Clustering of modifiable biobehavioral risk factors for chronic disease in US adults: a latent class analysis

Abstract

Aims: Examining the co-occurrence patterns of modifiable biobehavioral risk factors for deadly chronic diseases (e.g. cancer, cardiovascular disease, diabetes) can elucidate the etiology of risk factors and guide disease-prevention programming. The aims of this study were to (1) identify latent classes based on the clustering of five key biobehavioral risk factors among US adults who reported at least one risk factor and (2) explore the demographic correlates of the identified latent classes.

Methods: Participants were respondents of the National Epidemiologic Survey of Alcohol and Related Conditions (2004–2005) with at least one of the following disease risk factors in the past year (N = 22,789), which were also the latent class indicators: (1) alcohol abuse/dependence, (2) drug abuse/dependence, (3) nicotine dependence, (4) obesity, and (5) physical inactivity. Housing sample units were selected to match the US National Census in location and demographic characteristics, with young adults oversampled. Participants were administered surveys by trained interviewers.

Results: Five latent classes were yielded: ‘obese, active non-substance abusers’ (23%); ‘nicotine-dependent, active, and non-obese’ (19%); ‘active, non-obese alcohol abusers’ (6%); ‘inactive, non-substance abusers’ (50%); and ‘active, polysubstance abusers’ (3.7%). Four classes were characterized by a 100% likelihood of having one risk factor coupled with a low or moderate likelihood of having the other four risk factors. The five classes exhibited unique demographic profiles.

Conclusions: Risk factors may cluster together in a non-monotonic fashion, with the majority of the at-risk population of US adults expected to have a high likelihood of endorsing only one of these five risk factors.

INTRODUCTION

A promising method for studying the etiology of chronic diseases that plague the population, such as cardiovascular disease, cancer, and diabetes, involves investigating the modifiable biobehavioral risk factors that give rise to these diseases. Of the various modifiable biobehavioral risk factors for chronic disease, harmful alcohol and drug use, habitual cigarette smoking, obesity, and physical inactivity are important to study because they are highly prevalent in the US population.1-5 Furthermore, these particular risks are clear causes of numerous negative health outcomes, account for considerable economic and social burden, and are modifiable via intervention.6,7 These risk factors are also notable because they tend to co-occur in some individuals,8-10 and the clustering of these risk factors may have synergistic effects that disproportionately increase the likelihood of negative health outcomes.11

Examining the co-occurrence patterns of risk factors is important as such information could shed light on their etiology. For instance, if two risk factors cluster together with each other more than the other factors, this may suggest that they could be influenced by a common source. Furthermore, if certain cluster configurations are more common in certain subgroups (e.g. lower income bracket),...
characteristics common in that subgroup (e.g. reduced access to quality health care) may perhaps give rise to the co-occurrence of those risk factors.

Information on clustering may also guide designing disease-prevention programs. Intervention approaches that target multiple modifiable risk factors in a single program are gaining favor because they are potentially more efficient, cost-effective, and may have a greater public health impact than single-risk-factor approaches. Thus, data on the prevalence of subgroups with particular configurations of various risk factors and types of individuals represented in each of these subgroups would inform practitioners and policymakers regarding (1) the overall need for multi-risk factor interventions to prevent chronic disease, (2) the particular risk factors that should be grouped together in such interventions, and (3) the segments of the population that may benefit most from multi-factor intervention approaches.

These modifiable biobehavioral risk factors for chronic disease are unlikely to occur in an entirely uniform manner in population. That is, it is doubtful whether people can be accurately characterized into two groups – those that lead healthy versus those that lead unhealthy lifestyles. For instance, certain individuals who are physically active tend to use alcohol more frequently than their inactive counterparts (e.g. young adults). Also, tobacco use is associated with lower rates of obesity among some individuals (e.g. older adults). Thus, alcohol-, drug-, and tobacco-use disorders; physical inactivity; and obesity do not appear to co-occur in a monotonic fashion; rather, there are qualitative differences in profiles of heterogeneity, which may pose problems for designing disease-prevention programs. Indeed, a common challenge raised by public health professionals is that program development for multiple risk factor interventions is difficult due to the ‘infinite combinations of risk factors with which patients may present’.

Latent class analysis (LCA) is a useful method for identifying homogeneous subgroups within a heterogeneous population with regard to the manifestations of a set of categorical characteristics. Thus, instead of subgrouping every possible profile, LCA helps to reduce the data into the most parsimonious set of classes while accounting for measurement error. Despite the fact that there are a number of possible risk factors for chronic disease (e.g. family histories of cancer, heart disease, stroke, blood pressure, or blood cholesterol levels), focusing on modifiable factors that can be addressed via non-biologic treatment is important for public health.

There have only been a few prior studies that have applied LCA to characterize the clustering of modifiable behavioral risk factors for disease, which leaves several important gaps in the literature. First, to our knowledge, no study has used LCA to specifically investigate the co-occurrence of alcohol-, drug-, and tobacco-use disorder; physical inactivity; and obesity. Nonetheless, these specific five biobehavioral risk factors are critically important to study as a unique set of indicators using LCA because they tend to co-occur in a non-uniform fashion and are among the most common targets in behavioral interventions. Studies have often relied on non-diagnostic indicators of substance use, such as use status, frequency, or duration, which may be important predictors of disease, disability, and mortality risk and may be independent of one’s frequency or duration of substance use for the available studies using LCA with other biobehavioral risk factors are currently lacking. Third, there has been little prior research applying LCA to population-based samples, with the majority of prior work focusing on clinic-based samples or restricted community-based samples. By examining clustering of biobehavioral risk factors for chronic disease in a national population-based sample, the resultant findings may be relevant for guiding health promotion programs at the national community level.

This study identified latent classes based on the clustering of five key modifiable biobehavioral risk factors for chronic disease in a representative sample of US adults. We aimed to clarify (1) number of classes that could parsimoniously describe patterns of risk factor co-occurrence, (2) configurations of risk factor clustering represented in each class, and (3) the expected prevalence of each class. Once we established latent classes, we explored the demographic correlates of class membership to describe the makeup of each class, with the idea that this information could inform how prevention efforts are to be directed toward demographic cross-sections of the population with particular risk profiles. Given the potential for non-uniformity in patterns of co-occurrence, we did not hypothesize a particular number of classes or the risk factor configuration that would constitute each class. Rather, as in prior LCA work, we utilized an empirical data-driven approach for LCA.

METHODS
Sample and procedure
Participants were respondents in the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). The NESARC was selected because, to the best of our knowledge, the NESARC represents the most recent nationally representative study that yields clinical diagnoses of substance-use disorder as well as physical activity. The US Census Bureau and the US Office of Management reviewed and approved all consent procedures, research measures, and protocol.

For NESARC Wave 1 (2001–2002), the housing unit sampling frame was based to provide a systematic random sampling stratified by location and demographic factors to match the US Census 2000. All NESARC participants were civilian, non-institutionalized individuals aged 18 years and above. The following groups were oversampled because the NESARC aimed to inform public health substance-use trends in high-priority populations: African-Americans (1.55 oversampling ratio), Hispanic-Americans (1.59 oversampling ratio), and young adults aged 18–24 years (2.25 oversampling ratio). In each household, one adult was selected, and face-to-face interviews were conducted by lay interviewers or clinicians in

332 Perspectives in Public Health | November 2014 Vol 134 No 6
respondents’ homes following informed consent (N = 43,093). A second wave of assessment was conducted 3 years later (Wave 2, 2004–2005). The overall response rate for Wave 1 was 81.0%, and the subsequent response rate among those eligible for Wave 2 was 86.7% (N = 34,653). The current analyses utilize the Wave 2 sample only because physical activity was measured only at Wave 2. The sample was further restricted to those who reported having at least one of the risk factors (N = 22,789), which served as the five binary latent class indicators for the study: (1) past-year alcohol abuse/dependence, (2) past-year drug abuse/dependence, (3) past-year nicotine dependence, (4) current obesity, and (5) failed to meet federal government recommended weekly physical activity levels in the past year (see assessment section below for more information about the indicators). Requiring the presence of one or more risk factors to be included in the analysis increases the likelihood that a reasonable number of latent classes will be yielded with adequate distributions of individuals in each class. Because pregnancy can inflate body mass index (BMI), we excluded those who were currently pregnant (n = 348, 2.7%). Finally, participants younger than 21 years old were excluded because illegality of drinking may have impacted the accuracy of their reports (n = 24, 0.07%), leaving a final sample of 22,427 participants for analyses.

Assessment
The Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS-IV)37 was used to collect information on demographics, physical activity, weight/height, Diagnostic and Statistical Manual for Mental Disorders (4th ed.; DSM-IV)38 criteria for substance-use disorders, and other characteristics. Reliability and validity estimates for substance-use disorder diagnoses and other characteristics in this sample were adequate and available in a previous report.37

Substance-use disorders
Participants who met criteria for either abuse or dependence were combined into an abuse/dependence classification for both alcohol and drug indicators. We combined abuse and dependence because (1) both abuse and dependence are associated with negative health outcomes,38 thus any substance-use disorder may indicate increased disease risk; (2) abuse and dependence represent different severity levels of the same risk factor condition (i.e., substance-use disorder); and (3) it is likely that abuse and dependence will be combined into a single category in Diagnostic and Statistical Manual for Mental Disorders (5th ed.; DSM-V).39 Thus the present findings may have more relevance to future diagnostic classifications. In concordance with extant reports,1–3,40 all illicit and prescription drug-use disorders were analyzed as a combined drug abuse/dependence category to increase prevalence and reduce the number of indicators.

Obesity
BMI was computed based on the respondents’ self-reported body weight and height (kg/m²) because objective assessments were not collected in NESARC. Individuals with BMIs ≥ 30 kg/m² were classified as obese. We chose to focus on the distinction between obese versus not obese (rather than other cut-offs, such as overweight ≥25 kg/m²) based on evidence suggesting that obesity (but not overweight) is associated with increased mortality from cardiovascular disease and obesity-related cancers.41

Physical activity
Participants were asked to report their levels of physical activity over the past 12 months. Moderate physical activity was defined as ‘activities that caused only LIGHT sweating or SLIGHT to MODERATE increases in your breathing or heart rate’ and vigorous physical activity was defined as ‘activities that caused you to sweat HEAVILY or caused LARGE increases in your breathing or heart rate’. Participants were asked about the frequency (e.g., number of times per week/month/year) and average duration of a typical period of activity (answer in minutes). The US federal government currently recommends that for substantial health benefits, adults should engage in at least 150 minutes a week of moderate-intensity aerobic physical activity, 75 minutes a week of vigorous-intensity aerobic physical activity, or an equivalent combination of moderate- and vigorous-intensity aerobic activity.42 As in previous work,43 our physical activity indicator was whether individuals met some combination of this weekly recommendation on average over the past year.

Demographic characteristics
Participants reported their age in years, gender, marital status, race/ethnicity, education level, and annual household income. Marital status was recoded to a binary variable (married vs other). Education was an ordinal variable (less than high school, high school diploma or equivalent, or post–high school). Income was also a continuous variable (21 levels with increments ranging from US$3,000 to US$200,000+).

ANALYSIS
A series of latent class models were iteratively conducted, and the smallest set of latent classes that best fit the data was determined. To do so, we evaluated various model fit indices, including Pearson χ², likelihood ratio χ², Akaike information criterion (AIC) and Bayesian information criterion (BIC), Lo–Mendell–Rubin likelihood ratio (LMRLR); at likelihood ratio test for statistically significant differences in fit across models with different numbers of classes, and entropy values. Thereafter, the latent classes identified were then subjected to multinomial logistic regression to assess the association between demographic characteristics and latent classes. All analyses were conducted using Mplus Version 6.0 software program44 and Stata Version 11. The complex survey design of the NESARC was accounted for, and Wave 2 sampling weights were applied to approximate the demographics of US Census 2000 population. A two-sided .01 significance level was applied.

RESULTS
Unweighted distribution of risk factors
In the sample, 14.0%, 20.0%, and 3.3% met criteria for past-year alcohol-
Clustering of modifiable biobehavioral risk factors for chronic disease in US adults: a latent class analysis

Prevalence and characterization of five latent classes

For the 5-class solution, conditional item-response probabilities are presented in Table 2. A total of 23% of the sample was expected to belong to Class 1, which had a response pattern characterized as ‘obese, active non-substance abusers’. Class 2 (19%) was characterized as ‘nicotine-dependent, active, and non-obese’. Class 3 (6%) was characterized as ‘active, non-obese alcohol abusers’. Class 4 constituted the largest latent class (50%) and was characterized as ‘inactive, non-substance users’. Respondents expected to belong to Class 5 were characterized as ‘active, polysubstance abusers’ (3.7%).

Association between demographic characteristics and latent class membership

After the latent classes were identified, we examined the demographic correlates of class membership (results reported in Table 3); latent class was not contingent upon the demographic variables. Relative to the most prevalent Class 4 (inactive, non-substance abusers), older participants were less likely to belong to all other latent classes (see Table 3). Relative to Class 4, males were more likely to belong to all other classes than females were. Married respondents were less likely to belong to Classes 2 (nicotine-dependent, active, and non-obese), 3 (active, non-obese alcohol abusers), and 5 (active polysubstance abusers), relative to Class 4, than non-married respondents. The married participants, however, were 11% more likely to be classified as Class 1 (‘obese, active non-substance abusers’) versus Class 4. Relative to Class 4, respondents with a non-White racial background were less likely to belong to all the other classes except for the individuals with an American Indian/Alaska Native racial background. Also, African-Americans were 30% more likely to belong to Class 1 (‘obese, active non-substance abusers’) than White respondents. Relative to Class 4, participants with higher income were more likely to be classified as Classes 1 (obese, active, non-substance abusers) and 3 (active, non-obese alcohol abusers) but less likely to belong to Class 2 (nicotine-dependent, active, and non-obese). Compared with participants who did not attain a high school diploma, those with a high school diploma/General Educational Development (GED)

<table>
<thead>
<tr>
<th></th>
<th>Number of classes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Pearson $\chi^2$</td>
<td>478,360</td>
</tr>
<tr>
<td>LR $\chi^2$</td>
<td>550,124</td>
</tr>
<tr>
<td>df</td>
<td>20</td>
</tr>
<tr>
<td>No. of parameters</td>
<td>11</td>
</tr>
<tr>
<td>AIC</td>
<td>105,251.866</td>
</tr>
<tr>
<td>BIC</td>
<td>105,251.866</td>
</tr>
<tr>
<td>LMR testing the null hypothesis</td>
<td>Class 1 vs Class 2</td>
</tr>
<tr>
<td>LMR probability</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Entropy</td>
<td>.791</td>
</tr>
</tbody>
</table>

LMR: Lo–Mendell–Rubin; AIC: Akaike information criterion; BIC: Bayesian information criterion; N/A: Not Applicable.

Results from model fit indices. LMR illustrates that model fit improves up to 5 classes and then does not significantly improve fit beyond 5 classes. Hence, we selected the 5-class model as our final model.

nicotine-, and drug-use disorder, respectively. Approximately, 45% of the sample was currently obese, and 57.8% of the sample failed to meet recommended physical activity levels over the past year.

Determining the best-fitting model

As illustrated in Table 1, as the number of classes increased, AIC and BIC decreased, indicating improved model fit. The fit improvement was evidenced only up to the 5-class solution. LMRLR comparing $n$ versus $n - 1$ class models (i.e. reject the null hypothesis that $n - 1$ class fits the data better than $n$ class model, if $p < .01$).45 LMRLR test showed that the 5-class model was the best-fitting model and yielded highest value of entropy.

Prevalence and characterization of five latent classes

Table 1
were more likely to be classified as Classes 1 (obese, active, non-substance abusers) and 3 (active, non-obese alcohol abusers), relative to Class 4. Furthermore, respondents with higher levels of income were more likely to belong to Classes 1 (obese, active non-substance abusers) and 3 (active, non-obese alcohol abusers) relative to Class 4, yet less likely to belong to Class 2 (nicotine-dependent, active, and non-obese) relative to Class 4. Finally, those with some post-high school education were more likely to belong to Classes 1 (obese, active non-substance abusers) and 3 (active, non-obese alcohol abusers), but less likely to belong to Class 2 (nicotine-dependent, active, and non-obese).

**DISCUSSION**

In a population-based sample of US adults, five distinct latent subgroups adequately accounted for variation in co-occurrence patterns of five clinically important modifiable biobehavioral risk factors for chronic disease. Each of these classes exhibited unique risk factor configurations and demographic profiles from one another. Thus, chronic disease-prevention programming distributed to the at-risk population may be beneficial if tailored and targeted to those particular population subgroups and corresponding risk factor profiles illustrated in this analysis (see Tables 2 and 3).46,47

Among the US adults who endorsed at least one of the index risk factors studied in this report, four of the five classes were characterized by a 100% likelihood of having one of the risk factors coupled with a low or moderate likelihood of having the other four risk factors. Latent classes with 1.00 or 0.00 probabilities for indicators have been reported before in the literature55 and would be expected in studies using samples saturated for risk factor endorsement such as in this study, suggesting highly discriminative classes. Each of these four classes was typified by the presence of a different risk factor. Collectively, these four classes represented 96% of the overall sample. This finding has implications for chronic disease-prevention programming for these particular biobehavioral factors among those who endorsed at least one of these risk factors. One perspective is that programs focused primarily on single risk factor may be most efficient and practical in circumstances in which financial and temporal resources are limited.25 Another perspective is that targeting multiple health behaviors may nonetheless be useful, particularly when a single intervention model can be readily applied to multiple risk factors.51 Regardless, our findings suggest that for these five risk factors known to account for substantial public health burden,1–7 most members of the at-risk community are typified by a high likelihood of only a single risk factor.

Several high-prevalence classes were identified in this sample. About half of the respondents were expected to belong to a class of inactive individuals who did not have concurrent substance-use problems and had slightly lower probability of obesity in comparison to the overall sample. Thus, physical inactivity appears to be a highly prevalent factor, which underscores the potential utility of population-based physical activity programming.49 Almost a quarter of the sample comprised a class of obese individuals who were nevertheless highly likely to currently meet recommendations for weekly physical activity. Accordingly, members of this class may benefit most from obe-
Clustering of modifiable biobehavioral risk factors for chronic disease in US adults: a latent class analysis

A class of active, non-obese individuals with alcohol-use disorders predicted to represent 6% of the sample was identified. The profile of this class is consistent with recent work illustrating a counterintuitive positive association between alcohol use and physical activity. Recent evidence suggests that relation between physical activity and substance use appears to be specific to alcohol but not other substances even after controlling for the effects of other substance-use disorders, which is concordant with the current results illustrating that members of this class had little or no likelihood of comorbid tobacco- and drug-use disorder. Furthermore, extant data suggest that physical activity and alcohol use co-occurrence is more common among younger (vs older) and male (vs female) individuals. Consistent with past findings, demographic analyses of this class also suggest that physically active alcohol users tend to be younger and males. We also found that members of this class are more likely to be non-married, White, of a higher income bracket, or more educated. Thus, it may be important to assess for alcohol-use disorder among active adults from this demographic bracket, who may otherwise appear healthy and potentially overlooked in alcohol programming.

A class of polysubstance abusers expected to comprise 4% of the sample was yielded. This class was defined by meeting diagnostic criteria for an illicit drug-use disorder combined with a high likelihood of comorbid tobacco- and drug-use disorder. Furthermore, extant data suggest that physical activity and alcohol use co-occurrence is more common among younger (vs older) and male (vs female) individuals. Consistent with past findings, demographic analyses of this class also suggest that physically active alcohol users tend to be younger and males. We also found that members of this class are more likely to be non-married, White, of a higher income bracket, or more educated. Thus, it may be important to assess for alcohol-use disorder among active adults from this demographic bracket, who may otherwise appear healthy and potentially overlooked in alcohol programming.

### Table 3

<table>
<thead>
<tr>
<th>Class descriptors: Class 1 – active, obese non-substance abusers; Class 2 – nicotine-dependent active non-obese; Class 3 – active, non-obese alcohol abusers; Class 4 – inactive non-substance abusers; Class 5 – active, polysubstance abusers; RRR: relative risk ratio; GED: General Educational Development; SE: standard error.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class 1</strong></td>
</tr>
<tr>
<td><strong>RRR</strong> (SE)</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Married</td>
</tr>
<tr>
<td>Race/ethnicity</td>
</tr>
<tr>
<td>Black, non-Hispanic</td>
</tr>
<tr>
<td>American Indian/Alaska Native, non-Hispanic</td>
</tr>
<tr>
<td>Asian/Pacific Islander, non-Hispanic</td>
</tr>
<tr>
<td>Hispanic, any race</td>
</tr>
<tr>
<td>Income</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>High school diploma/GED</td>
</tr>
<tr>
<td>Post–high school education</td>
</tr>
</tbody>
</table>

*Relative to Class 4, the most prevalent class.

*Compared with ‘non-married’, including cohabitation, widowed, divorced, separated, and never married.

Reference category was non-Hispanic White.

Reference category was those who did not attain high school diploma or GED.

*p < .01; **p < .001.
class, this class was more likely to be younger, male, non-married, and White. The characteristics of this class are consistent with notions that the co-occurrence among multiple substance-use disorders may reflect an underlying syndrome of externalizing behavior that is more common in young men and may be associated with antisocial tendencies that could interfere with successful marriage.\textsuperscript{56,57} Clinically, these findings suggest that a small, but sizeable, portion of individuals may require multi-substance addiction interventions.\textsuperscript{58–60}

This study had limitations that impact interpretation of these findings. The goal of this study was to assess clustering of five putatively important modifiable biobehavioral risk factors for chronic disease. Thus, the findings primarily generalize to considerations of groupings of individuals with regard to these indicators, and the prevalence and profiles of classes may be different if other risk factors (e.g., diet) are considered in concert with the five we focused on. The intention of this study was to describe patterns of co-occurrence, and therefore, we used a cross-sectional design. However, this design does not elucidate causal and temporal mechanisms that may link risk factors to mechanisms that may link risk factors to disease.\textsuperscript{56,57,61–63} Thus, these findings are likely generalizable to today’s public health trends in the most part. This study also had several offsetting strengths, such as the use of a nationally representative sample, application of a current analytic methodology, and use of valid clinical intervention assessments to generate DSM-IV substance-use disorder diagnoses.

CONCLUSION

To our knowledge, this was first LCA study of these key modifiable biobehavioral risk factors, which were purposefully selected because of their public health significance and their tendency to co-occur in non-uniform configurations.\textsuperscript{1–7,24} Results suggest that risk factors cluster together in a unique fashion, with the majority of the at-risk population of US adults expected to have a high probability of endorsing only one of these five biobehavioral risk factors. Accordingly, health promotion programming for substance-use disorder, physical activity, and obesity that targets the population that endorses at least one of these risk factors should perhaps consider the utility of single risk factor interventions.

ACKNOWLEDGEMENTS

This research was supported by US National Institutes of Health grant DA025041.

DECLARATION OF CONFLICTING INTERESTS

The authors report no conflicts of interest related to the submission of this manuscript.

References


12. Huh J, Rigs NR, Spruijt-Metz D, Chou CP, Huang Z, Pentz M. Identifying patterns of eating behaviors that could interfere with successful marriage.\textsuperscript{56,57} Clinically, these findings suggest that a small, but sizeable, portion of individuals may require multi-substance addiction interventions.\textsuperscript{58–60}
Clustering of modifiable biobehavioral risk factors for chronic disease in US adults: a latent class analysis
