BMJ Open

# Identify the latent class of modifiable risk behaviours among diabetic and hypertensive patients in Northeast India: a population-base cross-sectional study

| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2021-053757 |
| Article Type: | Original research |
| Date Submitted by the Author: | 25-May-2021 |
| Complete List of Authors: | Marbaniang, Strong P; International Institute for Population Sciences, Public Health and Mortality Studies<br>Lhungdim, Hemkhothang; International Institute for Population Sciences, Public Health and Mortality Studies<br>Chungkham, Holendro; Indian Statistical Institute North-East Centre at Tezpur |
| Keywords: | Hypertension < CARDIOLOGY, General diabetes < DIABETES & ENDOCRINOLOGY, Epidemiology < TROPICAL MEDICINE |
| | |

*I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.*

*The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.*

*Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.*

# Identify the latent class of modifiable risk behaviours among diabetic and hypertensive patients in Northeast India: a population-base cross-sectional study

**Strong P Marbaniang[1*], Hemkhothang Lhungdim[1] , Holendro Singh Chungkham[2]**

[1]Department of Public Health & Mortality Studies, International Institute for Population Sciences, Mumbai-400088, Maharashtra, India.

[2]Indian Statistical Institute, North-East Centre, Tezpur-784501, Assam, India

*Corresponding author

**Strong P Marbaniang**
**Email address-** strongmarbaniang@yahoo.com

**MeSh Keywords:**

Type 2 Diabetes ; Hypertension, Alcohol Drinking, Smoking; Latent Class Analysis.

**Word Count**: 3281 words

## Abstract

**Objective**: To identify the latent class of modifiable risk factors among the diabetic and hypertensive patients based on the observed indicators variables: smoking, alcohol, aerated drinks, overweight or obesity, diabetes, and hypertension. It is hypothesized that the study population diagnosed with diabetes or hypertension is homogeneous with respect to the modifiable risk factors.

**Design:** A cross-sectional study was conducted using the national representative large scale survey data.

**Setting and participants:** The data comes from the Indian National Family Health Survey (NFHS-4), 2015-16. Participants aged 15-49 years who were diagnosed with either diabetes or hypertension, or both were included. The total sample is 22,249, out of which 3,284 were males, and 18,965 were females.

**Primary and secondary outcome measures:** The observed variables used as latent indicators are: smoking, alcohol, aerated drinks, overweight or obesity, diabetes, and hypertension. The concomitant variables include age, gender, education, marital status, and household wealth index. Latent class model was used to simultaneously identify the latent class and to determine the association between the concomitant variables and the latent class.

**Results:** Three latent classes were identified and labelled as Class 1: "Diabetic with low-risk lifestyle" (21%), Class 2: "High-risk lifestyle" (8%), and Class 3: "Hypertensive with low-risk lifestyle" (71%). Class 1 had a high probability of having diabetes and low probability of smoking and drinking alcohol. Class 2 respondents were characterized by high probability of smoking and drinking alcohol and Class 3 is characterized by high probability of having high blood pressure and low probability of smoking and drinking alcohol.

**Conclusions:** Male should manage smoking and alcohol consumption, while females should control body weight and blood pressure. Marital status and gender could be the preventive factors for high-risk smoking and alcohol drinking in Northeast India


**Keywords**: Modifiable risk; Diabetes; Hypertension; Latent class analysis; Northeast India.

**Strengths and limitations of the study**

- The sample for the study is relatively large as it used nationally representative large-scale data.
- Respondents' status of diabetes and hypertension is diagnosed using the biomarker measurement.
- Latent class analysis is a person-centred technique and the appropriate method for explaining the clustering of health risk behaviours.
- The cross-sectional design of the study restricts us to measure the causal effect.
- Modifiable risk factors were collected based on dichotomous response which limits us to measure the intensity of substance use.

## Introduction

Effective lifestyle modification plans and affordable approaches to control high blood sugar and high blood pressure level would benefit the at-risk population [1,2]. The leading causes of mortality due to chronic disease in adults, particularly older adults, are linked to unhealthy lifestyle and behaviours [3], like tobacco consumption, physical inactivity, excess alcohol consumption, and poor diet. Usually, most risk behaviours exist simultaneously or cluster within individuals [4]. It is argued that lifestyle risk factors within individuals are not random but more likely to cluster with other unhealthy behaviours [5,6]. In Asia, for example, almost 43.5% of Chinese adults and 37% of Korean adults had at least two cardiovascular risk factors [4,7]. Moreover, in India, 35.64% of adult men and 10% of adult women had at least three lifestyle risk factors [8].

Latent Class Analysis (LCA), based on structural equation modelling, is a person-centred analytical approach that allows the identification of homogeneous sub-groups in a heterogeneous population. LCA is similar to cluster analysis because individuals are classifying into homogenous unobserved (latent) groups based on the response pattern to a set of observed variables. However, in cluster analysis, objects are classified based on distance measures, whereas in LCA classification are based on probabilistic and finite mixture modelling approach [9,10]. A latent class model splits the population into mutually exclusive and exhaustive groups that are homogeneous within the groups but differ among them. Several studies have adopted the LCA approach to identify risk factors clustering within the heterogeneous population. In a study by Dey et al, based on seven observed risk factors, LCA was used to classify individuals into two latent class “susceptible to adverse health outcomes groups” and “not susceptible to adverse health outcomes groups” [11]. Another study from China applied LCA to ten observed complications and comorbidities of T2 diabetic patients and categorized the individuals into four latent classes, namely "complications and comorbidity groups", "high risk of complications group”, “high risk of comorbidities and Cardio Vascular Disease groups”, and “diabetes without complications and comorbidities group” [12]. A study in West Azerbaijan province among the hypertensive patients aged 50 years and above used four indicators such as dietary patterns, physical activity, tobacco use, and high blood pressure control to categorized the hypertensive patients into three latent classes [13].

The majority of studies reported a higher prevalence of smoking, excess alcohol consumption, physical inactivity, and poor diet among people with diabetes or hypertension

than those without diabetes or hypertension [14,15]. Notably, previous studies have examined diabetic or hypertensive patients as a single homogeneous group without considering the possible heterogeneity existing within the population [16]. Identifying the clustered pattern of modified risk factors for diabetes and hypertension and the socio-demographic factors associated with clusters of unhealthy lifestyle behaviours allows health intervention program to target these risk factors simultaneously and effectively.

Most of the existing studies on health risk behaviours have emphasized only changing a single unhealthy behaviour [15,17]. However, interventions that simultaneously targets these unhealthy behaviours are necessary from the policy point of view [4]. In addition, existing studies in India on clustering of Non Communicable disease (NCD) risk factors have used intraclass correlation and scoring methods [8,18,19]. However, studies have suggested that LCA is the most appropriate method for explaining the clustering of health risk behaviours [20,21]. LCA is a probabilistic model approach designed for identifying clusters based on dichotomous variables. This study is the first in Northeast India to explore the clustering of NCD risk factors using the LCA approach. This study aims to identify the latent class of modifiable risk factors among diabetic and hypertensive patients based on the observed indicators variables: smoking, alcohol, aerated drinks, overweight or obesity, diabetes, and hypertension. We applied LCA to investigate heterogeneity in the population of an individual with diabetes and hypertension and to identify possible latent class based on the response of observed risk factors. More specifically, the objective is to describe how individuals in Northeast India diagnosed with diabetes or hypertension are clustered according to the pattern of observed risk factors and identify the difference between these groups based on socio-demographic characteristics. We hypothesize that the study population diagnosed with diabetes or hypertension is homogeneous with respect to the modifiable risk factors.

## Methods

### *Study location*

The study focuses on the Northeastern region of India comprising eight states, namely Assam, Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, and Tripura (Figure 1). The region has a hilly terrain inhabited mainly by tribal people belonging to different cultures and ethnic communities [22]. It has over 45 million, which is 3.76% of India population and a population density of 159 persons per square km [23].

Figure 1: ***Map showing the location of the Study area***

***Design***

A cross-sectional study was conducted using the national representative large scale survey data. The analysis is based on the publicly available secondary data where the identification of respondent is concealed [24].

***Setting***

The data comes from the fourth round of the Indian National Family Health Survey (NFHS-4) conducted during the year 2015-16. The data was downloaded from the DHS website [25]. NFHS was initiated by the Ministry of Health and Family Welfare, Government of India, and coordinated by the International Institute for Population Sciences, Mumbai. The survey adopted a stratified two-stage sampling design. In the first stage, primary sampling units (PSUs) were selected based on probability proportional to population size. Rural PSUs were villages, while census enumeration blocks formed the PSUs in urban areas. In the second stage, systematic random sampling was done in each PSUs to select households for the sample. At the household level, information was sought from the women aged 15 – 49 years and men aged 15 – 54 years. A detailed description of the survey design can be found elsewhere [24].

***Sample***

The study consists of individuals aged 15-49 years who were diagnosed with either diabetes or hypertension, or both. The total sample size is 22,249, out of which 3,284 were males, and 18,965 were females.

***Patient and Public Invlovement***

No patient involved

***Measures***

**Diabetes:** Fingerstick blood is collected, and blood glucose level is determined using the FreeStyle Optium H Glucometer. Respondent is considered diagnosed with diabetes if the random blood sugar level is >140mg/dl [24].

**Hypertension:** Blood pressure level was measured using an OMRON Blood Pressure measuring device. Three separate blood pressure readings were taken with an interval of 5 minutes between readings. A respondent is considered hypertensive if the systolic blood

pressure ≥ 140 mmHg, or diastolic blood pressure ≥ 90 mmHg, or if the respondent was taking antihypertensive medication to lower blood pressure at the time of the survey [24].

**Overweight or obesity:** In this study, we categorized continuous body mass index (BMI) according to the WHO guidelines for the Asian population. A respondent was classified as overweight or obese if BMI≥23 kg/m$^2$ [26].

***Outcome Variables of Interest***

**Observed Indicators variables:** The observed variables used as latent indicators are cigarette's smoking (No=1, Yes=2), Alcohol consumption (No=1, Yes=2), Takes aerated drink (No=1, Yes=2), Overweight or obesity (No=1, Yes=2), Diabetes (No=1, Yes=2), Hypertension (No=1, Yes=2).

***Predictors or Concomitant variables***

**Concomitant variables for class determination:** The concomitant variables include age, gender, education, marital status, and household wealth index.

***Statistical Analysis***

Latent Class Analysis is an innovative statistical method used to identify the latent classes of homogeneous individuals in a heterogeneous population [12]. The latent class was performed using the six indicators variables mentioned above. LCA uses observed dichotomous indicators to identify the unobserved latent class in a heterogeneous population. To determine the optimal number of latent class, we examined one to four model. The optimal number of latent classes was chosen based on the model having the smallest Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (cAIC), Bayesian-Schwarz Information Criterion (BIC), adjusted Bayesian-Schwarz Information Criterion (aBIC), and Entropy [27,28]. A combination of parsimony and interpretability model selection criteria was used to carefully select the number of latent classes so that individuals are allocated to their most likely class. The output of the latent class analysis includes the number of latent classes, the latent class probability (i.e., the probability that an individual selected at random belonged to each latent class), and the conditional probability (i.e., the probability that an individual would give a particular response to a specific item of an observed indicator variable given that an individual belonged to a specific latent class).

To determine the association between the concomitant variables (i.e., age, gender, education, marital status, and household wealth index) and the latent class, we apply the one-step technique approach. In this approach, the concomitant variables are included in the latent class regression model, and their coefficient is estimated simultaneously as part of the latent class model [29]. This approach has been proved to provide the best and unbiased coefficient estimates of the concomitant variables as compared to other methods [30]. Statistical analysis was conducted in R statistical software, and LCA was analyze using the poLCA package (Polytomous Variable Latent Class Analysis) [29].

## Results

### *Descriptive Statistics*

A total of 22,249 participants with chronic disease, i.e., diagnosed with diabetes or hypertension, were included in the analysis (Table 1). The participants in the age group 45-49 and 15-19 constitute the highest (21.31%) and the lowest (5.04%) percentage shares out of the total sample. Over half (51.17%) of the sample have completed secondary education, 24.6 percent were illiterate, and 79.56 percent were currently married. Further, among the chronic patients, the prevalence of smoking, consumption of alcohol, and aerated drinks were 4.17, 14.8, and 68.57 percent, respectively. Also, 81.13 percent of the participants reported suffering from hypertension, and 27.7 percent suffered from diabetes.

**Table 1: Background characteristics of diabetes and hypertensive patients aged 15-49 years, NFHS-4, 2015-16.**

| Background Characteristics | % (n=22,249) |
|---|---|
| **Age** | |
| 15-19 | 5.04 |
| 20-24 | 8.38 |
| 25-29 | 12.17 |
| 30-34 | 14.96 |
| 35-39 | 19.19 |
| 40-44 | 18.95 |
| 45-49 | 21.31 |
| **Sex** | |
| Male | 13.88 |
| Female | 86.12 |
| **Education** | |
| Illiterate | 24.6 |
| Primary | 16.3 |

| | |
|---|---|
| Secondary | 51.17 |
| Higher Secondary | 7.93 |
| **Marital Status** | |
| Never married | 13.57 |
| Married | 79.56 |
| Widowed/Divorced/Separated | 6.87 |
| **Wealth Index** | |
| Poorest | 16.72 |
| Poorer | 34.55 |
| Middle | 22.44 |
| Richer | 17.5 |
| Richest | 8.79 |
| **Observed Indicators** | |
| Smoking | 4.17 |
| Drink alcohol | 14.8 |
| Take aerated drinks | 68.57 |
| Overweight or Obesity | 43.26 |
| Diabetes | 27.7 |
| Hypertension | 81.13 |

***Model fit and selection of Latent class***

The model selection statistics for models are shown in Table 2. We fit the LCA models with classes ranging from 1 to 4. For each LCA model, the model fit indicators BIC, AIC, cAIC, and Entropy were calculated. According to the model fit indicators, a 3-Class model was selected as the best fit model as it has the lowest BIC, cAIC, and Entropy values. These three latent classes were labelled as "Diabetic with low-risk lifestyle", "High-risk lifestyle", and "Hypertensive with low-risk lifestyle". Figure 2 portrays the graphical representation of item response probability for the indicator's variables across the latent classes.

**Table 2: Model fit statistics for the latent class models (n=22,249)**

| No of classes | log-likelihood | Residual df | BIC | aBIC | cAIC | likelihood-ratio | Entropy |
|---|---|---|---|---|---|---|---|
| **1** | -69173.11 | 57 | 138406.3 | 138387.2 | 138412.3 | 17350.762 | - |
| **2** | -61315.38 | 34 | 122921 | 122828.9 | 122950 | 1859.196 | 1 |
| **3** | -58920.21 | 11 | 118361 | 118195.7 | 118413 | 1168.644 | 0.91 |
| **4** | -59970.1 | -12 | 120691 | 120452.6 | 120766 | 2396.112 | 0.94 |

Latent class probability and the conditional probability of a "Yes" response for each indicator variables were summarized in Table 3. The last row in Table 3 indicates the

probability of class membership in each latent class. About 21 percent of the participants were expected to belong to the Class 1 "diabetic with low-risk lifestyle", 8 percent to Class 2 "high-risk lifestyle", and 71 percent to Class 3 "hypertensive with low-risk lifestyle". Members in Class 1 were likely to report having high blood sugar level (diabetic) and likely to report not having smoked or drinking alcohol. Members in Class 2 are more likely to smoke and drink alcohol, and members in Class 3 are likely to report having high blood pressure and likely to report not having smoked or drink alcohol.

Figure 2: **Item response conditional probability across the latent classes**

**Table 3: The conditional probability of the response item and the latent class probability**

| Indicator Variables | Diabetic with low-risk lifestyle (Class 1) | High-risk lifestyle (Class 2) | Hypertensive with low-risk lifestyle (Class 3) |
|---|---|---|---|
| Smoking | 0.04 | 0.55 | 0.02 |
| Drink alcohol | 0.12 | 0.81 | 0.14 |
| Take aerated drinks | 0.76 | 0.78 | 0.77 |
| Overweight or Obesity | 0.44 | 0.52 | 0.51 |
| Diabetic | 1.00 | 0.25 | 0.11 |
| Hypertensive | 0.00 | 0.88 | 1.00 |
| **Latent class probability** | **0.21** | **0.08** | **0.71** |

***Covariates predicting latent class membership***

Table 4 summarized the odds ratio (OR) from the latent regression model. Compared with respondents in the 15-19 years age group, those in the 20-29 age group were 4.1 times (OR=4.10) more likely to belong to class 2 (high-risk lifestyle) than to class 1 (diabetic with low-risk lifestyle). Similarly, compared with respondents in the 15-19 years age group, those in the 45-49 age group were 1.7 times (OR=1.73) more likely to belong to class 3 (hypertensive with low-risk lifestyle) than to class 1 (diabetic with low-risk lifestyle). Also, we observed that the odds of being in class 3 increases with the age of the respondents. With respect to males, females were less likely (OR=0.002) to belong to class 2 but more likely (OR=1.46) to belong to class 3. Higher odds of being in class 2-3 than to class 1 was observed among the married respondents (OR=1.54); however, widowed/divorced/separated were more likely to be in class 2 (OR=3.35), but less likely to be in class 3 (OR=0.98). Moreover, compared to the illiterate respondents, those with higher secondary education were less likely to belong to class 3 than

to class 1 (OR=0.79), and being in the richest quantile of wealth index lower the odds of being in class 3 than in class 1 (OR=0.69).

**Table 4: The latent classes of lifestyle behaviours and their associated covariates among individuals aged 15-49 years**

| Background Characteristics | Class 2 vs Class 1 | | Class 3 vs Class 1 | |
|---|---|---|---|---|
| | Odds Ratio | S. E | Odds Ratio | S. E |
| **Age** | | | | |
| 15-19 | Ref | | Ref | |
| 20-24 | 4.10 *** | 0.25 | 1.27 ** | 0.09 |
| 25-29 | 5.10 *** | 0.25 | 1.34 *** | 0.09 |
| 30-34 | 5.70 *** | 0.26 | 1.42 *** | 0.09 |
| 35-39 | 3.53 *** | 0.26 | 1.42 *** | 0.09 |
| 40-44 | 3.82 *** | 0.27 | 1.51 *** | 0.09 |
| 45-49 | 3.46 *** | 0.27 | 1.73 *** | 0.09 |
| **Gender** | | | | |
| Male | Ref | | Ref | |
| Female | 0.002 *** | 0.34 | 1.46 *** | 0.07 |
| **Education** | | | | |
| Illiterate | Ref | | Ref | |
| Primary | 1.03 | 0.19 | 0.73 *** | 0.06 |
| Secondary | 1.08 | 0.17 | 0.79 *** | 0.05 |
| Higher Secondary | 0.70 | 0.21 | 0.79 *** | 0.08 |
| **Marital Status** | | | | |
| Never married | Ref | | Ref | |
| Married | 1.54 ** | 0.15 | 1.16 ** | 0.06 |
| Widowed/Divorced/Separated | 3.35 *** | 0.35 | 0.98 | 0.08 |
| **Wealth Index** | | | | |
| Poorest | Ref | | Ref | |
| Poorer | 0.90 | 0.18 | 1.00 | 0.06 |
| Middle | 1.02 | 0.18 | 0.87 ** | 0.07 |
| Richer | 1.14 | 0.19 | 0.86 * | 0.07 |
| Richest | 1.13 | 0.22 | 0.69 *** | 0.08 |

*Ref: Reference group; ***p-value<0.01; **p-value<0.05; *p-value<0.1; S.E: Standard Error; Reference group: Class 1*

## Discussion

This study is the first in Northeast India that used latent class analysis to provides evidence about the pattern of modifiable risk factors among patients with diabetes or hypertension. Based on the response modifiable health risk behaviours, the LCA suggest three classes of respondents with diabetes or hypertension. These classes were characterized as “diabetic with low-risk lifestyle”, “high-risk lifestyle”, and “hypertensive with low-risk lifestyle”.

More than 70 percent of the respondents belonged to the class “hypertensive with low-risk lifestyle”, which is they have their blood pressure level higher than normal and less likely

to smoke. Several studies have acknowledged a significant elevation in blood pressure and an increase in the prevalence of hypertension following smoking cessation [31,32], which is possibly due to post-cessation weight gain [32]. Moreover, about 8 per cent of the respondents belonged to the class "high-risk lifestyle", mostly drink alcohol and smoke cigarette. This co-occurrence of alcohol use and cigarette smoking have been well documented in the existing studies [33,34]. In all the three identified classes, a vast majority of diabetic or hypertensive patients have a habit of consuming aerated drinks and have uncontrolled body weight. Studies suggest that sugar-sweetened beverages intake is the major contributor to weight gain and can increase the risk of T2 diabetes and hypertension [35,36]. Also, evidence from the study in Northeast India supports the positive correlation between overweight or obesity with diabetes and hypertension [16].

In line with the previous study [37], this study found that members aged 20 years and above are more likely to be in the high-risk cluster than the members below 20 years. Studies have explained that the teenage period is just the initial stage where an individual starts using or gets exposed to alcohol or smoking products. However, as this behaviour becomes a habit later, it may subsequently increase the intensity of substance use among older adults [38,39]. Also, it is well documented that starting to smoke and drink alcohol at an early age is associated with the number of cigarettes smoked and the quantity of alcohol consumed per day in adult age. According to Dawson, (2008) and Investigator, (1991), individuals who began to smoke or drink before the age of 20 years were most likely to consumed a large volume of alcohol and smoked more cigarettes per day as compared to those who started to smoke at the age of 20 or above [40,41]. Further, compared to young adults, older adults were more likely to report high blood pressure and less likely to smoke or drink alcohol. One reason could be the existing condition of high blood pressure; an older adult was most likely to have taken a precaution by lowering or quitting smoking and alcohol consumption to prevent further deterioration of health. Evidence from a recent study among older adults demonstrated that the main reason for smoking cessation attempt was the motivation towards better health [42]. It has been mentioned that hypertensive smokers were encouraged to quit smoking because of the risk of developing a severe form of hypertension such as malignant and renovascular hypertension [43]. Gender was found to be significantly associated with class 2 "high-risk lifestyle" and 3 "hypertensive with low-risk lifestyle". Consistent with the previous studies, in this study, female's respondents were less likely to be in the high-risk cluster than males' respondents [44]. One primary reason is that smoking and drinking alcohol by women is not socially and culturally

accepted by many [45], and women are more concerned about their health and tend to avoid unhealthy lifestyle behaviour [46]. In our study, individuals from a wealthier household with a higher level of education were less likely to belong to a cluster of hypertensives with low substance use. An individual with higher education and better income can afford to pay for a healthier lifestyle, including regular physical exercise, accessibility to advanced and quality healthcare services; such efforts may reduce the risk of hypertension [47,48]. Being widowed/divorced/separated increased the likelihood of membership in the high-risk cluster. Studies about substance use by marital status have indicated that widowed or divorced individuals are most likely to consumed alcohol, cigarette, and marijuana as compared to singles [49,50].

To our knowledge, no study in India or Northeast India has study the clustering of health risk factor using the latent class approach. However, our study has some limitation. Firstly, the cross-sectional design of the study restricts us to measure the causal effect. Secondly, due to the dichotomous response on alcohol consumption, we cannot measure the quantity of alcohol an individual consumed.

**Conclusions**

This study identified three classes based on modifiable risk factors closely related to the risk of chronic disease. In addition, the study also identified factors that uniquely distinguished the identified classes. The intake of aerated drinks and obesity are the common modifiable risk factors in all three classes. As an initiative to healthy lifestyle behaviours, the findings suggest that males should control smoking and alcohol consumption, while females should control body weight and blood pressure. Further, we found that marital status and gender could be the preventive factors for high-risk smoking and alcohol drinking in Northeast India.

**Contributors:** Conceived and designed the research paper: SPM, HL. Analysed the data: SPM, HSC. Contributed agents/materials/analysis tools: SPM, HL, HSC. Wrote the manuscript: SPM, HSC. Refined the manuscript: SPM, HL, HSC.

**Funding:** This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

**Competing Interest:** Non declared

**Patient consent for publication:** Not applicable

**Ethics approval**: This study is based on the secondary data which is publicly available from the Demographic Health Survey website and the respondents are de-identified, hence ethical approval is not required for this study.

**Data availability statement**: Data were collected and owned by the Demographic Health Survey authority. Data are available at https://dhsprogram.com/data/dataset/India_Standard-DHS_2015.cfm?flag=1 upon request.

**Word count**: 3093

**References**


1 Chobanian A V, Bakris GL, Black HR, *et al.* The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA* 2003;**289**:2560–72. doi:10.1001/jama.289.19.2560

2 International Diabetes Federation. Advocacy guide to the IDF Diabetes Atlas Ninth edition 2019. Brussels, Belgium:International Diabetes Federation: 2019. http://www.diabetesatlas.org/ (accessed 14 Nov 2019).

3 Ford ES, Zhao G, Tsai J, *et al.* Low-Risk Lifestyle Behaviors and All-Cause Mortality: Findings From the National Health and Nutrition Examination Survey III Mortality Study. *Am J Public Health* 2011;**101**:1922–9. doi:10.2105/AJPH.2011.300167

4 Ha S, Choi HR, Lee YH. Clustering of four major lifestyle risk factors among Korean adults with metabolic syndrome. *PLoS One* 2017;**12**:e0174567. doi:10.1371/journal.pone.0174567

5 Poortinga W. The prevalence and clustering of four major lifestyle risk factors in an English adult population. *Prev Med (Baltim)* 2007;**44**:124–8. doi:10.1016/j.ypmed.2006.10.006

6 Schuit AJ, van Loon AJM, Tijhuis M, *et al.* Clustering of Lifestyle Risk Factors in a General Adult Population. *Prev Med (Baltim)* 2002;**35**:219–24. doi:10.1006/pmed.2002.1064

7 Gao B, Zhang L, Wang H. Clustering of Major Cardiovascular Risk Factors and the Association with Unhealthy Lifestyles in the Chinese Adult Population. *PLoS One* 2013;**8**:e66780. doi:10.1371/journal.pone.0066780

8 Shaikh R, Khan J. Clustering of lifestyle risk factors among adult population in India:

A cross-sectional analysis from 2005 to 2016. *PLoS One* 2021;**16**:e0244559. doi:10.1371/journal.pone.0244559

9 Islam MM, Valderas JM, Yen L, *et al.* Multimorbidity and Comorbidity of Chronic Diseases among the Senior Australians: Prevalence and Patterns. *PLoS One* 2014;**9**:e83783. doi:10.1371/journal.pone.0083783

10 Brusco MJ, Shireman E, Steinley D. A comparison of latent class, K-means, and K-median methods for clustering dichotomous data. *Psychol Methods* 2017;**22**:563–80. doi:10.1037/met0000095

11 Dey A, Chakraborty A, Majumdar K, *et al.* Application of latent class analysis to estimate susceptibility to adverse health outcomes based on several risk factors. *Int J Community Med Public Heal* 2016;**3**:3423–9. doi:10.18203/2394-6040.ijcmph20164268

12 Gao F, Chen J, Liu X, *et al.* Latent class analysis suggests four classes of persons with type 2 diabetes mellitus based on complications and comorbidities in Tianjin, China: a cross-sectional analysis. *Endocr J* 2017;**64**:1007–16. doi:10.1507/endocrj.EJ17-0199

13 Ghanbari J, Mohammadpoorasl A, Jahangiry L, *et al.* Subgroups of lifestyle patterns among hypertension patients: a latent-class analysis. *BMC Med Res Methodol* 2018;**18**:127. doi:10.1186/s12874-018-0607-6

14 Sobngwi E, Mbanya J-CN, Unwin N, *et al.* Physical activity and its relationship with obesity, hypertension and diabetes in urban and rural Cameroon. *Int J Obes* 2002;**26**:1009–16. doi:10.1038/sj.ijo.0802008

15 Bani T Aeri DS. Risk Factors Associated with the Increasing Cardiovascular Diseases Prevalence in India: A Review. *J Nutr Food Sci* 2014;**05**:1–6. doi:10.4172/2155-9600.1000331

16 Marbaniang SP, Lhungdim H, Yadav B, *et al.* Overweight/obesity risks and prevalence of diabetes and hypertension in North Eastern India: An analysis using seemingly unrelated probit model. *Clin Epidemiol Glob Heal* 2021;**11**:100764. doi:10.1016/j.cegh.2021.100764

17 Umamahesh K, Vigneswari A, Surya Thejaswi G, *et al.* Incidence of cardiovascular diseases and associated risk factors among subjects with type 2 diabetes – An 11-year

follow up study. *Indian Heart J* 2014;**66**:5–10. doi:10.1016/j.ihj.2013.12.009

18 Bischops AC, De Neve J-W, Awasthi A, *et al.* A cross-sectional study of cardiovascular disease risk clustering at different socio-geographic levels in India. *Nat Commun* 2020;**11**:5891. doi:10.1038/s41467-020-19647-3

19 Sarveswaran G, Kulothungan V, Mathur P. Clustering of noncommunicable disease risk factors among adults (18–69 years) in rural population, South-India. *Diabetes Metab Syndr Clin Res Rev* 2020;**14**:1005–14. doi:10.1016/j.dsx.2020.05.042

20 Schreiber JB. Latent Class Analysis: An example for reporting results. *Res Soc Adm Pharm* 2017;**13**:1196–201. doi:10.1016/j.sapharm.2016.11.011

21 Weller BE, Bowen NK, Faubert SJ. Latent Class Analysis: A Guide to Best Practice. *J Black Psychol* 2020;**46**:287–311. doi:10.1177/0095798420930932

22 Marbaniang SP. Women Care and Practices in the Management of Childhood Diarrhea in Northeast India. *Child Care Pract* 2020;:1–13. doi:10.1080/13575279.2020.1812534

23 Dikshit KR, Dikshit JK. Population of the North-Eastern States of India. In: *North-East India: Land, People and Economy. Advances in Asian Human-Environmental Research*. Dordrecht: : Springer 2014. 421–56. doi:10.1007/978-94-007-7055-3_12

24 International Institute for Population Sciences (IIPS) and ICF. National Family Health Survey (NFHS-4), 2015-16: India. Mumbai:IIPS: 2017. http://rchiips.org/nfhs/NFHS-4Reports/India.pdf

25 DHS. The DHS Program-India: Standard DHS, 2015-16 Dataset,. Demogr. Heal. Surv. 2017.https://dhsprogram.com/data/dataset/India_Standard-DHS_2015.cfm?flag=1

26 Barba C, Cavalli-Sforza T, Cutter J, *et al.* Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* 2004;**363**:157–63. doi:10.1016/S0140-6736(03)15268-3

27 Sugiura N. Further analysts of the data by akaike' s information criterion and the finite corrections. *Commun Stat - Theory Methods* 1978;**7**:13–26. doi:10.1080/03610927808827599

28 Akaike H. A New Look at the Statistical Model Identification. 1974. 215–22.

doi:10.1007/978-1-4612-1694-0_16

29 Linzer DA, Lewis JB. poLCA : An R Package for Polytomous Variable Latent Class Analysis. *J Stat Softw* 2011;**42**. doi:10.18637/jss.v042.i10

30 Bolck A, Croon M, Hagenaars J. Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Polit Anal* 2004;**12**:3–27. doi:10.1093/pan/mph001

31 Jenner DA, Puddey IB, Beilin LJ, *et al.* Lifestyle- and occupation-related changes in blood pressure over a six-year period in a cohort of working men. *J Hypertens Suppl Off J Int Soc Hypertens* 1988;**6**:S605-7. doi:10.1097/00004872-198812040-00189

32 Gerace TA, Hollis J, Ockene JK, *et al.* Smoking cessation and change in diastolic blood pressure, body weight, and plasma lipids. MRFIT Research Group. *Prev Med (Baltim)* 1991;**20**:602–20. doi:10.1016/0091-7435(91)90058-c

33 Ho RKS, Fok PWY, Chan HCH. Pattern and determinants of alcohol and tobacco co-use and its relationship with smoking cessation in Hong Kong. *Tob Prev \& Cessat* 2021;**7**:1–9. doi:10.18332/tpc/132288

34 Jiang N, Lee YO, Ling PM. Association between tobacco and alcohol use among young adult bar patrons: a cross-sectional study in three cities. *BMC Public Health* 2014;**14**:500. doi:10.1186/1471-2458-14-500

35 Malik VS, Popkin BM, Bray GA, *et al.* Sugar-sweetened beverages, obesity, type 2 diabetes mellitus, and cardiovascular disease risk. *Circulation* 2010;**121**:1356–64. doi:10.1161/CIRCULATIONAHA.109.876185

36 Qin P, Li Q, Zhao Y, *et al.* Sugar and artificially sweetened beverages and risk of obesity, type 2 diabetes mellitus, hypertension, and all-cause mortality: a dose–response meta-analysis of prospective cohort studies. *Eur J Epidemiol* 2020;**35**:655–71. doi:10.1007/s10654-020-00655-y

37 Saikia N, Debbarma B. The socioeconomic correlates of substance use among male adults in Northeast India. *Clin Epidemiol Glob Heal* 2020;**8**:149–57. doi:10.1016/j.cegh.2019.06.004

38 IOM (Institute of Medicine). Patterns of Tobacco Use by Adolescents and Young Adults. In: Richard J. Bonnie, Kathleen Stratton LYK, ed. *Public Health Implications*

*of Raising the Minimum Age of Legal Access to Tobacco Products.* Washington, D.C.: : The National Academies Press 2015. https://www.ncbi.nlm.nih.gov/books/NBK310412/pdf/Bookshelf_NBK310412.pdf

39 Maggs JL, Schulenberg JE. Initiation and Course of Alcohol Consumption among Adolescents and Young Adults. In: *Recent Developments in Alcoholism*. New York: : Kluwer Academic Publishers-Plenum Publishers 2005. 29–47. doi:10.1007/0-306-48626-1_2

40 Investigators S. Effect of the Age at Which Smoking Begins on Frequency of Smoking in Adulthood. *N Engl J Med* 1991;**325**:968–9. doi:10.1056/NEJM199109263251318

41 Dawson DA, Goldstein RB, Patricia Chou S, *et al.* Age at First Drink and the First Incidence of Adult-Onset DSM-IV Alcohol Use Disorders. *Alcohol Clin Exp Res* 2008;**32**:2149–60. doi:10.1111/j.1530-0277.2008.00806.x

42 Tanihara S, Momose Y. Reasons for smoking cessation attempts among Japanese male smokers vary by nicotine dependence level: a cross-sectional study after the 2010 tobacco tax increase. *BMJ Open* 2015;**5**:e006658–e006658. doi:10.1136/bmjopen-2014-006658

43 Freitas SRS, Alvim RO. Smoking and blood pressure phenotypes: New perspective for an old problem. *Am J Hypertens* 2017;**30**:554–5. doi:10.1093/ajh/hpx039

44 Saikia B, Marbaniang SP, Kumar P, *et al.* Changing pattern of tobacco consumption and quitting behavior in Northeast India. *J Subst Use* 2021;:1–12. doi:10.1080/14659891.2021.1875068

45 Wilsnack RW, Wilsnack SC, Kristjanson AF, *et al.* Gender and alcohol consumption: patterns from the multinational GENACIS project. *Addiction* 2009;**104**:1487–500. doi:10.1111/j.1360-0443.2009.02696.x

46 Nolen-Hoeksema S. Gender differences in risk factors and consequences for alcohol use and problems. *Clin Psychol Rev* 2004;**24**:981–1010. doi:10.1016/j.cpr.2004.08.003

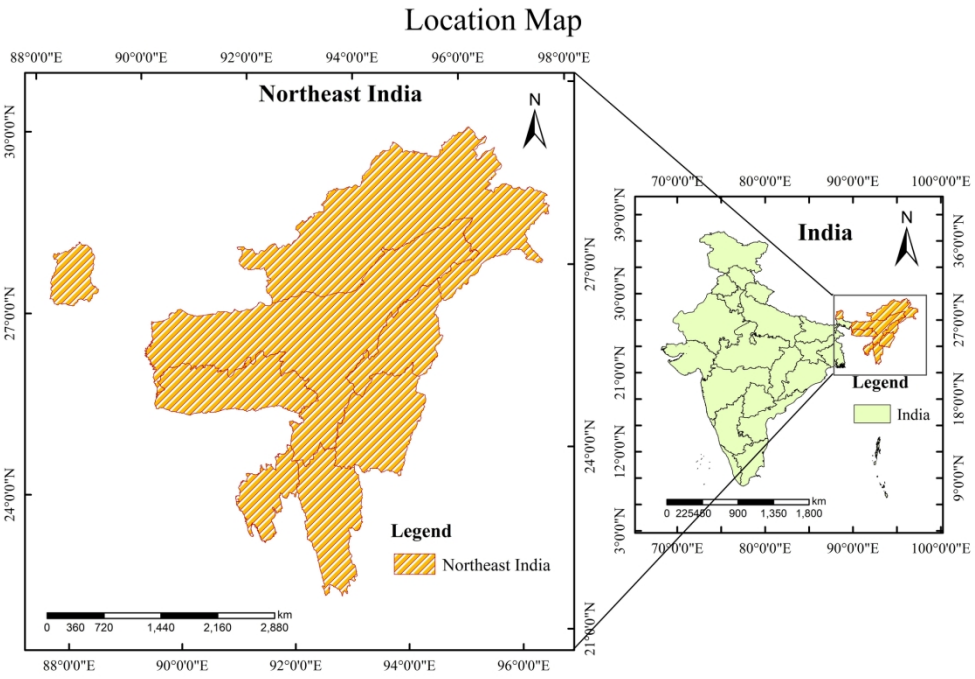
47 Rasiah R, Yusoff K, Mohammadreza A, *et al.* Cardiovascular disease risk factors and socioeconomic variables in a nation undergoing epidemiologic transition. *BMC Public Health* 2013;**13**:886. doi:10.1186/1471-2458-13-886

48 Winkleby MA, Jatulis DE, Frank E, *et al.* Socioeconomic status and health: how

education, income, and occupation contribute to risk factors for cardiovascular disease. *Am J Public Health* 1992;**82**:816–20. doi:10.2105/AJPH.82.6.816

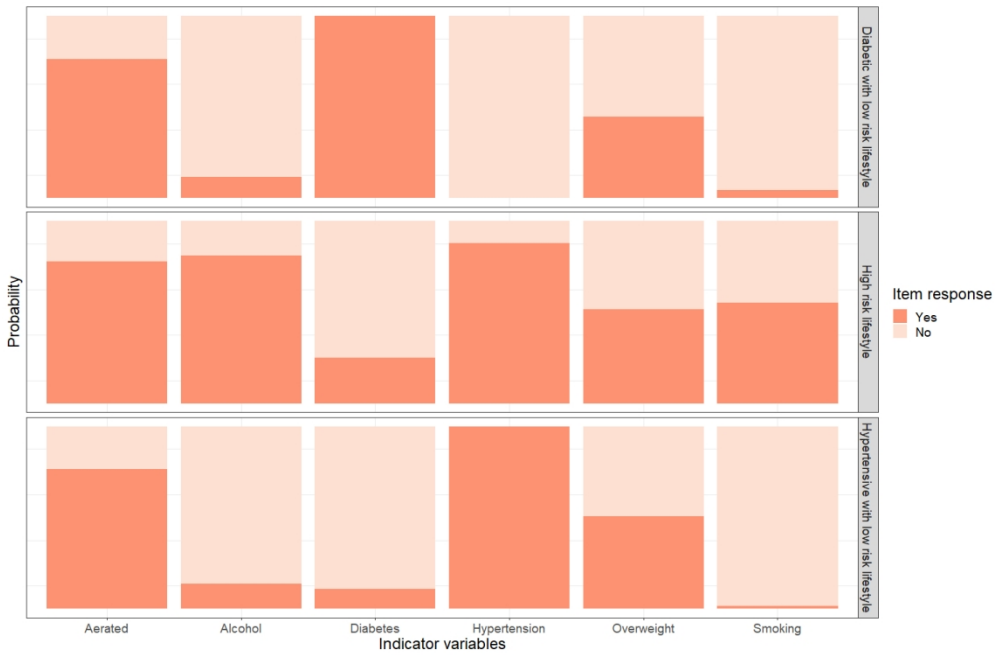
49 Menasco M, Blair S. Gender Differences in Substance Use Across Marital Statuses. *Int J Criminol Sociol* 2016;**5**:1–13. doi:10.6000/1929-4409.2016.05.01

50 Power C, Rodgers B, Hope S. Heavy alcohol consumption and marital status: disentangling the relationship in a national study of young adults. *Addiction* 1999;**94**:1477–87. doi:10.1046/j.1360-0443.1999.941014774.x




: Map showing the location of the Study area

190x131mm (300 x 300 DPI)




Item response conditional probability across the latent classes

129x84mm (300 x 300 DPI)

**STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies***

| Section/Topic | Item # | Recommendation | Reported on page # |
|---|---|---|---|
| **Title and abstract** | 1 | (*a*) Indicate the study's design with a commonly used term in the title or the abstract | 1 |
| | | (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found | 1 |
| **Introduction** | | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported | 4 |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses | 4 |
| **Methods** | | | |
| Study design | 4 | Present key elements of study design early in the paper | 5 |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | 5 |
| Participants | 6 | (*a*) Give the eligibility criteria, and the sources and methods of selection of participants | 5 |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | 6 |
| Data sources/ measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | 6 |
| Bias | 9 | Describe any efforts to address potential sources of bias | |
| Study size | 10 | Explain how the study size was arrived at | 5 |
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | |
| Statistical methods | 12 | (*a*) Describe all statistical methods, including those used to control for confounding | 6 |
| | | (*b*) Describe any methods used to examine subgroups and interactions | |
| | | (*c*) Explain how missing data were addressed | |
| | | (*d*) If applicable, describe analytical methods taking account of sampling strategy | |
| | | (*e*) Describe any sensitivity analyses | |
| **Results** | | | |

| | | | |
|---|---|---|---|
| Participants | 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed | |
| | | (b) Give reasons for non-participation at each stage | |
| | | (c) Consider use of a flow diagram | |
| Descriptive data | 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders | 7 |
| | | (b) Indicate number of participants with missing data for each variable of interest | |
| Outcome data | 15* | Report numbers of outcome events or summary measures | 7 |
| Main results | 16 | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included | 7 |
| | | (*b*) Report category boundaries when continuous variables were categorized | |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | |
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses | |
| **Discussion** | | | |
| Key results | 18 | Summarise key results with reference to study objectives | 8 |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias | 10 |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence | 9-10 |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results | 10 |
| **Other information** | | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | 11 |

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

# Identifying the latent classes of modifiable risk behaviours among diabetic and hypertensive individuals in North Eastern India: a population-based cross-sectional study

| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2021-053757.R1 |
| Article Type: | Original research |
| Date Submitted by the Author: | 01-Dec-2021 |
| Complete List of Authors: | Marbaniang, Strong P; International Institute for Population Sciences, Public Health and Mortality Studies<br>Lhungdim, Hemkhothang; International Institute for Population Sciences, Public Health and Mortality Studies<br>Chungkham, Holendro; Indian Statistical Institute North-East Centre at Tezpur |
| Primary Subject Heading: | Public health |
| Secondary Subject Heading: | Cardiovascular medicine, Epidemiology, Global health, Health policy, Smoking and tobacco |
| Keywords: | Hypertension < CARDIOLOGY, General diabetes < DIABETES & ENDOCRINOLOGY, Epidemiology < TROPICAL MEDICINE |
| | |

*I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.*

*The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.*

*Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.*

# Identifying the latent classes of modifiable risk behaviours among diabetic and hypertensive individuals in North Eastern India: a population-based cross-sectional study

**Strong P Marbaniang[1,3*], Hemkhothang Lhungdim[1] , Holendro Singh Chungkham[2,4]**

[1]Department of Public Health & Mortality Studies, International Institute for Population Sciences, Mumbai-400088, Maharashtra, India.

[2]Indian Statistical Institute, North-East Centre, Tezpur-784501, Assam, India

[3]Department of Statistics, Sankardev College, Shillong-793004, Meghalaya, India

[4]Stress Research Institute, Stockholm University, Stockholm, Sweden

*Corresponding author

**Strong P Marbaniang**
**Email address-** strongmarbaniang@yahoo.com

**MeSh Keywords:**

Type 2 Diabetes ; Hypertension, Alcohol drinking, Smoking; Latent Class Analysis; North Eastern India.

**Word Count**: 4464

## Abstract

**Objective**: To identify the latent classes of modifiable risk factors among the diabetic and hypertensive patients based on the observed indicators variables: smoking, alcohol, aerated drinks, overweight or obesity, diabetes, and hypertension. We hypothesized that the study population diagnosed with diabetes or hypertension is homogeneous with respect to the modifiable risk factors.

**Design:** A cross-sectional study using a stratified random sampling method, and a nationally representative large-scale survey.

**Setting and participants:** The data come from the fourth round of the Indian National Family Health Survey, 2015-16. Respondents age 15-49 years who were diagnosed with either diabetes or hypertension, or both were included. The total sample is 22,249, out of which 3,284 were males, and 18,965 females.

**Primary and secondary outcome measures:** The observed variables used as latent indicators are: smoking, alcohol, aerated drinks, overweight or obesity, diabetes, and hypertension. The concomitant variables include age, gender, education, marital status, and household wealth index. Latent class model was used to simultaneously identify the latent class and to determine the association between the concomitant variables and the latent classes.

**Results:** Three latent classes were identified and labelled as Class 1: "Diabetic with low-risk lifestyle" (21%), Class 2: "High-risk lifestyle" (8%), and Class 3: "Hypertensive with low-risk lifestyle" (71%). Class 1 is characterised by those with a high probability of having diabetes and low probability of smoking and drinking alcohol. Class 2 characterized by a high probability of smoking and drinking alcohol, and Class 3 by a high probability of having high blood pressure and low probability of smoking and drinking alcohol.

**Conclusions:** Males should control smoking and alcohol consumption, while females should control their body weight and blood pressure. Policy and intervention programmes in Norteastern India should focus on targeting multiple modifiable risk behaviours that are most likely to co-occur within an individual.


**Keywords**: Modifiable risk; Diabetes; Hypertension; Latent class analysis; Northeastern India.

**Strengths and limitations of the study**

- The sample for the study is relatively large as it uses a nationally representative large-scale data.
- Respondents' status of diabetes and hypertension is as diagnosed during fieldwork (on the day of interview) using a standard biomarker instrument and measurement.
- Latent class analysis (LCA) is a person-centred technique and the appropriate method for explaining the clustering of health risk behaviours.
- The cross-sectional design of the study restricts us to measure the causal effect.
- Modifiable risk factors were collected based on dichotomous responses, which limits measuring the intensity of substance use.

## Introduction

Effective lifestyle modification plans and affordable approaches to control high blood sugar and high blood pressure level benefit the at-risk population [1,2]. The leading causes of mortality due to chronic diseases in adults, particularly older adults, are linked to unhealthy lifestyle and behaviours [3], such as tobacco consumption, physical inactivity, excess alcohol consumption, and poor diet. An alarmingly high prevalence of hypertension in North Eastern (NE) India has been reported in recent studies [4-6]. Sikkim has the highest prevalence of hypertension among all the Indian states, while other NE states such as Nagaland, Assam, Arunachal Pradesh, Mizoram, and Tripura have prevalence above 11.0 percent, which is well above the national average [5,6]. Further, according to the Global Burden of Disease (GBD) study, the percentage change of diabetes prevalence between 1990 and 2016 has increased by more than 20 percent across the states of Northeastern India [7]. This percentage change was highest in the states of Sikkim, Arunachal Pradesh, and Nagaland.

Modifiable health risk behaviours such as smoking, heavy consumption of alcohol, low physical activity, and unhealthy diet characterized by high intake of sugar and fats, low consumption of fruits and vegetables are the major causes of non-communicable diseases (NCDs) [8], and strongly linked with early mortality [9,10]. Usually, these risk behaviours exist simultaneously or cluster within individuals [11]. It is argued that lifestyle risk factors within individuals are not random, but more likely to cluster with other unhealthy behaviours [12,13]. Further, when two risk behaviours cluster together with each other more than the other factors, it may suggests that they could be influenced by a common source [14]. Moreover, modifiable risk behaviours for chronic diseases such as excessive use of alcohol, cigarette smoking, obesity and unhealthy diets are unlikely to occur in an entirely uniform manner in a population. That is, it is uncertain whether people can be accurately classify into two groups, as those adapting to a healthy versus those unhealthy lifestyle. For example, certain individuals who engage in more vigorous physical activities tend to consume alcohol more frequently than their inactive counterparts [15]. Also, cigarette smoking is associated with lower rate of obesity among certain individuals [16]. In a study among the US adults, Leventhal et al. (2014) used modifiable risk factors like usage of alcohol, drug, nicotine, current obesity status, and weekly physical activity and found that the adult population clusters into five sub-groups based on the pattern of these modifiable risk factors [14]. Another study in India by Shaikh and Khan (2021),

found that hypertension is more likely to cluster with other modifiable risk behaviours such as smoking, alcohol, chewing tobacco, obesity, and unhealthy diets in both men and women [17]. Atorkey and Owiredua (2021), in their study that identified the clustering of five multiple health risk behaviours, also found that low vegetable and fruit intake, and low physical activity cluster together within an individual, whereas smoking tobacco and alcohol consumption co-occur together forming another cluster [18]. The pattern of clustering of modifiable risk behaviours could also differ by region and community level, as evident from previous studies [17,19]. The clustering of two or more modifiable risk behaviours is of great concern because it can intensify the risk of developing chronic diseases and cardiovascular mortality [20,21], and most intervention measures are specific to a single risk factors. For example, in Asia 43.5 percent of Chinese adults and 37 percent of Korean adults had at least two cardiovascular risk factors [11,22]. Moreover, in India, 35.6 percent of adult men and 10 percent adult women have at least three lifestyle risk factors [17]. It is also found that a combination of smoking and heavy drinking emerged as the riskiest behaviour for all causes of death [23].

The majority of studies reported a higher prevalence of smoking, excess alcohol consumption, physical inactivity, and poor diet among people with diabetes or hypertension than those without diabetes or hypertension [24,25]. Notably, previous studies have examined diabetic or hypertensive patients as a single homogeneous group without considering the possible heterogeneity existing within the population [26]. Most of the existing studies on health risk behaviours have emphasized only changing a single unhealthy behaviour [25,27]. However, interventions that simultaneously target these unhealthy behaviours are necessary from the policy perspective [11].

Existing studies in India on clustering of NCDs risk factors have used intraclass correlation and scoring methods [17,19,28]. However, studies have suggested that LCA is the most appropriate method for explaining the clustering of health risk behaviours [29,30]. LCA is a probabilistic model approach designed for identifying clusters based on dichotomous variables. This study is also an attempt to explore the clustering of NCD risk factors using the LCA approach in Northeastern India. The aims is to identify the latent classes of modifiable risk factors among diabetic and hypertensive patients based on the observed indicator variables such as smoking, alcohol, aerated drinks, overweight or obesity, diabetes, and hypertension. We applied LCA, also to investigate heterogeneity in the population of an individual with diabetes and hypertension based on the responses of observed risk factors. More specifically, the objective is to describe how individuals in Northeast India diagnosed with diabetes or

hypertension are clustered according to the pattern of observed risk factors and identify the difference between these groups based on socio-demographic characteristics. Information on clustering of health risk behaviours become vital as it guides in designing the disease control programs. Intervention approach that aimed at targeting multiple modifiable risk behaviours in a single program are gaining importance because they are potentially most cost-effective, efficient and may have a great public health impact than the single modifiable risk approaches. Therefore, we hypothesize that the study population diagnosed with diabetes or hypertension is homogeneous with respect to the modifiable risk factors.

## Methods

### *Study location*

The study focuses on the Northeastern region (NER) of India comprising eight states, namely Assam, Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, and Tripura (Figure 1). Most parts of the region has a hilly terrain and inhabited mainly by myriad tribes belonging to different cultures and ethnic communities [31]. The region has over 45 million population, which is 3.8 percent of India's population, with a density of 159 persons per $km^2$ [32].

.....Figure 1: ***Map showing the location of the Study area. Source:****Authors*.....**(here)**

### *Design and Setting*

The NFHS-4 (2015-16), a nationally representative large scale survey, is a cross-sectional study that adopted a stratified two-stage random sampling design for both urban and rural areas, including a systematic mapping and listing of households in the selected primary sampling units (PSUs). The analysis is based on the publicly available data, in which the identity of selected respondents, households, and geocodes of PSUs are concealed [33]. In the first stage, the PSUs were selected based on probability proportional to population size (PPS). Rural PSUs are the villages, while census enumeration blocks (CEBs) formed the PSUs in urban areas. In the second stage, systematic random sampling was done in each PSUs to select households for the sample. At the household level, information were sought from the heads, eligible women age 15 – 49 years, and men age 15 – 54 years, alongwith their biomarkers and blood samples on a specially designed Dried Blood Spots (DBS). A detailed description of the survey design can be found elsewhere [33].

The data used in this study is the fourth round of the Indian National Family Health Survey (NFHS-4) conducted during 2015-16. The data was downloaded from the DHS website [34]. The NFHS series of surveys has been initiated by the Ministry of Health and Family Welfare, Government of India, and coordinated by the International Institute for Population Sciences, Mumbai.

***Sample***

The study includes only individuals age 15-49 years who were diagnosed with either diabetes or hypertension, or both. The total sample size is 22,249, of which 3,284 were males, and 18,965 were females for the Northeastern region of India.

***Patient and Public Invlovement***

No patient involved

***Measures***

**Diabetes:** Fingerstick blood is collected, and blood glucose level is determined using the FreeStyle Optium H Glucometer. A respondent is considered diagnosed with diabetes if the random blood sugar level is >140mg/dl [33].

**Hypertension:** Blood pressure (BP) level was measured using an OMRON Blood Pressure measuring device. Three separate blood pressure readings were taken with an interval of 5 minutes between readings. A respondent is considered hypertensive if the systolic blood pressure is ≥ 140 mmHg, or diastolic blood pressure ≥ 90 mmHg, or if the respondent was taking antihypertensive medication to lower blood pressure at the time of the survey [33].

**Overweight or obesity:** In this study, we categorized continuous body mass index (BMI) according to the WHO guidelines for the Asian population. A respondent is classified as overweight or obese if BMI is ≥23 kg/m$^2$ [35].

***Outcome Variables of Interest***

**Observed Indicator variables:** The observed variables used as latent indicators are cigarette smoking (No=1, Yes=2), Alcohol consumption (No=1, Yes=2), Takes aerated drink (No=1, Yes=2), Overweight or obesity (No=1, Yes=2), Diabetes (No=1, Yes=2), and Hypertension (No=1, Yes=2).

***Predictors or Concomitant variables***

**Concomitant variables for class determination:** The concomitant variables include age, gender, education, marital status, and household wealth index.

***Statistical Analysis***

Latent Class Analysis (LCA), based on structural equation modelling, is a person-centred analytical approach that allows the identification of homogeneous sub-groups in a heterogeneous population. LCA is similar to cluster analysis because individuals are classifying into homogenous unobserved (latent) groups based on the response pattern to a set of observed variables. However, in cluster analysis, objects are classified based on distance measures, whereas in LCA classification are based on probabilistic and finite mixture modelling approach [36,37]. A latent class model splits the population into mutually exclusive and exhaustive groups that are homogeneous within the groups but differ among them. Several studies have adopted the LCA approach to identify risk factors clustering within the heterogeneous population. In a study by Dey et al (2016), based on seven observed risk factors, LCA was used to classify individuals into two latent class "susceptible to adverse health outcomes groups" and "not susceptible to adverse health outcomes groups" [38]. Another study in China applied LCA to ten observed complications and comorbidities of T2 diabetic patients and categorized the individuals into four latent classes, namely "complications and comorbidity groups", "high risk of complications group", "high risk of comorbidities and Cardio Vascular Disease groups", and "diabetes without complications and comorbidities group" [39]. A study in West Azerbaijan province among the hypertensive patients aged 50 years and above used four indicators such as dietary patterns, physical activity, tobacco use, and high blood pressure control to categorized the hypertensive patients into three latent classes [40].

Latent Class Analysis is an innovative statistical method used to identify the latent classes of homogeneous individuals in a heterogeneous population [39]. The latent class was performed using the six indicators variables mentioned above. LCA uses observed dichotomous indicators to identify the unobserved latent class in a heterogeneous population. To determine the optimal number of latent classes, we examined one to four models. The optimal number of latent classes was chosen based on the model having the smallest Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (cAIC), Bayesian-Schwarz Information Criterion (BIC), adjusted Bayesian-Schwarz Information Criterion

(aBIC), and Entropy [41,42]. A combination of parsimony and interpretability model selection criteria was used to carefully select the number of latent classes so that individuals are allocated to their most likely class. The outputs of the latent class analysis includes the number of latent classes, the latent class probability (i.e., the probability that an individual selected at random belonged to each latent class), and the conditional probability (i.e., the probability that an individual would give a particular response to a specific item of an observed indicator variable given that an individual belongs to a specific latent class).

To determine the association between the concomitant variables (i.e., age, gender, education, marital status, and household wealth index) and the latent class, we apply the one-step technique approach. In this approach, the concomitant variables are included in the latent class regression model, and their coefficient is estimated simultaneously as part of the latent class model [43]. This approach has been demonstrated to provide the best and unbiased coefficient estimates of the concomitant variables as compared to other methods [44]. Statistical analysis was conducted in R statistical software, and LCA was analyze using the poLCA package (Polytomous Variable Latent Class Analysis) [43].

Figure 2: ***Conceptual framework of the study. Source:*** *Authors*

## Results

### *Descriptive statistics*

A total of 22,249 individuals with chronic diseases, i.e., diagnosed with diabetes or hypertension, were included in the analysis (Table 1). Individuals in the age group 45-49 and 15-19 constitute the largest (21.3%) and the lowest (5.0%) respectively in the total sample. Over half (51.2%) of the sample have completed secondary education, 24.6 percent were illiterate, and 79.6 percent were currently married. Further, among the chronic disease patients, the prevalence of smoking, consumption of alcohol, and aerated drinks were 4.2, 14.8, and 68.6 percent respectively. Also, 81.1 percent of the participants reported suffering from hypertension, and 27.7 percent from diabetes. The sample distribution also indicates that chronic condition is much higher among females (86.1%) than males (13.9%), among older ages (age 35+), and among poorer households.

**Table 1: Percent distribution of adults of age 15-49 years with chronic condition (diabetic and hypertensive patients) by selected background characteristics, Northeastern India, NFHS-4, 2015-16.**

| Background Characteristics | % (n=22,249) |
|---|---|
| **Age** | |
| 15-19 | 5.04 |
| 20-24 | 8.38 |
| 25-29 | 12.17 |
| 30-34 | 14.96 |
| 35-39 | 19.19 |
| 40-44 | 18.95 |
| 45-49 | 21.31 |
| **Sex** | |
| Male | 13.88 |
| Female | 86.12 |
| **Education** | |
| Illiterate | 24.6 |
| Primary | 16.3 |
| Secondary | 51.17 |
| Higher Secondary | 7.93 |
| **Marital Status** | |
| Never married | 13.57 |
| Married | 79.56 |
| Widowed/Divorced/Separated | 6.87 |
| **Wealth Index** | |
| Poorest | 16.72 |
| Poorer | 34.55 |
| Middle | 22.44 |
| Richer | 17.5 |
| Richest | 8.79 |
| **Observed Indicators** | |
| Smoking | 4.17 |
| Drink alcohol | 14.8 |
| Take aerated drinks | 68.57 |
| Overweight or Obesity | 43.26 |
| Diabetes | 27.7 |
| Hypertension | 81.13 |



***Model fit and selection of Latent class***

The model fit/selection statistics to derive appropriate models are shown in Table 2. We fit the LCA models with classes ranging from 1 to 4. For each LCA model, the model fit indicators BIC, AIC, cAIC, and Entropy were calculated. According to the model fit indicators, a three-Class model has been selected as the best fit model as it has the lowest BIC and cAIC values. Smaller values on each indicator suggest a better model, or a model with few explanatory variables or parameters, and inform the decision of the best model to be retained [45]. Another diagnostic indicator is entropy [46], which indicates how accurately the model defines the

classes. In general, an entropy value close to 1.0 is considered as ideal [47], and above 0.8 is also acceptable. An entropy value of 1.0 would indicate that every individual has been perfectly classified within the classes based on the responses of the observed indicators. The three latent classes were labelled as "Diabetic with low-risk lifestyle", "High-risk lifestyle", and "Hypertensive with low-risk lifestyle". Figure 3 portrays the graphical representation of item response probability for the indicator's variables across the latent classes.

**Table 2: Model fit statistics for the latent class models (n=22,249)**

| No of classes | log-likelihood | Residual df | BIC | aBIC | cAIC | likelihood-ratio | Entropy |
|---|---|---|---|---|---|---|---|
| **1** | -69173.11 | 57 | 138406.3 | 138387.2 | 138412.3 | 17350.762 | - |
| **2** | -61315.38 | 34 | 122921 | 122828.9 | 122950 | 1859.196 | 1.0 |
| **3** | -58920.21 | 11 | 118361 | 118195.7 | 118413 | 1168.644 | 0.91 |
| **4** | -59970.1 | -12 | 120691 | 120452.6 | 120766 | 2396.112 | 0.94 |

Latent class probability and the conditional probability of a "Yes" response for each indicator variables were summarized in Table 3. The last row in Table 3 indicates the probability of class membership in each latent class. About 21 percent of the participants were expected to belong to the Class 1 "diabetic with low-risk lifestyle", 8 percent to Class 2 "high-risk lifestyle", and 71 percent to Class 3 "hypertensive with low-risk lifestyle". Members in Class 1 were likely to report having high blood sugar level (diabetic) and unlikely to smoke cigarette or drinking alcohol. Members in Class 2 are more likely to smoke cigarette and drink alcohol, and members in Class 3 are likely to report having high blood pressure and unlikely to smoke cigarette or drink alcohol.

Figure 3: **Item response conditional probability across the latent classes.** ***Source:*** *Authors*

**Table 3: The conditional probability of the response items and the latent class probability for study sample in Northeastern India, NFHS-4, 2015-16**

| **Indicator Variables** | **Diabetic with low-risk lifestyle (Class 1)** | **High-risk lifestyle (Class 2)** | **Hypertensive with low-risk lifestyle (Class 3)** |
|---|---|---|---|
| Smoking | 0.04 | 0.55 | 0.02 |
| Drink alcohol | 0.12 | 0.81 | 0.14 |
| Take aerated drinks | 0.76 | 0.78 | 0.77 |
| Overweight or Obesity | 0.44 | 0.52 | 0.51 |
| Diabetic | 1.00 | 0.25 | 0.11 |
| Hypertensive | 0.00 | 0.88 | 1.00 |
| **Latent class probability** | **0.21** | **0.08** | **0.71** |

Table 4 presents the distribution of socio-economic and demographic characteristics of the individuals by the three latent classes. Age, gender, education, marital status, and household wealth index were statistically found significantly associated with the three latent classes at 1 percent level of significant. The hypertensive with low-risk lifestyle class has the highest proportion of respondents aged 45-49 years, whereas both diabetic with low-risk lifestyle class and high-risk lifestyle class has the highest proportion of respondents aged 35-39 years (19.4% and 19.5% respectively). Both the diabetic with low-risk lifestyle class (10.6% males vs. 89.4% females) and the hypertensive with low-risk lifestyle class (6.2% males vs. 93.8% females) has the highest proportion of females; males constituted almost 100 percent of the high-risk lifestyle class (99.2% males vs. 0.8% females). Lower proportion of the respondents have completed higher secondary education in case of both diabetic with low-risk lifestyle class (10.5%) and hypertensive with low-risk lifestyle class (8.5%). Marital status shows a similar pattern in all the three classes; higher proportion of married respondents and lower proportion of widowed/divorced/separated respondents (<10%). The Wealth index shows that respondents belonging to the middle quintile constituted the largest proportion in the diabetic with low-risk lifestyle class (27.2%), and high-risk lifestyle class (26.6%). However, respondents belonging to the poorer wealth quintile have the highest proportion of hypertensive with low-risk lifestyle class (29.4%).

**Table 4: Percent distribution of diabetic and hypertensive patients age 15-49 years according to three latent classes by background characteristics, Northeastern India, NFHS-4, 2015-16**

| Background Characteristics | "Diabetic with low-risk lifestyle" class (n=4727, 21%) | "High-risk lifestyle" class (n=1828, 8%) | "Hypertensive with low-risk lifestyle" class (n=15,694, 71%) | P-value[a] |
|---|---|---|---|---|
| **Age** | | | | <0.01 |
| 15-19 | 8.08 | 3.06 | 4.96 | |
| 20-24 | 8.99 | 8.37 | 7.6 | |
| 25-29 | 12.88 | 13.51 | 11.93 | |
| 30-34 | 15.30 | 18.00 | 15.29 | |
| 35-39 | 19.40 | 19.53 | 19.46 | |
| 40-44 | 17.73 | 19.09 | 19.04 | |
| 45-49 | 17.62 | 18.44 | 21.72 | |
| **Gender** | | | | <0.01 |
| Male | 10.62 | 99.18 | 6.17 | |
| Female | 89.38 | 0.82 | 93.83 | |
| **Education** | | | | <0.01 |
| Illiterate | 17.96 | 11.65 | 25.17 | |
| Primary | 16.50 | 15.10 | 15.85 | |
| Secondary | 55.05 | 58.70 | 50.50 | |
| Higher Secondary | 10.49 | 14.55 | 8.47 | |
| **Marital Status** | | | | <0.01 |
| Never married | 19.93 | 21.55 | 14.30 | |

| | | | | |
|---|---|---|---|---|
| Married | 72.75 | 75.22 | 78.41 | |
| Widowed/Divorced/Separated | 7.32 | 3.23 | 7.30 | |
| **Wealth Index** | | | | <0.01 |
| Poorest | 10.94 | 12.14 | 12.94 | |
| Poorer | 26.06 | 24.84 | 29.43 | |
| Middle | 27.23 | 26.64 | 26.30 | |
| Richer | 22.49 | 23.34 | 21.25 | |
| Richest | 13.29 | 12.04 | 10.07 | |
| **Total** | 100.0 | 100.0 | 100.0 | |

*Note: [a]Based on chi-squared tests of association.*

### *Covariates predicting latent class membership*

Table 5 summarized the odds ratios (OR) results from the latent regression model. Compared with respondents in the 15-19 years age group, those in the 20-29 age group were 4.1 times (OR=4.10; p-<0.01) more likely to belong to class 2 (high-risk lifestyle) than to class 1 (diabetic with low-risk lifestyle). Similarly, compared with respondents in the 15-19 years age group, those in the 45-49 age group were 1.7 times (OR=1.73; p-<0.01) more likely to belong to class 3 (hypertensive with low-risk lifestyle) than to class 1 (diabetic with low-risk lifestyle). Also, we observed that the odds of being in class 3 increases with the age of the respondents. With respect to males, females were less likely (OR=0.002; p-<0.01) to belong to class 2 but more likely (OR=1.46; p-<0.01) to belong to class 3. Higher odds of being in class 2-3 than to class 1 was observed among the married respondents (OR=1.54). Similarly, widowed/divorced/separated were more likely to be in class 2 (OR=3.35; p-<0.01), but less likely to be in class 3 (OR=0.98; p-<0.01). Moreover, compared to the illiterate respondents, those with higher secondary education were less likely to belong to class 3 than to class 1 (OR=0.79; p-<0.01), and being in the richest wealth quintile lower is the odds of being in class 3 than in class 1 (OR=0.69; p-<0.01).

**Table 5: The odds for various latent classes of lifestyle behaviours and their associated covariates among individuals age 15-49 years, Northeastern India, NFHS-4, 2015-16**

| Background Characteristics | Class 2 vs Class 1 | | Class 3 vs Class 1 | |
|---|---|---|---|---|
| | Odds Ratio | S. E | Odds Ratio | S. E |
| **Age** | | | | |
| 15-19 | Ref | | Ref | |
| 20-24 | 4.10 *** | 0.25 | 1.27 ** | 0.09 |
| 25-29 | 5.10 *** | 0.25 | 1.34 *** | 0.09 |
| 30-34 | 5.70 *** | 0.26 | 1.42 *** | 0.09 |
| 35-39 | 3.53 *** | 0.26 | 1.42 *** | 0.09 |
| 40-44 | 3.82 *** | 0.27 | 1.51 *** | 0.09 |
| 45-49 | 3.46 *** | 0.27 | 1.73 *** | 0.09 |

| | | | | |
|---|---|---|---|---|
| **Gender** | | | | |
| Male | Ref | | Ref | |
| Female | 0.002 *** | 0.34 | 1.46 *** | 0.07 |
| **Education** | | | | |
| Illiterate | Ref | | Ref | |
| Primary | 1.03 | 0.19 | 0.73 *** | 0.06 |
| Secondary | 1.08 | 0.17 | 0.79 *** | 0.05 |
| Higher Secondary | 0.70 | 0.21 | 0.79 *** | 0.08 |
| **Marital Status** | | | | |
| Never married | Ref | | Ref | |
| Married | 1.54 ** | 0.15 | 1.16 ** | 0.06 |
| Widowed/Divorced/Separated | 3.35 *** | 0.35 | 0.98 | 0.08 |
| **Wealth Index** | | | | |
| Poorest | Ref | | Ref | |
| Poorer | 0.90 | 0.18 | 1.00 | 0.06 |
| Middle | 1.02 | 0.18 | 0.87 ** | 0.07 |
| Richer | 1.14 | 0.19 | 0.86 * | 0.07 |
| Richest | 1.13 | 0.22 | 0.69 *** | 0.08 |

*Ref: Reference group; ***p-value<0.01; **p-value<0.05; *p-value<0.1; S.E: Standard Error; Reference group: Class 1*

## Discussion

This study is the first attempt in Northeastern India, a region characterized by a seemingly homogenous racial group but inhabited by ethno-lingually diverse sub-populations, that applied latent class analysis to provide evidence about the pattern of modifiable risk factors among people suffering from diabetes or hypertension. Based on the observed modifiable health risk behaviours, the LCA suggests presence of three classes of respondents with diabetes or hypertension: "diabetic with low-risk lifestyle", "high-risk lifestyle", and "hypertensive with low-risk lifestyle". The findings portray that the 22,249 odd diabetic or hypertensive persons in the region can be categorized into three groups or classes, with homogeneity in their characteristics within the group but heterogeneous between the groups. One of the three latent classes has a 100 percent likelihood of having one of the risk factors combined with a low and moderate probability of having the other five risk factors. Each of the classes exhibit a unique risk factor configurations, and socio-economic and demographic profiles from one another. Thus, chronic diseases prevention programmes that specially target and reach out to the at-risk sub-populations will be beneficial and effective, as illustrated in this analysis (Table 3 and 5).

More than 70 percent of the respondents belong to the class "hypertensive with low-risk lifestyle" have their blood pressure level higher than normal and less likely to smoke and use alcohol. Evidence from a large body of existing literature supports that cigarette smoking leads to an immediate increase in blood pressure level and heart rate [48]. The effect of cigarette

smoking on blood pressure and heart rate is largely due to the nicotine content in the cigarettes [48]. However, studies also found a significant elevation in blood pressure and an increase in the prevalence of hypertension following smoking cessation [49,50], which could possibly be due to post-cessation weight gain [48,49]. Green et al. (1986) had concluded that smoking cessation in itself is not likely to result in a lowering blood pressure but on the contrary it may result even in higher blood pressure [51]. Moreover, about 8 percent of the respondents belonging to the "high-risk lifestyle" class mostly drink alcohol and smoke cigarettes. This co-occurrence of alcohol use and cigarette smoking have been well documented in the existing studies [52,53], and evidence suggest that a common genetic trait play a role in the co-occurance of smoking and alcohol use [54,55]. Epidemiological studies have also shown that individuals who are dependent on nicotine are more than four time more likely to be dependent on alcohol. Similarly, individuals who are dependent on alcohol are also more than three times more likely to be dependent on nicotine with reference to the general population [56]. Stress is another factor that influences use of tobacco and alcohol. For example, people who drink alcohol are most likely to smoke in an attempt to overcome stress [57]. Literature also suggest cross tolerance effect between alcohol and tobacco. It is found that the stimulating effect of nicotine in tobacco is restrained by the sedating effect of alcohol, which entices smokers to drink more alcohol [58]. Alcohol's sedating effect reduces the physiological effect of nicotine, thus increasing the consumption of tobacco.

In all the three identified classes, a vast majority of diabetic or hypertensive persons appear to have a habit of consuming aerated drinks and an uncontrolled body weight. Studies suggest that sugar-sweetened beverages intake is the major contributor to weight gain and can increase the risk of T2 diabetes and hypertension [59,60]. Cross-sectional studies acknowledge the positive association between overweight and obesity with the consumption of aerated drinks or sugar-sweetened beverages. For example, studies infer that women who consumed sweetened beverages more than once a week were more likely to be overweight than those who consumed sweetened beverages less than once a week [61]. Other studies found that women who consumed sugar-sweetened beverages regularly were heavier by 0.2 kg than the non-consumers [62]. The intake of sodas sweetened with high-fructose corn syrup is associated with hypertension among adolescents population [63]. Also, a spurt in the prevalence of hypertension have been observed among individuals who consumed more than average amount of fructose [64]. Evidence from a cross-sectional study in Northeast India supports the positive correlation between overweight or obesity with diabetes and hypertension [26]. The plausible

explanation is that both sugar sweetened beverages and artificial sweetened beverages are associated with the development of metabolic derangements that in turn might lead to elevated blood pressure [65]. High intake of sugar-sweetened beverages contributes to a higher total energy consumption, which increases energy intake without an increase in energy expenditure certainly results in weight gain. There is no dearth of scientific evidence which support the notion that decreasing aerated drinks consumption reduces the prevalence of overweight and obesity.

As observed in a previous study [66], this study also found that those aged 20 years and above are more likely to be in the 'high-risk cluster' than those below 20 years. Studies have elucidated that the adolescence or teenage stage is just the initial stage when an individual starts using or exposed to alcohol or tobacco (smoking) products. However, as this behaviour becomes a habit later on, it subsequently accentuates the intensity of substance (ab)use among older adults [67,68]. Also, it is well documented that initiation of smoking and drinking alcohol at an early age is associated with the number of cigarettes smoked and the quantity of alcohol consumed per day in adult age. According to Investigator (1991) and Dawson et al. (2008), individuals who began to smoke or drink before the age of 20 years were most likely to consumed a large volume of alcohol and smoked more cigarettes per day as compared to those who started after the age of 20 or above [69,70]. Further, compared to young adults, older adults were more likely to report high blood pressure and less likely to smoke or drink alcohol. One reason could be the existing condition of high blood pressure; an older adult was most likely to have taken a precaution by lowering or quitting smoking and alcohol consumption to prevent further deterioration of health. Evidence from a recent study among older adults demonstrated that the main reason for smoking cessation attempt was the motivation towards better health [71]. Hypertensive smokers were encouraged to quit smoking because of the risk of developing a severe form of hypertension such as malignant and renovascular hypertension [72]. Gender factors are also found to be significantly associated with class 2 "high-risk lifestyle" and 3 "hypertensive with low-risk lifestyle". Consistent with the previous studies, in this study also, female respondents were less likely to be in the high-risk cluster than males [73]. One primary reason stated is that smoking and drinking alcohol by women is not socially and culturally accepted by many [74], and women are more concerned about their health and tend to avoid unhealthy lifestyle behaviours [75]. In our study, individuals from a wealthier households with higher education were less likely to belong to a cluster of hypertensive with low substance use. An individual with higher education and better income can afford to pay for

a healthier lifestyle, including regular physical exercise, accessibility to advanced and quality healthcare services, as such efforts may reduce the risk of hypertension [76,77]. Being widowed/divorced/separated tend to increase the likelihood of falling into the high-risk cluster. Substance use by marital status indicated that widowed or divorced individuals are most likely to consumed alcohol, cigarette, and marijuana as compared to singles [78,79].

A search of similar studies indicates that no study in India or Northeast India has examined the clustering of health risk factors using the latent class approach. However, our study has some limitations. Firstly, the cross-sectional design of the study restricts us to measure the causal effect. Second, due to the dichotomous response on alcohol consumption, we cannot measure the quantity of alcohol an individual consumed.

**Conclusion**

In Northeastern India, as this study illustrates, there are three classes of modifiable risk factors closely related to the risk of chronic diseases (diabetes and hypertension). In addition, the study also identified factors that uniquely distinguished the identified classes. The intake of aerated drinks and obesity are the common modifiable risk factors in all three classes. As an initiative to healthy lifestyle behaviours, the findings suggest that males should control smoking and alcohol consumption, while females should control body weight and blood pressure. Further, we found that marital status and gender could be the catalyst to prevent high-risk smoking and alcohol drinking in Northeastern India. Also, the study observes smoking and alcohol use are the two modifiable risk behaviours, which tends to co-occur within an individual. Therefore, it is recommended that policy and intervention programmes in Norteastern India that promote healthy lifestyles should focus on targeting multiple modifiable risk behaviours that are most likely to co-occur within an individual. On the other hand, awareness about adapting to healthy diets and weight control along with physical exercises should be promoted with rigour in the region.

**Acknowledgement:** The authors are thankful to the two reviewers for their insightful comments and valuable suggestions, which hugely improved the analysis and the overall quality of the paper.

**Contributors:** Conceived and designed the research paper: SPM, HL. Analysed the data: SPM, HSC. Contributed agents/materials/analysis tools: SPM, HL, HSC. Drafting manuscript: SPM, HSC. Refined the manuscript: SPM, HL, HSC.

**Funding:** This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

**Competing Interest:** None declared.

**Patient consent for publication:** Not applicable

**Ethics approval**: This study is based on the secondary data which is publicly available from the Demographic Health Survey website and the respondents are de-identified, hence ethical approval is not required for this study.

**Data availability statement**: Data stored and maintained by the Demographic Health Survey office, and available to public at https://dhsprogram.com/data/dataset/India_Standard-DHS_2015.cfm?flag=1 upon request.

**Word count**: 4464

**References**

1 Chobanian A V, Bakris GL, Black HR, *et al.* The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA* 2003;**289**:2560–72. doi:10.1001/jama.289.19.2560

2 International Diabetes Federation. Advocacy guide to the IDF Diabetes Atlas Ninth edition 2019. Brussels, Belgium:International Diabetes Federation: 2019. http://www.diabetesatlas.org/ (accessed 14 Nov 2019).

3 Ford ES, Zhao G, Tsai J, *et al.* Low-Risk Lifestyle Behaviors and All-Cause Mortality: Findings From the National Health and Nutrition Examination Survey III Mortality Study. *Am J Public Health* 2011;**101**:1922–9. doi:10.2105/AJPH.2011.300167

4 Geldsetzer P, Manne-Goehler J, Theilmann M, et al. Geographic and sociodemographic variation of cardiovascular disease risk in India: A cross-sectional study of 797,540 adults. PLOS Med 2018;15:e1002581. doi:10.1371/journal.pmed.1002581

5 Ghosh S, Kumar M. Prevalence and associated risk factors of hypertension among persons aged 15–49 in India: a cross-sectional study. BMJ Open 2019;9:e029714. doi:10.1136/bmjopen-2019-029714

6 Gupta R, Gaur K, S. Ram CV. Emerging trends in hypertension epidemiology in India. J Hum Hypertens 2019;33:575–87. doi:10.1038/s41371-018-0117-3

7 Tandon N, Anjana RM, Mohan V, et al. The increasing burden of diabetes and variations

among the states of India: the Global Burden of Disease Study 1990-2016. Lancet Glob Heal 2018;6:e1352–62. doi:10.1016/S2214-109X(18)30387-5

8 WHO. (2014). Global status report on noncommunicable diseases 2014. In World Health Organization.

9 Kvaavik, E. (2010). Influence of Individual and Combined Health Behaviors on Total and Cause-Specific Mortality in Men and Women. Archives of Internal Medicine, 170(8), 711. https://doi.org/10.1001/archinternmed.2010.76

10 Loef, M., & Walach, H. (2012). The combined effects of healthy lifestyle behaviors on all cause mortality: A systematic review and meta-analysis. Preventive Medicine, 55(3), 163–170. https://doi.org/10.1016/j.ypmed.2012.06.017

11 Ha S, Choi HR, Lee YH. Clustering of four major lifestyle risk factors among Korean adults with metabolic syndrome. *PLoS One* 2017;**12**:e0174567. doi:10.1371/journal.pone.0174567

12 Poortinga W. The prevalence and clustering of four major lifestyle risk factors in an English adult population. *Prev Med (Baltim)* 2007;**44**:124–8. doi:10.1016/j.ypmed.2006.10.006

13 Schuit AJ, van Loon AJM, Tijhuis M, *et al.* Clustering of Lifestyle Risk Factors in a General Adult Population. *Prev Med (Baltim)* 2002;**35**:219–24. doi:10.1006/pmed.2002.1064

14 Leventhal AM, Huh J, Dunton GF. Clustering of modifiable biobehavioral risk factors for chronic disease in US adults: a latent class analysis. Perspect Public Health 2014;134:331–8. doi:10.1177/1757913913495780

15 Lisha NE, Martens M, Leventhal AM. Age and gender as moderators of the relationship between physical activity and alcohol use. Addict Behav 2011;36:933–6. doi:10.1016/j.addbeh.2011.04.003.

16 Akbartabartoori M, Lean MEJ, Hankey CR. Relationships between cigarette smoking, body size and body shape. Int J Obes 2005;29:236–43. doi:10.1038/sj.ijo.0802827

17 Shaikh R, Khan J. Clustering of lifestyle risk factors among adult population in India: A cross-sectional analysis from 2005 to 2016. *PLoS One* 2021;**16**:e0244559. doi:10.1371/journal.pone.0244559

18 Atorkey P, Owiredua C. Clustering of multiple health risk behaviours and association with socio-demographic characteristics and psychological distress among adolescents in Ghana: A latent class analysis. SSM - Popul Heal 2021;13:100707. doi:10.1016/j.ssmph.2020.100707

19 Bischops AC, De Neve J-W, Awasthi A, *et al.* A cross-sectional study of cardiovascular disease risk clustering at different socio-geographic levels in India. *Nat Commun* 2020;**11**:5891. doi:10.1038/s41467-020-19647-3

20 Laxer RE, Brownson RC, Dubin JA, et al. Clustering of risk-related modifiable behaviours and their association with overweight and obesity among a large sample of youth in the COMPASS study. BMC Public Health 2017;17:102. Doi:10.1186/s12889-017-4034-0

21 Lee DH, Nam JY, Kwon S, et al. Lifestyle risk score and mortality in Korean adults: a population-based cohort study. Sci Rep 2020;10:10260. doi:10.1038/s41598-020-66742-y

22 Gao B, Zhang L, Wang H. Clustering of Major Cardiovascular Risk Factors and the Association with Unhealthy Lifestyles in the Chinese Adult Population. *PLoS One* 2013;**8**:e66780. doi:10.1371/journal.pone.0066780

23 Hart CL, Davey Smith G, Gruer L, et al. The combined effect of smoking tobacco and drinking alcohol on cause-specific mortality: A 30 year cohort study. BMC Public Health 2010;10. doi:10.1186/1471-2458-10-789

24 Sobngwi E, Mbanya J-CN, Unwin N, *et al.* Physical activity and its relationship with obesity, hypertension and diabetes in urban and rural Cameroon. *Int J Obes* 2002;**26**:1009–16. doi:10.1038/sj.ijo.0802008

25 Bani T Aeri DS. Risk Factors Associated with the Increasing Cardiovascular Diseases Prevalence in India: A Review. *J Nutr Food Sci* 2014;**05**:1–6. doi:10.4172/2155-9600.1000331

26 Marbaniang SP, Lhungdim H, Yadav B, *et al.* Overweight/obesity risks and prevalence of diabetes and hypertension in North Eastern India: An analysis using seemingly unrelated probit model. *Clin Epidemiol Glob Heal* 2021;**11**:100764. doi:10.1016/j.cegh.2021.100764

27 Umamahesh K, Vigneswari A, Surya Thejaswi G, *et al.* Incidence of cardiovascular diseases and associated risk factors among subjects with type 2 diabetes – An 11-year follow up study. *Indian Heart J* 2014;**66**:5–10. doi:10.1016/j.ihj.2013.12.009

28 Sarveswaran G, Kulothungan V, Mathur P. Clustering of noncommunicable disease risk factors among adults (18–69 years) in rural population, South-India. *Diabetes Metab Syndr Clin Res Rev* 2020;**14**:1005–14. doi:10.1016/j.dsx.2020.05.042

29 Schreiber JB. Latent Class Analysis: An example for reporting results. *Res Soc Adm Pharm* 2017;**13**:1196–201. doi:10.1016/j.sapharm.2016.11.011

30 Weller BE, Bowen NK, Faubert SJ. Latent Class Analysis: A Guide to Best Practice. *J Black Psychol* 2020;**46**:287–311. doi:10.1177/0095798420930932

31 Marbaniang SP. Women Care and Practices in the Management of Childhood Diarrhea in Northeast India. *Child Care Pract* 2020;:1–13. doi:10.1080/13575279.2020.1812534

32 Dikshit KR, Dikshit JK. Population of the North-Eastern States of India. In: *North-East India: Land, People and Economy. Advances in Asian Human-Environmental Research*. Dordrecht: : Springer 2014. 421–56. doi:10.1007/978-94-007-7055-3_12

33 International Institute for Population Sciences (IIPS) and ICF. National Family Health Survey (NFHS-4), 2015-16: India. Mumbai:IIPS: 2017. http://rchiips.org/nfhs/NFHS-4Reports/India.pdf

34 DHS. The DHS Program-India: Standard DHS, 2015-16 Dataset,. Demogr. Heal. Surv. 2017.https://dhsprogram.com/data/dataset/India_Standard-DHS_2015.cfm?flag=1

35 Barba C, Cavalli-Sforza T, Cutter J, *et al.* Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* 2004;**363**:157–63. doi:10.1016/S0140-6736(03)15268-3

36 Islam MM, Valderas JM, Yen L, *et al.* Multimorbidity and Comorbidity of Chronic Diseases among the Senior Australians: Prevalence and Patterns. *PLoS One* 2014;**9**:e83783. doi:10.1371/journal.pone.0083783

37 Brusco MJ, Shireman E, Steinley D. A comparison of latent class, K-means, and K-median methods for clustering dichotomous data. *Psychol Methods* 2017;**22**:563–80. doi:10.1037/met0000095

38 Dey A, Chakraborty A, Majumdar K, *et al.* Application of latent class analysis to estimate susceptibility to adverse health outcomes based on several risk factors. *Int J Community Med Public Heal* 2016;**3**:3423–9. doi:10.18203/2394-6040.ijcmph20164268

39 Gao F, Chen J, Liu X, *et al.* Latent class analysis suggests four classes of persons with type 2 diabetes mellitus based on complications and comorbidities in Tianjin, China: a cross-sectional analysis. *Endocr J* 2017;**64**:1007–16. doi:10.1507/endocrj.EJ17-0199

40 Ghanbari J, Mohammadpoorasl A, Jahangiry L, *et al.* Subgroups of lifestyle patterns among hypertension patients: a latent-class analysis. *BMC Med Res Methodol* 2018;**18**:127. doi:10.1186/s12874-018-0607-6

41 Sugiura N. Further analysts of the data by akaike' s information criterion and the finite corrections. *Commun Stat - Theory Methods* 1978;**7**:13–26. doi:10.1080/03610927808827599

42 Akaike H. A New Look at the Statistical Model Identification. 1974. 215–22. doi:10.1007/978-1-4612-1694-0_16

43 Linzer DA, Lewis JB. poLCA : An R Package for Polytomous Variable Latent Class Analysis. *J Stat Softw* 2011;**42**. doi:10.18637/jss.v042.i10

44 Bolck A, Croon M, Hagenaars J. Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Polit Anal* 2004;**12**:3–27. doi:10.1093/pan/mph001

45 Rosato NS, Baer JC. Latent class analysis: A method for capturing heterogeneity. Soc Work Res 2012;36:61–9. doi:10.1093/swr/svs006

46 Wang M-C, Deng Q, Bi X, et al. Performance of the entropy as an index of classification accuracy in latent profile analysis: A Monte Carlo simulation study. Acta Psychol Sin 2017;49:1473. doi:10.3724/SP.J.1041.2017.01473

47 Celeux G, Soromenho G. An entropy criterion for assessing the number of clusters in a mixture model. J Classif 1996;13:195–212. doi:10.1007/BF01246098

48 Omvik P. How smoking affects blood pressure. Blood Press 1996;5:71–7. doi:10.3109/08037059609062111

49 Jenner DA, Puddey IB, Beilin LJ, *et al.* Lifestyle- and occupation-related changes in blood pressure over a six-year period in a cohort of working men. *J Hypertens Suppl Off J Int Soc Hypertens* 1988;**6**:S605-7. doi:10.1097/00004872-198812040-00189

50 Gerace TA, Hollis J, Ockene JK, *et al.* Smoking cessation and change in diastolic blood pressure, body weight, and plasma lipids. MRFIT Research Group. *Prev Med (Baltim)* 1991;**20**:602–20. doi:10.1016/0091-7435(91)90058-c

51 Green MS, Jucha E, Luz Y. Blood pressure in smokers and nonsmokers: Epidemiologic findings. Am Heart J 1986;111:932–40. doi:10.1016/0002-8703(86)90645-9

52 Ho RKS, Fok PWY, Chan HCH. Pattern and determinants of alcohol and tobacco co-use and its relationship with smoking cessation in Hong Kong. *Tob Prev \& Cessat* 2021;**7**:1–9. doi:10.18332/tpc/132288

53 Jiang N, Lee YO, Ling PM. Association between tobacco and alcohol use among young adult bar patrons: a cross-sectional study in three cities. *BMC Public Health* 2014;**14**:500. doi:10.1186/1471-2458-14-500

54 Slutske WS. Common Genetic Vulnerability for Pathological Gambling and Alcohol Dependence in Men. Arch Gen Psychiatry 2000;57:666–73. doi:10.1001/archpsyc.57.7.666

55 Bierut LJ, Schuckit MA, Hesselbrock V, et al. Co-occurring risk factors for alcohol dependence and habitual smoking. Alcohol Res Health 2000;24:233–41.http://www.ncbi.nlm.nih.gov/pubmed/15986718

56 Grant BF, Hasin DS, Chou SP, et al. Nicotine Dependence and Psychiatric Disorders in the United States. Arch Gen Psychiatry 2004;61:1107. doi:10.1001/archpsyc.61.11.1107

57 Lyon RJ, Tong JE, Leigh G, et al. The influence of alcohol and tobacco on the components of choice reaction time. J Stud Alcohol 1975;36:587–96. doi:10.15288/jsa.1975.36.587

58 de Fiebre CM, Collins AC. A comparison of the development of tolerance to ethanol and cross-tolerance to nicotine after chronic ethanol treatment in long- and short-sleep mice. J Pharmacol Exp Ther 1993;266:1398–406.

59 Malik VS, Popkin BM, Bray GA, *et al.* Sugar-sweetened beverages, obesity, type 2 diabetes mellitus, and cardiovascular disease risk. *Circulation* 2010;**121**:1356–64. doi:10.1161/CIRCULATIONAHA.109.876185

60 Qin P, Li Q, Zhao Y, *et al.* Sugar and artificially sweetened beverages and risk of obesity, type 2 diabetes mellitus, hypertension, and all-cause mortality: a dose–response meta-analysis of prospective cohort studies. *Eur J Epidemiol* 2020;**35**:655–71. doi:10.1007/s10654-020-00655-y

61 Liebman M, Pelican S, Moore SA, et al. Dietary intake, eating behavior, and physical activity-related determinants of high body mass index in rural communities in Wyoming, Montana, and Idaho. Int J Obes 2003;27:684–92. doi:10.1038/sj.ijo.0802277

62 French S, Jeffery R, Forster J, et al. Predictors of weight change over two years among a population of working adults: the Healthy Worker Project. Int J Obes Relat Metab Disord 1994;18 3:145–54.

63 Nguyen S, Choi HK, Lustig RH, et al. Sugar-Sweetened Beverages, Serum Uric Acid, and Blood Pressure in Adolescents. J Pediatr 2009;154:807–13. doi:10.1016/j.jpeds.2009.01.015

64 Jalal DI, Smits G, Johnson RJ, et al. Increased Fructose Associates with Elevated Blood Pressure. J Am Soc Nephrol 2010;21:1543–9. doi:10.1681/ASN.2009111111

65 Cohen L, Curhan G, Forman J. Association of sweetened beverage intake with incident hypertension. J Gen Intern Med 2012;27:1127–34. Doi:10.1007/s11606-012-2069-6

66 Saikia N, Debbarma B. The socioeconomic correlates of substance use among male adults in Northeast India. *Clin Epidemiol Glob Heal* 2020;**8**:149–57. doi:10.1016/j.cegh.2019.06.004

67 IOM (Institute of Medicine). Patterns of Tobacco Use by Adolescents and Young Adults. In: Richard J. Bonnie, Kathleen Stratton LYK, ed. *Public Health Implications of Raising the Minimum Age of Legal Access to Tobacco Products.* Washington, D.C.: : The National Academies Press 2015. https://www.ncbi.nlm.nih.gov/books/NBK310412/pdf/Bookshelf_NBK310412.pdf

68 Maggs JL, Schulenberg JE. Initiation and Course of Alcohol Consumption among

Adolescents and Young Adults. In: *Recent Developments in Alcoholism*. New York: : Kluwer Academic Publishers-Plenum Publishers 2005. 29–47. doi:10.1007/0-306-48626-1_2

69 Investigators S. Effect of the Age at Which Smoking Begins on Frequency of Smoking in Adulthood. *N Engl J Med* 1991;**325**:968–9. doi:10.1056/NEJM199109263251318

70 Dawson DA, Goldstein RB, Patricia Chou S, *et al.* Age at First Drink and the First Incidence of Adult-Onset DSM-IV Alcohol Use Disorders. *Alcohol Clin Exp Res* 2008;**32**:2149–60. doi:10.1111/j.1530-0277.2008.00806.x

71 Tanihara S, Momose Y. Reasons for smoking cessation attempts among Japanese male smokers vary by nicotine dependence level: a cross-sectional study after the 2010 tobacco tax increase. *BMJ Open* 2015;**5**:e006658–e006658. doi:10.1136/bmjopen-2014-006658

72 Freitas SRS, Alvim RO. Smoking and blood pressure phenotypes: New perspective for an old problem. *Am J Hypertens* 2017;**30**:554–5. doi:10.1093/ajh/hpx039

73 Saikia B, Marbaniang SP, Kumar P, *et al.* Changing pattern of tobacco consumption and quitting behavior in Northeast India. *J Subst Use* 2021;:1–12. doi:10.1080/14659891.2021.1875068

74 Wilsnack RW, Wilsnack SC, Kristjanson AF, *et al.* Gender and alcohol consumption: patterns from the multinational GENACIS project. *Addiction* 2009;**104**:1487–500. doi:10.1111/j.1360-0443.2009.02696.x

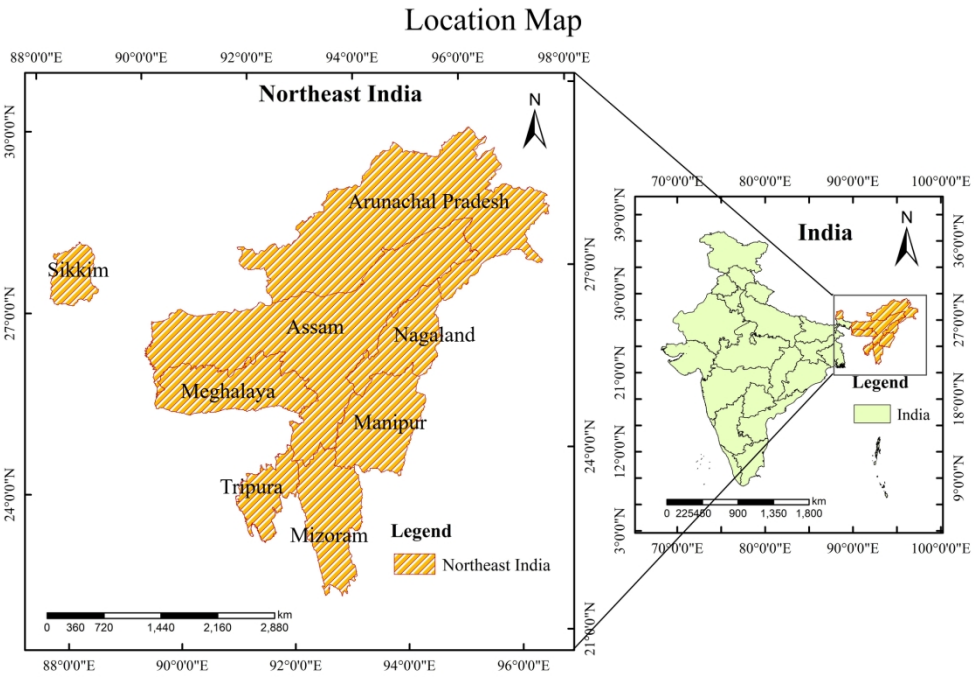
75 Nolen-Hoeksema S. Gender differences in risk factors and consequences for alcohol use and problems. *Clin Psychol Rev* 2004;**24**:981–1010. doi:10.1016/j.cpr.2004.08.003

76 Rasiah R, Yusoff K, Mohammadreza A, *et al.* Cardiovascular disease risk factors and socioeconomic variables in a nation undergoing epidemiologic transition. *BMC Public Health* 2013;**13**:886. doi:10.1186/1471-2458-13-886

77 Winkleby MA, Jatulis DE, Frank E, *et al.* Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. *Am J Public Health* 1992;**82**:816–20. doi:10.2105/AJPH.82.6.816

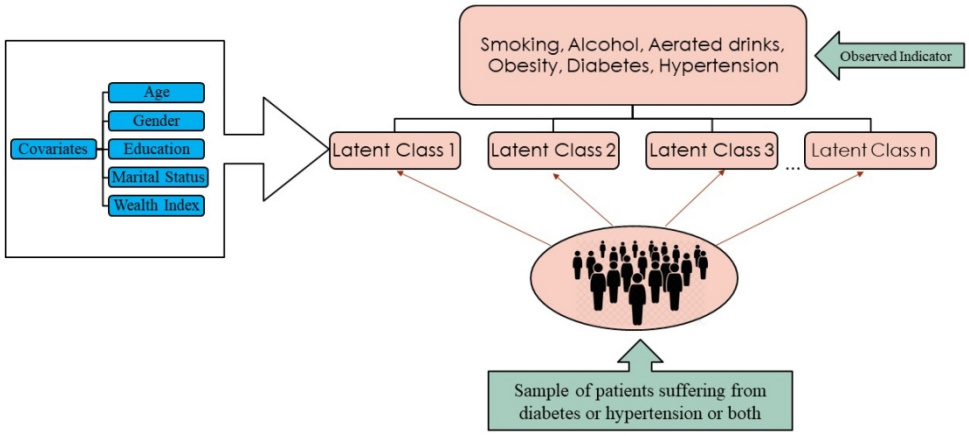
78 Menasco M, Blair S. Gender Differences in Substance Use Across Marital Statuses. *Int J Criminol Sociol* 2016;**5**:1–13. doi:10.6000/1929-4409.2016.05.01

79 Power C, Rodgers B, Hope S. Heavy alcohol consumption and marital status: disentangling the relationship in a national study of young adults. *Addiction* 1999;**94**:1477–87. doi:10.1046/j.1360-0443.1999.941014774.x



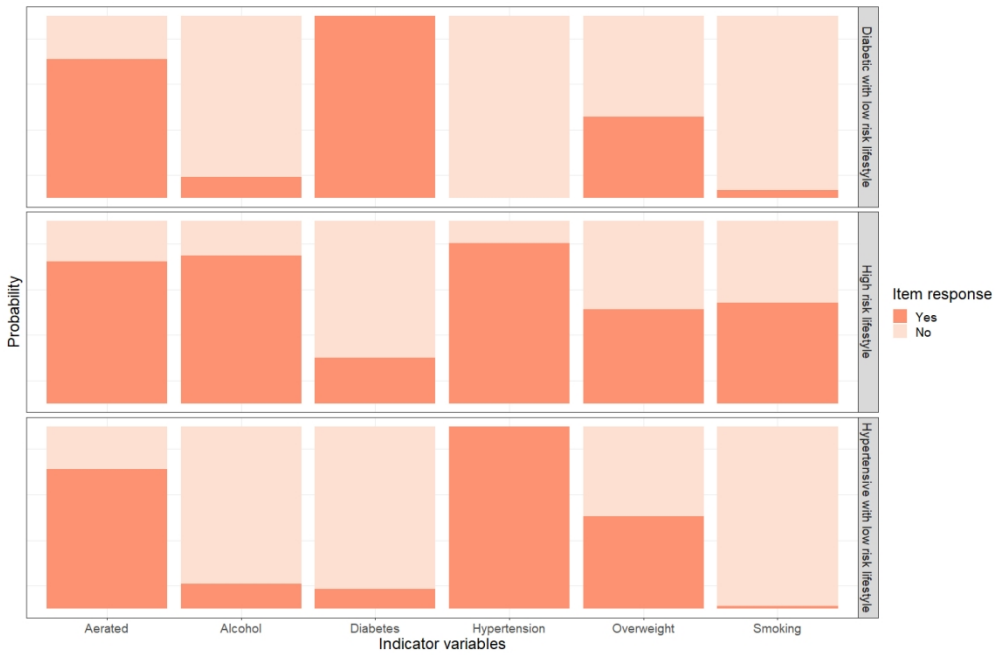

Map showing the location of the Study area. Source: Authors

595x412mm (96 x 96 DPI)




Conceptual framework of the study. Source: Authors

108x60mm (300 x 300 DPI)




Item response conditional probability across the latent classes. Source: Authors

129x84mm (300 x 300 DPI)

**STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies***

| Section/Topic | Item # | Recommendation | Reported on page # |
|---|---|---|---|
| **Title and abstract** | 1 | (*a*) Indicate the study's design with a commonly used term in the title or the abstract | 1 |
| | | (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found | 1 |
| **Introduction** | | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported | 4 |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses | 4 |
| **Methods** | | | |
| Study design | 4 | Present key elements of study design early in the paper | 5 |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | 5 |
| Participants | 6 | (*a*) Give the eligibility criteria, and the sources and methods of selection of participants | 5 |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | 6 |
| Data sources/ measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | 6 |
| Bias | 9 | Describe any efforts to address potential sources of bias | |
| Study size | 10 | Explain how the study size was arrived at | 5 |
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | |
| Statistical methods | 12 | (*a*) Describe all statistical methods, including those used to control for confounding | 6 |
| | | (*b*) Describe any methods used to examine subgroups and interactions | |
| | | (*c*) Explain how missing data were addressed | |
| | | (*d*) If applicable, describe analytical methods taking account of sampling strategy | |
| | | (*e*) Describe any sensitivity analyses | |
| **Results** | | | |

| | | | |
|---|---|---|---|
| Participants | 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed | |
| | | (b) Give reasons for non-participation at each stage | |
| | | (c) Consider use of a flow diagram | |
| Descriptive data | 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders | 7 |
| | | (b) Indicate number of participants with missing data for each variable of interest | |
| Outcome data | 15* | Report numbers of outcome events or summary measures | 7 |
| Main results | 16 | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included | 7 |
| | | (*b*) Report category boundaries when continuous variables were categorized | |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | |
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses | |
| **Discussion** | | | |
| Key results | 18 | Summarise key results with reference to study objectives | 8 |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias | 10 |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence | 9-10 |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results | 10 |
| **Other information** | | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | 11 |

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

# Identifying the latent classes of modifiable risk behaviours among diabetic and hypertensive individuals in North Eastern India: a population-based cross-sectional study



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2021-053757.R2 |
| Article Type: | Original research |
| Date Submitted by the Author: | 03-Feb-2022 |
| Complete List of Authors: | Marbaniang, Strong P; International Institute for Population Sciences, Public Health and Mortality Studies<br>Lhungdim, Hemkhothang; International Institute for Population Sciences, Public Health and Mortality Studies<br>Chungkham, Holendro; Indian Statistical Institute North-East Centre at Tezpur |
| Primary Subject Heading: | Public health |
| Secondary Subject Heading: | Cardiovascular medicine, Epidemiology, Global health, Health policy, Smoking and tobacco |
| Keywords: | Hypertension < CARDIOLOGY, General diabetes < DIABETES & ENDOCRINOLOGY, Epidemiology < TROPICAL MEDICINE |
| | |

*I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.*

*The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.*

*Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.*

# Identifying the latent classes of modifiable risk behaviours among diabetic and hypertensive individuals in North Eastern India: a population-based cross-sectional study

**Strong P Marbaniang[1,3*], Hemkhothang Lhungdim[1] , Holendro Singh Chungkham[2,4]**

[1]Department of Public Health & Mortality Studies, International Institute for Population Sciences, Mumbai-400088, Maharashtra, India.

[2]Indian Statistical Institute, North-East Centre, Tezpur-784501, Assam, India

[3]Department of Statistics, Sankardev College, Shillong-793004, Meghalaya, India

[4]Stress Research Institute, Stockholm University, Stockholm, Sweden

*Corresponding author

**Strong P Marbaniang**
**Email address-** strongmarbaniang@yahoo.com

**MeSh Keywords:**

Type 2 Diabetes; Hypertension, Alcohol drinking, Smoking; Latent Class Analysis; North Eastern India.

**Word Count**: 4464

## Abstract

**Objective**: To identify the latent classes of modifiable risk factors among the diabetic and hypertensive patients based on the observed indicators variables: smoking, alcohol, aerated drinks, overweight or obesity, diabetes, and hypertension. We hypothesized that the study population diagnosed with diabetes or hypertension is homogeneous with respect to the modifiable risk factors.

**Design:** A cross-sectional study using a stratified random sampling method, and a nationally representative large-scale survey.

**Setting and participants:** Data come from the fourth round of the Indian National Family Health Survey, 2015-16. Respondents aged 15-49 years who were diagnosed with either diabetes or hypertension or both were included. The total sample is 22,249, out of which 3,284 were males, and 18,965 were females.

**Primary and secondary outcome measures:** The observed variables used as latent indicators are: smoking, alcohol, aerated drinks, overweight or obesity, diabetes, and hypertension. The concomitant variables include age, gender, education, marital status, and household wealth index. Latent class model was used to simultaneously identify the latent class and to determine the association between the concomitant variables and the latent classes.

**Results:** Three latent classes were identified and labelled as Class 1: "Diabetic with low-risk lifestyle" (21%), Class 2: "High-risk lifestyle" (8%), and Class 3: "Hypertensive with low-risk lifestyle" (71%). Class 1 is characterised by those with a high probability of having diabetes and low probability of smoking and drinking alcohol. Class 2 characterized by a high probability of smoking and drinking alcohol, and Class 3 by a high probability of having high blood pressure and low probability of smoking and drinking alcohol.

**Conclusions:** Co-occurrence of smoking and alcohol consumption were prevalent in men, while excess body weight and high blood pressure were prevalent in women. Policy and programmes in Northeastern India should focus on targeting multiple modifiable risk behaviours that co-occur within an individual.


**Keywords**: Modifiable risk; Diabetes; Hypertension; Latent class analysis; Northeastern India.

**Strengths and limitations of the study**

- The sample for the study is relatively large as it uses a nationally representative large-scale data.
- Respondents' status of diabetes and hypertension is as diagnosed during fieldwork (on the day of interview) using a standard biomarker instrument and measurement.
- Latent class analysis (LCA) is a person-centred technique and the appropriate method for explaining the clustering of health risk behaviours.
- The cross-sectional design of the study restricts us to measure the causal effect.
- Modifiable risk factors were collected based on dichotomous responses, which limits measuring the intensity of substance use.

## Introduction

Effective lifestyle modification plans and affordable approaches to control high blood sugar and high blood pressure level benefit the at-risk population [1,2]. The leading causes of mortality due to chronic diseases in adults, particularly older adults, are linked to unhealthy lifestyle and behaviours [3], such as tobacco consumption, physical inactivity, excess alcohol consumption, and poor diet. An alarmingly high prevalence of hypertension in North Eastern (NE) India has been reported in recent studies [4-6]. Sikkim has the highest prevalence of hypertension among all the Indian states, while other NE states such as Nagaland, Assam, Arunachal Pradesh, Mizoram, and Tripura have prevalence above 11.0 percent, which is well above the national average [5,6]. Further, according to the Global Burden of Disease (GBD) study, the percentage change of diabetes prevalence between 1990 and 2016 has increased by more than 20 percent across the states of Northeastern India [7]. This percentage change was highest in the states of Sikkim, Arunachal Pradesh, and Nagaland.

Modifiable health risk behaviours such as smoking, heavy consumption of alcohol, low physical activity, and unhealthy diet characterized by high intake of sugar and fats, low consumption of fruits and vegetables are the major causes of non-communicable diseases (NCDs) [8], and strongly linked with early mortality [9,10]. Usually, these risk behaviours exist simultaneously or cluster within individuals [11]. It is argued that lifestyle risk factors within individuals are not random, but more likely to cluster with other unhealthy behaviours [12,13]. Further, when two risk behaviours cluster together with each other more than the other factors, it may suggests that they could be influenced by a common source [14]. Moreover, modifiable risk behaviours for chronic diseases such as excessive use of alcohol, cigarette smoking, obesity and unhealthy diets are unlikely to occur in an entirely uniform manner in a population. That is, it is uncertain whether people can be accurately classify into two groups, as those adapting to a healthy versus those unhealthy lifestyle. For example, certain individuals who engage in more vigorous physical activities tend to consume alcohol more frequently than their inactive counterparts [15]. Also, cigarette smoking is associated with lower rate of obesity among certain individuals [16]. In a study among the US adults, Leventhal et al. (2014) used modifiable risk factors like usage of alcohol, drug, nicotine, current obesity status, and weekly physical activity and found that the adult population clusters into five sub-groups based on the pattern of these modifiable risk factors [14]. Another study in India by Shaikh and Khan (2021),

found that hypertension is more likely to cluster with other modifiable risk behaviours such as smoking, alcohol, chewing tobacco, obesity, and unhealthy diets in both men and women [17]. Atorkey and Owiredua (2021), in their study that identified the clustering of five multiple health risk behaviours, also found that low vegetable and fruit intake, and low physical activity cluster together within an individual, whereas smoking tobacco and alcohol consumption co-occur together forming another cluster [18]. The pattern of clustering of modifiable risk behaviours could also differ by region and community level, as evident from previous studies [17,19]. The clustering of two or more modifiable risk behaviours is of great concern because it can intensify the risk of developing chronic diseases and cardiovascular mortality [20,21], and most intervention measures are specific to a single risk factors. For example, in Asia 43.5 percent of Chinese adults and 37 percent of Korean adults had at least two cardiovascular risk factors [11,22]. Moreover, in India, 35.6 percent of adult men and 10 percent adult women have at least three lifestyle risk factors [17]. It is also found that a combination of smoking and heavy drinking emerged as the riskiest behaviour for all causes of death [23].

The majority of studies reported a higher prevalence of smoking, excess alcohol consumption, physical inactivity, and poor diet among people with diabetes or hypertension than those without diabetes or hypertension [24,25]. Notably, previous studies have examined diabetic or hypertensive patients as a single homogeneous group without considering the possible heterogeneity existing within the population [26]. Most of the existing studies on health risk behaviours have emphasized only changing a single unhealthy behaviour [25,27]. However, interventions that simultaneously target these unhealthy behaviours are necessary from the policy perspective [11].

Existing studies in India on clustering of NCDs risk factors have used intraclass correlation and scoring methods [17,19,28]. However, studies have suggested that LCA is the most appropriate method for explaining the clustering of health risk behaviours [29,30]. LCA is a probabilistic model approach designed for identifying clusters based on dichotomous variables. This study is also an attempt to explore the clustering of NCD risk factors using the LCA approach in Northeastern India. The aims is to identify the latent classes of modifiable risk factors among diabetic and hypertensive patients based on the observed indicator variables such as smoking, alcohol, aerated drinks, overweight or obesity, diabetes, and hypertension. We applied LCA, also to investigate heterogeneity in the population of an individual with diabetes and hypertension based on the responses of observed risk factors. More specifically, the objective is to describe how individuals in Northeast India diagnosed with diabetes or

hypertension are clustered according to the pattern of observed risk factors and identify the difference between these groups based on socio-demographic characteristics. Information on clustering of health risk behaviours become vital as it guides in designing the disease control programs. Intervention approach that aimed at targeting multiple modifiable risk behaviours in a single program are gaining importance because they are potentially most cost-effective, efficient and may have a great public health impact than the single modifiable risk approaches. Therefore, we hypothesize that the study population diagnosed with diabetes or hypertension is homogeneous with respect to the modifiable risk factors.

## Methods

### *Study location*

The study focuses on the Northeastern region (NER) of India comprising eight states, namely Assam, Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, and Tripura (Figure 1). Most parts of the region has a hilly terrain and inhabited mainly by myriad tribes belonging to different cultures and ethnic communities [31]. The region has over 45 million population, which is 3.8 percent of India's population, with a density of 159 persons per $km^2$ [32].

### *Design and Setting*

The NFHS-4 (2015-16), a nationally representative large-scale survey, is a cross-sectional study that adopted a stratified two-stage random sampling design for both urban and rural areas, including a systematic mapping and listing of households in the selected primary sampling units (PSUs). The analysis is based on the publicly available data, in which the identity of selected respondents, households, and geocodes of PSUs are concealed [33]. In the first stage, the PSUs were selected based on probability proportional to population size (PPS). Rural PSUs are the villages, while census enumeration blocks (CEBs) formed the PSUs in urban areas. In the second stage, systematic random sampling was done in each PSUs to select households for the sample. At the household level, information were sought from the heads, eligible women age 15 – 49 years, and men age 15 – 54 years, along with their biomarkers and blood samples on a specially designed Dried Blood Spots (DBS). A detailed description of the survey design can be found elsewhere [33].

The data used in this study is the fourth round of the Indian National Family Health Survey (NFHS-4) conducted during 2015-16. The data was downloaded from the DHS website [34].

The NFHS series of surveys has been initiated by the Ministry of Health and Family Welfare, Government of India, and coordinated by the International Institute for Population Sciences, Mumbai.

***Sample***

The study includes only individuals age 15-49 years who were diagnosed with either diabetes or hypertension, or both. The total sample size is 22,249, of which 3,284 were males, and 18,965 were females for the Northeastern region of India.

***Patient and Public Involvement***

No patient involved

***Measures***

**Diabetes:** Fingerstick blood is collected, and blood glucose level is determined using the FreeStyle Optium H Glucometer. A respondent is considered diagnosed with diabetes if the random blood sugar level is >140mg/dl [33].

**Hypertension:** Blood pressure (BP) level was measured using an OMRON Blood Pressure measuring device. Three separate blood pressure readings were taken with an interval of 5 minutes between readings. A respondent is considered hypertensive if the systolic blood pressure is ≥ 140 mmHg, or diastolic blood pressure ≥ 90 mmHg, or if the respondent was taking antihypertensive medication to lower blood pressure at the time of the survey [33].

**Overweight or obesity:** In this study, we categorized continuous body mass index (BMI) according to the WHO guidelines for the Asian population. A respondent is classified as overweight or obese if BMI is ≥23 kg/m$^2$ [35].

***Outcome Variables of Interest***

**Observed Indicator variables:** The observed variables used as latent indicators are cigarette smoking (No=1, Yes=2), Alcohol consumption (No=1, Yes=2), Takes aerated drink (No=1, Yes=2), Overweight or obesity (No=1, Yes=2), Diabetes (No=1, Yes=2), and Hypertension (No=1, Yes=2).

***Predictors or Concomitant variables***

**Concomitant variables for class determination:** The concomitant variables include age, gender, education, marital status, and household wealth index. Figure 2 illustrates the conceptual framework of the study.

***Statistical Analysis***

Latent Class Analysis (LCA), based on structural equation modelling, is a person-centred analytical approach that allows the identification of homogeneous sub-groups in a heterogeneous population. LCA is similar to cluster analysis because individuals are classifying into homogenous unobserved (latent) groups based on the response pattern to a set of observed variables. However, in cluster analysis, objects are classified based on distance measures, whereas in LCA classification are based on probabilistic and finite mixture modelling approach [36,37]. A latent class model splits the population into mutually exclusive and exhaustive groups that are homogeneous within the groups but differ among them. Several studies have adopted the LCA approach to identify risk factors clustering within the heterogeneous population. In a study by Dey et al (2016), based on seven observed risk factors, LCA was used to classify individuals into two latent class "susceptible to adverse health outcomes groups" and "not susceptible to adverse health outcomes groups" [38]. Another study in China applied LCA to ten observed complications and comorbidities of T2 diabetic patients and categorized the individuals into four latent classes, namely "complications and comorbidity groups", "high risk of complications group", "high risk of comorbidities and Cardio Vascular Disease groups", and "diabetes without complications and comorbidities group" [39]. A study in West Azerbaijan province among the hypertensive patients aged 50 years and above used four indicators such as dietary patterns, physical activity, tobacco use, and high blood pressure control to categorized the hypertensive patients into three latent classes [40].

Latent Class Analysis is an innovative statistical method used to identify the latent classes of homogeneous individuals in a heterogeneous population [39]. The latent class was performed using the six indicators variables mentioned above. LCA uses observed dichotomous indicators to identify the unobserved latent class in a heterogeneous population. To determine the optimal number of latent classes, we examined one to four models. The optimal number of latent classes was chosen based on the model having the smallest Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (cAIC), Bayesian-Schwarz Information Criterion (BIC), adjusted Bayesian-Schwarz Information Criterion (aBIC), and Entropy [41,42]. A combination of parsimony and interpretability model selection

criteria was used to carefully select the number of latent classes so that individuals are allocated to their most likely class. The outputs of the latent class analysis includes the number of latent classes, the latent class probability (i.e., the probability that an individual selected at random belonged to each latent class), and the conditional probability (i.e., the probability that an individual would give a particular response to a specific item of an observed indicator variable given that an individual belongs to a specific latent class).

To determine the association between the concomitant variables (i.e., age, gender, education, marital status, and household wealth index) and the latent class, we apply the one-step technique approach. In this approach, the concomitant variables are included in the latent class regression model, and their coefficient is estimated simultaneously as part of the latent class model [43]. This approach has been demonstrated to provide the best and unbiased coefficient estimates of the concomitant variables as compared to other methods [44]. Statistical analysis was conducted in R statistical software, and LCA was analyze using the poLCA package (Polytomous Variable Latent Class Analysis) [43].

## Results

### *Descriptive statistics*

A total of 22,249 individuals with chronic diseases, i.e., diagnosed with diabetes or hypertension, were included in the analysis (Table 1). Individuals in the age group 45-49 and 15-19 constitute the largest (21.3%) and the lowest (5.0%) respectively in the total sample. Over half (51.2%) of the sample have completed secondary education, 24.6 percent were illiterate, and 79.6 percent were currently married. Further, among the chronic disease patients, the prevalence of smoking, consumption of alcohol, and aerated drinks were 4.2, 14.8, and 68.6 percent respectively. Also, 81.1 percent of the participants reported suffering from hypertension, and 27.7 percent from diabetes. The sample distribution also indicates that chronic condition is much higher among females (86.1%) than males (13.9%), among older ages (age 35+), and among poorer households.

**Table 1: Percent distribution of adults of age 15-49 years with chronic condition (diabetic and hypertensive patients) by selected background characteristics, Northeastern India, NFHS-4, 2015-16.**

| Background Characteristics | % (n=22,249) |
|---|---|
| **Age** | |
| 15-19 | 5.0 |
| 20-24 | 8.4 |
| 25-29 | 12.2 |
| 30-34 | 15.0 |
| 35-39 | 19.2 |
| 40-44 | 19.0 |
| 45-49 | 21.3 |
| **Sex** | |
| Male | 13.9 |
| Female | 86.1 |
| **Education** | |
| Illiterate | 24.6 |
| Primary | 16.3 |
| Secondary | 51.2 |
| Higher Secondary | 8.0 |
| **Marital Status** | |
| Never married | 13.6 |
| Married | 79.6 |
| Widowed/Divorced/Separated | 6.87 |
| **Wealth Index** | |
| Poorest | 16.7 |
| Poorer | 34.6 |
| Middle | 22.4 |
| Richer | 17.5 |
| Richest | 8.8 |
| **Observed Indicators** | |
| Smoking | 4.2 |
| Drink alcohol | 14.8 |
| Take aerated drinks | 68.6 |
| Overweight or Obesity | 43.3 |
| Diabetes | 27.7 |
| Hypertension | 81.1 |



***Model fit and selection of Latent class***

The model fit/selection statistics to derive appropriate models are shown in Table 2. We fit the LCA models with classes ranging from 1 to 4. For each LCA model, the model fit indicators BIC, AIC, cAIC, and Entropy were calculated. According to the model fit indicators, a three-Class model has been selected as the best fit model as it has the lowest BIC and cAIC values. Smaller values on each indicator suggest a better model, or a model with few explanatory variables or parameters, and inform the decision of the best model to be retained [45]. Another diagnostic indicator is entropy [46], which indicates how accurately the model defines the

classes. In general, an entropy value close to 1.0 is considered as ideal [47], and above 0.8 is also acceptable. An entropy value of 1.0 would indicate that every individual has been perfectly classified within the classes based on the responses of the observed indicators. The three latent classes were labelled as "Diabetic with low-risk lifestyle", "High-risk lifestyle", and "Hypertensive with low-risk lifestyle". Figure 3 portrays the graphical representation of item response probability for the indicator's variables across the latent classes.

**Table 2: Model fit statistics for the latent class models (n=22,249)**

| No of classes | log-likelihood | Residual df | BIC | aBIC | cAIC | likelihood-ratio | Entropy |
|---|---|---|---|---|---|---|---|
| **1** | -69173.1 | 57 | 138406.3 | 138387.2 | 138412.3 | 17350.8 | - |
| **2** | -61315.4 | 34 | 122921.0 | 122828.9 | 122950.0 | 1859.2 | 1.0 |
| **3** | -58920.2 | 11 | 118361.0 | 118195.7 | 118413.0 | 1168.6 | 0.9 |
| **4** | -59970.1 | -12 | 120691.0 | 120452.6 | 120766.0 | 2396.1 | 0.9 |

Latent class probability and the conditional probability of a "Yes" response for each indicator variables were summarized in Table 3. The last row in Table 3 indicates the probability of class membership in each latent class. About 21 percent of the participants were expected to belong to the Class 1 "diabetic with low-risk lifestyle", 8 percent to Class 2 "high-risk lifestyle", and 71 percent to Class 3 "hypertensive with low-risk lifestyle". Members in Class 1 were likely to report having high blood sugar level (diabetic) and unlikely to smoke cigarette or drinking alcohol. Members in Class 2 are more likely to smoke cigarette and drink alcohol, and members in Class 3 are likely to report having high blood pressure and unlikely to smoke cigarette or drink alcohol.

**Table 3: The conditional probability of the response items and the latent class probability for study sample in Northeastern India, NFHS-4, 2015-16**

| Indicator Variables | Diabetic with low-risk lifestyle (Class 1) | High-risk lifestyle (Class 2) | Hypertensive with low-risk lifestyle (Class 3) |
|---|---|---|---|
| Smoking | 0.04 | 0.55 | 0.02 |
| Drink alcohol | 0.12 | 0.81 | 0.14 |
| Take aerated drinks | 0.76 | 0.78 | 0.77 |
| Overweight or Obesity | 0.44 | 0.52 | 0.51 |
| Diabetic | 1.00 | 0.25 | 0.11 |
| Hypertensive | 0.00 | 0.88 | 1.00 |
| **Latent class probability** | **0.21** | **0.08** | **0.71** |

Table 4 presents the distribution of socio-economic and demographic characteristics of the individuals by the three latent classes. Age, gender, education, marital status, and household

wealth index were statistically found significantly associated with the three latent classes at 1 percent level of significant. The hypertensive with low-risk lifestyle class has the highest proportion of respondents aged 45-49 years, whereas both diabetic with low-risk lifestyle class and high-risk lifestyle class has the highest proportion of respondents aged 35-39 years (19.4% and 19.5% respectively). Both the diabetic with low-risk lifestyle class (10.6% males vs. 89.4% females) and the hypertensive with low-risk lifestyle class (6.2% males vs. 93.8% females) has the highest proportion of females; males constituted almost 100 percent of the high-risk lifestyle class (99.2% males vs. 0.8% females). Lower proportion of the respondents have completed higher secondary education in case of both diabetic with low-risk lifestyle class (10.5%) and hypertensive with low-risk lifestyle class (8.5%). Marital status shows a similar pattern in all the three classes; higher proportion of married respondents and lower proportion of widowed/divorced/separated respondents (<10%). The Wealth index shows that respondents belonging to the middle quintile constituted the largest proportion in the diabetic with low-risk lifestyle class (27.2%), and high-risk lifestyle class (26.6%). However, respondents belonging to the poorer wealth quintile have the highest proportion of hypertensive with low-risk lifestyle class (29.4%).

**Table 4: Percent distribution of diabetic and hypertensive patients age 15-49 years according to three latent classes by background characteristics, Northeastern India, NFHS-4, 2015-16**

| Background Characteristics | "Diabetic with low-risk lifestyle" class (n=4727, 21%) | "High-risk lifestyle" class (n=1828, 8%) | "Hypertensive with low-risk lifestyle" class (n=15,694, 71%) | P-value[a] |
|---|---|---|---|---|
| **Age** | | | | <0.01 |
| 15-19 | 8.08 | 3.06 | 4.96 | |
| 20-24 | 8.99 | 8.37 | 7.6 | |
| 25-29 | 12.88 | 13.51 | 11.93 | |
| 30-34 | 15.30 | 18.00 | 15.29 | |
| 35-39 | 19.40 | 19.53 | 19.46 | |
| 40-44 | 17.73 | 19.09 | 19.04 | |
| 45-49 | 17.62 | 18.44 | 21.72 | |
| **Gender** | | | | <0.01 |
| Male | 10.62 | 99.18 | 6.17 | |
| Female | 89.38 | 0.82 | 93.83 | |
| **Education** | | | | <0.01 |
| Illiterate | 17.96 | 11.65 | 25.17 | |
| Primary | 16.50 | 15.10 | 15.85 | |
| Secondary | 55.05 | 58.70 | 50.50 | |
| Higher Secondary | 10.49 | 14.55 | 8.47 | |
| **Marital Status** | | | | <0.01 |
| Never married | 19.93 | 21.55 | 14.30 | |
| Married | 72.75 | 75.22 | 78.41 | |
| Widowed/Divorced/Separated | 7.32 | 3.23 | 7.30 | |

| | | | | |
|---|---|---|---|---|
| **Wealth Index** | | | | <0.01 |
| Poorest | 10.94 | 12.14 | 12.94 | |
| Poorer | 26.06 | 24.84 | 29.43 | |
| Middle | 27.23 | 26.64 | 26.30 | |
| Richer | 22.49 | 23.34 | 21.25 | |
| Richest | 13.29 | 12.04 | 10.07 | |
| **Total** | 100.0 | 100.0 | 100.0 | |

*Note: [a]Based on chi-squared tests of association.*

### *Covariates predicting latent class membership*

Table 5 summarized the odds ratios (OR) results from the latent regression model. Compared with respondents in the 15-19 years age group, those in the 20-29 age group were 4.1 times (OR=4.10; p-<0.01) more likely to belong to class 2 (high-risk lifestyle) than to class 1 (diabetic with low-risk lifestyle). Similarly, compared with respondents in the 15-19 years age group, those in the 45-49 age group were 1.7 times (OR=1.73; p-<0.01) more likely to belong to class 3 (hypertensive with low-risk lifestyle) than to class 1 (diabetic with low-risk lifestyle). Also, we observed that the odds of being in class 3 increases with the age of the respondents. With respect to males, females were less likely (OR=0.002; p-<0.01) to belong to class 2 but more likely (OR=1.46; p-<0.01) to belong to class 3. Higher odds of being in class 2-3 than to class 1 was observed among the married respondents (OR=1.54). Similarly, widowed/divorced/separated were more likely to be in class 2 (OR=3.35; p-<0.01), but less likely to be in class 3 (OR=0.98; p-<0.01). Moreover, compared to the illiterate respondents, those with higher secondary education were less likely to belong to class 3 than to class 1 (OR=0.79; p-<0.01), and being in the richest wealth quintile lower is the odds of being in class 3 than in class 1 (OR=0.69; p-<0.01).

**Table 5: The odds for various latent classes of lifestyle behaviours and their associated covariates among individuals age 15-49 years, Northeastern India, NFHS-4, 2015-16**

| Background Characteristics | Class 2 vs Class 1 | | Class 3 vs Class 1 | |
|---|---|---|---|---|
| | Odds Ratio | S. E | Odds Ratio | S. E |
| **Age** | | | | |
| 15-19 | Ref | | Ref | |
| 20-24 | 4.10 *** | 0.25 | 1.27 ** | 0.09 |
| 25-29 | 5.10 *** | 0.25 | 1.34 *** | 0.09 |
| 30-34 | 5.70 *** | 0.26 | 1.42 *** | 0.09 |
| 35-39 | 3.53 *** | 0.26 | 1.42 *** | 0.09 |
| 40-44 | 3.82 *** | 0.27 | 1.51 *** | 0.09 |
| 45-49 | 3.46 *** | 0.27 | 1.73 *** | 0.09 |
| **Gender** | | | | |
| Male | Ref | | Ref | |
| Female | 0.002 *** | 0.34 | 1.46 *** | 0.07 |

| | | | | |
|---|---|---|---|---|
| **Education** | | | | |
| Illiterate | Ref | | Ref | |
| Primary | 1.03 | 0.19 | 0.73 *** | 0.06 |
| Secondary | 1.08 | 0.17 | 0.79 *** | 0.05 |
| Higher Secondary | 0.70 | 0.21 | 0.79 *** | 0.08 |
| **Marital Status** | | | | |
| Never married | Ref | | Ref | |
| Married | 1.54 ** | 0.15 | 1.16 ** | 0.06 |
| Widowed/Divorced/Separated | 3.35 *** | 0.35 | 0.98 | 0.08 |
| **Wealth Index** | | | | |
| Poorest | Ref | | Ref | |
| Poorer | 0.90 | 0.18 | 1.00 | 0.06 |
| Middle | 1.02 | 0.18 | 0.87 ** | 0.07 |
| Richer | 1.14 | 0.19 | 0.86 * | 0.07 |
| Richest | 1.13 | 0.22 | 0.69 *** | 0.08 |

*Ref: Reference group; ***p-value<0.01; **p-value<0.05; *p-value<0.1; S.E: Standard Error; Reference group: Class 1*

## Discussion

This study is the first attempt in Northeastern India, a region characterized by a seemingly homogenous racial group but inhabited by ethno-lingually diverse sub-populations, that applied latent class analysis to provide evidence about the pattern of modifiable risk factors among people suffering from diabetes or hypertension. Based on the observed modifiable health risk behaviours, the LCA suggests presence of three classes of respondents with diabetes or hypertension: "diabetic with low-risk lifestyle", "high-risk lifestyle", and "hypertensive with low-risk lifestyle". The findings portray that the 22,249 odd diabetic or hypertensive persons in the region can be categorized into three groups or classes, with homogeneity in their characteristics within the group but heterogeneous between the groups. One of the three latent classes has a 100 percent likelihood of having one of the risk factors combined with a low and moderate probability of having the other five risk factors. Each of the classes exhibit a unique risk factor configuration, and socio-economic and demographic profiles from one another. Thus, chronic diseases prevention programmes that specially target and reach out to the at-risk sub-populations will be beneficial and effective, as illustrated in this analysis (Table 3 and 5).

More than 70 percent of the respondents belong to the class "hypertensive with low-risk lifestyle" have their blood pressure level higher than normal and less likely to smoke and use alcohol. Evidence from a large body of existing literature supports that cigarette smoking leads to an immediate increase in blood pressure level and heart rate [48]. The effect of cigarette smoking on blood pressure and heart rate is largely due to the nicotine content in the cigarettes [48]. However, studies also found a significant elevation in blood pressure and an increase in

the prevalence of hypertension following smoking cessation [49,50], which could possibly be due to post-cessation weight gain [48,49]. Green et al. (1986) had concluded that smoking cessation in itself is not likely to result in a lowering blood pressure but on the contrary it may result even in higher blood pressure [51]. Moreover, about 8 percent of the respondents belonging to the "high-risk lifestyle" class mostly drink alcohol and smoke cigarettes. This co-occurrence of alcohol use and cigarette smoking have been well documented in the existing studies [52,53], and evidence suggest that a common genetic trait play a role in the co-occurrence of smoking and alcohol use [54,55]. Epidemiological studies have also shown that individuals who are dependent on nicotine are more than four time more likely to be dependent on alcohol. Similarly, individuals who are dependent on alcohol are also more than three times more likely to be dependent on nicotine with reference to the general population [56]. Stress is another factor that influences use of tobacco and alcohol. For example, people who drink alcohol are most likely to smoke in an attempt to overcome stress [57]. Literature also suggest cross tolerance effect between alcohol and tobacco. It is found that the stimulating effect of nicotine in tobacco is restrained by the sedating effect of alcohol, which entices smokers to drink more alcohol [58]. Alcohol's sedating effect reduces the physiological effect of nicotine, thus increasing the consumption of tobacco.

In all the three identified classes, a vast majority of diabetic or hypertensive persons appear to have a habit of consuming aerated drinks and an uncontrolled body weight. Studies suggest that sugar-sweetened beverages intake is the major contributor to weight gain and can increase the risk of T2 diabetes and hypertension [59,60]. Cross-sectional studies acknowledge the positive association between overweight and obesity with the consumption of aerated drinks or sugar-sweetened beverages. For example, studies infer that women who consumed sweetened beverages more than once a week were more likely to be overweight than those who consumed sweetened beverages less than once a week [61]. Other studies found that women who consumed sugar-sweetened beverages regularly were heavier by 0.2 kg than the non-consumers [62]. The intake of sodas sweetened with high-fructose corn syrup is associated with hypertension among adolescents population [63]. Also, a spurt in the prevalence of hypertension have been observed among individuals who consumed more than average amount of fructose [64]. Evidence from a cross-sectional study in Northeast India supports the positive correlation between overweight or obesity with diabetes and hypertension [26]. The plausible explanation is that both sugar sweetened beverages and artificial sweetened beverages are associated with the development of metabolic derangements that in turn might lead to elevated

blood pressure [65]. High intake of sugar-sweetened beverages contributes to a higher total energy consumption, which increases energy intake without an increase in energy expenditure certainly results in weight gain. There is no dearth of scientific evidence which support the notion that decreasing aerated drinks consumption reduces the prevalence of overweight and obesity.

As observed in a previous study [66], this study also found that those aged 20 years and above are more likely to be in the 'high-risk cluster' than those below 20 years. Studies have elucidated that the adolescence or teenage stage is just the initial stage when an individual starts using or exposed to alcohol or tobacco (smoking) products. However, as this behaviour becomes a habit later on, it subsequently accentuates the intensity of substance (ab)use among older adults [67,68]. Also, it is well documented that initiation of smoking and drinking alcohol at an early age is associated with the number of cigarettes smoked and the quantity of alcohol consumed per day in adult age. According to Investigator (1991) and Dawson et al. (2008), individuals who began to smoke or drink before the age of 20 years were most likely to consumed a large volume of alcohol and smoked more cigarettes per day as compared to those who started after the age of 20 or above [69,70]. Further, compared to young adults, older adults were more likely to report high blood pressure and less likely to smoke or drink alcohol. One reason could be the existing condition of high blood pressure; an older adult was most likely to have taken a precaution by lowering or quitting smoking and alcohol consumption to prevent further deterioration of health. Evidence from a recent study among older adults demonstrated that the main reason for smoking cessation attempt was the motivation towards better health [71]. Hypertensive smokers were encouraged to quit smoking because of the risk of developing a severe form of hypertension such as malignant and renovascular hypertension [72]. Gender factors are also found to be significantly associated with class 2 "high-risk lifestyle" and 3 "hypertensive with low-risk lifestyle". Consistent with the previous studies, in this study also, female respondents were less likely to be in the high-risk cluster than males [73]. One primary reason stated is that smoking and drinking alcohol by women is not socially and culturally accepted by many [74], and women are more concerned about their health and tend to avoid unhealthy lifestyle behaviours [75]. In our study, individuals from a wealthier household with higher education were less likely to belong to a cluster of hypertensive with low substance use. An individual with higher education and better income can afford to pay for a healthier lifestyle, including regular physical exercise, accessibility to advanced and quality healthcare services, as such efforts may reduce the risk of hypertension [76,77]. Being

widowed/divorced/separated tend to increase the likelihood of falling into the high-risk cluster. Substance use by marital status indicated that widowed or divorced individuals are most likely to consumed alcohol, cigarette, and marijuana as compared to singles [78,79].

A search of similar studies indicates that no study in India or Northeast India has examined the clustering of health risk factors using the latent class approach. However, our study has some limitations. Firstly, the cross-sectional design of the study restricts us to measure the causal effect. Second, due to the dichotomous response on alcohol consumption, we cannot measure the quantity of alcohol an individual consumed. Lastly, information on social groups and occupations could not be included in the analysis because of the large number of missing cases in these variables.

**Conclusion**

In Northeastern India, as this study illustrates, there are three classes of modifiable risk factors closely related to the risk of chronic diseases (diabetes and hypertension). In addition, the study also identified factors that uniquely distinguished the identified classes. The intake of aerated drinks and obesity are the common modifiable risk factors in all three classes. As an initiative to healthy lifestyle behaviours, the findings suggest that males should control smoking and alcohol consumption, while females should control body weight and blood pressure. Further, we found that marital status and gender could be the catalyst to prevent high-risk smoking and alcohol drinking in Northeastern India. Also, the study observes smoking and alcohol use are the two modifiable risk behaviours, which tends to co-occur within an individual. Therefore, it is recommended that policy and intervention programmes in Norteastern India that promote healthy lifestyles should focus on targeting multiple modifiable risk behaviours that are most likely to co-occur within an individual. On the other hand, awareness about adapting to healthy diets and weight control along with physical exercises should be promoted with rigour in the region.

**Acknowledgement:** The authors are thankful to the editor and the two reviewers for their insightful comments and valuable suggestions, which hugely improved the analysis and the overall quality of the paper.

**Contributors:** Conceived and designed the research paper: SPM, HL. Analysed the data: SPM, HSC. Contributed agents/materials/analysis tools: SPM, HL, HSC. Drafting manuscript: SPM, HSC. Refined the manuscript: SPM, HL, HSC.

**Funding:** This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

**Competing Interest:** None declared.

**Patient consent for publication:** Not applicable

**Ethics approval**: This study is based on the secondary data which is publicly available from the Demographic Health Survey website and the respondents are de-identified, hence ethical approval is not required for this study.

**Data availability statement**: Data stored and maintained by the Demographic Health Survey office, and available to public at https://dhsprogram.com/data/dataset/India_Standard-DHS_2015.cfm?flag=1 upon request.

**Word count**: 4464

**References**

1 Chobanian A V, Bakris GL, Black HR, *et al.* The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA* 2003;**289**:2560–72. doi:10.1001/jama.289.19.2560

2 International Diabetes Federation. Advocacy guide to the IDF Diabetes Atlas Ninth edition 2019. Brussels, Belgium:International Diabetes Federation: 2019. http://www.diabetesatlas.org/ (accessed 14 Nov 2019).

3 Ford ES, Zhao G, Tsai J, *et al.* Low-Risk Lifestyle Behaviors and All-Cause Mortality: Findings From the National Health and Nutrition Examination Survey III Mortality Study. *Am J Public Health* 2011;**101**:1922–9. doi:10.2105/AJPH.2011.300167

4 Geldsetzer P, Manne-Goehler J, Theilmann M, et al. Geographic and sociodemographic variation of cardiovascular disease risk in India: A cross-sectional study of 797,540 adults. PLOS Med 2018;15:e1002581. doi:10.1371/journal.pmed.1002581

5 Ghosh S, Kumar M. Prevalence and associated risk factors of hypertension among persons aged 15–49 in India: a cross-sectional study. BMJ Open 2019;9:e029714. doi:10.1136/bmjopen-2019-029714

6 Gupta R, Gaur K, S. Ram CV. Emerging trends in hypertension epidemiology in India. J Hum Hypertens 2019;33:575–87. doi:10.1038/s41371-018-0117-3

7 Tandon N, Anjana RM, Mohan V, et al. The increasing burden of diabetes and variations

among the states of India: the Global Burden of Disease Study 1990-2016. Lancet Glob Heal 2018;6:e1352–62. doi:10.1016/S2214-109X(18)30387-5

8 WHO. (2014). Global status report on noncommunicable diseases 2014. In World Health Organization.

9 Kvaavik, E. (2010). Influence of Individual and Combined Health Behaviors on Total and Cause-Specific Mortality in Men and Women. Archives of Internal Medicine, 170(8), 711. https://doi.org/10.1001/archinternmed.2010.76

10 Loef, M., & Walach, H. (2012). The combined effects of healthy lifestyle behaviors on all cause mortality: A systematic review and meta-analysis. Preventive Medicine, 55(3), 163–170. https://doi.org/10.1016/j.ypmed.2012.06.017

11 Ha S, Choi HR, Lee YH. Clustering of four major lifestyle risk factors among Korean adults with metabolic syndrome. *PLoS One* 2017;**12**:e0174567. doi:10.1371/journal.pone.0174567

12 Poortinga W. The prevalence and clustering of four major lifestyle risk factors in an English adult population. *Prev Med (Baltim)* 2007;**44**:124–8. doi:10.1016/j.ypmed.2006.10.006

13 Schuit AJ, van Loon AJM, Tijhuis M, *et al.* Clustering of Lifestyle Risk Factors in a General Adult Population. *Prev Med (Baltim)* 2002;**35**:219–24. doi:10.1006/pmed.2002.1064

14 Leventhal AM, Huh J, Dunton GF. Clustering of modifiable biobehavioral risk factors for chronic disease in US adults: a latent class analysis. Perspect Public Health 2014;134:331–8. doi:10.1177/1757913913495780

15 Lisha NE, Martens M, Leventhal AM. Age and gender as moderators of the relationship between physical activity and alcohol use. Addict Behav 2011;36:933–6. doi:10.1016/j.addbeh.2011.04.003.

16 Akbartabartoori M, Lean MEJ, Hankey CR. Relationships between cigarette smoking, body size and body shape. Int J Obes 2005;29:236–43. doi:10.1038/sj.ijo.0802827

17 Shaikh R, Khan J. Clustering of lifestyle risk factors among adult population in India: A cross-sectional analysis from 2005 to 2016. *PLoS One* 2021;**16**:e0244559. doi:10.1371/journal.pone.0244559

18 Atorkey P, Owiredua C. Clustering of multiple health risk behaviours and association with socio-demographic characteristics and psychological distress among adolescents in Ghana: A latent class analysis. SSM - Popul Heal 2021;13:100707. doi:10.1016/j.ssmph.2020.100707

19 Bischops AC, De Neve J-W, Awasthi A, *et al.* A cross-sectional study of cardiovascular disease risk clustering at different socio-geographic levels in India. *Nat Commun* 2020;**11**:5891. doi:10.1038/s41467-020-19647-3

20 Laxer RE, Brownson RC, Dubin JA, et al. Clustering of risk-related modifiable behaviours and their association with overweight and obesity among a large sample of youth in the COMPASS study. BMC Public Health 2017;17:102. Doi:10.1186/s12889-017-4034-0

21 Lee DH, Nam JY, Kwon S, et al. Lifestyle risk score and mortality in Korean adults: a population-based cohort study. Sci Rep 2020;10:10260. doi:10.1038/s41598-020-66742-y

22 Gao B, Zhang L, Wang H. Clustering of Major Cardiovascular Risk Factors and the Association with Unhealthy Lifestyles in the Chinese Adult Population. *PLoS One* 2013;**8**:e66780. doi:10.1371/journal.pone.0066780

23 Hart CL, Davey Smith G, Gruer L, et al. The combined effect of smoking tobacco and drinking alcohol on cause-specific mortality: A 30 year cohort study. BMC Public Health 2010;10. doi:10.1186/1471-2458-10-789

24 Sobngwi E, Mbanya J-CN, Unwin N, *et al.* Physical activity and its relationship with obesity, hypertension and diabetes in urban and rural Cameroon. *Int J Obes* 2002;**26**:1009–16. doi:10.1038/sj.ijo.0802008

25 Bani T Aeri DS. Risk Factors Associated with the Increasing Cardiovascular Diseases Prevalence in India: A Review. *J Nutr Food Sci* 2014;**05**:1–6. doi:10.4172/2155-9600.1000331

26 Marbaniang SP, Lhungdim H, Yadav B, *et al.* Overweight/obesity risks and prevalence of diabetes and hypertension in North Eastern India: An analysis using seemingly unrelated probit model. *Clin Epidemiol Glob Heal* 2021;**11**:100764. doi:10.1016/j.cegh.2021.100764

27 Umamahesh K, Vigneswari A, Surya Thejaswi G, *et al.* Incidence of cardiovascular diseases and associated risk factors among subjects with type 2 diabetes – An 11-year follow up study. *Indian Heart J* 2014;**66**:5–10. doi:10.1016/j.ihj.2013.12.009

28 Sarveswaran G, Kulothungan V, Mathur P. Clustering of noncommunicable disease risk factors among adults (18–69 years) in rural population, South-India. *Diabetes Metab Syndr Clin Res Rev* 2020;**14**:1005–14. doi:10.1016/j.dsx.2020.05.042

29 Schreiber JB. Latent Class Analysis: An example for reporting results. *Res Soc Adm Pharm* 2017;**13**:1196–201. doi:10.1016/j.sapharm.2016.11.011

30 Weller BE, Bowen NK, Faubert SJ. Latent Class Analysis: A Guide to Best Practice. *J Black Psychol* 2020;**46**:287–311. doi:10.1177/0095798420930932

31 Marbaniang SP. Women Care and Practices in the Management of Childhood Diarrhea in Northeast India. *Child Care Pract* 2020;:1–13. doi:10.1080/13575279.2020.1812534

32 Dikshit KR, Dikshit JK. Population of the North-Eastern States of India. In: *North-East India: Land, People and Economy. Advances in Asian Human-Environmental Research*. Dordrecht: : Springer 2014. 421–56. doi:10.1007/978-94-007-7055-3_12

33 International Institute for Population Sciences (IIPS) and ICF. National Family Health Survey (NFHS-4), 2015-16: India. Mumbai:IIPS: 2017. http://rchiips.org/nfhs/NFHS-4Reports/India.pdf

34 DHS. The DHS Program-India: Standard DHS, 2015-16 Dataset,. Demogr. Heal. Surv. 2017.https://dhsprogram.com/data/dataset/India_Standard-DHS_2015.cfm?flag=1

35 Barba C, Cavalli-Sforza T, Cutter J, *et al.* Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* 2004;**363**:157–63. doi:10.1016/S0140-6736(03)15268-3

36 Islam MM, Valderas JM, Yen L, *et al.* Multimorbidity and Comorbidity of Chronic Diseases among the Senior Australians: Prevalence and Patterns. *PLoS One* 2014;**9**:e83783. doi:10.1371/journal.pone.0083783

37 Brusco MJ, Shireman E, Steinley D. A comparison of latent class, K-means, and K-median methods for clustering dichotomous data. *Psychol Methods* 2017;**22**:563–80. doi:10.1037/met0000095

38 Dey A, Chakraborty A, Majumdar K, *et al.* Application of latent class analysis to estimate susceptibility to adverse health outcomes based on several risk factors. *Int J Community Med Public Heal* 2016;**3**:3423–9. doi:10.18203/2394-6040.ijcmph20164268

39 Gao F, Chen J, Liu X, *et al.* Latent class analysis suggests four classes of persons with type 2 diabetes mellitus based on complications and comorbidities in Tianjin, China: a cross-sectional analysis. *Endocr J* 2017;**64**:1007–16. doi:10.1507/endocrj.EJ17-0199

40 Ghanbari J, Mohammadpoorasl A, Jahangiry L, *et al.* Subgroups of lifestyle patterns among hypertension patients: a latent-class analysis. *BMC Med Res Methodol* 2018;**18**:127. doi:10.1186/s12874-018-0607-6

41 Sugiura N. Further analysts of the data by akaike' s information criterion and the finite corrections. *Commun Stat - Theory Methods* 1978;**7**:13–26. doi:10.1080/03610927808827599

42 Akaike H. A New Look at the Statistical Model Identification. 1974. 215–22. doi:10.1007/978-1-4612-1694-0_16

43 Linzer DA, Lewis JB. poLCA : An R Package for Polytomous Variable Latent Class Analysis. *J Stat Softw* 2011;**42**. doi:10.18637/jss.v042.i10

44 Bolck A, Croon M, Hagenaars J. Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Polit Anal* 2004;**12**:3–27. doi:10.1093/pan/mph001

45 Rosato NS, Baer JC. Latent class analysis: A method for capturing heterogeneity. Soc Work Res 2012;36:61–9. doi:10.1093/swr/svs006

46 Wang M-C, Deng Q, Bi X, et al. Performance of the entropy as an index of classification accuracy in latent profile analysis: A Monte Carlo simulation study. Acta Psychol Sin 2017;49:1473. doi:10.3724/SP.J.1041.2017.01473

47 Celeux G, Soromenho G. An entropy criterion for assessing the number of clusters in a mixture model. J Classif 1996;13:195–212. doi:10.1007/BF01246098

48 Omvik P. How smoking affects blood pressure. Blood Press 1996;5:71–7. doi:10.3109/08037059609062111

49 Jenner DA, Puddey IB, Beilin LJ, *et al.* Lifestyle- and occupation-related changes in blood pressure over a six-year period in a cohort of working men. *J Hypertens Suppl Off J Int Soc Hypertens* 1988;**6**:S605-7. doi:10.1097/00004872-198812040-00189

50 Gerace TA, Hollis J, Ockene JK, *et al.* Smoking cessation and change in diastolic blood pressure, body weight, and plasma lipids. MRFIT Research Group. *Prev Med (Baltim)* 1991;**20**:602–20. doi:10.1016/0091-7435(91)90058-c

51 Green MS, Jucha E, Luz Y. Blood pressure in smokers and nonsmokers: Epidemiologic findings. Am Heart J 1986;111:932–40. doi:10.1016/0002-8703(86)90645-9

52 Ho RKS, Fok PWY, Chan HCH. Pattern and determinants of alcohol and tobacco co-use and its relationship with smoking cessation in Hong Kong. *Tob Prev \& Cessat* 2021;**7**:1–9. doi:10.18332/tpc/132288

53 Jiang N, Lee YO, Ling PM. Association between tobacco and alcohol use among young adult bar patrons: a cross-sectional study in three cities. *BMC Public Health* 2014;**14**:500. doi:10.1186/1471-2458-14-500

54 Slutske WS. Common Genetic Vulnerability for Pathological Gambling and Alcohol Dependence in Men. Arch Gen Psychiatry 2000;57:666–73. doi:10.1001/archpsyc.57.7.666

55 Bierut LJ, Schuckit MA, Hesselbrock V, et al. Co-occurring risk factors for alcohol dependence and habitual smoking. Alcohol Res Health 2000;24:233–41.http://www.ncbi.nlm.nih.gov/pubmed/15986718

56 Grant BF, Hasin DS, Chou SP, et al. Nicotine Dependence and Psychiatric Disorders in the United States. Arch Gen Psychiatry 2004;61:1107. doi:10.1001/archpsyc.61.11.1107

57 Lyon RJ, Tong JE, Leigh G, et al. The influence of alcohol and tobacco on the components of choice reaction time. J Stud Alcohol 1975;36:587–96. doi:10.15288/jsa.1975.36.587

58 de Fiebre CM, Collins AC. A comparison of the development of tolerance to ethanol and cross-tolerance to nicotine after chronic ethanol treatment in long- and short-sleep mice. J Pharmacol Exp Ther 1993;266:1398–406.

59 Malik VS, Popkin BM, Bray GA, *et al.* Sugar-sweetened beverages, obesity, type 2 diabetes mellitus, and cardiovascular disease risk. *Circulation* 2010;**121**:1356–64. doi:10.1161/CIRCULATIONAHA.109.876185

60 Qin P, Li Q, Zhao Y, *et al.* Sugar and artificially sweetened beverages and risk of obesity, type 2 diabetes mellitus, hypertension, and all-cause mortality: a dose–response meta-analysis of prospective cohort studies. *Eur J Epidemiol* 2020;**35**:655–71. doi:10.1007/s10654-020-00655-y

61 Liebman M, Pelican S, Moore SA, et al. Dietary intake, eating behavior, and physical activity-related determinants of high body mass index in rural communities in Wyoming, Montana, and Idaho. Int J Obes 2003;27:684–92. doi:10.1038/sj.ijo.0802277

62 French S, Jeffery R, Forster J, et al. Predictors of weight change over two years among a population of working adults: the Healthy Worker Project. Int J Obes Relat Metab Disord 1994;18 3:145–54.

63 Nguyen S, Choi HK, Lustig RH, et al. Sugar-Sweetened Beverages, Serum Uric Acid, and Blood Pressure in Adolescents. J Pediatr 2009;154:807–13. doi:10.1016/j.jpeds.2009.01.015

64 Jalal DI, Smits G, Johnson RJ, et al. Increased Fructose Associates with Elevated Blood Pressure. J Am Soc Nephrol 2010;21:1543–9. doi:10.1681/ASN.2009111111

65 Cohen L, Curhan G, Forman J. Association of sweetened beverage intake with incident hypertension. J Gen Intern Med 2012;27:1127–34. Doi:10.1007/s11606-012-2069-6

66 Saikia N, Debbarma B. The socioeconomic correlates of substance use among male adults in Northeast India. *Clin Epidemiol Glob Heal* 2020;**8**:149–57. doi:10.1016/j.cegh.2019.06.004

67 IOM (Institute of Medicine). Patterns of Tobacco Use by Adolescents and Young Adults. In: Richard J. Bonnie, Kathleen Stratton LYK, ed. *Public Health Implications of Raising the Minimum Age of Legal Access to Tobacco Products.* Washington, D.C.: : The National Academies Press 2015. https://www.ncbi.nlm.nih.gov/books/NBK310412/pdf/Bookshelf_NBK310412.pdf

68 Maggs JL, Schulenberg JE. Initiation and Course of Alcohol Consumption among

Adolescents and Young Adults. In: *Recent Developments in Alcoholism*. New York: : Kluwer Academic Publishers-Plenum Publishers 2005. 29–47. doi:10.1007/0-306-48626-1_2

69 Investigators S. Effect of the Age at Which Smoking Begins on Frequency of Smoking in Adulthood. *N Engl J Med* 1991;**325**:968–9. doi:10.1056/NEJM199109263251318

70 Dawson DA, Goldstein RB, Patricia Chou S, *et al.* Age at First Drink and the First Incidence of Adult-Onset DSM-IV Alcohol Use Disorders. *Alcohol Clin Exp Res* 2008;**32**:2149–60. doi:10.1111/j.1530-0277.2008.00806.x

71 Tanihara S, Momose Y. Reasons for smoking cessation attempts among Japanese male smokers vary by nicotine dependence level: a cross-sectional study after the 2010 tobacco tax increase. *BMJ Open* 2015;**5**:e006658–e006658. doi:10.1136/bmjopen-2014-006658

72 Freitas SRS, Alvim RO. Smoking and blood pressure phenotypes: New perspective for an old problem. *Am J Hypertens* 2017;**30**:554–5. doi:10.1093/ajh/hpx039

73 Saikia B, Marbaniang SP, Kumar P, *et al.* Changing pattern of tobacco consumption and quitting behavior in Northeast India. *J Subst Use* 2021;:1–12. doi:10.1080/14659891.2021.1875068

74 Wilsnack RW, Wilsnack SC, Kristjanson AF, *et al.* Gender and alcohol consumption: patterns from the multinational GENACIS project. *Addiction* 2009;**104**:1487–500. doi:10.1111/j.1360-0443.2009.02696.x

75 Nolen-Hoeksema S. Gender differences in risk factors and consequences for alcohol use and problems. *Clin Psychol Rev* 2004;**24**:981–1010. doi:10.1016/j.cpr.2004.08.003

76 Rasiah R, Yusoff K, Mohammadreza A, *et al.* Cardiovascular disease risk factors and socioeconomic variables in a nation undergoing epidemiologic transition. *BMC Public Health* 2013;**13**:886. doi:10.1186/1471-2458-13-886

77 Winkleby MA, Jatulis DE, Frank E, *et al.* Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. *Am J Public Health* 1992;**82**:816–20. doi:10.2105/AJPH.82.6.816

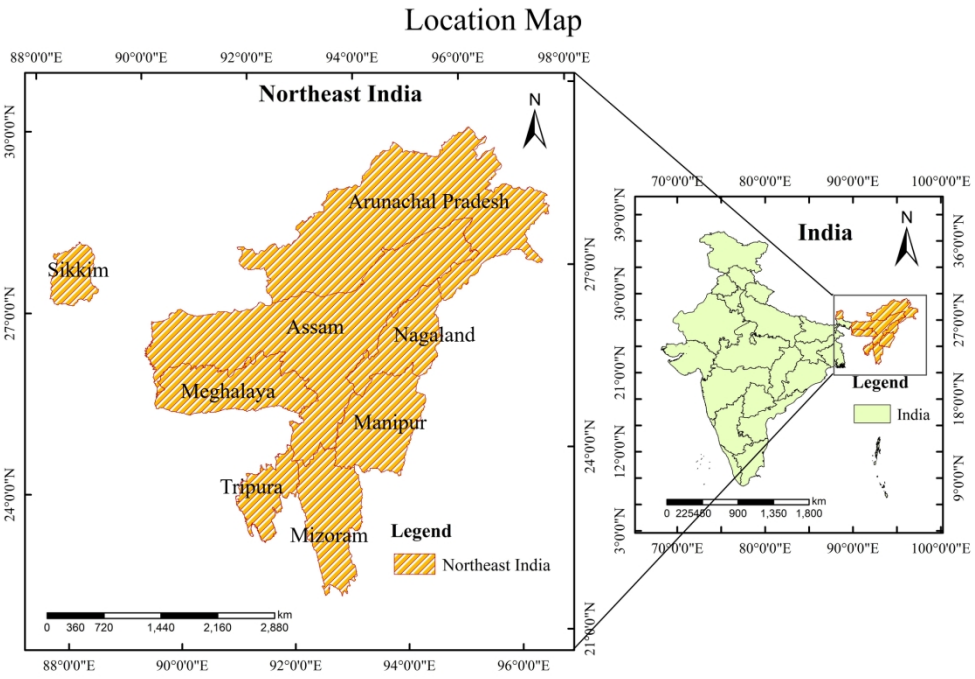
78 Menasco M, Blair S. Gender Differences in Substance Use Across Marital Statuses. *Int J Criminol Sociol* 2016;**5**:1–13. doi:10.6000/1929-4409.2016.05.01

79 Power C, Rodgers B, Hope S. Heavy alcohol consumption and marital status: disentangling the relationship in a national study of young adults. *Addiction* 1999;**94**:1477–87. doi:10.1046/j.1360-0443.1999.941014774.x

Figure 1: ***Map showing the location of the Study area. Source:*** *Authors*

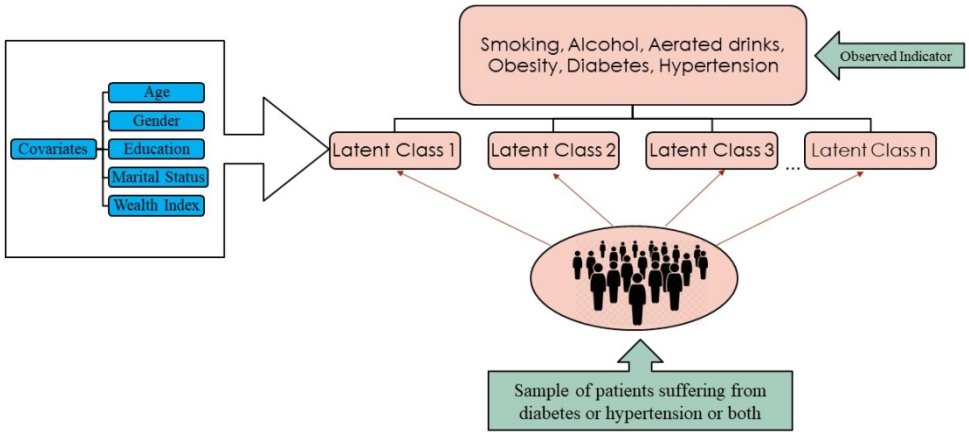
Figure 2: ***Conceptual framework of the study. Source:*** *Authors*

Figure 3: **Item response conditional probability across the latent classes. *Source:*** *Authors*



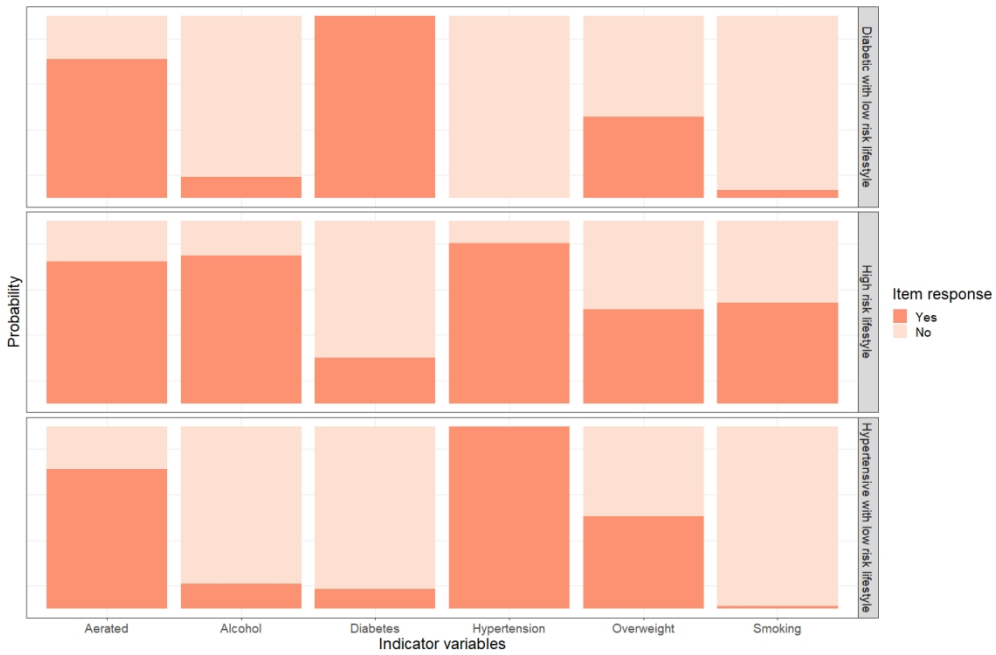

Map showing the location of the Study area. Source: Authors

595x412mm (96 x 96 DPI)




Conceptual framework of the study. Source: Authors

108x60mm (300 x 300 DPI)




Item response conditional probability across the latent classes. Source: Authors

129x84mm (300 x 300 DPI)

**STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies***

| Section/Topic | Item # | Recommendation | Reported on page # |
|---|---|---|---|
| **Title and abstract** | 1 | (*a*) Indicate the study's design with a commonly used term in the title or the abstract | 1 |
| | | (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found | 2 |
| **Introduction** | | | |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported | 4 |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses | 6 |
| **Methods** | | | |
| Study design | 4 | Present key elements of study design early in the paper | 6 |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | 6 |
| Participants | 6 | (*a*) Give the eligibility criteria, and the sources and methods of selection of participants | 7 |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | 8 |
| Data sources/ measurement | 8* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | 6 |
| Bias | 9 | Describe any efforts to address potential sources of bias | Not applicable |
| Study size | 10 | Explain how the study size was arrived at | 7 |
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | Not applicable |
| Statistical methods | 12 | (*a*) Describe all statistical methods, including those used to control for confounding | 8 |
| | | (*b*) Describe any methods used to examine subgroups and interactions | Not applicable |
| | | (*c*) Explain how missing data were addressed | Not applicable |
| | | (*d*) If applicable, describe analytical methods taking account of sampling strategy | Not applicable |
| | | (*e*) Describe any sensitivity analyses | Not applicable |
| **Results** | | | |

| | | | |
|---|---|---|---|
| Participants | 13* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed | 9 |
| | | (b) Give reasons for non-participation at each stage | Not applicable |
| | | (c) Consider use of a flow diagram | 9 |
| Descriptive data | 14* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders | 10 |
| | | (b) Indicate number of participants with missing data for each variable of interest | Not applicable |
| Outcome data | 15* | Report numbers of outcome events or summary measures | 11 |
| Main results | 16 | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included | Not applicable |
| | | (*b*) Report category boundaries when continuous variables were categorized | Not applicable |
| | | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | Not applicable |
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses | 12-13 |
| **Discussion** | | | |
| Key results | 18 | Summarise key results with reference to study objectives | 14 |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias | 17 |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence | 14-17 |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results | 17 |
| **Other information** | | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | 18 |

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.