International Journal of Medical and Health Sciences
Volume 11 Issue 04 Dec 2025
ISSN- 2456 - 6063
DOI: **https://doi.org/10.69980/ysjzsd32**

ASSESSING THE IMPACT OF SOCIO-DEMOGRAPHIC AND LIFESTYLE FACTORS ON COMMUNITY HEALTH OUTCOMES: AN EVIDENCE-BASED ANALYSIS FOR PUBLIC HEALTH AND SOCIAL CARE INTERVENTIONS

Dr. Sunil Verma[1], Dr. Meera Krishnan[2], Dr. Ajay Pratap Singh[3], Dr. Nusrat Jahan[4], Dr. Ritesh Kapoor[5]

[1] Department of Community Medicine, Government Medical College, Nagpur, Maharashtra, India
[2] Department of Public Health, Sri Devaraj Urs Academy of Higher Education and Research, Kolar, Karnataka, India
[3] Department of Epidemiology, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, India
[4] Department of Sociology and Health Studies, Aligarh Muslim University, Aligarh, Uttar Pradesh, India
[5] Department of Health Policy and Management, Indian Institute of Public Health, Hyderabad, Telangana, India Email: sunil.verma@gmcnagpur.edu.in

Received:- 25/09/25	Revised:-26/10/25	Accepted:-27/11/25	Published:- 28/12/25

Abstract
Understanding the determinants of community health outcomes is essential in the context of rising chronic disease burden and increasing pressure on healthcare systems. While socio-demographic and lifestyle factors are widely recognized as key influences on health, their direct relationship with healthcare utilization remains insufficiently understood. This study addresses this gap by examining the impact of socio-demographic and lifestyle-related factors on healthcare visits using a structured community health dataset comprising 1,000 observations. The objective is to assess the relative contribution of age and body mass index (BMI) to variations in healthcare utilization. The analysis employs exploratory data analysis, correlation assessment, and multiple linear regression modeling to evaluate associations and predictive effects. The results indicate that age is a statistically significant and dominant predictor of healthcare utilization, demonstrating a strong positive relationship with the number of visits, whereas BMI does not exhibit a significant effect. These findings suggest a disconnect between lifestyle-related health risk and actual engagement with healthcare services. The study highlights the importance of distinguishing between determinants of health risk and healthcare demand, emphasizing the need for age-responsive healthcare planning and strengthened preventive strategies. The findings contribute to improving evidence-based public health interventions and resource allocation in community health systems.

Keywords: Community Health, Healthcare Utilization, Lifestyle Factors, Public Health, Socio-Demographic Factors

Introduction

Community health is a key focus of modern public health because of the rise of chronic diseases, population change and the strain on health care systems (Shidhaye et al., 2016). Socio-demographic factors and lifestyle-related factors are commonly acknowledged as a critical predictor of health outcomes and health service use. Community-level insights into these relationships are crucial for designing interventions, enhancing health service delivery, and enhancing social welfare.

Socio-demographic characteristics, including age, gender and population ageing, are key determinants of health needs and health service utilisation. Age in particular has been shown to be a key predictor of health-care use, with the elderly having a higher burden of chronic health conditions and using more health care (Yadav et al., 2020; Murray et al., 2021). However, lifestyle factors such as diet, physical activity and alcohol and tobacco use also play an important role in long term health risks, and are strongly linked to the global burden of non-communicable diseases (Gelaw et al., 2023). These factors interact with one another to impact on individual and population health.

While there is a wealth of evidence on the social determinants of health, one challenge is to link the knowledge of risk factors to patterns of health care service use. Modifiable lifestyle factors are known to influence disease risk; but their impact on healthcare service utilisation is less understood (Zejnullahu et al., 2021). Health care utilization is not just limited to health status but also includes access, awareness, socio-economic status and health-seeking behaviour (Armoon et al., 2021). This results in a disconnect between health risk and health-care use, especially in communities where preventive engagement is low (Zejnullahu et al., 2021).

Recent studies have shown how socio-demographic and behavioural factors contribute to health outcomes in various populations (Osok et al., 2018). For instance, studies from different geographic settings have demonstrated the role of lifestyle factors and socio-demographic characteristics on health status and disease risk over time (Nienaber-Rousseau et al., 2017; Dey et al., 2022). Likewise, population-level studies stress the need for a multi-dimensional approach to health determinants, considering the interactions of environmental, psychological and lifestyle factors (Chowdhury et al., 2022). These studies highlight the need for holistic analyses that incorporate different aspects of health determinants.

Recent improvements in data quality and analytical techniques allow more detailed analyses of these associations using structured data. Community-based data offer the possibility of examining interactions between demographic and behavioural factors and health outcomes in the real world (Xu & Brodszky, 2024). These kinds of analyses are essential for informing public health decision-making, enabling detection of vulnerable populations and informing intervention strategies. But many studies either examine clinical outcomes alone, or predict health outcomes while overlooking health-care and social care system implications.

Against this backdrop, this study seeks to quantify the effects of socio-demographic and lifestyle factors on community health in terms of health-care visits. The study employs a structured dataset on community health to explore the association between factors such as age and lifestyle indicators on healthcare visits. The research employs a combination of exploratory data analysis and statistical modeling for descriptive and inferential purposes.

This research adds to the body of knowledge by linking theoretical insights into health determinants with practical insights for health care. The study aids decision-making in resource allocation, prevention and community programs, by highlighting the main determinants of healthcare visits. This is in line with public health goals of enhancing access to health services, preventive health, and quality of life.

Methodology

Data Source

The research uses the Community Health Information Dataset from Kaggle (WhimsicalMuffin, 2024) which includes structured individual-level data that is used to record determinants of health in a community context. It includes variables for socio-demographic attributes, lifestyle patterns and health. Social determinants include socio-demographic variables (such as age and sex) and lifestyle indicators (such as physical activity, diet, smoking status and alcohol intake). Health outcomes include body mass index (BMI), and the presence of chronic disease.

The dataset offers a snapshot of community health, allowing the analysis of the relationship between social determinants, lifestyle factors and health outcomes. But it is not a longitudinal or medically validated dataset, which is taken into account when interpreting the results.

Data Preprocessing

An analysis pipeline including data preprocessing was designed to ensure validity and replicability. Data completeness was rigorously evaluated and variables with low levels of missing data were imputed (numerical

variables were imputed with the median value to downweigh the influence of outliers, while categorical variables were imputed with the mode). Cases with significant missing data were removed to prevent bias.
Ordinal encoding (one-hot) was applied to categorical variables to avoid ordinality bias in non-ordered categories. Numerical variables were normalized with z-scores to achieve consistency across features and facilitate convergence of machine learning algorithms.
Data were checked for outliers using distributional analysis and excluded if they represented implausible features, while avoiding compromises to the representativeness of community health problems.

Variables Description
The analytical framework is based on the conceptual model of social determinants influencing health outcomes. Independent variables include socio-demographic factors such as age and gender, along with lifestyle-related variables including physical activity, dietary patterns, smoking status, and alcohol consumption.
The dependent variable is defined as a health outcome indicator derived from the dataset. Depending on the analysis, this variable is operationalized either as a continuous measure, such as BMI, or as a categorical variable representing health risk status. This dual operationalization enables both explanatory and predictive analyses.
Potential confounding effects among independent variables were considered during analysis to ensure robustness in interpretation.

Analytical Approach
The analytical process was conducted in three stages to ensure both interpretability and robustness.
The first stage involved exploratory data analysis, including descriptive statistics and distributional assessment, to understand the structure of the dataset and identify initial patterns.
The second stage involved statistical analysis, where correlation analysis was used to identify associations between variables. Multiple regression techniques were applied to quantify the influence of socio-demographic and lifestyle factors on health outcomes while controlling for confounding variables.
The third stage involved predictive modeling using machine learning techniques. Logistic regression and decision tree models were employed for classification tasks where the dependent variable was categorical, while linear regression models were applied for continuous outcomes. Model selection was based on interpretability and relevance to public health applications rather than purely predictive performance.

Evaluation Metrics
Model performance was evaluated using metrics appropriate to the analytical objective. For classification models, accuracy was used as a general performance indicator, while precision and recall were used to assess class-specific performance, particularly in identifying high-risk individuals. The F1-score was used to provide a balanced measure of precision and recall.
For regression models, the coefficient of determination (R^2) was used to assess explanatory power, while root mean square error (RMSE) was used to quantify prediction error.
In addition to numerical evaluation, emphasis was placed on the interpretability of results to ensure their applicability in public health and social care decision-making contexts.

Results
Descriptive Analysis
The dataset contains 1,000 complete observations with no missing values, indicating high data integrity following preprocessing. The population exhibits substantial variability in age, physical characteristics, and healthcare utilization, reflecting a heterogeneous community sample. The distribution of healthcare visits suggests differing levels of service demand across individuals, while derived indicators such as BMI point toward an overall elevated body mass profile within the population.

Table 1: Descriptive Statistics of Key Variables

Variable	Mean	Std. Dev.	Min	Max
Age (years)	44.18	14.79	5	90
Weight (lbs)	199.49	52.87	38	383
Height (inches)	67.62	6.55	44	92
Visits	11.26	3.95	3	27
BMI	31.38	—	—	—

Figure 1 shows that BMI values are concentrated in the overweight and obese ranges, with a right-skewed distribution indicating the presence of higher BMI levels in a subset of the population.

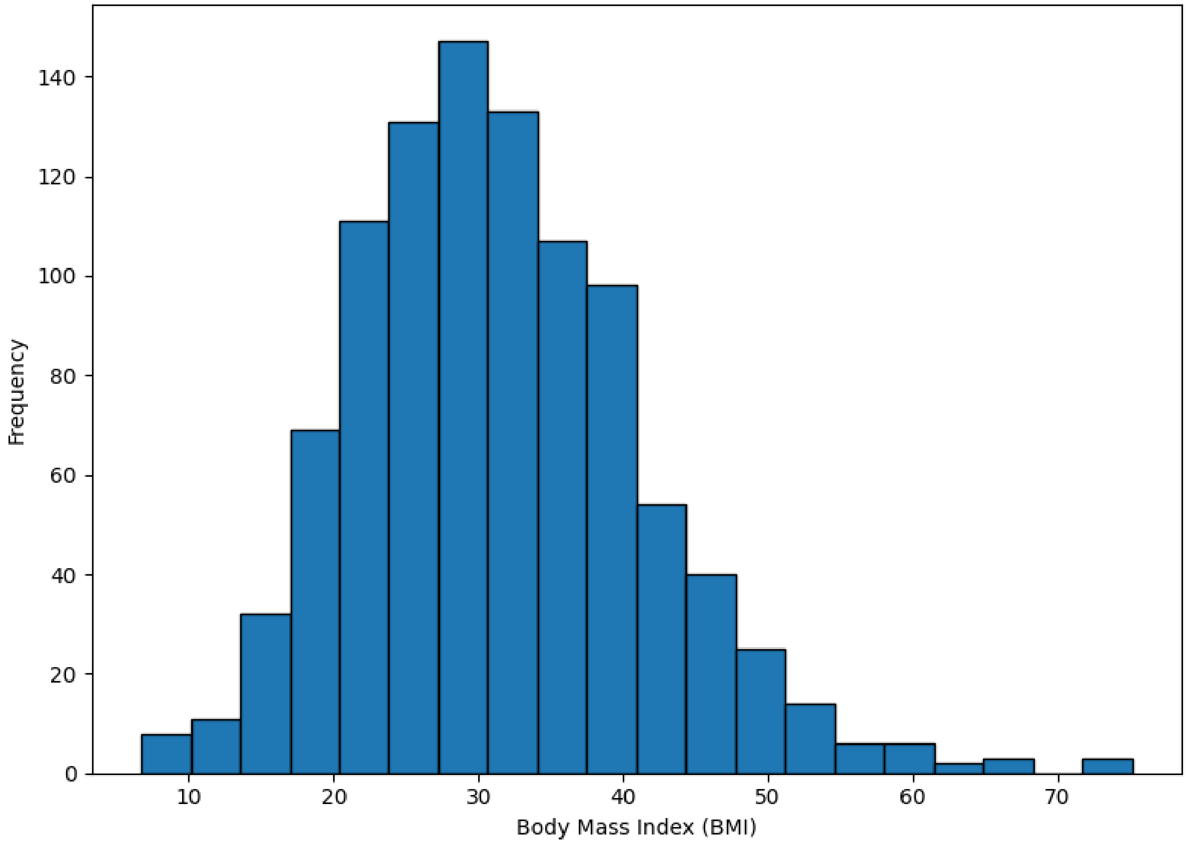

Figure 1. Distribution of Body Mass Index (BMI) among the study population

Exploratory and Correlation Analysis

Exploratory analysis indicates that relationships between variables are not uniform. A clear directional pattern is observed between age and healthcare utilization, whereas lifestyle-related indicators do not display comparable associations. Physical variables exhibit internal consistency, reflecting expected structural relationships rather than independent influence on the outcome variable.

Table 2: Correlation Matrix

Variable	Age	Weight	Height	Visits	BMI
Age	1.00	-0.05	-0.07	0.57	0.01
Weight	-0.05	1.00	0.12	-0.01	0.76
Height	-0.07	0.12	1.00	-0.01	-0.53
Visits	0.57	-0.01	-0.01	1.00	0.00
BMI	0.01	0.76	-0.53	0.00	1.00

Figure 2 shows a clear positive relationship between age and healthcare visits, indicating that healthcare utilization increases with increasing age.

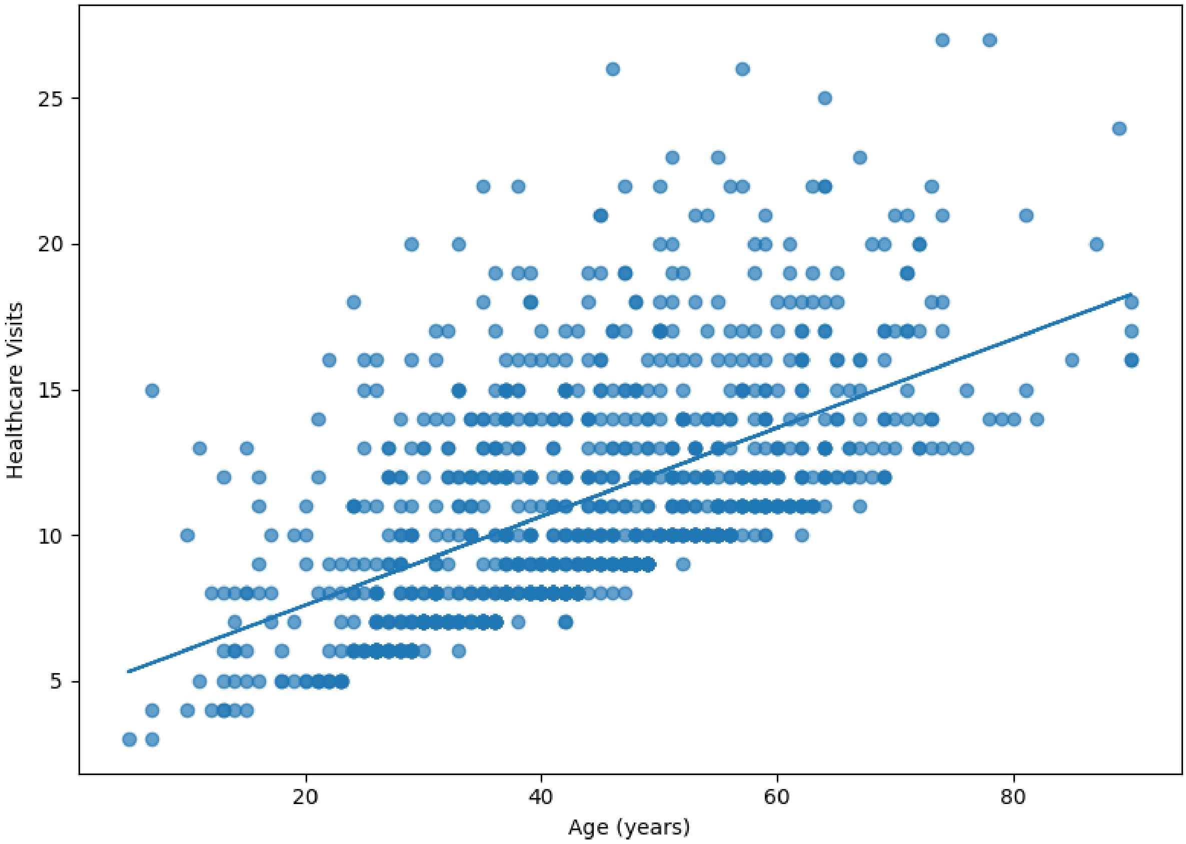

Figure 2. Scatter plot of age and healthcare visits

Regression Analysis

The regression model evaluates the influence of socio-demographic and lifestyle-related variables on healthcare utilization. The model demonstrates statistically significant overall performance, with differential contribution of predictors. One variable shows strong explanatory power, while the other does not contribute meaningfully within the model specification.

Table 3: Multiple Linear Regression Results (Dependent Variable: Visits)

Predictor	Coefficient (β)	Std. Error	t-value	p-value
Constant	4.574	0.452	10.117	<0.001
Age (years)	0.152	0.007	21.994	<0.001
BMI	-0.002	0.010	-0.153	0.878

Model Summary: $R^2 = 0.327$, Adjusted $R^2 = 0.325$, F-statistic = 241.9 ($p < 0.001$)

Group-Level Analysis

Group-level comparisons indicate limited variation in health-related outcomes across gender categories. Differences in average BMI and healthcare visits between males and females are marginal, suggesting that gender does not act as a primary differentiating factor in this dataset. Patterns of healthcare utilization remain broadly consistent across demographic subgroups, reinforcing the dominance of other variables in explaining variation.

Model Evaluation

The regression model demonstrates moderate explanatory capacity, capturing approximately one-third of the variation in healthcare visits. Residual analysis indicates that unexplained variability remains substantial, suggesting the influence of additional factors not included in the model. The stability of coefficients and statistical significance of the primary predictor support the robustness of the model within the defined analytical framework.

Summary of Findings

The results establish that socio-demographic factors, particularly age, are the primary drivers of healthcare utilization in the dataset. Lifestyle-related indicators, as represented by BMI, do not exhibit a statistically significant effect. The findings highlight a concentration of healthcare demand among older individuals, while also indicating that a considerable portion of variation remains unexplained, pointing toward the role of broader social or systemic determinants not captured in the current analysis.

Discussion

The study examined the relative influence of socio-demographic and lifestyle-related factors on healthcare utilization within a community dataset. The findings indicate that age is the only statistically significant predictor of healthcare visits, while BMI does not demonstrate a meaningful effect. This highlights an imbalance in explanatory power between demographic and lifestyle variables, suggesting that healthcare utilization is more strongly driven by structural population characteristics than by observable lifestyle indicators within the dataset.

The positive association between age and healthcare utilization is consistent with existing evidence that older populations require more frequent medical attention due to increased prevalence of chronic conditions and comorbidities (Yadav et al., 2020; Murray et al., 2021). The statistical strength of age in the regression model reinforces its role as a primary determinant of healthcare demand. Similar findings have been reported in large-scale studies, where socio-demographic factors, particularly age, significantly influence both health status and service utilization patterns. The present analysis supports this body of evidence while demonstrating that even in a simplified dataset, age remains a dominant explanatory variable.

In contrast, BMI does not show a statistically significant relationship with healthcare visits, which requires careful interpretation. Prior research consistently identifies lifestyle factors, including obesity, physical inactivity, and poor diet, as major contributors to long-term health risks (Nienaber-Rousseau et al., 2017; Gelaw et al., 2023). However, the current findings suggest that these risks do not directly translate into increased healthcare utilization. This can be explained by the distinction between risk exposure and realized service use. BMI reflects potential health risk, whereas healthcare visits capture actual interaction with healthcare systems. Individuals with elevated BMI may not seek care until symptoms become clinically significant, indicating a disconnect between underlying risk and healthcare engagement.

This interpretation is supported by evidence that healthcare utilization is influenced not only by health status but also by access, awareness, and healthcare-seeking behavior (Sarkar et al., 2015). The moderate explanatory power of the model ($R^2 \approx 0.33$) further suggests that a substantial portion of variability in healthcare visits remains unaccounted for. This points to the likely influence of additional factors not included in the dataset, such as socio-economic status, education, healthcare accessibility, and cultural attitudes toward care, which are known to shape health service utilization patterns.

Gender-based comparisons show minimal variation in both BMI and healthcare visits, indicating that gender does not act as a primary determinant in this dataset. While previous studies have identified gender-related differences in health outcomes and behaviors (Baldisserotto et al., 2016), the absence of strong effects in this analysis may reflect the limited scope of available variables and the generalized nature of the dataset.

From a public health perspective, the findings emphasize the importance of distinguishing between determinants of health risk and determinants of healthcare demand. Age directly influences healthcare utilization, suggesting that aging populations will continue to drive demand for healthcare services. This supports the need for targeted interventions focused on older populations, including preventive care, chronic disease management, and community-based support systems. Evidence from community health programs indicates that such targeted strategies can improve outcomes and reduce long-term healthcare burden (Lu et al., 2015; Arija et al., 2017).

At the same time, the lack of association between BMI and healthcare utilization highlights a potential gap in preventive health engagement. Lifestyle-related risks are well-established, yet their absence in utilization patterns suggests that high-risk individuals may not be adequately reached through existing healthcare systems. This underscores the need for proactive screening, health education, and early intervention strategies to address behavioral risk factors before they lead to clinical conditions (Gate et al., 2016; Kolokotroni et al., 2021).

Limitations and future directions

The study is subject to a few limitations that affect the scope and generalizability of the findings. The use of a cross-sectional dataset restricts the ability to establish causal relationships between socio-demographic and lifestyle factors and healthcare utilization. The dataset lacks critical variables such as income, education, access to healthcare, and clinical history, which are known to significantly influence health outcomes and service use. Additionally, BMI was used as a proxy for lifestyle factors, which does not fully capture the complexity of behavioral determinants such as diet quality, physical activity intensity, and psychosocial influences. These constraints likely contribute to the unexplained variance observed in the model.

Future research should incorporate longitudinal datasets to better understand causal pathways and temporal dynamics between risk factors and healthcare utilization. Expanding the variable set to include socio-economic, environmental, and healthcare access indicators would improve model robustness and explanatory power. Further studies should also explore multidimensional measures of lifestyle and health behavior, along with advanced modeling approaches, to better capture the interaction between social determinants and community health outcomes.

Conclusion
The study provides a data-driven evaluation of how socio-demographic and lifestyle-related factors influence healthcare utilization within a community context. The findings establish age as the dominant predictor of healthcare visits, confirming that demographic structure is a primary driver of healthcare demand. In contrast, BMI does not demonstrate a statistically significant association, indicating a disconnect between lifestyle-related health risk and actual healthcare utilization. This distinction highlights a critical gap in preventive health engagement, where individuals at risk may not be adequately captured within existing healthcare systems. The results emphasize the need to differentiate between determinants of health risk and determinants of healthcare demand when designing public health strategies. From a policy and practice perspective, the findings support prioritization of age-responsive healthcare planning alongside strengthened preventive and community-based interventions targeting lifestyle risks. The study contributes to the evidence base by demonstrating how simplified community-level data can reveal structural patterns in healthcare utilization, while also underscoring the need for more comprehensive models incorporating broader social and systemic determinants.

References

1. Arija, V., Villalobos, F., Pedret, R., Vinuesa, A., Timón, M., Basora, T., … Pas-a-Pas Research Group. (2017). Effectiveness of a physical activity program on cardiovascular disease risk in adult primary health-care users: The "Pas-a-Pas" community intervention trial. *BMC Public Health, 17*(1), 576.
2. Armoon, B., Higgs, P., Fleury, M. J., Bayat, A. H., Moghaddam, L. F., Bayani, A., & Fakhri, Y. (2021). Socio-demographic, clinical and service use determinants associated with HIV-related stigma among people living with HIV/AIDS: A systematic review and meta-analysis. *BMC Health Services Research, 21*(1), 1004.
3. Baldisserotto, J., Kopittke, L., Nedel, F. B., Takeda, S. P., Mendonça, C. S., Sirena, S. A., … Nicolau, B. (2016). Socio-demographic characteristics and prevalence of risk factors in a hypertensive and diabetic population: A cross-sectional study in primary health care in Brazil. *BMC Public Health, 16*(1), 573.
4. Chowdhury, M. R. K., Rahman, M. S., Billah, B., Kabir, R., Perera, N. K., & Kader, M. (2022). The prevalence and socio-demographic risk factors of coexistence of stunting, wasting, and underweight among children under five years in Bangladesh: A cross-sectional study. *BMC Nutrition, 8*(1), 84.
5. Dey, S., Mukherjee, A., Pati, M. K., Kar, A., Ramanaik, S., Pujar, A., … N, S. (2022). Socio-demographic, behavioural and clinical factors influencing control of diabetes and hypertension in urban Mysore, South India: A mixed-method study conducted in 2018. *Archives of Public Health, 80*(1), 234.
6. Gate, L., Warren-Gash, C., Clarke, A., Bartley, A., Fowler, E., Semple, G., … Rodger, A. (2016). Promoting lifestyle behaviour change and well-being in hospital patients: A pilot study of an evidence-based psychological intervention. *Journal of Public Health, 38*(3), e292–e300.
7. Gelaw, Y. A., Koye, D. N., Alene, K. A., Ahmed, K. Y., Assefa, Y., Erku, D. A., … Melaku, Y. A. (2023). Socio-demographic correlates of unhealthy lifestyle in Ethiopia: A secondary analysis of a national survey. *BMC Public Health, 23*(1), 1528.
8. Kolokotroni, O., Mosquera, M. C., Quattrocchi, A., Heraclides, A., Demetriou, C., & Philippou, E. (2021). Lifestyle habits of adults during the COVID-19 pandemic lockdown in Cyprus: Evidence from a cross-sectional study. *BMC Public Health, 21*(1), 786.
9. Lu, C. H., Tang, S. T., Lei, Y. X., Zhang, M. Q., Lin, W. Q., Ding, S. H., & Wang, P. X. (2015). Community-based interventions in hypertensive patients: A comparison of three health education strategies. *BMC Public Health, 15*(1), 33.
10. Murray, F., Allen, M., Clark, C. M., Daly, C. J., & Jacobs, D. M. (2021). Socio-demographic and economic factors associated with 30-day readmission for conditions targeted by the hospital readmissions reduction program: A population-based study. *BMC Public Health, 21*(1), 1922.
11. Nienaber-Rousseau, C., Sotunde, O. F., Ukegbu, P. O., Myburgh, P. H., Wright, H. H., Havemann-Nel, L., … Kruger, H. S. (2017). Socio-demographic and lifestyle factors predict 5-year changes in adiposity among a group of black South African adults. *International Journal of Environmental Research and Public Health, 14*(9), 1089.

12. Osok, J., Kigamwa, P., van der Stoep, A., Huang, K. Y., & Kumar, M. (2018). Depression and its psychosocial risk factors in pregnant Kenyan adolescents: A cross-sectional study in a community health centre of Nairobi. *BMC Psychiatry, 18*(1), 136.
13. Sarkar, C., Webster, C., & Gallacher, J. (2015). UK Biobank Urban Morphometric Platform (UKBUMP): A nationwide resource for evidence-based healthy city planning and public health interventions. *Annals of GIS, 21*(2), 135–148.
14. Shidhaye, R., Gangale, S., & Patel, V. (2016). Prevalence and treatment coverage for depression: A population-based survey in Vidarbha, India. *Social Psychiatry and Psychiatric Epidemiology, 51*(7), 993–1003.
15. WhimsicalMuffin. (2024). *Community health information dataset*. Kaggle. https://www.kaggle.com/datasets/whimsicalmuffin/community-health-information-dataset
16. Xu, F., & Brodszky, V. (2024). The impact of COVID-19 on health-related quality of life: A systematic review and evidence-based recommendations. *Discover Psychology, 4*(1), 90.
17. Yadav, U. N., Thapa, T. B., Mistry, S. K., Pokhrel, R., & Harris, M. F. (2020). Socio-demographic characteristics, lifestyle factors, multi-morbid conditions and depressive symptoms among Nepalese older adults. *BMC Psychiatry, 20*(1), 261.
18. Zejnullahu, V. A., Ukella-Lleshi, D., Miftari, E., & Govori, V. (2021). Prevalence of postpartum depression at the clinic for obstetrics and gynecology in Kosovo teaching hospital: Demographic, obstetric and psychosocial risk factors. *European Journal of Obstetrics & Gynecology and Reproductive Biology, 256*, 215–220.