Published in final edited form as:

Ann Epidemiol. 2018 March ; 28(3): 175–181.e4. doi:10.1016/j.annepidem.2018.01.001.

A demographic, clinical and behavioral typology of obesity in the United States: an analysis of NHANES 2011-2012

Marcia P. Jimenez, MSc, MA^a, Mark A. Green, PhD^b, S.V. Subramanian, PhD^c, and Fahad Razak, MD, MSc^{d,*}

^aDepartment of Epidemiology, Brown School of Public Health, Brown University, Providence RI 02912 USA

^bDepartment of Geography & Planning, University of Liverpool, Liverpool, UK

^cDepartment of Social and Behavioral Sciences, Harvard School of Public Health, Boston MA 02115 USA

^dLi Ka Shing Knowledge Institute of St Michael's Hospital, Toronto, Ontario, Canada; Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada; Harvard Centre for Population and Development Studies, Harvard University, Cambridge, MA

Abstract

Purpose—Public health reporting, randomized trials and epidemiologic studies of obesity tend to consider it as a homogeneous entity. However, obesity may represent a heterogeneous condition according to demographic, clinical and behavioral factors. We assessed the heterogeneity of individuals with obesity in the United States.

Methods—We analyzed data from the 2011-2012 wave of the National Health and Nutrition Examination Survey, a nationally representative sample of adults in the U.S. with detailed physical examination and clinical data (N=1 380). We used cluster analysis to identify sub-groups classified as obese according to demographic factors, clinical conditions and behavioral characteristics.

Results—We found significant heterogeneity among participants with obesity according to 6 distinct clusters ($p < 0.001$): Affluent men with sleep disorders (16% of sample); Elderly smokers with CVD (16%); Older women with poor mental health (20%); Healthy White women (13%); Healthy Non-White women (14%); and Active men who drink higher amounts of alcohol (21%).

Conclusions—Obesity in the U.S. is not a homogeneous condition. Current research and treatment may fail to account for complex and inter-related factors, with implications for prevention strategies and diverse risks of obesity.

Keywords

Obesity; Body mass index; Cluster Analysis; Population Heterogeneity

*Corresponding author.

INTRODUCTION

Obesity (body mass index - BMI ≥ 30 kg/m²) is a serious public health challenge and an important risk factor for chronic diseases such as arthritis, diabetes, hypertension (HTN), dyslipidemia, cardiovascular disease (CVD) and cancer (1). The American Medical Association recently classified obesity as a disease (2). However, a conceptualization of individuals with obesity as defined solely by BMI, may not reflect the true heterogeneity encountered in clinical practice (3). A better understanding of the heterogeneity of individuals with obesity may help public health practitioners, researchers and policymakers to establish tailored and appropriate goals for obesity treatment, and better design interventions and clinical trials around prevention and treatment.

There is an increasing recognition of obesity as a heterogeneous disease and the implications on chronic disease risk. For instance, the concept of the metabolically healthy obese suggests variability in cardiovascular disease risk factors and mortality risk beyond what is captured by BMI alone. Results from this type of research suggest that only obese and unfit individuals, but not obese and fit individuals, are at higher mortality risk than normal weight and fit individuals (3, 4). One study of weight loss maintenance identified four clusters that differed in terms of demographic characteristics, weight and health history, as well as weight-loss and weight-maintenance strategies and attitudes (5). A recent study examined the existence of subgroups of individuals with obesity in the Yorkshire region of England and found six distinct groups according to demographic, health and behavioral factors (3). Finally, recent studies also found obesogenic cluster patterns in children and adolescents with mixed physical activity/sedentary behaviors differing according to age, gender and socio-economic status (6, 7). The importance of modifiable factors in obesity etiology such as physical activity, diet, smoking and alcohol consumption is undeniable. However, there is evidence to suggest that many individuals do not meet recommendations from health providers, with multiple obesity risk behaviors often occurring together (6). The clustering of individuals with similar characteristics is a concept that has been applied to understanding the association among different health behaviors (6, 8, 9). However, there has been no previous investigation of the population level existence of clusters of health behaviors, demographic characteristics and clinical factors among adults classified as obese in the United States (U.S.). We employ this explorative research method to better understand dissimilarities among adults with obesity. The objective of this study is to use data-driven methodologies to discover and develop an understanding of the types of adult individuals classified as obese according to demographic, clinical and behavioral factors that have been demonstrated across a wide range of literature in multiple fields to be associated to and with obesity. To our knowledge, this is the first study to use the most comprehensive available dataset to look at heterogeneity in obesity at the population level in the U.S. using detailed clinical objective variables as well as self-reported characteristics.

SUBJECTS AND METHODS

Participants

We use data from the National Health and Nutrition Examination Survey (NHANES) from 2011-2012. NHANES uses a stratified multistage sampling design to produce data

representative of the U.S. civilian noninstitutionalized population (10). A detailed description of the methodology and procedures used in the survey has been previously published (11). A unique feature of NHANES is the collection of objective physical examination data for each respondent in the sample, which is carried out in mobile examination centers for standardization purposes. The 2011-2012 wave of NHANES collected data on 9 756 individuals, and we restricted our analysis to the 2 081 (21% of full sample) who were classified as obese using World Health Organization cut-offs for BMI (i.e. 30 kg/m^2). BMI values are calculated for NHANES participants using measured height and weight (12, 13). The analyses presented in this report included adult participants ≥ 20 years of age with obesity to ensure comparability to prior studies (14), giving a sample size of 1 873 (90% of participants classified as obese). Due to missing data in the covariates, the final analytical sample size was 1 380 (74% of intended sample, Supplementary Material).

Measures

The selection of the input variables was based on subject-matter knowledge and prior work focused on identifying types of individuals with obesity (3), to explore whether this approach is useful in the U.S. Specifically, we include: demographic indicators to assess the vulnerability of being obese among minority and low-socioeconomic-status groups. For example, we theorize that race/ethnicity plays a role in obesity typology through individual factors such as socio-economic status, as well as community-level characteristics such as the accessibility to recreational facilities. Behavioral factors we also added to evaluate how health behaviors cluster across individuals diagnosed with obesity; and respiratory, heart, circulatory and mental health variables to study individuals with certain biological predispositions. Prevalence of chronic conditions examined included asthma, arthritis, congestive heart failure, coronary heart disease, angina, heart attack, stroke, emphysema, cancer, pain score, anxiety score, diabetes, HTN, dyslipidemia, sleep disorder and depression score. Most chronic conditions were identified through self-report of previous diagnosis by a health care provider, except for dyslipidemia, HTN and diabetes, which were defined as either self-reported diagnosis or currently taking medication (15, 16). Well-being measures included anxiety, depression and pain. Anxiety was assessed based on how many days did the participant report feeling anxious during the past 30 days. Depression was assessed through the Patient Health Questionnaire (PHQ-9), a nine-item screening instrument that asked questions about the frequency of symptoms of depression over the past 2 weeks (17). Pain was assessed by the question “During the past 30 days, for about how many days did pain make it hard for you to do your usual activities, such as self-care, work, or recreation?”. Sleep disorder was also included as potential comorbidity of obesity (18). Finally, behavioral characteristics such as physical activity, diet, alcohol consumption and smoking were included. Physical activity was measured based on the Global Physical Activity Questionnaire as recommended by the World Health Organization (WHO), which incorporates person’s overall energy expenditure in moderate activities and in vigorous activities, where a higher score indicates higher level of physical activity (19). Diet was measured according to the Healthy Eating Index (HEI) which has been validated using NHANES data and has been shown to strongly predict risk of chronic disease (20). Alcohol consumption was split into 4 categories (nondrinkers, <1 drink per day, 1-2 drinks per day, and >2 drinks per day) (21). Smoking was assessed by reporting having smoked >100

cigarettes in life (22). A further detailed description of the variables used in this analysis is included in the Supplementary Material (SM).

We replicated the analysis in the prior wave of NHANES data from 2009-2010 to test for the robustness of our results. A detailed description of the replication analysis can be found in the SM (pp.7-10).

Statistical Analysis

Cluster analysis approaches have been useful in previous research at identifying groupings within data when there is no known structure to the data (3, 8, 9, 23). The approach takes individual-level data and groups individuals based on their similarities across multiple factors (as well as dissimilarities to the other groups).

Since our data contained a mixture of binary, categorical and continuous variables, conventional unsupervised classification methods were not applicable as they can only deal with a single data type within a model (23). To account for these multiple data types, we used a two-step cluster analysis approach (24). The approach first scans the entire data and merges cases that share similar values across their variables. The process reduces the data into a smaller set of 'dense' regions (known as 'cluster features') which reflect the main patterns. These cluster features are then used as the inputs to be clustered using an agglomerative hierarchical algorithm. The algorithm operates through first identifying the two most similar cluster features and joining them together into a single group. It then iteratively repeats this process until all cluster features or groups have been merged together into a single group. The different combination of groupings can then be evaluated to see which best describes the data. As the algorithm is only analysing the cluster features rather than the original data, it makes the algorithm efficient at processing large data sets. The log-likelihood is used to measure 'distance' (i.e. the similarity of variables) between cases to account for the different data types (24). Continuous variables were standardized using z-scores due to their differing scales (23).

To determine the number of clusters, we calculated two measures: the Schwartz's Bayesian Information Criterion (BIC) which is a measure of model fit, and the Silhouette measure which is a measure of cluster cohesion and separation. Figure 1 presents the change in BIC across a range of cluster solutions from 2 to 15. As the gradient of the change in BIC begins to level off at a 5 or 6 cluster solution, the plot suggests that subsequent solutions offer less relative information while increasing the complexity in the number of clusters. From Figure 1, it is clear that a five or six cluster solution would be most appropriate.

The silhouette measure (SM Figure S2) suggested similar findings. We selected a six cluster solution since it marginally performed better and the additional cluster adds detail. We have included the results of the five cluster solution in the supplement (SM Table S1) as well as a comparison table between both solutions (SM Table S2).

We used the dietary day one sample weights to ensure that our analyses remained nationally representative, as recommended by NHANES methodology (25). As our analytical approach does not deal directly with sample weights, we propagated our data set to represent the

sample weights, which altered our sample size from 9,756 to 2,239,682; with a sample of 560,928 cases defined as obese and with no missing data (further explanation of the use of sample weights in cluster analysis is provided in the SM page 6). All analyses were conducted in SPSS v.22.

RESULTS

Table 1 presents the weighted sample characteristics of individuals who were obese. The analytical database consisted of 560,928 men and women aged 20 years and older who were obese and had complete data for all variables (more detailed information can be found in the data flow chart in the SM).

Table 2 presents the results from the cluster analysis for the six cluster solution for the weighted sample, where cells in red represent values higher than the population mean, and cells in green represent values that are lower than the average within the same row. We performed statistical tests (ANOVA and Chi² tests) on the characteristics of individuals in the clusters in both the propagated and original data (using sample weights), and the results demonstrated that the clusters were significantly different for each variable (Table 2).

Mixed-sex cluster

“Elderly individuals who smoke with CVD” (16.3% of the population with obesity). The cluster is characterized by the complex list of comorbidities with the highest prevalence of the heart-related conditions, emphysema, diabetes and cancer. This cluster has the highest prevalence of smoking and the second lowest mean of HEI.

Female-Dominant Clusters

The cluster “Older women with poor mental health” (20.4%) is characterized by the lowest score for physical activity but the highest mean HEI. Health conditions are clustered around pain, anxiety, depression, hypertension, stroke, asthma and arthritis. The mean age of the participants in this cluster is the second highest and the mean BMI is the highest, compared to other clusters.

The cluster “Healthy White Women” (12.9%) consists mostly of females (90.8%) and has the highest prevalence of Whites compared to other female-dominant clusters. This cluster is characterized by its low prevalence of health conditions compared to the other clusters (particularly heart and circulatory related conditions), healthy behaviors (non-drinkers) and has a high mean HEI.

The cluster “Healthy Non-White Women” is the second smallest cluster (13.9%) and also consists mostly of females (77.5%). The individuals in this cluster are mostly Non-White, and have the lowest income. This cluster is characterized by good cardiovascular and respiratory health. The mean age of the participants in this cluster suggest that they are the youngest, compared to other clusters.

Male-Dominant Clusters

The cluster “Affluent men with sleep disorder” (15.6%) has the highest income. The demographic profile of the cluster is mostly White and middle-aged. They have low levels of anxiety and depression, and the highest burden of sleep disorder. About two thirds (59.9%) of the individuals in this cluster reported smoking. The individuals in this cluster had the second lowest score for physical activity.

Finally, “Active men who drink higher amounts of alcohol” (20.9%) has the highest prevalence of Mexican-Americans, low prevalence of all health conditions, low mean BMI, but highest prevalence of alcohol intake. The individuals in this cluster have the highest score for physical activity and the mean age suggest that they are the youngest, compared to other individuals in other clusters. It is important to note that the younger clusters also had the widest range of ages.

The replication analysis in the 2009-2010 wave of NHANES to establish the stability of clusters over time, also suggested that a six cluster solution was optimal (SM Table S3). The resulting clusters for the 2009-2010 wave shared similar patterns with our main results from 2011-2012. Most notably, the main health and behavioural clusters (i.e. a cluster of women with mental health disorders, a cluster of people who smoke with poor cardiovascular health, a cluster of men who drink high amounts of alcohol but are physically active, a cluster of affluent men and 2 clusters of women in good health) suggesting stability of the underlying clusters within the population over time.

DISCUSSION

Our findings suggest that obesity is not a homogeneous disease in the U.S. and that a BMI based definition of obesity may be limited. Our study shows that obesity patterns fall into six clusters: Affluent men with sleep disorder; Elderly individuals who smoke with CVD; Older women with poor mental health; Healthy White women; Healthy Non-White women; Active men who drink higher amounts of alcohol. The clusters were significantly different according to key demographic characteristics, clinical conditions, and behavioral factors. The clusters were replicated in the previous wave of NHANES data suggesting a stable underlying heterogeneity among people with obesity in the United States. These findings underscore the importance of the complex interaction of demographic, clinical and behavioral factors that intersect with obesity and are the first step towards future clinical trials to test specific tailored interventions. The relevance of our study is that public health reporting and epidemiologic studies on obesity may potentially benefit from stratifying by subtypes or by assessing different risks across distinctive groups of people, which is not the current practice (26). In particular, the significance of this type of research for public health is that identifying which behaviors need to be targeted simultaneously and in whom, will help to develop cost-effective targeted obesity prevention initiatives to those most in need. Since cluster analysis does not differentiate between dependent and independent variables, a full examination of interdependent relationships can be done simultaneously (5). This unique methodology allowed us to include factors shown to be causes of obesity, as well as factors known to be consequences of obesity, and examine the full range of public health burdens related to obesity.

Previous studies have shown that BMI does not differentiate between fat mass and lean muscle mass, or their distribution (27, 28). However, less research has been done on the heterogeneity among those classified as obese and the consequences for morbidity and prevention. Our results demonstrate the marked heterogeneity of individuals classified as obese suggesting a critical need to account for demographic, clinical and behavioral variables in obesity research and obesity treatment guidelines. To our knowledge, this is the first analysis of obesity patterns including demographic clinical and behavioral factors in adults in the U.S. The differences found among the clusters are important to understand in order to best tailor effective future strategies in response to the high levels of obesity in the U.S. Moreover, the clusters highlight important differences that would not be captured by a univariate analysis (e.g. age) alone and show a more complex perspective on individuals classified as obese. For example, we hypothesize that the two clusters of healthy women (Healthy White women and Healthy Non-White women) would likely benefit from health interventions for BMI reduction, since they could potentially have substantial life-time benefits from reduction in BMI. However, it is possible that women in this cluster may have higher BMI because they have higher muscle mass, which is not differentiated from fat mass based alone on BMI. Similarly, it would be ideal to intervene on BMI reduction in the cluster of “Active men who drink higher amounts of alcohol”, but in addition, we believe that important behavioral risk factors such as alcohol intake and smoking should be targeted and may be a higher initial priority for intervention given their likely stronger association with mortality and morbidity (29-31). In contrast, we hypothesize that although “Older women with poor mental health” would almost certainly benefit from BMI reduction, challenges with mental disorders and low levels of physical activity may represent crucial areas for intervention before addressing BMI reduction. Similarly, in “Elderly individuals who smoke with CVD” reducing BMI might be a likely distant goal compared to other more pressing medical issues. Furthermore, it is not even clear that individuals at an advanced age with high levels of comorbidities would receive tangible benefits by reducing BMI (32). Even though the evaluation of strategies across the sub-groups with obesity was not possible using our observational dataset, earlier studies have suggested that weight-loss strategies may be better tailored according to specific characteristics of groups of individuals (5, 33). This type of analysis provides a basis for upcoming studies to evaluate whether diverse strategies may be needed in the analysis of obesity.

Our results also shed light on an important gap in the way that obesity interventions and obesity guidelines are developed. It is important to note that while our results do not offer direct evidence that different treatment options are required for each cluster, the heterogeneity in obesity revealed is an important step in beginning the discussion towards this important area. The variety among clusters revealed may indicate that current guidelines and clinical trials that fail to consider complex and inter-related factors (34-36) would benefit from accounting for this heterogeneity.

Some of the clusters we observed represent known associations of obesity with other diseases. For example, the clusters of “Elderly individuals who smoke with CVD” and “Affluent men with sleep disorder” support the well documented association of obesity with chronic diseases (37-42). The clustering of high BMI in “Older women with poor mental health” with high levels of pain, anxiety, depression, asthma and stroke reflects the

association of obesity with depression (43), asthma (44) and the increase in women's stroke prevalence (45). However, in a cross-sectional study, it is impossible to know the directional association between comorbidities, behavioral factors and obesity, with each likely re-enforcing and worsening the other – a phenomenon known as the “Runaway Weight-Gain Train”(46).

This analysis helps to better understand population level differences of individuals classified as obese in the U.S., and draws some similarities to recent analysis in the United Kingdom. Green et al. studied the clustering of obesity characteristics in the Yorkshire region of England and also found six clusters, among which were “Poorest health”, “Unhappy anxious middle aged”, “Younger healthy females”, “Heavy drinking males” (3). These clusters are comparable to our “Elderly individuals who smoke with CVD”, “Older women with poor mental health”, “Healthy White women” or “Healthy Non-White women”, and “Active men who drink higher amounts of alcohol” respectively. The other two clusters found in England were “Physically sick but happy elderly” and “Affluent and healthy elderly” (*Ibid*). This strongly supports the presence of some degree of underlying generalizability to our observed obesity clusters given that Green's data used a different list of comorbidities, demographic and mental health measures, and that American and English populations have markedly different demographic and racial-ethnic composition. These differences may explain why the cluster of “Affluent men with sleep disorder” was found in the U.S. population but not England. In addition, our analysis uses data from a nationally representative population with objective physical examination data which allows us to extrapolate the results to the overall adult population classified as obese in the U.S., as opposed to restricting the interpretation of results based on self-reported data from participants in a region of a country.

There are several limitations to our study. First, due to the nature of this approach we cannot test for the robustness of our findings (23). However, the reproducibility of the clusters in the two most recent waves of complete data in NHANES, strengthen the stability of the clusters. Nevertheless, it is important for future research to explore whether these clusters can be replicated in other datasets. Second, the final decision of the number of clusters identified can often be subjective (23). To account for this, we looked for concordance between two measures of solution quality, and evaluated two solutions with different numbers of clusters suggesting stable results. While some solutions produced marginally stronger solutions by these metrics, we opted to select a parsimonious solution since a large number of clusters could hinder model interpretability and may not represent much of a data reduction (23). Third, the lack of longitudinal data did not allow us to address whether there are differences in course for the clusters. Furthermore, we were not able to test the stability of cluster membership over time, and identify sociodemographic predictors of cluster membership and cluster transition over time (47). Our team is currently working on an expansion of this study using long-term follow-up. Finally, we did not have information on medications which may have led to weight gain, such as oral hypoglycemic or use of inflammation suppressing agents such as steroids.

This study indicates that there was significant heterogeneity among adults with obesity in the U.S. according to demographic, clinical and behavioral factors. Our findings on the variability among individuals with obesity raise the important question of whether analyses

and treatment on obesity, and potential policy approaches for obesity prevention need to account for this heterogeneity. These findings and their implications require testing in rigorous prospective studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: This work was supported by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development of the National Institutes of Health [grant number T32HD007338] to MPJ, and the Robert Wood Johnson Foundation seed grant, “Analyzing Changes in Chronic Disease Risk Factor Distributions in the United States” to FR and SVS.

List of Abbreviations and Acronyms

BMI	Body Mass Index
NHANES	National Health and Nutrition Examination Survey
BIC	Bayesian Information Criterion
CVD	Cardiovascular Disease

References

1. Mokdad AH, Ford ES, Bowman BA, Dietz WH, Vinicor F, Bales VS, et al. Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. *Jama*. 2003; 289(1):76–9. [PubMed: 12503980]
2. Stoner L, Cornwall J. Did the American Medical Association make the correct decision classifying obesity as a disease? *Australas Med J*. 2014; 7(11):462–4. [PubMed: 25550718]
3. Green MA, Strong M, Razak F, Subramanian SV, Relton C, Bissell P. Who are the obese? A cluster analysis exploring subgroups of the obese. *Journal of public health*. 2015; 38(2):258–64. [PubMed: 25889387]
4. Stefan N, Haring HU, Hu FB, Schulze MB. Metabolically healthy obesity: epidemiology, mechanisms, and clinical implications. *Lancet Diabetes Endocrinol*. 2013; 1(2):152–62. [PubMed: 24622321]
5. Ogden LG, Stroebele N, Wyatt HR, Catenacci VA, Peters JC, Stuhlt J, et al. Cluster analysis of the national weight control registry to identify distinct subgroups maintaining successful weight loss. *Obesity*. 2012; 20(10):2039–47. [PubMed: 22469954]
6. Leech RM, McNaughton SA, Timperio A. The clustering of diet, physical activity and sedentary behavior in children and adolescents: a review. *The international journal of behavioral nutrition and physical activity*. 2014; 11:4. [PubMed: 24450617]
7. Ferrar K, Golley R. Adolescent diet and time use clusters and associations with overweight and obesity and socioeconomic position. *Health education & behavior : the official publication of the Society for Public Health Education*. 2015; 42(3):361–9. [PubMed: 25576499]
8. Pronk NP, Anderson LH, Crain AL, Martinson BC, O'Connor PJ, Sherwood NE, et al. Meeting recommendations for multiple healthy lifestyle factors. Prevalence, clustering, and predictors among adolescent, adult, and senior health plan members. *American journal of preventive medicine*. 2004; 27(2 Suppl):25–33. [PubMed: 15275671]
9. Schuit AJ, van Loon AJ, Tijhuis M, Ocke M. Clustering of lifestyle risk factors in a general adult population. *Preventive medicine*. 2002; 35(3):219–24. [PubMed: 12202063]

10. Centers for Disease Control and Prevention NCHS. , editor. National Health and Nutrition Examination Survey (NHANES). http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm2011-2012
11. NHANES 2011-2012 Overview: Centers for Disease Control and Prevention. Centers for Disease Control and Prevention; 2011. [Available from: http://www.cdc.gov/nchs/nhanes/nhanes2011-2012/overview_g.htm
12. NHANES 2011 - 2012 Data Documentation, Codebook, and Frequencies - Body Measures: Centers for Disease Control and Prevention 2015. [updated September 2015. Available from: http://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/BMX_G.htm#Component_Description
13. Clinical Trial Registration for Obesity Heterogeneity Study [Internet]. 2017. Available from: <https://clinicaltrials.gov/ct2/show/NCT03155880?term=NCT03155880&rank=1>
14. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011-2012. *Jama*. 2014; 311(8):806–14. [PubMed: 24570244]
15. Sarathy H, Henriquez G, Abramowitz MK, Kramer H, Rosas SE, Johns T, et al. Abdominal Obesity, Race and Chronic Kidney Disease in Young Adults: Results from NHANES 1999-2010. *PLoS one*. 2016; 11(5):e0153588. [PubMed: 27224643]
16. Wu B, Bell K, Stanford A, Kern DM, Tunceli O, Vupputuri S, et al. Understanding CKD among patients with T2DM: prevalence, temporal trends, and treatment patterns-NHANES 2007-2012. *BMJ Open Diabetes Res Care*. 2016; 4(1):e000154.
17. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire*. *Jama*. 1999; 282(18):1737–44. [PubMed: 10568646]
18. Okifuji A, Donaldson GW, Barck L, Fine PG. Relationship between fibromyalgia and obesity in pain, function, mood, and sleep. *J Pain*. 2010; 11(12):1329–37. [PubMed: 20542742]
19. Surveillance and Population-Based Prevention PoNDD Global Physical Activity Questionnaire (GPAQ) Analysis Guide. World Health Organization;
20. Chiuve SE, Fung TT, Rimm EB, Hu FB, McCullough ML, Wang M, et al. Alternative dietary indices both strongly predict risk of chronic disease. *J Nutr*. 2012; 142(6):1009–18. [PubMed: 22513989]
21. Wildman RP, Muntner P, Reynolds K, McGinn AP, Rajpathak S, Wylie-Rosett J, et al. The obese without cardiometabolic risk factor clustering and the normal weight with cardiometabolic risk factor clustering: prevalence and correlates of 2 phenotypes among the US population (NHANES 1999-2004). *Archives of internal medicine*. 2008; 168(15):1617–24. [PubMed: 18695075]
22. Kim KN, Choi YH, Lim YH, Hong YC. Urinary phthalate metabolites and depression in an elderly population: National Health and Nutrition Examination Survey 2005-2012. *Environmental research*. 2016; 145:61–7. [PubMed: 26624239]
23. Everitt, BLS., Leese, M., Stahl, D. *Cluster Analysis*. Wiley; 2001.
24. Chiu, T., Fang, D., Chen, J., Wang, Y., Jeris, C. Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. San Francisco, California. 502549: ACM; 2001. A robust and scalable clustering algorithm for mixed type attributes in large database environment; p. 263-8.
25. NHANES N. NHANES Weighting Module Centers for Disease Control and Prevention 2013. [updated May 10, 2013. Available from: <http://www.cdc.gov/nchs/tutorials/NHANES/SurveyDesign/Weighting/intro.htm>
26. Kirk SF, Penney TL, McHugh TL, Sharma AM. Effective weight management practice: a review of the lifestyle intervention evidence. *International journal of obesity*. 2012; 36(2):178–85. [PubMed: 21487396]
27. Gomez-Ambrosi J, Silva C, Galofre JC, Escalada J, Santos S, Millan D, et al. Body mass index classification misses subjects with increased cardiometabolic risk factors related to elevated adiposity. *International journal of obesity*. 2012; 36(2):286–94. [PubMed: 21587201]
28. Bastien M, Poirier P, Lemieux I, Despres JP. Overview of epidemiology and contribution of obesity to cardiovascular disease. *Prog Cardiovasc Dis*. 2014; 56(4):369–81. [PubMed: 24438728]

29. Flegal KM, Kit BK, Orpana H, Graubard BI. Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis. *Jama*. 2013; 309(1):71–82. [PubMed: 23280227]
30. Stahre M, Roeber J, Kanny D, Brewer RD, Zhang X. Contribution of excessive alcohol consumption to deaths and years of potential life lost in the United States. *Prev Chronic Dis*. 2014; 11:E109. [PubMed: 24967831]
31. Jha P, Ramasundarahettige C, Landsman V, Rostron B, Thun M, Anderson RN, et al. 21st-century hazards of smoking and benefits of cessation in the United States. *The New England journal of medicine*. 2013; 368(4):341–50. [PubMed: 23343063]
32. Oreopoulos A, Kalantar-Zadeh K, Sharma AM, Fonarow GC. The obesity paradox in the elderly: potential mechanisms and clinical implications. *Clin Geriatr Med*. 2009; 25(4):643–59. viii. [PubMed: 19944265]
33. Sacks FM, Bray GA, Carey VJ, Smith SR, Ryan DH, Anton SD, et al. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. *The New England journal of medicine*. 2009; 360(9):859–73. [PubMed: 19246357]
34. Jensen MD, Ryan DH, Apovian CM, Ard JD, Comuzzie AG, Donato KA, et al. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. *J Am Coll Cardiol*. 2014; 63(25 Pt B):2985–3023. [PubMed: 24239920]
35. The Practical Guide. Identification, Evaluation, and Treatment of Overweight and Obesity in Adults. NHLBI Obesity Education Initiative National Heart, Lung, and Blood Institute; 2000.
36. AACE 2015 guidelines Obesity Management: National Diabetes Education Initiative. 2015. [Available from: <http://www.ndei.org/AACE-guidelines-diabetes-obesity-treatment-algorithm.aspx>]
37. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *The New England journal of medicine*. 2003; 348(17):1625–38. [PubMed: 12711737]
38. Colditz GA, Willett WC, Rotnitzky A, Manson JE. Weight gain as a risk factor for clinical diabetes mellitus in women. *Ann Intern Med*. 1995; 122(7):481–6. [PubMed: 7872581]
39. Eckel RH, Krauss RM. American Heart Association call to action: obesity as a major risk factor for coronary heart disease. AHA Nutrition Committee. *Circulation*. 1998; 97(21):2099–100. [PubMed: 9626167]
40. Kenchaiah S, Evans JC, Levy D, Wilson PW, Benjamin EJ, Larson MG, et al. Obesity and the risk of heart failure. *The New England journal of medicine*. 2002; 347(5):305–13. [PubMed: 12151467]
41. Onat A, Hergenc G, Yuksel H, Can G, Ayhan E, Kaya Z, et al. Neck circumference as a measure of central obesity: associations with metabolic syndrome and obstructive sleep apnea syndrome beyond waist circumference. *Clin Nutr*. 2009; 28(1):46–51. [PubMed: 19010573]
42. Hargens TA, Kaleth AS, Edwards ES, Butner KL. Association between sleep disorders, obesity, and exercise: a review. *Nat Sci Sleep*. 2013; 5:27–35. [PubMed: 23620691]
43. Johnston E, Johnson S, McLeod P, Johnston M. The relation of body mass index to depressive symptoms. *Canadian journal of public health = Revue canadienne de sante publique*. 2004; 95(3): 179–83. [PubMed: 15191118]
44. Patra J, Maher YI, Mishra S, Bhatia M, Alam D, Malini DS, et al. Effects of body mass index, tobacco smoking, alcohol drinking and solid fuel use on the risk of asthma: Individual Participant Data (IPD) meta-analysis of 175 000 individuals from 51 nationally representative surveys. *BMJ Open Respir Res*. 2016; 3(1):e000121.
45. Towfighi A, Zheng L, Ovbiagele B. Weight of the obesity epidemic: rising stroke rates among middle-aged women in the United States. *Stroke*. 2010; 41(7):1371–5. [PubMed: 20508193]
46. Swinburn B, Egger G. The runaway weight gain train: too many accelerators, not enough brakes. *Bmj*. 2004; 329(7468):736–9. [PubMed: 15388619]
47. Walthouwer MJ, Oenema A, Soetens K, Lechner L, de Vries H. Are clusters of dietary patterns and cluster membership stable over time? Results of a longitudinal cluster analysis study. *Appetite*. 2014; 82:154–9. [PubMed: 25064671]

Figure 1. Change in Schwartz's Bayesian Information Criterion (BIC) across a range of cluster solutions.

Table 1

Sample Characteristics of Input Variables of the Analytical Sample NHANES 2011-2012

Variables			All individuals (n=2,239,682)	Sample aged >=20 and BMI >=30 (n= 560,928)
Demographic	Age	Mean Age	47.5	48.3
	Sex	Male (%)	48.0	46.7
		Female (%)	52.0	53.3
	Race	Mexican American (%)	7.7	9.1
		Other Hispanic (%)	6.6	6.5
		Non-Hispanic White (%)	66.4	65.5
		Non-Hispanic Black (%)	11.5	14.6
		Non-Hispanic Asian (%)	5.2	1.3
		Other Race (%)	2.6	3.0
	Income	<\$20,000 (%)	17.7	18.9
		\$20,000 - \$45,000 (%)	26.2	29.9
		\$45,000 - \$75,000 (%)	20.3	20.5
		\$75,000+ (%)	35.7	30.7
Health	Respiratory	Asthma (%)	14.8	29.8
		Emphysema (%)	2.1	2.3
	Heart & Circulatory	Angina (%)	2.3	3.7
		Congestive Heart Failure (%)	2.8	4.2
		Coronary Heart Disease (%)	3.0	3.8
		Dyslipdemia (%)	35.7	41.4
		Heart Attack (%)	3.2	4.0
		Hypertension (%)	34.6	47.6
		Stroke (%)	2.9	3.2
	Mental Health	Mean Anxiety (days in month)	5.7	5.9
		Mean Depression Score	3.0	3.5
	Chronic Diseases	Arthritis (%)	23.1	29.8
		Cancer (%)	9.7	9.9
		Diabetes (%)	21.9	28.8
		Sleep Disorder (%)	9.1	14.6
		Mean Pain (days in month)	3.3	4.0
Behavioural	Physical Activity	Mean physical activity (MET)	3057.00	3181.84
	Smoking	Smoked more than 100 cigs (%)	44.0	44.5
	Alcohol	Non-drinkers (%)	14.70	14.33
		<1 drink per day (%)	29.14	28.83
		1-2 drinks per day (%)	15.73	20.56
		>2 drinks per day (%)	35.88	36.27
	Diet	Mean Healthy Eating Index	52.50	52.55

Variables			All individuals (n=2,239,682)	Sample aged ≥ 20 and BMI ≥ 30 (n= 560,928)
		Mean Body Mass Index	28.7	35.8

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Mean Characteristics of the Clusters in the Weighted Sample NHANES 2011-2012.

Variables	Cluster	Cluster					P-value
		Affluent men with sleep disorder	Elderly smokers with CVD	Older women poor mental health	Healthy White women	Healthy Non-White women	
Personal	Age	87424	91611	114228	72656	77905	117104
		15.6	16.3	20.4	13.0	13.9	20.9
		34.5	34.7	37.8	36.2	36.0	34.9
		48.1	58.2	58.6	43.8	38.9	34.5
		74.8	68.5	23.0	9.2	22.5	75.3
		25.2	31.5	77.0	90.8	77.5	24.7
	Sex	1.0	2.8	6.8	0.0	16.9	21.8
		1.3	3.3	6.3	0.0	19.9	12.9
		97.7	85.4	61.9	93.9	0.0	52.3
		0.1	7.7	20.6	0.0	57.6	8.7
		0.0	0.7	4.5	6.1	5.6	4.2
		0.0	8.7	22.1	4.9	27.5	27.3
	Income	0.0	7.1	36.7	20.2	18.0	34.8
		7.4	27.6	22.7	25.5	19.8	31.9
		92.6	56.5	18.4	49.4	34.7	6.0
		17.0	34.2	75.8	18.6	12.8	3.9
		0.0	4.4	2.5	3.3	0.0	0.3
		4.3	8.9	6.5	0.0	0.0	0.0
Health	Respiratory	1.2	6.0	8.4	0.0	0.0	0.0
		0.0	12.3	4.3	0.0	0.0	0.5
	Heart & Circulatory	2.0	36.6	27.1	1.2	5.6	3.7

Variables	Cluster					P-value
	Affluent men with sleep disorder	Elderly smokers with CVD	Older women poor mental health	Healthy White women	Healthy Non-White women	
	27.1	94.8	64.0	3.6	17.5	<0.001
Health	Dyslipidemia (%)	0.0	12.4	4.3	0.0	<0.001
	Heart Attack (%)	33.2	82.8	84.7	23.1	<0.001
	Hypertension (%)	0.0	0.4	8.5	0.3	<0.001
	Stroke (%)	1.2	2.0	11.4	0.9	<0.001
	Mean Pain (days in month)	4.9	3.2	9.6	4.5	<0.001
Wellbeing	Mean Anxiety (days in month)	2.5	1.8	6.2	3.3	<0.001
	Mean Depression Score	17.0	34.2	75.8	12.8	<0.001
	Arthritis (%)	11.1	28.6	11.8	2.1	<0.001
	Cancer (%)	28.5	8.5	27.8	3.9	<0.001
	Sleep Disorder (%)	2140.3	2481.4	1466.2	2426.0	<0.001
Behavioural	Mean physical activity (MET)	53.0	52.7	53.5	53.0	0.002
	Mean Healthy Eating Score	59.9	65.0	38.7	17.1	<0.001
	Smoked > than 100 cigs (%)	0.0	1.8	24.6	28.4	<0.001
	Non-drinkers (%)	24.0	32.6	48.2	25.9	<0.001
	<1 drink per day (%)	36.9	22.0	12.1	30.1	<0.001
	1-2 drinks per day (%)	39.1	43.5	15.1	15.6	<0.001
	>2 drinks per day (%)					

Red cells represent values higher than the population mean & green cells values lower than the average, within a single row. Color intensity indicates being further from the mean (up or down). Please note that MET score and the healthy eating index were reverse color-coded to match the interpretation of the rest of the characteristics.