ORIGINAL ARTICLE

A Multivariate logistic regression model of Healthcare service utilization in district Kanpur Nagar, Uttar Pradesh

Amrish Chandra[1], Alpana Srivastava[2], Bhagwan Jagwani[3], Santosh Kumar Barman[4], Nidhi Nagar[5]

[1,2]Amity University, Uttar Pradesh, Lucknow campus, Lucknow, India
[3]School of Commerce & Business Management, Geeta University, Haryana, India
[4]Ganesh Shankar Vidyarthi memorial medical college, Kanpur, Uttar Pradesh, India
[5]D.A-V PG College (C.S.J.M.U), Kanpur, Uttar Pradesh, India

CORRESPONDING AUTHOR
Amrish Chandra, Research Scholar, Amity University Uttar Pradesh, Lucknow-226010, India
Email: amrishchandra@gmail.com

CITATION
Chandra A, Srivastava A, Jagwani B, Barman SK, Nagar N. A Multivariate logistic regression model of Healthcare service utilization in district Kanpur Nagar, Uttar Pradesh. Indian J Comm Health. 2025;37(3):453-458.
https://doi.org/10.47203/IJCH.2025.v37i03.016

ARTICLE CYCLE
Received: 16/04/2025; Accepted: 21/05/2025; Published: 30/06/2025

ABSTRACT

Background: To become a developed economy for good health of citizens is of prime importance. As healthcare facilities are getting costlier day by day, people are either avoiding to utilise health care services or they do incur lot of out of pocket expenditure (OOPE). Several independent factors also affect the utilization of healthcare facility by individual. Aim & Objective: The aim of this study was to explore factors significantly influencing healthcare utilization by constructing a predictive model using multivariate logistic regression. Methodology: A cross-sectional empirical study was conducted across urban and rural areas of district Kanpur Nagar using semi structured, pretested questionnaire. A sample of 400 respondents (264 urban and 136 rural) was selected using probability proportion to size, followed by multi stage random sampling. Analysis was done using different statistical tools. Results: Age, education and income were positively associated with dependent variable healthcare utilization while distance to healthcare facility is negatively associated with healthcare utilization. Conclusions: Result of multivariate logistic regression indicates various factors are significantly associated with healthcare utilization. Health insurance coverage to be increased for elderly and low income individuals, while accessibility for healthcare facility to be increased for rural areas.

KEYWORDS

Health Insurance, Utilization, Binary variable, Healthcare services

INTRODUCTION

Though utilization of healthcare facility is facilitated by health insurance, there are still many factors that affect the uptake of health insurance. Aging is an inevitable process with increasing health complexity and lower financial independence (1). Proportion of female visiting healthcare centre were higher than males (2). Married individuals preferred private healthcare facility as compared to widowed men and women (3). As per health dossier 2021: reflections on key health indicators- Uttar Pradesh, published by National Health Systems Resource centre (NHSRC), National family health survey 4 (NFHS 4, 2015-16) could found, 6.1% of households of Uttar Pradesh had at least one member covered under any type of health insurance or health financing scheme, which increased to 15.9% in NFHS 5. At national level these figures stood at 29% in NFHS-4 and 41% in NFHS-5 (13,14). These suggest that health insurance coverage in Uttar Pradesh is well below the National level. Despite Kanpur being one of the most populous and industrial hubs of Uttar Pradesh contributing around 3% of State's GSDP, number of families who had at least one member covered under any type of health insurance in Kanpur Nagar is 16.1% and in Kanpur Dehat 17.9%, which is very low as compared to national level (15). In view of these findings the present study was planned to

find out the status of health insurance coverage in district Kanpur Nagar.

Aim & Objective(s): The present study titled "Multivariate logistics regression modelling of healthcare service utilization in district Kanpur Nagar, U.P. was planned with following objectives 1. To find out the significant factors influencing healthcare services utilization. 2. To construct and evaluate a predictive model for the likelihood of healthcare service usage using multivariate logistic regression. 3. To develop model to estimate healthcare services utilization.

MATERIAL & METHODS

Study type: Observational study. **Study Design:** Analytical Cross-sectional study. **Study area:** Study was conducted among subjects residing in rural and urban areas of district Kanpur Nagar. **Study population and Duration:** Citizens of district Kanpur Nagar over 18 years of age **Duration of this study:** 06 months from September 2024 to March 2025. **Sample size:** As per health and morbidity survey 2014, by National Health Systems Resource Centre (NHSRC), proportion of ailing persons (per thousand) who utilised health care services stood at 91 for Uttar Pradesh, and 118 for overall India. Hence taking a prevalence of 0.09 and considering 50% response rate sample size is 400. Kanpur Nagar is divided in 10 blocks, 6 zones and 110 wards. Rural population of Kanpur Nagar district is 15,65,623 (34.17%), Urban population is 30,15,645 (65.83%), hence total sample of 400 respondents is divided into 136 (1/3rd) rural respondents and 264 (2/3rd) urban respondent.

Rural respondents - Out of 10 rural blocks, 2 rural blocks (Kalyanpur and Ghatampur) were selected randomly via lottery method, 5 villages per block and 27 respondents per village was selected which counts to 270 respondents. out of 270, 136 respondents have given complete information. **Urban respondent** - Out of 80 urban wards, 6 wards were selected randomly, 1 each from 6 zones namely Civil Lines zone1, Chakeri zone 2, Transport Nagar zone3, Ashok Nagar zone4, Panki zone 5, Kalyanpur Aawas Vikas zone 6. Further from each ward two mohalla's were selected via lottery method and 27 respondents from each mohalla which counts to 324 respondents. out of 324, 264 respondents have given complete information.

Inclusion and Exclusion criteria: Persons aged over 18 years and given informed consent to participate were included whereas sick respondents were excluded. **Data collection:** Data was collected using a semi structured predesigned and pretested questionnaire through direct interview. **Ethical Consideration:** This study did not involve any interventional procedure or drug administration and no personal identifiers were asked maintaining the anonymity of participants, as such ethical approval was not required. **Data Analysis:** Data collected was entered in excel and a master sheet was prepared, data was analysed using SPSS 19.0, percentages were calculated for descriptive statistics and multivariate logistics regression analysis was used to estimate the odds of utilizing healthcare services. Model fit statistics (e.g., Hosmer-Lemeshow test, AUC-ROC) to assess the accuracy and performance of the model.

RESULTS

Table 1: Demographic and Socio-economic profile of the respondents

	18-24	**14% (56)**
Age (in years)	25-54	42% (168)
	55-64	28% (112)
	65 and above	16% (64)
	Range 18 -80 years	Mean Age 47.5 years, S.D 16.52
Gender	Male	48% (192)
	Female	52% (208)
Marital Status	Single	41% (164)
	Married	59% (236)
	Illiterate	13% (52)
Educational Level	Primary	27% (108)
	Graduate	42% (168)
	PG & Above	18% (72)
Income Level	Low income group (LIG)	18% (72)
	Middle income group (MIG)	54% (216)
	High income group (HIG)	28% (112)
Health Insurance Coverage	No	63% (252)
	Yes	37% (148)
	No	55% (220)

Chronic Health Conditions/ Any morbidity	Yes	45% (180)
	0.5 -1.0	40 % (160)
Distance to nearest healthcare facility (in Km.)	1.0 – 5.0	45% (180)
	Above 5.0	15% (60)
	Range 0.5 km -20km	Mean Distance 4.2 km, S.D 3.91
Employment Status	Unemployed	34% (136)
	Employed	66% (264)
	Very Bad	14% (56)
	Bad	25% (100)
Self-rated Health	Moderate	35% (140)
	Good	21% (84)
	Very Good	5% (20)

The age of sample respondents had range between 18 to 80 years, with mean age 47.5 and standard deviation 16.52, while distance to nearest healthcare facility had range between 0.5 to 20 km, with mean distance 4.2 km and standard deviation 3.91. Fourteen percent of respondents belong to 15-24 years, 42% from 25-54 years, 28% from 55-64 years and 16% respondents were above 65 years' age. Fifty-two percent were females and 48% male while 59% were married and 41% were single. Thirteen percent respondents were illiterate, 27% had primary education, 42% graduate and 18% had PG and above qualification. Eighteen percent belong to LIG, 54% to MIG and 28% to HIG. Thirty-seven percent participants had health insurance coverage while 63% had no health insurance coverage. Forty-five percent respondents were suffering from chronic health conditions while 55% had no health related issues. Forty percent respondents claimed distance to nearest Healthcare facility between 0.5-1.0 km, 45% had healthcare facility between 1-5 km ,15% respondent had nearest health Care facility at more than 5 km distance. Thirty-four percent respondents were unemployed, while 66% were employed. Fourteen percent respondents reported their self-rated health very bad, 25% bad, 35% moderate, 21% good and 5% reported very good.

Table 2: Multivariate logistic regression results for healthcare service utilization

Independent Variables	Beta Coefficient	S.E.	p-value	Significance	Exp (Beta)/ Odds Ratio
Age	0.035	0.49	0.046	Significant	1.036
Gender	0.786	0.16	0.010	Significant	2.195
Marital Status	0.462	0.28	0.019	Significant	1.587
Educational Level	0.312	0.46	0.015	Significant	1.366
Income Level	1.249	0.36	0.001	Significant	3.487
Health Insurance coverage	1.408	0.26	0.001	Significant	4.088
Chronic Health Condition	1.019	0.20	0.012	Significant	2.770
Distance to Healthcare facility	-0.131	0.03	0.046	Significant	0.877
Employment Status	0.521	0.059	0.030	Significant	1.684
Self-rated Health	0.895	0.064	0.001	Significant	2.447

Logistic regression equation

Y= Logit ($\hat{p}$) = -3.450 + 0.035 X_1 (Age) + 0.786 X_2 (Gender) + 0.462 X_3 (Marital status) + 0.312 X_4 (Education level) + 1.249 X_5 (Income level) + 1.408 X_6 (Health Insurance Status) + 1.019 X_7 (Chronic health condition) – 0.131 X_8 (Distance to Healthcare facility) + 0.521 X_9 (Employment status) + 0.895 X_{10} (Self-rated Health)

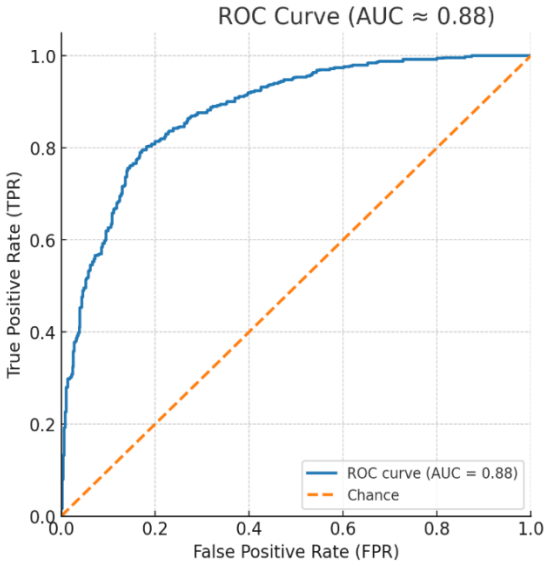
Model Evaluation

The model reliability was adjudged using the Hosmer-Lemeshow (H-L Statistic) goodness of fit. The null hypothesis was, there is no significant difference between the model predictions and observed values or the model fits the observed values well. The insignificance of H-L Statistic p=0.187 implies acceptance of the null hypothesis suggesting a good fit of model. Area Under the ROC Curve (AUC) was 0.88, indicating that the model has excellent discriminative ability.

In this case the Youden's J Index given as the difference of TPR & FPR is determined as ≈ 0.8695; which yields the threshold (cut-off probability) providing the best trade-off between sensitivity and specificity. Thus, a case is classified as "positive" by the model only when the predicted probability exceeds 0.87. In addition to this, at this optimal threshold: TPR (Sensitivity/Recall) = 0.794 which

means that about 79.4% of actual positives are correctly detected by the model and FPR = 0.172 i.e. about 17.2% of actual negatives are incorrectly classified as positives (or percent classified as false alarms).

Figure 1: ROC Curve



Nagelkerke R^2: 0.46, meaning that 46% of the variation in healthcare utilization is explained by the model. These pseudo-R^2 values give an idea of the model's explanatory power.

DISCUSSION

For independent variable age, (β1) = 0.035 and (p = 0.046) along with Exp(Beta)= 1.036 implies that age significantly affects healthcare utilization and for each additional year of age, the odds of utilizing healthcare services increase by 3.6% because of increased health complications with age, similarly Ratan S et al, in their study observed that elderly face financial crunch in utilising health care services (1,18). For independent variable gender, (β2) = 0.786 and (p = 0.010) along with Exp(Beta)=2.195 implies gender significantly affects healthcare utilization and females are 2.19 times more likely to utilize healthcare services compared to males, this finding is similar to the study done by Joseph j et al, where (34.7%) females utilised healthcare against (27.5%) males (2,18).

For independent variable marital status, (β3)=0.462 and (p=0.019) along with Exp(Beta)=1.587 implies marital status significantly affects healthcare utilization and married individuals are 1.59 times more likely to utilize healthcare services compared to singles as most of married individuals are covered through health insurance, these findings are similar to the observations by Hossain B et al, mentioning higher private healthcare expenditure for married elderly than widowed elderly (3).

For independent variable education, (β4) =0.312 and (p=0.015) along with Exp(Beta)=1.366 implies Education significantly affects healthcare utilization and Higher educated individuals are 1.37 times more likely to utilize healthcare services then less educated individuals, similar findings were reported by Kesarwani P et al, that 2/3rd of the mother with secondary education had recommended visit to healthcare services in NFHS 4 and 5 both (4,18).

For independent variable income level, (β5): = 1.249 and (p = 0.001) along with Exp (Beta)= 3.487 implies, income increases the odds by 24.9% and individuals with high income are 3.487 times more likely to utilize healthcare services compared to others, holding all other factors constant, similar trends were observed by Chaudhary A, showing a progressive distribution for healthcare utilisation with higher income quintiles (5).

For independent variable health insurance status, (β6) = 1.408 and (p = 0.001) along with Exp (Beta) = 4.088 implies having health insurance increases the likelihood of utilizing healthcare services by 40.8% and individuals with health insurance are 4.088 times more likely to utilize healthcare services compared to those without health insurance, holding all other factors constant, Similar observations were made by Krishnamoorthy Y et al, indicating higher utilisation of healthcare services by participants with health insurance coverage (55.2%) as compared to those without coverage (42.5%) (6).

For independent variable chronic health conditions, (β7) = 1.019 and (p = 0.012) along with Exp(Beta)= 2.770 implies it is significant and those with chronic conditions are 2.77 times more likely to utilize healthcare facility, similar trends were seen in the study by Goswami P et al, where 16% of individuals with diabetes were hospitalised compare to 9% without diabetes (7). For independent variable distance to healthcare facility, (β8) = -0.131 and (p = 0.046) along with Exp (Beta) = 0.877 implies it is significant and for each additional Km. of distance, the odds of utilizing healthcare services decrease by 12.3%, similar findings were reported by kesarwani P et al, which concluded from NFHS 4 and 5, distance to healthcare facility as a big problem with only 45.4% and 35.9% women had more than 4 ANC visit respectively (4).

For independent variable employment status, (β9) = 0.521 and (p = 0.03) along with Exp(Beta) = 1.684 implies it is significant and employed individuals are 1.68 times more likely to use healthcare services than unemployed, this finding is similar to the study by Yadav R et al, which found salaried employees

utilised healthcare services 2.3 times more than agriculture workers (8).

For independent variable self-rated health, (β10) = 0.895 and (p = 0.001) along with Exp(Beta) = 2.447 implies it is significant and individuals who rate their health as good, very good, or excellent are 2.45 times more likely to utilize healthcare services, Similar trends were observed by Dodd W et al in Krishnagiri district Tamilnadu where individuals with low self-reported morbidity (22.3%) delayed healthcare utilization (9).

Conclusion

Age, gender, marital status, educational level, income level, health insurance coverage, chronic health condition, distance to healthcare facility, employment status and self-rated health were found to be statistically significant factors affecting the utilization of healthcare facilities. Age, Education, Income were positively associated with dependent variable, healthcare utilization. Distance to healthcare facility, was negatively associated with healthcare utilization, exhibiting a distance-decay effect. Individuals who rate their health status as good tend to utilize more healthcare services, as they are more attentive to their health.

Recommendation

Based on findings, this study recommends better insurance coverage for elderly population, females and less educated section of society. As responsibility increases after marriage hence family health insurance plan must be promoted. Low income individuals utilize less healthcare facility, hence premium charged must be within reach of vulnerable class. Preventive healthcare services should be promoted through insurance to enable early detection of disease and prevent their progression into chronic condition. More healthcare service centre are needed to overcome challenges posed by distance.

Limitation of the Study

Study was conducted in Kanpur district only, limiting its generalizability and presence of recall bias cannot be ignored.

Relevance of the Study

This study would help health policymakers design policies to increase healthcare utilization through health insurance by identifying and prioritizing factors affecting healthcare utilization.

Authors Contribution

AC: concept and data collection, AS: conception and design of study BJ: drafting the article, SKB: revising the manuscript critically, NN: analysis and interpretation of data.

Financial Support and Sponsorship

Nil

Conflict of Interest

There are no conflicts of interest.

Acknowledgement

The authors are thankful to all the respondent for their participation.

Declaration of Generative AI and AI Assisted Technologies in the Writing Process

The authors haven't used any generative AI/AI assisted technologies in the writing process.

References

1. Ratan S, Nigam S, Midha T, Chaudhary S. A comparative study on treatment seeking behaviour of geriatric population in rural and urban areas of district Kanpur, Uttar Pradesh. Indian J Community Health. 2022;34(4):530–534.
2. Joseph J, Sankar H, Sharma SK, Nambiar D. Sex differences in disease burden, utilization, and expenditure on primary health care services in Kerala, India. Sci Rep. 2025 May 16;15(1):17116.
3. Hossain B, James KS, Nagargoje VP, Barman P. Differentials in private and public healthcare service utilization in later life: do gender and marital status have any association? J Women Aging. 2023 Mar-Apr;35(2):183–193.
4. Kesarwani P, Tripathi AK, Kumar A, Bihari A, Srivastava A. Enablers and barriers of antenatal health utilization among pregnant women in India: a comparative scrutiny from NFHS-4 & 5. Indian J Community Health. 2024;36(6):791–5.
5. Chaudhuri A. Socio economic inequity in health care utilization & expenditures in richer states in India. Indian J Med Res. 2012 Sep;136(3):368–369.
6. Krishnamoorthy Y, Kuberan D, Krishnan M, Sinha I, Kanth K, Samuel G. Impact of health insurance coverage on health care utilization during COVID 19 pandemic: a propensity score matched survey analysis in a target region in India. Int J Health Plann Manage. 2023 May;38(3):723–734.
7. Goswami P, Anand A. Impact of diabetes on healthcare utilization and expenditure among older adults in India. J Diabetes Metab Disord. 2023 Oct 13;23(1):593–601.
8. Yadav R, Zaman K, Mishra A, Reddy MM, Shankar P, Yadav P, et al. Health seeking behaviour and healthcare utilization in a rural cohort of North India. Healthcare (Basel). 2022 Apr 19;10(5):757.
9. Dodd W, King N, Humphries S, Little M, Dewey C. Self-reported morbidity and health service utilization in rural Tamil Nadu, India. Soc Sci Med. 2016 Jul; 161:118–25.
10. Kaushal SK, Maroof M, Singh LD, Gupta SK, Kumar N. A comparative study on prevalence of lifestyle diseases for National Family Health Survey rounds 5 and 4 in Uttar Pradesh. Indian J Community Health. 2024;36(3):351–358.
11. Rathore R, Nigam S, Barman SK. A study on utilization of maternal and infant health services in rural area of district Kanpur Nagar. Int J Community Med Public Health. 2023;10(4):1513–1516
12. Verma N, Bano T, Chopra H, Singh G, Mittal C. Implementation challenges of Pradhan Mantri Jan Arogya

Yojana: A cross-sectional study in Meerut, Uttar Pradesh. Indian J Community Health. 2022;34(3):330–333
13. National Health System Resource Centre. Health dossier 2021: Reflections on key health indicators. New Delhi: NHSRC; 2021. p. 18–20.
14. Ministry of Health and Family Welfare (MoHFW). National family health survey (NFHS-5). Mumbai: International Institute for Population Sciences (IIPS) and ICF; 2021. p. 171–172.
15. Ministry of Health and Family Welfare (MoHFW). State health account estimates for UP: 2019-20. New Delhi: NHSRC; 2020. p. 20–21
16. Kundapur R, Chandramohan S, Shetty V, Modi B. Health insurance in its way to combat financial crisis in breast cancer treatment. Indian J Community Health. 2019;31(3):396–400
17. Singh SK, Pandey P. Morbidity Profile, Healthcare utilization and associated out-of-pocket expenditure on health among elderly population of Lucknow district, northern India. Trop Med Surg. 2016;4(2):210
18. Dey DK, Mishra V. Determinants of choice of healthcare services utilization: empirical evidence from India. Indian J Community Health. 2014;26(4):356–363.
19. Vora K, Yasobant S, Mavalankar D. Predictors of availing maternal health schemes: a community-based study in Gujarat, India. Indian J Community Health. 2014;26(2):174–180.
20. Kanpur Nagar Nigam. Delimitation of zone 2017 [Internet]. Kanpur: Kanpur Nagar Nigam; 2017 [cited 2024 Dec 2]. Available from: http://kmc.up.nic.in/PDF_Files/others/zonewardlist.pdf [Accessed on 18/01/25].