

A Study on Days of hospitalization of insured with the claims data

G Srimannarayana^a, P Venkata Rajesh Kumar^b and Dr. P. Dhanavanthan^c

^aFaculty, Institute of Insurance & Risk Management, Hyderabad, India, srimannarayana@iirmworld.org.in

^bDentist, Texas, USA

^cProfessor, Department of Statistics, Pondicherry University, Puducherry

Article History: Received: 10 November 2020; Revised 12 January 2021 Accepted: 27 January 2021; Published online: 5 April 2021

Abstract: The health insurance sector has grown at a double digit growth rate in India in the past decade. The Government schemes for the individuals' insurance coverage from the low-income group have resulted in higher penetration. Raise in the disposable income of the middle income and awareness of health insurance has led to self-subscription by private individuals. An increase in insurance penetration has also led to an increment in the claims for the insurer. However, the insurance coverage has been actively subscribed by the population in the age group of 19-64 years. The average claim amount processed is higher in infants and aged people. Days of hospitalization for the insured treatment help the insurers derive the claims' amounts and hence can budget the reserves accordingly. Days of hospitalization is found to have a strong positive relationship with age. Logistic regression of Half-yearly bins data has the predictability of Days of hospitalization of claimants at 69%.

Keywords: Age, Claims, Days of hospitalization, Health insurance

1. Introduction

The united health insurance covered State-owned employees in India for a long time. Overburdened government hospitals in terms of treatment and lack of human resources and equipment have led the government to support the common man for alternative treatments. Huge advancements have taken place in medical treatments. Active incorporation of advanced technologies by private health centers and medical experts has motivated the common people to access the private medical infrastructure.

The government has taken multiple schemes to support the common people and government employees to access the medical infrastructure as per their convenience. Most of the schemes were insurance coverage extended to common people with the active participation of multiple insurers. Health insurance premium collected in India in 2019 is Rs. 27,400 crores. Government schemes have contributed 20% of total premiums and the rest by individuals' group health subscribers and retail policies.

Health insurance policies are offered by both the general insurers and standalone health insurers. Post-2000, there has been increased competition because of the entry of private insurers. Overall there are more than 30 players in the present Indian market. This increased competition has led to innovation in the health insurance policy features and provision of flexibility to the subscribers. The health insurance market is expected to quintuple its premium size in the next decade. Standalone health insurers are expected to grow at more than a 20% p.a. growth rate in the future.

Insurance penetration in India has hardly reached 5% of the population. Though there have been higher subscriptions in life insurance, subscription to health insurance is a new phenomenon. Public sector units and government entities had ensured to have group health schemes coverage for the employees from the initial days. Private firms in the manufacturing sector before 2000 had a lesser focus on the insurance coverage of the employees.

Post-2000, the Rapid growth of the service sector driven by private firms, has enabled the private employees to be covered by group health schemes. Both the state and central governments have increased efforts to provide health insurance coverage to low-income groups with insurers as partners. Education of the middle-income group has also increased awareness of health insurance for self-subscription by retail individuals.

An increase in awareness and disposable income has driven private individuals to be a part of either group health insurance coverage or retail policy coverage. Private medical infrastructure has a lesser load compared to government medical infrastructure. Focus on quality treatment has made high and middle-income groups to access the private medical infrastructure.

This rapid increment in the health insurance subscription has led to the increment in the claims by subscribers. There has been standardization undertaken by insurers in the past decade for a claims settlement process. Third part Administrators (TPA's) have actively involved in ensuring the documentation and processing of the insured's claims settlement by the insurers. TPA is playing a vital role in communicating between the insurer and the insured.

Insurers have also standardized the prices for standard medical treatments like cataracts etc. so that the insured can avail the better value from the insurance coverage. Price standardization also helps the insurers by preventing the malpractices of private health infrastructure. Insurance coverage has provided access to medical infrastructure to common people and has also burdened insurers with huge claim amounts.

The insurer's higher premium quotation may lead to losing the customer to the competitors, and a lower premium quotation would result in the insurer's losses. Hence, premium pricing becomes a vital activity of the insurers. It becomes essential for insurers to understand the pattern and amount of claims raised by different subscribers. These patterns help insurers to maintain the reserve for the claim settlement. Simultaneously, the study would also help the insurer rightly underwrite a policy and calculate the premium.

2. Literature Review

Days of hospitalization of the insured have a strong negative impact upon the insurer's profits; the longer the days of hospitalization, the higher is the burden upon the insurer by the claimants. It becomes vital for the insurers to understand the pattern of hospitalization days, which in turn helps in projecting the claims levels. Insurers would be able to keep the reserve aside in the case of prior information of claim levels. The age, comorbidities, readmissions, and preexisting health conditions impact the days of hospitalization of the policy subscribers. Preexisting health conditions increase the readmissions of the patients. High readmissions would result in a higher number of hospitalization days and hence a higher burden upon the insurer in terms of claims.

Health insurance has a higher impact on the health of vulnerable groups like infants, children, and individuals with chronic diseases [12]. However, there has little concrete proof on the impact of aged people by health insurance. Studies have also found that age and weight are not useful for predicting the days of hospitalization [1]. The study finds that the number of admissions to the hospital was high in preschool kids than the rest of the population. Though the patients have spent 23 days of hospitalization, the study finds no specific relation between the age and hospitalization days [6].

Predictive analytics helps in the detection of chronic diseases at an early stage. Early detection will help in timely treatment and decrease hospitalization days, lowering the burden upon the medical infrastructure [10]. Information of Past medical conditions and hospitalization days for treatment helps to predict the future hospitalization days for the patients, Model fits better for the population age above 63 years compared to rest of the population [18]. Multi morbidity increases contribution towards the hospitalization days and hence results in a higher burden upon the medical infrastructure and the insurers [17].

Patient & hospital characteristics, patient demography, and past medical information of the patients enable to predict the readmission of patients, especially for congestive heart failure [2]. The study finds the treatment of colorectal resection; there is high relation between the days of hospitalization and readmission [14]. Longer is the hospitalization; the higher is the probability of readmission. With multiple chronic conditions, young and adult populations with mental disorders, children transitioning to adulthood had higher readmissions than others in the USA [4]. There have been lesser readmissions in old age groups.

The study finds that medical insurance's continuous coverage for people between the ages of 55 and 65 would improve the health position from average to good and excellent [9]. This improved health would ensure that a lower number of aged people for medical treatment. They are resulting in a lower number of hospitalization in advanced ages. Similarly, Initiative help by nursing homes, enhanced staff skills, and perception of the facility will positively impact in-house patients' health [11].

New health information technologies had no significant difference in the impact on health between different age groups [16]. The study finds that older people are also adaptable to new health information technologies for self-care and health treatment. The aged group has utilized self-care resources significantly more than young adults. Better inpatient management with the objective of lower length of stay, avoidable hospitalization, reducing readmissions forms important parameters for reducing the hospitalization days, especially for the aged people [5].

Age-specific classification can also be undertaken with Age-Phenome Knowledge-base (APK) [8]. APK is derived for age categorization with respect to disease occurrence and treatment. APK categorization is found to be better than the established Medical Subject Headings (Mesh) categorization for the objective of medical research for disease occurrences.

Claims data is majorly used for the research in medical access. Claims data might not be useful to analyses of mortality and morbidity, as the data errors would have wrong implications [7]. The study analyses the data on the basis of bins of claims data. The data were sorted on the basis of customers and binned (accrued values of claims, days of hospitalization) in different periods. Analysis of half-yearly bins has a better ability to predict the number of days of hospitalization of the patients [19].

Bagging tree, Adaboost, and Random forest decision tree methods were used to project the hospitalization days. Adaboost decision tree method had better results than other tests in projecting the hospitalization days in the future based on past claims data [13]. Line regression models are used to predict the hospitalization days for the heart issue patients [15]. The predicted values had a precision of more than 60%. Prediction of hospitalization days was efficient in discharges of the patients; however, the study was not good enough in early discharges. Logistic regression methods with binning of data were used to predict the hospitalization days [3].

This study explores to understand the relation between the number of days of hospitalization, age, and the claimed amount.

3. Data & Methodology

Claims data for 2007, 2008 & 2009 years are collected from the Insurance Information Bureau of India (IIB). 72,114 records are processed for the analysis.

Data is classified into nine age groups as per Age-Phenome Knowledge-base (APK) [8]. Infant (0-2 years), preschool kids (3-5 years), child (6-13years), adolescent (14-18 years), Young adult (19-33 years), adult (34-48 years), middle-aged (49-64 years), aged (65-78 years), old (79 and above).

Descriptive and inferential statistics are applied for the analysis of the claims data in the initial step. In the second step, Collected data is sorted, w.r.t customers. Both the insurance coverage and claim amounts are converted into logarithmic values, and logarithmic values are used in the analysis. Regression models are applied to understand DOH's relation, claims amount, age, and the coverage amount. The number of days of hospitalization is binned in multiple ways. Bins are formed for half-yearly, quarterly, and bimonthly. i.e., for quarterly, the total number of days of hospitalization is calculated based on the total number of days the insured has spent in that quarter (days are accrued in case of multiple visits in the same quarter). So for half-yearly, there were two bins formed for each policy. Similarly, for quarterly, four bins were formed. Finally, for the Bimonthly period, six bins were formed.

4. Analysis

The average age of the treated patients has been on the lower side of each group's average group range. The highest number of claims are processed for Young adults. Middle-aged and adults groups also have raised a higher number of claims. More than 72% of the total number of claims are raised by groups in age, ranging from 19 to 64 years. The child group forms the highest number of claims for the patients below the age of 18 years. In comparison, aged has a higher number of claims for the patients above 65 years.

Category	Average Age	Number of Claims	Percentage of Total Number of Claims
Infant (0-2 years)	0.64	2,891	4.01%
Preschool Kids (3-5 years)	2.86	2,788	3.87%
Child (6-13 years)	8.21	4,107	5.70%
Adolescent (14-18 years)	15.05	2,018	2.80%
Young Adult (19-33 years)	26.34	20,676	28.67%
Adult (34-48 years)	39.78	15,710	21.78%
Middle Aged (49-64 years)	55.34	16,027	22.22%
Aged (65-78 years)	69.05	7,048	9.77%
Old (79 years and above)	81.21	849	1.18%

Table 1: Details of the number of claims

The middle-aged group has claimed the highest claim amount among all age groups (refer to table 2). Claims paid to the middle age group is around 27.82% of total claims. Young adults have been paid 24.11% of the total

claim amount. Adults were paid around 21.84% of the total claim amounts. That is, the people in the age group of 19-64 years are paid 73.76% of the total claims amount settled. 19-64 years age groups are more eligible for insurance coverage, as they have higher earning capability and fitness in health. As the 19-64 age group dominates insurance subscriptions, the same age group also forms the lion share in claims amounts settled (refer figure 1). Diseases related to the circulatory system are prevalent in middle-aged groups, and higher treatment costs result in higher claims.

Age Category	Average Claim Amount Paid	Percentage of Total Claims Paid
Infant (0-2 years)	11,468.73	2.48%
Preschool Kids (3-5 years)	8,007.76	1.67%
Child (6-13 years)	10,014.98	3.08%
Adolescent (14-18 years)	12,149.89	1.84%
Young Adult (19-33 years)	15,557.29	24.11%
Adult (34-48 years)	18,547.52	21.84%
Middle Aged (49-64 years)	23,158.08	27.82%
Aged (65-78 years)	28,249.70	14.92%
Old (79 years and above)	35,136.46	2.24%

Table 2: Details of the amount of claims paid

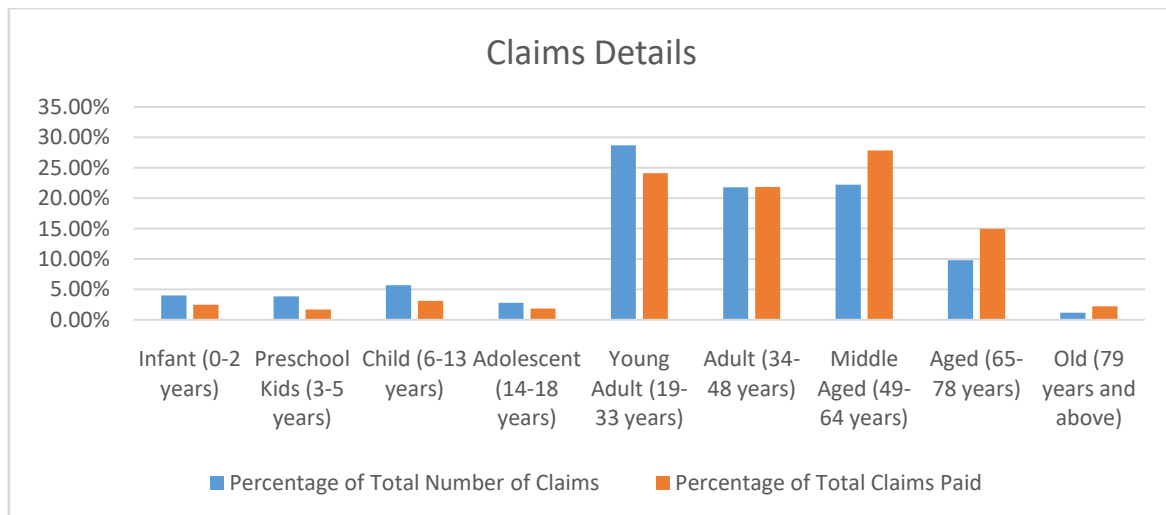


Figure 1: Claim details

Insurance coverage to retail policyholders for the aged and old is scarce (refer to table 3). However, they are generally covered in the group health insurance policies and government-driven schemes. 14.92% of total claim amounts are processed for aged, whereas claims of old were merely 2.24% of the total claim amount.

The average claim amount paid to the middle-aged is higher than the average amount paid to young adult and adult groups (refer to table 3). Though old age has formed a fraction of total claims amounts paid, the amount per claim is the highest among all age groups. Both the aged and old age groups have a higher amount settled per claim. This reflects that the treatment cost of treatment increases with age.

Infants require multiple hospital visits for vaccination and for the treatment of common diseases. However, speaking ability would ensure preschool kids have a lower number of visits to the hospitals. Individuals in adolescence would be aware of the hygiene and daily exercises to prevent common diseases resulting in a lower number of hospital visits. Preschool kids and adolescents are found to have claimed the least amount from the insurers.

The average claim amount paid to the infants is higher than the preschool kids and child groups. This is because of the higher costs incurred in the treatment of infants. Though the child group has a lower average claim amount, a higher number of claims resulted in higher claims amount paid.

Average insurance coverage has increased with the increment in the age group. Average insurance coverage is lowest for preschool kids, and the old group has the highest average insurance coverage. This is due to high medical expenses incurred for older people's treatment; hence coverage subscribed too is high.

Age Category	Average Insurance Coverage	Percentage of Total Insurance Coverage
Infant (0-2 years)	129,661	4.16%
Preschool Kids (3-5 years)	95,540	2.96%
Child (6-13 years)	101,091	4.61%
Adolescent (14-18 years)	99,562	2.23%
Young Adult (19-33 years)	116,628	26.77%
Adult (34-48 years)	122,524	21.37%
Middle Aged (49-64 years)	141,115	25.10%
Aged (65-78 years)	144,833	11.33%
Old (79 years and above)	156,853	1.48%

Table 3: Details of the insurance coverage amount

High subscription of the 19-64 age group has resulted in higher total insurance coverage for the same segment (more than 72% coverage). 11.33% of total insurance coverage is to the aged group, and a mere 1.48% of total insurance coverage is towards the old group. Preschool kids and adolescents have lower insurance coverage compared to the rest of the groups.

Age Category	Average Days in Hospital	Percentage of Total Days of Hospitalization
Infant (0-2 years)	36.19	3.90%
Preschool Kids (3-5 years)	34.69	3.61%
Child (6-13 years)	33.34	5.10%
Adolescent (14-18 years)	35.10	2.64%
Young Adult (19-33 years)	37.96	29.26%
Adult (34-48 years)	36.39	21.31%
Middle Aged (49-64 years)	38.15	22.79%
Aged (65-78 years)	38.10	10.01%
Old (79 years and above)	43.45	1.38%

Table 4: Details of Days of Hospitalization (DOH)

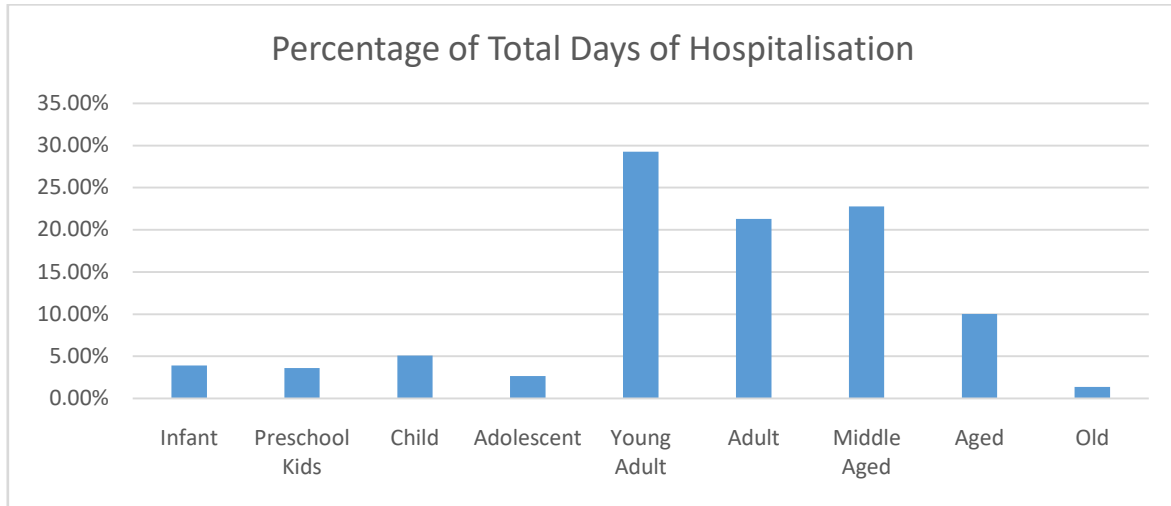


Figure 2: Percentage of total days of hospitalization

The average hospitalization for all the groups is above 33.34 days (refer tabl4 & figure 2). The old group has the highest average hospitalization days, and the least is with the Child group. However, with the people's highest claims between 19 to 64 years, total hospitalization days are highest for the same age groups. 73% of total hospitalization days have been incurred for the adult, middle-aged, and young adults.

5. Prediction of Days of Hospitalization

The regression model is applied to predict the number of days of hospitalization. The number of days of hospitalization (DOH) is the dependent variable, and the insurance coverage amount, claims amount, age is the independent variables. Both the claims amount and insurance coverage amount are transformed with a logarithmic function before applying the regression model.

$$DOH = \beta_1 * Age + \beta_2 * Ln(\text{Coverage amount}) + \beta_3 * Ln(\text{Claims amount}) + \alpha$$

	Model	Bin Period	Number of Observations	R square	Significance
1	Model 1	Half Yearly	35,593	69.00%	0.00
2	Model 2	Quarterly	37,825	60.70%	0.00
3	Model 3	Bimonthly	39,145	57.60%	0.00

Table 5: Summary of models

Model 1: Prediction with Half-yearly bins

There were around 35,593 half yearly bins formed based on the total policies considered. The model results were significant.

	Beta	t	Sig.
(Constant)	112.77	6.73	0.00
coverage	0.01	4.10	0.00
claims	-0.03	-10.47	0.00
age	0.84	274.40	0.00

Table 6: Coefficients of the independent variables in Model 1

Coverage amount and age had a positive relationship with the dependent variable. At the same time, claims have negative with minuscule sensitivity towards the DOH (refer to table 6). Age has a significant positive relation with DOH.

$$\text{DOH} = 0.84 * \text{Age} + 0.01 * \text{Ln (Coverage amount)} - 0.03 * \text{Ln (Claims amount)} + 112.77$$

69% variation of the DOH are explained with help of the Age, Coverage and claims amount.

Model 2: Prediction with quarterly bins

There were around 37,825 quarterly bins formed based on the total policies considered. The model results were significant.

	Beta	t	Sig.
(Constant)	3.27	0.26	0.80
coverage	0.01	4.29	0.00
claims	-0.01	-2.66	0.01
age	0.78	233.01	0.00

Table 7: Coefficients of the independent variables in Model 2

Coverage amount and age had a positive relationship with the dependent variable. At the same time, claims have negative with minuscule sensitivity towards the DOH (refer to table 7). Age has a significant positive relation with DOH.

$$\text{DOH} = 0.782 * \text{Age} + 0.014 * \text{Ln (Coverage amount)} - 0.01 * \text{Ln (Claims amount)} + 3.27$$

60.7% variation of the DOH are explained with help of the Age, Coverage and claims amount.

Model 3: Prediction with Bimonthly bins

There were around 39,125 two month bins formed based on the total policies considered. The model results were significant.

	Beta	t	Sig.
(Constant)	4.07	0.38	0.70
coverage	0.02	4.44	0.00
claims	-0.01	-2.77	0.01
age	0.76	220.78	0.00

Table 8: Coefficients of the independent variables in Model 3

Coverage amount and age had a positive relationship with the dependent variable. In comparison, claims have a negative with minuscule sensitivity towards the DOH (refer to table 8). Age has a significant positive relation with DOH.

$$\text{DOH} = 0.76 * \text{Age} + 0.02 * \text{Ln (Coverage amount)} - 0.01 * \text{Ln (Claims amount)} + 4.07$$

57.6% variation of the DOH are explained with help of the Age, Coverage and claims amount.

In summary, logistics regression with the half-yearly bins has better predictability of hospitalization days than quarterly or bi-monthly bins [15]. Age has a strong influence on the hospitalization days in the context of India. Both the coverage amount and claims amount has the least role in the prediction of the DOH.

6. Conclusion

The increment in the middle-income group's disposable income and rapid expansion of private health infrastructure has led to the health insurance subscriptions increment.

High subscription is found in the age group of 19-64 years as they have the high earning capability and physically fit; hence insurers are biased towards offering coverage to this segment. This high subscription has also resulted in more number of claims raised by the same age group. Insurers have also paid a major share claim amount to this particular segment. Middle-aged people have dominated in terms of the number of claims and claim amounts paid.

In the segment less than 19 years, the child group has a high number of claims and claim amounts paid. However, infants had a higher average claim amount. Similarly, in the segment with age greater than 64 years, it is the Aged which have a higher number of claims and claims amount disbursement. But, the average claim amount is higher for old. The treatment of old people is generally high because of comorbidities, preexisting medical conditions, and advanced age levels.

The average number of days in the hospital is similar across the age groups; older people do have a higher number of days for hospitalization. Peculiarly, it is the young adults who have spent more hospital days than the rest of the groups. 19-64 years groups were incurring more than 73% of all segments' total number of hospitalization days.

Logistic regression models have concluded that the DOH can be explained with the help of the age, coverage amount, and claimed amount of the policyholder. It is also found that the model with half-yearly bins has a higher explain ability than the quarterly or bimonthly bins. All the models have found that age has a strong positive relationship with the DOH. i.e., higher age leads to a higher number of days of hospitalization. Insurance coverage amount has a positive relation, with minuscule sensitivity. The coverage amount has a negative relation with minuscule sensitivity; this reflects a low moral hazard for the health insurers in the Indian context.

References

- Al-Radi, Osman O. Are Neonatal Age and Small Weight Predictive of In-Hospital Death and Prolonged Hospital Stay in Children Undergoing Heart Surgery?, 2019. *Cardiothoracic surgeon* 28
- Bardhan, Indranil, Jeong-ha C. Oh, Zhiqiang E. Zheng, and Kirk Kirksey. "Predictive Analytics for Readmission of Patients with Congestive Heart Failure." *Information Systems Research*. 26.1 (2015): 19-39. Print.
- Barnes, Sean, Eric Hamrock, Matthew Toerper, Sauleh Siddiqui, and Scott Levin. 2016. "Real-time prediction of inpatient length of stay for discharge prioritization". *Journal of the American Medical Informatics Association*. 23 (e1): e2-e10.
- Berry Jay G, Gay James C, Joynt Maddox Karen, Coleman Eric A, Bucholz Emily M, O'Neill Margaret R et al. Age trends in 30 day hospital readmissions: US national retrospective analysis *BMJ* 2018; 360 :k497
- Chi Leung Kwok, Carmen KM Lee, William TL Lo, and Paul SF Yip. 2017. "The Contribution of Ageing to Hospitalisation Days in Hong Kong: A Decomposition Analysis". *International Journal of Health Policy and Management*. 6 (3): 155-164.
- Dixon T, Shaw M, Frankel S, Ebrahim S. Hospital admissions, age, and death: retrospective cohort study. *BMJ*. 2004;328(7451):1288. doi:10.1136/bmj.38072.481933.EE
- Ferver, Kari, Bryan Burton, and Paul Jesilow. 2009. "The Use of Claims Data in Healthcare Research". *The Open Public Health Journal*. 2 (1): 11-24.
- Geifman N, Cohen R, Rubin E. Redefining meaningful age groups in the context of disease. *Age (Dordr)*. 2013;35(6):2357-2366. doi:10.1007/s11357-013-9510-6
- Hadley, J., & Waidmann, T. (2006). Health insurance and health at age 65: implications for medical care spending on new Medicare beneficiaries. *Health services research*, 41(2), 429–451. <https://doi.org/10.1111/j.1475-6773.2005.00491.x>
- K, Geetha Poornima and K, Krishna Prasad, Integrated Prediction System for Chronic Disease Diagnosis to Ensure Better Healthcare (April 2, 2020). *International Journal of Health Sciences and Pharmacy (IJHSP)*, 2020. Vol. 4 (1)
- Khatutsky, G, L E Bercaw, A Shah, K Porter, T Eng, C Ormond, P J Zoromski, and C Husick. 2018. "Evaluating Initiative to Reduce Avoidable Hospitalizations of Nursing Home Residents: Facility Perspective". *Innovation in Aging*. 2 (suppl_1): 88
- Levy, Helen & Meltzer, David (2008) *The Impact of Health Insurance on Health*, Annual Review of Public Health, 29, pp 399 409
- Ma, Fei & Yu, Limin & Ye, Lishan & Yao, David & Zhuang, Weifen. (2020). Length-of-Stay Prediction for Pediatric Patients With Respiratory Diseases Using Decision Tree Methods. *IEEE Journal of Biomedical and Health Informatics*. PP. 1-1. 10.1109/JBHI.2020.2973285.
- Maria Kelly, Linda Sharp, Fiona Dwane, Tracy Kelleher and Harry Comber : Factors predicting hospital length-of-stay and readmission after colorectal resection: A population-based study of elective and emergency admissions, *BMC Health Services Research* 2012, 12:77

- Pei-Fang, (Jennifer) T, Chen Po-Chia, Chen Yen-You, Song Hao-Yuan, Lin Hsiu-Mei, Lin Fu-Man, and Huang Qiou-Pieng. "Length of Hospital Stay Prediction at the Admission Stage for Cardiology Patients Using Artificial Neural Network." *Journal of Healthcare Engineering*. 2016
- Wagner, Lisa Smith, and Todd H. Wagner. 2003. "The Effect of Age on the Use of Health and Self-Care Information: Confronting the Stereotype". *The Gerontologist*. 43 (3): 318-324.
- Wei, Melissa Y, Nicholas Tilton, and Kenneth J Mukamal. 2019. "Quantifying the Burden of Hospitalized Days in Medicare Beneficiaries with Multimorbidity". *Innovation in Aging*. 3 (Supplement_1): S924.
- Xie, Yang & Schreier, Günter & Chang, David & Neubauer, Sandra & Liu, Ying & Lovell, Nigel. (2015). Predicting Days in Hospital Using Health Insurance Claims. *IEEE journal of biomedical and health informatics*. 19. 10.1109/JBHI.2015.2402692.
- Xie, Yang, Günter Schreier, Michael Hoy, Ying Liu, Sandra Neubauer, David C. W. Chang, Stephen J. Redmond, and Nigel H. Lovell. "Analyzing Health Insurance Claims on Different Timescales to Predict Days in Hospital." *Journal of Biomedical Informatics*. 60 (2016): 187-196. Print.