# Neighborhood and Individual Determinants of Cardiometabolic Health: Obesity and Blood Pressure

By Marcia I. Pescador Jiménez

MSc Katholieke Universiteit Leuven Belgium, 2010

MA Northeastern University, 2013

A dissertation submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the Brown University Graduate School and Brown University School of Public Health, Department of Epidemiology

Providence, Rhode Island

May 2018

This dissertation by Marcia I. Pescador Jiménez is accepted in its present form

by the Department of Epidemiology as satisfying the

dissertation requirement for the degree of Doctor of Philosophy.

| | |
|---|---|
| _____ | _____ |
| Date | Eric B. Loucks, Advisor |

Recommended to the Graduate Council

| | |
|---|---|
| _____ | _____ |
| Date | S.V. Subramanian, Reader |
| _____ | _____ |
| Date | Gregory A. Wellenius, Advisor |

Approved by the Graduate Council

| | |
|---|---|
| _____ | _____ |
| Date | Andrew Campbell, Dean of the Graduate School |

Brown University School of Public Health
Box G-S121-3, Providence, RI 02912, USA
617-413-3690
Marcia_pescador_jimenez@brown.edu

# Marcia P. Jiménez

## Education

**PhD in Epidemiology** 2014 – Present
Brown University Providence, RI

- Research areas: Life-course impact of geographic contextual factors on cardiovascular disease; Chronic disease heterogeneity according to social, clinical and behavioral factors.
- Cardiovascular and Spatial Epidemiology
- Co-Advisors: Dr. Eric B. Loucks & Dr. S.V. Subramanian

**Masters in Sociology** 2011-2013
Northeastern University Boston, MA

- Research area: Social determinants of health disparities in immigrant populations
- Research Assistant in the Health Sciences Department
- Co-Advisors: Dr. Luis M. Falcon & Dr. Katherine L. Tucker

**Masters in Statistics** 2008-2010
Katholieke Universiteit Leuven Leuven, Belgium

- Research track with specialization in "Social, Behavioral and Educational Statistics"
- Thesis Dissertation: "The use of the gradient function to check goodness-of-fit in statistical models"
- Co-advisors: Dr. Geert Verbeke and Dr. Geert Molenberghs

**B.S. in Actuarial Science** 2002-2006
Universidad Nacional Autónoma de México (UNAM) Mexico City, Mexico

- Thesis Dissertation: "A longitudinal study of the interaction between education and the entrance of children and young people to the labor force"
- Advisor: Dr. Edith Pacheco

## Research Fellowships & Awards

**Current Grants**

1. **F31 - Ruth L. Kirschstein National Research Service Award**
   2016-2018
   Individual Pre-doctoral Fellowship from the National Institutes of Health.
   $ 43,576 USD

**Completed Grants**

1. **Demography Trainee Fellowship at the Population Studies and Training Center at Brown University (PSTC)**
   2015-2016
   2nd year Funding Fellowship for Doctoral Studies at Brown University.

2. **Brown's Initiative to Maximize Student Development (IMSD)**
   2014-2015
   1st year Funding for Doctoral Studies at Brown University.
3. **Graduate Research Assistantship Award**
   2011-2013
   Full Funding for Masters Studies at Northeastern University, Boston, USA.
4. **Flemish Government Fellowship**
   2008-2010
   One of 10 nationally selected fellows.
   Funding for Master Studies at the Katholieke Universiteit Leuven, Belgium
   Renewed for the second year of studies
5. **Cuerpo Academico Fellowship**
   2007
   Funding for bachelor thesis investigation at the Colegio de México, México City, México

**Awards**

1. **Epidemiology Department Abstract Competition – Brown University School of Public Health**
   2017
   Award to support students presenting their work and experience the learning that takes place at conferences. $1,000.00 USD
2. **Population Studies and Training Center Award – Abstract Submission to Population Association of America**
   2016
   Award to nurture collaborative research papers between PSTC faculty associates and graduate student trainees. $1,000.00 USD
3. **Centers for Disease Control and Prevention (CDC) Millennial Health Leaders' Summit**
   2016
   Brown School of Public Health Awarded Delegate
   Award to attend a two–day intensive training for graduate and medical students on vital health equity topics
4. **S4 Institute for the Spatial Structures in the Social Sciences** 2014-2015
   3rd place in Student Paper Competition. Brown University. $250.00 USD

**Research Experience**

**Principal Investigator - F31** June 2016-Present
Project: "Sensitive Periods: Prospective Associations of Childhood and Adulthood Neighborhoods with Life Course Obesity and Blood Pressure Trajectories"
Department of Epidemiology, Brown University School of Public Health

- Lead Role in data collection, data geocoding, spatial analysis and creation of a new analytic dataset including neighborhood level variables for a cohort study.
- Mentorship of 5 undergraduate and graduate students, including teaching spatial analytical techniques in ArcGIS, statistical analysis in SAS and discussion of theoretical frameworks.
- Produce regular reports to funders, as needed.
- Extensive collaborations with senior faculty across universities and disciplines.

**Research Assistant** June 2013-June 2014
Harvard Center for Population and Development Studies

Harvard University Cambridge, MA

- Statistical analysis using SAS, R and STATA
- Spatial analysis using R
- Longitudinal analysis of data from the Health and Retirement Study (HRS)
- Age-period-cohort analysis using data from the General Social Survey (GSS)

**Research Associate and Data Manager** Jan. 2011-May 2013
Center for Population Health and Health Disparities
Health Sciences Department, Northeastern University Boston, MA

- Statistical analysis including Propensity Score Analysis, Mixed Models, Factor Analysis and Generalized Linear Models, using SAS
- Development of SAS macros
- Spatial Analysis using ArcGIS
- Directed and performed data cleaning, detection of data inconsistencies and outliers that might affect statistical analysis
- Responsible for releasing data to researchers
- Organized a group journal club

**Research Assistant** Sept. 2007-May 2008
Center of Demographic, Urban and Environmental Studies
Colegio de México Mexico City, Mexico

- Analysis of Nation-wide survey databases using SPSS
- Research on labor force in rural and urban areas, and gender inequality
- Use of Kaplan Meier methodology and Cox Regression to study entrance to labor force and school drop-out rates in Mexico

**Teaching Experience**

Brown University School of Public Health Fall 2016
Department of Epidemiology

- Social Determinants of Health (PHP 1920)
  i. Involved in course development, including syllabus and course content
  ii. Lead lecturer for 15% of course "Area-Level Determinants of Health" and "Policies as Tools". Wrote and graded exams for these modules. Teaching evaluations are available.
- Undergraduate and Masters Thesis mentor on:
  i. "Neighborhood Socioeconomic Status and Major Depressive Disorder: The New England Family Study" by Tammy Jiang, Master's Thesis in Public Health.
  ii. "Association between Neighborhood Access to Green Space in Childhood and Cholesterol Levels in Adulthood" by Anthony Mei. Undergraduate Thesis. Work in progress
  iii. "Association between Neighborhood Access to Green Space in Childhood and Diabetes Risk in Adulthood" by Pamela Acero. Master's Thesis in Behavioral and Social Sciences. Work in progress
- Direct Independent Study on "Statistical Methods" - Undergraduate and Graduate Level

Health Sciences Department Jan. 2011-July 2013
Northeastern University Boston, MA

- SAS workshop for undergraduate and graduate students
- Data management training for graduate students using SAS

**Publications**

1. **Jimenez, M.P.,** Green, M., Subramanian, S.V., and Razak, F. "A clinical, demographic and socioeconomic typology of obesity in the United States: an analysis of NHANES 2011-2012". *Annals of Epidemiology.*
2. Verma, A.A., **Jimenez, M.P**., Tangermann, R., Subramanian, S.V., Razak, F. "Insecurity, Polio Vaccination Rates, and Polio Incidence in Northwest Pakistan, An Epidemiological Analysis". *Proceedings of the National Academy of Sciences.*
3. Verma, A.A., **Jimenez, M.P.**, Subramanian, S.V., Sniderman, A.D., Razak, F. (2017) "Race and Socioeconomic Differences Associated with Changes in Statin Eligibility under the 2013 ACC/AHA Cholesterol Guidelines" *Circulation: Cardiovascular Quality and Outcomes.*
4. Noelke, C., McGovern, M., Corsi, D., **Jimenez, M.P**., Stern, A., Wing, I.S., Berkman, L. (2016) "Increasing Ambient Temperature Reduces Emotional Well-Being" *Environmental Research.*
5. Beltran-Sanchez, H., **Jimenez, M.P.**, and Subramanian S.V. (2016) "Assessing Morbidity Compression in two cohorts from the Health and Retirement Study" *Journal of Epidemiology and Community Health.*
6. Vable, A.M., Canning, D., Glymour, M.M., Kawachi, I., **Jimenez, M.P.** & Subramanian, S.V. (2016). "Are there spillover effects from the GI Bill? The mental health of wives of Korean War veterans" *PLOS ONE.*
7. Vable, A.M., Canning, D., Glymour, M.M., Kawachi, I., **Jimenez, M.P.** & Subramanian, S.V. (2015). "Can Social Policy Influence Socio-Economic Disparities? Korean War GI Bill Eligibility and Markers of Depression" *Annals of Epidemiology.*
8. Shaoyong, S., **Jimenez, M.P.**, Roberts, C. & Loucks, E.B. (2015). "The Role of Adverse Childhood Experiences in Cardiovascular Disease Risk: A Review with Emphasis on Plausible Mechanisms" *Current Cardiology Reports.*
9. **Jimenez, M.P.**, Osypuk, T., Arevalo, S., Tucker, K.L. & Falcon Luis. M. (2015). "Neighborhood socioeconomic context and change in allostatic load among older Puerto Ricans: The Boston Puerto Rican health study" *Health and Place.*
10. Todorova, I., Tucker K.L., **Jimenez, M.P.**, Lincoln, A., Arevalo, S. & Falcon, L.M. (2013). "Determinants of Self-rated Health and the role of acculturation: implications for health inequalities." *Ethnicity & health*.

**Manuscripts under review**

1. **Jimenez, M.P.**, Tucker, K.L. and Falcon Luis. M. "Proximity to green areas and change in Self-rated Health for older Puerto Ricans in the Boston Area."

**Manuscripts in progress**

1. **Jimenez, M.P.**, Lopez, L., Ramahlo, J., Galloway, Y., Rexrode, K.M., Jiménez, M.C. "Validation of the Framingham Stroke Risk Score among US adults by self-reported race/ethnicity ".
2. **Jimenez, M.P.,** Green, M., Verma, A.A., Loucks, E.B., Subramanian, S.V., and Razak, F. "Hypertension Clusters Across Time and Differential Risk of Mortality: National Health And Nutrition Examination Survey 1999 - 2012".
3. Kim, R., **Jimenez, M.P.,** Subramanian, S.V. "Body Mass index Variance Explained by Socio-demographic factors: National Health And Nutrition Examination Survey 1999 - 2012".
4. **Jimenez, M.P.,** Subramanian, S.V., Wellenius, G., Loucks, E.B., "Longitudinal association of life-course neighborhood socio-economic deprivation with blood pressure and obesity"

5. **Jimenez, M.P.,** Subramanian, S.V., Wellenius, G., Loucks, E.B., "Longitudinal association of life-course neighborhood accessibility to green space with blood pressure and obesity"
6. Jiang, T., **Jimenez, M.P.,** Loucks, E.B., Buka, S. "Neighborhood Socioeconomic Status and Major Depressive Disorder: The New England Family Study"
7. Mei, A., Acero, P., Loucks, E.B., Wellenius, G., **Jimenez, M.P.** "Association between access to green space in childhood and cholesterol level in adulthood"
8. Acero, P., Mei, A., Loucks, E.B., Akila, K., **Jimenez, M.P.** "Association between access to green space in childhood and risk of diabetes"

**Presentations**

1. June 2018. Society for Epidemiologic Research Annual Meeting. Oral Presentation. "Neighborhood green space access at birth, childhood and adulthood associated with blood pressure trajectories across the life course", **Jimenez, M.P.**, Subramanian, S.V., Wellenius, G., and Loucks E.B.
2. April 2018. Annual Meeting of the Population Association of America. Oral Presentation. "Longitudinal Association of Life-Course Neighborhood Deprivation with Blood Pressure", **Jimenez, M.P.**, Subramanian, S.V., Wellenius, G., and Loucks E.B.
3. March 2018. Annual Meeting of the American Heart Association EPI/Lifestyle. Poster Presentation. "Hypertension Clusters Across Time and Differential Risk of Mortality: National Health and Nutrition Examination Survey 1999 – 2012", **Jimenez, M.P.**, Green, M., Verma, A.A., Subramanian, S.V., and Razak, F.
4. October 2017. Second Annual Population Health Science Research Workshop at Boston University. Poster Presentation. "An Evaluation of the Predicted Risk of Death Due to Cerebrovascular Disease Among US Adults by Race/Ethnicity and Sex", **Jimenez, M.P.**, Loucks, E.B., and Jimenez, M.C.
5. July 2017. Brown University Summer Research Symposium. Poster Presentation. "The Association Between Early Childhood Access to Green Space & Adult Blood Pressure", Mei, A., Loucks, E.B., Acero, P., and **Jimenez, M.P**.
6. April 2017. Brown University Public Health Research Day. Poster Presentation. "Childhood Neighborhood of Residence and Risk of Depression", Jiang, T., **Jimenez, M.P.**, Loucks, E.B., and Buka, S.
7. April 2015. Population Association of America 2015 Annual Meeting (PAA). Presentation. "The Causal Effect of Age on Subjective Well-Being", Noelke, C., **Jimenez, M.P.**, and Corsi, D.
8. April 2015. Population Association of America 2015 Annual Meeting (PAA). Poster exposition. "Temperature and Well-Being in the U.S.: The Sub-Clinical Implications of Global Warming", Noelke, C., Corsi D., **Jimenez, M.P.** and Wing, I.S.
9. Nov. 2013. Aging and Public Health Program of the 141st Annual meeting of the American Public Health Association (APHA), Boston, MA, USA. Poster Exposition of work in progress: "Longitudinal analysis of the impact of neighborhood context on Allostatic Load in older Puerto Ricans"
10. Nov. 2013. Aging and Public Health Program of the 141st Annual meeting of the American Public Health Association (APHA), Boston, MA, USA. Poster Exposition of work in progress: "Change in social network based functional support and symptoms of depression"
11. May 2013. Center for Population Health and Health Disparities Annual Meeting, Boston, MA, USA. Poster Exposition on "Association between

proximity to green areas and Self-rated Health for older Puerto Ricans, a longitudinal perspective"

12. Mar. 2013. Eastern Sociological Society, Boston, MA, USA. Talk on "Association between proximity to green areas and Self-rated Health for older Puerto Ricans, a longitudinal perspective"
13. Mar. 2013. International Labor Process Conference, Rutgers University, New Jersey, USA. Talk on "The impact of labor market experiences on Allostatic Load for older Puerto Ricans"
14. Mar. 2012. Research, Innovation and Scholarship Expo, Northeastern University, Boston MA, USA. Poster exposition of work in progress: "The effects of labor market experiences on health outcomes for older Puerto Rican adults"
15. June 2011. Center for Population Health and Health Disparities Annual Meeting, Chicago Illinois, USA. Poster exposition of work in progress with Irina Todorova: "What do people integrate into their assessment of overall health? Determinants of Self-rated Health and implications for health inequalities of Puerto Rican adults"
16. Oct. 2008. Mexican Demography Society, IX National Reunion of Demographic Research in Mexico, Mérida Yucatán, México. Poster exposition of my bachelor thesis: "Longitudinal analysis of the entrance of children and youth into the labor force"

**Manuscript reviewer for**

- Advances in Life Course Research
- Journal of Epidemiology and Community Health
- PLOS Medicine
- Health and Place
- Population and Environment
- Journal of Immigrant and Minority Health

**Technical Skills**

- **Statistical Techniques:** Mixed Models, Generalized Linear Models, Factor Analysis, Longitudinal Analysis, Propensity Score Analysis, Age-Period Cohort Models
- **Spatial Analysis Techniques:** Geo-Coding, Network Analysis, Service Area Analysis, Spatial Correlation, Hot Spot Analysis
- **Programming Languages:** SAS, R, STATA, SPSS.
- **Spatial Analysis Software:** R, ArcGIS.
- **Languages:**
  Spanish : First language
  English : Full fluency
  French : Proficient
  German : Basic

# Acknowledgements

It has been a pleasure to work with my advisor Dr. Eric B. Loucks. I am truly grateful for the opportunity to have collaborated with him throughout the course of my doctorate. For his time, feedback, great support, both professional and personal, I extend my deepest appreciation. Additionally, I am grateful for the guidance and input from my Dissertation Advisory Committee members, Drs. Gregory A. Wellenius and S.V. Subramanian. Without their wisdom and thoughtful feedback, this dissertation would not be possible. I would also like to take this opportunity to thank Drs. Stephen Buka and John Logan who gave me access to the data necessary for this work.

I am forever grateful for my fellow cohort members Samantha Kingsley, Stephanie Shih, Kimberly Glazer, Brianna Osetinsky, Sarah Gordon and Sylvia Shangani, incredibly smart women who I am lucky to call my friends.

I am thankful for Brown's Initiative to Maximize Student Development, the Population Studies and Training Center, and my Ruth L. Kirschstein National Research Service Award from the NIH for their funding support.

For my family, I have the greatest thanks and love. Despite the distance, I know they are with me every day. Words cannot express how thankful I am to have my husband Jonathan and our daughter Emilia. Everything I do is with their love.

# Table of Contents

# List of Tables and Figures

# Introduction

Despite significant declines in mortality in the U.S., cardiovascular disease (CVD) remains a considerable burden, both in terms of health and costs. In 2011, CVD accounted for the underlying cause of death of 31.3% of all deaths in the U.S. [1]. According to the American Heart Association, the annual direct and indirect cost of CVD in the United States is an estimated $320.1 billion [1]. In particular, hypertension and obesity are 2 major health issues in the U.S., and risk factors of CVD. More than one-third (34.9% or 78.6 million) of U.S. adults are obese [2] and about a third (29% or 70 million) have hypertension [3, 4]. A growing social epidemiology literature is documenting that neighborhood and geographic context is associated with a range of adverse consequences on CVD [5-8]. A life course epidemiology literature has concurrently evolved by documenting the effects of early life exposures on cardiovascular health [7, 9-11]. However these areas of study have evolved separately. Incorporating both place and time for understanding causes of CVD is one of the key challenges for epidemiologic research.

A neighborhood's features can severely limit the choices and resources available to individuals. Recent research has shown how neighborhood factors (such as built environment or low NSES) are associated with the incidence and prevalence of CVD risk factors [7, 12-19]. The impact of area-level determinants of health has been considerably noticed, and policies have been proposed to make neighborhoods healthier places to live for everyone [20]. In fact, research has suggested that neighborhood interventions to increase access to recreation facilities would be cost-effective to improve health behavior [21]. Moreover, given the emerging recognition that CVD has childhood origins [10, 11, 22], an important question is whether neighborhood context

earlier in life exerts greater influence on CVD development. During the past two decades pathological data have shown that atherosclerosis, the leading cause of CVD, begins in childhood and that the extent of atherosclerotic change in children and young adults is associated with the presence of atherosclerotic risk in adults. It thus seems reasonable to consider whether initiating healthful lifestyle interventions in childhood could promote improved cardiovascular health in adult life. A life-course approach to chronic disease epidemiology recognizes the importance of time in understanding causal links between exposures and outcomes within an individual life-course. [10]

However, most of the research in neighborhood effects on health has been cross-sectional and thus susceptible to reverse causation. The development of strategies that allow linkage of cohort data to meaningful historical neighborhood data is crucial for advancing this area of research [6]. Improving causal inference in studies of neighborhood health effects requires not only greater methodological sophistication but also much more attention to the theoretical models underlying the research questions [6]. Another important obstacle for causal inference in neighborhood research is selection bias. Selection bias can be due to the fact that individuals may select their place of residence based on their health or based on their predisposition to certain behaviors (e.g. people who are more inclined to be physically active may choose to live in areas with better physical activity resources) [6, 23].

In the first two chapters of this project, we evaluate two theoretical models: the critical period model which states that exposures at a specific time have long-lasting effects on the function of the body (Figure 1) [24]; and the accumulation of risk model, which suggests that effects accumulate over the life course (Figure 2) [9, 24]. This is particularly important for diseases

that develop over a long period of time, as is the case for CVD. We evaluate neighborhood's features in childhood, when participants were not likely to select their place of residence. To our knowledge, this study's longitudinal design is the first to assess the association between objectively measured neighborhood characteristics from birth through adulthood, and CVD risk.

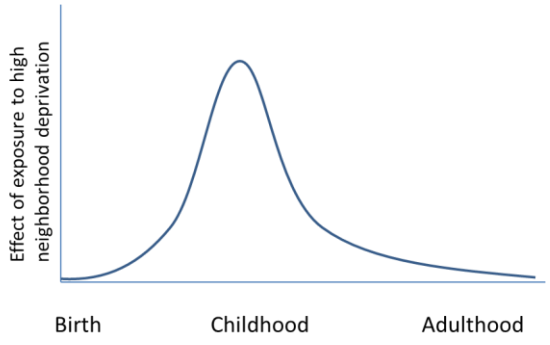



*Figure 1. Critical Period Model*

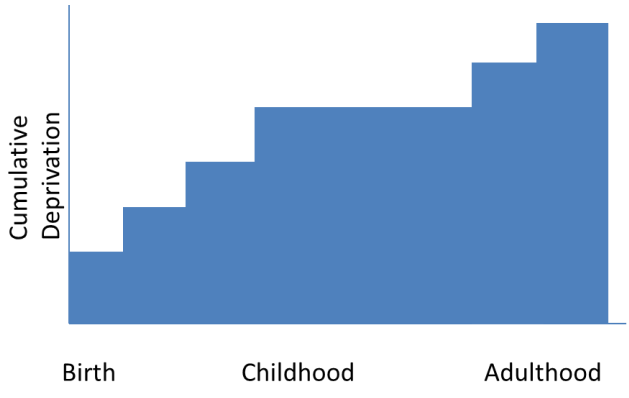



*Figure 2. Cumulative Risk Model*

On the third chapter, we focus on obesity, a serious public health challenge and an important risk factor for multiple chronic diseases besides CVD, such as arthritis, diabetes, hypertension (HTN), and dyslipidemia. The prevalence and severity of obesity have been dramatically increasing throughout the United States during recent decades [25, 26]. Currently, guidelines, randomized trials and epidemiologic studies of obesity tend to consider obesity as a homogeneous entity. However, obesity may represent a heterogeneous condition with respect to demographic, behavioral and clinical factors. Previous studies suggest that the association between neighborhood deprivation and obesity is not direct [27] and that researchers must look into other individual determinants of obesity, like race, physical activity, and clinical factors that

may deviate among individuals. 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 [28, 29]. In addition, studies have revealed that obesity varies significantly according to race/ethnicity, independent of sociodemographic characteristics and neighborhood context [20]. Another study found that people that drink higher amounts of alcohol and live in deprived communities were significantly more likely to be overweight [30]. In addition, whilst studies generally agree that neighborhood walkability influences levels of physical activity and obesity [31, 32], there is evidence that walkability may differentially impact residents depending on their age and sex [33]. Consequently, it will become critical to clinicians, public health practitioners, and policymakers to distinguish obese individuals at high risk for obesity-related chronic diseases, from those at lower risk. Using comprehensive and detailed population level data in the United States, we seek to examine complex interaction of biological, behavioral, and social factors that intersect with obesity. To our knowledge, this is the first analysis of obesity patterns in adults in the U.S. that identifies subgroups among individuals categorized as obese. Results from this analysis will be key to best tailor effective future interventions and clinical trials in response to the high levels of obesity in the U.S.

Overall, the goals of this project are to assess associations of neighborhood characteristics at birth, childhood and adulthood with blood pressure and obesity across the life course, and to evaluate groups of individuals classified as obese, by demographic factors, behavioral characteristics and clinical conditions. Individuals within different neighborhoods may be exposed to a variety of choices and resources available that subjects them to diverse risks of CVD; analogously, individuals classified as obese may have different demographic, clinical and

behavioral characteristics that may predispose them to distinctive risk of obesity. The contribution of the proposed research is expected to shed light on the interplay between (1) contextual neighborhood attributes across the lifespan, (2) clusters of individual characteristics, and cardio-metabolic disease development respectively. Results from the first two chapters of this project offer the potential to inform health professionals about modifiable features of neighborhoods at critical periods of susceptibility in life to target timely interventions of two major CVD risk factors in the United States: hypertension and obesity. These analyses may provide new evidence that greater access to neighborhood green space and higher NSES during specific periods in the life course may slow the development of CVD in middle-aged adults. Results from the third chapter of this project provides a basis for future studies to evaluate whether different strategies and goals may be needed in the clinical management of obesity. Finally, the addition of geographic data to the New England Family Study as a result of the first two chapters opened new opportunities for future research to understand the etiology of several diseases using spatial methods that is not available in most longitudinal observational studies.

# Chapter 1: Longitudinal associations of life-course neighborhood deprivation with blood pressure and obesity

## Abstract

Neighborhood is increasingly investigated as a determinant of cardiovascular disease (CVD). However, longitudinal studies on neighborhood are rare, hampering the ability to address questions on causality and critical stages in the life-course. We aim to investigate if higher neighborhood socioeconomic status (NSES) during early childhood, as a susceptible period, may slow the development of CVD in adults.

Using longitudinal data from the New England Family Study with a 48-year follow-up, we geocoded participants' home addresses throughout their life-course and used multilevel models to evaluate how NSES is associated with systolic blood pressure (SBP), diastolic blood pressure (DBP) and body mass index (BMI). Results suggested that one standard deviation higher of NSES at birth was associated with a 2.7 mmHg lower SBP in adulthood (95% CI: -5.1, -0.3), and that one standard deviation higher of NSES at adulthood was associated with 2.2 mmHg lower SBP (95%CI: -4.2, -0.3), 1.4 mmHg lower DBP (95%CI: -2.8, -0.1), and 1.1 lower BMI (95%CI: -2.2, -0.1), which remain statistically significant after controlling for confounders.

Results from this unique longitudinal study may provide insights into critical time periods during which individual and community actions can effectively improve public health and reduce neighborhood health disparities.

## Introduction

Despite significant declines in mortality in the U.S., cardiovascular disease (CVD) remains the leading cause of death with considerable burden on both health and costs. Hypertension and obesity, two major risk factors for adult CVD, are highly prevalent in the U.S. Specifically, an estimated 70 million (29%) of U.S. adults have hypertension [3, 4] and more than a third are classified as obese [2].

Multiple studies have found a gradient in the incidence, morbidity and mortality of CVD across the spectrum of socioeconomic status [34]. However, the additional contribution of the neighborhood of residence to the risk of CVD, independent of individual socioeconomic indicators is increasingly investigated. For example, comparisons of obesity prevalence using geographic criteria indicate that low income areas are associated with higher rates of obesity [35] and higher incidence of coronary heart disease [13]. Neighborhood may play a critical role in the development of CVD risk factors and produce long-term effects through a range of mechanisms, both physical and social. Neighborhood socioeconomic status (NSES) may affect health through systematic blockage of opportunities, such as few employment opportunities, and poor schools [36]. Furthermore, deprived neighborhoods are likely to experience financial disinvestment and low collective efficacy, leading to deteriorating buildings and dirty streets [37]. Research suggests that these conditions may contribute to socioeconomic inequalities in CVD risk [38], heart failure risk [39] and improvements in prevention and treatment of CVD [40].

However, most prior studies have only investigated cross-sectional associations between neighborhood deprivation and CVD risk factors, which are susceptible to reverse causation. As

the field moves to more detailed examination of life-course periods, researchers need to adopt methods that allow appropriate identification, including selection of covariates based on the hypothesized causal model, and specifying models so that the effects of each model might shed light for potential interventions targeted at life-course periods where they might be more effective [41]. The development of strategies that allow linkage of cohort data to meaningful historical neighborhood data is crucial for advancing this area of research [6]. In the present study, we compared the association between NSES at 3 different time points (birth, childhood and adulthood) and CVD risk factors, using multilevel models to appropriately account for within and between neighborhood variability. We hypothesize that participants living in economically deprived neighborhoods during early childhood are a high-risk subset of the population that should be prioritized.

## Data and Methods

### *Data*

We used data from the New England Family Study (NEFS), a longitudinal investigation with up to 48 years of follow-up of the offspring of pregnant women enrolled between 1959 and 1966 in the Collaborative Perinatal Project (CPP). The NEFS was initiated to locate and interview samples of the adult offspring at the Providence, Rhode Island, and Boston, Massachusetts sites (N=17,921 live births) [42]. Analyses for this project were achieved through the merging of data from 2 of the sub-samples that comprise NEFS: Longitudinal Effect on Aging Perinatal Project (LEAP) and Pathways Linking Education and Health in Middle Adulthood Project (Edhealth). Collectively, this yielded an analytical dataset of 931 participants. Participants were followed

since birth until adulthood in the metropolitan areas of Providence, RI and Boston, MA, two of the most densely inhabited areas in the Northeastern United States. Residential addresses for the participants were collected at 12 different occasions from birth (around 1960s) through childhood (around 1970s) and then again at adulthood (around 2000s); and were subsequently geocoded using Brown University Geocoding Service (96% of the total number of addresses) [43], and World Geocoding Services (3% of the total number of addresses, which did not report a street number) [44]. All spatial analysis was done using ArcGIS, Version 10.4.1 (Redlands, CA). The analytical sample included all NEFS subjects who resided in MA or RI at adulthood and who had complete data on systolic blood pressure (SBP), diastolic blood pressure (DBP) or body mass index (BMI) at adulthood respectively. Thus, our analyses are based on an analytic sample of 871 adults who lived in 255 census tracts at birth, 233 at childhood and 522 at adulthood (please refer to Figure 1 in the Supplementary Material).

*Measures*

The primary exposure was NSES measured at the census-tract (CT) level throughout the life-course. Spatial data on NSES were drawn from the National Historical Geographic Information System [45], ICPSR, and the Bureau of the Census, the final variables were created by the staff of the American Communities Project at Brown University [46]. Defining neighborhoods based on CT is commonplace among studies in the US [47]. Because of changing CT borders over time, we homogenized CT boundaries from 1960 and 1970. We assigned census data to subjects by waves according to their CT of residence and the date of their study examination, selecting the Census closest to the examination date.

NSES at each time point was measured as a standardized (mean=0, standard deviation=1) score of: percent with less than a high school education, percent of unemployment, and median household income [7, 48]. Household income from 1960s and 1970s were adjusted for inflation rates to 2000 [49]. Lower scores indicate lower NSES. NSES was considered as a time-varying covariate. We also dichotomized the NSES score below and above the mean (zero), where 0 indicates low NSES and 1 indicates high NSES. Finally, we constructed an indicator of NSES mobility as follows: (1) participants who remained categorized as low NSES throughout the life-course, (2) participants who moved from low to high NSES, (3) participants who moved from high to low NSES and (4) participants who remained in high NSES throughout the life-course.

Individual-level systolic blood pressure (SBP), diastolic blood pressure (DBP) and body mass index (BMI) were the outcome variables. SBP and DBP were clinically assessed, as the average of the mean of the second and third seated readings. Height and weight were measured by trained research technicians from participants wearing light clothing without shoes using a calibrated stadiometer and weighing scale. BMI was calculated as weight (kg) divided by height (m) squared. Individual-level covariates included the time-varying variable age, and time-constant variables individual socioeconomic status (SES) assessed as adult attained education (less than high-school, high-school, more than high-school), parental SES assessed as mother's education (less than high-school, high-school, more than high-school), and father's education (less than high-school, high-school, more than high-school), and demographic characteristics gender (male, female), and race (White, African-American, other). All time-constant variables were assessed at adulthood, except for parental SES which was assessed at birth.

*Statistical Analysis*

We fit two sets of models in the current analysis. The first set of models use a three-level multilevel model (MLM). As described in detail elsewhere [7–10], MLM assumes that observations are hierarchically nested, such that members of the lower level (i.e., level one) are nested in one and only one entity at the higher levels (i.e., levels two and three). We use three-level models to account for individuals (level 1) nested within families (level 2) within neighborhoods (level 3) at different time points and assess between and within variances at each level. The outcome for the first set of models is SBP, DBP and BMI at adulthood. These models are stratified by time, thus we have 3 models (one for birth, one for childhood and one for adulthood) to compare effect size across models and evaluate the critical period. All 3 models control for the following confounders: age, gender, race, individual SES, and parental SES. In addition, the childhood model was further adjusted for birth NSES, and the adulthood model was further adjusted for birth NSES and childhood NSES (for more detail please see the Directed Acyclic Graph in Supplementary Figure 3).

The second set of models evaluates the accumulation of risk model, where the main exposure is NSES mobility throughout the life-course. The outcome for the second set of models is time-varying SBP and BMI throughout the life-course (DBP was only measured at adulthood and thus was not included as outcome in these models). Since participants moved frequently, there was no clear nested hierarchy. Consequently, we used Cross-Classified Multilevel Models (CCMM) using Markov Chain Monte Carlo (MCMC) re-estimation to incorporate non-hierarchical nesting structures and account for these repeated measures with changing neighborhood membership over time [50]. MCMC is a simulation-based method estimating the parameters by re-sampling the data to produce more accurate estimates of the unknown parameters.

Observation occasion served as level 1, and individual and neighborhood were separate cross-classified levels for level 2 (Supplementary Figure 2) [50]. All models control for the following confounders: age, gender, race, individual SES, and parental SES.

Finally, since the data are unbalanced (individuals vary in their number of measurements by design or attrition), the first set of models was restricted to individuals who had complete data on all variables at all time-points to be able to compare effects across models (N=459). For the second set of models, we assume that the missing data are missing at random. Thus, within the likelihood framework of MLM, complete-case bias is eliminated by incorporating all available information [51]. Furthermore, we also conducted multiple imputation procedures with chained equations (MICE) as a sensitivity analysis (Supplementary Table 1). Statistical significance was assessed at 0.05 level. Data management was performed in SAS version 9.4 and statistical analyses were performed using R version 3.4.1 and MLWin 2.34.

## Results

The analytical sample included 871 individuals, of which 58% were female. The mean age was 44.5 years. Among included participants, 74.5% were White, 18.9% were African-American and 6.2% were categorized as other. About 15% of the sample had less than high-school education, and paternal education level was less than high-school for more than half of the sample (Table 1). The average values for the outcomes were: SBP 117.5 mmHg, DBP 76.3 mmHg and BMI 30 $kg/m^2$ (Table 1). Descriptive statistics of the NSES variables indicate that household income increased across time, while percentage of unemployment remained somewhat stable and percentage with less than high-school education decreased (Table 2). Indicators of neighborhood

socio-economic mobility showed that 25% of the participants lived in neighborhoods with low NSES at all time-points, while almost 30% lived in neighborhoods with high NSES throughout the life-course (Table 3).

Results for the first set of models for SBP indicated that one standard deviation higher of NSES at birth was significantly associated with a 2.7 mmHg lower SBP in adulthood (95%CI: -5.1, -0.3), after adjusting for confounders (Table 4). Higher NSES at adulthood was also significantly associated with 2.2 mmHg lower SBP in adulthood (95%CI: -4.2, -0.3), after adjustment for confounders. The intra-class correlation coefficient for SBP was 12.6% at birth, 9.7% at childhood and 4.7% at adulthood (Figure 1). Results for DBP indicated that one standard deviation higher of NSES at birth was associated with -1.59 mmHg of DBP (95%CI: -3.2, 0.04), however this association did not approach statistical significance. On the contrary, higher NSES at adulthood was significantly associated with 1.4 mmHg lower DBP (95%CI: -2.8, -0.1), after controlling for confounders (Table 4). The intra-class correlation coefficient for DBP was 8.9% at birth, 0.46% at childhood and 1.1% at adulthood (Figure 1). Results for BMI indicated that NSES at birth and childhood were not statistically associated with BMI at adulthood after controlling for confounders. However, one standard deviation higher of NSES at adulthood was significantly associated with 1.1 $kg/m^2$ lower BMI (95%CI: -2.2, -0.1) at adulthood after controlling for confounders (Table 4). The intra-class correlation coefficient for BMI was 7.9% at birth, 3.6% at childhood and 10.3% at adulthood (Figure 1).

In the CCMM analysis for SBP, results showed that after controlling for confounders, living in a neighborhood with low NSES across the life-course, compared to living in a neighborhood with high NSES across the life-course, was associated with a 1.6 mmHg increase of SBP (95%CI: -

0.3, 3.5), approaching significance. Results for BMI showed a consistent pattern suggesting that having stayed in low NSES or having experienced instability of NSES (fluctuated from low to high NSES or vice versa) was associated with a 0.5 $kg/m^2$ increase in BMI (95%CI: -0.04, 1.14), compared to having stayed in high NSES across the life-course, after adjusting for confounders. These effects were not statistically significant (Table 5).

## Discussion

We used data from a longitudinal cohort of 871 participants in Boston MA and Providence RI to investigate the association between NSES across the life-course and CVD risk factors. This paper adds to prior literature by directly testing NSES measures across 3 life-course stages and using multilevel models to control for the dependency of observations within a neighborhood. To our knowledge, this is the first study with objective measurements of neighborhood characteristics with a 48-year follow-up from birth through adulthood with clinical individual data necessary for this type of work. We found evidence that neighborhood socioeconomic disadvantage throughout the life-course is, in addition to and regardless of individual socioeconomic characteristics, related to increased risk of CVD. Our results provide critical evidence to suggest that neighborhood socioeconomic deprivation starting in early childhood may have severe life-long consequences on SBP and DBP, while neighborhood socioeconomic deprivation during adulthood may have a higher impact on BMI. In addition, our results suggest that cumulative exposure to low NSES, compared to cumulative exposure to high NSES was associated with a higher trend of SBP. Neighborhood at birth accounted for a larger proportion of variation in SBP and DBP than neighborhood at childhood or adulthood; while neighborhood at adulthood

accounted for a larger proportion of variation for BMI. Results from this study have important implications for preventive strategies that aim to reduce the burden of hypertension and obesity.

Prior evidence about the life-course associations between neighborhood and CVD risk factors is limited as most studies have reported cross-sectional results. The first set of models of the present study, tested the critical period model. Our results suggest that each of the 3 life-course measures of NSES contributed to an increased risk of hypertension or obesity later in life, although not all associations were statistically significant after accounting for confounders. We initially hypothesized childhood as a critical period and surprisingly found that critical periods may differ based on the outcome of interest. This observation is in accord with evidence from a study looking at major CHD risk factors present during childhood and adolescence, which concluded that the relative importance, of early and later-life exposures will differ by health outcome [10]. For instance, we found evidence that NSES at birth had a bigger effect on SBP and DBP, compared to NSES at childhood or adulthood; while NSES at adulthood had a bigger effect on BMI compared to NSES at birth and childhood. A reason for this finding might be that the participant's addresses at birth were collected since the third trimester of pregnancy through the first year of age; thus, the effect of NSES at birth on blood pressure might be picking up pre-natal and early growth exposure to adverse geographic factors, which have been found to increase the risk of hypertension [52]. Observations in animals show that the environment during development permanently changes the body's structure and function as well as its responses to environmental influences encountered in later life [53]. However, our study adds new evidence that exposure to low NSES at or around birth may be more detrimental than exposure to low NSES at other time point in life, for blood pressure. On the other hand, adulthood as a sensitive

period may be a result of a series of interactions occurring at different stages of development, such as exposure to neighborhood deprivation and biological changes over the life-course. Studies looking at child growth and socio-economic circumstances, suggest that the pathogenesis of CVD may be a product of branching paths of development that are triggered by the environment [52, 54]. Our results show a negative association of NSES and BMI at each time point, which may suggest that the pathway to obesity may originate in early life but that adverse exposure to NSES at different stages of life add to each other having a culmination point effect at adulthood. The latter might be explained by longer-term consequences of the complex accumulation and interaction, across generations, of early and later-life exposures [10].

The second set of models tested the accumulation of risk model, which is particularly important for diseases that develop over a long period of time, as is the case of CVD. Application of a CCMM enabled us simultaneously examine the fixed and random effects corresponding to time and neighborhood settings. Simultaneous examination of time and neighborhoods, is important because both settings can influence health behaviors through multiple pathways, including policies, normative behaviors, and access to resources [47]. Our results suggested that cumulative exposure to low NSES was associated with higher SBP but evidence for the association of NSES with BMI trend was less evident, suggesting that not only cumulative exposure to low NSES, but also the instability of NSES may impact BMI. For example, we found evidence that participants that moved from high NSEs to low NSES had a 0.54 kg/m$^2$ increase of BMI (95%CI: -0.04, 1,1). However, we also found evidence that suggested that upward NSES mobility did not offset the effects of low NSES earlier in life. Participants who moved from low NSES to high NSES also had a 0.49 kg/m$^2$ increase of BMI (95%CI: -0.07, 1.07). Neighborhood disadvantage has been

found to be associated with co-occurrence of poor health behaviors over and above individual's own socioeconomic status [48]. Thus, participants living in low NSES in early childhood may have been exposed to structural shortcomings (like access to healthy food) which may increase the likelihood of co-occurrence of poor health behaviors since early life and continue through adulthood even after moving to neighborhoods with higher NSES. The observed neighborhood effects, and the various possible reasons for them, suggest there are many issues policy makers could take into account when aiming at reducing neighborhood health disparities [48].

This study's longitudinal design relating NSES at birth, childhood and adulthood to CVD risk factors is among the few observational studies with detailed neighborhood- and individual-level data necessary for this work. Our analysis approach allows us to evaluate if causal processes operate at more than one level and describe the variability and heterogeneity in the population, rather than average values. Results suggested that a larger proportion of variation of blood pressure was accounted by neighborhoods at birth which may indicate that interventions at birth at the neighborhood level may be more effective than at other time point. However, improving causal inference in studies of neighborhood health effects requires not only greater methodological sophistication but a deeper examination of the theoretical models underlying the research question [6]. In this project, we evaluated 2 theoretical models: the critical period model which states that exposures at a specific time has long-lasting effects on the function of the body; and the accumulation of risk model, which suggests that effects accumulate over the life course. Finally, another important obstacle for causal inference in neighborhood research is selection bias due to the fact that individuals may select their place of residence based on their predisposition to certain health behaviors (e.g. people who are more inclined to be physically

active may choose to live in areas with better physical activity resources) [6, 23]. In this analysis, we evaluate neighborhood's features at birth and childhood, when participants were not able to select their place of residence.

Finally, there are several limitations to our research. First, we have missing data due to loss to follow-up that may result in lack of exchangeability between sampled and not sampled. However, MLM use maximum likelihood to estimate the parameters allowing for unbalanced data and thus, in the ignorable situation, this eliminate complete-case bias by incorporating all available information [51]. Moreover, we used MICE to account for the missing data as a sensitivity analysis and the results were similar. Future research should examine the model specifications more closely. Second, the results from any multilevel analyses are inherently sensitive to omitted units/scales. For instance, our definition of neighborhoods as CT may not be accurate and thus bias our results. Still, a strength of using CTs is that they can be appended to any database with addresses, and provide data for determining contextual as well as compositional neighborhood effects on health. Further studies should look at other definitions of neighborhood. Third, we might be missing to include relevant variables into our measurement of NSES, and thus the results can be biased. We looked into adding other socioeconomic variables, however unfortunately not all variables were measured consistently throughout time since 1960 and thus we were only able to include the 3 variables in the NSES score. Fourth, the analysis includes only participants who remained in MA or RI which limits generalizability of the study and may potentially induce selection bias. However, this limitation in generalizability also strengthens the plausibility of the association's being causal by removing confounding by design. Because all of the subjects were living in MA or RI, neighborhoods were more similar than those

of diffuse areas, and neighborhoods primarily differed with regard to the exposure rather than other, unobserved factors. Further analysis is needed to establish whether these findings generalize to states other than MA or RI, and to non-urban areas.

In conclusion, we found that NSES at birth had the strongest impact on SBP and DBP, whereas NSES at adulthood more strongly predicted BMI. In addition, we showed evidence that cumulative exposure to low NSES was associated with higher trend of SBP. This suggests that interventions in early life focused on NSES or its downstream consequences—such as health behaviors—have the potential to slow hypertension risk in adulthood; and that interventions in late life focused on NSES or its downstream consequences may reduce the burden of obesity. Future studies should attempt to more rigorously evaluate causality, tease out the potential mechanisms driving these associations, and assess whether these relationships differ by race/ethnicity, or sex.

*Table 1. Descriptive Characteristics of NEFS participants*

| | N | Mean or % | SD |
|---|---|---|---|
| Age | 871 | 44.46 | 2.91 |
| Female (%) | 509 | 58.37 | |
| Race (%) | | | |
| White | 650 | 74.5 | |
| African-American | 165 | 18.9 | |
| Other | 54 | 6.2 | |
| Education level (%) | | | |
| Less than High-school | 130 | 14.9 | |
| High-school | 366 | 42.0 | |
| More than High-school | 376 | 43.1 | |
| Mother's Education level | | | |
| Less than High-school | 461 | 52.9 | |
| High-school | 275 | 31.5 | |
| More than High-school | 136 | 15.6 | |
| Father's Education level | | | |
| Less than High-school | 492 | 56.4 | |
| High-school | 251 | 28.8 | |
| More than High-school | 129 | 14.8 | |
| Systolic Blood Pressure | 871 | 117.51 | 15.83 |
| Diastolic Blood Pressure | 871 | 76.27 | 15.84 |
| Body Mass Index | 871 | 30.02 | 7.76 |

*Table 2. Neighborhood socioeconomic characteristics*

| | Birth | | Childhood | | Adulthood | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| Median Household Income* | 26,095 | 8,350 | 39,534 | 9,300 | 47,620 | 21,136 |
| % Unemployed | 5.9 | 2.7 | 4.2 | 1.7 | 5.3 | 3.6 |
| % With High-school degree or less | 87.3 | 14.0 | 83.7 | 10.7 | 48.6 | 16.3 |

*Income from 1960s and 1970s were adjusted for inflation rates to 2000.

*Table 3. Neighborhood socioeconomic mobility of NEFS participants across the life-course*

| | N | % |
|---|---|---|
| Stayed in low NSES | 220 | 25.3 |
| Moved from low to high NSES | 182 | 20.9 |
| Moved from high to low NSES | 151 | 17.3 |
| Stayed in high NSES | 249 | 28.6 |

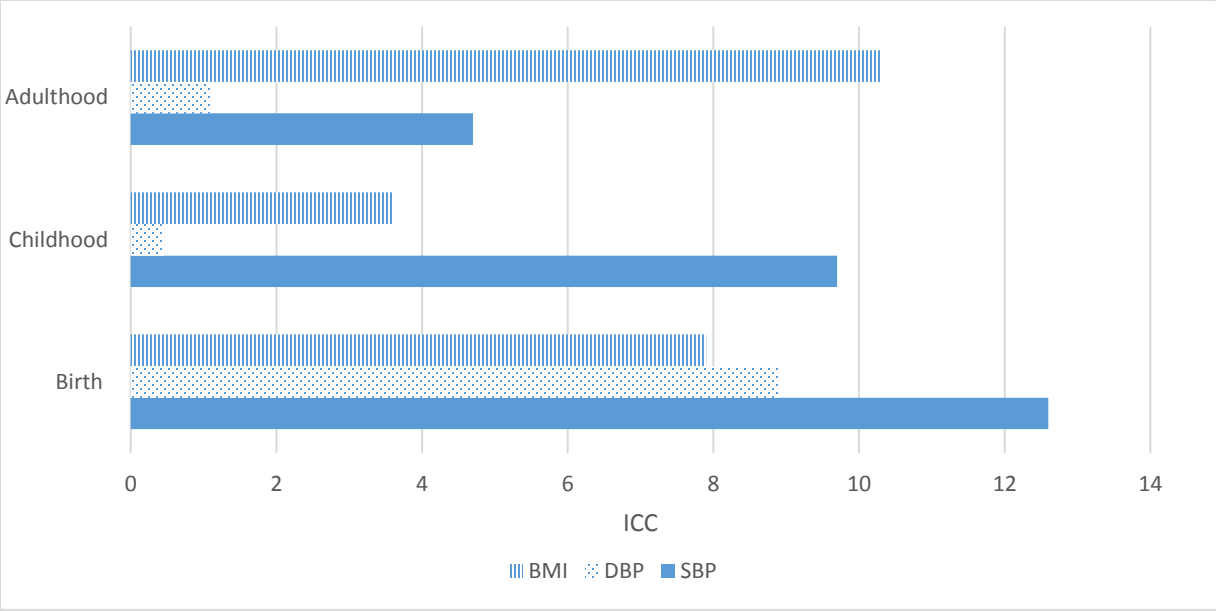



*Figure 1. Intra Class Correlation across the life-course*

*Table 4. Three-level Models for Adjusted Associations of Neighborhood Socioeconomic Status (NSES) and Adult Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP) and Body Mass Index (BMI) stratified by wave (N=459)*

| | SBP | | | | DBP | | | | BMI | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Estimate | 95% Confidence Limits | | | Estimate | 95% Confidence Limits | | | Estimate | 95% Confidence Limits | | |
| **NSES z-score at Birth [1]** | -2.68 | -5.10 | -0.26 | * | -1.59 | -3.22 | 0.04 | + | -0.78 | -2.05 | 0.50 | |
| **NSES z-score at Childhood [2]** | -1.47 | -3.57 | 0.62 | | -0.88 | -2.24 | 0.49 | | -0.48 | -1.56 | 0.60 | |
| **NSES z-score at Adulthood [3]** | -2.22 | -4.18 | -0.26 | * | -1.45 | -2.76 | -0.13 | * | -1.14 | -2.19 | -0.08 | * |

*[1] Adjusted for age, gender, race, individual and parental SES*

*[2] Adjusted for age, gender, race, individual SES, parental SES , and birth NSES*

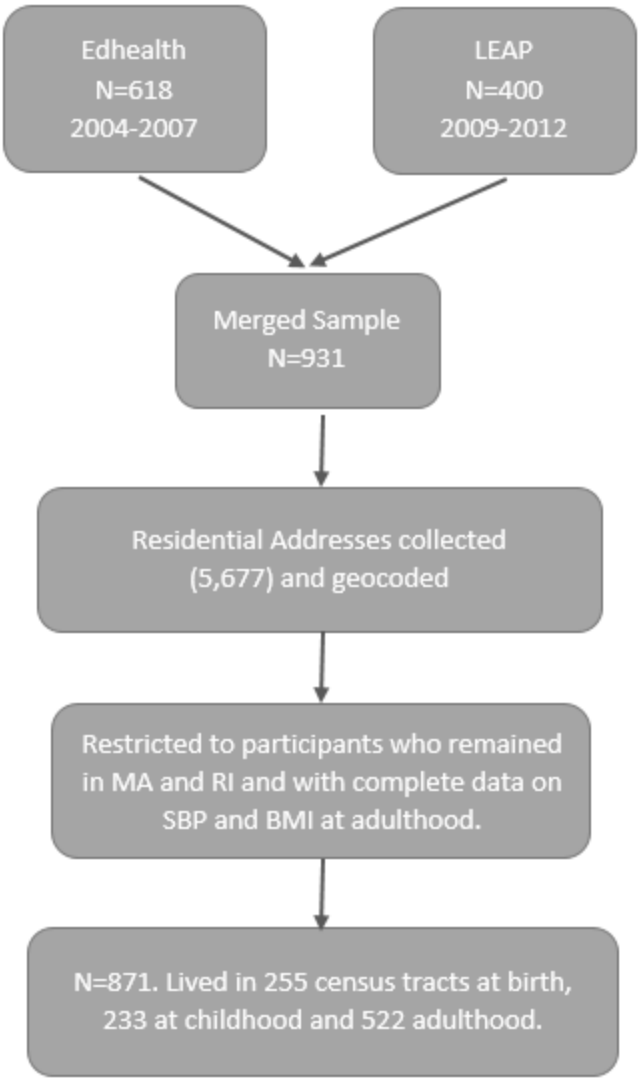
*[3] Adjusted for age, gender, race, individual SES, parental SES , birth and childhood NSES*

*Signif. codes: * <0.05 , + <0.1*

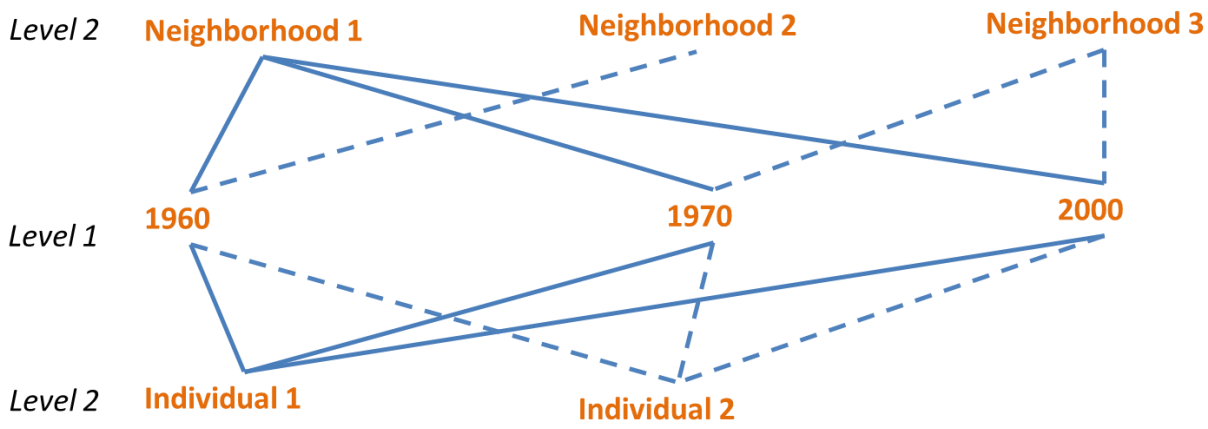
*Table 5. Cross-classified Multilevel Models. Adjusted Associations of Cumulative Neighborhood Deprivation across the life-course and Systolic Blood Pressure (SBP) and Body Mass index (BMI) (N=871)*

| | SBP | | | | BMI | | | |
|---|---|---|---|---|---|---|---|---|
| Models | Estimate | 95% Confidence Limits | | | Estimate | 95% Confidence Limits | | |
| NSES Mobility (ref: stayed in high NSES) | | | | | | | | |
| Stayed in low NSES | 1.60 | -0.26 | 3.526 | + | 0.54 | -0.04 | 1.14 | + |
| Moved from low to high NSES | 1.10 | -0.81 | 3.024 | | 0.49 | -0.07 | 1.07 | + |
| Moved from high to low NSES | 0.09 | -1.85 | 2.04 | | 0.54 | -0.04 | 1.12 | + |

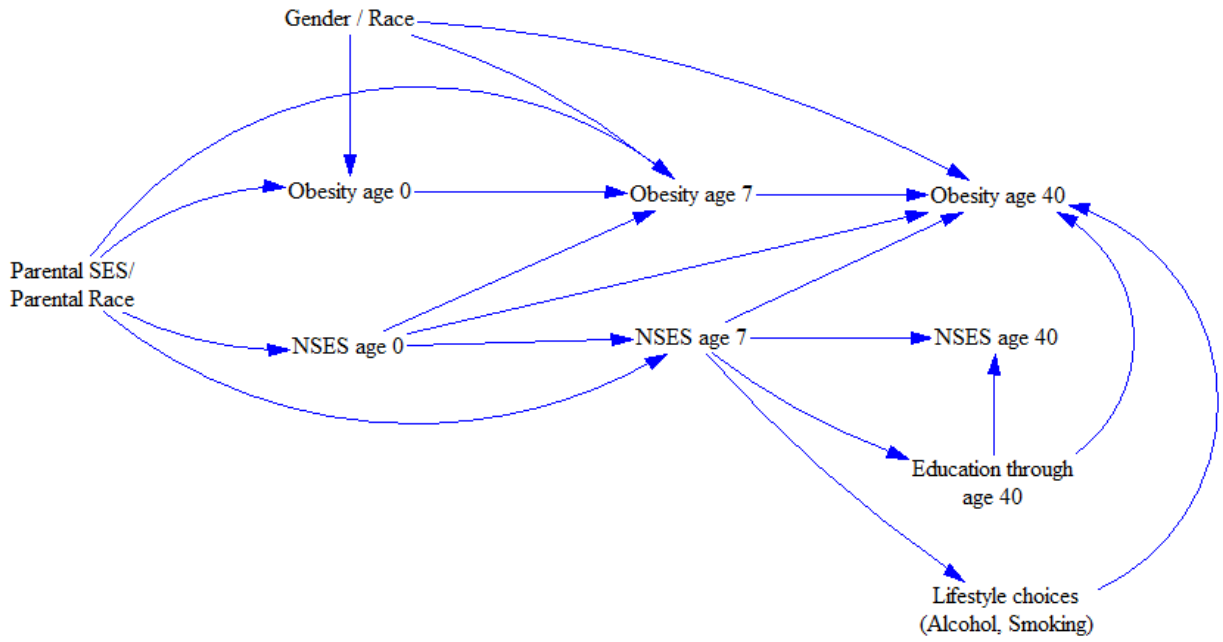
*Adjusting for age, sex, race, individual SES and parental SES*




*Supplemental Figure 1. Flow Chart*




*Supplemental Figure 2. Cross-classified Models*




*Supplemental Figure 3. Directed Acyclic Graph (DAG)*

*Supplemental Table 1. Three-level Models for Adjusted Associations of Neighborhood Socioeconomic Status (NSES) and Adult Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP) and Body Mass Index (BMI) stratified by wave using Multiple Imputation with Chained Equations (MICE)*

| | SBP | | | | DBP | | | | BMI | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Estimate | 95% Confidence Limits | | | Estimate | 95% Confidence Limits | | | Estimate | 95% Confidence Limits | | |
| NSES z-score at Birth [1] | -1.98 | -3.55 | -0.41 | * | -0.87 | -1.98 | 0.24 | | -0.22 | -1.06 | 0.62 | |
| NSES z-score at Childhood [2] | -0.52 | -1.91 | 0.87 | | -0.43 | -1.40 | 0.53 | | -0.42 | -1.09 | 0.24 | |
| NSES z-score at Adulthood [3] | -1.14 | -2.55 | 0.26 | | -2.45 | -1.19 | 0.70 | | -0.82 | -1.51 | -0.13 | * |

*[1] Adjusted for age, gender, race, individual and parental SES*

*[2] Adjusted for age, gender, race, individual SES, parental SES , and birth NSES*

*[3] Adjusted for age, gender, race, individual SES, parental SES , birth and childhood NSES*

*Signif. codes: * p<0.05*

# Chapter 2: Longitudinal associations between neighborhood access to green space in different periods in life, and blood pressure and body mass index

## Abstract

Background: Neighborhood access to green space is increasingly investigated as a determinant of cardiovascular disease (CVD). However, longitudinal studies on the effects of neighborhoods are rare, hampering the ability of researchers to address questions on causality and critical stages in the life-course. This research aims to evaluate the life course association between green space and systolic blood pressure (SBP) and body mass index (BMI), two major CVD risk factors, and assess vulnerable periods of life in which individuals might be more susceptible to their surroundings.

Methods: We used longitudinal data from the New England Family Study with a 48-year follow-up to evaluate how access to green space at birth, childhood and adulthood is associated with SBP and BMI growth across the life course. Access to green space at each time point was evaluated using three measures: distance to the closest green space, average area of green space, and green space density in the neighborhood. Growth Curve Models were used to examine the longitudinal association between time-varying markers of neighborhood accessibility to green space, and SBP and BMI.

Results: Findings suggest that living one mile farther away from a green space at childhood, is associated with a 3.4 mmHg increased rate of growth in SBP across the life-course (95% CI: 0.3,6.6), adjusting for age, race, sex, individual and parental socioeconomic status, and neighborhood socioeconomic status. Higher green space density at birth was also statistically associated with a decreased rate of growth of SBP, but at a lower extent (-0.2 mmHg, 95%CI:-0.3,

-0.02). Finally, green space did not show statistically significant associations with BMI growth rate.

Conclusion: Results suggest that childhood may be a critical time period where closer proximity to green space can reduce the burden of hypertension, a risk factor of CVD. Studies across diverse populations are needed to confirm or refute these novel findings.

## Introduction

The crucial role of green space on health has been considerably noticed, and policies have been proposed to make neighborhoods healthier places to live for everyone[55, 56]. Research shows that green space is linked to a remarkable breadth of positive health outcomes, such as improved mental health,[57] reduced blood pressure[58], improved self-perceived general health[59], and lower overall and cardiovascular mortality[60]. However, evidence on how to improve health outcomes with exposure to green space currently remains limited[57]; knowledge is lacking on physical attributes of green space such as size, as well as vulnerable periods of life in which individuals might be more susceptible to their natural surroundings.

Urbanization is emerging as a central global health challenge of the 21st century[61], with cities becoming epicenters for chronic diseases in adults[62]. In particular, in the United States, high blood pressure (BP) and obesity are two major public health burdens and major risk factors of chronic diseases such as arthritis, diabetes, dyslipidemia, and cardiovascular disease (CVD)[63]. Green space may influence CVD risk factors through a range of mechanistic pathways. For example, higher accessibility to green space may be linked to lower blood pressure through providing an aesthetic environment that can lead to relaxation and reduction in stress, or by encouraging interaction with nature and time spent outdoors that can lead to increased physical activity (Supplementary Figure 1)[8]. A recent study from Australia showed that people who made long visits to green spaces had lower rates of high blood pressure[57]. An experimental study concluded that, for coronary artery disease patients, restorative walking in a park significantly reduced stress and blood pressure compared to walking in an urban street

environment[64]. In addition, greenness has been shown to be associated with lower levels of obesity in adults[65-67] and children[68, 69].

However, most of the current research on green space and health has been cross-sectional and therefore susceptible to reverse causation. In addition, most studies have used a single or an aggregate measure of green space (e.g. percentage of green space within 1km radius of residence)[67], which may mask the effect of other features, such as type and quality of the green space[70]. Finally, although research into the influence of green space on adult health is increasing, evidence for children remains less developed. We used longitudinal data to evaluate the critical period model which states that exposures at a specific time have long-lasting effects on the function of the body[24]. To our knowledge, this innovative study is the first to longitudinally investigate critical stages of exposure to green space measured quantitatively, across the life-course. Our working hypothesis is that CVD risk in adulthood will be greater for participants who lived in neighborhoods with lower accessibility to green space during childhood. Our rationale is that childhood would be a sensitive period in the life course where healthy behaviors, such as increased levels of physical activity, are established and could extend into adulthood.

Results from this study provide new evidence that during critical periods of susceptibility in life, greater access to neighborhood green space may slow the development of CVD. Given the urgency brought about by global urban health challenges, studies like this are needed to directly inform city planning and design[71].

## Data and Methods

### *Data*

We used data from the New England Family Study (NEFS), a follow-up study to the National Collaborative Perinatal Project (NCPP), which was conducted from 1959-1966 in twelve cities throughout the United States. Twelve university-affiliated medical centers participated in this national study, two of which were in New England (Harvard Medical School and Brown University). The Project enrolled over 50,000 pregnancies, approximately 18,000 in New England, between 1959 and 1966, and evaluated offspring on 12 occasions prenatally through age 8 years.

Hypothesized social, behavioral, and biological pathways leading to adult obesity were assessed when offspring were interviewed as adults between 2005 and 2011. Analyses for this study was achieved through the merging of data from two of the sub-samples that comprise NEFS: Longitudinal Effects of Aging Perinatal Project (LEAP) and Pathways Linking Education and Health in Middle Adulthood (EdHealth). Collectively, this yields an analytical dataset of 931 participants who were followed from birth through adulthood. Participants' complete addresses were extracted from the interview files, as reported by their parents, from birth and age 7 and subsequently later geocoded (assigned latitude and longitude) at each time point using ArcGIS version 10.4.1. A detailed description of the geocoding process has been previously published[72]. We restricted the analytical sample to participants with a residential address in MA or RI throughout the life-course (n=803) and participants with complete address information available for all three time-points (n=421). Participants missing additional pertinent covariates

including education (n=2), were also excluded. The final analytic sample included 419 participants (see supplementary figure 3).

The primary exposure is accessibility to green space at birth (around 1960s), childhood (around 1970s), and adulthood (around 2000s). We define green space as "land that is partly or completely covered with grass, trees, shrubs, or other vegetation"[73], which can include parks, community gardens and cemeteries[70], and that is larger than 1 hectare[74]. A one-mile buffer was built around the geocoded NEFS participants at each time point and the accessibility to green space was estimated within the buffer areas[70, 75]. Although no common standard exists for proximity to parks, this one-mile buffer was thought to be a reasonable walking distance and has been used in several past studies of green space[75]. In addition, a national study recently showed that a one-mile buffer area is desirable to measure potential walkability within a neighborhood[76]. We created a historical database of green space in 1960 and 1970 within the buffer areas, by comparing historical topographic maps from the United States Geological Survey [77] to current recreational open space data layers from the Massachusetts GIS Department [78] and the Rhode Island Conservation Lands Public Access [79], and assessing whether existing green areas were available formerly. We further classified green space as follows: (1) Parks, (2) Playgrounds, (3) Blue Spaces and Beaches, (4) Swamps, (5) Golf-courses, country clubs, and zoos, and (6) Cemeteries. Our main analyses include green space from all categories except for blue spaces and swamps. The geographic databases used to classify the exposure in this study focused primarily on green space, and as such were an unreliable measure of blue spaces. Moreover, researchers have shown potential benefits of blue spaces on human health independent of green spaces[80]. Areas such as swamps were likely not readily accessible to participants, and therefore

not conducive to the types of activity that are hypothesized to improve cardiovascular health. At all time-points, access to green space was evaluated by using three objective measures obtained through GIS methodology: distance to the closest green space, the average area of green space within the neighborhood (or buffer area) and green space density. The distance from the participant's address to the closest park was calculated in miles using the street network distance[81], which has been shown to be more appropriate in simulating walking behavior than Euclidean distance[75]. Average area was assessed in hectares and green space density was evaluated as the number of green spaces within the buffer areas. Thus, access to green areas will be operationalized at the individual level by virtue of the fact that it is derived from both an individual level and a neighborhood environment measure.

The primary outcome variables are (a) systolic blood pressure (SBP), and (b) body mass index (BMI) assessed at the three time points birth, childhood and adulthood. The mean of the second and third seated readings on the non-dominant arm were averaged to calculate SBP. If the last two blood pressure readings varied by >5 mmHg, additional measures were taken at 5-minute intervals until two consecutive readings were within 5 mmHg, or a maximum of six resting measures were taken. Height and weight were directly assessed by trained research technicians, weight was measured using calibrated scales in participants in light clothing without shoes; height was assessed using calibrated stadiometers. BMI was calculated as weight (kg) divided by the square of the height ($m^2$).

Individual-level covariates included age, gender (male, female), and race (White, African-American, other). In addition, individual socioeconomic status (SES) was assessed as adult attained education (less than high-school, high-school, more than high-school), parental SES was

assessed as mother's education (less than high-school, high-school, more than high-school), and father's education (less than high-school, high-school, more than high-school). All individual-level covariates were assessed at adulthood, except for parental SES which was assessed at birth. Finally, neighborhood level covariates included neighborhood socio-economic status (NSES) measured at each time point as a standardized mean score of: percent with less than a high school education, percent of unemployment, and median household income[7, 48]. A detailed description of this measure can be found in Jimenez et al.[72].

*Statistical Analysis*

We used growth curve analysis to examine the longitudinal association between access to green space and the CVD risk factors: SBP and BMI. Growth curve analysis estimates mean growth trajectories based on the repeated measures of SBP and BMI over time and allows individual growth trajectories to randomly vary around the mean. This approach accounts for changes in scale and variance of growth measures over time and uses all available data for each individual [82]. Since we were interested in evaluating the critical time period where access to green space may have a stronger effect on CVD risk, the models were stratified by time; thus we have three models (one for birth, one for childhood and one for adulthood) per green space measure. Accordingly, row 1 in Table 3 evaluates the association between proximity to green space at birth and SBP, the second row in Table 3 evaluates the association between proximity to green space at childhood and SBP, and so on. All models controlled for the following confounders: age, gender, race, individual socioeconomic status, and parental socioeconomic status. Furthermore, to help account for neighborhood selection, wherein affluent neighborhoods may have higher

access to green space, we further adjusted for NSES at each time point. We ran the models sequentially, starting with individual covariates and adding neighborhood level covariates.

A critically important step in any growth model is the identification of the optimal functional form of the trajectory over time; that is, it must be established exactly how the repeated measures change as a function of time [83]. We evaluated distinct functional forms for time and compared Akaike's and Bayesian Information Criteria (AIC and BIC). Growth modeling is highly flexible in terms of the inclusion of a variety of complexities including partially missing data, unequally spaced time points, non-normally distributed or discretely scaled repeated measures, complex nonlinear or compound-shaped trajectories, time-varying covariates, and multivariate growth processes [83]. However, we further used multiple imputation by chained equations (MICE) to account for missing data as a sensitivity analysis. MICE operates under the assumption that given the variables used in the imputation procedure, the missing data are missing at random (MAR). We used subject-matter knowledge and the hypothesized DAG (Supplementary Figure 2), to maximize the number of predictors to be used in the multiple imputation procedure and have minimal bias and maximal certainty. Including as many predictors as possible tends to make the MAR assumption more plausible, thus reducing the need to make special adjustments for NMAR mechanisms[84]. Statistical significance was assessed at the 0.05 level. Spatial analysis was done using ArcGIS software version 10.4.1., data management was conducted in SAS version 9.4, and statistical analysis was performed in R software version 3.4.0.

## Results

The analytical sample included 419 individuals, out of which 58% were female. The mean age was 43.5 years. Among included participants, 75.7% were White, 17.9% were African-American and 6.4% were categorized as other. About 12% of the sample had less than a high-school education, and paternal education level was less than high-school for more than half of the sample (Table 1). The average values for the outcomes at each time point are presented in Table 2. SBP at adulthood was on average 117.8 mmHg (SD =16.5) and BMI was 30.1 (SD =8.0). Descriptive characteristics of the green space measures at the three time points showed that, on average, participants had six to seven green spaces in their neighborhoods. The average area of green space seemed to fluctuate, and the standard deviation for childhood and adulthood was large due to state parks included in our analysis. Finally, participants lived about 0.5 miles, on average, away from a green space at birth, childhood and adulthood (Table 2).

Table 3 shows the association between the SBP growth trajectories and access to green space at critical time periods. Results for proximity to green space suggested that living one mile farther from a green space at childhood was associated with an increase in growth rate of SBP of 3.2 mmHg (95%CI: 0.1, 6.3). This association was significant after controlling for confounders. Results for density of green space showed that having one more green space at birth was associated with a decrease in growth rate of SBP of -0.2 mmHg (95%CI: -0.4, -0.02). This association was significant after controlling for confounders. Results for average area of green space showed no significant associations with growth rates of SBP. After adjusting for NSES (Model 2, Table 3), the associations of green space density were attenuated by 1-18%, with greater attenuation for the association of green space density at adulthood. However, the

association of proximity to green space at childhood and SBP intensified to 3.5 mmHg (95%CI: 0.4, 6.7), after adjustment of NSES. The patterns of association were not substantially changed, where the effect size of proximity to green space was higher than the effect sizes of average area and green space density. Furthermore, childhood proximity to green space seemed to have the strongest association with SBP growth rates across all models.

Finally, Table 4 shows the association between BMI growth trajectories and access to green space at each time point. The results for proximity to green space suggested that living one mile farther from a green space at adulthood was associated with an increase in growth rate of BMI of 0.25 $kg/m^2$ (95%CI: -0.03, 0.53, $p<0.1$), after controlling for confounders. However, this result did not reach significance. Results for green space average area and green space density within the neighborhood did not show any significant associations with BMI growth trajectories. After adjusting for NSES (Model 2, Table 4), the association between proximity to green space in adulthood and BMI was attenuated by 2%, however this association was not significant. Associations of average area of green space and density of green space with BMI were not statistically significant after adjustment for NSES. Stronger associations were shown between proximity to green space and BMI, compared to average area or density, suggested by the effect sizes, however, none reached statistical significance. Lastly, AIC and BIC values for distinct functional forms of time (linear, quadratic and cubic) indicated that a linear association best fit the random intercept-only model.

## Discussion

We conducted a longitudinal analysis to evaluate the association between life-course accessibility to green space and the growth rate of CVD risk factors. Our analysis suggested that childhood proximity to green space had the strongest association with SBP, such that living one mile farther from a green space at childhood was associated with an increase of 3.5 mmHg (95%CI: 0.3, 6.6) rate of change in SBP. In addition, results suggested that proximity to green space had the sharpest impact on the slope of SBP across the life-course, compared to average area or density of green space within the neighborhood. Density of green space seemed to have the second largest impact, where having one more green space at birth was associated with a significant decrease of 0.2 mmHg in SBP slope across the life-course (95%CI: -0.4, -0.02) after adjusting for individual and neighborhood level confounders. However, the pattern of associations of green space across the life-course differed for SBP and BMI gain trajectories. For BMI, proximity to green space also seemed to have the sharpest impact on the slope, yet the effect size was considerably smaller and in some cases in the opposite direction to what was expected, although not statistically significant. For instance, living one mile farther away from a green space in childhood was associated with a decrease in the BMI trajectory of -0.4 $kg/m^2$ (95%CI: -0.9, 0.1). After controlling for NSES (Model 2, Table 4), results suggested that living one mile farther away from a green space at adulthood was associated with an increase of 0.2 (95%CI: -0.04, 0.5) in BMI trajectory ($p<0.1$). Results for the association of average area and density of green space in the neighborhood with BMI were not statistically significant.

Our finding that access to green space is distinctly associated with SBP and BMI has potentially important implications for the design of health interventions, and also reveals new

hypotheses that warrant further attention. Previous studies have also found diverse associations of green space with different outcomes. For instance, one recent study looking at the associations between the duration of exposure to nature and health outcomes, found that people who made long visits to green spaces had lower rates of depression and high blood pressure, but not greater social cohesion[57]. The null effect of green space on BMI found in this study might be due to low power or due to unmeasured mechanisms like physical activity. A recent U.S. based study looking at proximity to parks and physical activity among adults did not find a significant association either[75]. The relationship between green space and health may be mediated through a number of possible mechanisms including air quality, physical activity, social contacts, stress, and restoration. These mechanisms have a number of possible modifiers, such as distance to green space, accessibility factors, perceived safety in the green space, societal context, cultural context, gender, age, and socioeconomic status[70]. In addition, there is some indication in the literature that some groups of the population may benefit more than others, such as people with lower socio-economic positions and women, including pregnant women. Roe et al. (2013) found that greater percentages of green space had positive effects on cortisol levels in women, but not in men[85]. Perhaps the association between proximity to green space and BMI would be more accentuated in other populations. Future research should evaluate vulnerable groups of the population and the magnitude of the contribution of various mechanisms of the benefits of green space to health.

The prospective cohort design and long follow-up period from birth to adulthood (48-year follow-up), in addition to the distinctive measures of access to green space are strengths of this study. Only four studies since 2001 investigating the impact of green space on health have been

longitudinal[71]. In addition, our examination of the theoretical model of critical periods and the advanced methodology used to evaluate our research question is an improvement on the approaches used in previous studies[6]. These methods allowed us to identify critical life stages during which higher access to green space may lower the risk of CVD, and to assess which aspects of accessibility to green space are best suited to improve and/or sustain long-term health benefits[70]. Our analytical approach also enabled us to estimate between-person differences in within-person trajectories and their associations with differences in access to green space at three different time points.

Some limitations in our study also require discussion. First, although we adjusted for individual and neighborhood level characteristics, residual confounding cannot be excluded. Parental history of hypertension, a predictor of blood pressure of the offspring [86], for example, may also affect access to green space as a recent prospective study showed that health problems at baseline (high blood pressure among others) predicted subsequent neighborhood poverty[23], and thus likely lower access to green space. Second, some exposure misclassification is possible and thus results may be biased. However, any misclassification would be non-differential since the measurement errors for the outcome and the exposure would be independent as they come from different data sources. Classification of green space was an interesting challenge that we faced and our team is working on further analysis to evaluate the differential impact of types of green space on health. Third, we do not have information of workplace proximity to green space in adulthood (or school proximity in childhood) a possible source of unmeasured confounding. Fourth, we lack data on actual use of the greenspace. Fourth, we restricted the analysis to participants who remained in MA or RI with complete data which limits the internal

generalizability of our results and may potentially induce selection bias. However, we ran sensitivity analysis using MICE to account for hierarchy of the data and results were similar (please refer to Supplementary Tables 1 and 2). Finally, we need to consider issues of multiple statistical testing. In this study we evaluated 36 different models (18 versions of model 1 considering 3 different green space measurements at 3 different time points for 2 outcomes; and 18 versions for model 2 - model 1+ NSES), out of which we found 4 significant associations (11%), and thus it may be possible that we are incorrectly rejecting the null hypothesis. Future studies should consider the false discovery rate (expected proportion of type I errors) to account for multiple testing.

Our results shed light on the importance of proximity to green space which should be taken into consideration in the future planning of green spaces in urban areas. While we do not have data on type of green space, our analysis included distinct measures of access of green space that suggested that living closer to a green space may be more influential in reducing the burden of CVD risk than the size or the density of green space in a neighborhood. This type of study is vital to ensuring that investment in green space provision can be used effectively to meet current public health challenges. In the US alone, cities invested over $6 billion in 2015 for the provision, management, and enhancement of public green spaces[57]; and although research has suggested that neighborhood interventions to increase access to recreation facilities are a cost-effective means to improve health behavior, more information on the necessary features is fundamental as public parks and shaded areas may become even more important with ongoing climate change, further urban sprawl, and increasing obesity prevalence[87].

*Table 1.Characteristics of the Analytic Sample*

| | **Mean (SD)** |
|---|---|
| **Demographic Variables** | |
| Age | 43.5 (2.7) |
| Sex - % Female | 58.2 |
| Race | |
| % White | 75.7 |
| % African American | 17.9 |
| % Hispanic | 2.9 |
| Education | |
| % Less than High-school | 12.4 |
| % High-school education | 35.1 |
| % More than High-school | 52.5 |
| Mother's Education | |
| % Less than High-school | 50.4 |
| Father's Education | |
| % Less than High-school | 55.1 |

*Table 2. Blood pressure, body mass index, and green space characteristics across the life-course*

| | Birth | | Childhood | | Adulthood | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| Health Variables | | | | | | |
| Systolic Blood Pressure (mmHg) | 70.1 | 15.8 | 103.8 | 11.4 | 117.8 | 16.5 |
| Body Mass Index (kg/m2) | 12.7 | 1.6 | 16.3 | 2.1 | 30.1 | 8.0 |
| Green Space Variables | | | | | | |
| Green Space Density (n) | 6.4 | 5.4 | 6.6 | 5.2 | 7.1 | 5.6 |
| Green Space Area (Ha) | 9.3 | 14.8 | 31.1 | 319.1 | 16.5 | 81.2 |
| Closest Distance (mi) | 0.5 | 0.3 | 0.5 | 0.3 | 0.5 | 0.5 |

*Table 3. Conditional Growth Models for Adjusted Associations of Access to Green Space and Systolic Blood Pressure*

| Greenspace Measure | Time | Model 1 | | | | Model 2 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | β | 95% Confidence Limits | | | β | 95% Confidence Limits | | |
| **Distance (mi)** | **Birth** | 1.73 | -1.22 | 4.69 | | 1.71 | -1.25 | 4.66 | |
| | **Childhood** | 3.23 | 0.16 | 6.30 | * | 3.49 | 0.36 | 6.62 | * |
| | **Adulthood** | 1.27 | -0.61 | 3.14 | | 1.22 | -0.65 | 3.09 | |
| **Average Area (Ha)** | **Birth** | -0.04 | -0.10 | 0.01 | | -0.04 | -0.10 | 0.01 | |
| | **Childhood** | 0.00 | -0.01 | 0.00 | | 0.00 | -0.01 | 0.00 | |
| | **Adulthood** | 0.00 | -0.01 | 0.01 | | 0.00 | -0.01 | 0.02 | |
| **Density (n)** | **Birth** | -0.21 | -0.39 | -0.02 | * | -0.20 | -0.39 | -0.02 | * |
| | **Childhood** | -0.14 | -0.32 | 0.04 | | -0.16 | -0.34 | 0.02 | + |
| | **Adulthood** | -0.12 | -0.27 | 0.03 | | -0.10 | -0.25 | 0.06 | |

*Model 1: Adjusted for age, sex, race, individual SES and parental SES*

*Model 2: Model 1 + Neighborhood Socioeconomic status*

* $p<0.05$

+ $p<0.1$

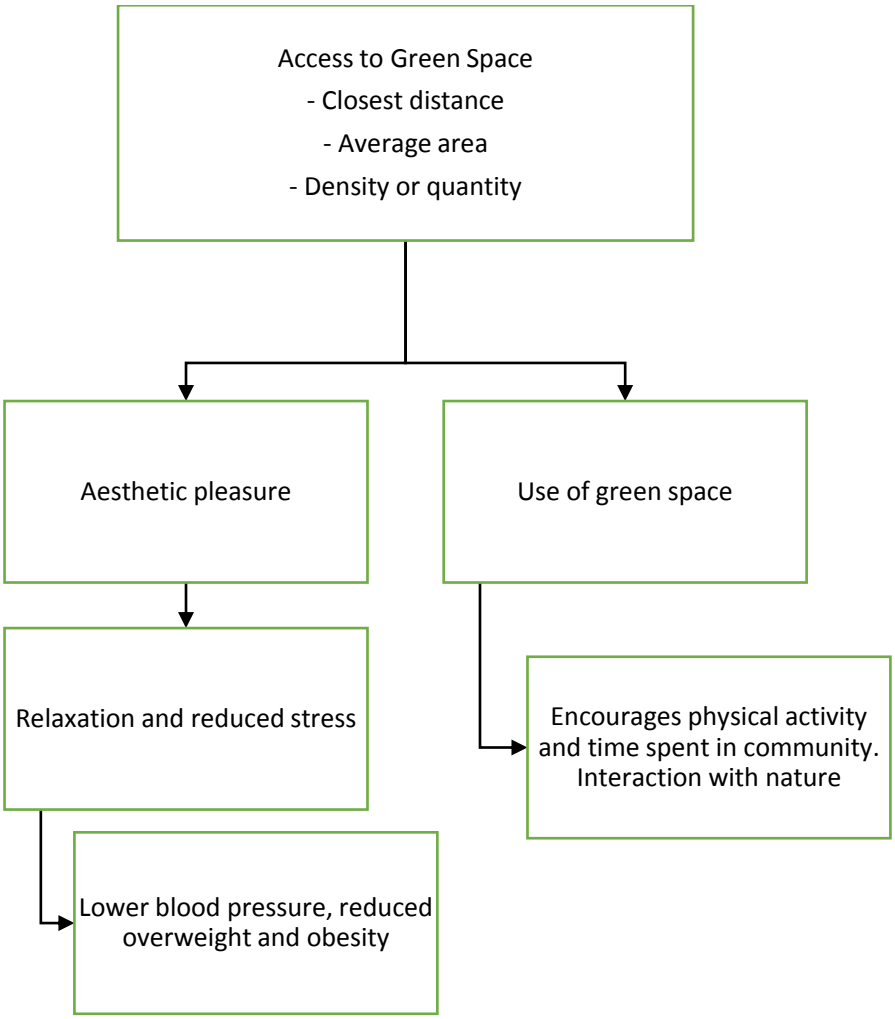
*Table 4. Conditional Growth Models for Adjusted Associations of Access to Green Space and Body Mass Index*

| Greenspace Measure | Time | Model 1 | | | | Model 2 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | β | 95% Confidence Limits | | | β | 95% Confidence Limits | | |
| **Distance (mi)** | **Birth** | 0.02 | -0.46 | 0.50 | | 0.02 | -0.46 | 0.49 | |
| | **Childhood** | -0.35 | -0.84 | 0.15 | | -0.43 | -0.93 | 0.07 | + |
| | **Adulthood** | 0.25 | -0.03 | 0.53 | + | 0.24 | -0.04 | 0.52 | + |
| **Average Area (Ha)** | **Birth** | -0.01 | -0.02 | 0.00 | | -0.01 | -0.02 | 0.00 | |
| | **Childhood** | 0.00 | 0.00 | 0.00 | | 0.00 | 0.00 | 0.00 | |
| | **Adulthood** | 0.00 | 0.00 | 0.00 | | 0.00 | 0.00 | 0.00 | |
| **Density (n)** | **Birth** | 0.00 | -0.03 | 0.03 | | 0.00 | -0.03 | 0.03 | |
| | **Childhood** | 0.00 | -0.03 | 0.03 | | 0.01 | -0.02 | 0.03 | |
| | **Adulthood** | -0.01 | -0.03 | 0.01 | | -0.01 | -0.03 | 0.02 | |

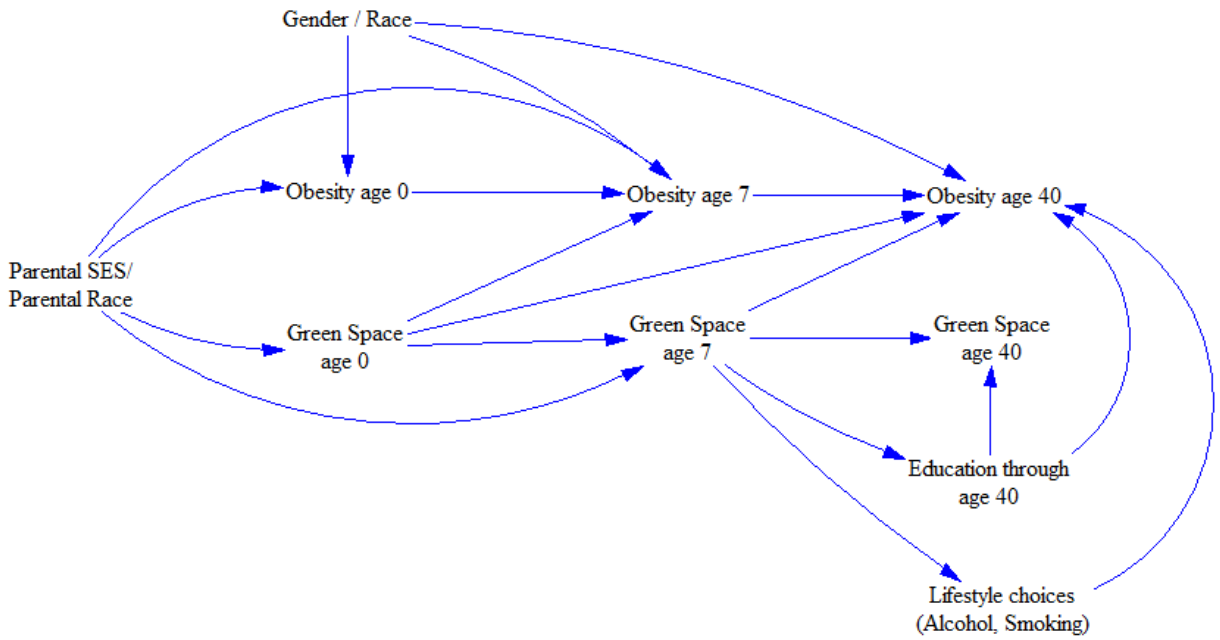
*Model 1: Adjusted for age, sex, race, individual SES and parental SES*

*Model 2: Model 1 + Neighborhood Socioeconomic status*

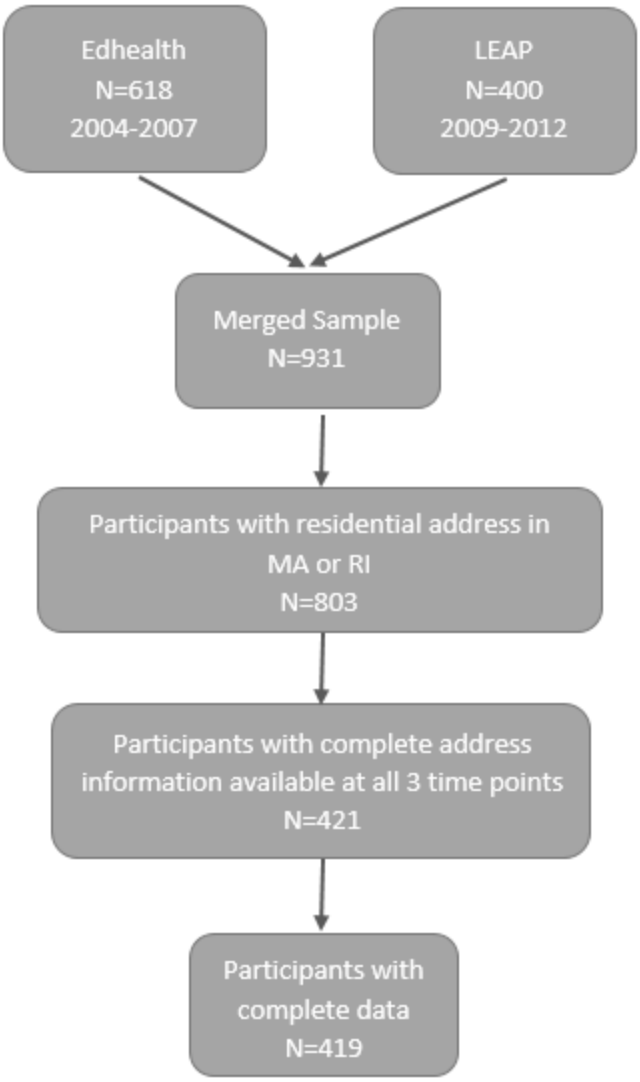
+ $p<0.1$




*Supplemental Figure 1. Pathways through which access to green space may affect blood pressure and obesity*




*Supplemental Figure 2. Directed Acyclic Graph*




*Supplemental Figure 3. Flow Chart*

*Supplementary Table 1. Conditional Growth Models for Adjusted Associations of Access to Green Space and SBP using MICE*

| **Supplementary Table 1. Conditional Growth Models for Adjusted Associations of Access to Green Space and Systolic Blood Pressure (SBP) using Multiple Imputation with Chained Equations** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Greenspace Measure** | **Time** | **Model 1** | | | | **Model 2** | | | |
| | | **β** | **95% Confidence Limits** | | | **β** | **95% Confidence Limits** | | |
| **Distance (mi)** | **Birth** | 1.07 | -1.04 | 3.18 | | 0.90 | -1.22 | 3.01 | |
| | **Childhood** | 2.01 | -0.89 | 4.91 | | 2.37 | -0.55 | 5.29 | |
| | **Adulthood** | 1.23 | -0.21 | 2.66 | | 1.26 | -0.19 | 2.70 | |
| **Average Area (Ha)** | **Birth** | -0.02 | -0.06 | 0.02 | | -0.02 | -0.06 | 0.02 | |
| | **Childhood** | 0.00 | 0.00 | 0.00 | | 0.00 | 0.00 | 0.00 | |
| | **Adulthood** | 0.00 | -0.01 | 0.01 | | 0.00 | -0.01 | 0.01 | |
| **Density (n)** | **Birth** | -0.19 | -0.32 | -0.06 | ** | -0.18 | -0.31 | -0.04 | ** |
| | **Childhood** | -0.07 | -0.22 | 0.09 | | 0.12 | -0.29 | 0.04 | |
| | **Adulthood** | -0.15 | -0.28 | -0.02 | * | -0.13 | -0.27 | 0.00 | ** |

*Model 1: Adjusted for age, sex, race, individual SES and parental SES*

*Model 2: Model 1 + Neighborhood Socioeconomic status*

* p<0.05, ** p<0.01

*Supplementary Table 2. . Conditional Growth Models for Adjusted Associations of Access to Green Space and BMI using MICE*

| **Supplementary Table 2. Conditional Growth Models for Adjusted Associations of Access to Green Space and Body Mass Index (BMI) using Multiple Imputation with Chained Equations** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Greenspace Measure** | **Time** | **Model 1** | | | **Model 2** | | |
| | | **β** | **95% Confidence Limits** | | **β** | **95% Confidence Limits** | |
| **Distance (mi)** | **Birth** | -0.03 | -0.38 | 0.32 | -0.05 | -0.40 | 0.30 |
| | **Childhood** | -0.08 | -0.49 | 0.32 | -0.38 | -0.84 | 0.08 |
| | **Adulthood** | 0.02 | -0.19 | 0.24 | 0.02 | -0.19 | 0.24 |
| **Average Area (Ha)** | **Birth** | -0.01 | -0.01 | 0.00 | -0.01 | -0.01 | 0.00 |
| | **Childhood** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | **Adulthood** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **Density (n)** | **Birth** | 0.01 | -0.01 | 0.03 | 0.01 | -0.01 | 0.03 |
| | **Childhood** | -0.01 | -0.03 | 0.02 | 0.00 | -0.02 | 0.03 |
| | **Adulthood** | -0.01 | -0.02 | 0.01 | 0.00 | -0.02 | 0.02 |

*Model 1: Adjusted for age, sex, race, individual SES and parental SES*

*Model 2: Model 1 + Neighborhood Socioeconomic status*

# Chapter 3: A demographic, clinical and behavioral typology of obesity in the United States: an analysis of NHANES 2011-2012

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.

## Introduction

Obesity (body mass index - BMI ≥30 kg/$m^2$) 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 [63]. The American Medical Association recently classified obesity as a disease [88]. However, a conceptualization of individuals with obesity as defined solely by BMI, may not reflect the true heterogeneity encountered in clinical practice [28]. 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 [28, 29]. 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 [89]. 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 [28]. Finally, recent studies also found

obesogenic cluster patterns in children and adolescents with mixed physicial activity/sedentary behaviors differing according to age, gender and socio-economic status [90, 91]. 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 [90]. The clustering of individuals with similar characteristics is a concept that has been applied to understanding the association among different health behaviors [90, 92, 93]. 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.

## Data 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 [94]. A detailed description of the methodology and procedures used in the survey has been previously published [95]. 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 [96, 97]. The analyses presented in this report included adult participants >=20 years of age with obesity to ensure comparability to prior studies [25], 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 [28], 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 [98, 99]. 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 [100]. 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 [101]. 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 [102]. 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 [103]. Alcohol consumption was split into 4 categories (nondrinkers, <1 drink per day, 1-2 drinks per day, and >2 drinks per day) [104]. Smoking was assessed by reporting having smoked >100 cigarettes in life [105]. 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 [28, 92, 93, 106]. 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 [106]. To account for these multiple data types, we used a two-step cluster analysis approach [107]. 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 [107]. Continuous variables were standardized using z-scores due to their differing scales [106].

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 [108]. 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 $Chi^2$ 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 [109]. 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 [89]. 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 [110, 111]. 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 [112-114]. 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 [115]. 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 [89, 116]. 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 [117-119] 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 [120-125]. 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 [126], asthma [127] and the increase in women’s stroke prevalence [128]. 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”[129].

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” [28]. 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 [106]. 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 [106]. 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 [106].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 [130]. 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.

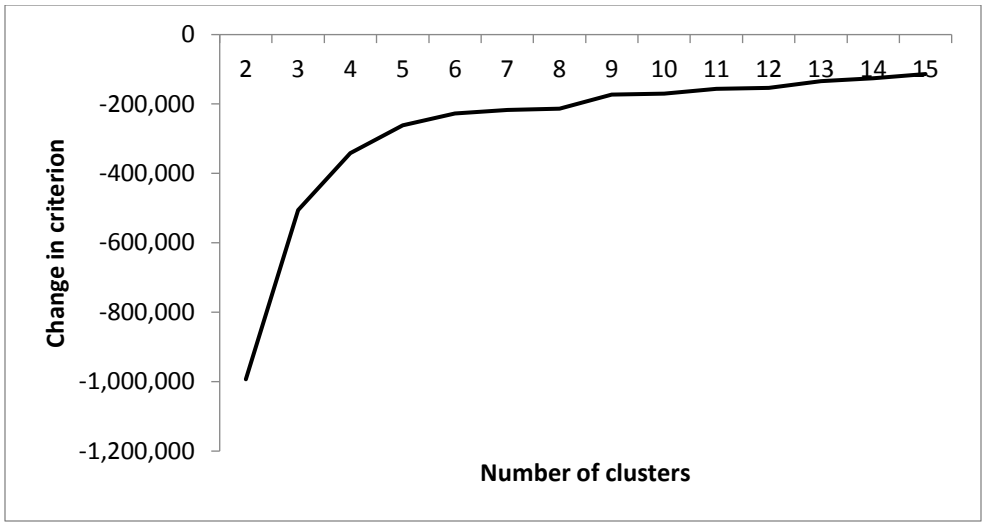



*Figure 1. Change in Schwartz's Bayesian Information Criterion (BIC) across a range of cluster solution*

*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 |
| | | Arthritis (%) | 23.1 | 29.8 |

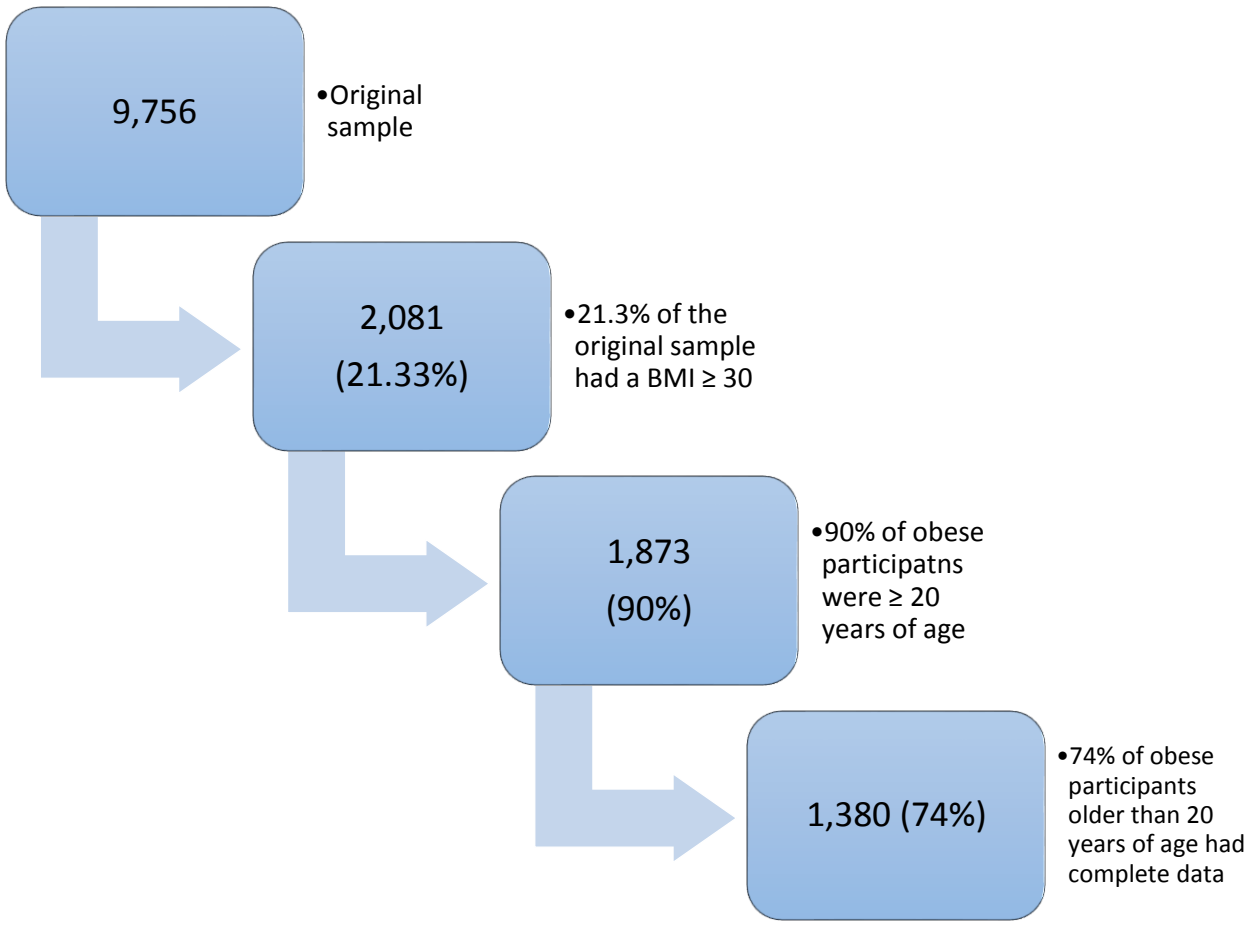
| | | | | |
|---|---|---|---|---|
| | Chronic Diseases | Cancer (%) | 9.7 | 9.9 |
| | | Diabetes (%) | 21.9 | 28.8 |
| | | Sleep Disorder (%) | 9.1 | 14.6 |
| | | Mean Pain (days in month) | 3.3 | 4.0 |
| Behavioural | Physical Activity | Mean physical activity (MET) | 3057.00 | 3181.84 |
| | Smoking | 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 |
| | | Mean Body Mass Index | 28.7 | 35.8 |

*Table 2. Mean Characteristics of the Clusters in the Weighted Sample NHANES 2011-2012*

| 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 | Active men who drink higher amounts of alcohol | |
| | | Sample Size (n) | 87424 | 91611 | 114228 | 72656 | 77905 | 117104 | |
| | | Sample Size (%) | 15.6 | 16.3 | 20.4 | 13.0 | 13.9 | 20.9 | |
| | | Mean Body Mass Index | 34.5 | 34.7 | 37.8 | 36.2 | 36.0 | 34.9 | <0.001 |
| Personal | Age | Mean Age | 48.1 | 58.2 | 58.6 | 43.8 | 38.9 | 34.5 | <0.001 |
| | Sex | Male (%) | 74.8 | 68.5 | 23.0 | 9.2 | 22.5 | 75.3 | <0.001 |
| | | Female (%) | 25.2 | 31.5 | 77.0 | 90.8 | 77.5 | 24.7 | |
| | Race | Mexican American (%) | 1.0 | 2.8 | 6.8 | 0.0 | 16.9 | 21.8 | <0.001 |
| | | Other Hispanic (%) | 1.3 | 3.3 | 6.3 | 0.0 | 19.9 | 12.9 | |
| | | Non-Hispanic White (%) | 97.7 | 85.4 | 61.9 | 93.9 | 0.0 | 52.3 | |
| | | Non-Hispanic Black (%) | 0.1 | 7.7 | 20.6 | 0.0 | 57.6 | 8.7 | |
| | | Other Race (%) | 0.0 | 0.7 | 4.5 | 6.1 | 5.6 | 4.2 | |
| | Income | <$20,000 (%) | 0.0 | 8.7 | 22.1 | 4.9 | 27.5 | 27.3 | <0.001 |
| | | $20,000 - $45,000 (%) | 0.0 | 7.1 | 36.7 | 20.2 | 18.0 | 34.8 | |
| | | $45,000 - $75,000 (%) | 7.4 | 27.6 | 22.7 | 25.5 | 19.8 | 31.9 | |
| | | $75,000+ (%) | 92.6 | 56.5 | 18.4 | 49.4 | 34.7 | 6.0 | |
| Health | Respiratory | Asthma (%) | 17.0 | 34.2 | 75.8 | 18.6 | 12.8 | 3.9 | <0.001 |
| | | Emphysema (%) | 0.0 | 4.4 | 2.5 | 3.3 | 0.0 | 0.3 | 0.003 |
| | Heart & Circulatory | Angina (%) | 4.3 | 8.9 | 6.5 | 0.0 | 0.0 | 0.0 | <0.001 |
| | | Congestive Heart Failure (%) | 1.2 | 6.0 | 8.4 | 0.0 | 0.0 | 0.0 | <0.001 |
| | | Coronary Heart Disease (%) | 0.0 | 12.3 | 4.3 | 0.0 | 0.0 | 0.5 | <0.001 |
| | | Diabetes (%) | 2.0 | 36.6 | 27.1 | 1.2 | 5.6 | 3.7 | <0.001 |
| | | Dyslipdemia (%) | 27.1 | 94.8 | 64.0 | 3.6 | 17.5 | 7.6 | <0.001 |

| Variables | | | Cluster | | | | | | P-value |
|---|---|---|---|---|---|---|---|---|---|
| | | | 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 | |
| Health | Heart & Circulatory | Heart Attack (%) | 0.0 | 12.4 | 4.3 | 0.0 | 0.0 | 0.0 | <0.001 |
| | | Hypertension (%) | 33.2 | 82.8 | 84.7 | 12.0 | 23.1 | 15.3 | <0.001 |
| | | Stroke (%) | 0.0 | 0.4 | 8.5 | 4.0 | 0.3 | 0.9 | <0.001 |
| | Wellbeing | Mean Pain (days in month) | 1.2 | 2.0 | 11.4 | 1.4 | 0.9 | 2.8 | <0.001 |
| | | Mean Anxiety (days in month) | 4.9 | 3.2 | 9.6 | 4.8 | 4.5 | 6.1 | <0.001 |
| | | Mean Depression Score | 2.5 | 1.8 | 6.2 | 2.1 | 3.3 | 3.5 | <0.001 |
| | Other | Arthritis (%) | 17.0 | 34.2 | 75.8 | 18.6 | 12.8 | 3.9 | <0.001 |
| | | Cancer (%) | 11.1 | 28.6 | 11.8 | 5.1 | 2.1 | 0.1 | <0.001 |
| | | Sleep Disorder (%) | 28.5 | 8.5 | 27.8 | 4.3 | 3.9 | 6.2 | <0.001 |
| Behavioural | Physical Acttivity | Mean physical activity (MET) | 2140.3 | 2481.4 | 1466.2 | 3198.4 | 2426.0 | 6673.4 | <0.001 |
| | Diet | Mean Healthy Eating Score | 53.0 | 52.7 | 53.6 | 53.5 | 53.0 | 50.1 | 0.002 |
| | Smoking | Smoked > than 100 cigs (%) | 59.9 | 65.0 | 38.7 | 0.0 | 17.1 | 58.0 | <0.001 |
| | Alcohol | Non-drinkers (%) | 0.0 | 1.8 | 24.6 | 28.2 | 28.4 | 6.9 | <0.001 |
| | | <1 drink per day (%) | 24.0 | 32.6 | 48.2 | 33.3 | 25.9 | 9.7 | |
| | | 1-2 drinks per day (%) | 36.9 | 22.0 | 12.1 | 24.5 | 30.1 | 6.7 | |
| | | >2 drinks per day (%) | 39.1 | 43.5 | 15.1 | 14.0 | 15.6 | 76.7 | |

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

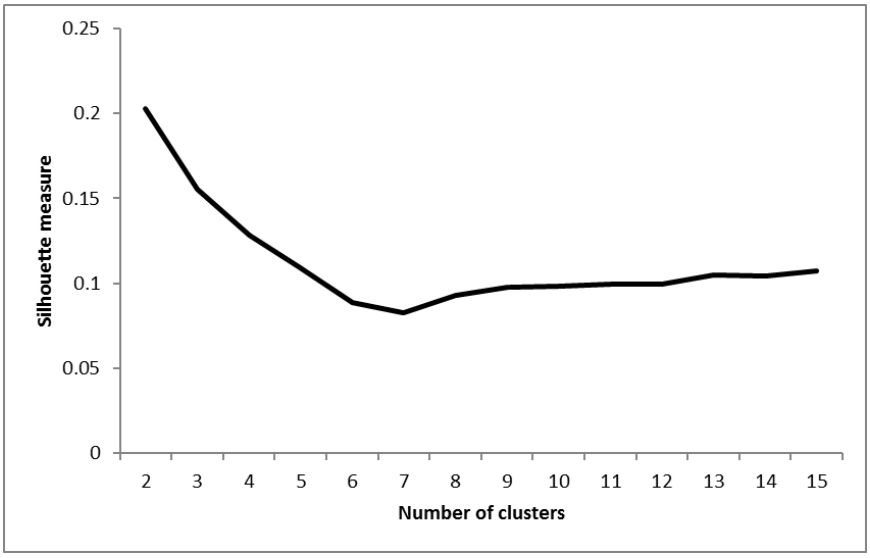



*Supplemental Figure S1. Missing Data Flow Chart*

We propagated our data set to represent the MEC sample weights and ensure that our analyses remained nationally representative, which altered our final sample size from1,380 to 560,928.

**Variables Included in Analysis**

Demographic characteristics included were age, sex, race and education level. Race was categorized as: White, Black, Mexican-American, other Latinos, and other. Annual household income was collapsed into roughly four evenly sized categories; $0-$20 000, $20 000-$45 000, $45 000-75 000 and $75 000+. Laboratory data was not included to define dyslipidemia, hypertension and diabetes since laboratory data in NHANES are only available among a sub-set of individuals which significantly reduces our analytic sample (to 49%). The depression score is based on the sum of the points in each item ranging from 0 to 27. Presence of a sleep disorder (yes vs no) was defined as an affirmative response to questions asking whether their doctor had ever told them they had sleep disorder [131, 132].




*Supplemental Figure 2. Silhouette Measure across a range of cluster solutions*

Figure S2 shows a higher value representing greater cohesion within clusters, as well as greater separation between the clusters. Whilst a two cluster solution scores the highest, this reflects a regression to the mean due to the small number of clusters [133]. There is a consistent decline in the silhouette measure with increasing clusters. A six cluster solution suggest a cessation in the decline, while a seven cluster solution presents an increase in the measure contrary to this trend.

*Supplemental Table S1. 5-Cluster Solution in the 2011-2012 NHANES*

| Variables | | | Clusters | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Affluent men with sleep disorder | Elderly smokers with CVD | Older women with poor mental health | Healthy women | Active men who drink higher amounts of alcohol | P-value |
| | | Sample Size (n) | 88581 | 92655 | 119581 | 134547 | 125564 | |
| | | Sample Size (%) | 15.8 | 16.5 | 21.3 | 24.0 | 22.4 | |
| | | Mean Body Mass Index | 34.3 | 34.7 | 37.8 | 35.9 | 35.2 | <0.001 |
| Personal | Age | Mean Age | 47.8 | 58.1 | 58.7 | 41.7 | 33.7 | <0.001 |
| | Sex | Male (%) | 75.1 | 68.6 | 22.6 | 11.9 | 74.8 | <0.001 |
| | | Female (%) | 24.9 | 31.4 | 77.4 | 88.1 | 25.2 | |
| | Race | Mexican American (%) | 1.2 | 2.8 | 6.8 | 7.9 | 21.8 | <0.001 |
| | | Other Hispanic (%) | 1.3 | 3.2 | 6.0 | 9.3 | 14.4 | |
| | | Non-Hispanic White (%) | 97.5 | 84.5 | 61.5 | 48.8 | 47.9 | |
| | | Non-Hispanic Black (%) | 0.1 | 8.8 | 21.3 | 28.6 | 10.8 | |
| | | Other Race (%) | 0.0 | 0.7 | 4.4 | 5.4 | 5.0 | |
| | Income | <$20,000 (%) | 0.0 | 7.3 | 31.7 | 17.0 | 44.2 | <0.001 |
| | | $20,000 - $45,000 (%) | 0.0 | 7.2 | 30.9 | 11.6 | 13.2 | |
| | | $45,000 - $75,000 (%) | 6.4 | 28.3 | 20.5 | 25.7 | 35.2 | |
| | | $75,000+ (%) | 93.6 | 57.2 | 16.9 | 45.6 | 7.4 | |
| Health | Respiratory | Asthma (%) | 15.4 | 34.0 | 75.6 | 15.1 | 4.0 | <0.001 |
| | | Emphysema (%) | 0.0 | 4.3 | 2.4 | 1.8 | 0.3 | 0.003 |
| | Heart & Circulatory | Angina (%) | 4.2 | 8.8 | 6.2 | 0.0 | 0.0 | <0.001 |
| | | Congestive Heart Failure (%) | 1.2 | 6.0 | 8.0 | 0.0 | 0.0 | <0.001 |
| | | Coronary Heart Disease (%) | 0.0 | 12.2 | 4.1 | 0.0 | 0.4 | <0.001 |
| | | Diabetes (%) | 2.0 | 36.3 | 26.1 | 3.1 | 4.1 | <0.001 |
| | | Dyslipdemia (%) | 25.5 | 94.7 | 61.9 | 10.8 | 7.7 | <0.001 |

*Supplemental Table S1. 5-Cluster Solution in the 2011-2012 NHANES (Continued)*

| Variables | | | Clusters | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Affluent men with sleep disorder | Elderly smokers with CVD | Older women with poor mental health | Healthy women | Active men who drink higher amounts of alcohol | P-value |
| Health | Heart & Circulatory | Heart Attack (%) | 0.0 | 12.2 | 4.1 | 0.0 | 0.0 | <0.001 |
| | | Hypertension (%) | 32.8 | 82.3 | 83.4 | 15.8 | 15.9 | <0.001 |
| | | Stroke (%) | 0.0 | 0.4 | 8.1 | 2.3 | 0.8 | <0.001 |
| | Wellbeing | Mean Pain (days in month) | 1.2 | 2.0 | 11.1 | 1.1 | 2.7 | <0.001 |
| | | Mean Anxiety (days in month) | 4.8 | 3.1 | 9.4 | 5.0 | 5.7 | <0.001 |
| | | Mean Depression Score | 2.4 | 1.8 | 6.0 | 2.8 | 3.4 | <0.001 |
| | Other | Arthritis (%) | 15.4 | 34.0 | 75.6 | 15.1 | 4.0 | <0.001 |
| | | Cancer (%) | 12.9 | 28.3 | 11.6 | 2.5 | 0.0 | <0.001 |
| | | Sleep Disorder (%) | 28.6 | 8.4 | 27.1 | 3.8 | 5.8 | <0.001 |
| Behavioural | Physical Acttivity | Mean physical activity (MET) | 2258.6 | 2475.1 | 1502.0 | 2451.1 | 6737.4 | 0.002 |
| | Diet | Mean Healthy Eating Score | 52.8 | 52.8 | 53.4 | 53.4 | 50.4 | <0.001 |
| | Smoking | Smoked more than 100 cigs (%) | 58.1 | 65.0 | 38.0 | 9.8 | 53.3 | <0.001 |
| | Alcohol | Non-drinkers (%) | 0.0 | 1.8 | 25.4 | 30.0 | 6.3 | <0.001 |
| | | <1 drink per day (%) | 25.2 | 32.7 | 47.2 | 30.9 | 8.8 | |
| | | 1-2 drinks per day (%) | 36.2 | 22.5 | 12.9 | 27.9 | 7.5 | |
| | | >2 drinks per day (%) | 38.6 | 43.0 | 14.4 | 11.2 | 77.3 | |

Cells in red represent values higher than the population mean, while cells in green represent values that are lower than the average within each row. Higher color intensity symbolizes values being further away from the mean (either up or down). Please not that the MET score and the healthy eating index were reverse color-coded to match the interpretation of the rest of the characteristics.

*Supplemental Table S2. Distribution Comparison Between the 5-Cluster Solution and the 6-Cluster Solution*

| | | | | 5 cluster solution | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | | | | 1 | 2 | 3 | 4 | 5 | |
| | | | | Affluent men with sleep disorder | Elderly smokers with CVD | Older women with poor mental health | Healthy women | Active men who drink higher amounts of alcohol | |
| **6 cluster solution** | 1 | Affluent men with sleep disorder | n | **85,311** | 0 | 0 | 2113 | 0 | 87,424 |
| | | | Row % | **97.6** | 0 | 0 | 2.4 | 0 | |
| | 2 | Elderly smokers with CVD | n | 0 | **91,611** | 0 | 0 | 0 | 91,611 |
| | | | Row % | 0 | **100** | 0 | 0 | 0 | |
| | 3 | Older women with poor mental health | n | 0 | 0 | **114,228** | 0 | 0 | 114,228 |
| | | | Row % | 0 | 0 | **100** | 0 | 0 | |
| | 4 | Healthy White women | n | 3,084 | 0 | 2,781 | **62,238** | 4,553 | 72,656 |
| | | | Row % | 4.2 | 0 | 3.8 | **85.7** | 6.3 | |
| | 5 | Healthy Non-White women | n | 186 | 1,044 | 2,572 | **64,495** | 9,608 | 77,905 |
| | | | Row % | 0.2 | 1.3 | 3.3 | **82.8** | 12.3 | |
| | 6 | Active men who drink higher amounts of alcohol | n | 0 | 0 | 0 | 5,701 | **111,403** | 117,104 |
| | | | Row % | 0 | 0 | 0 | 4.9 | **95.1** | |
| | | Total | | 88,581 | 92,655 | 119,581 | 134,547 | 125,564 | 560,928 |

Table S2 suggests that both 5 and 6 cluster solutions coincide in classifying more than 95% of the observations for clusters 1 (97.6%), 2 (100%), 3 (100%) and 6 (95.1%). Cluster 4 in the 5-cluster solution (Healthy women) is further divided in the 6-cluster solution into cluster 4 and cluster 5 providing more diversity according to race and income.

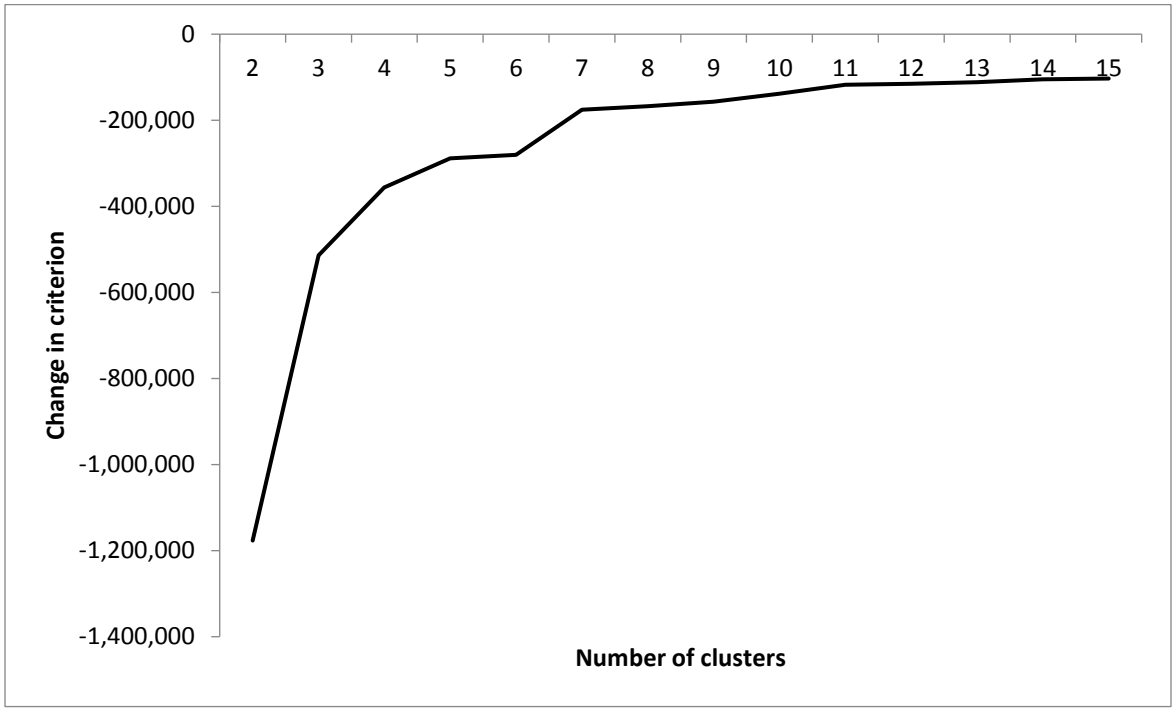
**Use of Sample Weights in Cluster Analysis**

Cluster analysis does not deal directly with sample weights, thus we propagated our data set to represent the sample weights. Propagating the data refers to weighting the data manually, as a cluster analysis procedure does not automatically take into account weights. It requires 'growing' out the data set by replicating individuals from the sample based on the sample weights. NHANES examination sample weights to ensure that our analyses remained nationally representative, as recommended by NHANES methodology[108]. The approach is essentially how sample weights are interpreted by statistical models. We truncated the weights to minimize the sample size so that the model would fit, whilst also being large enough to remain representative.

The propagation of the data orders the data by individuals (i.e. it would manually create *n* individuals corresponding to the weighting value of the first record, then do the same for the next individual, and so on). Therefore we randomized the data to scatter the observations randomly.

**Replication Analysis of the Existence of Clusters in the 2009-2010 Wave of NHANES**

To check for stability of clusters over time, we replicated the analysis using the previous wave of NHANES data (2009-2010) and compared results against the current wave of data. The 2009-2010 data contained all the same variables apart from the race variable where 'non-hispanic Asian' category was not captured seperately. Cluster weights were applied and the data were propagated similarly to the main solution (n = 605 793).




*Supplemental Figure S3. Change in Schwartz's Bayesian Information Criterion Across a Range of Cluster Solutions*

The Schwartz's Bayesian Information Criterion (BIC) for the same range of model solutions as in our main results were calculated to determine the number of clusters. Figure S3 presents the change in the BIC by each solution. There are clear leveling off in changes in BIC at cluster solutions five, six and sevensuggesting that either may be appropriate (although the seven does not result in a large improvement). Calculating the silhouette measure for each solution shows that a six-cluster performs the best on the measure. Since the six-cluster solution is the most parsimonious and other solutions do not appear to add more information it was selected as the most appropriate solution for the 2009-2010 data.

Table S3 presents the results of a five cluster solution. In comparison to the results presented in Table 2, we can see broad similarities in the clusters found in 2009-2010 compared to 2011-2012. We have given somewhat analogous names to the clusters in the main solution to reflect their similarities. The main difference between the clusters appears to be minor, and predominantly in the distribution of the gender and race variable, with the health and health-related behaviours remaining fairly consistent.

Female-dominant clusters: The cluster 'Women with poor mental health' have low levels of income and the lowest proportion of individuals who are non-Hispanic White. They have the lowest score for healthy eating, the highest mean BMI, and have high prevalence of some health conditions particularly anxiety and depression in comparison to the other clusters. 'Healthy White women' are the second eldest cluster and are primarily non-Hispanic White and of median income levels. They have high prevalence of a dyslipidemia, hypertension, arthritis and cancer. The women in this cluster have the lowest levels of physical activity but the highest healthy eating score. 'Young healthy women' are the largest cluster and are characterized by low prevalence of most health conditions, as well as higher proportions of Mexican Americans, non-Hispanic Blacks and other race.

Male-dominant clusters: The cluster 'Elderly smokers with CVD' are older than most other clusters, are primarily non-Hispanic White, have above average smoking, and the highest concentration of health and circulatory conditions. 'Active men who drink higher amounts of alcohol' are the second youngest cluster, they have low prevalence of most health conditions and have the highest score for physical activity, but have the highest level of alcohol consumption. 'Affluent men' has the highest income. The demographic profile of the cluster is mostly White and middle-aged. They have the lowest score for depression, and the highest healthy eating index. Slightly more than a third (37.7%) of the individuals in this cluster reported smoking. The individuals in this cluster had the second lowest score for physical activity and the lowest mean BMI.

*Supplemental Table S3. Mean cluster characteristics of a six cluster solution using the 2009-2010 NHANES*

| Variables | | | Cluster | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | | | Women with poor mental health | Elderly smokers with CVD | Young healthy women | Affluent men | Healthy White women | Active Men who drink higher amounts of alcohol | |
| | | Sample Size (n) | 84731 | 74786 | 132688 | 107710 | 104703 | 101175 | 605793 |
| | | Sample Size (%) | 14.0 | 12.3 | 21.9 | 17.8 | 17.3 | 16.7 | |
| | | Mean Body Mass Index | 38.0 | 35.5 | 36.5 | 33.7 | 36.0 | 35.3 | 35.8 |
| Personal | Age | Mean Age | 49.2 | 63.6 | 37.3 | 43.7 | 62.1 | 38.6 | 47.8 |
| | Sex | Male (%) | 25.7 | 84.5 | 0.0 | 100.0 | 0.2 | 99.4 | 48.4 |
| | | Female (%) | 74.3 | 15.5 | 100.0 | 0.0 | 99.8 | 0.6 | 51.6 |
| | Race | Mexican American (%) | 11.2 | 1.8 | 13.2 | 3.9 | 3.1 | 20.1 | 9.3 |
| | | Other Hispanic (%) | 10.9 | 1.3 | 5.9 | 3.6 | 1.0 | 4.0 | 4.5 |
| | | Non-Hispanic White (%) | 44.1 | 86.4 | 53.8 | 82.8 | 86.3 | 57.3 | 67.8 |
| | | Non-Hispanic Black (%) | 30.2 | 8.0 | 21.5 | 7.2 | 9.1 | 14.0 | 15.1 |
| | | Other Race (%) | 3.6 | 2.5 | 5.5 | 2.5 | 0.5 | 4.6 | 3.3 |
| | Income | <$20,000 (%) | 32.3 | 4.8 | 22.8 | 0.0 | 4.3 | 20.6 | 14.3 |
| | | $20,000 - $45,000 (%) | 31.0 | 12.3 | 22.0 | 0.0 | 12.5 | 33.0 | 18.3 |
| | | $45,000 - $75,000 (%) | 21.9 | 36.6 | 22.1 | 0.0 | 37.1 | 46.2 | 26.6 |
| | | $75,000+ (%) | 14.8 | 46.4 | 33.0 | 100.0 | 46.1 | 0.2 | 40.8 |
| Health | Respiratory | Asthma (%) | 62.6 | 56.7 | 4.2 | 17.9 | 55.8 | 8.5 | 30.9 |
| | | Emphysema (%) | 3.4 | 5.1 | 0.0 | 0.0 | 0.0 | 0.0 | 1.1 |
| | Heart & Circulatory | Angina (%) | 5.2 | 13.7 | 0.0 | 0.0 | 0.7 | 0.2 | 2.6 |
| | | Congestive Heart Failure (%) | 6.2 | 17.0 | 0.1 | 0.0 | 0.0 | 0.0 | 3.0 |
| | | Coronary Heart Disease (%) | 1.5 | 31.6 | 0.0 | 0.0 | 0.0 | 0.0 | 4.1 |
| | | Diabetes (%) | 27.0 | 47.0 | 1.3 | 1.4 | 16.4 | 2.1 | 13.3 |
| | | Dyslipdemia (%) | 55.8 | 71.2 | 9.6 | 20.9 | 49.6 | 14.3 | 33.4 |

*Supplemental Table S3. Mean cluster characteristics of a six cluster solution using the 2009-2010 NHANES (Continued)*

| Variables | | | Cluster | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | | | Women with poor mental health | Elderly smokers with CVD | Young healthy women | Affluent men | Healthy White women | Active Men who drink higher amounts of alcohol | |
| Health | Heart & Circulatory | Heart Attack (%) | 5.1 | 26.9 | 0.0 | 0.0 | 0.0 | 0.0 | 4.0 |
| | | Hypertension (%) | 70.7 | 81.4 | 9.0 | 29.2 | 68.5 | 24.3 | 43.0 |
| | | Stroke (%) | 8.2 | 9.9 | 0.0 | 0.5 | 0.7 | 0.3 | 2.6 |
| | Wellbeing | Mean Pain (days in month) | 12.4 | 6.6 | 2.2 | 3.0 | 3.9 | 2.5 | 4.7 |
| | | Mean Anxiety (days in month) | 17.9 | 4.3 | 8.2 | 5.6 | 3.6 | 4.8 | 7.3 |
| | | Mean Depression Score | 9.1 | 2.6 | 3.5 | 2.1 | 2.3 | 2.6 | 3.6 |
| | Other | Arthritis (%) | 62.6 | 56.7 | 4.2 | 17.9 | 55.8 | 8.5 | 30.9 |
| | | Cancer (%) | 8.8 | 24.5 | 1.8 | 4.8 | 23.5 | 2.8 | 10.0 |
| | | Sleep Disorder (%) | 29.6 | 27.3 | 3.0 | 8.5 | 13.0 | 3.7 | 12.5 |
| Behavioural | Physical Acttivity | Mean physical activity (MET) | 2850.4 | 1971.0 | 2019.5 | 4372.7 | 1081.0 | 6443.3 | 3124.8 |
| | Diet | Mean Healthy Eating Score | 49.6 | 52.8 | 52.7 | 51.0 | 57.4 | 50.5 | 52.4 |
| | Smoking | Smoked more than 100 cigs (%) | 57.0 | 67.0 | 36.0 | 37.7 | 26.9 | 35.0 | 41.3 |
| | Alcohol | Non-drinkers (%) | 13.1 | 5.6 | 14.5 | 2.4 | 23.1 | 8.1 | 11.5 |
| | | <1 drink per day (%) | 39.0 | 49.2 | 26.3 | 24.3 | 55.7 | 21.9 | 34.9 |
| | | 1-2 drinks per day (%) | 20.9 | 27.0 | 23.2 | 25.2 | 20.3 | 13.5 | 21.6 |
| | | >2 drinks per day (%) | 26.9 | 18.2 | 36.0 | 48.1 | 0.9 | 56.5 | 32.1 |

Cells in red represent values higher than the population mean, while cells in green represent values that are lower than the average within each row. Higher color intensity symbolizes values being further away from the mean (either up or down). Please not that the MET score and the healthy eating index were reverse color-coded to match the interpretation of the rest of the characteristics.

# Conclusions

This dissertation examined the heterogeneity of cardiometabolic health at the individual level and at the contextual level. Individuals classified as obese were found to be significantly different based on demographic characteristics, behavioral patterns and clinical conditions. The individual heterogeneity of obesity found provides a basis for future studies to evaluate whether different strategies may be needed in the clinical management of obesity. The life-course analysis of cardio-metabolic health suggested that exposures at the contextual level during early childhood were associated with a decrease of blood pressure in adulthood. In addition, the higher proportion of variability of blood pressure explained by the contextual level at the earliest years of life suggests intervention time periods that may be more effective to reduce cardiometabolic risk.

In Chapter 1, "Longitudinal associations of life-course NSES with BP and BMI", results of the critical period model suggested that higher NSES at birth was associated with a decrease of 2.68 mmHg of SBP and of 1.59 mmHg of DBP in adulthood. In addition, higher NSES at adulthood was associated with a decrease of 1.1 kg/$m^2$ of BMI in adulthood, controlling for confounders. However, results of the accumulation of risk model were not statistically significant for BP or BMI.

In Chapter 2, "Longitudinal associations between neighborhood access to green space in different periods in life, and blood pressure and body mass index", we found that living one mile farther from a green space at childhood was associated with an increase in growth rate of SBP of 3.5 mmHg. Additionally, having one more green space at birth was associated with a decrease in growth rate of SBP of -0.2 mmHg. Results of the association between access to green space and

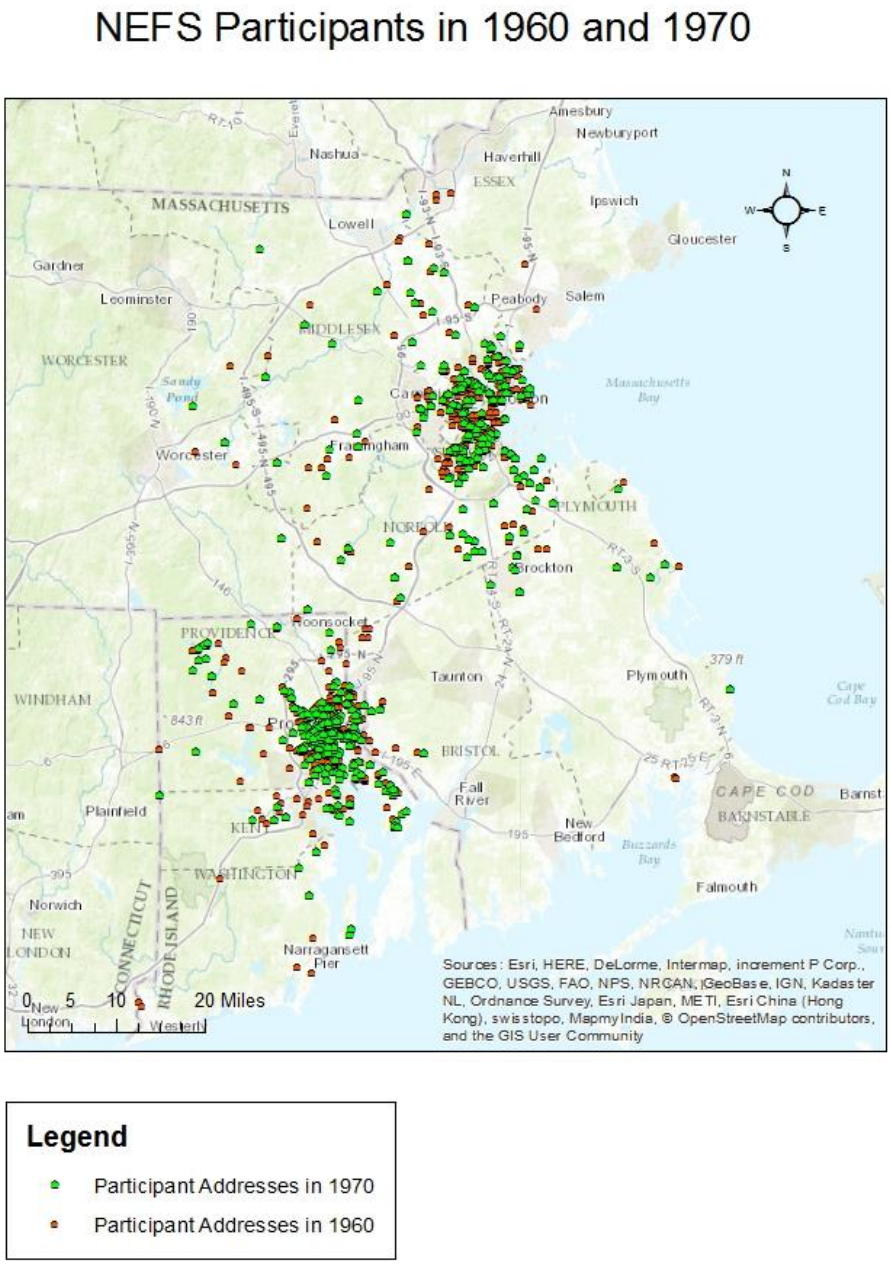
BMI across the life-course were not statistically significant. Results from the first 2 chapters of this dissertation provide evidence that during critical periods of susceptibility in life, greater access to neighborhood green space and higher NSES may slow the development of CVD in middle-aged adults.

In chapter 3 “A demographic, clinical, and behavioral typology of obesity in the United States: an analysis of NHANES 2011-2012”, we found 6 different clusters of individuals classified as obese, suggesting that obesity is not a homogeneous disease in the U.S. These findings were replicated in the prior wave indicating that the heterogeneity among obese clusters was constant across time. The clusters help to identify which behaviors need to be targeted simultaneously and in whom. Furthermore, results show that subgroups of individuals classified as obese may benefit from tailored obesity treatment strategies and guidelines for clinical practice.

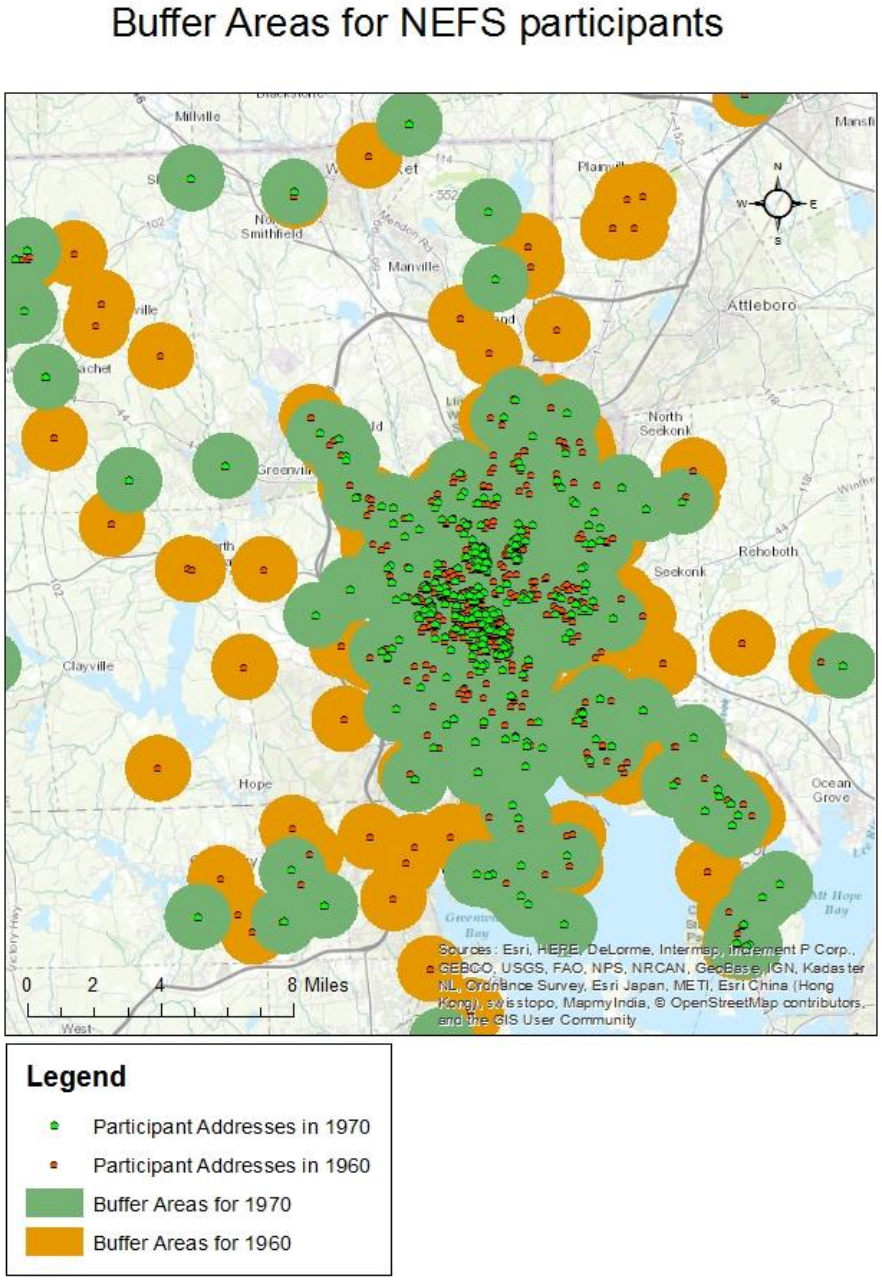
The understanding of the time of exposure at which neighborhood social and physical context impacts the development of cardio-metabolic disease is limited. However, our findings suggest that neighborhood social context at birth and neighborhood physical context at childhood may be crucial sensitive periods. Further investigation is needed to fully elucidate these findings. In particular, future studies should conduct a formal mediation analysis to estimate the underlying mechanisms of neighborhood on cardio-metabolic health. Additionally, future studies should collect information on frequency of use of neighborhood resources. Information on amount of time spent at home, as well as work geographic locations at adulthood or school locations at childhood are also important to correctly classify participants’ full exposure. In conclusion, this work suggests that to understand cardiometabolic risk, it is necessary to examine heterogeneity

at 2 levels: individual and contextual. Distinctive groups of individuals diagnosed with obesity and deprived neighborhoods might be at higher cardiometabolic risk and should be prioritized for public health assistance. Neighborhood factors are a ubiquitous exposure and improving access at sensitive periods in the life-course would greatly reduce cardio-metabolic risk.

Supplemental Material




**Supplemental Figure 1.1**: Participants Addresses




**Supplemental Figure 2.1**: Buffer Areas

# References


1. Mozaffarian, D., et al., *Heart disease and stroke statistics--2015 update: a report from the American Heart Association.* Circulation, 2015. **131**(4): p. e29-322.
2. *Adult Obesity Facts*. Data and Statistics [cited 2015 Nov. 2015 ]; Available from: http://www.cdc.gov/obesity/data/adult.html.
3. *High Blood Pressure Facts* Statistics and Maps [cited 2015 Nov., 2015 ]; Available from: http://www.cdc.gov/bloodpressure/facts.htm.
4. Nwankwo, T., et al., *Hypertension among adults in the United States: National Health and Nutrition Examination Survey, 2011-2012.* NCHS Data Brief, 2013(133): p. 1-8.
5. Chaix, B., *Geographic life environments and coronary heart disease: a literature review, theoretical contributions, methodological updates, and a research agenda.* Annu Rev Public Health, 2009. **30**: p. 81-105.
6. Diez Roux, A.V. and C. Mair, *Neighborhoods and health.* Ann N Y Acad Sci, 2010. **1186**: p. 125-45.
7. Halonen, J.I., et al., *Childhood Psychosocial Adversity and Adult Neighborhood Disadvantage as Predictors of Cardiovascular Disease: A Cohort Study.* Circulation, 2015.
8. James, P., et al., *A Review of the Health Benefits of Greenness.* Curr Epidemiol Rep, 2015. **2**(2): p. 131-142.
9. Ben-Shlomo, Y. and D. Kuh, *A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives.* Int J Epidemiol, 2002. **31**(2): p. 285-93.
10. Lynch, J. and G.D. Smith, *A life course approach to chronic disease epidemiology.* Annu Rev Public Health, 2005. **26**: p. 1-35.
11. Su, S., et al., *The Role of Adverse Childhood Experiences in Cardiovascular Disease Risk: a Review with Emphasis on Plausible Mechanisms.* Curr Cardiol Rep, 2015. **17**(10): p. 645.
12. Diez-Roux, A.V., et al., *Neighborhood environments and coronary heart disease: a multilevel analysis.* Am J Epidemiol, 1997. **146**(1): p. 48-63.
13. Diez Roux, A.V., et al., *Neighborhood of residence and incidence of coronary heart disease.* N Engl J Med, 2001. **345**(2): p. 99-106.
14. Borrell, L.N., et al., *Neighbourhood characteristics and mortality in the Atherosclerosis Risk in Communities Study.* Int J Epidemiol, 2004. **33**(2): p. 398-407.
15. Lee, H., K.M. Harris, and J. Lee, *Multiple levels of social disadvantage and links to obesity in adolescence and young adulthood.* J Sch Health, 2013. **83**(3): p. 139-49.
16. Lopez, R.P. and H.P. Hynes, *Obesity, physical activity, and the urban environment: public health research needs.* Environ Health, 2006. **5**: p. 25.
17. Fisher, K.J., et al., *Neighborhood-level influences on physical activity among older adults: a multilevel analysis.* J Aging Phys Act, 2004. **12**(1): p. 45-63.
18. Viruell-Fuentes, E.A., N.A. Ponce, and M. Alegria, *Neighborhood context and hypertension outcomes among Latinos in Chicago.* J Immigr Minor Health, 2012. **14**(6): p. 959-67.
19. Bodicoat, D.H., et al., *The association between neighbourhood greenspace and type 2 diabetes in a large cross-sectional study.* BMJ Open, 2014. **4**(12): p. e006076.
20. Mujahid, M.S., et al., *Neighborhoods and racial/ethnic differences in ideal cardiovascular health (the Multi-Ethnic Study of Atherosclerosis).* Health Place, 2017. **44**: p. 61-69.
21. Longo, A., et al., *Demand response to improved walking infrastructure: A study into the economics of walking and health behaviour change.* Soc Sci Med, 2015. **143**: p. 107-116.

22. Berenson, G.S., et al., *Association between multiple cardiovascular risk factors and atherosclerosis in children and young adults. The Bogalusa Heart Study.* N Engl J Med, 1998. **338**(23): p. 1650-6.
23. Arcaya, M.C., et al., *Role of health in predicting moves to poor neighborhoods among Hurricane Katrina survivors.* Proc Natl Acad Sci U S A, 2014. **111**(46): p. 16246-53.
24. Kuh, D., et al., *Life course epidemiology.* J Epidemiol Community Health, 2003. **57**(10): p. 778-83.
25. Ogden, C.L., et al., *Prevalence of childhood and adult obesity in the United States, 2011-2012.* JAMA, 2014. **311**(8): p. 806-14.
26. Ogden, C.L., et al., *Prevalence of Obesity Among Adults and Youth: United States, 2011-2014.* NCHS Data Brief, 2015(219): p. 1-8.
27. Schuurman, N., P.A. Peters, and L.N. Oliver, *Are obesity and physical activity clustered? A spatial analysis linked to residential density.* Obesity (Silver Spring), 2009. **17**(12): p. 2202-9.
28. Green, M.A., et al., *Who are the obese? A cluster analysis exploring subgroups of the obese.* J Public Health (Oxf), 2015. **38**(2): p. 258-64.
29. Stefan, N., et al., *Metabolically healthy obesity: epidemiology, mechanisms, and clinical implications.* Lancet Diabetes Endocrinol, 2013. **1**(2): p. 152-62.
30. Bellis, M.A., et al., *The alcohol harm paradox: using a national survey to explore how alcohol may disproportionately impact health in deprived individuals.* BMC Public Health, 2016. **16**: p. 111.
31. Michael, Y., et al., *Measuring the influence of built neighborhood environments on walking in older adults.* J Aging Phys Act, 2006. **14**(3): p. 302-12.
32. Frank, L.D., M.A. Andresen, and T.L. Schmid, *Obesity relationships with community design, physical activity, and time spent in cars.* Am J Prev Med, 2004. **27**(2): p. 87-96.
33. Richardson, A.S., et al., *One size doesn't fit all: cross-sectional associations between neighborhood walkability, crime and physical activity depends on age and sex of residents.* BMC Public Health, 2017. **17**(1): p. 97.
34. Min, Y.I., et al., *Cardiovascular Disease Burden and Socioeconomic Correlates: Findings From the Jackson Heart Study.* J Am Heart Assoc, 2017. **6**(8).
35. Psaltopoulou, T., et al., *Socioeconomic status and risk factors for cardiovascular disease: Impact of dietary mediators.* Hellenic J Cardiol, 2017. **58**(1): p. 32-42.
36. Wilson, W.J., *The Truly Disadvantaged. The inner city, the underclass and public policy.* . 1987, Chicago: The University of Chicago Press.
37. Sampson, R.J., S.W. Raudenbush, and F. Earls, *Neighborhoods and violent crime: a multilevel study of collective efficacy.* Science, 1997. **277**(5328): p. 918-24.
38. Hussein, M., et al., *Unequal Exposure or Unequal Vulnerability? Contributions of Neighborhood Conditions and Cardiovascular Risk Factors to Socioeconomic Inequality in Incident Cardiovascular Disease in the Multi-Ethnic Study of Atherosclerosis.* Am J Epidemiol, 2017.
39. Akwo, E.A., et al., *Neighborhood Deprivation Predicts Heart Failure Risk in a Low-Income Population of Blacks and Whites in the Southeastern United States.* Circ Cardiovasc Qual Outcomes, 2018. **11**(1): p. e004052.
40. Oudin Astrom, D., J. Sundquist, and K. Sundquist, *Differences in declining mortality rates due to coronary heart disease by neighbourhood deprivation.* J Epidemiol Community Health, 2018.
41. Marden, J.R., et al., *Contribution of Socioeconomic Status at 3 Life-Course Periods to Late-Life Memory Function and Decline: Early and Late Predictors of Dementia Risk.* Am J Epidemiol, 2017. **186**(7): p. 805-814.
42. Slopen, N., et al., *Early origins of inflammation: An examination of prenatal and childhood social adversity in a prospective cohort study.* Psychoneuroendocrinology, 2015. **51**: p. 403-13.

43. *Earth Lab Basic GIS How to*. [cited 2017 July]; Available from: http://www.brown.edu/Research/Earthlab/basicgishowto.html.
44. [cited 2017 July]; Available from: https://developers.arcgis.com/rest/geocode/api-reference/overview-world-geocoding-service.htm.
45. *National Historical Geographic Information System (NGHIS): Version 11.0 [Database]*. 2016, Minnesota Population Center: Minneapolis: University of Minnesota.
46. Logan, J. *MapUSA: A Human-Mapping Project (1940-2000)*. [cited 2017.
47. Dunn, E.C., et al., *Using cross-classified multilevel models to disentangle school and neighborhood effects: an example focusing on smoking behaviors among adolescents in the United States.* Health Place, 2015. **31**: p. 224-32.
48. Halonen, J.I., et al., *Quantifying neighbourhood socioeconomic effects in clustering of behaviour-related risk factors: a multilevel analysis.* PLoS One, 2012. **7**(3): p. e32937.
49. *Inflation Calculator* [cited 2017 July]; Available from: https://www.dollartimes.com/inflation/inflation.php?amount=1&year=1960.
50. Block, J.P., et al., *Proximity to food establishments and body mass index in the Framingham Heart Study offspring cohort over 30 years.* Am J Epidemiol, 2011. **174**(10): p. 1108-14.
51. Ibrahim, J.G. and G. Molenberghs, *Missing data methods in longitudinal studies: a review.* Test (Madr), 2009. **18**(1): p. 1-43.
52. Barker, D.J., *The developmental origins of adult disease.* J Am Coll Nutr, 2004. **23**(6 Suppl): p. 588S-595S.
53. Bateson P., M., P., *Design for a Life: How Behaviour Develops*. 1999, London: Jonathan Cape
54. Barker, D.J., et al., *Growth and living conditions in childhood and hypertension in adult life: a longitudinal study.* J Hypertens, 2002. **20**(10): p. 1951-6.
55. *Mayor de Blasio, Council Speaker Mark-Viverito, Parks Commissioner Silver Announce $150 Million in Funding for Five Anchor Parks*. 2016: The Official Website of the City of New York.
56. *Healthy Neighborhoods Initiative*. 2016: Official Website of the City of Milwaukee.
57. Shanahan, D.F., et al., *Health Benefits from Nature Experiences Depend on Dose.* Sci Rep, 2016. **6**: p. 28551.
58. de Blasio, B., *Healthier neighbourhoods through healthier parks.* Lancet, 2016. **388**(10062): p. 2850-2851.
59. Kardan, O., et al., *Neighborhood greenspace and health in a large urban center.* Sci Rep, 2015. **5**: p. 11610.
60. Gascon, M., et al., *Residential green spaces and mortality: A systematic review.* Environ Int, 2016. **86**: p. 60-7.
61. *Urban Health*. 2018 03/07/2018]; Available from: http://www.who.int/topics/urban_health/en/.
62. Beltran-Heredia, J., J. Sanchez-Martin, and M.A. Davila-Acedo, *Optimization of the synthesis of a new coagulant from a tannin extract.* J Hazard Mater, 2011. **186**(2-3): p. 1704-12.
63. Mokdad, A.H., et al., *Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001.* JAMA, 2003. **289**(1): p. 76-9.
64. Grazuleviciene, R., et al., *Tracking Restoration of Park and Urban Street Settings in Coronary Artery Disease Patients.* Int J Environ Res Public Health, 2016. **13**(6).
65. Rundle, A., et al., *Associations between body mass index and park proximity, size, cleanliness, and recreational facilities.* Am J Health Promot, 2013. **27**(4): p. 262-9.
66. Browning, M. and K. Lee, *Within What Distance Does "Greenness" Best Predict Physical Health? A Systematic Review of Articles with GIS Buffer Analyses across the Lifespan.* Int J Environ Res Public Health, 2017. **14**(7).

67. Astell-Burt, T., X. Feng, and G.S. Kolt, *Greener neighborhoods, slimmer people? Evidence from 246,920 Australians.* Int J Obes (Lond), 2014. **38**(1): p. 156-9.
68. Sanders, T., et al., *Green Space and Child Weight Status: Does Outcome Measurement Matter? Evidence from an Australian Longitudinal Study.* J Obes, 2015. **2015**: p. 194838.
69. Sanders, T., et al., *Greener neighbourhoods, slimmer children? Evidence from 4423 participants aged 6 to 13 years in the Longitudinal Study of Australian children.* Int J Obes (Lond), 2015. **39**(8): p. 1224-9.
70. Nieuwenhuijsen, M.J., et al., *Fifty Shades of Green: Pathway to Healthy Urban Living.* Epidemiology, 2017. **28**(1): p. 63-71.
71. Zhang, L., Yok Tan, P., Diehl, J.A. , *A conceptual framework for studying urban green spaces effects on health.* Journal of Urban Ecology, 2017. **3**(1): p. 1-13.
72. Jimenez, M.P., Subramanian, S.V., Wellenius, G., and Loucks, E.B., *Longitudinal association of life-course neighborhood deprivation with blood pressure and obesity* In progress.
73. *What is Open Space/Green Space?* 2017 03/07/2018]; Available from: http://www3.epa.gov/region1/eco/uep/openspace.html.
74. Tamosiunas, A., et al., *Accessibility and use of urban green spaces, and cardiovascular health: findings from a Kaunas cohort study.* Environ Health, 2014. **13**(1): p. 20.
75. Kaczynski, A.T., et al., *Are park proximity and park features related to park use and park-based physical activity among adults? Variations by multiple socio-demographic characteristics.* Int J Behav Nutr Phys Act, 2014. **11**(1): p. 146.
76. Leslie, E., et al., *Walkability of local communities: using geographic information systems to objectively assess relevant environmental attributes.* Health Place, 2007. **13**(1): p. 111-22.
77. *Historical Topographic Map Collection*. June 2017 ]; Available from: https://ngmdb.usgs.gov/topoview/.
78. *MassGIS Data: Protected and Recreational OpenSpace*. June 2017 ]; Available from: https://docs.digital.mass.gov/dataset/massgis-data-protected-and-recreational-openspace.
79. *Rhode Island Conservation Lands Public Access*. June 2017 ]; Available from: https://www.arcgis.com/home/item.html?id=e31d50b387e04b3a9c025856c015923c.
80. Gascon, M., et al., *Outdoor blue spaces, human health and well-being: A systematic review of quantitative studies.* Int J Hyg Environ Health, 2017. **220**(8): p. 1207-1221.
81. *HERE Map Data. Harnessing the power of location data*. June 2017 ]; Available from: https://www.here.com/en/products-services/map-content/here-map-data.
82. Yang, S., et al., *Pre-natal and post-natal growth trajectories and childhood cognitive ability and mental health.* Int J Epidemiol, 2011. **40**(5): p. 1215-26.
83. Curran, P.J., K. Obeidat, and D. Losardo, *Twelve Frequently Asked Questions About Growth Curve Modeling.* J Cogn Dev, 2010. **11**(2): p. 121-136.
84. Van Buuren, S., Groothuis-Oudshoorn, K. , *MICE: Multivariate Imputation by Chained Equations in R* Journal of Statistical Software, 2011. **45**(3).
85. Roe, J.J., et al., *Green space and stress: evidence from cortisol measures in deprived urban communities.* Int J Environ Res Public Health, 2013. **10**(9): p. 4086-103.
86. Lascaux-Lefebvre, V., et al., *Influence of parental history of hypertension on blood pressure.* J Hum Hypertens, 1999. **13**(9): p. 631-6.
87. Heckman, C.J., *Public Parks and Shady Areas in Times of Climate Change, Urban Sprawl, and Obesity.* American Journal of Public Health, 2017. **107**(12): p. 1856-1858.
88. Stoner, L. and J. Cornwall, *Did the American Medical Association make the correct decision classifying obesity as a disease?* Australas Med J, 2014. **7**(11): p. 462-4.

89. Ogden, L.G., et al., *Cluster analysis of the national weight control registry to identify distinct subgroups maintaining successful weight loss.* Obesity (Silver Spring), 2012. **20**(10): p. 2039-47.
90. Leech, R.M., S.A. McNaughton, and A. Timperio, *The clustering of diet, physical activity and sedentary behavior in children and adolescents: a review.* Int J Behav Nutr Phys Act, 2014. **11**: p. 4.
91. Ferrar, K. and R. Golley, *Adolescent diet and time use clusters and associations with overweight and obesity and socioeconomic position.* Health Educ Behav, 2015. **42**(3): p. 361-9.
92. Pronk, N.P., et al., *Meeting recommendations for multiple healthy lifestyle factors. Prevalence, clustering, and predictors among adolescent, adult, and senior health plan members.* Am J Prev Med, 2004. **27**(2 Suppl): p. 25-33.
93. Schuit, A.J., et al., *Clustering of lifestyle risk factors in a general adult population.* Prev Med, 2002. **35**(3): p. 219-24.
94. *National Health and Nutrition Examination Survey (NHANES)*, N.C.f.H.S. Centers for Disease Control and Prevention, Editor. 2011-2012: http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm.
95. *NHANES 2011–2012 Overview*. 2011 May, 2015 ]; Available from: http://www.cdc.gov/nchs/nhanes/nhanes2011-2012/overview_g.htm.
96. *NHANES 2011 - 2012 Data Documentation, Codebook, and Frequencies - Body Measures*. 2015 September 2015; Available from: http://wwwn.cdc.gov/Nchs/Nhanes/2011-2012/BMX_G.htm#Component_Description.
97. Jimenez, M.P., Green, M.A., Subramanian, S.V., and Razak, F. , *Clinical Trial Registration for Obesity Heterogeneity Study*. 2017: ClinicalTrials.gov.
98. Sarathy, H., et al., *Abdominal Obesity, Race and Chronic Kidney Disease in Young Adults: Results from NHANES 1999-2010.* PLoS One, 2016. **11**(5): p. e0153588.
99. Wu, B., 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): p. e000154.
100. Spitzer, R.L., K. Kroenke, and J.B. Williams, *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): p. 1737-44.
101. Okifuji, A., et al., *Relationship between fibromyalgia and obesity in pain, function, mood, and sleep.* J Pain, 2010. **11**(12): p. 1329-37.
102. Surveillance and Population-Based Prevention, P.o.N.D.D., *Global Physical Activity Questionnaire (GPAQ) Analysis Guide*. World Health Organization.
103. Chiuve, S.E., et al., *Alternative dietary indices both strongly predict risk of chronic disease.* J Nutr, 2012. **142**(6): p. 1009-18.
104. Wildman, R.P., 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).* Arch Intern Med, 2008. **168**(15): p. 1617-24.
105. Kim, K.N., et al., *Urinary phthalate metabolites and depression in an elderly population: National Health and Nutrition Examination Survey 2005-2012.* Environ Res, 2016. **145**: p. 61-7.
106. Everitt, B.L., S.; Leese, M.; Stahl, D. , *Cluster Analysis*. Wiley Series in Probability and Statistics. 2001: Wiley.
107. Chiu, T., et al., *A robust and scalable clustering algorithm for mixed type attributes in large database environment*, in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. 2001, ACM: San Francisco, California. p. 263-268.

108. NHANES, N. *NHANES Weighting Module* Continuous NHANES Web Tutorial 2013 May 10, 2013 [cited 2016 March 11, 2016]; Available from: http://www.cdc.gov/nchs/tutorials/NHANES/SurveyDesign/Weighting/intro.htm.
109. Kirk, S.F., et al., *Effective weight management practice: a review of the lifestyle intervention evidence.* Int J Obes (Lond), 2012. **36**(2): p. 178-85.
110. Gomez-Ambrosi, J., et al., *Body mass index classification misses subjects with increased cardiometabolic risk factors related to elevated adiposity.* Int J Obes (Lond), 2012. **36**(2): p. 286-94.
111. Bastien, M., et al., *Overview of epidemiology and contribution of obesity to cardiovascular disease.* Prog Cardiovasc Dis, 2014. **56**(4): p. 369-81.
112. Flegal, K.M., et al., *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): p. 71-82.
113. Stahre, M., et al., *Contribution of excessive alcohol consumption to deaths and years of potential life lost in the United States.* Prev Chronic Dis, 2014. **11**: p. E109.
114. Jha, P., et al., *21st-century hazards of smoking and benefits of cessation in the United States.* N Engl J Med, 2013. **368**(4): p. 341-50.
115. Oreopoulos, A., et al., *The obesity paradox in the elderly: potential mechanisms and clinical implications.* Clin Geriatr Med, 2009. **25**(4): p. 643-59, viii.
116. Sacks, F.M., et al., *Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates.* N Engl J Med, 2009. **360**(9): p. 859-73.
117. Jensen, M.D., 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): p. 2985-3023.
118. *The Practical Guide. Identification, Evaluation, and Treatment of Overweight and Obesity in Adults*, M.D. F.Xavier Pi-Sunyer, M.P.H. , Editor. 2000, National Heart, Lung, and Blood Institute: NHLBI Obesity Education Initiative
119. *AACE 2015 guidelines Obesity Management*. 2015; Available from: http://www.ndei.org/AACE-guidelines-diabetes-obesity-treatment-algorithm.aspx.
120. Calle, E.E., et al., *Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults.* N Engl J Med, 2003. **348**(17): p. 1625-38.
121. Colditz, G.A., et al., *Weight gain as a risk factor for clinical diabetes mellitus in women.* Ann Intern Med, 1995. **122**(7): p. 481-6.
122. Eckel, R.H. and R.M. Krauss, *American Heart Association call to action: obesity as a major risk factor for coronary heart disease. AHA Nutrition Committee.* Circulation, 1998. **97**(21): p. 2099-100.
123. Kenchaiah, S., et al., *Obesity and the risk of heart failure.* N Engl J Med, 2002. **347**(5): p. 305-13.
124. Onat, A., 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): p. 46-51.
125. Hargens, T.A., et al., *Association between sleep disorders, obesity, and exercise: a review.* Nat Sci Sleep, 2013. **5**: p. 27-35.
126. Johnston, E., et al., *The relation of body mass index to depressive symptoms.* Can J Public Health, 2004. **95**(3): p. 179-83.

127. Patra, J., 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): p. e000121.
128. Towfighi, A., L. Zheng, and B. Ovbiagele, *Weight of the obesity epidemic: rising stroke rates among middle-aged women in the United States.* Stroke, 2010. **41**(7): p. 1371-5.
129. Swinburn, B. and G. Egger, *The runaway weight gain train: too many accelerators, not enough brakes.* BMJ, 2004. **329**(7468): p. 736-9.
130. Walthouwer, M.J., et al., *Are clusters of dietary patterns and cluster membership stable over time? Results of a longitudinal cluster analysis study.* Appetite, 2014. **82**: p. 154-9.
131. Beydoun, H.A., et al., *Bisphenol-A and Sleep Adequacy among Adults in the National Health and Nutrition Examination Surveys.* Sleep, 2016. **39**(2): p. 467-76.
132. Cunningham, T.D. and B.S. Di Pace, *Is Self-Reported Sleep Duration Associated with Osteoporosis? Data from a 4-Year Aggregated Analysis from the National Health and Nutrition Examination Survey.* J Am Geriatr Soc, 2015. **63**(7): p. 1401-6.
133. Christensen, A.C.R.a.W.F., *Methods of Multivariate Analysis, Third Edition*. Wiley Series in Probability and Statistics, ed. Wiley. 2012.