Clustering of health-related behavior patterns using latent class analysis: Results from a working population of 12,168 Japanese at risk for noncommunicable diseases

Takahiro Miki (tkhr.mk@gmail.com)  
PREVENT Inc.

Kojiro Yamamoto  
PREVENT Inc.

Masashi Kanai  
Konan Women's University

Kento Takeyama  
PREVENT Inc.

Maki Iwatake  
PREVENT Inc.

Yuta Hagiwara  
PREVENT Inc.

Research Article

Keywords: Noncommunicable Diseases, Health behavior, Latent class analysis,

Posted Date: October 26th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2170299/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background:

Noncommunicable diseases (NCDs) have become a significant global problem. Since health behaviors are associated with NCDs, and it is more effective to characterize populations using a public health population approach and provide specific interventions according to those characteristics, it is useful to examine the formation of clusters of health behavior combinations. Therefore, this study is intended to conduct clustering analysis by health behavior of Japanese people in the working population who are at risk of NCDs, through the use of latent class analysis.

Methods:

Participants were individuals who had not previously been diagnosed with NCDs but were at risk for NCDs. Participants were assessed on the basis of several demographic data and several health behaviors. The clustering study used latent class analysis (LCA) based on health behavior characteristics for individuals at risk of NCDs among the Japanese working population. All statistical analyses were conducted using R (Version 4.0.4) and the “poLCA” package (Version 1.6.0).

Results:

This study’s participants was 12,168. Regarding LCA, the study compared models with one to six latent classes. The five classes were closest to the most appropriate model because of the Bayesian Information Criterion, Akaike Information Criterion, and G^2 values and the more distinguishable cluster characteristics. Cluster 1 was “having healthy lifestyles but disliking hospitals”; Cluster 2 was “women with healthy lifestyle behaviors”; Cluster 3 was “general population”; and Cluster 4 was “middle-aged group in need of lifestyle improvement.” Cluster 5 was “a group of treatment for lifestyle-related diseases.”

Conclusions:

This study reveals more discernible health behavior patterns in a sample of the Japanese population using large real-world data. This may suggest the effectiveness of each distinctive approach in considering a population approach to public health.

Background

Noncommunicable diseases (NCDs), commonly known as "lifestyle diseases," such as diabetes, hypertension, and heart disease, have become more frequent in recent years and a significant global problem [1]. It is estimated that approximately one billion people worldwide suffer from hypertension and that 60% of adults will suffer from hypertension by 2050 [2]. The number of diabetics in the United States is projected to rise to 48.3 million by 2050 [3]. In India, 10% of adults suffer from hypertension, and 25 to 30 million people have diabetes. Three out of every 1,000 people suffer from stroke. Heart attack deaths
are projected to rise from 1.2 million to 2 million by 2010 [4]. The risk of stroke is also high in Japan. Mortality rates were higher in Japan than in the West, although the Japanese mortality rate declined from one-third to one-fourth between 1970 and 2005. Heart disease is also the second most common cause of death among the Japanese population [5].

The accumulation of poor health behaviors causes NCDs. Health behaviors are health-related behavioral habits, such as frequency of alcohol consumption, smoking, and exercise. These poor health behaviors are associated with poor health status, including early mortality [6,7]. For example, it is widely known that smoking is a risk factor for coronary heart disease [8]. The frequency of alcohol consumption affects the incidence of many diseases, including cancer, neuropsychiatric disorders, and many cardiovascular and gastrointestinal infections [9]. Health behavior is a widely used proxy for proneness to NCDs [10,11]. A combination of unhealthy habits, rather than a single habit, is a risk factor. Furthermore, the combinations often exhibit specific common patterns [12,13]. For example, cigarette smokers drink large amounts of alcohol [14,15]. Therefore, it is crucial for epidemiologists to investigate the formation of common patterns and clustering of combinations of health behaviors [16,17]. Especially in public health population approaches, it is more effective to understand the population's characteristics and provide specific interventions that meet those characteristics.

Many recent studies have examined clustering based on health behavior for various countries’ populations [12,16,18–20]. For example, a study in Ireland reported that several clusters can be formed according to health behaviors and that such clusters are similar to those of other European populations [16]. Despite some differences according to age and gender in the UK study, they formed a clustering with some consistent characteristics [21]. However, these were relatively small sample sizes or studies focused on Europe and the United States. It is not appropriate to generalize these results to different countries, including Japan, since different countries have different cultures and lifestyles. In addition, study samples have included a broad general population, including healthy individuals or those who already have severe NCDs. Since a vital perspective concerning NCDs includes early prevention, it would be beneficial to examine the characteristics of multiple health behaviors in a working population at risk for NCDs before they become diagnosed [22,23]. More studies are needed to examine clusters based on health behavior characteristics in the Japanese population. In particular, although most work environments mandate regular annual health screening, no studies have been conducted on approaches to secondary prevention for those in the workforce who require treatment above the reference values.

Therefore, this study is intended to conduct clustering analysis by health behavior of Japanese people in the working population who are at risk of NCDs to understand the characteristics of each subgroup.

**Method**

**Study Design and Participants**

This study was a cross-sectional study. Participants were selected based on data collected by PREVENT Inc. (Nagoya, Japan), which conducted medical data analyses between April 2013 and July 2021.
Included were individuals who had not previously been diagnosed with NCDs and who met one of the following criteria.

- Blood pressure: systolic blood pressure 160 mmHg or higher and/or diastolic blood pressure 100 mmHg or higher
- Lipid: Low-Density Lipoprotein (LDL) cholesterol level 180 mg/dL or higher and/or High-Density Lipoprotein (HDL) cholesterol level 30 mg/dL or lower and/or triglycerides 500 mg/dL or higher
- Blood glucose: fasting blood glucose 140 mg/dL or higher and/or Hemoglobin ba1c (Hba1c) 7.0% or higher

These reference values are based on the values recommended by the national government and various academic societies in Japan [24–26].

Measurements of Demographic Characteristics

Participant demographics included age, gender, and BMI. In addition, triglycerides, HDL cholesterol, LDL cholesterol, fasting blood glucose, Hba1c, and blood pressure (systolic/diastolic) were included.

Measurements of Health-Related Behaviors

The underlying latent variable in the study is health-related behaviors. Health behaviors described in the following were evaluated.

**Exercise Habits**

Exercise habits were ascertained by asking whether the respondents exercised at least twice a week for at least 30 minutes. The participants were asked to choose between “Yes” and “No.”

**Smoking**

The participants were asked if they were current smokers. They were asked to choose between “Yes” and “No.”

**Alcohol Consumption**

For alcohol consumption, participants were asked to choose from the options “daily,” “some time,” and “rarely or cannot.”

**Healthy Eating Habits**

Regarding healthy eating habits, those who ate dinner within two hours of bedtime at least three times a week were asked to select “Yes,” and those who ate less than that were asked to select “No.”

**Willingness to Improve Lifestyle**
Participants were asked to select “Yes” if they wanted to improve their lifestyle habits and “No” if they did not.

**Number of Hospital Visits within the Last Year**

The participants were asked to select the number of visits to a hospital within the last year from “fewer than 3,” “3 to 5,” “6 to 8,” “9 to 11,” and “12 or more.”

**Habits of Taking Medications within the Last Year**

The participants were asked to answer “Yes” or “No” to whether or not they had taken any medications within the last year.

**History of Vascular Diseases**

The participants were asked to answer “Yes” or “No” to whether they had a history of vascular disease.

**Statistical Analysis**

Descriptive statistical analysis was conducted to describe and interpret the different classes. This study was analyzed using latent class analysis (LCA) to identify clusters based on the participants' health behaviors. LCA is a statistical method for finding latent classes of related cases from multivariate categorical data. LCA differs from cluster analysis and factor analysis because it allows for the inclusion of discrete and dichotomous variables [27]. It is an appropriate method when variables are categorical rather than continuous. Furthermore, this method is flexible, and the choice of cluster criteria is less arbitrary [28].

LCA is considered more informative for describing health behaviors [29]. This is because, according to one study, it is more appropriate for analyzing health behaviors as categorical variables [30]. Concerning variable selection, in addition to health behaviors, age and gender were added to the variables in the study because they are widely known to be important factors [31].

The model selection criteria were based on the Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and G^2 [32]. For AIC and BIC, smaller values indicate a better fit of the model. BIC is the most reliable fit statistic and should be routinely reported [33,34]. Another study reported that the preferred model minimizes BIC and/or AIC values [35]. The goal is also to select a model that minimizes G^2 without estimating an excessive number of parameters [35]. In addition, the number of models in the cluster was determined by taking into account the actual practical interpretation [36].

All statistical analyses were conducted using R (Version 4.0.4). In addition, the LCA was performed using the “poLCA” package (Version 1.6.0). Under the assumption that missing data were missing at random, the likelihood was calculated, including missing data based on the previous study [35].
Results

Participant Characteristics

This study's participants was 12,168. Table 1 summarizes the characteristics of the participants in the study.
<table>
<thead>
<tr>
<th>Variables</th>
<th>Overall</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
<th>Cluster 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>12,168</td>
<td>2,389</td>
<td>2,215</td>
<td>4,395</td>
<td>2,098</td>
<td>1,071</td>
</tr>
<tr>
<td>sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>man</td>
<td>9,703 (80%)</td>
<td>2,348 (98%)</td>
<td>27 (1.2%)</td>
<td>4,350 (99%)</td>
<td>2,031 (97%)</td>
<td>947 (88%)</td>
</tr>
<tr>
<td>Woman</td>
<td>2,465 (20%)</td>
<td>41 (1.7%)</td>
<td>2,188 (99%)</td>
<td>45 (1.0%)</td>
<td>67 (3.2%)</td>
<td>124 (12%)</td>
</tr>
<tr>
<td>BMI</td>
<td>25.0 (3.9)</td>
<td>24.6 (3.3)</td>
<td>23.9 (4.2)</td>
<td>25.9 (4.0)</td>
<td>24.6 (3.4)</td>
<td>25.5 (3.8)</td>
</tr>
<tr>
<td>age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;30</td>
<td>186 (1.5%)</td>
<td>0 (0%)</td>
<td>42 (1.9%)</td>
<td>126 (2.9%)</td>
<td>15 (0.7%)</td>
<td>3 (0.3%)</td>
</tr>
<tr>
<td>30 ~ 39</td>
<td>1,098 (9.0%)</td>
<td>0 (0%)</td>
<td>152 (6.9%)</td>
<td>725 (16%)</td>
<td>201 (9.6%)</td>
<td>20 (1.9%)</td>
</tr>
<tr>
<td>40 ~ 49</td>
<td>4,726 (39%)</td>
<td>21 (0.9%)</td>
<td>636 (29%)</td>
<td>2,685 (61%)</td>
<td>1,110 (53%)</td>
<td>274 (26%)</td>
</tr>
<tr>
<td>50 ~ 59</td>
<td>4,784 (39%)</td>
<td>1,508 (63%)</td>
<td>1,117 (50%)</td>
<td>859 (20%)</td>
<td>737 (35%)</td>
<td>563 (53%)</td>
</tr>
<tr>
<td>60over</td>
<td>1,374 (11%)</td>
<td>860 (36%)</td>
<td>268 (12%)</td>
<td>0 (0%)</td>
<td>35 (1.7%)</td>
<td>211 (20%)</td>
</tr>
<tr>
<td>Neutral fat</td>
<td>185 (197)</td>
<td>174 (183)</td>
<td>120 (95)</td>
<td>193 (195)</td>
<td>238 (265)</td>
<td>201 (192)</td>
</tr>
<tr>
<td>(missing)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HDL</td>
<td>58 (17)</td>
<td>58 (17)</td>
<td>68 (18)</td>
<td>52 (15)</td>
<td>60 (17)</td>
<td>57 (17)</td>
</tr>
<tr>
<td>(missing)</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LDL</td>
<td>157 (44)</td>
<td>147 (43)</td>
<td>173 (38)</td>
<td>163 (42)</td>
<td>144 (46)</td>
<td>151 (45)</td>
</tr>
<tr>
<td>(missing)</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LDL/HDL ration</td>
<td>2.91 (1.04)</td>
<td>2.71 (1.03)</td>
<td>2.70 (0.90)</td>
<td>3.31 (1.01)</td>
<td>2.54 (0.97)</td>
<td>2.85 (1.03)</td>
</tr>
<tr>
<td>(missing)</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fasting blood glucose</td>
<td>106 (29)</td>
<td>115 (35)</td>
<td>99 (22)</td>
<td>105 (29)</td>
<td>105 (26)</td>
<td>110 (28)</td>
</tr>
<tr>
<td>Variables</td>
<td>Overall</td>
<td>Cluster 1</td>
<td>Cluster 2</td>
<td>Cluster 3</td>
<td>Cluster 4</td>
<td>Cluster 5</td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>(missing)</td>
<td>1,154</td>
<td>207</td>
<td>211</td>
<td>421</td>
<td>221</td>
<td>94</td>
</tr>
<tr>
<td>Hba1c</td>
<td>5.86 (1.09)</td>
<td>6.10 (1.12)</td>
<td>5.77 (0.94)</td>
<td>5.85 (1.01)</td>
<td>5.69 (1.39)</td>
<td>5.93 (0.87)</td>
</tr>
<tr>
<td>(missing)</td>
<td>1,004</td>
<td>217</td>
<td>188</td>
<td>322</td>
<td>166</td>
<td>111</td>
</tr>
<tr>
<td>SBP</td>
<td>133 (21)</td>
<td>138 (21)</td>
<td>129 (25)</td>
<td>130 (19)</td>
<td>138 (20)</td>
<td>134 (19)</td>
</tr>
<tr>
<td>DBP</td>
<td>85 (15)</td>
<td>87 (14)</td>
<td>79 (16)</td>
<td>84 (15)</td>
<td>90 (15)</td>
<td>86 (14)</td>
</tr>
<tr>
<td>exercise habits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>9,098 (78%)</td>
<td>1,452 (64%)</td>
<td>1,749 (82%)</td>
<td>3,499 (83%)</td>
<td>1,616 (78%)</td>
<td>782 (75%)</td>
</tr>
<tr>
<td>yes</td>
<td>2,627 (22%)</td>
<td>816 (36%)</td>
<td>377 (18%)</td>
<td>717 (17%)</td>
<td>460 (22%)</td>
<td>257 (25%)</td>
</tr>
<tr>
<td>(missing)</td>
<td>5,557</td>
<td>1,219</td>
<td>1,112</td>
<td>1,929</td>
<td>751</td>
<td>546</td>
</tr>
<tr>
<td>Healthy eating habits</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>7,229 (62%)</td>
<td>2,030 (90%)</td>
<td>1,773 (84%)</td>
<td>2,121 (50%)</td>
<td>632 (30%)</td>
<td>673 (65%)</td>
</tr>
<tr>
<td>yes</td>
<td>4,460 (38%)</td>
<td>225 (10.0%)</td>
<td>348 (16%)</td>
<td>2,085 (50%)</td>
<td>1,441 (70%)</td>
<td>361 (35%)</td>
</tr>
<tr>
<td>(missing)</td>
<td>479</td>
<td>134</td>
<td>94</td>
<td>189</td>
<td>25</td>
<td>37</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>everyday</td>
<td>3,301 (28%)</td>
<td>794 (35%)</td>
<td>90 (4.2%)</td>
<td>0 (0%)</td>
<td>2,083 (100%)</td>
<td>334 (32%)</td>
</tr>
<tr>
<td>sometimes</td>
<td>4,017 (34%)</td>
<td>799 (35%)</td>
<td>596 (28%)</td>
<td>2,275 (54%)</td>
<td>0 (0%)</td>
<td>347 (33%)</td>
</tr>
<tr>
<td>rarely,cannot</td>
<td>4,450 (38%)</td>
<td>675 (30%)</td>
<td>1,450 (68%)</td>
<td>1,963 (46%)</td>
<td>0 (0%)</td>
<td>362 (35%)</td>
</tr>
<tr>
<td>(missing)</td>
<td>400</td>
<td>121</td>
<td>79</td>
<td>157</td>
<td>15</td>
<td>28</td>
</tr>
<tr>
<td>Willingness to improve lifestyle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>2,729 (23%)</td>
<td>629 (28%)</td>
<td>311 (15%)</td>
<td>952 (23%)</td>
<td>637 (31%)</td>
<td>200 (19%)</td>
</tr>
<tr>
<td>yes</td>
<td>9,010 (77%)</td>
<td>1,643 (72%)</td>
<td>1,818 (85%)</td>
<td>3,269 (77%)</td>
<td>1,441 (69%)</td>
<td>839 (81%)</td>
</tr>
<tr>
<td>(missing)</td>
<td>429</td>
<td>117</td>
<td>86</td>
<td>174</td>
<td>20</td>
<td>32</td>
</tr>
<tr>
<td>Variables</td>
<td>Overall</td>
<td>Cluster 1</td>
<td>Cluster 2</td>
<td>Cluster 3</td>
<td>Cluster 4</td>
<td>Cluster 5</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>---------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>8,171 (68%)</td>
<td>1,600 (68%)</td>
<td>2,083 (95%)</td>
<td>2,646 (62%)</td>
<td>1,066 (51%)</td>
<td>776 (73%)</td>
</tr>
<tr>
<td>yes</td>
<td>3,801 (32%)</td>
<td>737 (32%)</td>
<td>103 (4.7%)</td>
<td>1,652 (38%)</td>
<td>1,026 (49%)</td>
<td>283 (27%)</td>
</tr>
<tr>
<td>(missing)</td>
<td>196</td>
<td>52</td>
<td>29</td>
<td>97</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Number of hospital visits within the last year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>5,076 (42%)</td>
<td>1,761 (74%)</td>
<td>775 (35%)</td>
<td>1,749 (40%)</td>
<td>791 (38%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>&lt;= 3</td>
<td>3,977 (33%)</td>
<td>464 (19%)</td>
<td>692 (31%)</td>
<td>1,804 (41%)</td>
<td>953 (45%)</td>
<td>64 (6.0%)</td>
</tr>
<tr>
<td>&lt;= 6</td>
<td>1,641 (13%)</td>
<td>164 (6.9%)</td>
<td>432 (20%)</td>
<td>559 (13%)</td>
<td>257 (12%)</td>
<td>229 (21%)</td>
</tr>
<tr>
<td>&lt;= 9</td>
<td>780 (6.4%)</td>
<td>0 (0%)</td>
<td>190 (8.6%)</td>
<td>153 (3.5%)</td>
<td>81 (3.9%)</td>
<td>356 (33%)</td>
</tr>
<tr>
<td>&lt;= 12</td>
<td>694 (5.7%)</td>
<td>0 (0%)</td>
<td>126 (5.7%)</td>
<td>130 (3.0%)</td>
<td>16 (0.8%)</td>
<td>422 (39%)</td>
</tr>
<tr>
<td>Habits of taking medications within the last year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>11,601 (95%)</td>
<td>2,389 (100%)</td>
<td>2,202 (99%)</td>
<td>4,395 (100%)</td>
<td>2,098 (100%)</td>
<td>517 (48%)</td>
</tr>
<tr>
<td>yes</td>
<td>567 (4.7%)</td>
<td>0 (0%)</td>
<td>13 (0.6%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>554 (52%)</td>
</tr>
<tr>
<td>History of vascular disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>11,511 (95%)</td>
<td>2,371 (99%)</td>
<td>2,094 (95%)</td>
<td>4,275 (97%)</td>
<td>2,018 (96%)</td>
<td>753 (70%)</td>
</tr>
<tr>
<td>yes</td>
<td>657 (5.4%)</td>
<td>18 (0.8%)</td>
<td>121 (5.5%)</td>
<td>120 (2.7%)</td>
<td>80 (3.8%)</td>
<td>318 (30%)</td>
</tr>
</tbody>
</table>

\(^1\) n (%) or Mean (SD)

Abbreviation: BMI, body mass index; HDL, high-density lipoprotein; LDL, low-density lipoprotein; SBP, systolic blood pressure; DBP, diastolic blood pressure;
The study compared models with one to six latent classes. The results of this study are provided in Table 2. We determined that classes 5 and 6 have smaller values of BIC, AIC, and G^2, thus making them much closer to the most appropriate model. Furthermore, after careful consideration of both the 5-class and 6-class models, the 5-class model was selected because it more clearly distinguishes between the characteristics of the groups.

Table 2
Model fit for multiple models

<table>
<thead>
<tr>
<th>Cluster model</th>
<th>bic</th>
<th>aic</th>
<th>G^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>166825.8745</td>
<td>166699.963</td>
<td>9786.17619</td>
</tr>
<tr>
<td>2</td>
<td>163982.1129</td>
<td>163722.883</td>
<td>6889.67201</td>
</tr>
<tr>
<td>3</td>
<td>162519.6342</td>
<td>162127.086</td>
<td>5360.66962</td>
</tr>
<tr>
<td>4</td>
<td>162050.0189</td>
<td>161524.153</td>
<td>4740.9934</td>
</tr>
<tr>
<td>5</td>
<td>161917.9136</td>
<td>161258.729</td>
<td>4445.74064</td>
</tr>
<tr>
<td>6</td>
<td>161927.7694</td>
<td>161135.267</td>
<td>4280.28902</td>
</tr>
</tbody>
</table>

Bic, ayesian Information Criterion; aic, Akaike Information Criterion

**Description of the Identified Groups from the Latent Class Analysis**

Each cluster was created based on the participants’ demographic and health behavior characteristics. Their clusters are shown in Fig. 1. The characteristics of each of the five clusters are also described in the following.

**Cluster 1**

Cluster 1 included 2,389 participants. They tend to have healthy lifestyles, exercise, and not eat late at night. However, 74% (1,749 persons) had not visited a hospital in the previous year, indicating reluctance to visit hospitals. Hence, the label “having healthy lifestyles but disliking hospitals” was used.

**Cluster 2**

Cluster 2 had 2,215 participants. Ninety-five percent (n = 2,083) did not smoke, and 68% (n = 1,450) drank little or not at all. As in Cluster 1, the frequency of having dinner late in the evening was low. This group was the only one with a high percentage of women: 99% (4,350). Therefore, this group was labeled “women with healthy lifestyle behaviors.”

**Cluster 3**
Cluster 3 had the largest number of participants (4,395). They were generally characterized by relative youth, infrequent drinking, and infrequent hospital visits. Therefore, this group was labeled “general population.”

Cluster 4

Cluster 4 had 2,098 participants. They were middle-aged, drank alcohol daily, and had dinner frequently at night and half of them, 1,026 (49%), smoked. On the basis of these characteristics, we named this group the “middle-aged group in need of lifestyle improvement.”

Cluster 5

Cluster 5 consisted of 1,071 participants, of whom 774 (73%) were 50 years of age or older (“the older age group”). The percentage of participants taking lifestyle-related medications was as high as 53% (554), and the percentage of participants with an actual history of lifestyle-related diseases was also high at 30% (318). In other words, this group was an older group with more frequent hospital visits and lifestyle-related diseases. Hence, this group was labeled “a group of treatment for lifestyle-related diseases.”

Discussion

The study was conducted as clustering using LCA based on health behavior characteristics for individuals at risk of NCDs in the Japanese working population. Cluster 1 was “having healthy lifestyles but disliking hospitals,” Cluster 2 was “Women with healthy lifestyle behaviors,” Cluster 3 was “general population,” and Cluster 4 was “middle-aged group in need of lifestyle improvement.” Cluster 5 was “a group of treatment for lifestyle-related diseases.”

The clustering was at the extremes, with clusters 1 and 2 being healthy clusters and clusters 4 and 5 being unhealthy features.

Healthy clusters were observed in studies covering a variety of countries [13,15,17]. The formation of clusters of healthy people was typical in various population samples. Consider clusters 4 and 5, which have multiple unhealthy behaviors. This suggests that unhealthy behaviors tend to overlap. For example, a person with a drinking habit is likely to have an unhealthy diet. This clustering of multiple unhealthy behaviors was also observed in studies of other countries [12,13].

One of the characteristics that separates the healthy groups, Cluster 1 and Cluster 2, is “the number of visits to a hospital.” “The number of visits to a hospital” was also the key variable separating 4 and 5. Furthermore, the variable “the number of visits to a hospital” distinguishes Cluster 1 from Cluster 5. These results suggest that “the number of visits to a hospital” may be one of the key health behavior variables in creating clusters for this study’s sample. One study reported that the frequency of hospital visits tended to be higher among older people [37]. Cluster 5 was the older age group, which suggests that age may be related to the frequency of hospital visits.
Gender was a valuable variable for forming clusters, such as “women with healthy lifestyle behaviors” in Cluster 2. The current result suggests that gender may be necessary for forming clusters. In fact, a study in the United Kingdom reported that men were more likely to form unhealthy clusters than women [13].

Regarding age, there were entire age groups in the healthy clusters and the unhealthy clusters. Thus, age was not a characteristic variable in forming clusters in the study. Studies have indicated that age is not a variable that characterizes clusters [13,31]. The same results were obtained for those individuals at risk of NCDs in the Japanese working population for the study, suggesting that age was an important factor even for the different population targets. The previous study also concluded that alcohol consumption and smoking were important factors in forming the clusters [31]. Alcohol consumption and smoking were also factors in this study as variables that classified clusters 2 and 4. Similar results were found in many countries, such as the Netherlands, Hong Kong, South Korea, Ireland, and Australia [12,16,18–20]. The findings of the current study, which included a different population, would further reinforce the importance of alcohol consumption and smoking as factors in forming the clusters. In addition, the habit of exercising at least twice a week for at least 30 minutes was a variable that distinguished Cluster 1. Exercise is effective in preventing or improving NCDs [38]. Several studies reported that exercise habits are an important health behavior variable [20,31,39]. However, they were not shown to affect the other clusters in this study significantly. This may be because other unhealthy behaviors, such as alcohol consumption and smoking, substantially affect clustering.

The study has several limitations. First, since the health behaviors used as variables are self-reported outcomes, bias may have been introduced because the respondents may have unconsciously given socially desirable responses. In addition, the health behavior variable employed a dichotomous or ordinal scale, which may more inaccurate than free description [40]. Another limitation was the limited number of health behavior variables. Studies reported that educational background and nutritional status were effective [12]. The study could not address all health behavior variables. However, the health behavior variables used in the current study were based on those that had been key factors in different countries in the previous years and whose selection was thus valid. Using these variables also has the advantage of cross-national comparison. As an additional limitation, the target population for the study was set by adopting a slightly stricter criterion than the national recommendation for requiring medical examinations. This may reflect a need for additional rationales in selecting the participants. However, the reference values for the inclusion of participants in the study reflected the values established by Japanese conferences and guidelines and are evidence-based. In addition, although the reference value was unique to Japan, it caused a few problems because all the subjects in the study were Japanese. Despite such limitations, the study has sufficient strengths. The study focuses on the health behavior of a large sample of more than 10,000 working-age Japanese. It is further focused on the population that needed care for NCDs. Virtually no studies meet these criteria. Another important feature of the analysis is that it was performed using LCA. This allowed for a more flexible and distinctive cluster classification. The study’s results suggest the need for interventions tailored to these characteristics in a population approach to public health. They also provide valuable information when recommending health consultations to prevent severe disease. In this regard, this study has significant value. Future studies
may lead to additional findings by investigating the association of each cluster with outcomes such as quality of life and the effects of interventions associated with the characteristics of this cluster.

**Conclusion**

In conclusion, this study reveals more discernible patterns in health behavior in a population sample with extensive real-world data in Japan. Knowledge of clusters common to large populations is important information for public health policies to improve health. The study results revealed that five clusters based on health behaviors form among the population of those in need of treatment for NCDs. This result suggested the need for an approach tailored to each group’s characteristics, rather than treating them as a uniform group, when considering population approaches in public health.

**Abbreviations**

NCDs  
Noncommunicable Diseases  
LDL  
Low-Density Lipoprotein  
HDL  
High-Density Lipoprotein  
LCA  
The latent class analysis  
BIC  
The Bayesian Information Criterion  
AIC  
Akaike Information Criterion

**Declarations**

Ethics approval and consent to participate

It was conducted by the Declaration of Helsinki and was approved by the Konan Women’s University Research Ethics Committee (approval number 2022021) and Informed consent was obtained from all participants.

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.
Competing interests

TM and MK. has received consulting fees from PREVENT Inc. and is a non-regular staff member of PREVENT Inc. YH is a founder and stockholder of PREVENT Inc. KT, KY and MI are employees of PREVENT Inc.

Funding

Not applicable

Authors’ contributions

All authors contributed significantly to this article. TM and YH conceived and designed the study; TM, MK and KY analyzed the data and wrote the manuscript.

MK, KT, MI and TH collected the data. MK and YH supervised the implementation; All authors provided critical comments on the manuscript and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References


Figures

**Figure 1**

Overview of the distribution in the five latent classes.

This figure shows the demographics of the participants in the five classes and the health behavior characteristics of each.