrics is really huge with multiple tables in their dataset and is confidential. My first step was to identify possible factors which lead to GP visits and hence result in consequent hospitalisations. Clearly, based on the in-depth literature review done, the following factors of age, meteorological conditions, previous cases of hospital admissions, usage and dependency of assistive technology, existing chronic medical conditions, presence of multi-morbidity and comorbidity, lack of functional ability, increased need for self care, provision and availability of telemedicine services, educational background, smoking and consumption of alcohol, proximity to an ED, economic conditions, intake of multiple prescription medicines, recent cases of strokes and falls, social isolation, marital status and finally a lack of family/social support were identified as important factors in determining hospital admissions. My second next step would be to detect each of these mentioned factors and possibly more in the huge dataset of Health Metrics. Once these factors have been located in the database, I will have to analyse the correlation of this factor and its corresponding table with the Hospital Admissions table. Locating, extracting and analysing these factors from the corresponding tables would require substantial effort and a major chunk of the project. Since the database provided by Health Metrics is confidential, I may not be able to access all the factors that I need for this model development. My third next step would be to get this approved authorised information in the form of a CSV file, ready for data cleaning and data wrangling. After substantial data cleaning and data wrangling is done, my fourth step would be to analyse and understand the data. This can be carried out through data visualisations. My fifth and final step would be in applying various models, mostly multivariate logistic regression models in order to predict the hospital admissions. This step would also entail understanding the effect and importance of the various factors using p value analysis, c statistics and f statistics.

Thus, the project can be divided into five subsections which need to be delivered over a period of three months. The following captures the key timeline of my proposed project. The first step which was Identification of factors, took around one month to complete starting from March 1st to April 1st. This was done using extensive review of more than thirty existing research papers. The second step which is Detection of factors in the Database will take around fifteen days to complete and would be done by April end.This step would completely revolve around exploring and understanding the database and is the most crucial step of the project. Once the factors are identified, detected in the database, the third and fourth step is Data Wrangling and Data Analysis 15 days to complete and I expect to complete it by mid May. My fifth and final step which is Model Development, would consume another 15 days. Overall, the tasks of Data Cleaning, Data Analysis and Model Development would be done simultaneously over a period of one month and the intended date of completion is May End.

Now, based on the findings from the Literature Review, I can clearly see that certain aspects of socio-economic and socio-demographic features of the population have not been included while implementing the Logistic Regression Models. I plan on analyzing the importance and effectiveness of factors which mainly revolve around the psychological aspects of the elderly population. Thus, I intend to cover the implications of family/social support, companionships and depression tendencies. These mental problems can cause disturbing health problems, which in turn lead to an increasing number of hospital admissions. I intend on covering the effect of these factors as the admissions caused due to these parameters are completely avoidable, if appropriate measures and mental health services are provided. These would be implemented using various models of Univariate and Multivariate Linear and Logistics Regressions. I also intend on studying the various factors using the methods of p value analysis and c statistics.

1.5 Conclusion

In conclusion, my project revolves around Predicting Hospital Admissions of Residents from Aged Care Homes, which is in collaboration with Health Metrics which provides software solutions to the Aged Care Sector. The Aged Care industry covers all the services of Home Care, Residential Care and Disability Care. This report first provides a substantial, thorough literature review. This review has been done after having an in-depth understanding and analysis of more than 30 Research Papers. The review covers studies which have been carried out across all age groups, mostly in the elderly section. It covers studies which have been conducted in different countries around the world ranging from the United States, Canada, United Kingdom, Germany, Italy, Singapore, Spain, Thailand and Switzerland. Finally, the review also throws light on the various factors which were identified to be crucial for predicting and preventing the hospital admissions such as age, recent cases of hospital admissions, use of assistive technology, existing chronic medical conditions, presence of multi-morbidity, increased need for care, availability of telemedicine services, educational background, lifestyle habits, proximity to an ED, economic conditions, intake of multiple medicines, recent cases of strokes and falls, marital status and finally a lack of family/social support.

The report then outlines the various limitations and gaps which were observed in the existing literature review, mainly there were issues in implementing the effects of socio-economic and socio-demographic data while incorporating the predictive model. The report finally throws light on my intended project proposal. This highlights the various tasks which have to be implemented along with the intended/proposed timeline of the project, keeping in mind the data security and confidentiality needs of Health Metrics.

Part 2: The Research Paper

Abstract

Background: There is a need for various factors, tools and models to be developed in order to identify and predict the elders in aged care homes who need to be admitted to hospitals so that appropriate measures and timely services can be taken and directed towards them respectively.
Objective: To determine the importance of the various factors which help in the identification of elderly residents from aged care facilities, who are at high risk of hospital admissions.
Design and Methods: Data that was that wrangled, explored and analysed was the test, sample data provided by HealthMetrics. Logistic regression analysis of this test data was performed on half of the data in order to identify the importance of high-risk factors which led to repeated and multiple hospital admissions. Then, these various short-listed factors were then tested on the other half of the sample data to understand and locate the residents who were at a greater risk of admissions.
Setting: Sample Aged Care Resident Data provided by HealthMetrics,Victoria, Australia.
Participants: A sample test data of around 766 residents, in the age group - 60 to 100 years, belonging to the various aged care facilities, provided by HealthMetrics.
Results: Ten factors were determined which had significant effects on hospital admissions of the resident population from aged care homes. The identified factors were older age group ($p < 1.92e-10$), gender (0.1919) , partnership status (0.4487) , index of family support (0.798), index of social support ($p < 3.85e-10$) , an increased need for care ($p < 2e-16$) , index of functional abilities ($< 2e-16$) , presence of multi-morbidity ($< 2e-16$), lifestyle patterns such as smoking ($< 2e-16$) and alcohol consumptions ($< 2e-16$), and recent history of hospitalisation.
Conclusions: Thus, ten factors were ascertained to identify elderly and senior residents who had greater chances of getting hospitalized. In the future, these factors can be further used to determine and predict in advance the residents who need to be admitted, so that they can be looked after with greater care and timely services and assistance can be provided so that temporary hospital visits/admissions or even deaths can be averted.

Keywords: Aged, Elderly, Older People,Risk Factors, Admissions, Hospitalizations.

2.1 Introduction

2.1.1 Contributions of the Research Project

This section throws light on my contribution to the research work conducted.

This Research Project revolved primarily around analysing the importance of various factors which led to Hospital Admissions of Residents from Australian Aged Care Homes. It has been repeatedly observed through numerous studies conducted that an increased need of dependency, greater disability and presence of diseases are resulting in greater emergency and hospital admissions - both temporary and permanent(Wanless, 2002).

Demographic factors such as older age, gender, social factors such as marital status, index of social and family support, lifestyle factors such as smoking history, drinking patterns, medical factors such as an increased need for care, poor functional abilities, presence of multiple diseases, recent history of hospitalisations, educational backgrounds, recent cases of strokes and falls, social isolation, existing medical conditions such as heart diseases, respiratory diseases, and diabetes, proximity of an existing emergency department, availability and provisions of telemedicine services were short-listed, noted and identified as some of the most common and important factors which lead to increased use of medical services and consequently hospital admissions(Cornette, D'Hoore, Malhomme, Van Pee, Meert & Swine, 2005).

It is also quite commonly observed that these hospital admissions also have an adverse effect on the health of individuals, especially the elderly group in terms of mental wellbeing, mortality and functional decline (Walter et al., 2001; Ponzetto et al., 2003a; Graf, 2006; Palmisano-Mills, 2007).

This first step of my Research Project was spent in identifying, researching and analysing the various low risk and high risk factors which were responsible for causing these emergency temporary and permanent admissions(O'Caoimh, Cornally, Weathers, O'Sullivan, Fitzgerald, Orfila & Molloy, 2015). The factors were determined after a thorough literature review of more than 30 Research Papers which gave me a peek into the research which was previously done in this area of healthcare. A substantial amount of research was done across all age groups, but mainly targeting the elderly population, aged 60 years and above. Research and studies were also over a period of twenty years from 1980s to 2000s in countries around the world which included Singapore, Thailand, Switzerland, Germany, Italy, United Kingdom, Spain and the United States.

After identifying these factors, the second major part of my project involved looking into the various models which have already been developed in order to assess the importance of each of these factors, which caused hospital admissions. The most common methods included p value analysis and Logistic Regression, followed by p value analysis of the various factors(Zhang, Kim, Patzer, Pitts, Patzer & Schrager, 2017).

My dataset was the sample test data provided by HealthMetrics. HealthMetrics was established in 2008 and it provides software solutions and supports the operational and clinical departments of Residential Aged Care, Disability Support, Community and Retirement Living. After the identification of factors and models developed, the sample test resident aged care data was explored and wrangled further in order to locate the above mentioned factors in the dataset. This stage of Data Wrangling was the hardest part in the project. Once the dataset was sufficiently wrangled and the different factors were located in the different tables of the database, the data from the different contributing factors was then combined into multiple datasets for analysis. Finally, logistic regression and p-value analysis was performed in order to assess the importance of the factors identified and located.

Thus, the objective of this research was to identify the various factors which contribute towards the hospital admissions, determine the degree of importance of each factors, locate the presence of the various identified factors in the existing sample test dataset and finally analyze the degree of importance of each of the identified factors with respect to our sample dataset.

2.1.2 Thesis Report Organisation

This section revolves around how the remaining contents of the report are arranged. Section 2.2 will provide the readers with an overview and background of the research work carried out. It explains the reasons and benefits of carrying out the research work. Section 2.3 discusses the methodology followed in order to conduct the research. It discusses the design, setting, data source and study variables. It also gives the readers a detailed description of the tools used and steps followed for the development of the model, followed by the statistical observations and analysis. The results and discussions of the results are discussed in detail in Section 2.4. Section 2.5 highlights the limitations of the research work and possible applications and opportunities in the future research. I have concluded the paper with the conclusion in Section 2.7.

2.2 Background

This section revolves around the condition of the present healthcare system, the admission of elders into emergency departments of hospitals and consequently the need and motive for carrying out the research work.

2.3.1 Overview of the Elderly Population in Healthcare

The elderly population is constantly increasing worldwide and this further leads to increased healthcare related needs, increased healthcare related costs and consequently increased hospital admissions(WHO, 2015). With every change in the elderly population, there is a significant impact on the healthcare policies, economy and consequently in the system.

The number of emergency hospital admits in hospitals has been rising and these admissions are common among individuals who have already been either admitted previously or are suffering from an existing long-term medical condition(Bridges & Meyer, 2000).

This admission of the elderly population accounts for shortage in hospital beds and growth in emergency admissions(Thompson, Shaw, Harrison, Gunnel & Verne, 2004). Research has shown that these hospital admissions, planned or unplanned, can be definitely avoided. It has also shown that when the elders are admitted, they are more likely to be readmitted. These admissions and readmissions can be contained and reduced by better management of healthcare facilities, combined with solid support from the government (Lyon, Lancaster, Taylor, Dowrick & Chellaswamy, 2007).

The present healthcare system in most of the developing countries and even in some developed countries is not sufficient enough and is not designed as per the needs of their aging population and healthcare needs(Banerjee, 2015). Thus, the healthcare system is at large considered to be a disorganized sector. However, at the same time, there are many studies and reports which indicate otherwise (Soong, Poots, Scott, Donald & Bell, 2015). Majority of the older population consider themselves to be perfectly healthy and are satisfied with their health (Nagga, Dong, Marcussion, Skoglund & Wressle, 2012).

2.3.2 Need and motive of the research work

The elders present in the various aged care facilities are also prone to hospital admissions based on various above mentioned factors. These admissions can be either temporary or permanent. Temporary admissions are where the resident returns back to the aged care facility after taking the necessary medications and treatments. Permanent admissions are slightly more serious, where the patient may need permanent support of medical services and can eventually die. Sometimes, because of the increasing number of hospital admissions, there can be a shortage of beds in the emergency department and ED overcrowding. The patients admitted can also have post-admission effects of poor mental health and social isolation. There are also cases of residents with poor immune systems acquiring diseases after getting admitted. Also, healthcare services are an expensive affair and increased admissions can also lead to increased expenditure(Collerton, Davies, Jagger, Kingston, Bond, Eccles & Kirkwood, 2009).

Hence, it is becoming increasingly important to prevent these hospital admissions. It’s crucial to have a set of factors, tools, models which can help early detection of hospital admissions from aged care homes. In the various previous studies which have been conducted, statistical prediction models have been developed to accurately predict and identify the older residents who are in greater need of medicare services and consequently at a greater risk of admissions(Guideline, N.I.C.E, 2016). In my present research study, I am working on a larger sample, test data, to see if I can develop a statistical model for the older population at a greater risk of hospital admissions which can then be applied in routine healthcare.

Furthermore, If we are able to successfully identify the elders present in the various aged care homes, who are at higher risk of hospital admissions, then we can provide them with increased primary care, attention and proactive measures can be taken before hospital care steps in to provide care(Marcusson, Nord, Johansson, Alwin, Levin, Dannapfel & Anderson, 2019). Also, the people who are in need for hospital level care and treatment, can be moved to appropriate respective clinics, instead of rushing them all to the emergency departments.

Thus, in effect, these frequently admitted elderly residents can enjoy improved health if these residents are identified at an earlier stage and treated. This can reduce their risk of temporary or permanent admissions and consequently improve the quality of their mental health and increase their life-span.

2.3 Methodology

The section revolves around the various methods adopted in order to go about this research. It demonstrates the aim and design of the study. It later highlights the data source and study variables. Furthermore, it demonstrates the tools used and steps followed to implement the various models. Finally, it describes in detail the statistical observations and subsequent analysis.

2.3.1 Aim

The aim of this research project was to first identify the various high risk factors which were crucial indicators of hospital admissions from Aged Care Homes. Following the identification of the factors, the second aim was to then assess the degree of importance of each of these identified factors using regression methods and p-value analysis.

2.3.2 Design

This study included the resident information stored as a part of the sample test data provided by HealthMetrics. This included data of 766 sample residents divided aged mostly between 60-100 years in the Victorian State of Australia.

2.3.3 Tools Used and Steps Followed

The tools used in this study were SQL, Microsoft Excel, Tableau, Python and R.

The main methods followed were as follows:-
Step 1 : Identification of the different contributing factors which cause potential hospital admissions in aged care homes. Factors were identified after a thorough and comprehensive literature review of more than 40 academic papers of the existing work which had been done in this area. The different factors which were identified to be analysed and tested on the HealthMetrics sample data are as follows :

- Demographic factors such as older age, gender and marital status.
- Geographic factors such as meteorological conditions, proximity to an emergency department of a hospital and urban habitats.
- Social factors such as index of social and family support and social isolation.
- Medical factors such as recent history and cases of hospital admissions, recent cases of strokes and falls, use of supportive and assistive technology, presence of multi-morbidity, poor functional ability, increased need for care, provision and availability of telemedicine services and general medical educational knowledge.
- Lifestyle factors such as smoking and consumption of alcohol and economic conditions.

Step 2 : Location of factors in the HealthMetrics Database
I then tried to locate the above 20 shortlisted factors in the sample HealthMetrics Database. However, I was only able to successfully locate and extract ten following factors :- age, gender, partnership

status, functional abilities, increased need for care, smoking and drinking history, index of social and family support, presence of multimorbidity and recent history and cases of previous hospital admissions. It was difficult to locate the other identified factors of economic conditions, proximity to emergency services and meteorological conditions affecting the hospital admissions of the elderly people. A detailed explanation is further provided in section 2.5, which entails the limitations and possible future work of the report.

Step 3 : Resident Data Extraction and Resident Data Wrangling
After the above mentioned were identified and located in the HealthMetrics database, the sample resident data was extracted from the database tables using SQL Queries. It was then wrangled and cleaned for analysis using Microsoft Excel and Python Programming.

Step 4 : Resident Data Exploration, Data Visualization and Data Analysis
The different features of the resident data were explored and analysed using Python Programming and Pandas. The spread and distribution of data was analyzed using Python and the data was further analyzed through visualizations in Tableau.

Step 5 : Regression Analysis and P Value Analysis
The importance of the different factors was assessed by applying logistic regression models and having a p value analysis. This model development step was done using R Programming.

2.3.4 Data Source and Study Variables

My dataset was the sample test Residential Aged Care Data provided by HealthMetrics. For example, corresponding to every resident, there are records of information around resident ID, resident name, gender, date of birth, marital status, country of origin, country of birth, current address, billing address etc.

The sample test dataset accurately represented the actual database accurately in terms of the number, types and names of the tables. The schema was exactly the same as that of the Health Metrics's real world database. Only the data which was present in the database was created for testing and confidentiality purposes. This sample test data was considerably huge with thousands of tables. The demographic factors around age, gender and partnership status was easily available in the resident information. Age was extracted from date of birth and partnership statuses were derived from their respective marital status.

The data to get an idea of increased need for care and poor functional ability was analysed through the appraisal records. These two factors were further analysed to be interdependent on each other, i.e poor functional ability lead to an increase in the need for care.

The data which provided an index of social and family support was collected through the records of social leaves, contacts and corresponding relationships. This index, which indicated the degree of social and family support that the resident received was calculated using the parameters of number of records of visits made by the family members and the number of leaves taken by the residents to visit their friends, close immediate family members and relatives.

The data around the lifestyle factors were taken from the records around smoking and consumption of alcohol assessments.

The data surrounding the index of economic conditions was gathered through the calculator which was used to determine the fees charged for various hospital services. With this, we were able to get a general idea of the resident's income and financial conditions based on the fees charged for the various medical services used by the resident.

The data indicating the intake of multiple medicines was accessed and taken through the records of medication modules pertaining to each of the residents.

2.3.5 Model Development

The sample dataset was randomly divided into two halves (50% each) - a training dataset and the test or validation dataset. I used the training dataset in order to build a prediction model and the test dataset was used to validate the prediction model. The prediction model was developed using a logistic regression algorithm. The aim of the model was to successfully identify the residents who were at a greater risk of getting admitted. The model development was completely done using R Programming.

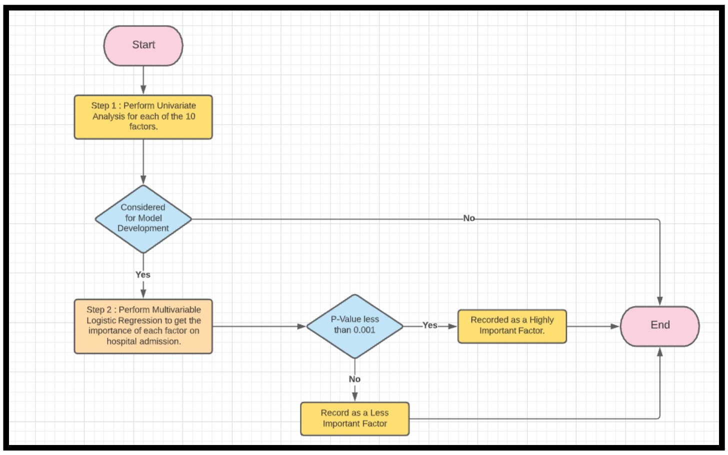
2.3.6 Statistical Explorations, Observations and Analysis

There were around 766 records to be exact which were observed for analysis. At the same time, I was going to analyse and assess the importance of the ten short-listed, identified factors which I was able to successfully locate in the sample test database.

The first critical step was to perform an univariate association for each of the 10 factors. Univariate analysis is one of the simplest forms of analysing datasets. This takes in the data, provides a summary of the data and is useful for finding patterns in the data. As we had only ten factors which were considered for model development, factors having p values of greater than 0.001 were also considered for model development.

The second step was the application of multivariable logistic regression in order to identify the most significant contributing factors which were high indicators of possible hospital admissions. This model-building stage revolved around three steps: selection of variables, building and testing of various models and validation of the model. Also, during the stage of data wrangling, I had converted all my variables into numerical variables for easier regression analysis and results. The different models which were developed were the Linear Regression Model and SVM Models. We know that the value of the Adjusted R-Square increases when a new predictor added improves the model, more than what is expected and it's value decreases when the predictor factor improves the model by less than what is expected. In order to test the accuracy of the model, values of Adjusted R- Squared were observed. Finally, factors which had low values (less than 0.001) of the p value parameter were marked as highly important factors, which were important indicators of the patients who were in need of hospital admissions. The above steps of the model developed are highlighted in the flowchart below.

Flowchart 1 : Flowchart indicating a high level overview of the model development stages.



Finally, the basic advantage of incorporating a Linear Regression model is that it has a simpler approach and is easy to read, understand and interpret. However it does not take into consideration the non-linear aspects and assumes that there is always going to be a linear relationship between the variables. This can lead to over-simplification of the problem. SVM overcomes this shortcoming of the Linear Model by taking into account the different nonlinear aspects and making the model more memory efficient. Finally, I implemented the Linear Regression Model on the validation data.

Linear Model

Basic Idea of the model:

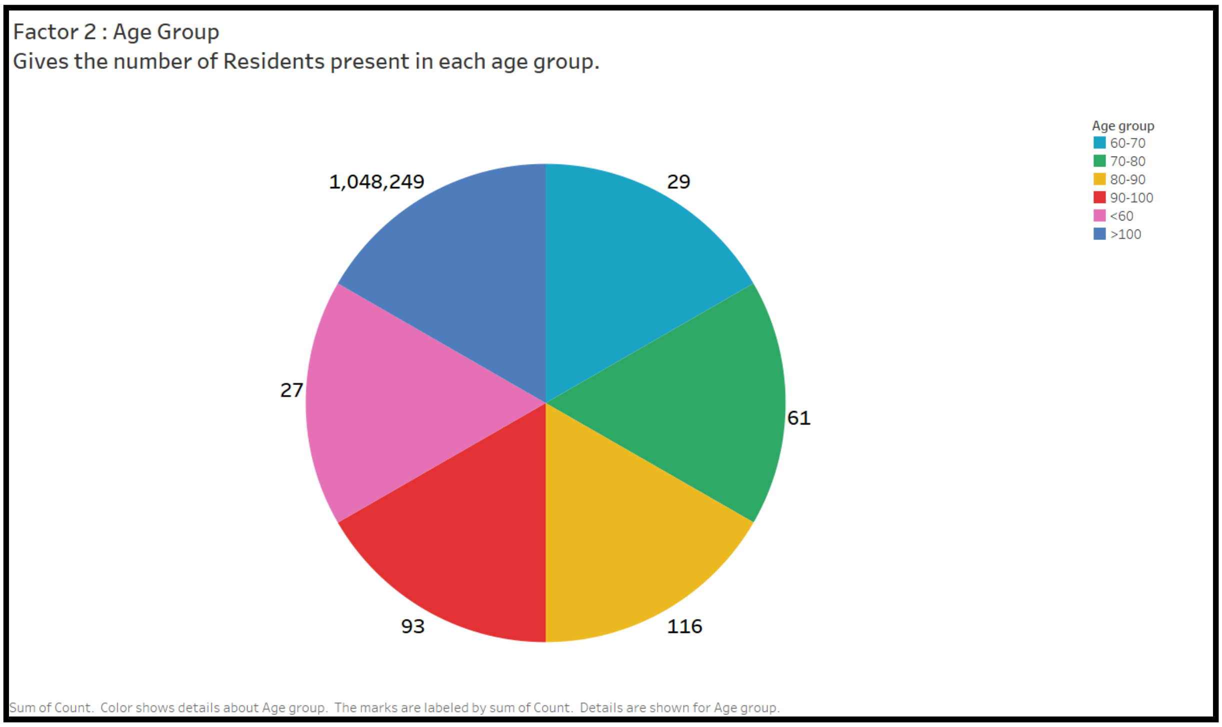
The basic idea of the logistic regression model is that it uses a logistic function through which we can model a binary, dependent variable. This model uses a supervised classification algorithm. Logistic Regression is used to define a relationship between one dependent variable which is binary in nature and one or more independent variables. Here, in this model our dependent variable is the index of hospital admissions and the independent variables are our factors such as age, gender, marital status, functional ability, need for care, smoking and drinking habits.

Construction of the Model:

Factor 1 : Age

The age of the participants was calculated from the date of birth data. This was done using the tools in Microsoft Excel. In order to further analyse the contribution of the age factor on hospital admissions, the participant data was further divided into age groups of less than 60, between 60-70, 70-80, 80-90, 90-100 and greater than 100. Most of the residents were present in the above 100 years of age category and the least number of residents were found in the below 60 years of age category. Considering that the data taken was from the sample test database, these unusual findings were quite normal. The resident age data, before the data cleaning of missing and redundant data, has been visualised in the table below, created using Tableau.

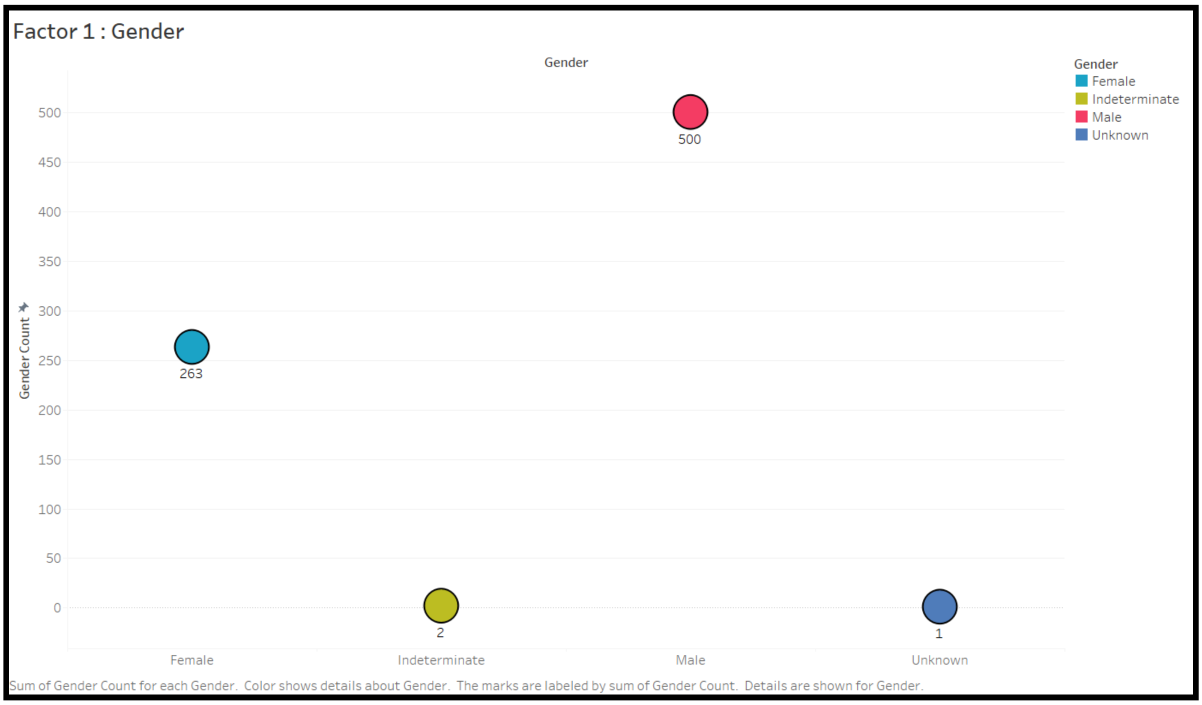
Table 1 : Visual Representation of the Age-Group Distribution of the sample data, before cleaning.



Factor 2 : Gender

In addition, the two genders analysed were the Male and Female Gender as the dataset predominantly revolved around these two genders. There were 500 and 263 records centred around Male and Female residents respectively. There were only 3 records, where the genders were not mentioned and these records were removed for analysis as they were not in a significant number. The gender distribution of the dataset, after performing vigorous data-cleaning is visualised below in Table 1 and was generated using the Tableau tool.

Table 2 : Visual Representation of the Gender Distribution of the sample data, after cleaning.



Factor 3 : Partnership Status
The different partnership information present was in the categories of unknown, single, married, divorced, widowed, De facto, life partner, separated and in a committed relationship. In order to further simplify the process of analysis. I further divided these ten categories into two main categories, named as 1 and 0, where 1 signified that the resident had a partner and 0 suggested otherwise.

Factor 4 : Functional Ability
Furthermore, unlike the above factors, the factors of need for care, functional ability, index of family and social support and drinking and smoking habits were not readily and directly available in the database. These factors had to be derived for further analysis. The factors of functional ability were assigned values from 0 to 5, indicating an index, where a value of 0 indicated poor functional ability and a value of 5 represented residents in excellent functional capacity. It was observed that for most of the residents in the sample dataset, an increasing age led to a decrease in the functional abilities. Residents aged below 60 years of age, were comparatively in better health when compared to residents above 90 years of age with respect to their functional abilities.

Factor 5 : Need for Care
The need for care factor was also assigned the values from 0 to 5 which represented the following.
0 : Indicated null or almost no need for care.
1 : Indicated little need for care, almost every 5-6 months
2 : Residents need care every 3-4 months.
3 : Residents need care every 2-3 months.
4 : Residents needed care almost every month
5 : Indicated a high need for care.
Similar to the above findings, it was highlighted that older residents, with poor functional abilities needed greater care and attention and this resulted in a higher number of hospital admissions.

Factor 6 : Family Support

The family support index factor was determined after calculating the number of visits from the resident's family members and the number of leaves the resident's took in order to visit their family. Based on the above conditions and results, a family support index was calculated and assigned from 0 to 5, which had the following representation and indications.

0 : Indicates null or almost no family visits and hence a very low family support index
1 : Indicates occasional family visits, almost every 5-6 months
2 : Indicates family visits every 3-4 months
3 : Indicates family visits every 2-3 months
4 : Indicates family visits almost every month
5 : Indicates family visits almost every month and hence a very high family support index

Factor 7 : Social Support

Similar to the family support index, the social support index was also determined after assessing the number of visits from friends and the leaves taken in order to meet friends. The social support index was calculated and values were 0 to 5 were assignment, which conveyed the following:

0 : Indicates null or almost no friends visits and hence a very low social support index
1 : Indicates occasional visits from friends, almost every 5-6 months
2 : Indicates visits from friends every 3-4 months
3 : Indicates visits from friends every 2-3 months
4 : Indicates visits from friends almost every month
5 : Indicates visits from friends almost every month and hence a very high social support index

Factor 8 : Multimorbidity

This factor was calculated using the records of resident medication intake which highlighted the medicines consumed on a daily, weekly and monthly basis. There was also information which pertained to the reason for consumption of medicine. These records were analysed to create a multimorbidity index.

0 : Indicated the consumption of almost 0 medicines.
1 : Indicated intake of medicine for 1 ailment.
2 : Indicated intake of medicine for 2 ailments.
3 : Indicated intake of medicine for 3 ailments.
4 : Indicated intake of medicine 4 or more than 4 ailments.

Factor 9 : Lifestyle Habit of Smoking

This factor was assessed from the records of lifestyle habits. This parameter was also assigned the values of 0 and 1. Here, 0 indicated that the resident did not indulge in smoking and had no history of smoking at all. The value 1 indicated that the resident indulged in smoking or had a history of smoking in the near past. It was a general observation that male residents had higher smoking habits when compared to female residents.

Factor 10 : Lifestyle Habit of Alcohol Consumption

Similar to the factor of smoking, the alcohol consumption factor was also calculated from the tables of lifestyle habits of the residents. This parameter too, was given values of 0 and 1, where 0 referred to no history of past and current drinking habits and 1 indicated a recent history and current lifestyle of drinking. Based on the resident data provided, it was generally observed that male residents were more likely to have drinking habits when compared to female residents.

Assumptions:
Here, we have assumed that the different observations of the dataset are independent of each other.

In conclusion, I developed four linear models in order to get an importance of each of the identified factors. The first model took into account the factors of age, gender, marital status, smoking and drinking indexes in order to predict the hospital admissions. The second model analysed the importance of functional ability and family support index. Finally, the third model analysed the importance of need for care and social support index on the hospital admissions of the residents.

2.4 Results and Discussions

This section throws light on the results obtained from the research work conducted, followed by a detailed discussion of the results.

2.4.1 Results

Overall, there were around 766 records of resident information which were analysed from the sample database. From the studies, it is evident that age plays an important role in determining hospital admissions. Thus, a resident aged 90 years and above is more likely to get admitted than a resident aged in the 60-70 age group.

Results of Model 1:

Table 3 : Factors alongside their corresponding p values for Model 1

Index	Factors Identified	P Value Analysis (>)
1	Age	1.92e-19***
2	Gender	0.1919
3	Partnership Status	0.4457
4	Alcohol Consumption Index	<1.92e-16 ***
5	Smoking Index	<1.92e-16 ***

Thus, we can see from the above table that age, smoking index and alcohol consumption indexes are strong indicators of hospital predictions with low p values - 1.92e-19. Clearly, gender and partnership status are poor indicators and factors causing hospital admissions, with p values of 0.1919 and 0.4457 respectively.

Results of Model 2 :

Table 4 : Factors alongside their corresponding p values for Model 2

Index	Factors Identified	P Value Analysis (>)
1	Functional Ability	<2e-16***
2	Family Support Index	0.798

Thus, we can see from the results of the above model 2, that functional ability has a low p value of <2e-16, indicating that it is a very significant factor contributing to hospital admissions. Similarly, the family support index has a relatively higher p value of 0.798, indicating that it is not a significant factor which leads to hospital admissions.

Results of Model 3 :

Table 5 : Factors alongside their corresponding p values for Model 3

Index	Factors Identified	P Value Analysis (>)

1	Need for care	<2e-16***
2	Social Support Index	3.85e-10***

Finally, based on the above results, we can see that need for care has a low p value of <2e-16 and the social support index also has a low p value of 3.85e-10. This indicates that both of them are significant factors when it comes to predicting hospital admissions of the residents.

Thus, to summarise the results, based on our analysis and findings, the most important, significant factors which lead to hospital admissions are age, functional ability, need for care, social support index, alcohol and smoking habits. The factors which are not that significant are gender and partnership status. These results obtained are also in alignment to the prior studies conducted in this area, the detailed description of which is provided in the discussion section below.

2.4.2 Discussions

I have followed a systematic approach in order to determine the importance of various risk factors. These risk factors which were identified and chosen were from reviews of prior research work and hence the factors are based on evidence. Also, the results which were obtained as a part of my research are in alignment with the results obtained from the academic review of the existing literature. The in-text references in the following discussion sections point towards the existing academic papers which have demonstrated similar significant results.

I used the sample test data provided by HealthMetrics in order to develop a predictive model for hospital admissions of elderly people from aged care homes. Older age, male gender, presence of a partner, lesser index of social and family support, increased need for care, poor functional abilities, increased history of smoking and drinking, greater index of multimorbidity and recent cases of emergency hospital admissions were identified as the ten most important contributing high risk factors for determining possible hospital admissions from these aged care homes for the model.

From our analysis, it can be observed that age continues to be one of the major factors leading to hospital admissions(Lyon, Lancaster, Taylor, Dowrick & Chellaswamy, 2007). Residents above 80 and 90 years of age continue to be at greater risks of hospital admissions. Another important contributing factors are increased need for care and poor functional abilities. These factors are also further interlinked with each other. Poor functional abilities lead to an increase in the need for care. Furthermore, this also leads to the increased use of assistive devices which consequently lead to greater number of hospital admissions (Nägga, Dong, Marcusson, Skoglund, & Wressle, 2012).

The presence of morbidity and multimorbidity also has significant effects on hospital admissions. An increased number of medical conditions leads to a greater number of local hospital visits. These small, local visits can have an adverse effect on the mental health and frailty of these individuals which in turn lead to more serious hospital admissions(Marcusson, Nord, Dong & Lyth, 2020). It has also been observed from studies performed that these patients with multiple diseases have restricted and limited movements in general and their many temporary visits which are a result of comorbidities, lead to increased hospital admissions (Shelton, Sager & Schraeder, 2000).

Similar to the above findings and results observed from my analysis, it was noted that a lack of social and family support contributed towards increased hospital admissions (Doñate-Martínez, Ródenas-Rigla & Garcés-Ferrer, 2017). It was also found that males, who were living independently and by themselves also frequently visited the hospitals.This could be due to additional factors such as social isolation and depression. An increase in the depression levels caused an increase in the alcohol consumption which further led to an increased number of hospital admissions(Hu, Gonsahn & Nerenz, 2014).

Furthermore, it was also observed that residents who were married or widowed were more likely to visit the hospitals. This can be explained by the fact that residents who did not have partners are more independent and self-sufficient than those with partners. Similarly, elderly residents may also have partners who are suffering from medical problems which can lead to multiple GP visits and eventually lead to a hospital admission.

The other variables which were shortlisted could not be located effectively in the database for them to be analysed. The main reason for prediction was to use it in possible clinical settings (Marcussion, Nord, Johansson, Alwin, Levin, Dannapfel & Andersson, 2019). One of the major strengths of this study is that these identified factors and models developed can be used as a tool in various routine healthcare services, in order to timely predict people in need of care, treatment and attention and prevent admissions. In clinical applications, the population which was predicted to be most likely to be admitted can be taken to primary, preliminary care centres where care can be provided through GP visits, home visits and support through phones. As a result, the medical providers can now direct their healthcare services towards people who are in greater need of these services (Marcussion, Nord, Johansson, Alwin, Levin, Dannapfel & Andersson, 2019).

Similarly, if we determine possible hospital admissions through the concept of frailty, it has limitations as this needs the presence of face to face interactions with the General Practioners. Based on the studies conducted in this area, it has poor accuracy while predicting hospital admissions (Wou, Gladman, Bradshaw, Franklin, Edmans & Conroy, 2013). Thus, in a practical clinical setting, if we apply these predictive models, along with measures of frailty observed from GP assessments, this method is likely to be more useful than measuring the index of frailty alone (Marcussion, Nord, Johansson, Alwin, Levin, Dannapfel & Andersson, 2019).

2.5 Limitations and Future Work

Through this research project, I initially short-listed more than 15 factors, through which I wanted to identify patients from the test sample data which were at a greater risk of admissions. However, I was only able to successfully identify the importance of ten major factors which played a pivotal role in determining the elderly residents who were at a greater risk of hospital admissions. Now, this section revolves around the various limitations and gaps which were observed in the research work conducted.

The first major limitation observed was the difficulty in locating and analysing the top ten analysed factors themselves! I was able to detect and locate demographic factors of age and gender and social factors of partnerships and marital status quite easily in the database. But, medical factors such as multimorbidity, reduced functional abilities and increased need for care took a substantial amount of time. Furthermore, the factors which were hardest to identify, analyse, visualise and assign an index were social factors such as presence of family support, social support and lifestyle factors such as habits of smoking and drinking.

The second major limitation observed was the lack of available data for meteorological conditions. The only way by which I could form a relation between the meteorological data and hospital admissions was through different datasets from different geographic locations with clearly different climatic conditions. However, since this entire research centred around Victorian Aged Care Homes, the climate is nearly the same and clearly not a contributing factor. Hence, the importance of this factor could not be determined. Similarly, there wasn't sufficient data around provision and availability of telemedicine services and presence and location of nearby emergency departments and I was unable to assess the importance of these factors.

The third major limitation was the lack of substantial information about the resident's current economic conditions and existing medical knowledge background. Firstly, there was data about the fees which the residents were charged for the various medical services provided. Hence, I was able to form an estimation of the minimum income which the resident should have in order to afford the services. However, there was no specific data which gave an indication of the resident's economic conditions. Based on my earlier reviews conducted, elderly people who enjoy excellent financial abundance, do not hesitate to visit hospitals at the slightest inconvenience (Fitzpatrick, Powe, Cooper, Ives & Robbins, 2004). Thus, good economic conditions lead to increased hospital admissions (Kingsley, 2015).Maybe, one of the reasons that there isn't much information is because residents may not be comfortable sharing their economic details and hence I was unable to analyse the importance of this factor.

Similarly, based on my earlier research review conducted, there was evidence of studies which suggested that residents with good knowledge of medical diseases and healthcare services were more likely to get admitted in order to get timely treatment(Soong, Poots, Scott, Donald & Bell, 2015). Similarly, elderly residents who had lesser knowledge about the services were more likely to resort to athome solutions and care, thereby resulting in lesser hospital admissions. Though I was able to detect data which centred around residents consuming multiple medicines, residents having poor or decreased functional ability and even residents needing increased amount of care, I was unable to locate data which gave me an index of the resident's medical awareness and knowledge and as a result I was not able to analyse the importance of this factor.

The fourth limitation of the study is that the outcomes and results of my model were similar to those conducted using similar conditions(Kansagara, Englander, Salanitro, Kagen, Theobald, Freeman & Kripalani, 2011). Thus, I believe that the results obtained from my dataset can be easily generalised.

The fifth and final limitation of the research is that this study was conducted on the sample test data, not on the actual data of the HealthMetrics database. This sample test data had the same database schema as that of the actual data and only the data used was created, in order to protect the privacy and confidentiality of the residents. Hence, the results obtained were on the sample data and not on the actual data.

In the future, the commonly identified and analysed factors can be further used to predict in advance the elderly and aged residents who are at a greater risk of admissions. These identified residents can then be looked after with greater care and timely medical services and assistance can be provided. This will consequently prevent temporary hospital visits/admissions and even avert permanent admissions and untimely deaths (Marcusson, Nord, Johansson, Alwin, Levin, Dannapfel & Anderson, 2019).

2.6 Conclusion

In conclusion, this research project throws light on the degree of importance of various identified factors in Predicting Hospital Admissions of Residents from Aged Care Homes, which is in collaboration with the Health Metrics Database.

The various ten contributing factors were successfully identified and located in the Health Metrics Sample Database. The factors were then merged together into different datasets in order to perform regression analysis on them and their importance was assessed using p value analysis. Furthemore, the limitations of the research project were also identified along with the different scopes and possibilities for the future.

The importance and analysis of the associated factors tell us that hospitalisations are not only due to medical conditions, but also due living conditions, i.e greater and increased use of assistive technology and assistance from the community (Parker & Thorslund, 2005). The possible step can be to further delve deeper into the factors like diseases which are a consequence of living conditions which influence healthcare services in the elderly population.

The methods and analysis used can be definitely generalized and they can be implemented in most of the aged care homes which have their data digitally stored (Garrard, Cox, Dodds, Roberts & Sayer, 2020).

Thus, effective predictions of the elderly group in need of immediate care and medication, can lead to timely precautions, assessments and interventions, which further enables the different healthcare professionals to direct their resources towards people who are in a greater need of those resources, thereby improving the efficiency of the healthcare system at large(Ellis, Gardner, Tsiachristas, Langhorne, Burke, Harwood & Shepherd, 2017).

2.7 References

1. Kassomenos, P., Papaloukas, C., Petrakis, M., & Karakitsios, S. (2008). Assessment and prediction of short term hospital admissions: the case of Athens, Greece. *Atmospheric environment*, *42*(30), 7078-7086.

2. Nägga, K., Dong, H. J., Marcusson, J., Skoglund, S. O., & Wressle, E. (2012). Health-related factors associated with hospitalization for old people: comparisons of elderly aged 85 in a population cohort study. *Archives of gerontology and geriatrics*, *54*(2), 391-397.

3. Sun, Y., Heng, B. H., Tay, S. Y., & Seow, E. (2011). Predicting hospital admissions at emergency department triage using routine administrative data. *Academic Emergency Medicine*, *18*(8), 844-850.

4. Donzé, J., Aujesky, D., Williams, D., & Schnipper, J. L. (2013). Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. *JAMA internal medicine*, *173*(8), 632-638.

5. Marcusson, J., Nord, M., Dong, H. J., & Lyth, J. (2020). Clinically useful prediction of hospital admissions in an older population. *BMC geriatrics*, *20*(1), 1-9.

6. Boult, C., Dowd, B., McCaffrey, D., Boult, L., Hernandez, R., & Krulewitch, H. (1993). Screening elders for risk of hospital admission. *Journal of the American Geriatrics Society*, *41*(8), 811-817.

7. Chenore, T., Pereira Gray, D. J., Forrer, J., Wright, C., & Evans, P. H. (2013). Emergency hospital admissions for the elderly: insights from the Devon Predictive Model. *Journal of Public Health*, *35*(4), 616-623.

8. Coleman, E. A., Wagner, E. H., Grothaus, L. C., Hecht, J., Savarino, J., & Buchner, D. M. (1998). Predicting hospitalization and functional decline in older health plan enrollees: Are administrative data as accurate as self-report?. *Journal of the American Geriatrics Society*, *46*(4), 419-425.

9. Wagner, J. T., Bachmann, L. M., Boult, C., Harari, D., Von Renteln-Kruse, W., Egger, M., ... & Stuck, A. E. (2006). Predicting the risk of hospital admission in older persons—validation of a brief self-administered questionnaire in three European countries. *Journal of the American Geriatrics Society*, *54*(8), 1271-1276.

10. LaMantia, M. A., Platts-Mills, T. F., Biese, K., Khandelwal, C., Forbach, C., Cairns, C. B., ... & Kizer, J. S. (2010). Predicting hospital admission and returns to the emergency department for elderly patients. *Academic emergency medicine*, *17*(3), 252-259.

11. Purdey, S., & Huntley, A. (2013). Predicting and preventing avoidable hospital admissions: a review. *The journal of the Royal College of Physicians of Edinburgh*, *43*(4), 340-344.

12. Assantachai, P., & Maranetra, N. (2005). Factors determining hospital admission of Thai elderly by a mailed survey. *J Med Assoc Thai*, *88*(8), 1051-6.

13. Shelton, P., Sager, M. A., & Schraeder, C. (2000). The community assessment risk screen (CARS): identifying elderly persons at risk for hospitalization or emergency department visit. *Am J Manag Care*, *6*(8), 925-33.

14. Doñate-Martínez, A., Ródenas-Rigla, F., & Garcés-Ferrer, J. (2017). Development of a Predictive Model of Elderly Patients at Risk of Future Hospital Admission at Primary Care Centres in Valencia (Spain). *European Journal of Interdisciplinary Studies*, *3*(2), 13-25.

15. Wallace, E., Stuart, E., Vaughan, N., Bennett, K., Fahey, T., & Smith, S. M. (2014). Risk prediction models to predict emergency hospital admission in community-dwelling adults: a systematic review. *Medical care*, *52*(8), 751.

16. Pacala, J. T., Boult, C., & Boult, L. (1995). Predictive validity of a questionnaire that identifies older persons at risk for hospital admission. *Journal of the American Geriatrics Society*, *43*(4), 374-377.

17. McCusker, J., Cardin, S., Bellavance, F. O., & Belzile, E. (2000). Return to the emergency department among elders: patterns and predictors. *Academic Emergency Medicine*, *7*(3), 249-259.

18. Cornette, P., D'Hoore, W., Malhomme, B., Van Pee, D., Meert, P., & Swine, C. (2005). Differential risk factors for early and later hospital readmission of older patients. *Aging clinical and experimental research*, *17*(4), 322-328.

19. O'Caoimh, R., Cornally, N., Weathers, E., O'Sullivan, R., Fitzgerald, C., Orfila, F., ... & Molloy, D. W. (2015). Risk prediction in the community: a systematic review of case-finding instruments that predict adverse healthcare outcomes in community-dwelling older adults. *Maturitas*, *82*(1), 3-21.

20. Lyon, D., Lancaster, G. A., Taylor, S., Dowrick, C., & Chellaswamy, H. (2007). Predicting the likelihood of emergency admission to hospitals of older people: development and validation of the Emergency Admission Risk Likelihood Index (EARLI). *Family Practice*, *24*(2), 158-167.

21.

22. Hong, W. S., Haimovich, A. D., & Taylor, R. A. (2018). Predicting hospital admission at emergency department triage using machine learning. *PloS one*, *13*(7), e0201016.

23. Zhang, X., Kim, J., Patzer, R. E., Pitts, S. R., Patzer, A., & Schrager, J. D. (2017). Prediction of emergency department hospital admission based on natural language processing and neural networks. *Methods Inf Med*, *56*(5), 377-89.

24. Aminzadeh, F., & Dalziel, W. B. (2002). Older adults in the emergency department: a systematic review of patterns of use, adverse outcomes, and effectiveness of interventions. *Annals of emergency medicine*, *39*(3), 238-247.

25. McCusker, J., Karp, I., Cardin, S., Durand, P., & Morin, J. (2003). Determinants of emergency department visits by older adults: a systematic review. *Academic Emergency Medicine*, *10*(12), 1362-1370.

26. Salvi, F., Morichi, V., Grilli, A., Giorgi, R., De Tommaso, G., & Dessi-Fulgheri, P. (2007). The elderly in the emergency department: a critical review of problems and solutions. *Internal and Emergency Medicine*, *2*(4), 292-301.

27. McCusker, J., Bellavance, F., Cardin, S., Trepanier, S., Verdon, J., & Ardman, O. (1999). Detection of older people at increased risk of adverse health outcomes after an emergency visit: the ISAR screening tool. *Journal of the American Geriatrics Society*, *47*(10), 1229-1237.

28. Miller, E. A., & Weissert, W. G. (2000). Predicting elderly people's risk for nursing home placement, hospitalization, functional impairment, and mortality: a synthesis. *Medical care research and review*, *57*(3), 259-297.

29. Hasan, O., Meltzer, D. O., Shaykevich, S. A., Bell, C. M., Kaboli, P. J., Auerbach, A. D., ... & Schnipper, J. L. (2010). Hospital readmission in general medicine patients: a prediction model. *Journal of general internal medicine*, *25*(3), 211-219.

30. Donzé, J., Lipsitz, S., Bates, D. W., & Schnipper, J. L. (2013). Causes and patterns of readmissions in patients with common comorbidities: retrospective cohort study. *Bmj*, *347*.

31. Hu, J., Gonsahn, M. D., & Nerenz, D. R. (2014). Socioeconomic status and readmissions: evidence from an urban teaching hospital. *Health Affairs*, *33*(5), 778-785.

32. Wanless, D. (2002). Securing our future health: taking a long-term view.

33. Walter, L. C., Brand, R. J., Counsell, S. R., Palmer, R. M., Landefeld, C. S., Fortinsky, R. H., & Covinsky, K. E. (2001). Development and validation of a prognostic index for 1-year mortality in older adults after hospitalization. *Jama*, *285*(23), 2987-2994.

34. [WHO] World Health Organization. (2015). World report on aging and health.

35. Banerjee, S. (2015). Multimorbidity—older adults need health care that can count past one. *The Lancet*, *385*(9968), 587-589.

36. Soong, J., Poots, A. J., Scott, S., Donald, K., & Bell, D. (2015). Developing and validating a risk prediction model for acute care based on frailty syndromes. *BMJ open*, *5*(10).

37. Nägga, K., Dong, H. J., Marcusson, J., Skoglund, S. O., & Wressle, E. (2012). Health-related factors associated with hospitalization for old people: comparisons of elderly aged 85 in a population cohort study. *Archives of gerontology and geriatrics*, *54*(2), 391-397.

38. Collerton, J., Davies, K., Jagger, C., Kingston, A., Bond, J., Eccles, M. P., ... & Kirkwood, T. B. (2009). Health and disease in 85 year olds: baseline findings from the Newcastle 85+ cohort study. *Bmj*, *339*.

39. Guideline, N. I. C. E. (2016). Multimorbidity: clinical assessment and management.

40. Marcusson, J., Nord, M., Johansson, M. M., Alwin, J., Levin, L. Å., Dannapfel, P., ... & Andersson, A. (2019). Proactive healthcare for frail elderly persons: study protocol for a prospective controlled primary care intervention in Sweden. *BMJ open*, *9*(5), e027847.

41. Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., & Kripalani, S. (2011). Risk prediction models for hospital readmission: a systematic review. *Jama*, *306*(15), 1688-1698.

42. Fitzpatrick, A. L., Powe, N. R., Cooper, L. S., Ives, D. G., & Robbins, J. A. (2004). Barriers to health care access among the elderly and who perceives them. *American journal of public health*, *94*(10), 1788-1794.

43. Kingsley, D. E. (2015). Aging and health care costs: narrative versus reality. *Poverty & Public Policy*, *7*(1), 3-21.

44. Ellis, G., Gardner, M., Tsiachristas, A., Langhorne, P., Burke, O., Harwood, R. H., ... & Shepperd, S. (2017). Comprehensive geriatric assessment for older adults admitted to hospital. *Cochrane database of systematic reviews*, (9).

45. Garrard, J. W., Cox, N. J., Dodds, R. M., Roberts, H. C., & Sayer, A. A. (2020). Comprehensive geriatric assessment in primary care: a systematic review. *Aging clinical and experimental research*, *32*(2), 197-205.

46. Wou, F., Gladman, J. R., Bradshaw, L., Franklin, M., Edmans, J., & Conroy, S. P. (2013). The predictive properties of frailty-rating scales in the acute medical unit. *Age and ageing*, *42*(6), 776-781.

47. Parker, M. G., Ahacic, K., & Thorslund, M. (2005). Health changes among Swedish oldest old: prevalence rates from 1992 and 2002 show increasing health problems. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 60(10), 1351-1355.

48. Bridges, J., & Meyer, J. (2000). Older people in accident and emergency: the use of action research to explore the interface between services in an acute hospital. *Educational Action Research*, *8*(2), 277-289.

49. Thompson, A., Shaw, M., Harrison, G., Ho, D., Gunnell, D., & Verne, J. (2004). Patterns of hospital admission for adult psychiatric illness in England: analysis of Hospital Episode Statistics data. *The British Journal of Psychiatry*, *185*(4), 334-341.

50. Lyon, D., Lancaster, G. A., Taylor, S., Dowrick, C., & Chellaswamy, H. (2007). Predicting the likelihood of emergency admission to hospital of older people: development and validation of the Emergency Admission Risk Likelihood Index (EARLI). *Family Practice*, *24*(2), 158-167.

Part 3: Appendices

This section contains snippets of the code used for data cleaning and model development. There were many data cleaning activities performed and I have highlighted data cleaning for the factors of age and marital status.

Data Cleaning

Assigning Age-Groups from Age :

The below figure provides the code for how the age of the participants was converted into a numeric datatype for easier analysis. Later, I created a new column, named as the age group column, which assigned an age group to every record. Later, I have even displayed the first five rows, which have the age groups along with the age mentioned.

In [26]:
```
# Changing the Datatype
my_age_data["Age"] = pd.to_numeric(my_age_data["Age"])
```

In [27]:
```
# Assigning Age Groups
my_age_data.loc[(my_age_data.Age < 60),  'AgeGroup'] = '< 60'
my_age_data.loc[( (my_age_data.Age >= 60) & (my_age_data.Age < 70 )),  'AgeGroup'] = '(60 - 70)'
my_age_data.loc[( (my_age_data.Age >= 70) & (my_age_data.Age < 80 )),  'AgeGroup'] = '(70 - 80)'
my_age_data.loc[( (my_age_data.Age >= 80) & (my_age_data.Age < 90 )),  'AgeGroup'] = '(80 - 90)'
my_age_data.loc[( (my_age_data.Age >= 90) & (my_age_data.Age < 100 )),  'AgeGroup'] = '(90 - 100)'
my_age_data.loc[(my_age_data.Age >= 100),  'AgeGroup'] = '> 100'
```

In [28]:
```
my_age_data.head()
```

Out[28]:

	ID	LastName	FirstName	DateOfBirth	Today	Age	AgeGroup
0	143.0	Turner	Marion	16/08/1932	19/05/2021	88	(80 - 90)
1	7864.0	Pollock	James	3/03/1901	19/05/2021	120	> 100
2	7701.0	SMART	IISA	2/02/1902	19/05/2021	119	> 100
3	174.0	Gomez	Felix	19/04/1945	19/05/2021	76	(70 - 80)
4	90.0	Morris	Donna	16/08/1930	19/05/2021	90	(90 - 100)

Marital Status :

Our data had different values of partnership status and I merged them all into two categories, of having a partner -1 or not having a partner -0.

In [66]:
```
# Removing rows with Unknown Values for Marital Status
my_married_data = my_married_data[my_married_data.MaritalStatus != 1]
```

In [68]:
```
# Assigning Partner = 0 for Single/Divorced/Widowed
my_married_data.loc[(my_married_data.MaritalStatus == 2) | (my_married_data.MaritalStatus == 4) |
                    (my_married_data.MaritalStatus == 5) | (my_married_data.MaritalStatus == 6),  'Partner'] = 0

# Assigning Partner = 1 for Married/Committed/Partner Relationships
my_married_data.loc[(my_married_data.MaritalStatus == 3) | (my_married_data.MaritalStatus == 9) |
                    (my_married_data.MaritalStatus == 11) ,  'Partner'] = 1
```

In [69]:
```
# Checking for Marital Status after cleaning
my_married_data[["ID","LastName","FirstName","MaritalStatus","Partner"]].head()
```

Out[69]:

	ID	LastName	FirstName	MaritalStatus	Partner
34	7575	kx	jo	2.0	0.0
47	6513	Sheppard	Shirley	5.0	0.0
93	7435	DREWETT	PETER	5.0	0.0
126	6550	Healy	Ivan	5.0	0.0
172	7515	Treseder	Wendy	11.0	1.0

Model Development

Model 1 :

This model takes into account the factors of age, gender and partnership status in order to predict the hospital admissions.

```
In [1]: # Model 1 : Age, Gender, Partner

        # Reading in the data into the dataframe fuel
        model_1 <- read.csv("Model_Group_1.csv")

In [2]: # Our Target : Previous_Admission

        # Fitting in a multiple linear model
        # "Previous_Admission ~." expression indicates all variables other than Previous_Admission that will be used for prediction
        fit1 <- lm(Previous_Admission ~ ., model_1)

In [3]: # Details of the Linear Model
        summary(fit1)

        Call:
        lm(formula = Previous_Admission ~ ., data = model_1)

        Residuals:
             Min       1Q   Median       3Q      Max
        -1.20114 -0.22599  0.00925  0.33464  0.92593

        Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
        (Intercept) -1.613329   0.625784  -2.578   0.0162 *
        Gender       0.289470   0.215835   1.341   0.1919
        Age          0.054231   0.005288  10.255 1.92e-10 ***
        Partner      0.241115   0.311208   0.775   0.4457
```

Model 2 :

This model includes the factors of functional ability and family support index in order to predict the hospital admissions.

```
In [4]: # Model 2 : Functional ability and Family Support Index

        # Reading in the data into the dataframe model_2
        model_2 <- read.csv("Model_Group_2.csv")

In [5]: # Our Target : Previous_Admission

        # Fitting in a multiple linear model
        # "Previous_Admission ~." expression indicates all variables other than Previous_Admission that will be used for prediction
        fit2 <- lm(Previous_Admission ~ ., model_2)

In [6]: # Details of the Linear Model
        summary(fit2)

        Warning message in summary.lm(fit2):
        "essentially perfect fit: summary may be unreliable"

        Call:
        lm(formula = Previous_Admission ~ ., data = model_2)

        Residuals:
               Min         1Q     Median         3Q        Max
        -2.589e-15  0.000e+00  1.415e-16  1.762e-16  1.993e-16

        Coefficients:
                               Estimate Std. Error    t value Pr(>|t|)
        (Intercept)           5.000e+00  3.060e-15  1.634e+15   <2e-16 ***
        Functional_Ability   -1.000e+00  6.532e-16 -1.531e+15   <2e-16 ***
        Family_Support_Index  1.568e-16  6.061e-16  2.590e-01    0.798
        ---
        Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 3 :

This model included the factors of need for care and the social support index, in order to predict the hospital admissions.

In [7]:
```r
# Model 3 : Need for care and Social Support Index

# Reading in the data into the dataframe named model_3
model_3 <- read.csv("Model_Group_3.csv")
```

In [8]:
```r
# Our Target : Previous_Admission

# Fitting in a multiple linear model
# "Previous_Admission ~" expression indicates all variables other than Previous_Admission that will be used for prediction
fit3 <- lm(Previous_Admission ~ ., model_3)
```

In [9]:
```r
# Details of the Linear Model
summary(fit3)
```

```
Warning message in summary.lm(fit3):
"essentially perfect fit: summary may be unreliable"
```

```
Call:
lm(formula = Previous_Admission ~ ., data = model_3)

Residuals:
       Min         1Q     Median         3Q        Max 
-1.996e-16  7.425e-18  8.811e-18  1.020e-17  1.020e-17 

Coefficients:
                       Estimate Std. Error    t value Pr(>|t|)    
(Intercept)           0.000e+00  2.897e-17  0.000e+00        1    
Need_for_care         1.000e+00  4.977e-17  2.009e+16  < 2e-16 ***
Social_Support_Index -4.487e-16  4.618e-17 -9.716e+00 3.85e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```