Published in final edited form as:

Ann Epidemiol. 2013 July ; 23(7): 381–387. doi:10.1016/j.annepidem.2013.04.007.

Implementation of Permutation Testing to Determine Clustering of Social and Behavioral Risk Factors for Coronary Heart Disease, NHANES 2001-2004

Nicholas J. Everage, PhD^a, Crystal D. Linkletter, PhD^{b,c}, Annie Gjelsvik, PhD^{a,d}, Stephen T. McGarvey, PhD, MPH^{a,e}, and Eric B. Loucks, PhD^{a,f}

^aDepartment of Epidemiology, Warren Alpert Medical School, Brown University, Providence, RI

^bCenter for Statistical Sciences, Brown University, Warren Alpert Medical School, Brown University, Providence, RI

^cDepartment of Epidemiology, Biostatistics Section, Warren Alpert Medical School, Warren Alpert Medical School, Brown University, Providence, RI

^dDiabetes Prevention and Control Program, Rhode Island Department of Health, Providence, RI

^eInternational Health Institute, Warren Alpert Medical School, Brown University, Providence, RI

^fCenter for Population Health & Clinical Epidemiology, Warren Alpert Medical School, Brown University, Providence, RI

Abstract

Purpose—To determine if social and behavioral risk factors for CHD, including education, physical activity (PA), fruit/vegetable intake and smoking, cluster (i.e. co-occur more than expected due to chance) in US adults.

Methods—The study included 4,305 males and 4,673 females aged ≥20 years from the National Health and Nutrition Examination Survey. Risk factors included: HS diploma/GED; <150 minutes of moderate/vigorous PA per week; <3 or <2 servings of vegetables and fruit, respectively, per day; and smoking cigarettes. Indicator variables were summed into a sociobehavioral risk index (SRI, range 0 (no risk factors) to 4 (all risk factors)). Ratios of observed-to-expected prevalence (under the assumption of independence) of the SRI were assessed. Statistical significance was evaluated using randomly permuted average observed-to-expected SRI ratios and 95% CIs.

Results—In males, the ratio of observed-to-expected prevalence of SRI=0 was 1.70 (permuted ratio=1.00, 95% CI: 0.92, 1.08), and SRI=4 was 2.10 (permuted ratio=1.00, 95% CI: 0.86, 1.14), demonstrating significant clustering. In females, the ratio of observed-to-expected prevalence of SRI=0 was 1.67 (permuted ratio=1.00, 95% CI: 0.92, 1.08), and SRI=4 was 1.86 (permuted ratio=1.00, 95% CI: 0.85, 1.15).

Conclusions—Social and behavioral risk factors for CHD cluster in this sample of United States adults.

MeSH Keywords

Heart diseases; risk factors

Despite years of increased attention to prevention and decreasing overall mortality rates, coronary heart disease (CHD) remains the leading cause of death in the US [1, 2]. Consequently, important changes must be made to curtail the loss of life and productivity brought about by this disease. Intervening upstream with lifestyle improvements, such as diet, increased physical activity (PA), and smoking cessation may yield substantial CHD benefits. Underlying lifestyle risk factors, such as smoking, low PA, higher alcohol consumption, low fiber intake, and high trans-fat and saturated fat diets, have been demonstrated to be responsible for a substantial proportion of CHD events [3, 4]. A number of individual- and community-based trials, including the Stanford Five-City Project, the Pawtucket Heart Health Program, and MRFIT have attempted to change lifestyle risk factors/behaviors related to CHD [5-10]. Systematic reviews of these cardiovascular disease prevention interventions have demonstrated either little effect or small but favorable reductions in cardiovascular disease risk in response to these programs [10, 11]. Similarly, social factors such as education may be important risk markers for CHD [12-15]. Over 3 decades in the US, educated and wealthier groups experienced greater decreases in CHD risk factors, such as smoking prevalence and cholesterol levels, compared to their less educated, poorer counterparts [16]. There have been advancements emphasizing the importance of social ecological intervention models. These models take into account the social context such as socioeconomic position (e.g. education), race/ethnicity, neighborhood characteristics and social network transmission of health behaviors, which may shape the success of health behavior interventions, or the behaviors themselves [17-21]. In considering interventions to prevent CHD, it may be helpful to consider the potential mutually reinforcing characteristics of both social and behavioral risk factors. This could facilitate the creation of more effective interventions, for example, if interventions on a single risk factor (e.g. PA) were substantially affected by other co-occurring risk factors such as diet, smoking and socioeconomic position. Understanding which social and behavioral risk factors might mutually influence each other could substantially inform etiologic understanding of CHD, and identify possible interventions aimed at addressing the mutually reinforcing causes of CHD. Unlike advances with metabolic syndrome, which have demonstrated that biological CHD risk factors (e.g. blood pressure, central obesity, fasting glucose and lipids) cluster (i.e. co-occur more often than would be expected due to chance) [22-24], there have been relatively few studies evaluating whether health behaviors cluster [25-27]. Furthermore, to our knowledge, no studies to date have statistically evaluated whether both social and behavioral CHD risk factors cluster. Therefore, the objective of this study was to determine if social and behavioral risk factors for CHD (specifically education, PA, fruit/vegetable intake, and smoking) cluster and are statistically significant in a population of US adults.

Materials and Methods

Study sample

The study sample included participants from the 2001-2004 National Health and Nutrition Examination Survey (NHANES). Between 2001 and 2004, participants underwent household interviews and physical examinations. The study was reviewed and approved by the National Center for Health Statistics Institutional Review Board. The sample size for adults 20 years old was 10,452.

Physical activity

For PA, the recommendations for the time period of the study (2001-2004) were from the Centers for Disease Control and the American College of Sports Medicine guidelines, which suggested “adults should accumulate 30 minutes or more of moderate-intensity PA on most, preferably all, days of the week [28].” This was interpreted as exercising 5 days per week for at least 30 minutes. Consequently the cut-point of 150 minutes moderate and/or vigorous of PA per week was considered as satisfying the recommendation [29]. Total weekly PA was estimated based on responses to the following questions: “Over the past 30 days, have you walked or bicycled as part of getting to and from work, or school, or to do errands?” and “Over the past 30 days, did you do any tasks in or around your home or yard for at least 10 minutes that required moderate or greater physical effort [30]?” In addition, participants reported time spent in 45 moderate or vigorous leisure-time physical activities [31]. Respondents reported frequency and average duration of all physical activities which was summed to calculate time per week in moderate/vigorous activity.

Fruit/vegetable intake

The NHANES used twenty-four hour dietary recalls to determine food and nutrient consumption, described elsewhere [32, 33]. In 2001-2002, one 24-hour recall was collected; in 2003-2004, participants reported two 24-hour recalls 3-10 days apart. Approximately 87% of participants in the 2003-2004 data completed 2 days of valid dietary recall data. From NHANES dietary data, the US Department of Agriculture calculated the number of cups of each food group consumed into a database called the MyPyramid Equivalents Database [34]. Participants were considered to have met US dietary recommendations (specifically the 1992 Food Guide Pyramid recommendations) if they consumed 3 servings (1.5 cups) of vegetables and 2 servings (1 cup) of fruit per day (based on the one-day recall for 2001-2002 and the mean of the two-day recall for 2003-2004) according to the aforementioned database [35].

Socioeconomic position

Education was self-reported and categorized as high school (i.e. high school diploma, GED, or less) versus >high school (i.e. some college, associate's degree, college or postgraduate) based on previous literature [36, 37], *allowing for adequate distribution of sample size in the two education comparison groups.*

Smoking status

Participants reporting they currently smoked 'every day' or 'some days' were considered current smokers. Nonsmokers reported were those that no longer smoked (i.e. 'not at all') or had not smoked at least 100 cigarettes.

Covariates

Age and race/ethnicity were collected by NHANES based on participant self-report. Race/ethnicity was self-reported as non-Hispanic black (n=1,719), non-Hispanic white (n=4,752), Mexican-American/Other Hispanic (n=2,174), or Other Race (n=333), which includes all other non-Hispanics who are not black or white. Due to fairly low sample size, participants indicating race/ethnicity of "other race" were excluded from analyses stratified by race/ethnicity.

Statistical analyses

These analyses were conducted from 2010-2011. As described above, the presence of each CHD risk factor (i.e. low educational attainment, not meeting fruit/vegetable guidelines, not meeting PA guidelines, and current smoking) was specified by a dichotomous indicator variable (i.e. 1=Yes, 0=No). These indicators were summed into a sociobehavioral risk index (SRI) ranging from 0-4 (0=No risk factors; 4=All risk factors). There were 16 unique combinations of the four risk factors, and the prevalence of each of the 16 risk factor combinations was assessed [25, 27, 38]. Furthermore, the observed prevalence of participants with 0, 1, 2, 3 or 4 risk factors was calculated. The expected prevalence estimates were developed under the assumption of independence between risk factors. Thus, the expected prevalence estimates were the product of the prevalence, or fixed marginal probabilities, for each combination of risk factors. Take, for example, a sample of males with the following prevalence estimates for each risk factor: Did not comply with PA guidelines=55%, did not meet fruit/vegetable intake guidelines=45%, current smokers=25%, and had a high school diploma/GED or less education=40%. The expected prevalence of having all of these risk factors would be the product of the prevalence estimates, or $0.55 \times 0.45 \times 0.25 \times 0.40 = 0.025$ or 2.5%. Clustering was considered present when the observed prevalence was greater than the expected prevalence of participants with the risk factors [25, 27].

In order to evaluate whether clustering was statistically significant, the analysis used 1,000 replicates of the NHANES sample using random permutation with the same fixed marginal probabilities and sample size as the original sample. For each replicate the prevalence of each combination of risk factors was calculated and then summed the prevalence of the combinations into the observed prevalence of participants with 0 risk factors, 1 risk factor, etc. Next, the analysis computed the ratio of the observed (randomly permuted) prevalence of participants with N risk factor(s) to the expected prevalence of participants with N risk factors(s) for each replicate. The mean of these ratios across all 1,000 replicates was estimated and a 95% confidence interval (CI) was developed around each mean. The ratios of observed-to-expected prevalence of the risk factors from the original sample were compared to these means and 95% CIs. If the 95% confidence intervals of the randomly

permuted observed/expected ratio prevalence were outside of the original observed/expected ratio prevalence, then clustering was considered to be statistically significant.

Sex-, racial/ethnic- and age-stratified analyses were used to understand potential effect modification of risk factor clustering by these variables. Analyses were conducted using SAS 9.2 (Carey, NC).

Results

Less than 1% of participants 20 years were missing data on education (n=22) and smoking (n=35); approximately 3% lacked information on PA (n=306); and 12% did not have data on fruit/vegetable intake (n=1,260). The final analytic sample included 8,978 participants who had data for all four sociobehavioral variables. Compared to those excluded from the sample, included participants were younger (49 vs. 57 years; $p<0.0001$), and had higher smoking prevalence (22.6% vs. 18.9% smokers; $p=0.002$), lower levels of education attainment (53.3% vs. 62.7% with more than a high school education; $p<0.0001$), less likely to have met fruit/vegetable intake guidelines (53.6% vs. 61.2%; $p=0.03$), and PA guidelines (47.6% vs. 55.2%; $p=0.03$). There were no significant differences between included and excluded participants by race/ethnicity ($p=0.09$).

Characteristics of the study sample are shown in Table 1. Smoking, fruit/vegetable guideline compliance, and PA guideline compliance were generally lower in females than males. Male and female participants were fairly similar in age (mean age: 50.1 years for males and 48.5 years for females), racial/ethnic background, and education. The distribution of age, race/ethnicity, smoking, education, PA and fruit/vegetable intake according to SRI (Table 2) suggested that healthful risk factor clustering (SRI=0) was more likely to occur in white participants than black or Mexican American/Hispanic participants.

In this study, 15.4% of males had healthful SRI=0, compared with the expected (and randomly permuted) SRI prevalence of 9.1% (Table 3, Figure 1a). The ratio of observed/expected SRI prevalence was 1.70. Given that the observed/expected ratio of 1.70 was outside the 95% confidence intervals for the randomly permuted ratio (permuted ratio=1.00, 95% CI: 0.92, 1.08), this demonstrated statistically significant clustering of healthful social and behavioral risk factors in males (Table 3, Figure 1a). Similarly, for females, the observed healthful SRI=0 was 14.1% compared with an expected (and randomly permuted) SRI=0 prevalence of 8.4%, giving a ratio of observed-to-expected of 1.67 (randomly permuted ratio: 1.00, 95% CI: 0.86, 1.14; Table 3, Figure 1b). For unhealthy risk factor clustering, 7.1% of males had all four sociobehavioral risk factors (SRI=4), compared to an expected (and randomly permuted) prevalence of 3.3% in this sample. This yielded an observed/expected ratio of 2.10, which was well outside the 95% confidence intervals for the randomly permuted ratio of 1.00, 95% CI: 0.86, 1.14, demonstrating significant clustering of unhealthy sociobehavioral risk factors in males. Similarly in females, 5.1% had an SRI of 4, compared with the expected (and randomly permuted) prevalence of 2.7%, yielding an observed/expected ratio of 1.86 which was well outside the 95% confidence intervals for the randomly permuted ratio (1.00, 95% CI: 0.85, 1.15), again demonstrating significant clustering (Table 3, Figure 1b).

Further analyses stratified by race/ethnicity demonstrated there was statistically significant clustering among all racial/ethnic groups (i.e. non-Hispanic whites, non-Hispanic blacks and Mexican-American/Other Hispanics; Table 4) and all age groups (20-34, 35-49, 50-64 and >65 years; Table 5).

Discussion

Overall, this study demonstrated that social and behavioral CHD risk factors (education, smoking, fruit/vegetable intake and PA) cluster in males and females, regardless of age or racial group. There was generally clustering of both healthy (i.e. >high school education, 3 vegetables and 2 fruits per day, non-smoking and 150 minutes of moderate/vigorous PA per week) and unhealthy (i.e. high school education, <3 vegetables or <2 fruits per day, smoking and <150 minutes of moderate/vigorous PA per week) social and behavioral CHD risk factors.

Prior literature

While some studies have evaluated behavioral risk factor clustering, our study uniquely addresses the clustering of both social and behavioral risk factors. Furthermore, very few studies have attempted to robustly evaluate the statistical significance of clustering [27, 39]. Though our study is not completely analogous to others focused on CHD risk factor clustering, our findings are generally similar with the suggestion of clustering of behavioral risk factors. Fine et al. demonstrated evidence of clustering of alcohol consumption, overweight, smoking and physical inactivity using the National Health Interview Survey, but did not measure the statistical significance of the clustering directly [26]. Schuit et al. similarly discovered that behavioral risk factors for heart disease, specifically smoking, excessive alcohol intake, low PA, and low consumption of fruits and vegetables, clustered in a Dutch population, without direct statistical significance testing of the clustering [25]. Tobias et al. found that healthy and unhealthy risk factors, including fruit/vegetable intake, PA, alcohol consumption, and tobacco use, clustered in a statistically significant manner in a national sample of New Zealanders [39]. Furthermore, Alamian and Paradis evaluated the statistical significance of the clustering using bootstrap techniques [27] and found that the clustering of chronic disease behavioral risk factors, including physical inactivity, sedentary behavior, ever smoking, ever drinking and high BMI, among adolescents in Canada was significant, especially among children of low socioeconomic position. Our study builds on the literature through the novel inclusion of both social and behavioral risk factors in clustering schemas as well as by putting forth a unique and relevant form of performing statistical significance testing of clustering.

Potential mechanisms

The clustering of the social and behavioral risk factors seen in this current study may occur due to mutually reinforcing associations between each of the risk factors. That is, the underlying associations between each of these risk factors may bolster the presentation of clustering. For example, physical inactivity and smoking have been found to cluster with each other [26, 40], which may be due to the somatic impact of smoking. It has been shown that that smoking leads to decreased cardiovascular capacity and therefore increased

difficulty and less benefit from performing cardiovascular-supporting PA[41, 42]. Smoking has also been associated with decreased fruit/vegetable intake for several hypothesized reasons, including decreased taste and smell of consumed foods or even distaste for sweeter items [43]. Educational attainment strongly predicts income and occupation, which affect the quality and safety of neighborhoods that one can afford to live in. Perceived neighborhood safety has been directly associated with pedometer-determined PA, particularly among females [44]. Educational attainment positively affects health literacy, which in turn is related to fruit/vegetable consumption [45]. In these ways education, diet, PA and smoking may influence each other, resulting in observed clustering of the social and behavioral factors.

Potential common prior causes to education, smoking, diet and PA may be responsible for observed sociobehavioral risk factor clustering [46]. Link and Phelan discuss the concept of “fundamental determinants of health” which are upstream fundamental causes of health, some of which may be social policies, socioeconomic position, or intelligence [47]. The observed sociobehavioral clustering in this study may in part be due to fundamental common prior causes to diet, PA, smoking and education, such as childhood socioeconomic circumstances and neighborhood characteristics.

This study evaluated four sociobehavioral CHD risk factors that have fairly robust evidence to predict CHD (PA, smoking, fruit/vegetable intake, and education; note that education would be better termed a “CHD risk marker” as, although it predicts CHD, evidence on its causal role in CHD is still lacking) [48, 49]. However, the sociobehavioral clustering may not be limited to these factors alone. Other social/psychosocial constructs and health behaviors such as occupation, neighborhood safety, health literacy, intelligence, depression, social integration, alcohol consumption, caloric intake and medication adherence, may well cluster in addition to the variables assessed in this study. Similarly to the metabolic syndrome where many biological risk factors cluster even outside of standard definitions of the metabolic syndrome (e.g. inflammatory markers and clotting factors cluster with more traditional metabolic syndrome variables such as lipids, glucose, blood pressure and central obesity) [23], it is expected that many social and behavioral CHD risk factors/markers cluster. Future research can evaluate the breadth of sociobehavioral clustering, including determining whether the aforementioned additional social and behavioral factors cluster.

Clinical and population health implications

There have been substantial advancements emphasizing the importance of social ecological intervention models account for the social context which may shape the success of health behavior interventions, or the behaviors themselves [17-19, 21, 50]. In considering interventions to prevent CHD, it may be helpful to consider the potential mutually reinforcing characteristics of both social and behavioral risk factors. It is important to note there is evidence for clustering of positive health factors, as well as clustering of negative health factors. The mechanisms responsible for positive health clustering (e.g. improved feelings of well-being) and negative health clustering (e.g. neighborhood adversity) may be distinct. Greater awareness of the potential mutually reinforcing characteristics of both social and behavioral risk factors could help to create more effective interventions, for

example if interventions on a single risk factor (e.g. PA) may be substantially affected by co-occurring other risk factors such as diet, smoking and socioeconomic position. Understanding which social and behavioral risk factors may mutually influence each other could substantially inform etiologic understanding of CHD, and identify possible interventions that will aim to address the mutually reinforcing causes of CHD.

Strengths and limitations

Measurement error is inherent in measures of PA and diet, consequently some misclassification is expected in this study. In the case of food intake, reliance upon memory recall of consumed foods and their associated quantity typically leads to an underestimation of total nutrient intake [51]. Multiple 24-hour recalls are preferred over dietary data from a single 24-hour period. In this study, NHANES increased the number of dietary recall days from 1 day in 2001-2002 to 2 days in 2003-2004. In NHANES, when compared to accelerometer use, research has shown that self-reported PA overestimates guideline compliance.[52] Thus, the number of people meeting PA guidelines, and consequently healthy sociobehavioral clustering, may be overestimated. Future work using more objective measures of physical activity will provide more accurate estimates. Because the NHANES data is cross-sectional in nature, causation cannot be determined between the sociobehavioral risk factors. Finally, to our knowledge, random permutation testing does not allow for weighting to be used, consequently findings from this study are from a nationally-derived sample, but cannot be considered to be nationally representative.

A strength of our study was the use of NHANES data, with its high level of quality control and assurance [53]. The study innovatively used permutation testing to evaluate the statistical significance of clustering in addition to assessing how social and behavioral risk factors co-occurred more frequently than expected.

This study demonstrated that social and behavioral risk factors for coronary heart disease (specifically education, smoking, fruit/vegetable consumption, and PA) cluster. In considering interventions to prevent coronary heart disease, it may be helpful to consider the potential mutually reinforcing characteristics of these risk factors.

References

1. Lloyd-Jones DM, Wilson PWF, Larson MG, Leip E, Beiser A, D'Agostino RB, et al. Lifetime risk of coronary heart disease by cholesterol levels at selected ages. *Arch Intern Med.* 2003 Sep 8; 163(16):1966–72. [PubMed: 12963571]
2. Ford ES, Capewell S. Coronary heart disease mortality among young adults in the U.S. from 1980 through 2002: Concealed leveling of mortality rates. *J Am Coll Cardiol.* 2007 Nov 27; 50(22): 2128–32. [PubMed: 18036449]
3. Hu FB, Manson JE, Stampfer MJ, Colditz G, Liu S, Solomon CG, et al. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N Engl J Med.* 2001 Sep 13; 345(11):790–7. [PubMed: 11556298]
4. Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC. Primary prevention of coronary heart disease in women through diet and lifestyle. *N Engl J Med.* 2000 Jul 06; 343(1):16–22. [PubMed: 10882764]

5. Carleton RA, Lasater TM, Assaf AR, Feldman HA, McKinlay S. The pawtucket heart health program: Community changes in cardiovascular risk factors and projected disease risk. *Am J Public Health.* 1995; 85(6):777–85. [PubMed: 7762709]
6. Luepker RV, Murray DM, Jacobs DR Jr, Mittelman MB, Bracht N, Carlaw R, et al. Community education for cardiovascular disease prevention: Risk factor changes in the minnesota heart health program. *Am J Public Health.* 1994; 84(9):1383–93. [PubMed: 8092360]
7. Farquhar JW, Fortmann SP, Flora JA, Taylor CB, Haskell WL, Williams PT, et al. Effects of communitywide education on cardiovascular disease risk factors: The stanford five-city project. *JAMA.* 1990; 264(3):359–65. [PubMed: 2362332]
8. Stamler J, Neaton JD. The multiple risk factor intervention trial (MRFIT)—importance then and now. *JAMA.* 2008 Sep 17; 300(11):1343–5. [PubMed: 18799447]
9. Chiuve SE, McCullough ML, Sacks FM, Rimm EB. Healthy lifestyle factors in the primary prevention of coronary heart disease among men: Benefits among users and nonusers of lipid-lowering and antihypertensive medications. *Circulation.* 2006; 114(2):160–7. [PubMed: 16818808]
10. Pennant M, Davenport C, Bayliss S, Greenheld W, Marshall T, Hyde C. Community programs for the prevention of cardiovascular disease: A systematic review. *Am J Epidemiol.* 2010; 172(6): 501–16. [PubMed: 20667932]
11. Ebrahim S, Taylor F, Ward K, Beswick A, Burke M, Davey Smith G. Multiple risk factor interventions for primary prevention of coronary heart disease. *Cochrane Database Syst Rev.* 2011; 1(1)
12. Lleras-Muney A. The relationship between education and adult mortality in the united states. *Rev Econ Stud.* 2005; 72(1):189–221.
13. Loucks EB, Lynch JW, Pilote L, Fuhrer R, Almeida ND, Richard H, et al. Life-course socioeconomic position and incidence of coronary heart disease: The framingham offspring study. *Am J Epidemiol.* 2009 Apr 1; 169(7):829–36. [PubMed: 19179358]
14. Kaplan G, Keil J. Socioeconomic factors and cardiovascular disease: A review of the literature. *Circulation.* 1993 Oct 1; 88(4):1973–98. [PubMed: 8403348]
15. González MA, Artalejo FR, Calero JdR. Relationship between socioeconomic status and ischaemic heart disease in cohort and case-control studies: 1960–1993. *Int J Epidemiol.* 1998 Jun 01; 27(3): 350–8. [PubMed: 9698119]
16. Kanjilal S, Gregg EW, Cheng YJ, Zhang P, Nelson DE, Mensah G, et al. Socioeconomic status and trends in disparities in 4 major risk factors for cardiovascular disease among US adults, 1971–2002. *Arch Intern Med.* 2006 Nov 27; 166(21):2348–55. [PubMed: 17130388]
17. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. *N Engl J Med.* 2008 May 22; 358(21):2249–58. [PubMed: 18499567]
18. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *N Engl J Med.* 2007 Jul 26; 357(4):370–9. [PubMed: 17652652]
19. Emmons KM, Barbeau EM, Gutheil C, Stryker JE, Stoddard AM. Social influences, social context, and health behaviors among working-class, multi-ethnic adults. *Health Educ Behav.* 2007 Apr 01; 34(2):315–34. [PubMed: 16740510]
20. Sorensen G, Barbeau E, Hunt MK, Emmons K. Reducing social disparities in tobacco use: A social-contextual model for reducing tobacco use among blue-collar workers. *Am J Public Health.* 2004 Feb 1; 94(2):230–9. [PubMed: 14759932]
21. Emmons, KM. Health behaviors in a social context. In: Berkman, LF.; Kawachi, I., editors. *Social Epidemiology.* New York, NY: Oxford University Press; 2000. p. 242–66.
22. Lusi AJ, Attie AD, Reue K. Metabolic syndrome: From epidemiology to systems biology. *Nat Rev Genet.* 2008; 9(11):819–30. [PubMed: 18852695]
23. Grundy SM, Cleeman JI, Daniels SR, Donato KA, Eckel RH, Franklin BA, et al. Diagnosis and management of the metabolic syndrome: An american heart Association/National heart, lung, and blood institute scientific statement. *Circulation.* 2005 Oct 25; 112(17):2735–52. [PubMed: 16157765]
24. Sundström J, Risérus U, Byberg L, Zethelius B, Lithell H, Lind L. Clinical value of the metabolic syndrome for long term prediction of total and cardiovascular mortality: Prospective, population based cohort study. *BMJ.* 2006 Apr 15; 332(7546):878–82. [PubMed: 16510492]

25. Schuit AJ, van Loon AJM, Tijhuis M, Ocké MC. Clustering of lifestyle risk factors in a general adult population. *Prev Med.* 2002; 35(3):219–24. [PubMed: 12202063]
26. Fine LJ, Philogene GS, Gramling R, Coups EJ, Sinha S. Prevalence of multiple chronic disease risk factors 2001 national health interview survey. *Am J Prev Med.* 2004; 27(2S):18–24. [PubMed: 15275670]
27. Alamian A, Paradis G. Clustering of chronic disease behavioral risk factors in canadian children and adolescents. *Prev Med.* 2009; 48(5):493–9. [PubMed: 19254742]
28. Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, et al. Physical activity and public health. A recommendation from the centers for disease control and prevention and the american college of sports medicine. *JAMA.* 1995 Feb 1; 273(5):402–7. [PubMed: 7823386]
29. Haskell WL, Lee I, Pate RR, Powell KE, Blair SN. Physical activity and public health: Updated recommendation for adults from the american college of sports medicine and the american heart association. *Circulation.* 2007 Aug 28; 116(9):1081–93. [PubMed: 17671237]
30. SP questionnaire component: Physical activity questionnaire data [Internet]. 2005. [updated Novemeber; cited 8/16/2010]. Available from: http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/paq_c.pdf
31. SP questionnaire component: Physical activity individual activities data [Internet]. 2005. [updated Novemeber; cited 8/16/2010]. Available from: http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/paqiaf_c.pdf
32. What we eat in america, NHANES [Internet]. 2010. [updated 8/12; cited 9/18/2010]. Available from: <http://www.ars.usda.gov/Services/docs.htm?docid=13793>
33. NHANES - documentation, codebook, and frequencies - dietary interview - individual foods (first day) - 2003-2004 [Internet]. 2007. [updated 11; cited 9/18/2010]. Available from: http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/dr1iff_c.pdf
34. Food surveys products and services- what we eat in america [Internet]. 2010. [updated 09/29; cited 1/18/2011]. Available from: <http://www.ars.usda.gov/Services/docs.htm?docid=13793>
35. Welsh S, Davis C, Shaw A. Development of the food guide pyramid. *Nutr Today.* 1992; 26(6):12–23.
36. Explaining the rise in educational gradients in mortality. NBER working paper no 15678 [Internet]. 2010. [updated January; cited 8/16/2010]. Available from: <http://www.nber.org/papers/w15678.pdf>
37. Montez JK, Hayward MD, Brown DC, Hummer RA. Why is the educational gradient of mortality steeper for men? *J Gerontol B Psychol Sci Soc Sci.* 2009; 64(5):625–34. [PubMed: 19307285]
38. Galan I, Rodriguez-Artalejo F, Tobias A, Diez-Ganan L, Gandarillas A, Zorrilla B. Clustering of behavior-related risk factors and its association with subjective health. *Gac Sanit.* 2005 Sep-Oct; 19(5):370–8. [PubMed: 16242095]
39. Tobias M, Jackson G, Yeh LC, Huang K. Do healthy and unhealthy behaviours cluster in new zealand? *Aust N Z J Public Health.* 2007; 31(2):155–63. [PubMed: 17461007]
40. Chiolero A, Wietlisbach V, Ruffieux C, Paccaud F, Cornuz J. Clustering of risk behaviors with cigarette consumption: A population-based survey. *Prev Med.* 2006; 42(5):348–53. [PubMed: 16504277]
41. Bernaards CM, Twisk JOSWR, Van Mechelen W, Snel J, Kemper HANCG. A longitudinal study on smoking in relationship to fitness and heart rate response. *Med Sci Sports Exerc.* 2003; 35(5):793–800. [PubMed: 12750589]
42. Gidding SS, Xie X, Liu K, Manolio T, Flack JM, Gardin JM. Cardiac function in smokers and nonsmokers: The CARDIA study. *J Am Coll Cardiol.* 1995 Jul; 26(1):211–6. [PubMed: 7797754]
43. Serdula MK, Gillespie C, Kettel-Khan L, Farris R, Seymour J, Denny C. Trends in fruit and vegetable consumption among adults in the united states: Behavioral risk factor surveillance system, 1994-2000. *Am J Public Health.* 2004; 94(6):1014–8. [PubMed: 15249308]
44. Bennett GG, McNeill LH, Wolin KY, Duncan DT, Puleo E, Emmons KM. Safe to walk? neighborhood safety and physical activity among public housing residents. *PLoS Med.* 2007; 4(10):1599–606. [PubMed: 17958465]

45. von Wagner C, Knight K, Steptoe A, Wardle J. Functional health literacy and health-promoting behaviour in a national sample of british adults. *J Epidemiol Community Health*. 2007 Dec 01; 61(12):1086–90. [PubMed: 18000132]
46. Gilman SE, Martin LT, Abrams DB, Kawachi I, Kubzansky L, Loucks EB, et al. Educational attainment and cigarette smoking: A causal association? *Int J Epidemiol*. 2008 Jun 01; 37(3):615–24. [PubMed: 18180240]
47. Link BG, Phelan J. Social conditions as fundamental causes of disease. *J Health Soc Behav*. 1995; 35:80–94. [PubMed: 7560851]
48. Albouy V, Lequien L. Does compulsory education lower mortality? *J Health Econ*. 2009; 28(1): 155–68. [PubMed: 18976824]
49. Madsen M, Andersen AMN, Christensen K, Andersen PK, Osler M. Does educational status impact adult mortality in denmark? A twin approach. *Am J Epidemiol*. 2010; 172(2):225–34. [PubMed: 20530466]
50. Emmons KM, McBride CM, Puleo E, Pollak KI, Marcus BH, Napolitano M, et al. Prevalence and predictors of multiple behavioral risk factors for colon cancer. *Prev Med*. 2005; 40(5):527–34. [PubMed: 15749134]
51. Biro G, Hulshof K, Ovesen L, Cruz JAA. Selection of methodology to assess food intake. *Eur J Clin Nutr*. 2002; 56:25–32.
52. Troiano RP, Berrigan D, Dodd KW, MÂSSE LC, Tilert T, McDowell M. Physical activity in the united states measured by accelerometer. *Medicine & Science in Sports & Exercise*. 2008; 40(1): 181. [PubMed: 18091006]
53. NHANES 2001-2002 - manuals, brochures, and consent documents [Internet]. 2010. [updated 05/14; cited 8/15/2010]. Available from: http://www.cdc.gov/nchs/nhanes/nhanes2001-2002/current_nhanes_01_02.htm

List of Abbreviations

CI	Confidence Interval
CHD	Coronary Heart Disease
NHANES	National Health and Nutrition Examination Survey
SRI	Sociobehavioral Risk Index

Figure 1. Sociobehavioral risk index (SRI) ratios for the observed/expected prevalence ratio and the randomly permuted observed/expected prevalence ratio (error bars represent 95% confidence intervals) among (A) males and (B) females, NHANES 2001-2004.

Table 1
Characteristics of Study Participants Stratified by Sex, NHANES 2001-2004

	Males (n=4,305)		Females (n=4,673)	
	%	95% CI ^a	%	95% CI
Age				
20-29	17.7	15.0, 20.4	20.5	18.0, 23.1
30-39	15.7	13.0, 18.4	18.4	15.8, 21.0
40-49	18.0	15.3, 20.7	16.4	13.8, 19.0
50-59	13.7	10.9, 16.5	12.1	9.4, 14.8
60-69	14.7	12.0, 17.5	14.7	12.0, 17.3
70-79	12.7	9.9, 15.5	9.9	7.1, 12.6
80	7.4	4.6, 10.3	8.0	5.3, 10.8
Race/Ethnicity				
Non-Hispanic, white	54.8	52.8, 56.9	55.1	50.5, 54.5
Non-Hispanic, black	19.9	17.2, 22.6	19.9	24.6, 29.7
Mexican American/Other Hispanic	25.3	22.6, 27.9	25.0	15.8, 21
Education				
HS Diploma/GED or less Education	54.2	52.1, 56.2	52.5	50.5, 54.5
More than HS Diploma/GED	45.8	43.6, 48	47.5	45.4, 49.6
Current Smoker	27.2	24.6, 29.7	18.4	15.8, 21.0
Did Not Meet Fruit/Vegetable Guidelines^b	51.7	49.6, 53.8	55.4	53.5, 57.3
Did Not Meet Physical Activity Guidelines^c	43.8	41.5, 46.0	51.2	49.2, 53.2

^aConfidence Interval

^b<3 vegetable servings or <2 fruit servings per day.

^c<150 minutes moderate or vigorous physical activity per week.

Table 2
Distribution of Study Sample Characteristics According to Sex and Sociobehavioral Risk Index, NHANES 2001-2004

	Males (n=4,305)					Females (n=4,673)				
	Sociobehavioral Risk Index (%) ^a					Sociobehavioral Risk Index (%)				
	0	1	2	3	4	0	1	2	3	4
Age										
20-34	11.9	27.7	31.6	21.1	7.7	15.0	29.0	31.1	19.8	5.2
35-49	15.8	26.3	28.2	21.0	8.6	13.8	28.3	28.2	23.0	6.8
50-64	19.6	25.5	26.8	21.1	6.9	15.8	27.8	29.8	20.4	6.1
65+	14.9	29.1	31.5	19.3	5.1	11.7	23.9	33.8	28.2	2.5
Race/Ethnicity										
White, NH	20.0	32.5	26.8	15.8	4.8	18.4	29.2	29.2	18.5	4.7
Black, NH	9.7	23.0	32.8	24.9	9.7	9.5	24.6	30.3	28.0	7.6
Mexican American/Other Hispanic	9.1	18.9	32.9	28.9	10.1	7.0	24.8	34.9	29.2	4.1
Current Smoker										
Yes		8.6	27.5	37.7	26.1		7.8	24.7	39.8	27.7
No	21.1	34.2	30.5	14.2		17.2	31.7	32.1	18.9	
Education										
High School Diploma/GED		16.3	36.8	33.7	13.1		14.4	36.4	39.5	9.7
>High School	33.6	40.1	21.2	5.1		29.6	41.6	24.6	4.2	
Did not meet Physical Activity Guidelines										
Yes		13.8	33.7	36.4	16.2		15.5	37.2	37.4	10.0
No	27.3	37.7	26.6	8.3		28.8	39.7	24.1	7.5	
Did not meet Fruit/Vegetable Guidelines										
Yes		19.4	33.3	33.6	13.7		18.7	34.1	38.0	9.2
No	31.8	35.7	25.8	6.7		31.5	38.0	26.6	3.8	

^aThe sociobehavioral risk index is the sum of the number of risk factors each person has accumulated. Risk factors include low-educational attainment (High School Diploma/GED), current smoking, not meeting fruit/vegetable guidelines (<3 vegetable servings or <2 fruit servings per day), and not meeting physical activity guidelines (<150 minutes moderate or vigorous physical activity per week).

Table 3
Observed and Randomly Permuted Clustering of Social and Behavioral Risk Factors, Sex Stratified, NHANES 2001-2004

SRI	Education ^a	Physical Activity ^b	Fruit and Vegetables ^c	Current Smoker ^d	Observed Risk Factor Combination Prevalence	Observed SRI Prevalence	Expected and Permuted SRI Prevalence	Ratio of Observed/Expected SRI Prevalence	Permuted Mean Ratio of Observed/Expected SRI 95% CI ^e
Males (n=4,305)									
0	f	-	-	-	15.4%	15.4%	9.1%	1.70	1.00 0.92, 1.08
	+	-	-	-	8.9%				
1	-	+	-	-	6.0%	27.3%	30.9%	0.88	1.00 0.96, 1.04
	-	-	+	-	10.0%				
	-	-	-	+	2.4%				
	+	+	-	-	7.8%				
	+	-	+	-	9.1%				
2	+	-	-	+	3.0%	29.7%	37.6%	0.79	1.00 0.96, 1.04
	-	+	+	-	5.3%				
	-	+	-	+	1.6%				
	-	-	+	+	2.8%				
	+	+	+	-	10.4%				
3	+	+	-	+	3.2%	20.6%	19.1%	1.08	1.00 0.95, 1.05
	+	-	+	+	4.7%				
	-	+	+	+	2.3%	7.1%	3.3%	2.10	1.00 0.86, 1.14
	+	+	+	+	7.1%				
Females (N=4,673)									
0	-	-	-	-	14.1%	14.1%	8.4%	1.67	1.00 0.92, 1.08
	+	-	-	-	7.6%				
1	-	+	-	-	7.9%	27.3%	30.6%	0.89	1.00 0.96, 1.04
	-	-	+	-	10.4%				
	-	-	-	+	1.4%				
	+	+	-	-	9.7%				
	+	-	+	-	8.2%				
2	+	-	-	+	1.2%	30.8%	38.8%	0.79	1.00 0.96, 1.03

SRI	Education ^a	Physical Activity ^b	Fruit and Vegetables ^c	Current Smoker ^d	Observed Risk Factor Combination Prevalence	Observed SRI Prevalence	Expected and Permutated SRI Prevalence	Ratio of Observed/Expected SRI Prevalence	Permutated Mean Ratio of Observed/Expected SRI 95% CI ^e
	-	+	+	-	8.4%				
	-	+	-	+	1.0%				
	-	-	+	+	2.4%				
	+	+	+	-	15.4%				
3	+	+	-	+	1.7%	22.8%	19.4%	1.17	1.00
	+	-	+	+	3.6%				0.96, 1.04
	-	+	+	+	2.0%				
4	+	+	+	+	5.1%	5.1%	2.7%	1.86	1.00
									0.85, 1.15

SRI, Sociobehavioral Risk Index

^a High School/GED vs. >High School;

^b Met vs. Did not meet Physical Activity Guidelines (i.e. 150 vs. <150 minutes moderate or vigorous physical activity per week);

^c Met vs. Did not meet Fruit/Vegetable Guidelines (i.e. 3 vegetable servings and 2 fruit servings per day vs. <3 vegetable servings or <2 fruit servings per day);

^d Current vs. Never or Former Smokers;

^e 95% Confidence Interval;

^f +=has risk factor, - = does not have risk factor.

Author Manuscript

 $c_{95\% \text{ Confidence interval}}$

Table 5
Observed (O) Versus Expected (E) Sociobehavioral Risk Index (SRI) Ratios, and Randomly Permuted Mean Ratios of Observed Versus Expected SRI, Stratified by Sex and Age, NHANES 2001-2004

Male												
20-34 (n=1,102)			35-49 (n=1,112)			50-64 (n=924)			65+ (n=1,167)			
SRI	O/E ^a	Permuted Mean O/E ^b (95% CI) ^c	O/E	Permuted Mean O/E (95% CI)	O/E	Permuted Mean O/E (95% CI)	O/E	Permuted Mean O/E (95% CI)	O/E	Permuted Mean O/E (95% CI)		
0	1.50	1.00	0.84, 1.16	1.80	1.00	0.84, 1.15	1.88	1.00	0.86, 1.15	1.65	1.00	0.85, 1.15
1	0.96	1.00	0.92, 1.08	0.88	1.00	0.92, 1.08	0.79	1.00	0.92, 1.08	0.90	1.00	0.93, 1.07
2	0.84	1.00	0.93, 1.08	0.76	1.00	0.93, 1.07	0.73	1.00	0.92, 1.08	0.80	1.00	0.93, 1.07
3	1.00	1.00	0.91, 1.09	1.05	1.00	0.91, 1.09	1.20	1.00	0.89, 1.11	1.10	1.00	0.91, 1.10
4	1.80	0.99	0.74, 1.25	2.20	1.00	0.73, 1.26	2.33	1.00	0.68, 1.32	2.89	1.00	0.61, 1.39
Female												
20-34 (n=1,418)			35-49 (n=1,167)			50-64 (n=949)			65+ (n=1,139)			
SRI	O/E	Permuted Mean O/E (95% CI)	O/E	Permuted Mean O/E (95% CI)	O/E	Permuted Mean O/E (95% CI)	O/E	Permuted Mean O/E (95% CI)	O/E	Permuted Mean O/E (95% CI)		
0	1.54	1.00	0.88, 1.12	1.67	1.00	0.85, 1.16	1.71	1.00	0.84, 1.16	1.99	0.99	0.79, 1.19
1	0.90	1.00	0.94, 1.06	0.95	1.00	0.93, 1.08	0.88	1.00	0.92, 1.08	0.86	1.00	0.93, 1.08
2	0.83	1.00	0.94, 1.06	0.74	1.00	0.93, 1.07	0.78	1.00	0.92, 1.08	0.79	1.00	0.94, 1.06
3	1.12	1.00	0.91, 1.09	1.14	1.00	0.91, 1.09	1.12	1.00	0.90, 1.10	1.25	1.00	0.92, 1.08
4	1.96	1.00	0.71, 1.29	1.94	1.00	0.75, 1.26	2.36	1.00	0.65, 1.35	1.80	0.99	0.56, 1.43

^aRatio of observed/expected SRI prevalence

^bRandomly permuted mean ratio of observed/expected SRI (95% CI)

^c95% Confidence interval