Life's Simple 7 in Two U.S. Populations Facing Cardiovascular Disease- and Cancer-Related Health Disparities

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Rory Cusack Weier, M.P.H.

Graduate Program in Public Health

The Ohio State University

2015

Dissertation Committee:

Electra D. Paskett, Ph.D., Advisor

Randi E. Foraker, Ph.D.

Michelle L. Kaiser, Ph.D.

Abigail B. Shoben, Ph.D.
Copyright by
Rory Cusack Weier
2015
Abstract

Background

Despite significant gains in cardiovascular disease (CVD) prevention and control over the past four decades, CVD remains the leading cause of mortality in the United States (U.S.).\(^1\) Cancer is a significant cause of morbidity in the U.S. and is poised to overtake CVD as the nation’s leading cause of mortality within two decades.\(^2\) These diseases are important public health issues for the general U.S. population as well as among subgroups in the U.S. that experience differences in the incidence, prevalence, mortality and overall burden of CVD and cancer that should not exist, also known as health disparities.\(^3\)

Smoking, overweight and obesity, inadequate physical activity, poor diet, high blood pressure, high total cholesterol and high plasma glucose are modifiable risk behaviors and factors that contribute to CVD and cancer outcomes through similar inflammatory pathways. The status of these overlapping risk behaviors and factors among populations with disparate CVD and cancer health outcomes need to be elucidated, as does the impact of the social determinants of health on CVD and cancer outcomes in these populations.\(^4,\)\(^5\) This insight is needed to inform intervention efforts to address health disparities and make progress toward CVD and cancer public health benchmarks among populations disproportionately burdened by these diseases.
The conceptual basis for this study was the Centers for Population Health and Health Disparities (CPHHD) model for the analysis of population health and health disparities (Figure A.4). This model allows social health determinants to influence CVD and cancer disparities independently and directly or through interactions across three levels: distal (i.e., population), intermediate (i.e., area) and proximal (i.e., individual). Distal health determinants include social conditions and policies like poverty, culture, discrimination, and the healthcare system. Intermediate health determinants include the social and physical contexts and social relationships, such as neighborhood poverty level, racial/ethnic integration, and access to grocery stores and supermarkets. Proximal health determinants include individual demographics, individual risk behaviors, biological responses and biologic/genetic pathways, such as age, household income, marital status, acculturation, and diet.

The metric that served as the outcome of interest for this study was the American Heart Association’s (AHA’s) Life’s Simple 7 (LS7) of Cardiovascular Health (CVH). LS7 includes well-defined poor, intermediate and ideal levels of four health behaviors (smoking, body mass index (BMI), physical activity and diet) and three health factors (blood pressure, total cholesterol and plasma glucose), respectively represented in the conceptual model as proximal-level risk behaviors and biological responses that contribute the CVD and cancer health disparities. These risk behaviors and factors are associated with CVD and cancer outcomes.

Using the LS7 metric, this study assessed the distribution of overlapping CVD and cancer risk behaviors and factors in two relatively understudied populations facing
health disparities, Puerto Ricans living in the contiguous U.S. (herein referred to as mainland Puerto Ricans) and residents of Appalachia. Furthermore, findings were placed in a national context by making comparisons to one another and to a research sample representative of civilian, non-institutionalized non-Hispanic white (NHW) men and women of similar age living in the U.S. In addition to risk behaviors and biological factors, this approach took into account place, culture, age and sex, which are all social determinants of health also contribute to disparate CVD and cancer health outcomes.

Since diet is the leading cause of the overall disease burden in the U.S. and the least prevalent ideal LS7 component among American adults, this study independently assessed multi-level social determinants of health that have the potential to contribute to poor diet among mainland Puerto Ricans and residents of Appalachia.\(^9\)\(^{-}\)\(^{14}\) Given the breadth of the diet literature, the multi-level social determinants explored were limited to facilitators and barriers pertaining to distance to food retailers that influence food purchasing behaviors and consumption that were particularly relevant to mainland Puerto Ricans and residents of Appalachia.\(^5\)\(^,\)\(^{15}\) Geographic food access, a characteristic of the physical context in which an individual lives, has received increased empirical attention due to burgeoning use of geographic information systems (GIS) and availability of incentives to improve access to healthy food in underserved areas. However, the literature lacks consistency regarding the relationship between GIS-based geographic food access measures and diet including intake of fruits and vegetables and overall diet quality.\(^{16}\)\(^,\)\(^{17}\) Furthermore, it is unclear from the current literature whether acculturation a form of informational food access should be taken into consideration when studying the
relationship between geographic food access and diet among ethnic minorities. Acculturation is known to have a protective effect on dietary components such as fruit and vegetable consumption and fat intake among ethnic minorities. Therefore, acculturation is also a social determinant of health that has the potential to influence to CVD and cancer health disparities through its impact on diet.

Similarly, although county- and individual-level geographic food access have the potential to independently influence diet, and an individual’s geographic food access is impacted by where they live, county-level geographic food access measures based on traditional food retailers continue to be used as proxies of individual-level geographic food access in rural and urban/rural heterogeneous areas. Studying the relationship between geographic food access and diet in these areas using county-level proxy measures based on distance to traditional food retailers like grocery stores and supermarkets is problematic given the closing of grocery stores and supermarkets in rural and urban/rural heterogeneous areas due to population loss, commuting patterns of residents to areas outside of local towns and counties and increased connectivity between rural and urban areas due to increased car ownership and roadways/highways.\(^\text{262, 263}\) It is unclear from the literature if an improved understanding of geographic food access and diet in rural and urban/rural heterogeneous areas can be achieved by focusing on the relationship between individual-level geographic food access and diet, after accounting for county-level geographic food access. Further, causal models for these individual- and county-level relationships with food access have not previously been specified.
The three aims of this study were to:

1. Quantify the distribution of each LS7 component and an overall LS7 score within research samples of mainland Puerto Ricans and residents of Appalachia and compare the distributions to one another and national data;

2. Quantify the association between a census tract (CT)-level food access metric and individual-level LS7 healthy diet score in a research sample of mainland Puerto Ricans, controlling for individual- and CT-level sociodemographic characteristics and assessing for effect measure modification (EMM) by individual- and CT-level acculturation; and

3. Compare nested models adjusted for individual- and area-level sociodemographic characteristics to determine whether using a county-level food access metric or individual-level home-to-preferred food store distance as the exposure of interest is more strongly associated with individual-level LS7 healthy diet score in a research sample of Appalachian residents.

Methods

Aim 1

Using 2004-2009 baseline data from the Boston Puerto Rican Health Study (BPRHS), 2012-2013 baseline data from the Appalachia Community Cancer Network (ACCN) study and 2007-2010 National Health and Nutrition Examination Survey (NHANES) data, this study quantified the distribution of each LS7 component and an overall LS7 score within each research sample. Different data restrictions were applied to
foster comparisons between BPRHS and ACCN study participants, BPRHS and NHW NHANES participants and ACCN study and NHW NHANES participants. Differences in age- and sex-adjusted prevalence estimates of poor, intermediate and ideal levels of each LS7 component (or some other combination of levels given data available) between BPRHS and ACCN study participants were assessed using Wald test for proportions and two sample t tests at the 0.003 level for an overall Type 1 error rate of 0.05. Differences in age- and sex-adjusted prevalence estimates of poor, intermediate and ideal levels of each LS7 component (or some other combination of levels given data available) between BPRHS and NHANES participants and between ACCN study and NHANES participants were assessed by comparing 99% confidence intervals (CIs) to minimize overall Type 1 error.

Aim 2

Using 2004-2009 baseline data from the BPRHS, generalized estimating equations (GEE) were used to yield population averaged models of the association between a publicly-available census tract (CT)-level geographic food access metric and diet, as measured by LS7 healthy diet score, a composite score based on intake of fruit/vegetable, fish, sodium, sweets/sugar-sweetened beverages and whole grains. Confounding by sociodemographic characteristics identified by a directed acyclic graph (DAG) and effect measure modification (EMM) by culturally-relevant sources of informational food access of the relationship between geographic food access and diet were assessed.
Aim 3

A DAG was used to identity individual- and area-level sociodemographic characteristics that had the potential to confound the relationship between each geographic food access metric and diet, as measured by the LS7 healthy diet score. Using 2012-2013 baseline data from the ACCN study, GEE were used to yield nested county- and individual-level geographic food access models adjusted for DAG-identified potential confounders. Effect estimate magnitude and statistical significance were compared across the models as were the quasi-likelihood under the independence model criterion (QIC) values of each model.

Results

Aim 1

There were significant differences in the distribution of several LS7 components and overall LS7 score between BPRHS and ACCN study participants. BPRHS participants had a more favorable blood pressure distribution, but ACCN participants had more favorable distributions of smoking, physical activity, diet and plasma glucose. ACCN participants also had a more favorable distribution of categorical overall LS7 score and a higher average continuous overall LS7 score.

There were significant differences between BPRHS and NHW NHANES participants across at least one level of available LS7 components except for diet. BPRHS participants had more favorable distributions of physical activity and total cholesterol, whereas NHW NHANES participants had more favorable distributions of smoking, BMI,
and plasma glucose. However, the distribution of categorical overall LS7 score and mean continuous overall LS7 scores did not differ substantially between the study populations.

Excluding diet, there were significant differences between ACCN study and NHW NHANES participants across at least one level of remaining LS7 components. ACCN participants tended to have more favorable distributions of smoking, physical activity and plasma glucose, whereas NHW NHANES participants had more favorable distributions of BMI and blood pressure. Although the distribution of categorical overall LS7 score was more favorable among ACCN study participants, mean continuous overall LS7 score did not differ substantially between the study populations, likely because there were more NHW NHANES participants with overall LS7 scores in the upper and lower tails of the distribution while ACCN participants tended to have overall LS7 scores in the middle of the distribution.

Aim 2

In the primary analysis, an absolute geographic food access metric representing low grocery store or large supermarket access at the CT-level was not associated with a significant change in the odds of a poor LS7 healthy diet score (odds ratio (OR): 1.20, 95% CI: 0.88-1.64, p-value: 0.25) among BPRHS participants. In the secondary analysis, a relative geographic food access metric representing the balance of healthy and unhealthy food available at the CT-level was not associated with a significant change in the odds of a poor LS7 healthy diet score at $\alpha=0.05$ (OR: 0.99, 95% CI: 0.96-1.03, p-value: 0.63). The final primary and secondary models were unadjusted given a lack of confounding and EMM.
**Aim 3**

Adjusted for potential confounders, county-level geographic food access was not associated with a significant change in the odds of poor LS7 healthy diet score among ACCN study participants (OR: 0.98, 95% confidence interval (CI): 0.96-1.00, p-value: 0.07, QIC: 507.64). Individual-level geographic food access was not associated with a significant change in the odds of poor LS7 healthy diet score among ACCN study participants, accounting for potential confounders, including county-level geographic food access (OR: 0.99, 95% CI: 0.97-1.02, p-value: 0.64, QIC: 508.90). Based on QIC values, the model that included individual-level food access did not improve model fit, and in all cases, effect estimates for the food access measures were weak in magnitude and lacked statistical significance.

**Conclusions**

In the first aim of the current study, the LS7 metric was used to determine precisely how BPRHS and ACCN study participants differed from each other and from a nationally representative research sample of NHWs in terms of the distribution of CVD and cancer risk as represented by health behaviors like smoking, diet, physical activity and BMI; and biological responses like blood pressure, total cholesterol and plasma glucose, while taking place, culture, age and sex, into account. Results suggest that smoking, BMI, and plasma glucose warrant further investigation among BPRHS participants, and BMI and blood pressure warrant further investigation among ACCN study participants. Diet was also identified as a risk behavior for prioritization among both research samples. To date, the LS7 metric has typically been explored in one
population at a time. However, as suggested by this study, LS7 can be used to identify where to focus efforts to reduce CVD and cancer risk and related disparities across diverse populations. However, an important next step in this line of inquiry is to utilize data from a variety of sources, such as electronic health records, the U.S. Census and North American Industry Classification System (NAICS) business codes (commonly used to determine access to health promoting resources such as grocery stores, hospitals, health clubs and recreational sports centers), to further understand the social determinants of health that impact CVD- and cancer-related health disparities and to inform interventions to address these disparate health outcomes.\textsuperscript{18,19}

The second aim of the study found no relationship between area-level geographic food access and LS7 healthy diet score among BPRHS participants, using publicly-available CT-level exposure measures. No confounding or EMM was found, despite using an explicit method to identify potential confounders and exploration of culturally relevant forms of informational food access that may potentially modify the geographic food access and diet relationship. Given the lack of research using multiple forms of food access, future studies should further examine the impact of culturally relevant forms of informational food access on the relationship between geographic food access and diet since informational food access, geographic food access and diet are social determinants of health that have the potential to independently and jointly contribute to CVD and cancer health disparities.

The third aim of the study found that neither county-level nor individual-level geographic food access significantly influenced diet, as measured by the LS7 healthy diet
score, among ACCN study participants. The primary data collected to derive the individual-level geographic food access measure provided insight into non-geographic factors influencing food purchasing behavior, such as low prices and good selection, and raised questions regarding the interaction between geographic food access and food purchasing behavior on diet in a rural population. With regard to the specification of causal relationships, future research should assess whether geographic food access is more appropriately treated as an EMM of the relationship between food purchasing behavior and diet, rather than as the exposure of interest, since food purchasing behavior, diet and geographic food access are social determinants of health that may independently and jointly contribute to CVD and cancer health disparities.

It is important to note limitations of this study, particularly the limited generalizability of results given characteristics of the research samples from populations facing health disparities employed. As noted in the literature, publicly-available health data on the health status of mainland Puerto Ricans and residents of Appalachia are limited, and so exploration of specific research questions may only be possible by using secondary analyses of data collected from studies focused on a very specific subset of these populations. Although NHW NHANES participants were representative of civilian, non-institutionalized U.S. NHWs, BPRHS participants were 45 to 75 years of age and were recruited using approaches such as door-to-door enumeration of randomly selected census blocks with 10 or more Hispanics within the target age range. ACCN study participants were adults recruited from churches located in select Appalachian counties, agreed to participate and had to have a BMI of at least 25 to participate, and were thus
not representative of the Appalachian population. Therefore, future studies should ideally utilize data from research samples that are more representative of U.S. populations facing health disparities. Furthermore, surveillance efforts should be adapted to include sufficient numbers of Hispanic subgroups so that reliable estimates are easily accessible for Puerto Ricans, and an indicator of Appalachian status to foster comparisons between residents of the Appalachian counties of 13 states and their non-Appalachian counterparts.

Public Health Implications

The findings of this study can be used by researchers, public health officials and policy makers to prioritize modifiable sources of CVD and cancer risk, for which BPRHS and ACCN study participants had unfavorable distributions and, for future research and interventions to ultimately reduce health disparities. In addition to diet, for which the majority of both research samples failed to reach ideal status, smoking, BMI and plasma glucose need to be further explored among BPRHS participants, and BMI and blood pressure need to be further elucidated among ACCN study participants. Importantly, exploration of these research questions should always consider how multi-level social determinants of health contribute to CVD and cancer health disparities through their impact on the health behaviors and risk factors that comprise LS7.

Despite the null results of this study and others, geographic food access remains an expanding area of research in the context of diet. Consequently, greater empirical attention is needed to clarify nuances of the relationships between diet, geographic food
access, acculturation and food purchasing behaviors related to quality and affordability. More research reflecting the complexity of the way in which geographic food access and other social determinants of health such as acculturation, cultural norms, income, personal food shopping behaviors and perceptions of food quality and price, contribute to CVD and cancer health disparities through diet, the county’s leading modifiable chronic disease risk factor, is necessary to better inform intervention and policy efforts seeking to improve diet among populations facing health disparities.
Dedicated to:

My husband, Jason, and English bulldog, Colonel Mustard, for their patience, encouragement and unconditional love each and every day,

My parents, Bill and Jeri, for allowing their daughters to pave their own paths and supporting me through all the twists and turns,

&

Captain Jack, my belated Pa, who ran a tight ship and never missed an opportunity to stress the importance of training. Here is that “report” we often talked about!
Acknowledgments

It is difficult to describe how it feels to have accomplished this daunting personal and professional goal and to articulate my gratitude to all those who have been a part of the journey. Firstly, I would like to thank my committee chair, Dr. Electra Paskett, a force to be reckoned with in the realm of cancer control. It has been inspiring to witness firsthand her unwavering commitment to reduce and eliminate cancer health disparities among underserved populations. Among other opportunities that she has provided to challenge me and foster my growth as an independent researcher, Dr. Paskett allowed me to collect primary data as part of my dissertation and encouraged me to apply to the Centers of Population Health and Health Disparities Cross-Center Training Program. I am thankful she saw potential in me four years ago and chose to invest in me.

I would also like to thank the additional members of my committee, Drs. Randi Foraker, Michelle Kaiser and Abigail Shoben. Dr. Foraker exposed me to the Life’s Simple 7 metric in my first quarter at Ohio State during her thought-provoking Cardiovascular Epidemiology course. I am extremely grateful that she invited me to take part in a number of collaborative projects over the years and I value the mentorship she has provided. I was exposed to a food mapping initiative led by Dr. Kaiser two years ago and was immediately impressed by her ability to foster collaboration across academic disciplines and among academic, government and community partners. Although happily
done with coursework, I am envious of students who have the opportunity to benefit from Dr. Kaiser’s food accessibility expertise by enrolling in her new applied food system service-learning course. I have benefited from Dr. Shoben’s expertise in Biostatistics and approachability for more than a year. I appreciate the time she generously spent with me talking through unexpected and unfamiliar issues so that I could independently tackle data analyses with confidence. I thank her for answering too many questions to count.

I am thankful for the people I met and the experiences I had while beginning my public health career in Boston, Massachusetts that piqued my interest in pursuing a doctoral degree. My eyes were opened to issues on the front line of community health and health disparities by the legendary Elmer Freeman and employment opportunities within Martha Eliot Health Center and Upham’s Corner Community Health Center. I am very fortunate to have the continued support of inspiring researchers who mentored me as a graduate student at Northeastern University, notably Drs. Theresa Osypuk, Roger Edwards, Dharma Cortes, Lindsay Rosenfeld and Shan Mohammed.

I am equally appreciative of all those with whom I have had the opportunity to work while being part of the Division of Population Sciences for the past four years. My supervisor, Dr. Jay Fisher, has been an exceptional mentor and calming force. He provided me with opportunities to gain undergraduate teaching experience and I will dearly miss working with him and other collaborators to develop Ohio Cancer Incidence Surveillance System reports. I am thankful to have received support from Drs. Adana Llanos, Jessica Krok-Schoen and Jesse Plascak and doctors-in-training, Sarah Reisinger, Brittany Brewer and Melinda Hill. I would like to thank Cecelia DeGraffenreid for
teaching me how to navigate the institutional review board, Evelyn Walters for working her calendar magic at my frequent request and my lunch pals for prying me away from the computer every now and again.

It would not have been possible to complete my dissertation without the support of the Appalachian Community Cancer Network, under the direction of Dr. Mark Dignan, and the Network’s research study team, particularly Ryan Baltic, Cathy Tatum, Darla Fickle, Ann Jena, Greg Young and Dr. Mira Katz. I am equally grateful for the Centers of Population Health and Health Disparities Cross-Center Training Program, through which I reconnected with Dr. Katy Tucker, who generously allowed me to use data from her Boston Puerto Rican Health Study for my dissertation.

I am very lucky to have been part of a doctoral Epidemiology cohort that created a wonderfully collaborative environment in which we could learn from one another and share a few laughs over the rollercoaster that was the past four years. I would like to give a special thanks to Lindsay Asti, for being my qualifying exam study buddy and networking asset, and to Thalia Farietta, for being my partner in crime, from the computer lab to the gym.

Last but not least, I must thank my immediate and extended family. Jason, my kind and patient husband, has been my rock throughout this entire experience and, unlike me, never once complained. He has set the bar very high for being a supportive spouse and I look forward to returning the favor should he choose to pursue a doctoral degree. I am thankful for my loveable English bulldog, Colonel Mustard, who is happy to see me, no matter what kind of day I have had, and the countless therapeutic snuggle sessions he
has provided. I am indebted to my parents, Bill and Jeri, who always encouraged my sisters and me to pursue our vastly different academic, professional and personal goals. I do not take for granted the freedom and flexibility they gave me to choose and adjust my own path. As the baby of the family, I have turned to my older and wiser sisters, Katherine and Annie, for advice. Both they and their families have been an amazing source of love and laughter.

I also must thank my in-laws, Richard and Liz, and acquired siblings, Jen and John, and their families for sending encouragement from afar. I would like to express my gratitude to the entire Cusack clan of Braintree, Massachusetts for taking such excellent care of me when I was in Boston for the Centers of Population Health and Health Disparities Cross-Center Training Program and a variety of public health conferences. As always, my extended family greeted me with open arms and made me feel right at home. I am especially thankful for my loving and witty Nana, Mary, who can lift my spirits with just the sound of her voice. She has provided many a pep talk in person and over the phone. I must also thank my late grandparents, Janet, Bernie and Jack, who were the first college graduates in their families, in spite of humble beginnings and serving their country in World War II. The sacrifices they made have opened doors for their children, grandchildren and great-grandchildren…a legacy to which I aspire using my newly earned doctorate.
Vita

June 2015 - present……………………… Data/Statistics Analyst, Nationwide Children’s Hospital, Partners for Kids, Columbus, OH

June 2013 – August 2015………………..Resource Planning Analyst, The Ohio State University, Comprehensive Cancer Center, Population Sciences, Columbus, OH

January 2013 - May 2013………………..Graduate Research Associate, The Ohio State University, Kirwan Institute for the Study of Race and Ethnicity, Columbus, OH

June 2011 - May 2013………………….. Graduate Research Associate, The Ohio State University, Comprehensive Cancer Center, Population Sciences, Columbus OH

January 2010……………………………...M.P.H. Urban Health, Northeastern University, Boston, MA

May 2008……………………………….. B.A. Psychology, Kenyon College Gambier, OH

2006/2007 Academic Year………………..Undergraduate Study Abroad Experience, National University of Ireland, Galway, Galway, Ireland
Publications

Baltic, RD, **Weier, RC**, Katz, ML, Kennedy, S, Lengerich, EJ, Lesko, SM, Reese, D, Roberto, KA, Schoenberg, NE, Young, GS, Dignan, MB and Paskett, ED. Study design, intervention, and baseline characteristics of a faith-based healthy eating and physical activity intervention (Walk by Faith) to reduce overweight and obesity and cancer risk among Appalachian adults. Contemporary Clinical Trials. doi:10.1016/j.cct.2015.06.017


Krok-Schoen, JL, Kurta, ML, **Weier, RC**, Young, GS, Carey, AB, Tatum, CM and Paskett, ED. (2014) Clinic type and patient characteristics affecting time to resolution after an abnormal cancer-screening exam. Cancer Epidemiology, Biomarkers & Prevention. Published OnlineFirst October 13, 2014; doi: 10.1158/1055-9965.EPI-14-0692

Plascak, JJ, Llanos, AA, Pennell, ML, **Weier, RC** and Paskett, ED. (2014). Neighborhood factors associated with time to resolution following an abnormal breast or cervical cancer screening test. Cancer Epidemiology, Biomarkers & Prevention. Published OnlineFirst September 9, 2014; doi:10.1158/1055-9965.EPI-14-0348


Fields of Study

Major Field: Public Health
Table of Contents

Abstract............................................................................................................................................ii
Dedication.................................................................................................................................xv
Acknowledgments.....................................................................................................................xvi
Vita................................................................................................................................................xx
List of Abbreviations..................................................................................................................xxxiii
List of Figures................................................................................................................................xxxvi
List of Tables............................................................................................................................xxxviii
Chapter 1: Introduction..............................................................................................................1
  1.1 Background of problem......................................................................................................1
  1.2 Justification for study........................................................................................................1
  1.3 Purpose of study................................................................................................................3
Chapter 2: Background and significance....................................................................................7
  2.1 Preface................................................................................................................................7
  2.2 Populations facing health disparities..............................................................................8
    2.2.1 Mainland Puerto Ricans.........................................................................................8
    2.2.2 Residents of Appalachia......................................................................................11
  2.3 Cardiovascular disease (CVD)......................................................................................14
    2.3.1 Epidemiology.......................................................................................................15
2.3.1.1 Current and predicted global morbidity and mortality……15
2.3.1.2 Current and predicted U.S. morbidity and mortality……16
   2.3.1.2.1 Mainland Puerto Ricans…………………………19
   2.3.1.2.2 Residents of Appalachia………………………19
2.4 Cancer……………………………………………………………20
   2.4.1 Epidemiology………………………………………………21
      2.4.1.1 Current and predicted global morbidity and mortality…21
      2.4.1.2 Current and predicted U.S. morbidity and mortality……23
         2.4.1.2.1 Mainland Puerto Ricans……………………26
         2.4.1.2.2 Residents of Appalachia……………………26
2.5 Overlapping risk behaviors and factors…………………………27
   2.5.1 Prevention recommendations………………………………27
   2.5.2 Risk behaviors………………………………………………29
      2.5.2.1 Smoking………………………………………………30
         2.5.2.1.1 Global burden……………………………………30
         2.5.2.1.2 U.S. burden………………………………………31
            2.5.2.1.2.1 Mainland Puerto Ricans…………………31
            2.5.2.1.2.2 Residents of Appalachia………………31
      2.5.2.2 Body mass index (BMI)……………………………32
         2.5.2.2.1 Global burden……………………………………32
         2.5.2.2.2 U.S. burden………………………………………33
            2.5.2.2.2.1 Mainland Puerto Ricans…………………33
            2.5.2.2.2.2 Residents of Appalachia………………34

xxiii
3.2.2 Measures..................................................................................................................77
  3.2.2.1 Life’s Simple 7 (LS7) health behaviors........................................77
  3.2.2.2 Life’s Simple 7 (LS7) health factors.............................................78
  3.2.2.3 Additional variables.................................................................78

3.3 The Appalachian Community Cancer Network (ACCN) Study..............78
  3.3.1 Study population and design..........................................................78
  3.3.2 Measures.............................................................................................80
    3.3.2.1 Life’s Simple 7 (LS7) health behaviors.............................80
    3.3.2.2 Life’s Simple 7 (LS7) health factors.................................81
    3.3.2.3 Additional variables..........................................................81

3.4 External data sources.......................................................................................82
  3.4.1 U.S. Department of Agriculture (USDA) Food Environment
       Atlas..................................................................................................................82
  3.4.2 Centers for Disease Control and Prevention (CDC).........................83
  3.4.3 U.S. Census.............................................................................................84
  3.4.4 Appalachian Regional Commission (ARC)......................................84
  3.4.5 National Health and Nutrition Examination Survey (NHANES)......85
    3.4.5.1 Life’s Simple 7 (LS7) health behaviors..................................85
    3.4.5.2 Life’s Simple 7 (LS7) health factors......................................86

3.5 Aim 1......................................................................................................................86
  3.5.1 Data Sources..............................................................................................86
    3.5.1.1 Boston Puerto Rican Health Study (BPRHS).........................86
3.5.1.2 Appalachia Community Cancer Network (ACCN) Study.................................................................86
3.5.1.3 National Health and Nutrition Examination Survey (NHANES).....................................................................................86
3.5.2 Analyses.........................................................................................................................87
  3.5.2.1 Individual LS7 Components and Overall LS7 Score............87
  3.5.2.2 Comparison between BPRHS and ACCN .....................88
  3.5.2.3 Comparisons between BPRHS and NHANES and ACCN and NHANES .................................................................90
3.6 Aim 2........................................................................................................................................92
  3.6.1 Data...............................................................................................................................92
    3.6.1.1 Outcome.................................................................................................................92
    3.6.1.2 Exposure................................................................................................................93
    3.6.1.3 Effect measure modifiers.................................................................93
    3.6.1.4 Potential confounders....................................................................................94
  3.6.2 Analyses..................................................................................................................................95
    3.6.2.1 Primary Analysis........................................................95
    3.6.2.2 Secondary Analysis.................................................................96
3.7 Aim 3........................................................................................................................................98
  3.7.1 Data...............................................................................................................................99
    3.7.1.1 Outcome.................................................................................................................99
    3.7.1.2 Exposure................................................................................................................99
3.7.1.3 Potential confounders.............................................99

3.7.2 Analyses.................................................................100

Chapter 4: The Prevalence of Life’s Simple 7 of Cardiovascular Health Metric in The
Boston Puerto Rican Health Study and The Appalachia Community Cancer Network
Study.................................................................103

4.1 Abstract..............................................................................103

4.2 Introduction.................................................................107
  4.2.1 Purpose.........................................................................109

4.3 Methods...........................................................................110
  4.3.1 Study Populations.......................................................110
    4.3.1.1 Boston Puerto Rican Health Study (BPRHS)............110
    4.3.1.2 Appalachia Community Cancer Network (ACCN)
    Study.............................................................................111
    4.3.1.3 National Health and Nutrition Examination Survey
    (NHANES)....................................................................113
  4.3.2 Analyses.................................................................114
    4.3.2.1 Individual LS7 Components and Overall LS7 Score .....114
      4.3.2.1.1 Health Behaviors............................................115
      4.3.2.1.2 Health Factors............................................115
      4.3.2.1.3 Overall LS7 Score....................................116
    4.3.2.2 Comparison between BPRHS and ACCN........116
5.3.1 Study Population.........................................................141
5.3.2 Outcome of Interest...................................................142
5.3.3 Exposure of Interest..................................................143
5.3.4 Effect Measure Modifiers of Interest..............................144
5.3.5 Potential Confounders.............................................145
5.3.6 Analyses.................................................................146
    5.3.6.1 USDA Low Access Primary Analysis.......................146
    5.3.6.2 mRFEI Sensitivity Analysis..................................147
5.4 Results........................................................................149
    5.4.1 USDA Low Access Primary Analysis...........................149
        5.4.1.1 Participant Characteristics.................................149
        5.4.1.2 Diet Characteristics........................................150
        5.4.1.3 Model..........................................................150
        5.4.1.4 Model Diagnostics...........................................151
    5.4.2 mRFEI Sensitivity Analysis......................................151
        5.4.2.1 Participant Characteristics.................................151
        5.4.2.2 Diet Characteristics........................................152
        5.4.2.3 Model..........................................................152
        5.4.2.4 Model Diagnostics...........................................152
5.5 Discussion.....................................................................154
    5.5.1 Strengths.............................................................155
    5.5.2 Limitations..........................................................157

xxx
Chapter 6: The Relationship between Geographic Food Access and Life’s Simple 7 Healthy Diet Score in The Appalachia Community Cancer Network Study

6.1 Abstract

6.2 Introduction

6.2.1 Purpose

6.3 Methods

6.3.1 Study Population

6.3.2 Outcome of Interest

6.3.3 Exposures of Interest

6.3.3.1 County-Level

6.3.3.2 Individual-Level

6.3.4 Potential Confounders

6.3.5 Analyses

6.4 Results

6.4.1 Participant Characteristics

6.4.2 Diet Characteristics

6.4.3 Food Purchasing Behaviors

6.4.4 Models

6.4.5 Model Diagnostics

6.5 Discussion

6.5.1 Strengths
6.5.2 Limitations......................................................................................186
6.6. Conclusion.......................................................................................189
Chapter 7: Conclusions and Implications................................................191
  7.1 Aim 1 conclusions and implications...............................................192
    7.1.1 Strengths and limitations of Life’s Simple 7 (LS7).................193
  7.2 Aim 2 and 3 justification.................................................................194
  7.3 Aim 2 conclusions and implications...............................................195
    7.3.1 Suggestions for future research...............................................196
  7.4 Aim 3 conclusions and implications...............................................198
    7.4.1 Suggestions for future research...............................................199
  7.5 Broader conclusions and implications.............................................201
    7.5.1 Advancing the geographic food access literature to inform policy........................................................................................................201
    7.5.2 Improving diet in the interim......................................................203
    7.5.3 Collecting more representative data on populations facing health disparities........................................................................................................204
  7.6 Closing remarks..............................................................................205
Appendix A: Figures..............................................................................209
Appendix B: Tables................................................................................224
Appendix C: ACCN Food Access Questionnaire.....................................238
Bibliography..........................................................................................244

xxxii
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCN</td>
<td>Appalachia Community Cancer Network</td>
</tr>
<tr>
<td>ACS</td>
<td>American Cancer Society</td>
</tr>
<tr>
<td>AmCS</td>
<td>American Community Survey</td>
</tr>
<tr>
<td>AHA</td>
<td>American Heart Association</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike information criterion</td>
</tr>
<tr>
<td>AOR</td>
<td>Adjusted odds ratio</td>
</tr>
<tr>
<td>ARC</td>
<td>Appalachian Regional Commission</td>
</tr>
<tr>
<td>ARIC</td>
<td>Atherosclerosis Risk in Communities</td>
</tr>
<tr>
<td>BAS</td>
<td>Bi-dimensional Acculturation Scale</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index</td>
</tr>
<tr>
<td>BPRHS</td>
<td>Boston Puerto Rican Health Study</td>
</tr>
<tr>
<td>BRFSS</td>
<td>Behavioral Risk Factor Surveillance System</td>
</tr>
<tr>
<td>CB</td>
<td>Census block</td>
</tr>
<tr>
<td>CBG</td>
<td>Census block group</td>
</tr>
<tr>
<td>CDC</td>
<td>Centers for Disease Control and Prevention</td>
</tr>
<tr>
<td>CES-D</td>
<td>Center for Epidemiology Studies-Depression</td>
</tr>
<tr>
<td>CHD</td>
<td>Coronary heart disease</td>
</tr>
<tr>
<td>CPHHD</td>
<td>Centers for Population Health and Health Disparities</td>
</tr>
<tr>
<td>CSA</td>
<td>Community supported agriculture</td>
</tr>
<tr>
<td>CT</td>
<td>Census tract</td>
</tr>
<tr>
<td>CVD</td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>CVH</td>
<td>Cardiovascular health</td>
</tr>
<tr>
<td>DAG</td>
<td>Directed acyclic graph</td>
</tr>
<tr>
<td>DALYs</td>
<td>Disability-adjusted life years</td>
</tr>
<tr>
<td>DF</td>
<td>Degrees of freedom</td>
</tr>
<tr>
<td>EHES</td>
<td>European Health Examination Survey</td>
</tr>
<tr>
<td>EMM(s)</td>
<td>Effect measure modification/effect measure modifiers(s)</td>
</tr>
<tr>
<td>FFQ</td>
<td>Food-frequency questionnaire</td>
</tr>
<tr>
<td>FY</td>
<td>Fiscal year</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographic Information Systems</td>
</tr>
<tr>
<td>HCHS/SOL</td>
<td>Hispanic Community Health Study/Study of Latinos</td>
</tr>
<tr>
<td>HHANES</td>
<td>Hispanic Health and Nutrition Examination Survey</td>
</tr>
</tbody>
</table>
HONU  Heart of New Ulm
HPSA  Health Professional Shortage Area
IPAQ  International Physical Activity Questionnaire
IQR   Interquartile range
IRB   Institutional Review Board
LS7   *Life's Simple 7*
MAUP  Modifiable Areal Unit Problem
MESA  Multi-ethnic Study of Atherosclerosis
mg/dL  Milligram per deciliter
MLE   Maximum likelihood estimator
MMSE  Mini Mental State Examination
mRFEI  Modified Retail Food Environment Index
NAICS  North American Industry Classification System
NCHS  National Center for Health Statistics
NCI   National Cancer Institute
NHANES  National Health and Nutrition Examination Survey
NHIS  National Health Interview Survey
NIH   National Institutes of Health
NHW(s)  Non-Hispanic white(s)
PAR-Q  Physical Activity Readiness Questionnaire
PLANET  Plan, Link, Act, Network with Evidence-based Tools
PIR   Poverty-to-income ratio
PSS   Perceived Stress Scale
QIC   Quasi-likelihood under the independence model criterion
RTIPs  Research-tested intervention programs
SD    Standard deviation
SEER  Surveillance, Epidemiology, and End Results
SE    Standard error
SES   Socioeconomic status
SNAP  Supplemental Nutrition Assistance Program
U.S.  United States
USDA  U.S. Department of Agriculture
USDHHS  U.S. Department of Health and Human Services
USPTF  U.S. Preventive Task Force
WCRF  World Cancer Research Fund
WHF   World Heart Federation
WHI   Women's Health Initiative
WHO   World Health Organization
WIC   Women, Infants and Children Program

xxxiv
YLDs  Years lived with disability
YLLs  Years of life lost to premature mortality
# List of Figures

| A.1  | Appalachian County FY2012 Economic Status                      | Page 210 |
| A.2  | ARC-defined Appalachian sub-regions                            | Page 211 |
| A.3  | Relationship between modifiable CVD and cancer risk factors and behaviors | Page 212 |
| A.4  | CPHHD model for analysis of population health and health disparities | Page 213 |
| A.5  | Initial DAG for Aim 2                                          | Page 214 |
| A.6  | Final DAG for Aim 2                                            | Page 215 |
| A.7  | Initial DAG for Aim 3                                          | Page 216 |
| A.8  | Final DAG for Aim 3                                            | Page 217 |
| A.9  | Unadjusted prevalence of poor, intermediate and ideal LS7 metric components of BPRHS participants | Page 218 |
| A.10 | Unadjusted prevalence of poor, intermediate and ideal LS7 metric components of ACCN study participants | Page 219 |
| A.11 | Unadjusted prevalence of poor, intermediate and ideal LS7 metric components of NHW 2007-2010 NHANES participants | Page 220 |
| A.12 | Comparison of age- and sex-adjusted LS7 metric component prevalence among BPRHS and ACCN study participants | Page 221 |
| A.13 | Comparison of age- and sex-adjusted LS7 metric component prevalence among BPRHS and NHW 2007-2010 NHANES participants | Page 222 |
A.14 Comparison of age- and sex-adjusted LS7 metric component prevalence among ACCN study and NHW 2007-2010 NHANES participants........................223
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.1</td>
<td>AHA LS7 of CVH</td>
<td>225</td>
</tr>
<tr>
<td>B.2</td>
<td>ACS guidelines on nutrition and physical activity for cancer prevention: recommendations for individual choice</td>
<td>226</td>
</tr>
<tr>
<td>B.3</td>
<td>Summary of relationship between data sources and specific aims</td>
<td>227</td>
</tr>
<tr>
<td>B.4</td>
<td>LS7 components derivable from BPRHS and ACCN datasets</td>
<td>228</td>
</tr>
<tr>
<td>B.5</td>
<td>Comparison of age- and sex-adjusted LS7 metric component prevalence among BPRHS and ACCN study participant</td>
<td>229</td>
</tr>
<tr>
<td>B.6</td>
<td>Comparison of age- and sex-adjusted LS7 metric component prevalence among BPRHS and NHW 2007-2010 NHANES participants</td>
<td>230</td>
</tr>
<tr>
<td>B.7</td>
<td>Comparison of age- and sex-adjusted LS7 metric component prevalence among ACCN study and NHW 2007-2010 NHANES participants</td>
<td>231</td>
</tr>
<tr>
<td>B.8</td>
<td>Baseline BPRHS participant sociodemographic characteristics (Primary Analysis)</td>
<td>232</td>
</tr>
<tr>
<td>B.9</td>
<td>Baseline BPRHS participant diet characteristics (Primary Analysis)</td>
<td>233</td>
</tr>
<tr>
<td>B.10</td>
<td>Baseline BPRHS participant sociodemographic characteristics (Secondary Analysis)</td>
<td>234</td>
</tr>
<tr>
<td>B.11</td>
<td>Baseline BPRHS participant diet characteristics (Secondary Analysis)</td>
<td>235</td>
</tr>
<tr>
<td>B.12</td>
<td>Baseline ACCN study participant sociodemographic characteristics</td>
<td>236</td>
</tr>
</tbody>
</table>
B.13 Baseline ACCN study participant diet characteristics…………………………237
Chapter 1: Introduction

1.1 Background of problem

Smoking, overweight and obesity, inadequate physical activity, poor diet, high blood pressure, high total cholesterol and high plasma glucose are well-defined, modifiable risk behaviors and factors influencing cardiovascular (CVD) and cancer outcomes through similar inflammatory pathways. These mutual risk behaviors and factors highlight the potential for cost-effective interventions to prevent CVD and cancer outcomes and reduce related health disparities, differences in the incidence, prevalence, mortality and overall burden of CVD and cancer that should not exist. However, in order to inform intervention efforts and make progress toward public health benchmarks and CVD and cancer prevention recommendations, the status of overlapping CVD and cancer risk behaviors and factors among populations facing health disparities, and the social determinants of health that contribute to CVD and cancers through their impact on these risk behaviors and factors need to be elucidated.

1.2 Justification for study

This study assessed the distribution of overlapping CVD and cancer risk behaviors and factors in two relatively understudied populations facing health disparities: Puerto Ricans living in the contiguous U.S. (herein referred to as mainland Puerto Ricans) and residents of Appalachia. Puerto Ricans, the second largest U.S. Hispanic subgroup, are greatly overshadowed in the epidemiologic literature by Mexican
Americans, the country’s largest Hispanic subgroup. However, Puerto Ricans experience worse health outcomes and higher prevalence of acute and chronic medical conditions than non-Hispanic whites (NHWs), Mexican Americans and other Hispanic subgroups. Health disparities related to CVD and cancer have also been documented between island and mainland Puerto Ricans.

Appalachia is a federally-designated region of the U.S. along the Appalachian mountains home to over 25 million residents. Designation of Appalachian status occurs at the county level on the basis of socioeconomic indicators, and there are marked differences in economic strength, sociodemographic characteristics and the availability of health-promoting resources between Appalachia and the U.S. as well as within Appalachia. These differences likely contribute to documented CVD and cancer disparities between Appalachia and the U.S. and within the region.

U.S. public health benchmarks to address CVD and cancer disparities include the American Heart Association’s (AHA) 2020 Impact Goal “to improve the cardiovascular health of all Americans by 20% while reducing deaths from cardiovascular diseases and stroke by 20%” and the U.S. Department of Health and Human Services’ (USDHHS) Healthy People 2020 objective to reduce the cancer mortality rate by 10% from 179.3 per 100,000 in 2007 to 161.4 per 100,000 in 2020. Life’s Simple 7 (LS7) of cardiovascular health (CVH), a metric to monitor the AHA 2020 Impact Goal progress, is well-suited to quantify gradients in CVD and cancer risk. This metric includes well-defined poor, intermediate and ideal levels of four health behaviors (smoking, body mass index,
physical activity, and diet) and 3 health factors (total cholesterol, blood pressure and plasma glucose) and is associated with both CVD and cancer outcomes. 1, 5, 7, 8

The majority of studies that have used the LS7 metric to quantify the distribution of overlapping CVD and cancer risk factors and behaviors among research samples from populations facing CVD and cancer disparities have not examined multi-level social determinants of LS7 components or an overall LS7 score. 9-12 A recent analysis of Multi-ethnic Study of Atherosclerosis (MESA) data found neighborhood environmental characteristics such as favorable food stores (i.e. food stores likely carrying fresh produce), physical activity resources, walking/physical activity environment and neighborhood SES were associated with higher odds of having an ideal overall LS7 score. 10 Although informative, exploring determinants of an overall LS7 score within a population facing health disparities does not offer insight into multi-level social determinants at the same level of granularity as examining LS7 metric components individually. Among the seven LS7 components, diet was the leading cause of the overall disease burden in the U.S. between 1990 and 2010, and national estimates from the 2003-2008 National Health and Nutrition Examination Survey (NHANES) stratified by age indicate that an ideal LS7 healthy diet score is the least prevalent measure among U.S. adults (range: 0.2% to 2.6%) 13, 14

1.3 Purpose of study

Disparities in CVD and cancer health outcomes are influenced by a myriad of social determinants of health that can be categorized into three levels: distal (i.e., population), intermediate (i.e., area) and proximal (i.e., individual) (Figure A.4). In brief,
distal factors like policy exist at the population-level and are considered fundamental causes of health disparities. Intermediate health determinants include the social and physical contexts and social relationships connecting distal health determinants to the level of the individual such as culture, place, area-level income and access to grocery stores. Proximal health determinants (e.g., sociodemographic characteristics like age, sex, and household-level income; health behaviors like smoking, diet, physical activity and BMI; and biological responses like blood pressure, total cholesterol and plasma glucose) occur at the individual level and have an independent effect from those at the intermediate and distal levels. Social determinants at each of the three levels have the ability to influence CVD and cancer health disparities independently and directly or through interactions across three levels.

To address the need for improved understanding of the social determinants of health that the impact CVD and cancer health disparities, the overall goals of this study were to quantify the distribution of the LS7 metric among mainland Puerto Ricans and residents of Appalachia, taking into account place, culture, age and sex, social determinants of health that also contribute to disparate CVD and cancer health outcomes, and to prioritize LS7 components for future research. Given that the LS7 healthy diet score is the measure most Americans fail to meet, this study also sought to assess geographic food access as a social determinant of health with the potential to contribute to CVD and cancer disparities through its impact on diet, as measured by the LS7 healthy diet score, while accounting for potential confounding by area- and individual-level sociodemographics. With increased empirical attention given to geographic food access
due to the burgeoning use of geographic information systems (GIS) and availability of federal, state and local incentives to improve access to healthy food in underserved areas, geographic food access, distance-based facilitators and barriers that influence food purchasing and consumption, raises conceptual and methodological issues particularly relevant to the populations facing health disparities of interest.5,15-17

The specific aims of this study and associated hypotheses were:

1. To quantify the distribution of each LS7 component and an overall LS7 score within research samples of mainland Puerto Ricans and residents of Appalachia and compare the distributions to one another and national data (NHANES).

   a. **Hypothesis:** A large proportion of BPRHS and ACCN study participants would be categorized into the poor level of available LS7 behaviors and factors and have low overall LS7 scores. BPRHS and ACCN study participants will be relatively similar in terms of the distribution of LS7 components, overall LS7 score categories, and in terms of mean overall LS7 score. When comparing LS7 participant classification and overall LS7 score between NHANES and BPRHS participants and between NHANES and ACCN participants, there would be statistically significant differences in the proportion of participants classified into poor, intermediate and ideal levels (or some other combination of levels due to data availability) between the NHANES sample and each research sample.
2. To quantify the association between a census tract (CT)-level food access metric and individual-level LS7 healthy diet score in a research sample of mainland Puerto Ricans, controlling for individual- and CT-level sociodemographic characteristics and assessing for effect measure modification (EMM) by individual- and CT-level acculturation.

   a. **Hypothesis:** The CT-level geographic food access metric would be significantly associated with LS7 healthy diet score, after adjusting for confounding variables, and at least one indicator of acculturation (individual- or CT-level) will significantly modify the adjusted association between the CT-level geographic food access metric and LS7 score among BPRHS participants.

3. To compare nested models adjusted for individual- and area-level sociodemographic characteristics to determine whether using a county-level food access metric or individual-level home-to-preferred food store distance as the exposure of interest is more strongly associated with individual-level LS7 healthy diet score in a research sample of Appalachian residents.

   a. **Hypothesis:** The adjusted model using individual-level home-to-preferred food store distance as the exposure of interest would yield a larger, statistically significant association with LS7 healthy diet score, and would be shown to fit the data better than the adjusted model using the county-level food access measure.
Chapter 2: Background and significance

2.1. Preface

As leading causes of morbidity and mortality globally, CVD and cancer have disparate effects in low- and middle-income countries compared to high income nations as well as within high-income nations. Given the overlap between a number of well-defined, modifiable risk behaviors and factors that influence CVD and cancer outcomes, there is potential to prevent negative CVD and cancer outcomes and reduce related health disparities. However, the social determinants of health that contribute to CVD and cancer health disparities through their impact on overlapping CVD and cancer risk behaviors and factors are not well understood. This study assessed the distribution of overlapping CVD and cancer risk behaviors and factors among mainland Puerto Ricans and Appalachian residents, two populations facing health disparities that warrant increased empirical attention.

In addition to providing more background information on the two populations facing health disparities, the burden of CVD and cancer from global and national perspectives are detailed in this chapter with an emphasis on disparities. The global and national epidemiology of mutual, modifiable risk behaviors and factors are presented in the context of CVD and cancer prevention recommendations published by the AHA and the American Cancer Society (ACS). This background information sets the stage for
exploration of differences between mainland Puerto Ricans and residents of Appalachia in contrast to the remainder of the U.S in terms the AHA-defined LS7 metric associated with CVD and cancer outcomes.\textsuperscript{5, 7, 13, 39} Area and individual-level determinants of LS7 healthy diet score have been explored among mainland Puerto Ricans and residents of Appalachia since diet, a shared CVD and cancer risk factor, was the largest contributor to overall U.S. disease burden between 1990 and 2010 and the LS7 component for which the fewest American adults are ideal.\textsuperscript{14}

2.2 Populations facing health disparities

2.2.1 Mainland Puerto Ricans

Compromising 16.4\% (50.7 million) of the U.S. population in 2010, Hispanics are the nation’s largest minority and immigrant group and are expected to constitute 29\% of the country’s population by 2050.\textsuperscript{22, 40} Mexican Americans, the largest Hispanic subgroup in the U.S., account for 65\% of the nation’s Hispanic population whereas Puerto Ricans, the second largest Hispanic subgroup, account for just 9\%.\textsuperscript{22} Consequently, Mexican Americans have primarily been the focus of epidemiological research, an approach that ignores the heterogeneity of the Hispanic population in terms of geographic concentration, nativity, educational attainment, English proficiency, citizenship, income, poverty and health indicators, among other characteristics.\textsuperscript{22}

These subgroup differences are particularly important in light of the “Hispanic health paradox”, a controversial phenomenon regarding patterns of morbidity and mortality that are better than expected given social gradients in health and SES-related mortality and morbidity patterns of other minority populations that have low average SES
observed among Hispanics, who, on average, have low SES.\textsuperscript{41-43} This term has also been used to describe a potential “residual protective effect of Latino (or foreign-born) status that cannot be accounted for by measured demographic, socioeconomic, behavioral, and/or medical risk factors.”\textsuperscript{41} Both meanings are subject to a number of biases, such as the “salmon bias”, the notion that Hispanics return to their country of origin in their later years so that their death is not captured in U.S. vital statistic records, resulting in an artificially low mortality rate. The “healthy migrant hypothesis”, on the other hand, speculates that individuals who migrate to the U.S. tend to be healthier than those in their country of origin who do not migrate to the U.S.\textsuperscript{44} However, these biases are not plausible in all Hispanic subgroups. For example, the salmon bias is less likely to occur among Cubans, for whom return has historically been difficult, or for Puerto Ricans, who come from a U.S. commonwealth included in U.S. vital statistics.\textsuperscript{44}

Treating this diverse ethnic population as a homogenous group suggests that the “Hispanic paradox” applies equally to all subgroups, which the literature has shown not be the case, particularly for Puerto Ricans.\textsuperscript{24} Puerto Ricans experience higher prevalence of acute and chronic medical conditions, including CVD and cancer, and worse health outcomes than NHWs, Mexican Americans, Cuban Americans, Central and South Americans and other Hispanic subgroups.\textsuperscript{23} Furthermore, health disparities between Puerto Ricans living on the island of Puerto Rico (island Puerto Ricans) and those living in the contiguous U.S. (mainland Puerto Ricans) have been documented and are often in the opposite direction implied by the Hispanic health paradox, such that island Puerto Ricans fare better than their mainland counterparts.\textsuperscript{24-27} Despite the disparate health
outcomes they face, including those related to CVD, Puerto Ricans as a whole and mainland Puerto Ricans specifically, remain relatively understudied Hispanic subgroups.\textsuperscript{24, 45} Mainland Puerto Ricans tend to reside in urban areas in the Northeastern U.S. In fact, in 2000, an estimated 50\% of mainland Puerto Ricans live in New York, New Jersey and Connecticut.\textsuperscript{27} In light of health differences between Florida and Northeast dwelling Puerto Ricans, location-specific research that takes into account the environment in which individuals live is needed to better understand disparate health outcomes and risk behaviors and patterns of mainland Puerto Ricans.\textsuperscript{24}

It should be noted that the study of specific Hispanic subgroups, such as Puerto Ricans, is challenging since data tends to only be collected in sufficient volume to speak to Hispanics overall or Hispanics of Mexican descent.\textsuperscript{46} From 1982-1984, The Hispanic Health and Nutrition Examination Survey (HHANES), a national probability sample of 16,000 Hispanics age 6 months to 74 years, was conducted to produce stable estimates pertaining to the health of Hispanics overall, Puerto Ricans, Mexican Americans, and Cuban Americans.\textsuperscript{47} However, HHANES was not continued. Presently, the continuous NHANES oversamples all Hispanic persons rather than just Mexican-American, but reliable estimates are only available for Hispanics overall and Mexican Americans.\textsuperscript{48} Therefore, data on the incidence and prevalence of CVD and cancer are limited for Hispanic subgroups other than Mexican Americans, and data needed to stratify by length of residence in the U.S. and acculturation are not readily available.\textsuperscript{46} Currently, the Hispanic Community Health Study (HCHS)/Study of Latinos (SOL) is the largest and most comprehensive prospective population-based cohort study of 15,079
Hispanics of Mexican, Puerto Rican, Cuban, Dominican, Central American and South American ethnicity. However, HCHS/SOL data are not publicly available and participants were recruited from only four areas of the U.S., which may limit their representativeness.\textsuperscript{49}

\textbf{2.2.2. Residents of Appalachia}

Appalachia, a federally-designated region of the U.S. along the Appalachian mountains, is composed of 420 contiguous counties spanning 13 states from Mississippi to New York and home to over 25 million residents.\textsuperscript{28} Of Appalachia’s population, 60% live in metropolitan counties, 25% live in counties adjacent to metropolitan counties and the remainder live in more remote, rural locations.\textsuperscript{31} In addition to geographic diversity, the region is characterized by economic, sociodemographic and resource-based diversity.

Appalachian status designation occurs at the county level on the basis of socioeconomic indicators. Each fiscal year (FY), counties are classified into five levels of economic strength by the Appalachian Regional Commission (ARC) using poverty, unemployment, and per capita market income indicators relative to the national average: distressed, at-risk, transitional, competitive or attainment (Figure A.1).\textsuperscript{31} Counties that have (1) a poverty rate greater than or equal to 150% of the national average, per capita market income less than or equal to two thirds of the national average and unemployment rate greater than or equal to 150% of the national average or (2) a poverty rate greater than or equal to 200% that of the national average but meet one of the other criteria are considered distressed, the weakest category of economic strength.\textsuperscript{31} Of the nation’s counties, distressed counties rank among the worst 10% in terms of economic strength,
whereas counties that are at-risk of becoming distressed rank between the worst 10-25%, counties transitioning between strong and weak economies rank between worst 25% to best 25%, counties that are competitive in the national economy rank between the best 10-25% and attainment counties rank among the best 10%. 31 Nationally, in 1960, 29% of U.S. counties met the ARC’s distressed criteria, but only 12% did so in 2000. In Appalachia, 54.4% of counties were distressed in 1960 whereas 21.7% of counties met this designation in 2000. Economic progress has not been felt equally across the region such that 76% of counties in Southeastern Ohio, West Virginia, and Kentucky that were designated as distressed in 2000 also met this designation in four of the five past decennial censuses. 29

Appalachian counties are also organized by sociodemographic composition similarities into Northern, Central and Southern sub-regions (Figure A.2) 50. As with economic strength, there are sociodemographic differences between the rest of the country and the Appalachian region, as well as within Appalachia. For example, per 2007-2011 American Community Survey (AmCS) data, 64.2% of U.S. residents and 83.9% of Appalachian residents were NHW, 14.6% of U.S. residents and 16.5% of Appalachian residents ages 25 years and older had not attained a high school diploma, and the poverty rate was 14.3% among U.S. residents and 16.1% among Appalachian residents. 30 Within Appalachia, the percentage of NHW was lowest in the Southern sub-region (70.4%) and highest in the Central sub-region (95.5%). The percentage of residents not attaining a high school diploma was lower than the national estimate in Northern Appalachia (11.8%) and highest in Central Appalachia (27.2%), and the poverty
rate was lower than the U.S. in Northern Appalachia (13.8%) and highest in the Central Appalachia (23.5%).

Access to health-promoting resources like healthcare services have been found to be unevenly distributed between Appalachia and the country, as well as within the region. As a group, Appalachian counties experience disparate healthcare access compared to the states in which they are located and the national average. Within Appalachia, of 406 Appalachian counties examined in a study using 2000 Health Professional Shortage Area (HPSA) codes, 108 had county-wide health professional shortages, in which the ratio of primary care physicians to population was less than 1 to 4,000, and 189 had shortages in a portion of the county, meaning only 109 counties had no shortages. Central West Virginia, Eastern Kentucky, Northeastern Mississippi, and Central Alabama were areas in which county-wide health professional shortages were most concentrated. This study also found that of 406 counties, 81 had no hospitals and 203 had a single hospital. Large numbers of hospitals were limited to Appalachian counties with large metropolitan areas.

Differences in economic strength, sociodemographic characteristics and the availability of health-promoting resources between Appalachia and the U.S. and within Appalachia sub-regions likely contribute to differential health outcomes, including cancer, and related burdens of risk behaviors and factors. However, data to monitor Appalachian-specific cancer and risk factor and behavior trends are limited in volume and quality and available data are often outdated. For example, there are no public health surveillance efforts spanning the Appalachian region in its entirety, sub-regions have not
been studied equally and it is not standard for state or national surveys like the Behavioral Risk Factor Surveillance System (BRFSS) or NHANES to identify Appalachian residents. Among other items, research spanning Appalachian sub-regions that takes the environment in which individuals live is needed to better understand disparate health outcomes and risk behaviors and patterns of Appalachian residents.

2.3 Cardiovascular disease (CVD)

CVD encompasses diseases of the heart and blood vessels, several of which are caused by atherosclerosis. Atherosclerosis is the chronic pathological process through which plaque, deposits of fatty material and cholesterol, modify the inner surface of arteries, medium- and large-sized blood vessels, thereby narrowing the artery opening and impeding blood flow. If ruptured, atherosclerotic plaques can lead to the formation of blood clots, causing acute CVD events. Atherosclerosis-related CVDs include ischemic heart disease or coronary heart disease (CHD), which are diseases of the blood vessels that supply blood to the heart muscle such as myocardial infarction (heart attack); cerebrovascular diseases, which are diseases of the blood vessels that supply blood to the brain such as stroke; and diseases of the arteries that supply blood to the arms and legs, such as peripheral vascular disease. Although previously viewed as passive accumulation of plaque, atherosclerosis is now considered an ongoing inflammatory response. Research has implicated inflammation as a key contributor throughout the atherosclerotic process, from early atherogenesis to the development of plaque and subsequent complications.
Preclinical atherosclerosis begins as early as childhood, meaning individuals may be at risk for decades before experiencing a clinical CVD event.\textsuperscript{54} An analysis that included ante mortem risk factor and behavior data on 93 autopsied individuals age 2 to 39 at death indicated that several modifiable cardiovascular risk behaviors and factors were significantly associated with the extent of atherosclerotic lesions in the aorta and coronary arteries, and the more risk behaviors and factors a person exhibited, the larger the percentage of the intimal surface of the aorta or coronary arteries that were covered with atherosclerotic plaques.\textsuperscript{54}

2.3.1 Epidemiology

A detailed summary of the current and predicted CVD burdens of morbidity and mortality are presented at the global and national scale, and available data on Mainland Puerto Ricans and residents of Appalachia are presented below. Where possible, the most recent data pertaining to CVD as a broad category of disease were used. However, there were instances for which data were only available for a specific form of CVD (e.g., CHD) or available data were not up to date given aforementioned limitations of studying Mainland Puerto Ricans and residents of Appalachia.

2.3.1.1 Current and predicted global morbidity and mortality

CVD is the leading cause of morbidity and mortality globally.\textsuperscript{55} Disability-adjusted life years (DALYs), a summary population health metric representing the number of years lived with disability (YLDs) and years of life lost to premature mortality (YLLs)), facilitate the examination of CVD at a global scale. In 2010, there were 2.490 billion global DALYs, of which cardiovascular and circulatory diseases accounted for
More specifically ischemic heart disease accounted for 5.2% of DALYs, hemorrhagic stroke accounted for 2.5% of DALYs, ischemic stroke accounted for 1.6% of DALYs, and hypertensive heart disease accounted for 0.6% of DALYs. Worldwide, 30% of all deaths occurring in 2008 (17.3 million) were due to CVDs, of which 7.3 million and 6.2 million were due to CHD and stroke, respectively.

The burden of CVD is not distributed evenly across nations of differing economic position. In the past several decades, the rate of CVDs doubled in low- and middle-income countries, such that incidence rates for stroke and heart attack now exceed those in high-income countries. Furthermore, more than 80% of all CVD-related deaths occur in low- and middle-income countries.

By 2030, an estimated 23.3 million people around the world will die from CVDs. Across contrasting models based on economic and social development predictions, ischemic heart disease and cerebrovascular disease are projected to be two of the four main global causes of death in 2030. In low income countries, more deaths will be attributable to CVD than infectious diseases, maternal and perinatal conditions, and nutritional disorders combined. The global cost of CVD is expected to increase 22% between 2010 and 2030 from $863 billion to $1,044 billion in 2030 (in U.S. dollars).

2.3.1.2 Current and predicted U.S. morbidity and mortality

CVD is also the leading cause of morbidity in the U.S. In terms of prevalence, more than a third of adults have one or more types of CVD in 2014. Roughly 7.6 million Americans have experienced a heart attack and 6.8 million have experienced a stroke. In
2008, there were 785,000 incident heart attacks, 610,000 incident strokes, 470,000 recurrent heart attacks and 185,000 recurrent stroke events.\textsuperscript{59}

Despite significant gains in CVD prevention and control over the past four decades, CVD remains the leading cause of mortality in the U.S.\textsuperscript{1} For example, between 1980 and 2000, the age-adjusted CHD mortality rate decreased from 542.9 to 266.8 deaths per 100,000 and from 263.3 to 134.4 deaths per 100,000 among men and women, respectively. An estimated 44\% of the decrease in CHD mortality rates was attributed to improvements in terms of risk factors, whereas 47\% was attributed to treatment advancements.\textsuperscript{60} Furthermore, between 2000 and 2010, the morality rate attributable to CVD overall decreased by 31.0\%. However, CVD was still responsible for approximately 1 of every 3 deaths in 2010.\textsuperscript{1} Based on the 2010 overall rate of CVD-attributable death (23.5 per 100,000), there were at least 2,150 CVD-related deaths each day, over a third of which occurred before age 75, 3.7 year short of the nation’s average life expectancy.\textsuperscript{1} More people presently die from CVD each year than from cancer and chronic lower respiratory disease combined.\textsuperscript{1}

The burden of CVD is not distributed evenly within the U.S. The average annual CVD incidence rate among men ranges from 3 per 1,000 for ages 35 to 44 to 74 per 1,000 for ages 85 to 94. Comparable rates are not experienced among women until 10 years later in life, so the gap between men and women gradually narrows over the lifespan.\textsuperscript{1} The results of the 2012 National Health Interview Survey (NHIS) demonstrate racial and ethnic differences in estimated age-adjusted prevalence of two specific types of CVD: CHD and stroke.\textsuperscript{61} For example, the prevalence of CHD was 4.5\% among Asians, 5.3\%
among Hispanics/Latinos, 6.1% among NHWs, 6.5% among African Americans, 8.1% among American Indians/Alaska Natives and 10.3% among Native Hawaiians/other Pacific Islanders. Ischemic heart disease and stroke have previously been found to be inversely related to indications of SES like education, income, and poverty status. Per an analysis of 2002 NHIS data, the prevalence of these CVDs were higher among individuals with less than a high school education and incomes below or within 100% to 199% of the poverty threshold than among individuals with at least a high school education and income at least 200% of the poverty threshold. In terms of CVD-related hospitalizations, geographic clustering has been found among Medicare beneficiaries age 65 or older such that high acute myocardial infarction and stroke prevalence rates of hospitalizations were concentrated in the southeastern U.S. Furthermore, 2001 mortality data compiled by the National Center for Health Statistics of the Centers for Disease Control and Prevention (CDC) indicated that CVD mortality rates were frequently highest among blacks across all age groups.

As a result of the country’s aging and increasingly diverse population, cancer is projected to be the leading cause of death in the U.S. by 2030. However, CVD will remain a significant health issue in the U.S. since advanced age is one of many risk factors shared by CVD and cancer. A projected 43.9% of the country’s population will have some form of CVD by 2030, which will cause the total direct costs of CVD to increase to $918 billion and indirect costs attributable to lost productivity to increase to $290 billion in 2030. Between 2013 and 2030, the prevalence of CHD will increase approximately 18% and the prevalence of stroke is expected to increase 20.5% between
2012 and 2030.\textsuperscript{1} Given the aforementioned existing CVD disparities by race and ethnicity, SES and geographic location, it is likely that these populations will continue to be disproportionately burdened by CVD moving forward.

\textbf{2.3.1.2.1 Mainland Puerto Ricans}

An analysis of data from HCHS/SOL specifically reported on CHD, not CVDs as a whole. The overall prevalence of CHD was 4\% among men and 2\% among women.\textsuperscript{63} However, the prevalence of CHD was highest among Puerto Rican, Cuban and Dominican men and Puerto Rican women (5\%). The overall prevalence of stroke was 2\% among men and 1\% among women with Dominican men (4\%) and Puerto Rican women (2\%) experiencing the highest stroke prevalence.\textsuperscript{63}

Although an analysis of vital statistics records from U.S., Mexico and Cuba for the year 2000 found the pattern of CVD to be similar between Puerto Ricans and NHWs, mainland Puerto Ricans were found to experience higher rates of CHD rates than island Puerto Ricans.\textsuperscript{25} In terms of CVD mortality, mainland Puerto Ricans experienced the second highest all-cause CVD mortality rate of all U.S. racial and ethnic subgroups studied (288 per 100,000), a far second to the rate among blacks (472 per 100,000).\textsuperscript{25}

\textbf{2.3.1.2.2 Residents of Appalachia}

Although known as a population facing cancer-related health disparities, some of the country’s highest rates of CHD occur among Appalachia populations, specifically those in southern Ohio, eastern Kentucky and West Virginia.\textsuperscript{36} In fact, in 2008, the state with the highest percentage of residents reporting a diagnosis of angina or CHD in the U.S. was West Virginia (8.1\%), almost double the national average (4.3\%).\textsuperscript{36}
Furthermore, four of the five states with the highest heart disease death rates, Alabama, Mississippi, Pennsylvania and West Virginia, contain Appalachian counties.  

2.4 Cancer

Cancer captures more than 100 diseases caused by the uncontrolled division of abnormal cells that can spread throughout the body via the blood and lymph systems. Cancers are typically named to indicate the organ or cell of origin. Broad categorizations of cancer include carcinoma, which originates in skin or tissues that line or cover internal organs; sarcoma, which originates in bone, cartilage, fat, muscle, blood vessels, or other connective or supportive tissue; leukemia, which originates in blood-forming tissue (e.g. bone marrow) and results in the production of large numbers of abnormal blood cells that enter the blood stream; lymphoma and myeloma, which originate in the immune system; and central nervous system cancers, which originate in the brain and spinal cord.

As with atherosclerosis, inflammation has also been shown to affect cancer risk and development. For example, inflammatory diseases increase the bladder, cervical, gastric, intestinal, esophageal, ovarian, prostate and thyroid cancer risk, and use of non-steroidal anti-inflammatory drugs can reduce colon and breast cancer risk and mortality. In animal and human studies, inflammatory cells and mediators have been found in the microenvironment of tumors, even during early development. Targeting inflammatory cells and mediators can decrease cancer incidence and spread. Although inflammation exemplifies just one component of the multifactorial etiologies underlying different types of cancer, there are several modifiable risk behaviors and factors known to increase cancer risk and others that may affect cancer risk.
2.4.1 Epidemiology

As with CVD, a detailed summary of the current and predicted cancer burdens of morbidity and mortality are presented at the global and national scale, and available data on Mainland Puerto Ricans and residents of Appalachia are presented below. Where possible, the most recent data pertaining to cancer as a broad category of disease were used. However, there were instances for which data were only available for specific cancer types or available data were not up to date given noted data limitations of studying Mainland Puerto Ricans and residents of Appalachia.

2.4.1.1 Current and predicted global morbidity and mortality

Cancer is the leading cause of mortality in economically developed countries and the second leading cause of mortality in developing countries, where the incidence rate for all cancers combined is nearly half of that of economically developed countries.\textsuperscript{67, 68} In 2008, there were 12.7 million new cases of cancer and 7.6 million cancer-related deaths across the world. Lung cancer was the most common cancer in term of incidence (1.6 million cases, 12.7% of total) and mortality (1.4 million deaths, 18.2% of total), whereas breast cancer was the second most incident cancer (1.4 million cases, 10.9% of total) but the 5th leading cause of mortality (458,000, 6.1% of total). Other common forms of cancer include colon and rectum cancer (1.2 million incident cases and 608,000 deaths), stomach cancer (990,000 incident cases and 738,000 deaths), prostate cancer (913,000 incident cases and 261,000 deaths) and liver cancer (748,000 cases and 695,000 deaths).\textsuperscript{69}
When broken down by gender, there are differences in cancer incidence and mortality rankings. Worldwide, among men, the five leading causes of cancer incidence are cancers of the lung and bronchus (1,095,200 cases), prostate (903,500 cases), colon and rectum (663,600 cases), stomach (640,600 cases), and liver (522,400 incident cases). Among women, the five leading causes of cancer incidence are cancers of the breast (1,383,500 cases), colon and rectum (570,100 cases), cervix uteri (529,800 cases), lung and bronchus (513,600 cases) and stomach (349,000 cases). Among men, the five leading causes of cancer mortality are cancers of the lung and bronchus (951,000 deaths), liver (478,300 deaths), stomach (464,400 deaths), colon and rectum (320,600 deaths), and esophagus (276,100 deaths). Cancers of the breast (458,400 deaths), lung and bronchus (427,400 deaths), colon and rectum (288,100), cervix uteri (275,100 deaths), and stomach (273,600 deaths) compromise the five leading causes of cancer-related death among women.

There are noticeable differences in morbidity and mortality patterns for the major forms of cancer between developed and developing nations. In upper-middle-income and high-income nations, prostate and lung are the most common cancers among men, whereas breast and colorectal are the most common cancers among women. Lung cancer is the most common cause of cancer death, while prostate, breast and colorectal cancers also contribute highly to cancer-related deaths in these nations. Although cancer morbidity is much lower in low-income countries, more than two thirds of all cancer deaths occur in low- and middle-income countries. As in higher income nations, lung and breast cancers are leading causes of cancer incidence and mortality in lower income
nations. However, in lower income countries, cancer burden varies as a function of underlying risk such that cancers with infectious etiology (e.g. cancers of the cervix, stomach and liver) are also prominent.\textsuperscript{38}

Assuming no change in cancer risk from 2008, 21.4 million new cases of cancer and 13.1 million cancer-related deaths are projected globally for the year 2030.\textsuperscript{70} This equates to a 69% increase in cancer incidence and a 72% increase in cancer mortality.\textsuperscript{70} Forecasted increases in cancer incidence and mortality will have differential impacts. The estimated percentage increase in cancer incidence by 2030 will be 82% in low and 70% in lower-middle income countries compared to 58% in the upper-middle and 40% high income countries.\textsuperscript{38} The global cost of cancer is expected to increase from U.S.$ 290 billion in 2010 to U.S.$ 458 billion in 2030.\textsuperscript{58}

2.4.1.2 Current and predicted U.S. morbidity and mortality

Cancer is a significant cause of morbidity in the U.S., and is poised to overtake CVD as the nation’s leading cause of mortality within the next two decades.\textsuperscript{2} Roughly 1,665,540 new cases of invasive cancer will be diagnosed in 2014.\textsuperscript{71} Cancers of the prostate, lung and bronchus, and colon and rectum will account for half of all diagnoses among men and cancers of the breast, lung and bronchus, and colon and rectum will account for half of all diagnoses among women. In men, 27% of incident cases will be due to prostate cancer alone, whereas 29% of incident cases in women will be due to breast cancer alone.\textsuperscript{71} Although survival statistics vary greatly by cancer type and stage at diagnosis, the 5-year relative survival rate for all cancers has increased from 49% in
Advancements in early detection and treatment likely underlie survival improvement.\textsuperscript{71} Due to progress made in cancer prevention, early detection and treatment, the overall rate of cancer mortality dropped from its peak of 215.1 per 100,000 in 1991 to 171.8 per 100,000 in 2000, a 20\% decrease. However, cancer is presently the second leading cause of death in the U.S., accounting for nearly 1 of every 4 deaths.\textsuperscript{71} In 2014, about 585,720 Americans are expected to die of cancer, almost 1,600 people per day. Nearly half of all cancer deaths can be attributed to the most common causes of cancer death among men (i.e. cancers of the lung and bronchus, prostate, and colon and rectum) and among women (i.e. cancers of the lung and bronchus, breast, and colon and rectum). However, over 25\% of all cancer deaths are due to lung cancer.\textsuperscript{71}

As with CVD, the nation’s cancer burden is not evenly distributed. For example, prostate cancer is the most common form of cancer among men. However, between 2006 and 2010, African-American men had the highest incidence rate of prostate cancer (220.0 per 100,000) compared to NHW (138.6 per 100,000), Asian American/Pacific Islanders (75.0 per 100,000), American Indian/Alaska Natives (104.1 per 100,000), and Hispanics (124.2 per 100,000).\textsuperscript{71} Similar racial/ethnic incidence patterns are observed for lung and bronchus incidence among men and colon and rectum cancer incidence among both sexes. Although, among women, the breast cancer incidence rate is lower among African Americans than NHWs.\textsuperscript{71, 72} In addition to race and ethnicity, there are also differences by nativity, geography and SES, among other factors. For cancers of the uterine cervix, stomach and liver, which have infectious etiology, lower incidence rates
have been observed in select first generation Hispanic populations compared to U.S.-born NHWs.\textsuperscript{72, 73} At a national scale, overall cancer incidence among American Indian/Alaska Natives is lower than among NHWs, but there are significant geographic differences between these groups in terms of specific cancers like colon and rectal cancer, most notably in Alaska and the Northern and Southern Plains.\textsuperscript{72, 74} An analysis of 24 cancer types across 21 countries including the U.S., found, with consistency, excess risk of cancers of the nose, larynx, lung, oral cavity and pharynx, esophagus, stomach, as well as all sites combined in men of low SES and excess risk of cancers for the esophagus, stomach and cervix uteri in women of low SES.\textsuperscript{72, 75}

In terms of mortality, an analysis of mortality data from the Surveillance, Epidemiology, and End Results (SEER) 9 registries of the National Cancer Institute (NCI) spanning 1975-2009 found that African American men experienced 33% higher rate of overall cancer mortality than white men, a difference driven by colorectal, lung and bronchus, and prostate mortality rates. African American women were found to have a 16% higher rate of overall cancer mortality, a difference driven by breast and colon and rectum cancer mortality rates, despite lower incidence of breast cancer.\textsuperscript{72, 76} Similar to incidence, select Hispanic subpopulations experience higher mortality rates for cancers linked with infectious agents than NHWs.\textsuperscript{72, 77} Hispanics tend to have similar 5-year cancer-specific survival rates to NHWs, but experience higher mortality rates for cancers of the gallbladder, stomach, liver and bile duct. Conversely, African-American 5-year survival rates are lower than NHW for almost all cancer types.\textsuperscript{71, 72} As with cancer incidence, poverty and a variety of social factors contribute to differences in cancer
mortality and survival making it difficult to isolate the effect of race and ethnicity alone. \cite{72, 78, 79}

In total, cancer incidence is projected to rise from 1.6 million in 2010 to 2.3 million in 2030, a 45% increase driven by the aging and diversification of the U.S. population. The number of cancer-related deaths projected to occur in the U.S. in 2030 is 620,000.\cite{80} Subgroups that tend to experience health disparities will be disproportionately affected by the nation’s growing cancer burden. For example, cancer incidence is projected to increase by 11% among younger adults versus 67% among older adults and 31% in NHW versus a 99% increase among minorities.\cite{2}

### 2.4.1.2.1 Mainland Puerto Ricans

Although known as a population facing CVD-related health disparities compared to NHWs, mainland Puerto Ricans also experience disparate cancer-related outcomes when compared to island Puerto Ricans.\cite{26} An analysis of incidence data from Puerto Rican, New York, New Jersey and Connecticut cancer registries found that among men and women, cancer incidence rates tended to be highest among NHWs, followed by mainland Puerto Ricans then island Puerto Ricans.\cite{26} For men and women, differences between island and mainland Puerto Rican population cancer incidence rates (34% among men, 36% among women) exceeded those between mainland Puerto Ricans and NHWs (16% among men, 21% among women) (p< 0.05).\cite{26}

### 2.4.1.2.2 Residents of Appalachia

Cancer is the leading cause of mortality in Appalachia.\cite{33} In terms of incidence, between 2001 and 2003, Appalachian residents experienced higher cancer incidence rates
than the rest of the country. There were also sub-regional differences in cancer incidence rates during this time period. The Southern sub-region experienced lower cancer incidence rates than the Central and Northern sub-regions. In terms of mortality, an analysis of National Center for Health Statistics (NCHS) cancer mortality data from 2003 to 2007 found the all-cause cancer mortality rate was 7% higher in the 13 states included in the Appalachian region than in the rest of the U.S. Furthermore, across these 13 states, the all-cause cancer mortality rate was 5% higher in Appalachian counties than in non-Appalachian counties.

2.5 Overlapping risk behaviors and factors

Although factors like family history, gender, age and genetics that contribute to CVD and cancer cannot be modified, more than 70% of CVD- and 34% of cancer-related deaths globally can be attributed to a limited number of risk behaviors and factors amenable to change, including tobacco use, high body mass index (BMI), low fruit and vegetable intake and physical inactivity and alcohol use. Moreover, these common behaviors and risk factors have similar inflammatory properties (Figure A.3). In order to address the overall global and national burdens of CVD and cancer and subsequent health disparities and reach public health benchmarks, modifiable health behaviors and factors contributing to these diseases need to be clearly defined and monitored using prevention recommendations.

2.5.1 Prevention recommendations

Both globally and within the U.S., there are a plethora of CVD and cancer prevention recommendations published by different organizations, e.g., World Health
Organization (WHO), World Cancer Research Fund (WCRF), World Heart Federation (WHF), the U.S. Preventive Task Force (USPTF), NCI, ACS and AHA, just to name a few, and these recommendations tend to have a substantial amount of overlap, regardless of the disease of interest.\textsuperscript{5, 52, 84-88} Similarly, the numerous risk appraisal tools that can be used in clinical settings to predict an individual’s risk of developing and dying from CVD and cancer have commonalities, including the Framingham Risk Score and the Harvard Cancer Risk Index.\textsuperscript{89, 90}

Population-level prevention recommendations and individual-level risk appraisal tools represent the complementary prevention approaches famously articulated by Geoffrey Rose, i.e., the population-wide strategy, which seeks to shift the general population’s risk distribution in a more favorable direction, and the high-risk strategy, which targets individuals identified through screening as being at increased risk of developing a given disease.\textsuperscript{91} Both strategies are necessary to address the country’s growing burden of CVD and cancer and related disparities.\textsuperscript{92} However, when specifically seeking to reduce and ultimately eliminate CVD and cancer health disparities, the approach supported by the literature is the high-risk strategy, with emphasis on the social determinants of health. By addressing the social determinants of health underlying disparate CVD and cancer burdens among a given population that experiences disparities, there is potential to shift the distribution of CVD and cancer risk within these populations to a more favorable direction.\textsuperscript{93, 94}

For the purpose of this paper, the AHA CVD (Table B.1) and ACS cancer (Table B.2) prevention recommendations are explored, noting similarities and differences.
Compared to the recommendations made by USPTF and NCI, which are broken down by disease subtype, the AHA and ACS provide more globalized CVD and cancer prevention recommendations. In 2010, the AHA released recommendations in the form of the LS7 of cardiovascular health (CVH), a metric that includes poor, intermediate and ideal levels of four health behaviors (smoking, BMI, physical activity, and diet) and 3 health factors (total cholesterol, blood pressure and plasma glucose). Ideal CVH is defined as having ideal levels of each behavior and factor. Every five years, the ACS reviews the scientific evidence surrounding diet, physical activity and cancer risk and releases updated guidelines on nutrition and physical activity for cancer prevention. The current recommendations were published in 2012.

Modifiable overlapping CVD and cancer risk factors are presented at the global and national scale in the context of AHA CVD and ACS cancer prevention recommendations. Although presented one at a time, it is important to remember that these factors and behaviors are often experienced concurrently by an individual and may act synergistically to increase one’s risk of developing CVD and/or cancer. Since prevention recommendations are made at the population level, only modifiable, overlapping risk factors and behaviors affecting the population at large are areas of focus.

2.5.2 Risk behaviors

A detailed summary of each of the four risk factors included in the LS7 metric are presented at the global and national scale, and available data on Mainland Puerto Ricans and residents of Appalachia are presented below. Where possible, the most recent data pertaining to each risk behavior were used. However, there were instances for which
available data were up to date, given aforementioned limitations of studying Mainland Puerto Ricans and residents of Appalachia.

### 2.5.2.1 Smoking

In terms of LS7, the AHA views a current smoking status of never or quit within the past 12 months to be ideal. In contrast, the ACS recommends the avoidance of tobacco use altogether as well as secondhand exposure. As summarized in the 2004 Surgeon General’s Report on the Health Consequences of Smoking, there is sufficient evidence to support a causal relationship between smoking, subclinical atherosclerosis, CHD, cerebrovascular disease and cancers of the lung and bronchus, larynx, oral cavity and pharynx, esophagus, pancreas, bladder and kidney, cervix, stomach, and acute myeloid leukemia and insufficient but suggestive evidence of a causal relationship between smoking and colon and rectum and liver cancers.\(^9^5\) Toxins contained within cigarette smoke can induce chronic inflammation of oral, nasal and airway mucosal surfaces and change how the body’s immune system responds to external antigens.\(^9^6\) Globally, smoking is estimated to cause 10% and 71% cases of CVD and lung cancer, respectively.\(^3^8\) In 2010, the adjusted estimated population attributable fraction of CVD deaths due to smoking was 13.2% in the U.S. (95% Confidence Interval (CI): 3.5%–29.2%).\(^5\) Of 585,720 cancer deaths that occurred in the U.S. in 2014, 176,000 were projected to be due to tobacco use.\(^9^7\)

#### 2.5.2.1.1 Global burden

Around the world, roughly one billion people are smokers. At the global scale, smoking prevalence is highest among upper-middle-income countries. However, the
highest incidence of smoking is actually found among men in lower-middle-income countries.\textsuperscript{38} Annually, six million people die directly from tobacco use or indirectly from second hand smoke exposure.\textsuperscript{38} By 2030, the number of tobacco-related deaths across the world is projected to surpass 8 million deaths per year.\textsuperscript{38,98}

2.5.2.1.2 U.S. burden

Although the prevalence of current cigarette smokers declined 25\% between 1998 (24.1\%) and 2012 (18.1\%) in the U.S., cigarette smoking remains a leading cause of preventable morbidity and mortality.\textsuperscript{61,95,99} Among adults 18 years of age or older, a larger percentage of men (20.5\%) were cigarette smokers than women (15.9\%).\textsuperscript{61} In addition to gender, there are also differences in the prevalence on the basis of geography, race and ethnicity, and education. In 2012, the smoking prevalence was highest in Kentucky (28.3\%) and lowest in Utah (10.6\%).\textsuperscript{1} In 2008, American Indians/Alaska Natives had the highest prevalence (32.4\%), whereas Asians had the lowest (9.9\%) and adults 25 years of age or older with a General Educational Development certificate [GED] had the highest prevalence of smoking (41.3\%), whereas those with a graduate degree had the lowest (5.7\%).\textsuperscript{99}

2.5.2.1.2.1 Mainland Puerto Ricans

An analysis of data from participants of Cuban, Dominican, Mexican, Puerto Rican, Central American and South American descent enrolled in the Hispanic Community Health Study/Study of Latinos found that roughly 26\% of men and 15\% of women were current smokers. Puerto Ricans had the highest prevalence of smoking among men (35\%) and women (32\%).\textsuperscript{63}
2.5.2.1.2.2 Residents of Appalachia

An analysis of 2009 state BRFSS data found that the average county-level percentage of adults age 18 and older in the Appalachian region who smoke was 24.0% (standard deviation (SD): 9.2). According to state 2003-2007 BRFSS data, the smoking prevalence in the Appalachian regions of Kentucky, Ohio, Pennsylvania and Virginia ranged from 25.9% (Virginia) and 33.6% (Kentucky) among men and from 25.9% (Virginia) and 29.0% (Kentucky) among women. In Kentucky, Ohio, Pennsylvania, Virginia, the prevalence of smoking for men and women was higher in the Appalachian portion of the state compared to the non-Appalachian portion.

2.5.2.2 Body mass index (BMI)

Among adults, BMI, a ratio of weight in kilograms (kg) to height in meters (m) squared, between 25.0 and 29.9 is considered overweight and BMI greater than or equal to 30.0 is considered obese. In terms of LS7, the AHA considers BMI less than 25 to be ideal. The ACS recommends being lean without being underweight and avoiding excess weight gain, similarly suggesting a target BMI less than 25. CHD and ischemic stroke risk is positively correlated with increasing BMI. Overweight and obesity are also known to increase risk of esophageal, pancreatic, colon and rectum, female (postmenopausal) breast, endometrial, kidney, thyroid, and gallbladder cancers. There is evidence suggesting overweight and obesity may increase the risk of additional cancers including prostate, other male genitals, ovary, non-Hodgkin’s lymphoma, leukemia, liver, and hemangioma. The inflammatory mechanism underlying obesity is unique in that it is metabolic in nature. This form of inflammation, metaflammation, is a low-grade,
chronic form of inflammation initiated by metabolic cells in response to excess nutrients in the metabolic cells.\textsuperscript{105} Approximately 5\% of global CVD deaths are attributed to overweight and obesity.\textsuperscript{57} In the U.S., over the past 25 years, obesity caused approximately 14\% and up to 20\% of all cancer deaths in men and women, respectively.\textsuperscript{106}

### 2.5.2.2.1 Global burden

In 2008, across the world, 33.6\% of men and 35\% of women over the age of 20 were overweight and 9.8\% of men and 13.8\% of women were obese.\textsuperscript{38} Each year, there are 2.8 million or more global deaths due to overweight and obesity.\textsuperscript{38} Compared to low- and lower-middle-income countries, the prevalence of overweight and obesity in high-income and upper-middle income countries is more than double and triple those estimates, respectively. In high-income countries, lower SES is associated with a higher prevalence of obesity, whereas in medium- and low-income countries, there is a positive relationship between SES and obesity.\textsuperscript{107, 108}

### 2.5.2.2.2 U.S. burden

Results from the 2009-2010 National Health and Nutrition Examination Survey (NHANES) show that 68.7 \% of U.S. adults age 20 years and older were overweight or obese.\textsuperscript{109} Although the obesity prevalence has not changed recently at the national level or among women and girls, obesity prevalence among men and boys has increased significantly in the past 10 years. In addition to gender differences, there are also differences by race, ethnicity, income and geography, among other factors. Minority and low-SES groups are disproportionately affected by obesity across all age groups and
Southeastern states generally have higher prevalence rates of obesity when compared to states on the west and northeast coasts and those in the Midwest.\textsuperscript{110} The prevalence of obesity and severe obesity (BMI $\geq$40) are projected to increase by 33% and 130%, respectively, by 2030, which means that 51% of the population will be obese and 11% will be severely obese.\textsuperscript{111}

2.5.2.2.1 Mainland Puerto Ricans

An analysis of data from the HCHS/SOL found that, among men, the overall obesity prevalence was 37%, and Puerto Rican men experienced the highest obesity prevalence (41%). Among women, the overall obesity prevalence was 43%, and Puerto Rican women experienced the highest obesity prevalence (51%).\textsuperscript{63}

2.5.2.2.2 Residents of Appalachia

According to national 2007 BRFSS data, the areas of the US with the highest prevalence of obesity ($\geq$30.9%) include West Virginia and Appalachian counties in Tennessee and Kentucky.\textsuperscript{112} Over 80% of counties in Kentucky, Tennessee, and West Virginia and 77% of counties in Alabama, Georgia, Louisiana, Mississippi, and South Carolina, many of which lie within the Appalachian region, experienced an obesity prevalence that exceeded 60% of all U.S. counties.\textsuperscript{112} An analysis of 2001 state BRFSS data (the most recent data available) found that the average county-level percentage of obese adults age 18 and older in the Appalachian region was 30.8% (SD: 5.4).\textsuperscript{100} According to 2004-2007 state BRFSS data from Kentucky, Ohio, Pennsylvania and Virginia, the obesity prevalence in the Appalachian portion of the state ranged from 27% (Pennsylvania) and 34.7% (Kentucky) among men and from 26.0% (Pennsylvania,
Virginia) and 31.7% (Kentucky) among women.\textsuperscript{101} Except for Pennsylvania, the obesity prevalence in the Appalachian region exceeded that in the non-Appalachian region.

2.5.2.3 Physical activity

The AHA and ACS are in agreement in terms of physical activity recommendations. Both organizations recommend adults participate in at least 75 minutes of vigorous intensity physical activity or 150 minutes of moderate intensity physical activity or an equivalent combination on a weekly basis.\textsuperscript{5, 86} Studies have found a dose-response association between participation in physical activity and reduced risk of developing CHD and dying from CHD.\textsuperscript{38, 87, 113-116} There is evidence that participation in 150 minutes of moderate physical activity each week (or equivalent) can reduce the risk of ischemic heart disease by roughly 30%.\textsuperscript{87} In terms of cancer, there is strong evidence of a dose-response association between physical activity and reduced risk of colon and pre- and postmenopausal breast cancer.\textsuperscript{117} Evidence to support a causal relationship between physical activity and low-grade inflammation is insufficient, but the relationship between suppression of systemic low-grade inflammation and regular participation in physical activity is supported by the literature.\textsuperscript{118} Approximately 6% of global CVD deaths can be attributed to physical inactivity (6%).\textsuperscript{98} In 2010, the adjusted estimated population attributable fractions of CVD deaths due to insufficient physical activity was 11.9% in the U.S. (95% CI: 1.3–22.3%).\textsuperscript{5}

2.5.2.3.1 Global burden

Each year, insufficient physical activity causes approximately 3.2 million deaths and 32.1 million DALYs (2.1% of global total).\textsuperscript{98} In 2008, among adults age 15 or older
across the world, 31.3% were insufficiently physically active, but women (34.4%) were found to be less active than men (28.2%).\textsuperscript{38} The prevalence of insufficient physical activity in high-income countries among men (41%) and women (48%) is more than double that of low-income countries (18% of men and 21% of women).\textsuperscript{38}

2.5.2.3.2 U.S. burden

According to 2011 NHIS data, 32% of adults age 18 or older in the U.S. do not engage in daily sessions of light/moderate or vigorous physical activity at least 10 minutes in length but the prevalence varies by gender, age and race and ethnicity.\textsuperscript{61} Age-adjusted prevalence of physical inactivity was higher among women (33.2%) than men (29.9%), and the prevalence of physical inactivity was found to increase from 26.1% among adults age 18-44, 33.4% among adults age 45-64, 40.0% among adults age 65-74 and 52.4% among adults 75 years of age or older ≥75. Furthermore, age-adjusted prevalence of physical inactivity was higher among non-Hispanic black adults (41.1%) and Hispanic adults (42.2%) than among NHW adults (27.7%).

2.5.2.3.2.1 Mainland Puerto Ricans

Per 2000-2003 NHIS data, the percentage of adults 18 years and older participating in no leisure-time physical activity was 35.8% among NHW and 52.4% among Puerto Ricans. Although other Hispanic subpopulations like Cubans (66.3%) had higher rates of no leisure-time physical activity, differences between the physical activity patters of NHW and Puerto Rican adults nevertheless contribute to differences in health outcomes between these groups.\textsuperscript{119}
2.5.2.3.2 Residents of Appalachia

An analysis of data from the 2008 BRFSS indicate that the prevalence of leisure-time physical inactivity was highest in select Southern and Appalachian counties. In fact, four of the six states in which 70% of counties had a physical inactivity prevalence greater than or equal to 29.2% were within the Appalachian region: Alabama, Kentucky, Mississippi and Tennessee. An analysis of 2009 state BRFSS data found that the average county-level total minutes of exercise per week by adults age 18 and older in the Appalachian region was 368.8 (SD: 157.3). State BRFSS data from 2003 to 2007 indicate that the prevalence of no physical activity in the past month among men and women in the Appalachian regions of Kentucky, Ohio, Pennsylvania and Virginia were lowest in Pennsylvania (24% and 29.0%, respectively) and highest in Kentucky (36.8% and 41.1%, respectively). In all of the states examined, the prevalence of no physical activity in the past month among men and women were higher in the Appalachian portion of the state than in the non-Appalachian portion.

2.5.2.4 Diet, including alcohol intake

The diet component of the AHA LS7 metric is actually a composite score that includes five components worth one point each: ≥4.5 cups per day of fruits/vegetables, ≥2 servings of fish per week (3.5-oz servings), <1500 mg/d of sodium, ≤450 kcal (36 oz) per week of sweets/sugar-sweetened beverages and ≥3 servings per day of whole grains (1.1 g of fiber in 10 g of carbohydrate; 1-oz equivalent servings). In terms of LS7, a healthy diet score of 4 or 5 is considered ideal. The ACS recommendations emphasize plant based foods and components of a health promoting diet (e.g. limited processed and
red meat consumption, fruit and vegetable consumption and whole grain consumption), but they are less precise than the AHA recommendations. Fruit and vegetable consumption is the only ACS recommendation that includes an amount, but the ACS distinguishes between recommendations for cancer risk reduction and overall. For cancer risk reduction, at least 2.5 cups of vegetables and fruits are recommended per day whereas for overall health, at least 2 cups of vegetables and 1.5 cups of fruit are recommended per day.\(^{86}\) Alcohol intake was not included in the AHA’s healthy diet score, whereas ACS recommends that current drinkers limit consumption to no more than 1 drink per day (if female) or 2 drinks per day (if male).\(^{5,86}\)

The relationship between diet and CVD and cancer is challenging to summarize for a number of reasons, including the lack of consistency in dietary recommendations and methods used to measure diet, as well as a shift in research focus from specific nutrients to food groups and dietary patterns.\(^{121-123}\) In terms of food groups, research has consistently found fruits and vegetables, fish and whole grains to have a strong protective association with CHD risk.\(^{121}\) High sodium consumption has been linked to greater risk of CVD, particularly stroke, whereas the consumption of sugar sweetened beverages (SSB) has been linked to increased risk of stroke among women, but not men.\(^{124,125}\) The relationship between SSBs and ischemic heart disease is inconclusive, raising questions of reverse causation, a form of bias that is a common concern in nutritional epidemiology, since those with pre-existing conditions may have altered their SSB consumption because of their health status.\(^{124}\) Among other cancer sites, adequate consumption of fruit and vegetables has also been found to reduce the risk of cancers of
the stomach, colon and rectum, and breast.\textsuperscript{52, 126, 127} Consumption of whole grains is inversely associated with risk of cancer-related death, whereas the evidence regarding fish consumption and cancer mortality is inconsistent, whether looking at all sites combined or site specific cancers.\textsuperscript{128-131} The relationships between sodium and SSBs and cancer risk tend to be site specific. For example, high sodium intake is associated with increased risk with gastric cancers specifically, whereas SSB consumption was shown to increase risk of non-Hodgkin Lymphoma among men, but not among women.\textsuperscript{124, 132}

High alcohol consumption, which is sometimes treated as a component of diet or a separate form of risk, increases risk of acute myocardial infarction and cancers of the oral cavity and pharynx, larynx, esophagus, liver, colon and rectum, and female breast cancer, among other adverse health outcomes.\textsuperscript{133} The literature has U- or J-shaped relationships between alcohol intake and CHD and certain forms of cancer. This suggests that unlike no or heavy alcohol consumption, which are associated with increased CVD and cancer risk, moderate alcohol consumption is associated with decreased CVD and cancer risk.\textsuperscript{121, 122} Research on the relationship between diet and inflammation has shifted from a narrow focus on specific dietary elements (e.g. glycemic index and load, fiber, fatty acid composition, magnesium, carotenoids, and flavonoids.) to a broader focus of dietary patterns.\textsuperscript{134, 135} For example, compared to typical North American and Northern European dietary patterns, observational and interventional studies have found traditional Mediterranean dietary patterns, which include an abundance of fruits, vegetables, legumes, grains and fish, demonstrate superior anti-inflammatory effects.\textsuperscript{136}
2.5.2.4.1 Global burden

Globally, there are 1.7 million deaths and 16 million DALYs attributable to low fruit and vegetable consumption each year.\textsuperscript{98} In 2010, 4.9 million global deaths (95% CI: 3.8–5.9 million) and 4.2% of global DALYs (95% CI: 3.3–5.0) were attributable to diets low in fruit consumption.\textsuperscript{137} However, there were differential impacts by gender such that the fraction of disease burden attributable to diet low in fruits among men was 1.5 times that among women.\textsuperscript{137} In terms of income, a study using 2002-2003 World Health Survey data found that 78% of participants from predominately low- and middle-income countries failed to meet recommended daily servings of fruits and vegetables, a finding that is similar to research conducted in higher income nations.\textsuperscript{138-140}

Each year, alcohol causes roughly 2.3 million deaths (3.8% of all global deaths) and an estimated 4.5% of global DALYs, the bulk of which is related to cancer, CVD and cirrhosis of the liver.\textsuperscript{68} Adult per capita consumption of alcohol is highest in high-income countries, but is not far behind in populous upper-middle-income countries.\textsuperscript{38} However, in 2004, alcohol was the leading cause of DALYs in middle-income countries, and the second leading in high-income countries, but it was the eighth leading cause in low-income countries.\textsuperscript{98} In addition to differences by country economic status, there are also gender differences at the global scale. In 2010, alcohol use accounted for 5.4% of the worldwide disease burden among men but only 2.0% among women.\textsuperscript{137}

2.5.2.4.2 U.S. burden

In the U.S., suboptimal dietary habits surpassed tobacco as the country’s leading cause of mortality and DALYs between 1990 and 2010.\textsuperscript{13} In 2010, there were 678,000
deaths attributable to suboptimal diet, and the adjusted estimated population attributable fraction of CVD deaths due to poor diet was 13.2% (95% CI: 3.5–29.2%). A study that used 2005 Healthy Eating Index scores to describe the diet of participants in the 2003-2004 NHANES found that the diet quality of the study population as a whole was suboptimal, but there were still meaningful differences in diet quality by age, sex, race/ethnicity and education level.

Estimates of alcohol-attributable fractions of cancer mortality in the U.S. indicate that cancers of the oral cavity and pharynx, larynx, and esophagus accounted for between 53-71% of alcohol attributable deaths among men whereas breast cancer accounted for 56-66% of alcohol-attributable deaths among women. Cancer risk was found to increase with higher alcohol consumption levels, but the greatest percentage of total alcohol related cancer deaths occurred among those who consumed low level of alcohol, which represents the majority of drinkers. In the U.S., the findings of national surveys demonstrate differences in drinking, alcohol use disorders, alcohol problems, and treatment use between racial and ethnic groups, as well as within groups on the basis of gender, age and other sociodemographic characteristics.

**2.5.2.4.2.1 Mainland Puerto Ricans**

An analysis of data from the HCHS/SOL found that total fruit intake was highest among Dominicans and South Americans (2.4 servings/day and 2.3 servings/day, respectively) and lowest among Puerto Ricans (1.4 servings/day). Total vegetable intake was highest among Cubans (3.8 servings/day) and lowest among Dominicans and Puerto Ricans (2.5 servings/day). Sugar-sweetened beverage consumption was lowest.
among Dominicans (1.1 servings/day) and highest among Puerto Ricans (1.8 servings/day). Across the Hispanic/Latino subgroups, mean alcohol consumption was 0.3 servings/day, and alcohol intake was highest among Cuban and Mexicans (0.4 servings/day) and lowest among Dominicans, Central Americans and Puerto Ricans (0.2 servings/day).  

2.5.2.4.2.2 Residents of Appalachia

In terms of fruit and vegetable consumption, an analysis of national BRFSS data from 2000 to 2006 found the prevalence of fruit and vegetable consumption in the Mississippi Delta, Appalachian Mountains, and Great Plains was lower than the prevalence on the West Coast and in the Northeast and parts of the South. State BRFSS data from 2002 to 2007 indicate that the prevalence of inadequate fruit and vegetable consumption among Appalachian regions of Kentucky, Ohio, Pennsylvania and Virginia among men were lowest in Pennsylvania (52%) and highest in Ohio (88.9%). Among women, the prevalence of inadequate fruit and vegetable consumption was lowest in Pennsylvania (36.0%) and highest in Kentucky (80.2%). In all of the states examined, the prevalence of inadequate fruit and vegetable consumption were higher in the Appalachian portion of the state than in the non-Appalachian portion. In terms of alcohol consumption, an analysis of 2009 state BRFSS data found that the average county-level percentage of adults age 18 and older in the Appalachian region who consumed alcohol was 30.7% (standard deviation: 16.6).
2.5.3 Risk factors

In addition to the four health behaviors associated with CVD and cancer risk described above, for which there are both AHA and ACS recommendations, the AHA LS7 metric includes three additional health factors traditionally associated with CVD that are gaining increasing attention due to their suggestive, yet less delineated relationship with cancer outcomes: blood pressure, total cholesterol and plasma glucose. Given subsequent differences in the state of the literature regarding these factors and CVD and cancer, a brief overview of these factors are presented at the global and national scale, and the most recent data available on Mainland Puerto Ricans and residents of Appalachia are presented below.

When examining the relationship between these health factors and cancer, it is important to be aware of the epidemiological issue of competing risks. In brief, among other diseases, CVD and cancer compete to be the cause of death for an individual. Therefore, individuals who take drugs to treat health factors associated with CVD like blood pressure, total cholesterol and plasma glucose may be less likely to die from CVD and more likely to die from a competing risk, such as cancer. The concept of competing risks must be kept in mind when reviewing the literature, since these health factors have the potential to impact the risk of select cancers independently, or as a result of treatment.

2.5.3.1 Blood pressure

In terms of LS7, untreated blood pressure for which systolic blood pressure is less than 120 millimeters of mercury (mm Hg) and diastolic blood pressure is less than 80
mm Hg is considered ideal. Potential mechanisms linking inflammation, elevated blood pressure and CVD suggested by the literature include an imbalance between agents that narrow a blood vessel opening (vasoconstrictors) and those that widen them (vasodilators), amplified formation of blood clots and activation of platelets, or through the direct effect of inflammatory mediators. Increasing blood pressure is positively correlated with CHD, stroke and peripheral vascular disease risk. Hypertension (high blood pressure) is also correlated with increased risk of select cancers such as endometrial cancer and prostate cancer. Furthermore, anticancer medications like sorafenib, sunitinib, and pazopanib are positively correlated with risk of high blood pressure, whereas there are mixed results regarding whether antihypertensive drug use and risk of certain cancers, such as breast cancer, are positively correlated.

Annually, 16.5% of all global deaths (9.4 million) can be attributed to hypertension. Hypertension is an underlying cause of 51% of stroke deaths and 45% CHD deaths across the world. In 2010, the adjusted estimated population attributable fractions of CVD deaths due to hypertension was 40.6% in the U.S. (95% CI: 24.5-54.6%). Globally, the overall prevalence of hypertension in adults aged 25 and over was around 40% in 2008, but the prevalence is higher among low-, lower-middle- and upper-middle income countries (40% for men in women) than among high-income countries (35% for men and women). In the U.S., one out of every three adults has hypertension (roughly 78 million). However, there are differences by a number of sociodemographic characteristics, the most striking of which is found among Blacks, who experience the highest prevalence of hypertension in the world (44%).
An analysis of data from the HCHS/SOL found that among men, the overall prevalence of hypertension was 25% and 27% among Puerto Rican men, specifically. Among women, the overall hypertension prevalence was 24%, and Puerto Rican women experienced the highest prevalence (29%). An analysis of 2001 and 2009 state BRFSS data found that the average percentage of adults age 18 and older in the Appalachian region with hypertension was 29.7% (SD: 4.6).

### 2.5.3.2 Total cholesterol

In terms of LS7, untreated total cholesterol that is less than 200 milligrams per deciliter (mg/dL) is considered ideal. One third of the global ischemic heart disease burden is attributable to high total cholesterol. Statins, a form of lipid lowering drugs prescribed to individuals at high risk of developing CVD, target low-density lipoprotein, a component of total cholesterol that is considered “bad” cholesterol. Research suggests that statins also demonstrate anti-inflammatory effects, independent of their effect on lipids. Low levels of serum lipids have been found to moderately reduce risk of obesity-related cancers, whereas research on the association between statin use and risk of cancers is not consistent across cancers of the breast, esophagus and colon and rectum.

At a global scale, high total cholesterol is estimated to cause 2.6 million deaths (4.5% of total deaths) and 29.7 million DALYs (2% of total DALYs) each year. In 2008, the prevalence of high total cholesterol among adults was 9.7% overall, 8.5% for males and 10.7% for females. The prevalence of high cholesterol among adults is higher in high income countries (>50%) than in low-income countries (25%).
U.S., nearly 13.8% of adults 20 years of age and older (nearly 32 million) had high total cholesterol in 2010.\textsuperscript{1} Although total blood cholesterol levels in the U.S. adult population declined between 1960 and 1994, total blood cholesterol levels have changed since, despite progress in cholesterol screening and awareness.\textsuperscript{158-160} A potential explanation for this plateau is observed disparities in cholesterol screening on the basis of age, sex, and race/ethnicity. From 1999-2002, increasing age was associated with the likelihood of undergoing a blood cholesterol screening within the previous 5 years. NHWs were more likely to have undergone a blood cholesterol screening within the previous 5 years than non-Hispanic blacks (adjusted OR (AOR): 0.70, 95% CI: 0.57-0.84) and Mexican Americans (AOR: 0.43, 95% CI: 0.35-0.53) as were women compared to men (AOR: 1.20, 95% CI:1.03-1.39).\textsuperscript{161}

An analysis of data from the HCHS/SOL found that, among men, the overall prevalence of hypercholesterolemia was 52%, but Dominican and Puerto Rican men experienced the lowest prevalence (48%). Among women, the overall prevalence of hypercholesterolemia was 37%, but Puerto Rican women experienced the highest prevalence (41%).\textsuperscript{63} Unlike other behaviors and factors, the prevalence of hypercholesterolemia in Appalachia has not been elucidated. However, a health needs assessment of seven Ohio Appalachia counties found the prevalence of hypercholesterolemia to be 37.6%, and the prevalence of this risk factor exceeding that of the national average may be common among other Appalachian counties.\textsuperscript{162} Furthermore, between 1995 and 2007, gains in hypercholesterolemia greater than the national average
(gain of 9.5%) were experienced by nine of the thirteen states within the Appalachian region.\textsuperscript{36}

\subsection*{2.5.3.3 Plasma glucose}

In terms of LS7, untreated plasma glucose that is less than 100 milligrams per deciliter (mg/dL) is considered ideal.\textsuperscript{5} In addition to being a risk factor for diabetes (having a fasting plasma glucose value $\geq 126$ mg/dl), impaired glucose tolerance and impaired fasting glucose are also risk factors for CVD.\textsuperscript{87} Inflammatory responses are triggered by glycemic fluctuations and chronic high blood sugar. Furthermore, the molecular pathways underlying high blood sugar contribute to the dysfunction of the inner lining of blood vessels (endothelium), the first step of atherogenesis.\textsuperscript{163} Diabetes has also been found to increase risk of specific cancers, including pancreatic cancer, kidney cancer and hepatocellular carcinoma, the most common form of liver cancer.\textsuperscript{164-166} Additionally, diabetes is associated with increased risk of cancer-specific mortality, such as hepatocellular carcinoma and breast cancer.\textsuperscript{166, 167} Observational studies have found diabetic patients using metformin, an antidiabetic drug, to have significantly lower risk of cancer incidence and mortality than diabetic patients not using metformin.\textsuperscript{168}

In 2010, the adjusted estimated population attributable fractions of CVD deaths due to abnormal blood glucose levels was 8.8\% in the U.S. (95\% CI: 2.1-15.4\%) \textsuperscript{5} In 2008, there were 1.3 million diabetes-related deaths globally, but this is likely an underestimation of the magnitude of glucose tolerance abnormalities since categories of “impaired fasting” and “impaired glucose tolerance” were excluded. Diabetes prevalence is similar between low-income countries (8\% for men and women) and upper-middle-
income countries (10% for men and women). In 2010, an estimated 19.7 million adults in the U.S. (8.3%) had diagnosed diabetes. However, diabetes is not distributed evenly such that African Americans, Hispanics/Latinos and other ethnic minorities are disproportionately burdened.¹

An analysis of data from the HCHS/SOL found 17% of men and women had diabetes. However, the prevalence was highest among Mexican men and women and Puerto Rican women (19%).⁶³ An analysis of 2009 state BRFSS data found that the average county-level percentage of adults age 18 and older in the Appalachian region with diabetes was 15.5% (SD: 6.5).¹⁶⁹

2.5.4 Prevention recommendation summary

There is substantial overlap between the CVD prevention recommendations of the AHA and the cancer prevention recommendations of the ACS in terms of their content, as well as their association with CVD and cancer outcomes. Among 115,306 Women’s Health Initiative (WHI) observational study participants free of CVD at baseline, those with 0 to 1 ideal LS7 factors and behaviors had a 6.83 times higher instantaneous risk of an incident CVD event than those with 6 to 7 ideal LS7 factors and behaviors over 18 to 22 years of follow-up (95% CI: 5.83-8.00). Among 129,149 WHI participants free of cancer at baseline, those with 0 to 1 ideal LS7 factors and behaviors were 52% more likely to have a non-melanomatous cancer event than those with 6 to 7 ideal LS7 factors and behaviors (95% CI: 1.35-1.72).⁷ Similarly, among 13,253 adults participating in the Atherosclerosis Risk in Communities (ARIC) study, a graded, inverse association was observed between ideal LS7 metrics at baseline and cancer incidence later in life (median
follow up duration: 18.7 years). Although only 2.7% of the study population had 6 or 7 ideal cardiovascular health metrics, their risk of incident cancer was 51% lower than among those with 0 ideal metrics. When smoking was removed from the count of ideal LS7 metrics, those with 5 to 6 ideal LS7 metrics had 25% lower risk than those with 0 ideal metrics.\textsuperscript{170} In terms of ACS recommendations, among 111,966 nonsmoking adults participating in the Cancer Prevention Study-II Nutrition Cohort in the early 1990s, adherence to obesity, diet, physical activity, and alcohol consumption related-recommendations decreased risk of cancer, CVD, and all-cause mortality.\textsuperscript{8}

Despite their similarities, however, the AHA LS7 metric is more clearly defined in terms of risk behaviors than ACS cancer prevention recommendations. Furthermore, the AHA LS7 metric includes blood pressure, total cholesterol and plasma glucose. Although the relationship between cancer and these health factors is not as clear as the relationship between cancer and smoking, BMI, physical activity and diet, blood pressure, cholesterol and plasma glucose have the potential to impact the risk of select cancers independently, or as a result of treatment for these health factors, through similar inflammatory pathways. The differentiation between ideal, intermediate and poor levels of each health behavior and factor made by the AHA LS7 metric is better suited to quantify gradients in disease risk and monitor progress made toward public health benchmarks compared to the ACS recommendations. As a result, the AHA LS7 metric was used for the purposes of this paper to determine the prevalence of overlapping CVD and cancer risk behaviors and factors.
2.6 Social determinants of overlapping risk behaviors and factors

Whereas the relationships between overlapping health behaviors and factors and CVD and cancer outcomes are well established, understanding of the relationship between social determinants of health, “the conditions in the environments in which people are born, live, learn, work, play, worship, and age”, and these behaviors and factors is more limited. The term “place” is often used to capture the social, economic and physical environmental conditions, from patterns of social engagement to resources like the availability of healthy foods, that impact health through psychosocial, behavioral and biological pathways.

Of studies that have used the LS7 metric to quantify the distribution of overlapping CVD and cancer risk factors and behaviors among community samples from populations facing CVD and cancer disparities, only a recent analysis of MESA data applied a multi-level perspective. This study found neighborhood environment characteristics such as food stores more likely to carry fresh produce, physical activity resources, walking/physical activity environment and neighborhood socioeconomic status (SES) were associated with higher odds of having an ideal overall LS7 score. As suggested by the findings, improved understanding of how populations disproportionately burdened by CVD and cancer experience place and how place impacts disparate health CVD and cancer outcomes through LS7 is needed.

2.6.1 Conceptual model

The model for the analysis of population health and health disparities (Figure A.4) developed by investigators within the Centers for Population Health and Health
Disparities (CPHHD) served as the conceptual basis for this study.\textsuperscript{6} Funded by the National Institutes of Health (NIH), CPHHD is the first federal initiative to address disparities and inequities in heart disease and cancer incidence and outcomes through multi-level transdisciplinary research.\textsuperscript{173} This multi-level conceptual model builds upon the strength of established social determinants of health-related theories and frameworks.\textsuperscript{6, 37, 171, 174-179} For example, Marmot and Wilkinson’s social determinants of health (SDH) framework encompasses a variety of factors that may impact health, beginning in early life and through accumulation across the life course: social gradients, vulnerability, support and patterning of individual behaviors; poverty and economic disadvantage, unemployment and job insecurity; work environments; transportation; food accessibility; neighborhood, housing and sexual behavior. However, unlike the established theories and frameworks which serve as its basis, the CPHHD model allows social health determinants to influence health disparities independently and directly or through interactions across three levels: distal (i.e., population), intermediate (i.e., area) and proximal (i.e., individual).

\textbf{2.6.1.1 Distal health determinants}

Since health disparities exist at the population level, distal factors are considered fundamental causes and determinants of health inequities, not health differences.\textsuperscript{6, 37} Distal health determinants include population social conditions and policies and the institutional context in which they exist. Social conditions and policies include poverty, SES, public policy, culture, norms, discrimination and prejudice, whereas the institutional
context includes family, organized religion, and the healthcare, economic, legal, media and political system.\(^6\)

2.6.1.2 Intermediate health determinants

Intermediate health determinants include the social and physical contexts and the social relationships connecting distal health determinants to the level of the individual. The social context includes collective efficacy, social capital, social cohesion, poverty level, racial/ethnic integration, and social/economic gradients. Social relationships include social networks, social support, social isolation, social influence, social engagement, religious participation, civic engagement and employment. The physical environment includes building quality, pollution, business, transit access, orderliness, graffiti, cleanliness, sidewalks, open space, parks and neighborhood stability.\(^6\)

2.6.1.3 Proximal health determinants

Proximal health determinants occur at the individual level and have an independent effect from those at the intermediate and distal levels. Proximal health determinants include individual demographics, individual risk behaviors, biological responses and biologic/genetic pathways. Individual demographics include characteristics like age, SES, health status, education, race/ethnicity and acculturation that can influence an individual’s ability to respond to challenges in their environment. Individual cultural beliefs and risk behaviors like tobacco use, alcohol use, diet, sexual behavior, loneliness and trust in the health care system also influence health disparities, as do biological responses (e.g. depression, stress, previous illness, chronic lung disease and alcoholism)
and biologic/genetic pathways (e.g. allostatic load, biologic processes, genetic ancestry and genetic mechanisms).  

2.6.2 Strengths and Limitations of Using a Multi-Level Approach

Ecological studies, in which groups are the unit of analysis, are subject to the ecological fallacy. This is a form of bias that arises when inferences regarding the variability between individuals are made using group-level data since group- and individual-level measures do not necessarily capture the same construct. Ecological studies are also subject to the sociologistic fallacy since ecological studies ignore the role that individual-level measures play in explaining variability at the group-level and may fail to account for individual-level confounders or effect measure modification (EMM). Confounding occurs when a third variable that is independently associated with the exposure and disease of interest, but not on the causal pathway between exposure and disease, distorts the exposure-disease relationship. EMM occurs when the strength of the association between exposure and disease differs by the level of a third variable. Individual-level studies, in which individuals are the unit of analysis, are subject to the counterpart of the ecological fallacy: the automistic fallacy. This fallacy arises when inferences regarding the variability between groups are made using individual-level data since individual-level studies may lack group-level confounders and/or EMMs.

In theory, unlike ecological and individual-level studies, a multi-level approach allows for the estimation of group- and individual-level effects so that inferences regarding the causes of inter-group and inter-individual variation (and the extent to which
they are explained by individual- and/or group-level variables) can be drawn by controlling to multi-level confounders and allowing for multi-level EMMs. Furthermore, as opposed to ecological and individual-level studies, a multi-level approach accounts for the nesting of individuals within groups, which violates the independence of observation assumption underlying traditional analysis methods.

In practice, however, the utility of a multi-level approach is often hampered by current understanding of relationships between constructs measured at different levels as well as data availability. Multi-level data assumed to capture the same construct may not be easily accessible or readily available, commonly resulting in the use of data from one level as a proxy for data from another. Data availability issues may therefore lead to forms of bias common among ecological and individual-level studies, such as the ecological and sociologistic fallacies, despite employment of a multi-level approach.

2.6.3 Conceptual model in the context of Life’s Simple 7 (LS7)

The LS7 metric has been evaluated in distinct research samples, to date without much consideration of the social determinants of health. LS7 metric components are proximal determinants of CVD and cancer health disparities. More specifically, smoking, BMI, physical activity and diet are considered to be risk behaviors and blood pressure, total cholesterol and plasma glucose are considered to be biological responses. The conceptual model can be used to consider the social determinants of health contributing to CVD and cancer health disparities, including LS7 risk behaviors and factors. It can also be adapted to consider the social determinants of health that contribute to CVD and
cancer health disparities through their impact on the modifiable risk behaviors and biological responses represented by the LS7 component.

2.6.4 Conceptual model in the context of diet

Between 1990 and 2010, each LS7 component contributed to U.S. mortality and morbidity, but diet was found to be the leading cause of the overall disease burden in the U.S.\textsuperscript{13} As a result, this paper specifically explored select social determinants of health that impact diet in the U.S. Given the breadth of the diet literature and range of multi-level social determinants of diet available for study, this paper explored a limited number of intermediate- and proximal-level social determinants of diet related to food access that are particularly relevant to mainland Puerto Ricans and residents of Appalachia yet understudied in these populations.\textsuperscript{15} There are three main forms of food access: geographic, informational and economic. Geographic access includes facilitators and barriers pertaining to distance that influence food purchasing and consumption. Informational access includes educational, cultural and social facilitators and barriers to food and economic access includes financial facilitators and barriers that impact food access.\textsuperscript{15} Of the three forms of food access, geographic food access has received increased empirical attention due to burgeoning use of GIS and availability of federal, state and local incentives, such as financing, to improve access to healthy food in underserved areas.\textsuperscript{16, 17} More specifically, this paper focused on the intricacies of using intermediate-level geographic food access as a proxy of proximal-level geographic food access when studying the relationship between geographic food access and individual-
level LS7 healthy diet score among mainland Puerto Ricans and residents of Appalachia using a multi-level approach.

2.6.4.1 Intermediate-level geographic food access

A systematic review of food access articles published from 1990-2007 defined the food environment as community-level organizations and food resources falling into four categories: the food store environment (e.g., grocery stores, supermarkets, specialty food stores, farmers’ markets, food pantries); the restaurant food environment (e.g., fast food restaurants, full-service restaurants); the school food environment (e.g., cafeterias, vending machines, snack shops in daycares, schools, and/or colleges) and the worksite food environment (e.g., cafeterias, vending, snack shops). Of these categories, the food store and restaurant food environments have received the most empirical attention. Facets of intermediate-level geographic food access like geographic diversity, proximity, and/or variety of food stores and restaurants are increasingly being assessed using GIS. Geographic diversity refers to the density of food stores or restaurants within a defined spatial region, shopping catchment, buffer area, or population. Coverage methods to determine the number of food stores or restaurants include the count within a given radius, count per population, or count per square linear area (e.g. square mile). Geographic proximity refers to the nearest distance to food stores or restaurants, often from place of residence, using methods such as shortest path, a weighted calculation of the shortest distance between two vertices or nodes; Euclidean distance, calculated using the Pythagorean formula and also referred to “as the crow flies”; or Manhattan distance, calculated by city block or street. Geographic variety refers to the assessment of the
availability, pricing and quality of different types of foods stores or restaurants within a geographic region. A systematic review of the nearly 40 studies on the relationship between intermediate-level geographic food access and diet found that over two thirds of the studies reviewed used GIS-based measures. In terms of geographic diversity, GIS-based methods were the most commonly applied to explore the relationship between store density and diet and 13 of 20 GIS-based studies found a significant association. However, these studies counted the number of food stores and restaurants using a range of buffer distances, geographic units of analysis and definitions of population density. In terms of geographic proximity, there was a lack of consistency such that over half of the 13 studies that examined the relationship between food store distance and diet found null associations whereas the remaining studies demonstrated associations in mixed directions. Some studies assessed geographic variety using an index of food prices in participants’ areas of residence, participants’ perceptions of produce affordability or store auditors’ food prices accounts and found that lower area food prices were associated with at least one measure of dietary health in every study, whereas measures of perceived produce affordability and food audit affordability were inconsistently associated with diet.

2.6.4.1.1 Methodological concerns

2.6.4.1.1.1 Modifiable areal unit problem

As suggested by the systematic reviews, measures of geographic diversity, proximity and variety are impacted by the modifiable areal unit problem (MAUP), a form
of statistical bias that arises because there is no natural geographic unit of analysis. Consequently, the results of analyses using count data aggregated to arbitrary, modifiable units of geography are unit-dependent.\textsuperscript{183} The MAUP is related to edge effects, a form of bias that occurs when the study boundary affects the estimation of a measure and inferences drawn from the measure, which has been shown to greatly impact food accessibility measures.\textsuperscript{184, 185}

In addition to buffer distances and county boundaries, common geographic units of analysis used to capture intermediate-level geographic food access include U.S. census geographic entities like CTs, small, statistical county (or equivalent entity) subdivisions composed of somewhat socially and economically homogeneous populations bound by visible features (e.g. streets, roads, streams, railroad tracks) ranging in size from 1,200 to 8,000 people, and census block groups (CBGs), statistical areas bounded by visible features or non-visible boundaries (e.g. property lines, city, township, school district, and county limits) that nest within all other census-based geographic entities.\textsuperscript{186}

The geographic unit of analysis has important implications for food access research conducted in urban, suburban, rural and remote areas. For example, while urban census blocks (CBs) may be relatively consistent in terms of population size and shape, suburban and rural area CBs may be large and irregular and remote area CBs may cover hundreds of square miles.\textsuperscript{186} Although there may be a disconnect between neighborhoods defined by census-based geographic entities and those defined by residents across settings, this discrepancy may be more pronounced in less densely populated areas.\textsuperscript{187} Therefore, the intermediate-level food environment captured by
census-based geographic entities may be less relevant to suburban, rural and remote area residents in terms of geographic diversity, proximity and variety.

2.6.4.1.1.2 Absolute and relative approaches

Another methodological concern is the conceptualization and measurement of food access as an absolute versus a relative phenomenon, which could have important implications for the study of intermediate-level geographic food access and diet. Absolute food access measures determine access in relation to one type of store (i.e. large grocery store or supermarket) and are the most prominent across the literature.\textsuperscript{16, 182} Though less common, relative food access measures also take food retail locations like fast food establishments and convenience stores into account.\textsuperscript{188} Each approach is summarized below, including their respective strengths and limitations.

2.6.4.1.1.2.1 Absolute measure

An example of an absolute food access measure is the U.S. Department of Agriculture Economic Research Service (herein referred to as USDA) CT-level binary food desert measure, which served as the basis of the \textit{USDA Food Desert Locator} created in 2009.\textsuperscript{189} The USDA defines a food desert as CTs that are low income in which a substantial number or percentage of residents have low access to a supermarket or large grocery. CTs are considered low-income if the poverty rate is at least 20% or the median family income is 80% or lower than that of the state or metro-area in which it is located. Low access is defined differently for urban and rural CTs. A CT is considered urban if its centroid is within an urban area, defined by the U.S. Census as having 50,000 or more people, or within an urban cluster of at least 2,500 people and less than 50,000 people.
CTs in areas not meeting the urban area or urban cluster of the U.S. Census definitions are considered rural.\textsuperscript{190} Urban CTs are considered low access if 500 or more people or 33\% of those residing in the CT live more than one mile from the nearest large grocery store or supermarket. Rural CTs are considered low access if 500 or more people or 33\% of those residing in the CT live more than ten miles from the nearest large grocery store or supermarket.\textsuperscript{191} Supermarkets and large grocery stores included in the USDA food desert measure must report at least $2 million in annual sales and contain all the major food departments found in a traditional supermarket (e.g. fresh meat and poultry, dairy, dry and packaged foods, frozen foods).\textsuperscript{190, 191}

Population and income data to calculate this measure were obtained from the 2000 Census of Population and Housing. The 1-km square grid data from which distance to nearest supermarket or large grocery store is calculated was obtained from the Socioeconomic Data and Applications Center. Eligible supermarket and large grocery store data were obtained from a directory based on data from stores authorized to receive Supplemental Nutrition Assistance Program (SNAP) benefits in 2006 and 2006 proprietary data from Trade Dimensions TDLinx.\textsuperscript{189}

Using 2000 Decennial U.S. Census data, 2010 CT definitions and 2006 food retailer data, 6,529 US CTs (9.6 \% of all CTs) home to 13.6 million residents were classified by the USDA as food deserts. Of the food deserts identified, 2,204 were in rural areas and 4,175 were in urban areas. Food desert CTs tended to have smaller populations, higher rates of abandoned or vacant homes, and higher proportions of residents with lower levels of education, lower incomes, and higher unemployment than
CTs that were not classified as food deserts. In both rural and densely populated urban areas, CTs with higher poverty rates were more likely to be food deserts than low-income CTs to which they were otherwise comparable. In less densely populated urban areas, CTs with higher concentrations of minority populations were more likely to be food deserts. CTs experiencing substantial minority population loss between 1990 and 2000 were less likely to be food deserts.\textsuperscript{191}

It should be noted that the USDA’s food desert designation has important policy implications since this geographic food access measure is used to determine eligibility for and prioritization of federal funding to increase access to healthy, affordable food. Since 2011, more than $140 million has been awarded to add or expand upon healthy food options in communities designated as food deserts. In addition, there are also a growing number of state and local incentives with varying eligibility requirements to develop, expand and preserve access to healthy food in underserved areas.\textsuperscript{17} However, as noted in the literature, the policy implications of the food desert metric highlight the need for this measure to be validated, particularly in urban versus rural environments.\textsuperscript{192}

The \textit{USDA Food Desert Locator} was archived in March 2013. Data derived with slightly modified methods are now available through the \textit{USDA Food Access Research Atlas}.\textsuperscript{193} The updated measure is based on 2010 Decennial U.S. Census data, 2010 food retailer data, 2010 Census tract definitions and 2006-2010 AmCS data. According to the updated measure, there were 8,959 USDA food deserts (12.3 \% of all CTs) home to over 18 million residents in 2010. The increase in food desert CTs between 2006 and 2010 was driven by the growing number of low-income CTs.\textsuperscript{194} County-level data similar to
the updated CT-level food desert metric are available through the *USDA Food Environment Atlas* (e.g. count and percentage of people in a county with low income and living more than 1 mile from a supermarket or large grocery store (if in an urban area) or more than 10 miles from a supermarket or large grocery store (if in a rural area)) and may be better suited for the study of food deserts in rural areas.\textsuperscript{195,196}

Strengths of the USDA’s absolute food access measure include the fact that they take neighborhood income into account as well as distance to the nearest supermarket or large grocery store. Additionally, the USDA’s absolute food access measures have been updated with US Census geographic boundary definitions per decennial census. The most recently available measures are based on 2010 Decennial U.S. Census data, 2010 food retailer data, 2010 Census tract definitions and 2006-2010 AmCS data.

In terms of limitations, as implied by its categorization, the USDA’s absolute food access measures only take supermarkets and large grocery stores into account, and may therefore fail to capture that variety of stores that carry grocery items from which individuals may choose to shop. Additionally, absolute food access measures reflect perceived nearest supermarket and large grocery store access, which may not equal realized access due to reasons such as the price of those stores compared to other grocery shopping alternatives, or travel outside of the unit of geography (e.g. CT) that affords grocery shopping alternatives not captured by these measures.

\textbf{2.6.4.1.2.2 Relative measure}

An example of a relative food access measure is the CT-level modified retail food environment index (mRFEI) employed by the CDC’s Division of Nutrition, Physical
Activity and Obesity in a number of reports. The mRFEI combines the concept of a “food desert” with another absolute food access concept, “food swamp”, an area plagued with an overabundance of energy-dense snack food, into a single measure. The mRFEI considers the number of healthy food retailers as well as the number of less healthy food retailers within a CT or ½ mile from the CT boundary. The mRFEI also differs in its store classification approach by utilizing North American Industry Classification System (NAICS) codes to categorize food retailers. Healthy food retailers include supermarkets and larger grocery stores (NAICS code 445110; specifically supermarkets with >= 50 annual payroll employees and grocery stores with 10–49 employees); fruit and vegetable markets (NAICS code 445230) and warehouse clubs (NAICS code 452910). Less healthy food retailers include fast food restaurants (NAICS code 722211), small grocery stores (NAICS code 445110, specifically grocery stores with three or fewer employees), and convenience stores (NAICS code 445120).

To calculate the CDC mRFEI score, supermarket, supercenter, and produce store data were obtained from the proprietary InfoUSA business database (2009), fast food retail data were obtained from the proprietary NavTeq database (2009) and convenience store data were obtained from Homeland Security Information Program database, (2008). Although not explicitly stated, 2000 Census tract definitions were used. The mRFEI score, which can range from 0 to 100, represents the percentage of healthy food retailers out of the total number of healthy and less healthy food retailers in a CT or ½ mile from the CT boundary. For example, a score of 10 means that 10 out of every
100 stores in the CT likely offer healthy foods like fruits and vegetables, low-fat dairy items, meat products, and whole grain foods. Presently, country-wide CDC mRFEI measure estimates have only been published at the state-level. According to the 2011 Children’s Food Environment State Indicator report, which presented the median score of all CTs scores in a given state, the nation median mRFEI score was 10. The District of Columbia and 32 states had median scores at or below the national median. State-specific median scores ranged from 16 in Montana and 15 in Maine to 5 in Rhode Island and 4 in the District of Columbia. Unlike the USDA food desert measures, characteristics of CTs associated with lower or higher mRFEI scores are less understood. Since the CDC mRFEI measure is well-suited for the study of food access in urban environments, additional research is needed to identify neighborhood characteristics related to the mRFEI measure. Strengths of the CDC’s relative food access measure include the fact that they take numerous food outlets into account, not just the nearest supermarket or large grocery store. Additionally, the CDC’s relative food access measure is continuous rather than binary, which is more informative than binary measures when comparing geographic food access across different geographies, including rural and urban settings. In terms of limitations, as opposed to the USDA’s absolute food access measures, the CDC’s relative food access measure has not been updated for 2010 US Census CT definitions, which may limit its relevancy given the availability of more recent data than what was used to define the mRFEI. Unlike the USDA’s absolute food access measure, the CDC’s relative food access measure only reflects geographic food access, meaning
relevant neighborhood characteristics like income are not taken into account. Additionally, like the USDA’s absolute food access measures, the CDC’s relative food access measure also reflects perceived geographic food access as opposed to realized geographic access.

### 2.6.4.1.2.3 Impact on results

The conceptual differences between the USDA food desert and CDC mRFEI measures extend beyond absolute and relative food access. For example, the USDA food desert measure was mandated by the 2008 Farm Bill to identify low-income tracts that have low access and as a result, it places an emphasis on socioeconomically disadvantaged populations.\(^\text{190, 191}\) In contrast, the CDC mRFEI, which was first used in the context of a national report on state-level fruit and vegetable consumption, does not include income or distance indicators nor does it distinguish between rural and urban CTs.\(^\text{197}\) Given the numerous conceptual differences between the two measures, food access study results likely vary between use of the USDA food desert measure and CDC mRFEI score.

The first systematic comparison of the USDA food desert and CDC mRFEI measures was recently completed and offers insight into the how the use of these measures impacts results.\(^\text{201}\) This study used data from a 2009 food environment validation study conducted in 8 South Carolina counties to replicate and compare the measures with respect to the number of CTs classified as low access via the USDA food desert definition or a mRFEI score of 0. To account for border effects, a 10-mile external buffer based on Euclidean distance was created using non-validated state licensed food...
services data and data from InfoUSA, a commercial listing of food outlets. The replicated USDA food desert and CDC mRFEI measures gave identical classifications to 120 (71%) of the 169 CTs included in the study area: 101 CTs (59.8%) were identified as not having poor food access by both measures and 19 CTs (11.2%) were identified as having poor food access by both measures. For the remaining 49 CTs (29.0%), one but not both measures identified the CTs as poor access. The consistency between the two measures was much higher in urban areas (83.5%) than in rural areas (60.0%). Sensitivity analyses found consistency improved when comparing the complete USDA food desert criteria (i.e. low access and low income) to CDC mRFEI score of 0. A possible explanation for the increase in consistency between the two measures is that the CDC mRFEI score of 0, which reflects the absence of a healthier food retail outlet within a CT, may be associated with area-level socioeconomic indicators, such as the low income component of the USDA food desert definition.

In addition to needing to replicate these findings, future research is needed to examine how USDA food desert and CDC mRFEI measures differ in relation to diet when controlling for individual- and area-level covariates, including urban and rural designation. Furthermore, moving forward, additional research is needed to determine the utility of the antiquated CDC mRFEI measure, which is based on 2000 US CT definitions and 2009 NAICS codes and for which an update is unavailable, compared the USDA food desert and related measures, which has been updated since its inception using the
current decennial US Census CT definition and recent population and food retail industry data.

2.6.4.1.2 Use as proxy for proximal-level geographic food access

Unlike the raw data on individual food retailers that serve as their basis, the USDA’s food desert and the CDC’s mRFEI measures are free and readily available for use at national and sub-national levels. However, the use of these measures, in both the absence and presence of less readily available proximal-level geographic food access data, requires clarification of whether intermediate- (i.e. area) and proximal (i.e., individual)-level geographic food access measures capture similar or different concepts. Unfortunately, this conceptual issue has not been clearly addressed in the literature.

2.6.4.1.2.1 Mainland Puerto Ricans

Given the limited disconnect between the local food environment defined by census-based geographic entity boundaries like CTs and the perceptions of urban residents, the USDA’s food desert measure and the CDC’s mRFEI may be proxies of proximal-level geographic food access in urban environments in which mainland Puerto Ricans live. However, these convenient geographic food access measures do not address the importance of informational food access in shaping the diet of ethnic minority populations. Consequently, immigrant enclaves, an intermediate-level indicator of informational food access, and acculturation, an individual-level indicator of informational food access, is now reviewed with emphasis on mainland Puerto Ricans given the potential of immigrant enclaves and acculturation to modify the intermediate geographic food access indicator-LS7 healthy diet score relationship.
2.6.4.1.2.1.1 Immigrant enclaves

When studying the relationship between geographic food access and diet among mainland Puerto Ricans, immigrant enclaves, neighborhoods with high proportions of immigrants possibly of the same ethnicity, are an important source of informational food access to consider. Immigrant enclaves have the potential to modify the relationship between intermediate-level geographic food access and diet through pathways such as social networks, resources like ethnic food stores which carry culturally appropriate foods, fewer communication barriers for non-English language speakers and insulation from discrimination induced-stress.

An analysis of MESA data examined the relationship between neighborhood immigrant composition and diet that included 1,191 Hispanic Americans recruited from New York, Los Angeles and St. Paul, of whom 117 were mainland Puerto Ricans. The study found that living in census tracts with higher proportions of immigrants born in Latin America was associated with lower high-fat food consumption, controlling for age, gender, income, education, neighborhood poverty and acculturation. A similar study of 345 Hispanic women in New York City, of whom 41 were mainland Puerto Ricans, examined the relationship between neighborhood linguistic isolation, a census term to describe households in which English is not spoken very well by any person 14 years of age or older, and diet. Adherence to a healthy dietary pattern consisting of high consumption of vegetables, legumes, potatoes, fish and other seafood was positively associated with neighborhood linguistic isolation and negatively associated with
neighborhood poverty, adjusting for individual and neighborhood sociodemographic characteristics.\textsuperscript{203}

2.6.4.1.2.1.2 Acculturation

Acculturation describes changes in attitudes, values, behaviors and cultural identity that an individual or group may experience upon entering a new and different cultural context.\textsuperscript{204} Some researchers view acculturation as a continuum ranging from the immersion in the person’s culture of origin to immersion in the dominant or host culture, while others view it as two distinct, independent dimensions: dominant culture adherence and culture of origin maintenance.\textsuperscript{204} One-dimensional measures of acculturation are derived from language acquisition and usage, frequency of participating in cultural practices, interpersonal relationships, cultural identity, family beliefs, and adherence to traditional values, or replaced by proxies like generational status, age at immigration, proportion of life spent in the U.S., years lived in the new country, place of birth, and place of education.\textsuperscript{204} Two-dimensional measures of acculturation produce independent measures for dominant culture adherence and culture of origin maintenance across different domains like language use, food and music preference and attitudes toward both cultures.\textsuperscript{204}

A systematic review of the relationship between acculturation and diet among Latinos in the U.S. found a number of findings with consistency, even though different studies employed different acculturation measures.\textsuperscript{205} For example, the lack of a relationship between acculturation and dietary fat intake or percent energy from fat were common findings, as were results demonstrating that individuals who were less
acculturated consumed more fruit, rice and beans and less sugar and sugar-sweetened beverages than those who were acculturated. Unfortunately, few studies provided sufficient information to draw any meaningful conclusions regarding whether the relationship between acculturation and diet differs by Latino subgroup, such as mainland Puerto Ricans.  

2.6.4.1.2.1.3 Immigrant enclaves and acculturation

The association between immigrant enclaves and diet may be partially explained by the individual-level acculturation of residents since immigrant enclave residents may be less acculturated than residents of neighborhoods with fewer immigrants. However, it is unclear from the literature whether immigrant enclaves are proxies for individual-level acculturation or whether immigrant enclaves and individual-level acculturation have independent effects on diet. This conceptual issue is further complicated by evidence of a synergistic effect of individual-level acculturation and immigrant enclaves on health. Research on the relationship between immigrant enclaves and BMI among an ethnically and racially diverse sample of 13,011 New York City residents found that individual-level acculturation, as measured by place of birth, did not modify the neighborhood proportion of foreign born residents-BMI relationship. However, place of birth was found to be an EMM of the relationship between the neighborhood linguistic isolation and BMI such that the association between place of birth and BMI was stronger in neighborhoods with lower levels of linguistic isolation and weaker in neighborhoods with high levels of linguistic isolation, albeit not statistically significant (p=0.27).
2.6.4.1.2.2 Residents of Appalachia

Appalachia is a distinct region in which the relationship between geographic food access and diet has received little empirical attention.\textsuperscript{207} Furthermore, it is heterogeneous in terms of urban and rural areas, making it challenging to determine what geographic unit of analysis is best suited to study the relationship between geographic food access and diet among residents of the region.\textsuperscript{31} As noted previously, the use of intermediate-level geographic food access measures as proxies for proximal-level geographic food access may be especially problematic outside of urban areas given differences between the local food environment defined by census-based geographic entity boundaries and the perception of residents.\textsuperscript{187} There is even evidence to suggest the use of county-level data may not adequately capture the local food environment of residents of non-metropolitan areas. An analysis of national 2000-2006 BRFSS data found that fruit and vegetable consumption was higher in metropolitan areas than in nonmetropolitan areas and county-level population-weighted mean distance to supermarkets was greater in nonmetropolitan areas compared to metropolitan areas. However, fruit and vegetable consumption decreased as a function of population-weighted mean distance to supermarkets in metropolitan areas only.\textsuperscript{208}

Since the healthcare access literature indicates travel distance and time are inversely related to population density, these findings may be attributed to less frequent travel to supermarkets by nonmetropolitan residents, who may purchase more processed foods and fewer perishable food items than metropolitan residents.\textsuperscript{208} Another explanation is the high prevalence of car ownership in nonmetropolitan areas since rural
residents rely on private automobiles to meet travel needs, irrespective of age, race, and income, and car ownership may mitigate the importance of distance as a barrier to supermarket shopping. Farmer’s markets, fruit and vegetable stands, home gardens, general merchandise stores and church and community center food distribution efforts in nonmetropolitan areas, which may go undetected depending on the store definitions or NAICS codes employed, may also reduce the importance of supermarket accessibility in determining diet in non-metropolitan areas.

As suggested by the authors, there is a need for more detailed, localized studies in non-metropolitan areas to improve our understanding of the effect of geographic food access on diet. Given the paucity of research on this topic specific to Appalachia and characteristics of the region that make it distinct from other parts of the U.S., this suggestion should be extended to the heterogeneous Appalachia region. Primary data collected at the individual-level are needed to shed light on the geographic food access realities of this distinct region, while accounting for geographic food access at the county level. Furthermore, primary geographic food access data collected at the individual-level allow for the limitations of using an area-level geographic food access measure as proxy for proximal-level geographic food access among residents of Appalachia to be explored while accounting for multi-level confounding by factors within the various levels of the social determinants of health relevant to the region.

2.6.5 Summary

In summary, CVD is the leading cause of mortality in U.S., but cancer is poised to overtake CVD as the nation’s leading cause of mortality within two decades.
Moreover, certain populations suffer a disproportionate burden of these diseases, leading to health disparities. There are four modifiable health behaviors (smoking, overweight and obesity, inadequate physical activity and poor diet) and three modifiable health factors (high blood pressure, high total cholesterol and high plasma glucose) that contribute to CVD and cancer outcomes through similar inflammatory pathways. The AHA’s LS7 of CVH is a metric that includes well-defined poor, intermediate and ideal levels of these behaviors and factors, is associated with CVD and cancer outcomes, and is well suited to track progress toward the elimination of health disparities.1, 5, 7, 8

As demonstrated by the Figure A.4, social health determinants at distal, intermediate and proximal levels have the ability to influence disparities in health outcomes independently and directly or through interactions across distal, intermediate and proximal levels. To address the need for improved understanding of how the social determinants of health impact overlapping CVD and cancer risk behaviors and factors, this study first quantified the distribution of the LS7 metric among research samples from mainland Puerto Ricans and residents of Appalachia, two relatively under-researched populations facing health disparities, so that sources of risk for which unfavorable distributions were observed could be prioritized for future research. Of the LS7 components, ideal healthy diet score was the measure most Americans failed to meet.

To better inform research and policy intervention efforts seeking to improve diet among populations facing health disparities, research reflecting the complexity of the way in which diet, the county’s leading modifiable risk factor, is influenced by social determinants of health is needed. There has been increased empirical attention given to
geographic food access, distance-based facilitators and barriers that influence food purchasing and consumption, due to burgeoning use of GIS and availability of federal, state and local incentives to improve access to healthy food in underserved areas. Consequently, this study also sought to explore the relationship between geographic food access, an intermediate-level social determinant, and LS7 healthy diet score among research samples from mainland Puerto Rico and residents of Appalachia. Relevant sources of potential confounding at the area- and individual-level were identified for each research sample and, among the research sample of mainland Puerto Ricans, along with area- and individual-level indicators of acculturation suggested by the literature to have a protective effect on diet were explored as EMM. 

This research provides important information for researchers, public health officials and policy makers related to gaps in the CVD- and cancer-related health of these populations so that efforts to address disparities might be implemented.
Chapter 3: Methodology

3.1 Preface

The overall goals of this study were to quantify the distribution of LS7 among mainland Puerto Ricans and residents of Appalachia and to assess multi-level social determinants of LS7 healthy diet score in these populations facing health disparities. These goals were accomplished by addressing three specific aims: (1) using available Boston Puerto Rican Health Study (BPRHS) and Appalachia Community Cancer Network (ACCN) data, quantify the distribution of each LS7 component and an overall LS7 score within each research sample and compare the distributions to one another and national data; (2) using available BPRHS data, quantify the association between a CT-level food access metric and individual-level LS7 healthy diet score, controlling for individual- and CT-level sociodemographic characteristics and assessing for EMM by individual- and CT-level acculturation; and (3) using available ACCN study data, compare nested models adjusted for individual- and area-level sociodemographic characteristics to determine whether using county-level food access metric or individual-level home-to-preferred food store distance as the exposure of interest is more strongly associated with individual-level LS7 healthy diet score in a research sample of Appalachian residents.

The mainland Puerto Rican and Appalachian research populations that served as the basis of this study are described in terms of study design and measures collected.
Sources of external supplemental data are also summarized. A detailed description of the analyses for each of the three specific aims is provided below and Table B.3 provides a graphical summary of the relationship between the multiple data sources and specific aims.

3.2 The Boston Puerto Rican Health Study (BPRHS)

3.2.1 Study population and design

The BPRHS is an ongoing longitudinal study to examine the role of psychosocial stress on the presence and development of allostatic load and other health outcomes like CVD among Puerto Ricans living in the Greater Boston area. A detailed overview of the study population and design has been published elsewhere. In brief, from June 2004 to October 2009, a total of 2,170 Puerto Rican adults age 45-75 years were identified through door-to-door enumeration, community events, referrals, and calls to the study office. In addition to age, eligibility criteria included: self-identified Puerto Rican descent, able to answer questions in English or Spanish, a Mini Mental State Examination (MMSE) score greater than 10 and living in the Boston, MA metropolitan area at the time of the study and not planning to move within the next two years. Of 1,802 eligible individuals, 1,500 (83.2%) completed the baseline interview, which took place after the initial screening contact, written informed consent was obtained and neuropsychological testing was conducted to identify those who needed assistance from a proxy or were ineligible due to low MMSE score.

The BPRHS protocol was approved through the institutional review board (IRB) of Tufts University and Northeastern University. Approval for this study was obtained
from the IRBs of Tufts University and the Ohio State University. A BPRHS data use agreement was also completed, in which appropriate confidential handling of data during the study and the disposal of the data after the project is completed were outlined.

3.2.2 Measures

Baseline questionnaires and tests were administered by trained, bilingual interviewers in the participant’s home following procedures from NHANES II and the MacArthur Studies of Successful Aging. Age, education level, household income, migration, employment history, family size and food security information were collected using questionnaires modeled after NHANES III, the HANES and the NHIS Supplement on Aging. Additionally, participants provided self-reported chronic condition diagnoses, self-rated health status, health insurance information and health care satisfaction information.

3.2.2.1 *Life’s Simple 7 (LS7)* health behaviors

Smoking frequency, history, and type of smoking were assessed and standing height and weight were measured in duplicate. Current physical activity was captured using a modified Paffenbarger questionnaire of the Harvard Alumni Activity Survey previously tested in an elderly Puerto Rican population. Dietary intake was assessed using a semi-quantitative 126-item food-frequency questionnaire (FFQ) revised to include appropriate foods and portion sizes. The modified FFQ was shown to capture intakes reported in 24-hour recalls more accurately than the NCI-Block FFQ format questionnaire after which it was modeled. This FFQ has been validated against dietary biomarkers in Hispanics aged ≥ 60 years.
3.2.2.2 Life’s Simple 7 (LS7) health factors

The interviewer collected detailed information on prescription and over-the-counter medications using medical bottles presented by participants. Systolic and diastolic blood pressures were measured in duplicate, at three time points during the interview and the second and third readings were averaged. A day after completing the baseline questionnaire, a certified phlebotomist returned to participants’ homes to collect the urine and saliva samples and draw blood in the home. Blood samples were analyzed for plasma lipids and glucose, among other outcomes.

3.2.2.3 Additional variables

A questionnaire adapted from the Bi-dimensional Acculturation Scale for Hispanics (BAS) was used to assess linguistic acculturation. The Bi-dimensional Acculturation Scale for Hispanics (BAS) yields two scores that rank participant’s acculturation in Hispanic and non-Hispanic domains. An acculturation scale of psychological attachment to either culture, previously validated in three samples of Puerto Ricans in the greater Boston area, was also administered. Place of birth and years in the U.S., commonly used as proxies for acculturation, were also assessed.

3.3 The Appalachian Community Cancer Network (ACCN) Study

3.3.1 Study population and design

The ACCN, a joint effort of the University of Kentucky, The Ohio State University, Pennsylvania State University, Virginia Tech University, and West Virginia University, is currently conducting a group randomized trial in the five-state area to test a previously piloted ACCN faith-based intervention focused on individual and
environmental level changes to increase physical activity and improve healthy food choices using eHealth technology. Using county as the unit of randomization, 13 churches were randomly assigned to receive *Walk by Faith*, the dietary and physical activity faith-based intervention program, and 15 were randomly assigned to receive *Ribbons of Faith*, a comparison condition focused on cancer screening. Intervention effectiveness and sustainability will be tested and the results will be used to disseminate *Walk by Faith* to the comparison churches, other Appalachian and rural churches and the ‘Cancer Control Plan, Link, Act, Network with Evidence-based Tools (PLANET)’ research-tested intervention programs (RTIPs) website.

The study was promoted in church bulletins and announcements from the pulpit. Group recruitment events and individual appointments were held at participating churches, at which time an individual could be screened to determine eligibility to participate. Eligibility criteria included: at least 18 years of age, attending services at least 4 times in the past two months at a participating church, able to understand and read English, cognitively able to provide informed consent, resident of an Appalachian county, not residing in a nursing facility or residential home, not planning to move away from the study area, willing to use a computer, no dietary restrictions prescribed for weight-loss or part of a formal weight-loss program, less than 400 pounds, having a BMI of at least 25, and, if female, not pregnant, breastfeeding or less than 9 months post-partum or planning to become pregnant during study period. Questions modified from the Physical Activity Readiness Questionnaire (PAR-Q) were used to determine whether medical clearance from a physician was required to participate. In total, 427 *Walk by Faith* and 237
Ribbons of Faith participants completed the baseline questionnaires. For the purpose of this study, the ACCN study population was treated as a cohort of 663 people.

The ACCN protocol was approved through the IRBs of the Ohio State University, the University of Kentucky, Pennsylvania State University, Virginia Polytechnic Institute and State University and West Virginia University. Approval for the purposes of this study was obtained from the Ohio State University and an ACCN data use agreement was executed.

3.3.2 Measures

Those meeting eligibility requirements completed a phone interview, in-person computer assisted interview and take-home survey prior to the start of the Walk by Faith and Ribbons of Faith programs. All interviewers were trained centrally by research staff at the Ohio State University, the coordinating center for the study, using a standardized protocol.

3.3.2.1 Life’s Simple 7 (LS7) health behaviors

Smoking frequency, history, and type were assessed during the phone interview. Height and weight measurement protocols were adapted from the PhenX Toolkit version 4.5, July 29, 2011. Standing height was measured twice using a portable stadiometer. If the difference between the two measurements was greater than 0.1 cm, a third measurement was taken. The measurements were then averaged to determine height. Weight was measured twice using a calibrated portable scale. If the difference between the two measurements was greater than 0.1 kg, a third measurement was taken. The measurements were then averaged to determine weight. VioActive, a proprietary survey
adapted from the International Physical Activity Questionnaire, (IPAQ) by Viocare, Inc., a privately-owned wellness systems technology company, was administered online during the in-person interview to assess physical activity.\textsuperscript{231, 232} Viocare’s proprietary graphic FFQ, VioScreen, was also administered online during the in-person interview to assess diet.\textsuperscript{233}

\subsection*{3.3.2.2 Life’s Simple 7 (LS7) health factors}

Systolic and diastolic blood pressure were measured twice on each arm using a protocol adapted from recommendations of the Subcommittee of Professional and Public Education of the American Heart Association Council on High Blood Pressure Research, the PhenX Toolkit version 4.5, July 29, 2011 and the European Health Examination Survey (EHES) Manual version July 28, 2011.\textsuperscript{230, 234, 235} Since blood samples were not collected, total cholesterol and plasma glucose were not assessed. However, it was possible to determine use of blood pressure, cholesterol and glucose lowering medication from responses to medical clearance questions.\textsuperscript{229}

\subsection*{3.3.2.3 Additional variables}

Questions pertaining to food purchasing behavior, including the name and location of preferred food shopping store, store-to-home distance, method of transportation and reasons for patronizing, were part of the take-home surveys given to participants in Ohio, Kentucky, Virginia and West Virginia. Some food purchasing behavior questions were adapted from food-related questions included in the Boehmer/Brownson physical environment survey.\textsuperscript{236}
3.4 External data sources

3.4.1 U.S. Department of Agriculture (USDA) Food Environment Atlas

For Aim 2, the exposure of interest was the USDA’s low access status for urban CTs using a 0.5-mile distance. Urban CTs are considered low access if 500 or more people or 33% of those residing in the CT live a half mile or more from the nearest supermarket or large grocery store. This variable was derived from a directory of supermarkets and large grocery stores within the U.S. created by merging a 2010 directory of stores authorized to accept SNAP benefits and the 2010 Trade Dimensions TDLine directory of stores. To be included, stores must report at least $2 million in annual sales and contain all the major food departments found in a traditional supermarket (e.g. fresh meat and poultry, dairy, dry and packaged foods and frozen foods). The Census Bureau’s Urban Area definition was used to designate urban status. Stores were geocoded to the street address. Census block-level population data from the 2010 Census of Population and Housing were aerially allocated to ½-kilometer-square grids across the U.S., and distance was determined by calculating the distance from the geographic center of each ½- kilometer-square grid cell to the center of the grid cell containing the nearest store.

USDA data on the percentage of county residents living more than 1 mile from a supermarket or large grocery store, if in an urban area, or more than 10 miles from a supermarket or large grocery store, if in a rural area, was used to supplement the ACCN study data. This variable was derived from a directory of supermarkets and large grocery stores within the U.S. created by merging a 2010 directory of stores authorized to
accept SNAP benefits and the 2010 Trade Dimensions TDLinx directory of stores. To be included, supermarkets and large grocery stores must report at least $2 million in annual sales and contain all the major food departments found in a traditional supermarket (e.g. fresh meat and poultry, dairy, dry and packaged foods and frozen foods).

The Census Bureau’s Urban Area definition was used to designate rural and urban status. Low-income was defined as annual family income less than or equal to 200 percent of the Federal poverty threshold, taking family size into account. Supermarkets and large grocery stores listed were geocoded to the street address. Census block-level population data from the 2010 Census of Population and Housing were aerially allocated to ½-kilometer-square grids across the U.S. Distance was determined by calculating the distance from the geographic center of each ½-kilometer-square grid cell to the center of the grid cell containing the nearest supermarket or large grocery store. To obtain this variable, a county-level aggregate of the number of individuals living more than 1 mile from a supermarket or large grocery store in urban areas and more than 10 miles from a supermarket or large grocery store in rural areas was divided by the county population size.195

3.4.2 Centers for Disease Control and Prevention (CDC)

CT-level mRFEI scores were obtained from the CDC website to supplement the BPRHS data as part of a sensitivity analysis among participants for whom the definition of the CT in which they lived at baseline did not change between the 2000 and 2010 US Census.237 Using NAICS codes, the mRFEI considers the number of healthy and less healthy food retailers within a CT or ½ mile from the CT boundary, using 2000 CT
definitions. The mRFEI score ranges from 0 to 100, indicating the percentage of healthy food retailers out of the total number of healthy and less healthy food retailers in a census tract or ½ mile from the tract boundary.\textsuperscript{188}

3.4.3 U.S. Census

To supplement the BPRHS data, five-year population estimates from the 2007-2011 AmCS (Total Population, Hispanic or Latino by Specific Origin, Ratio of Income in 2009 to Poverty Level, Number Spanish-Speaking Linguistically Isolated Households) data were obtained from Social Explorer, an online subscription demographic research tool.\textsuperscript{238} However, Spanish-speaking linguistically isolated households estimates were unstable and could not be used for the purpose of this study.

To facilitate the calculation of individual-level poverty within each research sample, poverty thresholds by size of family were obtained from the U.S. Census website for the years in which BPRHS and ACCN study data were collected.\textsuperscript{239}

3.4.4 Appalachian Regional Commission (ARC)

Fiscal year 2012 county economic strength data compiled by the ARC were used to supplement ACCN study data.\textsuperscript{31} As described previously, distressed counties have (1) a poverty rate greater than or equal to 150% of the national average, per capita market income less than or equal to two thirds of the national average, and unemployment rate greater than or equal to 150% of the national average, or (2) a poverty rate greater than or equal to 200% that of the national average and meets one of the other criteria.\textsuperscript{31} There are five possible categories of economic strength: distressed, at-risk, transitional, competitive and attainment.\textsuperscript{31}
3.4.5 National Health and Nutrition Examination Survey (NHANES)

NHANES encompasses a series of cross-sectional, survey-based studies that assess the health and nutritional status among U.S. adults and children using two data collection methods: interviews and physical examinations. NHANES employs a complex, multistage, probability design to select participants from geographically-defined strata (e.g. single or contiguous counties for which probability of selection is proportional to size) and proportions of minority populations so that the resulting sample is nationally representative of the country’s civilian, non-institutionalized population.

Demographic, socioeconomic, dietary, and health-related information is collected in the interview portion whereas medical, dental, and physiological measurements and laboratory tests are administered in the physical examination portion. Since NHW are commonly used as the group to which populations facing health disparities are compared in the literature, 2007-2010 Continuous NHANES data were restricted to NHW for the purpose of this study.

3.4.5.1 Life’s Simple 7 (LS7) health behaviors

Use of cigarettes, pipes, and cigars was assessed via questionnaire to determine smoking status and height and weight were measured during the clinical examination. Self-reported frequency and duration of specific moderate-intensity and vigorous-intensity activities over the past week or month and transportation and household activities were assessed via questionnaire. Participants completed two 24-hour dietary recalls administered by an interviewer.
3.4.5.2 Life’s Simple 7 (LS7) health factors

Blood lipids and fasting plasma glucose were determined from blood samples collected during the clinical examination. Three consecutive blood pressure readings (or four, if a measurement was interrupted or incomplete) were collected during the clinical examination. Medication use was assessed during the home interview.²⁴⁰

3.5 Aim 1

The first aim of this study was to quantify the distribution of each LS7 component and an overall LS7 score within community samples of mainland Puerto Ricans and residents of Appalachia and compare the distributions to one another and national data.

3.5.1 Data Sources

3.5.1.1 Boston Puerto Rican Health Study (BPRHS)

2004-2009 baseline data available for BPRHS participants on smoking, BMI, physical activity, diet, systolic and diastolic blood pressure, plasma lipids and glucose were used to derive individual LS7 components and overall LS7 score.

3.5.1.2 Appalachia Community Cancer Network (ACCN) Study

2012-2013 baseline data on smoking, BMI, physical activity, dietary intake and systolic and diastolic blood pressure available for ACCN study participants were used to derive individual LS7 components and overall LS7 score.

3.5.1.3 National Health and Nutrition Examination Survey (NHANES)

2007-2010 NHANES data from NHW participants pertaining to smoking, BMI, physical activity, diet, systolic and diastolic blood pressure, plasma lipids and glucose were used to derive individual LS7 components and overall LS7 score.
3.5.2 Analyses

3.5.2.1 Individual LS7 Components and Overall LS7 Score

Unless otherwise specified, baseline BPRHS and ACCN study data and 2007-2010 Continuous NHANES data were used to derive individual LS7 components in line with the AHA definition. For the BPRHS and ACCN study population, the definition of ideal, intermediate and poor smoking differed slightly from the AHA definition because of the data available. Ideal smoking was defined as never smoking or smoking less than 100 cigarettes in one’s entire life, intermediate was defined as smoked in the past but not currently, and poor was defined as currently smoking. Only poor and intermediate levels of BMI were estimated since ACCN participants had to have had a BMI of at least 25 to be eligible for the study. In terms of healthy diet score components among ACCN participants, daily fruit and vegetable servings were not based on cups but rather standard portion sizes, four ounces was considered a standard serving of fish rather than three and a half ounces. Ideal, intermediate and poor sweets/sugar-sweetened beverages were derived from a variable of gram and teaspoon equivalents of added sugars among ACCN and NHANES participants, respectively. Given the categories available in the USDA’s Food Patterns and Equivalent Databases for 2007-2008 and 2009-2010, NHANES fish consumption had to be defined as at least two 3.5 oz servings of seafood high in n-3 fatty acids and seafood low in n-3 fatty acids. Cholesterol and plasma glucose were not assessed among ACCN study participants. However, the use of glucose lowering medication was derived from a response to a medical clearance question regarding
whether a doctor had prescribed drugs to control diabetes so that ACCN participants could be classified as ideal or not.\textsuperscript{243}

For each study population, an overall LS7 was calculated using the approach by Folsom et al. (2011) such that the overall LS7 score was based on the number of ideal LS7 metrics present at baseline.\textsuperscript{170} Among BPRHS and NHANES participants, there were 8 possible strata ranging from 0 ideal LS7 components to 7 ideal LS7 components. Among ACCN participants, who were required to have a BMI \( \geq 25 \) to participate and were not subject to cholesterol or glucose testing, there were 6 possible strata for the overall LS7 score ranging from 0 ideal LS7 components to 5 ideal LS7 components. Achievement of poor, intermediate and ideal LS7 components, as defined by the AHA (Table B.1) was calculated. Continuous variables were described using means (SDs), and categorical variables were described using percentages.

\textbf{3.5.2.2 Comparison between BPRHS and ACCN}

Although there are differences in study design and subject representativeness between the BPRHS and ACCN study, comparisons between the research samples are nevertheless informative in light of the health disparities in the U.S. and limited use of the LS7 metric among research samples from populations facing CVD and cancer disparities. Given the differing eligibility criteria between the two studies, BPRHS and ACCN data were restricted to participants age 45-75 with a BMI of at least 25 to foster comparison between the two research samples, resulting in a sample of 937 BPRHS and 313 ACCN study participants.
For smoking, physical activity, diet and blood pressure, the proportion of participants from both studies who were classified into poor, intermediate, and ideal categories was calculated. For plasma glucose, the proportion of participants from both studies who were classified into not ideal and ideal categories were calculated. For BMI, the proportion of participants from both studies who were classified into poor and intermediate categories was calculated. For each study population, an overall LS7 score was calculated, as described previously. There were 6 possible strata for the overall LS7 score ranging from 0 ideal LS7 components to 5 ideal LS7 components.\textsuperscript{170}

Adjustment for age and sex to the 2000 U.S. standard population was performed using the direct method.\textsuperscript{244} Age- and sex-adjusted estimates and standard errors of the estimates made it impossible to make compare BPRHS and ACCN participants using traditional contingency table analysis. Consequently, Wald tests for proportions were used with an understanding that a difference in the proportion of BPRHS and ACCN participants classified at one level suggested differences in the proportion of BPRHS and ACCN participants classified at other available level(s). Wald tests for proportions were not treated as separate, independent tests.\textsuperscript{245} For smoking, physical activity, blood pressure and diet, the age- and sex-adjusted proportion of participants classified into poor, intermediate and ideal levels were compared between the two research samples using three Wald tests for proportions based on the age- and sex-weighted frequencies. The proportion of participants classified into poor and intermediate BMI levels and the proportion of participants classified into ideal and not ideal plasma glucose levels were compared between the two research samples using one Wald test for proportions based
on the age- and sex-weighted frequencies. Overall LS7 score was examined categorically and continuously. The overall LS7 score was collapsed into three groups based on the distributions observed (i.e., 0-1, 2-3 and 4) and compared using three Wald tests for proportions based on the age- and sex-weighted frequencies. Difference in mean continuous overall LS7 score for each research sample was assessed using a two sample t test of the sex and age-adjusted mean for BPRHS and ACCN study participants. All statistical tests were two-sided. For the comparisons of overall LS7 score and available LS7 components, the Bonferroni multiple testing correction method was applied to determine the significance for an individual comparison given an overall Type I error of 0.05. All analyses were performed using SAS version 9.3 (SAS Institute, Inc., Cary, NC) and Stata SE version 12 (StataCorp, College Station, Texas).

The following formula was used to determine the size of the effect estimate (Δ) that could be obtained by this study: \[ \Delta^2 = \frac{2\sigma^2(Z_p+Z_{a/2})^2}{n}, \] with \( \sigma = 1 \) to obtain the effect size (i.e., mean difference/SD). Given a sample size of at least 313 in each study population and an overall Type 1 error of 0.05 using the Bonferroni correction (i.e., a p value of 0.003 for an individual comparison), this study had 80% power to detect an effect size of 0.305 pertaining to the difference in mean continuous overall LS7 score between BPRHS and ACCN study participants.

3.5.2.3 Comparisons between BPRHS and NHANES and ACCN and NHANES

BPRHS and ACCN study restrictions were applied separately to NHANES data for NHW to foster comparisons with each research sample. Adjustment for age and sex to
the 2000 U.S. standard population was performed using the direct method.\textsuperscript{244} It was impossible to compare NHANES estimates to age and sex adjusted BPRHS or ACCN estimates and standard errors using traditional contingency table analysis, given the NHANES’ complex, multistage sampling design. NHANES estimates and standard errors of the estimates accounted for age and sex adjustment as well as survey weighting, clustering and stratification. Consequently, differences in the prevalence of poor, intermediate and ideal levels of each LS7 component (or some other combination of levels) were assessed by comparing overlapping 99\% confidence intervals for sex and age-adjusted estimates of each component at each level with an understanding that non-overlap at one level suggested differences, at least qualitatively, in the proportion of individuals in each study at other available level(s). To minimize Type 1 error, 99\% CIs were used, rather than 95\% CIs.

Overall LS7 score was collapsed into three groups based on the distributions observed.\textsuperscript{7} Differences in the proportion of NHANES participants classified into each overall LS7 score category were assessed by comparing overlapping 99\% confidence intervals for each category. Non-overlapping 99\% confidence intervals suggested a difference in the proportion of individuals in each study population categorized in a given category. Differences in mean overall LS7 score for each research sample were assessed by comparing overlapping 99\% confidence intervals with an understanding that non-overlap at one categorical level suggested differences, at least qualitatively, in the proportion of individuals in each study at other categorical levels. All analyses were
performed using SAS version 9.3 (SAS Institute, Inc., Cary, NC) and Stata SE version 12 (StataCorp, College Station, Texas).

The following formula was used to determine the size of the effect estimate ($\Delta$) that could be obtained by this study: $\Delta^2 = \frac{2 \sigma^2 (Z_\beta + Z_\alpha)^2}{n}$, with $\sigma=1$ to obtain the effect size (i.e., mean difference/SD). This study had 80% power to detect an effect estimate of 0.198 pertaining to the difference in mean continuous overall LS7 score between BPRHS and NHANES participants given a sample size of at least 587 in each study population and an overall Type 1 error of 0.01. This study had 80% power to an effect estimate of 0.235 pertaining to the difference in mean continuous overall LS7 score between ACCN study and NHANES participants given a sample size of at least 417 in each study population and an overall Type 1 error of 0.01.

3.6 Aim 2

The second aim of this study was to quantify the association between a CT-level food access metric and individual-level LS7 healthy diet score in a research sample of mainland Puerto Ricans, controlling for individual- and CT-level sociodemographic characteristics and assessing for EMM by individual- and CT-level acculturation.

3.6.1 Data

3.6.1.1 Outcome

The outcome of interest was LS7 healthy diet score, which was derived from dietary intake collected at baseline of the BPRHS using a culturally tailored FFQ. Given the low prevalence of LS7 healthy diet score in the BPRHS (per descriptive statistics run
by BPRHS staff), LS7 healthy diet score was treated as a binary variable (<2 or ≥2 or (i.e., poor vs. not poor)).

3.6.1.2 Exposure

For the purpose of this study, a USDA variable representing whether an urban CT is considered as having low food access, a 0.5-mile distance was used as the exposure of interest. Urban CTs are considered low access if 500 or more people or 33% of those residing in the CT live a half mile or more from the nearest supermarket, supercenter or large grocery store. 

3.6.1.3 Effect measure modifiers

Based on the literature, individual and CT-level indicators of acculturation have been chosen a priori for exploration as EMMs of the relationship between CT-level USDA low access status and individual-level LS7 healthy diet score. Linguistic acculturation, an indicator of acculturation at the individual-level, was explored as an EMM using a questionnaire adapted from the BAS. The BAS yields two scores to rank participant’s acculturation in the Hispanic and the non-Hispanic domains. Linguistic acculturation score ranges from 0% (fully unacculturated, speaking Spanish only) to 100% (fully acculturated, speaking fluent English) and was treated as a binary variable (<50 vs. ≥50). In terms of acculturation indicators at the CT-level to be explored as EMMs, the percent of the total population that was Hispanic/Latino and Puerto Rican specifically were derived from the 2010 decennial U.S. Census data available from Social Explorer, an online subscription demographic research tool. Both area-level measures
were treated as continuous variables and explored as EMMS because it was unclear from the literature which is more closely associated with diet among mainland Puerto Ricans.

3.6.1.4 Potential confounders

A Directed Acyclic Graph (DAG) was used to determine whether a number of available BRPHS and U.S. Census covariates representing area- and individual-level social determinants of health met criteria for confounding (Figures A.5 and A.6). To be a confounder, a covariate must (1) be associated with LS7 healthy diet score; (2) be associated with USDA low access status, and (3) not be affected by USDA low access status. Plausible causal relationships (represented by unidirectional arrows) were determined on the basis of extant literature and biological plausibility. Potential confounders considered for model building must (1) have an arrow pointing in either direction between it and LS7 healthy diet score; and (2) have an arrow pointing from it to USDA low access status (arrows pointing from USDA low access status suggest the covariate is on the causal pathway between USDA low access status, and that LS7 healthy diet score and treating the covariate as a confounder may induce bias). As shown in Figure A.6, individual-level potential confounders considered were age, sex and marital status, and poverty was a potential confounder at the level of the individual and CT.

Sex was treated as a binary variable, and age at baseline was treated as a continuous variable. Marital status was treated as a binary variable: married/living as married/spouse in household or not. Individual-level poverty was treated as a binary variable and was derived from family size, total household income, and poverty
thresholds by size of family obtained from the U.S. Census website for the year in which
data were collected.\(^{239}\) The indicator of area-level poverty was treated as a continuous
variable. CT-level percent of population with a poverty-to-income ratio (PIR) below the
poverty line was derived from five-year 2007-2011 AmCS population estimates of the
ratio of 2009 family or unrelated individual income to the appropriate poverty
threshold.\(^{238}\) A ratio below 1.00 indicated income below the poverty level.

3.6.2 Analyses

3.6.2.1 Primary Analysis

Descriptive statistics were used to provide overall sample characteristics. The
method of Generalized Estimating Equations (GEE) was used to yield a population
averaged model of the association between USDA low access status and LS7 healthy diet
score with robust standard errors (SEs). Bernoulli was specified as the family (the
assumed distribution of LS7 healthy diet score), logit was specified as the link function,
and an exchangeable working correlation structure was specified to reflect the average
dependence among correlated LS7 healthy diet scores among participants in the same
CT.

GEE models with robust SEs unadjusted for potential confounders were created
for (1) the association between USDA low access status and LS7 healthy diet score, (2)
the association between USDA low access status and LS7 healthy diet score, assessing
for EMM by individual-level linguistic acculturation, (3) the association between USDA
low access status and LS7 healthy diet score, assessing for EMM by CT-level percent
Latino, and (4) the association between USDA low access status and LS7 healthy diet
score, assessing for EMM by CT-level percent Puerto Rican. Prior to the addition of all the EMMs into a GEE model of the association between USDA low access status and LS7 diet adjusted for potential confounders identified by the DAG, collinearity of EMMs at the CT-level was assessed. Interaction terms were tested individually and deleted if not significant at $\alpha=0.10$ until all remaining interactions in the model were statistically significant. Potential confounding variables resulting from the DAG exercise were included in the final model if there was at least a 10% change in the effect estimate between models with and without controlling for the variable. For EMMs, $p<0.10$ was considered statistically significant and $p<0.05$ was considered statistically significant for USDA food desert status, the sole covariate, and 95% CIs were calculated. Model diagnostics included (1) the effects of potentially influential points, (2) whether linearly modeled variables were approximately linear in the logit and (3) a sensitivity analysis excluding large clusters. All analyses were performed using SAS version 9.3 (SAS Institute, Inc., Cary, NC) and Stata SE version 12 (StataCorp, College Station, Texas).

Given the use of secondary data analysis, this study was not planned for confirmatory statistical inference. Rather, it was planned to produce estimates of associations and assess potential confounders to inform power calculations for future studies.

### 3.6.2.2 Secondary Analysis

Geographic food access can be conceptualized and measured as an absolute or a relative phenomenon. Absolute food access measures like the USDA low access variable determine access in relation to one type of store (i.e. large grocery store or supermarket)
and are the most prominent across the literature. Though less common, relative food access measures also take food retail locations like fast food establishments and convenience stores into account. A sensitivity analysis was conducted using the CDC’s mRFEI, a relative food access measure that considers the number of healthy food retailers, as well as the number of less healthy food retailers within a CT or ½ mile from the CT boundary. Healthy food retailers include supermarkets and larger grocery stores (NAICS code 445110; specifically supermarkets with ≥ 50 annual payroll employees and grocery stores with 10–49 employees); fruit and vegetable markets (NAICS code 445230) and warehouse clubs (NAICS code 452910). Less healthy food retailers include fast food restaurants (NAICS code 722211), small grocery stores (NAICS code 445110, specifically grocery stores with three or fewer employees), and convenience stores (NAICS code 445120). The mRFEI score, which can range from 0 to 100, represents the percentage of healthy food retailers out of the total number of healthy and less healthy food retailers in a CT or ½ mile from the CT boundary. Since mRFEI was created using 2000 CT definitions and BPRHS participants were geocoded to 2010 CT definitions, only participants for whom the definition of the CT in which they lived at baseline did not change between the 2000 and 2010 US Census were included in the sensitivity analysis.

Descriptive statistics were used to provide overall sample characteristics. The method of GEE was used to yield a population averaged model of (1) the association between CDC mRFEI score and LS7 healthy diet score (dichotomized into poor (>2) vs. not poor (≥2)), (2) the association between CDC mRFEI score and LS7 healthy diet score,
assessing for EMM by individual-level linguistic acculturation, and (3) the association between CDC mRFEI score and LS7 healthy diet score, assessing for EMM by CT-level percent Latino. Prior to the addition of all the EMMs into a GEE model of the association between CDC mRFEI score and LS7 diet adjusted for potential confounders identified by the DAG, collinearity of EMMs at the CT-level was assessed. Interaction terms were tested individually and deleted if not significant at $\alpha=0.10$ until all remaining interactions in the model were statistically significant. Potential confounding variables resulting from the DAG exercise were included if there $10\%$ change in the effect estimate between models with and without controlling for the variable. For EMMs, $p<0.10$ was considered statistically significant and $p<0.05$ was considered statistically significant for mRFEI score, the sole covariate, and 95% CIs were calculated. Model diagnostics included (1) the effects of potentially influential points, (2) whether linearly modeled variables were approximately linear in the logit, and (3) a sensitivity analysis excluding large clusters.

3.7 Aim 3

The third aim of this study was to compare nested models adjusted for individual- and area-level sociodemographic characteristics to determine whether using county-level food access metric or individual-level home-to-preferred food store distance as the exposure of interest was more strongly associated with individual-level LS7 healthy diet score in a research sample of Appalachian residents. Since questions pertaining to food purchasing behavior were not given to Pennsylvania participants, this aim was limited to data from participants in Ohio, Kentucky, Virginia and West Virginia.
3.7.1 Data

3.7.1.1 Outcome

The outcome of interest was the LS7 healthy diet score, which was derived from dietary intake collected at baseline of the ACCN study using a computer-based visual FFQ. Given the low prevalence of LS7 healthy diet score exhibited in the Appalachian literature, LS7 healthy diet score was treated as a binary variable (i.e., <2 (poor) vs. ≥2 (not poor)).

3.7.1.2 Exposure

There were two exposures of interest: (1) the USDA county-level variable capturing the percentage of county living more than 1 mile from a supermarket or large grocery store if in an urban area, or more than 10 miles from a supermarket or large grocery store if in a rural area, and (2) the self-reported, individual-level home-to-preferred food store distance.

3.7.1.3 Potential confounders

A DAG was used to determine whether a number of available ACCN, ARC and U.S. Census covariates representing area- and individual-level social determinants of health met criteria for confounding (Figures A.7 and A.8). To be a confounder, a covariate must (1) be associated with LS7 healthy diet score; (2) be associated with the exposures of interest, and (3) not be affected by the exposures of interest. Plausible causal relationships (represented by unidirectional arrows) were determined on the basis of extant literature and biological plausibility. Potential confounders considered for model building must (1) have an arrow pointing in either direction between it and LS7
healthy diet score; and (2) have an arrow pointing from it to the county-level USDA food access variable and/or individual-level home-to-preferred store distance (arrows pointing from either exposure suggest the covariate is on the causal pathway between an exposure and LS7 healthy diet score, and that treating the covariate as a confounder may induce bias).²⁵⁰ As shown in Figure A.8, individual-level potential confounders of the relationship between the county-level USDA variable and LS7 healthy diet score variable were age, sex, marital status and household income, whereas economic strength is a potential confounder at the level of the county. For the relationship between individual-level, preferred home-to-food store distance and LS7 healthy diet score, individual-level potential confounders were age, sex marital status and household income whereas economic strength and the USDA variable were potential confounders at the level of the county.

3.7.2 Analyses

Descriptive statistics were used to provide overall ACCN participant characteristics. GEE yielded population average models of the association between the exposures of interest and LS7 healthy diet score with robust SEs. Bernoulli was specified as the family (the assumed distribution of LS7 healthy diet score), logit was specified as the link function, and a working exchangeable correlation structure was specified to reflect the average dependence among correlated LS7 healthy diet scores among participants in the same county.

As shown in Figure A.8, the adjusted model using the USDA variable as the exposure of interest includes fewer parameters than the adjusted model using individual-
level, preferred home-to-food store distance as the exposure of interest and is thus nested within the adjusted model using individual-level, preferred home-to-food store distance as the exposure of interest. Given the nested nature of the two models, all potential confounders identified by the DAG were included in the GEE models. In terms of model diagnostics, fractional polynomials were used to assess whether county-level USDA percent low food access and individual-level food distance in miles were linear in the logit.

For classic linear regression with independent data, the Akaike's Information Criterion (AIC) is a common model-selection criterion that combines model fit and complexity. AIC is based on the likelihood and asymptotic properties of the maximum likelihood estimator (MLE) method used to estimate statistical model parameters. AIC cannot be used in the context of GEE since no distribution assumptions regarding the outcome are made, thus there is no likelihood function. Instead, the quasi-likelihood under the independence model criterion (QIC) was used for GEE model selection. As with AIC, when comparing two models, the model with the smaller QIC value is preferred, meaning it is a better fit taking model complexity into account.

For the effect estimate of the exposure of interest, p<0.05 was considered statistically significant and 95% CIs were calculated. The magnitude and statistical significance of the effect estimate of the exposures of interest and QIC were compared across the two models. All analyses were performed using SAS version 9.3 (SAS Institute, Inc., Cary, NC) and Stata SE version 12 (StataCorp, College Station, Texas).
Given the use of secondary data analysis, this study was not planned for statistical inference, but to produce precise estimates. The reported estimates can be used to inform power calculations for future studies on the relationship between geographic food access and diet.
Chapter 4: The Prevalence of *Life’s Simple 7 of Cardiovascular Health* Metric in The Boston Puerto Rican Health Study, The Appalachia Community Cancer Network Study, and the National Health and Nutrition Examination Survey

4.1 Abstract

**Background:** Within two decades, cardiovascular disease (CVD) will be surpassed by cancer as the leading cause of death in the U.S.\(^1\)\(^2\) The American Heart Association’s (AHA’s) *Life’s Simple 7* (LS7) of cardiovascular health (CVH) metric captures four behaviors (smoking, body mass index (BMI), physical activity, and diet) and factors (total cholesterol, blood pressure and plasma glucose) associated with CVD and cancer risk.\(^5\) By differentiating between poor, intermediate and ideal levels of each health behavior and factor, this metric is well-suited to quantify gradients in disease risk in the general population and monitor progress toward public health benchmarks. Furthermore, by quantifying LS7 across diverse populations, this metric can also be used to prioritize specific risk factors and behaviors for future research and interventions to address the social determinants of health contributing to CVD- and cancer-related health disparities. This study focused on the distribution of the LS7 metric among research samples of Puerto Ricans living in the contiguous U.S. (herein referred to as mainland Puerto Ricans); residents of Appalachia, two relatively understudied populations facing CVD-
and cancer-related health disparities; and a nationally representative sample of non-Hispanic whites (NHWs).

Methods: Using 2004-2009 baseline data from the Boston Puerto Rican Health Study (BPRHS), 2012-2013 baseline data from the Appalachia Community Cancer Network (ACCN) study and 2007-2010 National Health and Nutrition Examination Survey (NHANES) data, this study quantified the distribution of each LS7 component and an overall LS7 score within each study population. Different data restrictions were applied to foster comparisons between BPRHS and ACCN study participants, BPRHS and NHW 2007-2010 NHANES participants and ACCN study and NHW 2007-2010 NHANES participants. Differences in age- and sex-adjusted prevalence estimates of poor, intermediate and ideal levels of each LS7 component (or some other combination of levels given data available) between BPRHS and ACCN study participants were assessed using Wald test for proportions and two sample t tests with an overall Type 1 error of 0.05 (i.e., p value of 0.003 for an individual comparison). Differences in age- and sex-adjusted prevalence estimates of poor, intermediate and ideal levels of each LS7 component (or some other combination of levels given data available) between BPRHS and NHANES participants and between ACCN study and NHANES participants were assessed by comparing 99% confidence intervals (CIs) to minimize overall Type 1 error.

Results: There were statistically significant differences between BPRHS and ACCN study participants across at least one level of available LS7 components, categorical overall LS7 score (≤1, 2 to 3, 4), and continuous overall LS7 score (range: 0-4). Whereas BPRHS participants had a more favorable blood pressure distribution, ACCN participants
had more favorable distributions of smoking, physical activity, diet and plasma glucose. ACCN participants also had a more favorable categorical overall LS7 score distribution and a higher average continuous overall LS7 score. Based on the 99% CIs, there were differences between BPRHS and NHW NHANES participants across at least one level of all LS7 components except diet. BPRHS participants had more favorable distributions of physical activity and total cholesterol, whereas NHW NHANES participants had more favorable distributions of smoking, BMI, and plasma glucose. However, the distribution of categorical overall LS7 score (≤2, 3 to 4, and ≥5) and mean continuous overall LS7 scores (range: 0-6) did not differ substantially between the study populations. Based on the 99% CIs, there were significant differences between ACCN study and NHW NHANES participants across at least one level of all LS7 components except diet. ACCN participants tended to have more favorable distributions of smoking, physical activity and plasma glucose, whereas NHW NHANES participants had more favorable distributions of BMI and blood pressure. Although the distribution of categorical overall LS7 score was more favorable among ACCN study participants, mean continuous overall LS7 score did not differ substantially between the study populations, likely because there were more NHW NHANES participants with overall LS7 scores ≥4 and ≤1 while ACCN participants tended to have overall LS7 scores of 2 and 3.

**Discussion:** Using LS7, this study identified precisely where CVD and cancer risk differs between research samples from two populations facing health disparities and between each research sample and a nationally representative sample of NHWs. As demonstrated by this study, two populations that face CVD and cancer health disparities are not
necessarily similar in terms of the distribution of mutual CVD and cancer risk factors. Differences between a research sample from a population facing health disparities and a nationally representative research sample of NHWs are not always in the direction that may be expected. Given the way in which research samples were recruited, the results of this study are limited in terms of generalizability, and conclusions are limited in terms of how the results apply to all mainland Puerto Ricans and residents of Appalachia. This study’s approach can be nevertheless be used to prioritize specific modifiable CVD and cancer risk factors for future research and intervention and track progress toward the elimination of health disparities within underserved populations and between underserved populations and the rest of the U.S. In addition to using the LS7 to identify areas for future research and intervention to impact health disparities, an important next step in health disparity research is to integrate the LS7 metric with diverse data sources to better understand the role of social determinants of CVD and cancer risk factors and behaviors and inform intervention strategies.\textsuperscript{18, 19}
4.2 Introduction

Despite significant gains in CVD prevention and control over the past four decades, CVD is still the leading cause of mortality in the U.S.\textsuperscript{1} However, cancer is poised to overtake CVD as the nation’s leading cause of mortality within the next two decades.\textsuperscript{2} U.S. public health benchmarks to address rising CVD and cancer burdens include the AHA 2020 Impact Goal “to improve the cardiovascular health of all Americans by 20% while reducing deaths from cardiovascular diseases and stroke by 20%” and the U.S. Department of Health and Human Services’ (USDHHS) Healthy People 2020 objective to reduce the cancer mortality rate by 10% from 179.3 per 100,000 in 2007 to 161.4 per 100,000 in 2020.

More than 70% of CVD- and 34% of cancer-related deaths globally can be attributed to sources of risk that are amenable to change.\textsuperscript{81-83} Smoking, overweight and obesity, inadequate physical activity, poor diet, high blood pressure, high total cholesterol and high plasma glucose are modifiable risk behaviors and factors are conceptualized as individual-level risk behaviors and biological responses that contribute to CVD and cancer health disparities independently and through interaction with social determinants of health at the individual-, area- and population-levels (Figure A.4). These risk behaviors and factors influence CVD and cancer outcomes through similar inflammatory pathways, highlighting the potential for cost-effective interventions to prevent CVD and cancer outcomes in the general population, as well as to reduce related health disparities to reach the aforementioned U.S. public health benchmarks.\textsuperscript{20, 21} However, a better
understanding of the distribution of these overlapping risk behaviors and factors among the populations experiencing disparate CVD and cancer outcomes is needed.\textsuperscript{4, 5}

The AHA’s LS7 of CVH metric includes four health behaviors (smoking, BMI, physical activity, and diet) and 3 health factors (total cholesterol, blood pressure and plasma glucose).\textsuperscript{5} Ideal CVH, defined as having ideal levels of each of the seven components, is associated with CVD and cancer outcomes.\textsuperscript{1, 7} Furthermore, the differentiation between well-defined ideal, intermediate and poor levels of each health behavior and factor makes this metric well-suited to quantify gradients in disease risk in the general population and monitor progress made toward public health benchmarks.\textsuperscript{1, 5, 7, 8}

Although LS7 can identify precisely where populations differ in terms of CVD and cancer risk, studies have typically only quantified this metric in a single population and not made comparisons between groups.\textsuperscript{12, 14, 170, 256}

By quantifying LS7 across diverse populations, this metric can also be used to prioritize specific modifiable CVD and cancer risk factors for future research and intervention to address the social determinants of health contributing to CVD- and cancer-related health disparities. This study focused on two relatively understudied populations facing CVD- and cancer-related health disparities: mainland Puerto Ricans and residents of Appalachia. Puerto Ricans, the second largest U.S. Hispanic subgroup, are greatly overshadowed in the epidemiologic literature by Mexican Americans, the country’s largest Hispanic subgroup.\textsuperscript{22} However, Puerto Ricans living in the U.S. experience worse CVD- and cancer-related outcomes and higher prevalence of acute and chronic medical conditions than non-Hispanic NHWs, Mexican Americans and
other Hispanic subgroups. Health disparities related to CVD and cancer have also been documented between island and mainland Puerto Ricans. Appalachia is a federally-designated region of the U.S. along the Appalachian mountains home to over 25 million residents. Differences in economic strength, sociodemographic characteristics and the availability of health-promoting resources between Appalachia and the remainder of the U.S. and within Appalachia likely contribute to documented CVD and cancer disparities between Appalachia and the U.S. and within the distinct region.

4.2.1 Purpose

This study sought to quantify the distribution of each LS7 component and an overall LS7 score within research samples from mainland Puerto Ricans and Appalachian residents and compare the distributions to one another and national data from NHANES. The three hypotheses for this study were that: 1) a large proportion of BPRHS and ACCN study participants would be categorized into the poor level of available LS7 behaviors and factors and have low overall LS7 scores; 2) BPRHS and ACCN study participants would be relatively similar in terms of the distribution of LS7 components, overall LS7 score categories, and in terms of mean overall LS7 score; and 3) there would be differences in the proportion of participants classified into poor, intermediate and ideal (or some other combination of levels due to data availability) between the NHW NHANES sample and each research sample.
4.3 Methods

4.3.1 Study Populations

4.3.1.1 The Boston Puerto Rican Health Study (BPRHS)

The BPRHS is an ongoing longitudinal study to examine the role of psychosocial stress on the presence and development of allostatic load and other health outcomes like CVD among Puerto Ricans living in the Greater Boston area. A detailed overview of the study population and design has been published elsewhere. In brief, from June 2004 to October 2009, a total of 2,170 Puerto Rican adults age 45-75 years were identified through door-to-door enumeration of randomly selected census blocks with 10 or more Hispanics within the target age range, community events, referrals, and calls to the study office. In addition to age, eligibility criteria included: self-identified Puerto Rican descent, able to answer questions in English or Spanish, a MMSE score greater than 10, living in the Boston, MA metropolitan area at the time of the study and not planning to move within the next two years. Of 1,802 eligible individuals, 1,500 (83.2%) completed the baseline interview, which took place after the initial screening contact, completion of written informed consent and neuropsychological testing to identify those who needed assistance from a proxy or were ineligible due to low MMSE score. The BPRHS protocol was approved through the IRB of Tufts University and Northeastern University.

Trained, bilingual interviewers administered baseline questionnaires and tests in the participant’s home. Age, education level, household income, migration, employment history, family size and food security information were collected using
questionnaires modeled after national surveys. Smoking frequency, history, and type of smoking were assessed and standing height and weight were measured in duplicate. Current physical activity was captured using a modified Paffenbarger questionnaire of the Harvard Alumni Activity Survey previously tested in an elderly Puerto Rican population. Dietary intake was assessed using a semi-quantitative 126-item FFQ revised to include appropriate foods and portion sizes. The modified FFQ was shown to capture intakes reported in 24-hour recalls more accurately than the NCI-Block FFQ format questionnaire after which it was modeled. This FFQ has been validated against dietary biomarkers in Hispanics aged ≥ 60 years. The interviewer collected detailed information on prescription and over-the-counter medications using medical bottles presented by participants. Systolic and diastolic blood pressures were measured in duplicate, at three time points during the interview, and the second and third readings were averaged. A day after completing the baseline questionnaire, a certified phlebotomist returned to participants’ homes to collect the urine and saliva samples and draw blood in the home. Blood samples were analyzed for plasma lipids and glucose.

4.3.1.2 Appalachia Community Cancer Network (ACCN) Study

The ACCN, a joint effort of the University of Kentucky, The Ohio State University, Pennsylvania State University, Virginia Tech University, and West Virginia University, is currently conducting a group randomized trial in a five-state area to test a previously piloted ACCN faith-based intervention focused on individual and environmental level changes to increase physical activity and improve healthy food choices using eHealth technology. Using county as the unit of randomization, 13
churches were randomly assigned to receive *Walk by Faith*, the dietary and physical activity faith-based intervention program, and 15 were randomly assigned to receive *Ribbons of Faith*, a comparison condition focused on cancer screening.²²⁷

The study was promoted in church bulletins and announcements from the pulpit. Group recruitment events and individual appointments were held at participating churches, at which time an individual could be screened to determine eligibility to participate. Eligibility criteria included: at least 18 years of age, attending services at least 4 times in the past two months at a participating church, able to understand and read English, cognitively able to provide informed consent, resident of an Appalachian county, not residing in a nursing facility or residential home, not planning to move away from the study area, willing to use a computer, no dietary restrictions prescribed for weight-loss or part of a formal weight-loss program, less than 400 pounds, having a BMI of at least 25, and, if female, not pregnant, breastfeeding or less than 9 months post-partum or planning to become pregnant during the study period. Questions modified from the physical activity readiness questionnaire (PAR-Q) were used to determine whether medical clearance from a physician was required to participate.²²⁹ In total, 427 *Walk by Faith* and 237 *Ribbons of Faith* participants completed the baseline questionnaires between 2012 and 2013. For the purpose of this study, the ACCN study population was treated as a cohort of 663 people. The ACCN protocol was approved through the IRBs of the Ohio State University, the University of Kentucky, Pennsylvania State University, Virginia Polytechnic Institute and State University and West Virginia University.

112
Smoking frequency, history, and type were assessed during the phone interview. Height and weight measurement protocols were adapted from the PhenX Toolkit version 4.5, July 29, 2011. Standing height was measured twice using a portable stadiometer. If the difference between the two measurements was greater than 0.1 cm, a third measurement was taken. The measurements were then averaged to determine height. Weight was measured twice using a calibrated portable scale. If the difference between the two measurements was greater than 0.1 kg, a third measurement was taken. The measurements were then averaged to determine weight. VioActive, a proprietary survey adapted from the international physical activity questionnaire (IPAQ) by Viocare, Inc., a privately-owned wellness systems technology company, was administered online during the in-person interview to assess physical activity. Viocare’s proprietary graphic FFQ, VioScreen, was also administered online during the in-person interview to assess diet. Systolic and diastolic blood pressure were measured twice on each arm using adapted protocol. Since biospecimens were not collected, total cholesterol and plasma glucose were not assessed. The use of blood pressure and glucose lowering medication was determined using responses to medical clearance questions.

4.3.1.3 National Health and Nutrition Examination Survey (NHANES)

NHANES encompasses a series of cross-sectional, survey-based studies that assess the health and nutritional status among U.S. adults and children using two data collection methods: interviews and physical examinations. NHANES employs a complex, multistage, probability design to select participants from geographically-defined strata (e.g. single or contiguous counties for which probability of selection is
proportional to size) and proportions of minority populations so that the resulting sample is nationally representative of the country’s civilian, non-institutionalized population. Demographic, socioeconomic, dietary, and health-related information is collected in the interview portion whereas medical, dental, and physiological measurements and laboratory tests are administered in the physical examination portion.

Continuous 2007-2010 NHANES data restricted to NHW participants were used for the purposes of this study. Use of cigarettes was assessed via questionnaire to determine smoking status; height and weight were measured during the clinical examination. Self-reported frequency and duration of specific moderate-intensity and vigorous-intensity activities over the past week or month and transportation and household activities were assessed via questionnaire. Participants completed two 24-hour dietary recalls administered by an interviewer. Blood lipids and fasting plasma glucose were determined from blood samples collected during the clinical examination. Three consecutive blood pressure readings (or four, if a measurement was interrupted or incomplete) were collected during the clinical examination. Medication use was assessed during the home interview.

### 4.3.2 Analyses

#### 4.3.2.1 Individual LS7 Components and Overall LS7 Score

Unless otherwise specified, baseline BPRHS and ACCN study data and NHANES data were used to derive individual LS7 components in line with the AHA definition (Table B.1).
4.3.2.1.1 Health Behaviors

For the BPRHS and ACCN study population, the definition of ideal, intermediate and poor smoking status differed slightly from the AHA definition because of the data available. Ideal smoking was defined as never or less than 100 cigarettes in one’s entire life, intermediate was defined as smoked in the past but not currently, and poor was defined as currently smoking. Only poor and intermediate levels of BMI (BMI ≥ 30 kg/m² and BMI between 25 and 30 kg/m², respectively) were estimated since ACCN participants had to have had a BMI of at least 25 to be eligible for the study. In terms of healthy diet score components among ACCN participants, daily fruit and vegetable servings were not based on cups but rather standard portion sizes, and four ounces was considered a standard serving of fish rather than three and a half ounces. Ideal, intermediate and poor sweets/sugar-sweetened beverages were derived from a variable of gram and teaspoon equivalents of added sugars among ACCN and NHANES participants, respectively. Given the categories available in the USDA’s Food Patterns and Equivalent Databases for 2007-2008 and 2009-2010, NHANES fish consumption had to be defined as at least two 3.5 oz servings of seafood high in n-3 fatty acids and seafood low in n-3 fatty acids.²⁴¹,²⁴²

4.3.2.1.2 Health Factors

Cholesterol and plasma glucose were not assessed among ACCN study participants. However, the use of glucose-lowering medication was derived from response to a medical clearance question regarding if a doctor prescribed drugs to control diabetes so that ACCN participants could be classified as ideal or not.²⁴³
4.3.2.1.3 Overall LS7 Score

For each study population, an overall LS7 score was calculated using the approach by Folsom et al. (2011) such that the overall LS7 score was based on the number of ideal LS7 metrics present at baseline.\textsuperscript{170} Among BPRHS and NHANES participants, there were 8 possible strata ranging from 0 ideal LS7 components to 7 ideal LS7 components. ACCN participants were required to have a BMI \(\geq 25\) to participate and were not subject to cholesterol or glucose testing. Consequently, there were 6 possible strata for the overall LS7 score ranging from 0 ideal LS7 components to 5 ideal LS7 components among ACCN study participants. Achievement of poor, intermediate and ideal LS7 components, as defined by the AHA (Table B.1) was calculated. Continuous variables were described using means (SDs), and categorical variables were described using percentages.

4.3.2.2 Comparison between BPRHS and ACCN

Given the differing eligibility criteria between the two studies, BPRHS and ACCN data were restricted to participants age 45-75 with a BMI of at least 25. Adjustment for age and sex to the 2000 U.S. standard population was performed using the direct method to foster comparison between the two research samples.\textsuperscript{244}

As summarized in Table B.4, for smoking, physical activity, diet and blood pressure, the proportion of participants from both studies who are classified into poor, intermediate, and ideal categories was calculated. For plasma glucose, the proportion of participants from both studies who are classified into not ideal and ideal categories was calculated. For BMI, the proportion of participants from both studies who were classified...
into poor and intermediate categories was calculated. For each study population, an overall LS7 score was also be calculated as described previously. There were 6 possible strata for the overall LS7 score ranging from 0 ideal LS7 components to 5 ideal LS7 components.\textsuperscript{170}

Age- and sex-adjusted estimates and standard errors of the estimates made it impossible to make compare BPRHS and ACCN participants using traditional contingency table analysis. Consequently, Wald tests for proportions were used with an understanding that a difference in the proportion of BPRHS and ACCN participants classified at one level suggested differences in the proportion of BPRHS and ACCN participants classified at other available level(s). Wald tests for proportions were not treated as separate, independent tests.\textsuperscript{245} For smoking, physical activity, blood pressure and diet, the age- and sex-adjusted proportion of participants classified into poor, intermediate and ideal levels were compared between the two research samples using three Wald tests for proportions based on the age- and sex-weighted frequencies.\textsuperscript{257} The proportion of participants classified into poor and intermediate BMI levels and the proportion of participants classified into ideal and not ideal plasma glucose levels were compared between the two research samples using one Wald test for proportions based on the age- and sex-weighted frequencies. Overall LS7 scores were examined categorically and continuously. The overall LS7 score was collapsed into three groups based on the distributions observed (i.e., 0-1, 2-3 and 4) and compared using three Wald tests for proportions based on the age- and sex-weighted frequencies. Difference in mean continuous overall LS7 scores for each research sample was assessed using a two sample
t test of the sex and age-adjusted mean for BPRHS and ACCN study participants. All statistical tests were two-sided. For the comparisons of overall LS7 scores and available LS7 components, overall Type I error was set to <0.05, and the Bonferroni multiple testing correction method was applied to determine the significance for an individual comparison (i.e., p value of 0.003).\textsuperscript{246} All analyses were performed using SAS version 9.3 (SAS Institute, Inc., Cary, NC) and Stata SE version 12 (StataCorp, College Station, Texas).

### 4.3.2.3 Comparisons between BPRHS and NHANES and ACCN and NHANES

BPRHS and ACCN study restrictions were applied separately to NHANES data for NHW participants to foster comparison with each research sample. Adjustment for age and sex to using the 2000 U.S. standard population was performed using the direct method.\textsuperscript{244} It was difficult to compare NHANES estimates to age and sex adjusted BPRHS or ACCN estimates and standard errors using traditional statistical testing given the NHANES’ complex, multistage sampling design, NHANES estimates and standard errors of the estimates, which accounted for age and sex adjustment, as well as survey weighting, clustering and stratification. Consequently, differences in the prevalence of poor, intermediate and ideal levels of each LS7 component (or some other combination of levels) were assessed by comparing overlapping 99\% CIs for sex and age-adjusted estimates of each component at each level. 99\% CIs were used to control type I error. Non-overlapping 99\% confidence intervals suggested a difference in the proportion of individuals in each study population categorized as poor, intermediate or ideal (or some other combination of levels) with an understanding that non-overlap at one level
suggested differences, at least qualitatively, in the proportion of individuals in each study at other available level(s).

Overall LS7 scores were collapsed into three groups based on the distributions observed.\textsuperscript{7} Differences in the proportion of NHW NHANES participants classified into each overall LS7 score category were assessed by comparing overlapping 99\% confidence intervals for each category. Non-overlapping 99\% confidence intervals suggested a difference in the proportion of individuals in each study population categorized at a given category with an understanding that non-overlap at one categorical level suggested differences, at least qualitatively, in the proportion of individuals in each study at other available categorical levels. Differences in mean overall LS7 scores for each research sample were assessed by comparing overlapping 99\% confidence intervals. Non-overlapping 99\% confidence intervals suggested a difference in mean overall LS7 scores between the study populations. All analyses were performed using SAS version 9.3 (SAS Institute, Inc., Cary, NC) and Stata SE version 12 (StataCorp, College Station, Texas).

4.4 Results

4.4.1 Individual LS7 Components and Overall LS7 Score

4.4.1.1 Boston Puerto Rican Health Study (BPRHS)

Of 1,500 baseline BPRHS participants, there were 1,071 (71.4\%) with complete LS7-related data of whom 71.7\% percent were female and among whom the average age was 57.1 years (SD: 7.6). The sex- and age-unadjusted prevalence of individual LS7 components among BPRHS participants at baseline are visually summarized in Figure
A.9. In terms of the proportion of BRPHS participants that were classified as poor for each component, 68.1% had poor diet, 57.9% had poor BMI, 25.6% had poor blood pressure, 24.6% had poor plasma glucose, 23.8% had poor smoking status, 23.6% had poor physical activity and 8.7% had poor total cholesterol. Out of a possible score of 7, overall LS7 score ranged from 0 to 6, with an average score of 2.46 (SD: 1.30).

4.4.1.2 Appalachia Community Cancer Network (ACCN) Study

Of 663 baseline ACCN study participants, there were 420 (63.3%) with complete LS7-related data of whom 71.9% percent were female and among whom the average age was 53.4 years (SD: 12.6). The sex- and age-unadjusted prevalence of individual LS7 components among ACCN participants at baseline are visually summarized in Figure A.10. In terms of the proportion of ACCN participants that were classified as poor for each component, 65.0% had poor BMI, 59.5% had poor diet, 15.5% had poor blood pressure, 10% had poor plasma glucose, 2.6% had poor smoking status and 1.2% had poor physical activity. Out of a possible score of 5, overall LS7 score ranged from 0 to 4, with an average score of 2.4 (SD: 0.7).

4.4.1.3 National Health and Nutrition Examination Survey (NHANES)

There were 1,128 NHW participants in NHANES 2007-2010, representing 71,819,035 civilian, non-institutionalized U.S. NHWs. The majority of participants were female (56.1%) and the average age was 50.3 years (95% CI: 49.0 - 51.6). The sex- and age-unadjusted prevalence of individual LS7 components among NHW who participated in NHANES between 2007 and 2010 are visually summarized in Figure A.11. In terms of the proportion of NHW NHANES participants that were classified as poor for each
component, 67.4% had poor diet, 32.1% had poor BMI, 25.1% had poor physical activity, 20.5% had poor blood pressure, 16.6% had poor total cholesterol, 16.5% had poor smoking status and 8.3% had poor plasma glucose. Out of a possible score of 7, overall LS7 score ranged from 0 to 6, with an average score of 2.79 (95% CI: 2.61-2.96).

4.4.2 Comparison between BPRHS and ACCN

BPRHS and ACCN study data were restricted to participants age 45-74 with a BMI of at least 25 and age- and sex-adjusted to the 2000 U.S. standard population using the direct method, resulting in a sample of 934 BPRHS and 313 ACCN study participants. As summarized in Figure A.12 and Table B.5, the differences in the proportion of participants categorized as having ideal and poor smoking status were significant with more ACCN study participants classified as having ideal smoking status (Wald-statistic: 54.34, DF: 1, p<0.003), but more BPRHS participants classified as having poor smoking status (Wald-statistic: 95.86, DF: 1, p<0.003). There were no significant differences in the proportion of participants categorized as intermediate (or poor) BMI at α=0.003 for an individual comparison. For physical activity, the differences in the proportion of participants categorized as intermediate and poor were significant such that there were more ACCN study participants classified as having intermediate physical activity (Wald test statistic: 64.79, DF: 1, p<0.003), whereas there were more BPRHS participants classified as having poor physical activity (Wald-statistic: 178.73, DF: 1, p<0.003).
Differences in the proportion of participants classified as having intermediate and poor healthy diet score were significant. There were more ACCN study participants classified as having intermediate diet (Wald-statistic: 10.16, DF: 1, p<0.003), but more BPRHS participants classified as having poor diet (Wald-statistic: 11.65, DF: 1, p<0.003). For blood pressure, the differences in the proportion of participants categorized as ideal and intermediate were significant with more BPRHS participants classified as having ideal blood pressure (Wald-statistic: 322.40, DF: 1, p<0.003), and more ACCN study participants classified as having intermediate blood pressure (Wald-statistic: 254.11, DF: 1, p<0.003). In terms of plasma glucose, there were significant differences in the proportion of participants categorized as ideal and not ideal such that more ACCN participants classified as having ideal plasma glucose (Wald-statistic: 918.66, DF: 1, p<0.003), whereas there were more BPRHS participants classified as having not ideal plasma glucose (Wald-statistic: 918.66, DF: 1, p<0.003).

The proportion of participants classified into each overall LS7 score categorical level was compared using three Wald tests for proportions based on the age- and sex-weighted frequencies at each available level. There were significant differences in the proportion of participants with an overall LS7 score of 0 to 1 and of 2 to 3 with more BPRHS participants having an overall LS7 of 0 to 1 (Wald-statistic: 85.83, DF: 1, p<0.003), but more ACCN study participants with an overall LS7 score of 2 to 3 (Wald-statistic: 92.47, DF: 1, p<0.003). The mean overall LS7 score was 1.90 (SD: 0.04) and 2.42 (SD:0.04) among BPRHS and ACCN study participants, respectively, a difference that was statistically significantly (t-statistic: 7.38, DF: 1245, p<0.003).
4.4.3 Comparison between BPRHS and NHANES

BPRHS and 2007-2010 NHANES data age were restricted to participants 45 to 74 years of age, resulting in a sample of 1,067 BPRHS participants and 587 NHW 2007-2010 NHANES participants (representing 37,662,918 civilian, non-institutionalized U.S. NHWs). As summarized in Figure A.13 and Table B.6, the 99% confidence intervals did not overlap between the two study populations at any of the three smoking levels, suggesting the proportion of individuals categorized each level was different between BPRHS and NHW NHANES participants. Of note, the prevalence of ideal smoking status was 82.6% (99% CI: 76.1-89.2%) among NHW NHANES participants but only 43.0% among BPRHS participants (99% CI: 38.8-47.3%). In terms of BMI, 99% CI overlap between the two study populations was limited to the proportion of study participants classified as having intermediate BMI. The proportion of participants classified as having ideal BMI was higher among NHW NHANES participants while the proportion of participants classified as having poor BMI was higher among BPRHS participants (54.2, 99% CI: 50.0-58.6%). For physical activity, there was overlap between the 99% CIs at the poor level, but the proportion of individuals classified as having ideal physical activity was higher among BPRHS participants (74.2%, 99% CI: 70.5-77.9%), and the proportion of individuals classified as having intermediate physical activity was higher among NHW NHANES participants (24.7%, 99% CI: 18.5-30.8%). There was complete overlap of 99% CI in terms of healthy diet score, and in both study populations the prevalence of ideal diet was just 0.4%. The 99% CIs for ideal, intermediate and poor blood pressure also overlapped between the two study populations with almost half of the
participants in each study having ideal blood pressure. In terms of total cholesterol, 99% CI overlap between the two study populations was limited to the intermediate level. The proportion of participants classified as having ideal total cholesterol was higher among BPRHS participants (38.5%, 99% CI: 34.3-42.7%), but the proportion of participants classified as having poor total cholesterol was higher among NHW NHANES participants (19.1%, 99% CI: 15.2-23.1%). For plasma glucose, overlap between 99% CIs occurred only at the intermediate level. The proportion of participants classified as having ideal plasma glucose was higher among NHW NHANES participants (43.7%, 99% CI: 36.8-50.6%), and the proportion of participants classified as having poor plasma glucose was higher among BPRHS participants (23.9%, 99% CI: 20.3-27.6%).

There was overlap between the 99% CIs for each category of overall LS7 score (i.e., 0-2, 3-4 and ≥5) and for mean overall LS7 score, suggesting the overall LS7 score distribution did not differ substantially between the two study populations.

4.4.4 Comparison between ACCN and NHANES

ACCN and NHANES data were restricted to participants age 18-84 with a BMI of at least 25 and age- and sex-adjusted to the 2000 U.S. standard population using the direct method, resulting in a sample of 417 ACCN study participants and 925 NHW NHANES participants (representing 57,517,123 civilian, non-institutionalized U.S. NHWs). As summarized in Figure A.14 and Table B.7, the 99% confidence intervals overlapped between the two study populations at only the ideal level of smoking, whereas the proportion of participants classified as having intermediate smoking status was higher among ACCN study participants (23.7%, 99% CI: 18.7-28.8%), and the proportion of
participants classified as having poor smoking status was higher among NHW NHANES participants (19.9%, 99% CI: 16.3-23.5%). In terms of BMI, the proportion of participants classified as having intermediate BMI was higher among NHW NHANES participants (55.2%, 99% CI: 49.7-60.7%). For physical activity, there was overlap between the 99% CIs at the intermediate level, but the proportion of individuals classified as having ideal physical activity was higher among ACCN study participants (70.3%, 99% CI: 63.3-77.4%) and the proportion of individuals classified as having poor physical activity was higher among NHW NHANES participants (25.5%, 99% CI: 20.7-30.3%).

There was complete overlap of 99% CI in terms of healthy diet score, and in both study populations the prevalence of ideal diet was less than 1%. A slightly higher proportion of ACCN participants were classified as having intermediate diet and a slightly higher proportion of NHW NHANES participants were classified as having poor diet. The 99% CIs for poor blood pressure also overlapped between the two study populations, but the proportion of participants classified as having ideal blood pressure was higher among NHW NHANES participants (60.0%, 99% CI: 55.1-64.8%), whereas the proportion of participants classified as having intermediate blood pressure was higher among ACCN study participants (76.9, 99% CI: 68.8-85.0%). There was no overlap between the 99% CIs for the two available levels of plasma glucose between the two study populations. The proportion of participants classified as having ideal plasma glucose was higher among ACCN study participants (90.4%, 99% CI: 86.3-94.6%), and the proportion of participants classified as having not-ideal plasma glucose was higher among NHW NHANES participants (50.5%, 99% CI: 43.9-57.1%).
At the three levels of categorical overall LS7 score (i.e., \( \leq 1 \), 2-3 and 4), there was no overlap between the 99% CIs, suggesting the overall LS7 score distribution differed substantially between the two study populations. More NHW NHANES participants had an overall score of 4, more ACCN study participants had an overall score of 2 to 3, and more NHANES participants had an overall score of \( \leq 1 \). However, there was overlap between the 99% CIs for mean overall LS7 score, suggesting the mean overall LS7 score did not differ substantially between the two study populations.

4.5 Discussion

This study sought to quantify the distribution of each LS7 component and an overall LS7 score within research samples from two populations facing CVD and cancer-related health disparities, mainland Puerto Ricans and residents of Appalachia, and compare the distributions to one another and national data. Although it was hypothesized that a large proportion of BPRHS and ACCN study participants would be categorized into the poor level of available LS7 behaviors, diet and BMI were the only components for which the majority of BPRHS and ACCN study participants were categorized into the poor level. However, overall LS7 scores were low among both groups, as hypothesized. It was hypothesized that BPRHS and ACCN study participants would be relatively similar in terms of the distribution of LS7 components, overall LS7 score categories, and in terms of mean overall LS7 score, but there were statistically significant differences between the BPRHS and ACCN study participants in terms of the distribution of all available age- and sex-adjusted LS7 components, excluding BMI. Whereas BPRHS participants had a more favorable blood pressure distribution, ACCN participants had
more favorable distributions of smoking, physical activity, diet, and plasma glucose. ACCN participants also had a more favorable categorical overall LS7 score distribution and a higher average continuous overall LS7 score.

Lastly, it was hypothesized that there would be differences in the proportion of participants classified into poor, intermediate and ideal levels (or some other combination due to data availability) between the NHW NHANES sample and each research sample. Excluding diet, there were significant differences between BPRHS and NHW NHANES participants across at least one level of remaining LS7 components. BPRHS participants had more favorable distributions of physical activity and total cholesterol, whereas NHANES participants had more favorable distributions of smoking, BMI, and plasma glucose. However, the distribution of categorical overall LS7 score and mean continuous overall LS7 scores did not differ substantially between the study populations. Although there was complete overlap of 99% CIs for all levels of diet, there were significant differences between ACCN study and NHW NHANES participants across at least one level of remaining LS7 components and across categorical overall LS7 score. ACCN participants tended to have more favorable distributions of smoking, physical activity and plasma glucose, and NHANES participants had more favorable distributions of BMI and blood pressure. The distribution of categorical overall LS7 score appeared to differ substantially between the two study populations, but the mean continuous overall LS7 score did not, likely because there were more NHW NHANES participants with overall LS7 scores ≥4 and ≤1 while ACCN participants tended to have overall LS7 scores of 2 and 3.
The age- and sex-adjusted prevalence estimates obtained for NHW 2007-2010 NHANES participants are in line with published age- and sex-stratified estimates for NHW 2003-2008 NHANES participants.\textsuperscript{14} However, given the lack of LS7-related research comparing the distribution of LS7 components and an overall LS7 score across research populations, including mainland Puerto Ricans and residents of Appalachia, it is difficult to compare and contrast findings of the study pertaining to BPRHS and ACCN study participants to the current literature. However, the unadjusted prevalence of LS7 components and overall LS7 score among BPRHS participants can be compared to findings from a study of LS7 among the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) participants.\textsuperscript{256} Compared to BPRHS participants, HCHS/SOL participants had a higher prevalence of ideal smoking status (76.6%), ideal plasma glucose (68.9%), ideal blood pressure (53.4%), ideal BMI (23.0%) and ideal diet (2.5%), but a lower prevalence of ideal physical activity (51.3%). The prevalence of ideal cholesterol among HCHS/SOL participants was not reported.\textsuperscript{256}

Similarly, the unadjusted prevalence of LS7 components and overall LS7 score among ACCN study participants can be compared to findings from a study of LS7 among 4,754 rural Minnesota adults screened for CVD through the Heart of New Ulm (HONU) Project.\textsuperscript{12} Compared to ACCN study participants, HONU project participants had a higher prevalence of ideal smoking status (90.0%), ideal plasma glucose (68.9%), ideal blood pressure (53.4%), intermediate BMI (34.8%), and ideal blood pressure (30.9%), but a lower prevalence of ideal physical activity (63.7%). Of healthy diet score components, only fruit and vegetable consumption was assessed among HONU project
participants, of whom 16.4% met the daily recommendation, which is slightly less than the proportion among ACCN study participants (18.6%).

As demonstrated by this study, research samples from two populations that face CVD and cancer health disparities are not necessarily similar in terms of the distribution of mutual CVD and cancer risk factors, and differences between a research sample from a population facing health disparities and nationally representative research sample of NHWs are not always in the expected direction. These finding may be due to inherent dissimilarities between research samples, due to geographic location, culture and exposure to other social determinants of health, and the fact that BPRHS and ACCN study participants are research samples and not representative of the target populations. However, the results of this study are nevertheless useful since quantification of the LS7 metric in BPRHS, the ACCN study and NHW NHANES participants demonstrates precisely where these populations differ in terms of CVD and cancer risk and can be used to direct future research. For example, since BPRHS participants fared worse than ACCN study participants and NHW NHANES participants in terms of smoking, BMI, and plasma glucose and ACCN study participants fared worse than BPRHS and NHW NHANES participants in terms of BMI and blood pressure, these behaviors and factors should be prioritized for future research and intervention. Diet should also be prioritized in all populations given the universal low prevalence of ideal diet observed and since diet was the leading cause of the overall U.S. disease burden.
4.5.1 Strengths

To date, the majority of studies that have quantified the LS7 metric have done so in a single population without making comparisons to other groups.\textsuperscript{12, 14, 170, 256} A strength of this study is that it quantified the LS7 metric in research samples from two populations facing health disparities and compared the distribution of age- and sex-adjusted LS7 components and categorical overall LS7 score and mean continuous overall LS7 between the two groups so that similarities and differences between the two health disparity population research samples could be explored. Additionally, this study compared each research sample from a population facing health disparities to a research sample representative of civilian, non-institutionalized U.S. NHW men and women of similar age, thereby placing results in a national context.

In addition to being able to compare and contrast one research sample to another, this study’s approach allowed the differences between the research samples to be explored at a fine level of detail, given the ideal, intermediate and poor levels (or some other combination of levels based on data availability) of the mutual, modifiable CVD and cancer risk factors captured by the LS7 metric. As demonstrated by this study, the LS7 metric identified precisely where in BPRHS, ACCN study and NHW NHANES participants differ in terms of CVD and cancer risk.

4.5.2 Limitations

By using cross-sectional data, this study is limited in that it cannot speak to changes in the distribution of LS7 components or overall LS7 score within a research
sample or between research samples over time, or causality. The results of this study are also limited in terms of generalizability given the research samples from populations facing health disparities employed. Although NHW NHANES participants were representative of civilian, non-institutionalized U.S. NHWS, BPRHS participants were 45 to 75 years of age and were recruited using approaches such as door-to-door enumeration of randomly selected census blocks with 10 or more Hispanics within the target age range. ACCN study participants regularly attended churches located in select Appalachian counties in a five-state area and had to have a BMI of at least 25 to participate. Study participants had higher education and income levels and a lower prevalence of current smokers than what was reported previously among the Appalachian population as a whole.\textsuperscript{258} Therefore, BPRHS and ACCN study participants are not representative of all mainland Puerto Ricans and all Appalachian residents in the U.S.

BPRHS participants were homogenous in terms of geography given residence in the Greater Boston area. However, ACCN participants spanned multiple states and sub-regions of Appalachia known to differ in terms of education, poverty, race/ethnicity and access to healthcare.\textsuperscript{30, 31} Although the prevalence of LS7 components were not stratified by state of residence among ACCN participants, a sensitivity analysis was conducted to examine state-specific mean overall LS7 score, unadjusted for age and sex. There did not appear to be meaningful differences between participants who lived in Kentucky (2.4, SD: 0.8), Ohio (2.5, SD: 0.7), Pennsylvania (2.4, SD: 0.7), Virginia (2.4, SD: 0.7) and West Virginia (2.5, SD: 0.7).
In addition to adding data restrictions to foster comparisons between the diverse research samples, this study excluded BPRHS and ACCN study participants for whom LS7-related data was missing, which impacted sample sizes. Furthermore, this study is limited in that comparisons were made using data that were not collected uniformly, and modifications to the AHA’s LS7 definition for various components, which may have induced information bias. For example, because biospecimens were not collected as part of the ACCN study protocol, participants could only be categorized as ideal or not ideal based on their response to a question concerning use of diabetes controlling medication that was collected as part of the ACCN study protocol. Since the distribution of ideal and not ideal plasma glucose among ACCN study participants differed greatly from that of BPRHS and NHW NHANES participants, the difference in data completeness in terms of laboratory-confirmed plasma glucose levels may lead to misclassification of ACCN participants. When overall score was limited to the four components for which there were complete data among ACCN study participants (i.e., omitting BMI and glucose as well as total cholesterol), there were statistically significant differences in mean overall LS7 score that differed from the results of the primary analysis, such that mean overall LS7 score was higher among BPRHS participants (1.6, 99% CI: 1.6-1.7) than ACCN study participants (1.5, 99% CI: 1.4-1.6, t-statistic: -2.4, DF: 1245, p<0.02). Additionally, unlike before, there was not overlap between the 99% CI for mean overall LS7 score between ACCN and NHW NHANES participants, suggesting the overall LS7 score was higher among NHW NHANES participants (1.9, 99% CI: 1.8-1.9) than ACCN study participants (1.5, 99% CI: 1.4-1.6).
Another limitation was that diet and physical activity behaviors were self-reported by participants across all three studies, which likely caused participants to overestimate their dietary intake and physical activity participation (and lead to misclassification). Misclassification of diet and physical activity due to participant self-report was likely dissimilar across the BPRHS, ACCN study and NHANES. For example, compared to BPHRS participants who enrolled in a longitudinal cohort study of allostatic load and NHANES participants who enrolled in a cross-sectional study that broadly assessed health and nutritional status, overestimation of physical activity was likely highest among ACCN study participants who enrolled in a faith-based group randomized trial to improve physical activity and diet. Due to differential misclassification, the results of comparisons of physical activity between ACCN study participants and the other research samples may have been biased away from the null.

Similarly, although the definitions of ideal, intermediate and poor levels of physical activity were not modified for any of the studies, each study collected physical activity data in a slightly different manner. Among all activities completed the previous week, BPRHS participants were asked how much time they spent in hours on a usual weekday participating in vigorous activity and moderate physical activity and how much time they spent in hours on a usual weekend day participating in vigorous activity and moderate physical activity. However, ACCN participants were separately asked how many days during a typical week did they do vigorous physical activity as part of their work, in the garden or yard, for recreation or exercise and moderate physical activity as part of their work, in the garden, for recreation or exercise, and inside their home. For
each type of vigorous and moderate physical activity a participant reported taking part in at least one day a week, the hours and minutes spent doing that activity were recorded. Conversely, NHANES participants were asked if their work involves vigorous-intensity for at least 10 continuous minutes, how many days they do vigorous-intensity activities as part of their work in a typical week, and the amount of time spent doing vigorous-intensity activities at work on a typical day in hours or minutes. This process was repeated for moderate-intensity activities as part of their work and vigorous- and moderate-intensity sports, fitness or recreational activities. Differences between BPRHS, ACCN study and NHANES participants in terms of the distribution of ideal, intermediate and poor levels of LS7-defined physical activity may partially be an artifact of the different questionnaires.

4.6 Conclusion

Using LS7, this study identified precisely where CVD and cancer risk differs between research samples from two populations facing health disparities and between each research sample and a nationally representative sample of NHWs. This approach can be used to prioritize specific modifiable CVD and cancer risk factors for future research and intervention and track progress toward the elimination of health disparities among underserved populations. In addition to risk behaviors and biological factors, this approach took into account place, culture, age and sex, social determinants of health that may contribute to disparate CVD and cancer health outcomes. An important next step in this line of inquiry is to integrate the LS7 metric with data from electronic health records, the U.S. Census and North American Industry Classification System (NAICS) business
codes (commonly used to determine access to health promoting resources such as grocery stores, hospitals, health clubs and recreational sports centers) to better understand the role of multi-level social determinants of CVD and cancer disparities (Figure A.4). This information is needed to inform intervention strategies that target relevant social determinants of health within the populations facing health disparities in order to shift the distribution of CVD and cancer risk in a more favorable direction among these high-risk groups.\textsuperscript{18, 19}
5.1 Abstract

Background: Census tract (CT)-level geographic food access measures are often used as a proxy for individual-level geographic food access in urban areas. These measures ignore culturally relevant facilitators and barriers of food purchasing and consumption (i.e., informational food access), which may modify the geographic food access and healthy diet relationship among ethnic minority populations known to live in predominately urban areas. This study explored the relationship between CT-level geographic food access and diet and effect measure modification (EMM) of this relationship by culturally-relevant forms of informational food access among Puerto Ricans, a relatively understudied ethnic minority population despite being the second largest U.S. Hispanic subgroup. Puerto Ricans experience disparate burdens of cardiovascular disease (CVD), cancer, and mutual modifiable risk factors, such as diet.\textsuperscript{23, 24, 45, 144}

Methods: Using data from the Boston Puerto Rican Health Study (BPRHS), generalized estimating equations (GEE) were used to yield population averaged models of the association between a publicly-available CT-level geographic food access metric and diet, as measured by \textit{Life’s Simple 7} (LS7) healthy diet score, a composite score based on intake of fruit/vegetable, fish, sodium, sweets/sugar-sweetened beverages and whole
grains. Confounding by sociodemographic characteristics identified by a directed acyclic graph (DAG) and EMM by culturally-relevant sources of informational food access of the relationship between geographic food access and diet were assessed.

Results: In the primary analysis, an absolute geographic food access metric representing low grocery store or large supermarket access at the CT-level was not associated with a significant change in the odds of a poor LS7 healthy diet score (OR: 1.20, 95% CI: 0.88-1.64, p-value: 0.25). In the secondary analysis, a relative geographic food access metric representing the balance of healthy and unhealthy food available at the CT-level was not associated with a significant change in the odds of a poor LS7 healthy diet score at $\alpha=0.05$ (OR: 0.99, 95% CI: 0.96-1.03, p-value: 0.63). The final primary and secondary models were unadjusted given a lack of confounding and EMM.

Discussion: Using publicly-available CT-level exposure measures and LS7 healthy diet score, the current study did not find a significant relationship between area-level geographic food access and healthy diet among a research sample of older, urban mainland Puerto Ricans. No confounding and EMM were evident, despite using an explicit method to identify potential confounders and exploring culturally relevant forms of informational food access suggested by the literature to potentially modify the geographic food access and diet relationship. Given that this is an expanding area of scientific inquiry for which mainland Puerto Ricans have been understudied, the DAG used by this study and the reported effect estimates can be used to inform future research on the relationship between geographic food access and diet among research samples that are more representative of this population facing health disparities, and incorporating
individual- as well as area-level geographic food access data to further examine the impact of informational food access.
5.2 Introduction

Geographic food access, distance-based facilitators and barriers of food purchasing and consumption, is a characteristic of the physical context in which an individual lives that has the potential to contribute to CVD and cancer health disparities through diet, a mutual, modifiable chronic disease risk behavior at the individual-level (Figure A.4). Geographic food access has received increased empirical attention due to the burgeoning use of geographic information systems (GIS) and incentives at the federal-, state- and municipal-level to encourage improved access to food in underserved areas. However, the literature lacks consistency regarding the relationship between GIS-based geographic food access measures and diet, yielding null associations or significant associations in mixed directions. CTs, small, statistical county (or equivalent entity) subdivisions composed of somewhat socially and economically homogeneous populations bound by visible features (e.g. streets, roads, streams, railroad tracks) ranging from 1,200 to 8,000 people in size, are common units of analysis to capture area-level geographic food access in the U.S., particularly in urban areas. In the absence of individual-level data, it may be appropriate to use CT-based geographic food access measures as a proxy for individual-level geographic food access in urban areas.

There are a number of publicly-available CT-level geographic food access measures that, although convenient, fail to address the importance of educational, cultural and social facilitators and barriers (i.e., informational food access) relevant to the population being studied. For example, immigrant enclaves, neighborhoods with high proportions of immigrants, possibly of the same ethnicity, are not reflected in readily
available CT-level geographic food access measures. Immigrant enclaves have the potential to modify the relationship between intermediate-level geographic food access and diet through numerous pathways, such as ethnic food stores carrying culturally appropriate foods. Previous research has found that living in a CT with a higher proportion of immigrants born in Latin America was associated with lower high-fat food consumption, controlling for multi-level confounders like age, gender, income, education, neighborhood poverty and acculturation.

Acculturation encompasses changes in attitudes, values, behaviors and cultural identity that an individual or group may experience upon entering a new and different cultural context. Of various individual-level acculturation measures, language use is most closely aligned with health and health behaviors of Hispanics living in the U.S. For example, language is associated with dietary habits, possibly because language influences knowledge and understanding of available healthful foods and dietary information.

Despite being the second largest Hispanic subgroup in the U.S., constituting 50.7 million people in 2010, Puerto Ricans are relatively understudied. Compared to non-Hispanic Whites (NHWs), Mexican Americans, the nation’s largest Hispanic subgroup, and other Hispanic subgroups, Puerto Ricans experience disparate burdens of CVD, cancer, and mutual modifiable risk factors, such as diet. Health disparities between island Puerto Ricans and mainland Puerto Ricans have also been documented, such that island Puerto Ricans fare better than their mainland counterparts. Unlike the relationship between acculturation and diet, the relationship between geographic food
access and diet, and the degree to which this is modified by acculturation, has not yet been explored in this population.\textsuperscript{249}

5.2.1 Purpose

Since it is unclear whether area-level and individual-level indicators of acculturation modify the relationship between area-level geographic food access and individual-level diet, this study sought to quantify the association between a CT-level geographic food access metric and individual-level LS7 healthy diet score in a research sample of an urban, mainland Puerto Ricans, controlling for individual- and CT-level social determinants of health identified as potential confounders by a DAG and assessing EMM by area-level immigrant enclave and individual-level linguistic acculturation status.

5.3 Methods

5.3.1 Study Population

The Boston Puerto Rican Health Study (BPRHS) is an ongoing longitudinal study of the role of psychosocial stress on the presence and development of allostatic load and other health outcomes like CVD among mainland Puerto Ricans living in the Greater Boston area. A detailed overview of the BPRHS population and design has been published elsewhere.\textsuperscript{45} In brief, from June 2004 to October 2009, a total of 2,170 Puerto Rican adults age 45-75 years were identified through door-to-door enumeration, community events, referrals, and calls to the study office. In addition to age, eligibility criteria included: self-identified Puerto Rican descent, able to answer questions in English
or Spanish, a Mini Mental State Exam (MMSE) score greater than 10, living in the Boston, Massachusetts metropolitan area at the time of the study and not planning to move within the next two years. Of 1,802 eligible individuals, 1,500 (83.2%) completed the baseline interview. The protocols for the BPRHS and this study were approved by the institutional review boards (IRBs) of Tufts University and Northeastern University. This study was also approved by The Ohio State University IRB.

5.3.2 Outcome of Interest

The outcome of interest was a healthy diet score, as defined by the American Heart Association’s (AHA’s) LS7 of cardiovascular health (CVH) metric, which includes well-defined poor, intermediate and ideal levels of four health behaviors (smoking, BMI, physical activity, and diet) and 3 health factors (total cholesterol, blood pressure and plasma glucose). LS7’s diet component is a composite score of five components worth one point each (≥4.5 servings of fruits/vegetables per day, ≥2 servings of fish per week, <1500 mg of sodium per day, ≤450 kcal (36 oz) per week of added sugars (sweets/sugar-sweetened beverages) and ≥3 servings per day of whole grains). An ideal diet was defined as a score ≥4, whereas intermediate diet was defined as a score between 3 and 2, and a score less than 2 was considered poor.

The LS7 healthy diet scores of BPRHS participants were derived from a semi-quantitative 126-item FFQ revised to include appropriate foods and portion sizes that has been validated against dietary biomarkers in Hispanics aged ≥60 years. Since previous research found the prevalence of poor LS7 healthy diet score among BPRHS
study participants to be 68.1%, LS7 healthy diet score was dichotomized into poor (>2) or not poor (≤2) for the purpose of this study.\textsuperscript{259}

5.3.3 Exposure of Interest

The USDA variable representing whether an urban CT is considered as having low food access using a 0.5-mile distance was used as the exposure of interest. This variable was used since BPRHS data were geocoded to the 2010 U.S. decennial census CT definition, all CTs met the US Census urban definition, and very few CTs met the USDA’s food desert criteria. A CT was considered urban if its centroid was within an urban area, defined by the U.S. Census as having 50,000 or more people, or within an urban cluster of at least 2,500 people and less than 50,000 people. The USDA defined a food desert as CTs that were low income in which a substantial number or percentage of residents had low access to a supermarket or large grocery. CTs were considered low-income if the poverty rate was at least 20% or the median family income was 80% or lower than that of the state or metro-area in which it was located. Low access was defined differently for urban and rural CTs. Urban CTs were considered low access if 500 or more people or 33% of those residing in the CT lived a half mile or more from the nearest supermarket, supercenter or large grocery store.\textsuperscript{193} This variable was derived by the USDA from a directory of supermarkets, supercenters and large grocery stores within the U.S. created by merging a 2010 directory of stores authorized to accept SNAP benefits and the 2010 Trade Dimensions TDLinx directory of stores.

In brief, per USDA methods, stores must have reported at least $2 million in annual sales and contain all the major food departments found in a traditional
supermarket (e.g. fresh meat and poultry, dairy, dry and packaged foods and frozen foods) to be included in the measure. The Census Bureau’s urban area definition was used to designate urban status. Stores were then geocoded to the street address. Census block-level population data from the 2010 Census of Population and Housing were aerially allocated to ½-kilometer-square grids across the country. Distance was determined by calculating the distance from the geographic center of each ½-kilometer-square grid cell to the center of the grid cell containing the nearest store.195

5.3.4 Effect Measure Modifiers of Interest

Based on the literature, individual and CT-level indicators of acculturation were chosen a priori for exploration as EMMs of the relationship between CT-level USDA low access status and individual-level LS7 diet score.247-249 Linguistic acculturation, an indicator of acculturation at the individual-level, was derived from a questionnaire adapted from the BAS.223-225 The BAS yields two scores which rank acculturation in the Hispanic and the non-Hispanic domains.225 The linguistic acculturation score ranged from 0% (fully unacculturated, speaking Spanish only) to 100% (fully acculturated, speaking fluent English) and was treated as a binary variable (<50 vs. ≥50). Because it is unclear which is more closely associated with diet among mainland Puerto Ricans, the percent of the total population that was Hispanic/Latino and the percent of the total population that was Puerto Rican were explored as EMMs. These measures were derived from the 2010 decennial U.S. Census data available from Social Explorer and treated as continuous variables.238
5.3.5 Potential Confounders

A DAG was used to determine whether a number of available BRPHS and U.S. Census covariates meet criteria for confounding (Figure A.5). To be considered a potential confounder, a covariate must (1) have been associated with LS7 healthy diet score; (2) have been associated with USDA low access status and (3) have not been affected by USDA low access status.\(^\text{181}\) Plausible causal relationships (represented by unidirectional arrows) were determined on the basis of extant literature and biological plausibility. Potential confounders considered for model building must (1) have had an arrow pointing in either direction between it and LS7 healthy diet score; and (2) have had an arrow pointing from it to USDA low access status. Arrows pointing from USDA low access status would have suggested the covariate was on the causal pathway between USDA low access status and LS7 healthy diet score. Thus treating the covariate as a confounder may have induced bias.\(^\text{250}\) Even after meeting three requirements, a variable was removed from the DAG if controlling for another potential confounder would block the direct pathway between the variable and LS7 healthy diet score and/or the exposure of interest.\(^\text{250}\)

Since individual-level food security may be on the causal pathway between the USDA low access status and LS7 healthy diet score, it was not included in the final DAG (Figure A.6). Although they met the requirements to be considered as a potential confounder, individual-level education and insurance status were also left out of the final DAG since controlling for individual-level poverty blocked the direct pathways between these variables and USDA low access status. As shown in Figure A.6, individual-level
potential confounders were age, sex and marital status. Poverty was a potential confounder at the level of the individual and CT.

Age at baseline was treated as a continuous variable and sex was treated as a binary variable. Marital status was treated as a binary variable: married/living as married/spouse in household or not. Individual-level poverty was treated as a binary variable and was derived from family size, total household income, and poverty thresholds by size of family obtained from the U.S. Census website for the year in which data were collected. The indicator of CT-level poverty was treated as a continuous variable. CT-level percent of population with a poverty-to-income ratio (PIR) below the poverty level was derived from five-year 2007-2011 AmCS population estimates of the ratio of 2009 family or unrelated individual income to the appropriate poverty threshold. A ratio below 1.00 indicates income below the poverty level.

5.3.6 Analyses

5.3.6.1 USDA Low Access Primary Analysis

Descriptive statistics were used to provide overall sample characteristics. The method of GEE yielded a population averaged model of the association between USDA low access status and LS7 healthy diet score with robust SEs. Bernoulli was specified as the family (the assumed distribution of LS7 healthy diet score), logit was specified as the link function, and a working exchangeable correlation structure was specified to reflect the average dependence among correlated LS7 healthy diet scores among participants in the same CT. GEE was used to yield an unadjusted population averaged model of the association between USDA low access status and LS7 healthy diet score and models
adjusted for EMM by individual-level linguistic acculturation, CT-level percent Hispanic/Latino and CT-level percent Puerto Rican. Prior to the addition of all EMMs into a GEE model of the association between USDA low access status and LS7 diet adjusted for potential confounders identified by the DAG, collinearity of CT-level EMMs was assessed. Interaction terms were tested individually and deleted if not significant at \( \alpha = 0.10 \) until all remaining interactions in the model were statistically significant. Potential confounding variables resulting from the DAG exercise were included if there was at least a 10% change in the effect estimate between models with and without controlling for the variable.\(^{251}\) For EMMs, \( p<0.10 \) was considered statistically significant and \( p<0.05 \) was considered statistically significant for covariates, and 95% CIs were calculated.\(^{181}\) Model diagnostics included a sensitivity analysis excluding the four largest clusters.\(^{252}\) All analyses were performed using SAS version 9.3 (SAS Institute, Inc., Cary, NC) and Stata SE version 12 (StataCorp, College Station, Texas).

5.3.6.2 mRFEI Sensitivity Analysis

Geographic food access can be conceptualized and measured as an absolute or a relative phenomenon. Absolute food access measures like the USDA low access variable determine access in relation to one type of store (i.e. large grocery store or supermarket) and are the most prominent across the literature.\(^ {16,182}\) Though less common, relative food access measures also take food retail locations like fast food establishments and convenience stores into account.\(^ {188}\) A sensitivity analysis was conducted using the CDC’s mRFEI, a relative food access measure that considers the number of healthy food retailers as well as the number of less healthy food retailers within a CT or \( \frac{1}{2} \) mile from the CT.
boundary. Healthy food retailers include supermarkets and larger grocery stores (North American Industry Classification System (NAICS) code 445110; specifically supermarkets with \( \geq 50 \) annual payroll employees and grocery stores with 10–49 employees); fruit and vegetable markets (NAICS code 445230); and warehouse clubs (NAICS code 452910). Less healthy food retailers include fast food restaurants (NAICS code 722211), small grocery stores (NAICS code 445110, specifically grocery stores with three or fewer employees), and convenience stores (NAICS code 445120). The mRFEI score, which can range from 0 to 100, represents the percentage of healthy food retailers out of the total number of healthy and less healthy food retailers in a CT or \( \frac{1}{2} \) mile from the CT boundary. Since mRFEI was created using 2000 CT definitions and BPRHS participants were geocoded to 2010 CT definitions, only participants for whom the definition of the CT in which they lived at baseline did not change between the 2000 and 2010 US decennial census were included in the sensitivity analysis.

Descriptive statistics were used to provide overall sample characteristics. GEE was used to yield an unadjusted population averaged model of the association between USDA low access status and LS7 healthy diet score and models adjusted for EMM by individual-level linguistic acculturation, CT-level percent Hispanic/Latino and CT-level percent Puerto Rican. Prior to the addition of all the EMMs into a GEE model of the association between CDC mRFEI score and LS7 diet adjusted for potential confounders identified by the DAG, collinearity of CT-level EMMs was assessed. Interaction terms were tested individually and deleted if not significant at \( \alpha=0.10 \) until all remaining interactions in the model adjusted for potential confounders were statistically significant.
Potential confounding variables resulting from the DAG exercise were included if there was a 10% change in the effect estimate between models with and without controlling for the variable. For EMMs, \( p<0.10 \) was considered statistically significant, and \( p<0.05 \) was considered statistically significant for covariates and 95% CIs were calculated. Model diagnostics included (1) the effects of potentially influential points, (2) whether linearly modeled variables were approximately linear in the logit and (3) a sensitivity analysis excluding the four largest clusters. All analyses were performed using SAS version 9.3 (SAS Institute, Inc., Cary, NC) and Stata SE version 12 (StataCorp, College Station, Texas).

5.4 Results

5.4.1 USDA Low Access Primary Analysis

5.4.1.1 Participant Characteristics

Of 1,500 baseline participants, 429 were excluded due to missing dietary information, 56 were excluded due to missing individual poverty information and 4 were excluded due to missing marital status information, resulting in a study population of 1,011 participants (67.4% of original study population). The average age was 56.2 years (SD: 7.6), the majority were female (71.9%), and 30.2% were married or living as married with a spouse in the household (Table B.8). The majority of BRPHS participants were living at or below the poverty line (58.2%). On average, participants lived in CTs in which the percent of the CT with a PIR at or below the poverty line was 30.2% (SD: 13.9), the percent of the population that was Hispanic/Latino was 36.2% (SD: 24.7) and the percent of the population that was Puerto Rican specifically was 11.6% (SD: 7.7). All
participants lived in urban CTs and 43.4% lived in urban CTs designated as low access by the USDA using a 0.5-mile distance.

5.4.1.2 Diet Characteristics

As shown in Table B.9, out of a possible score of 5, the average overall LS7 healthy diet score was 1.21 (SD: 0.8). The majority had a poor LS7 healthy diet score (68.2%) whereas only 5 participants (0.5%) had an ideal LS7 diet score. All participants exceeded, and, therefore, did not meet the daily sodium intake recommendation, but the majority met recommended weekly intake of sweets/sugar sweetened beverages (73.8%). However, the majority did not meet recommended daily fruit and vegetable intake (81.1%), weekly fish intake (75.0%), or daily whole grain intake (95.9%).

5.4.1.3 Model

Since the Spearman’s rank correlation coefficient for the association between CT-level percent Hispanic/Latino and CT-Level percent Puerto Rican was 0.97 (p-value <0.001), only the interaction terms for individual-level linguistic acculturation and CT-level percent Puerto Rican were individually tested in models adjusted for potential confounders. The interaction terms were subsequently deleted since neither was significant at α=0.10, as were DAG-identified potential confounding variables since none of them produced a 10% or larger change in the effect estimate between models with and without controlling for the variable.251

In the adjusted model that included interaction terms for individual-level linguistic acculturation and CT-level percent Puerto Rican, the odds of a poor LS7 healthy diet score did not differ significantly between participants living in and outside of
low access urban CTs at $\alpha=0.05$ (OR: 0.96, 95% CI: 0.60-1.53, p-value: 0.86). In the fully adjusted model without interaction terms, the odds of a poor LS7 healthy diet score did not differ significantly between participants living in and outside of low access urban CTs at $\alpha=0.05$ (OR: 1.06, 95% CI: 0.78-1.45, p-value: 0.71). In the final, unadjusted model, the odds of a poor LS7 healthy diet score did not differ significantly between participants living in and outside of low access urban CTs at $\alpha=0.05$ (OR: 1.20, 95% CI: 0.88-1.64, p-value: 0.25).

5.4.1.4 Model Diagnostics

CT clusters ranged in size from 1 to 64 with an average cluster size of 5.9 participants. A sensitivity analysis was conducted after excluding the four largest CT clusters, which were 64, 51, 38 and 34 participants in size. As in the primary analysis, no interaction terms were significant at 0.10, and no confounders changed the unadjusted estimate (OR: 1.13, 95% CI: 0.84-1.55, p-value: 0.41) by 10% or more.

5.4.2 mRFEI Sensitivity Analysis

5.4.2.1 Participant Characteristics

The sensitivity analysis included 680 participants for whom there was no change in CT definition between the 2000 and 2010 US decennial census (45.3% of original study population). The average age was 57.1 years (SD: 7.6), the majority were female (71.5%), and 29.4% were married or living as married with a spouse in the household (Table B.10). The majority of BPRHS participants were living at or below the poverty line (57.9%). On average, participants lived in CTs in which the percent of the CT with a PIR at or below the poverty line was 28.8% (SD: 11.5), the percent of the population that
was Hispanic/Latino was 37.2% (SD: 26.8), the percent of the population that was Puerto Rican specifically was 11.6% (SD: 8.4) and the mRFEI score was 6.8 (SD: 5.1).

5.4.2.2 Diet Characteristics

As shown in Table B.11, the average overall LS7 healthy diet score was 1.20 (SD: 0.8). The majority had a poor LS7 healthy diet score (68.8%) whereas only 3 participants (0.4%) had an ideal LS7 healthy diet scores. Although all participants exceeded the daily sodium intake recommendation, the majority met recommended weekly intake of sweets/sugar sweetened beverages (73.8%). However, the majority did not meet recommended daily intake of fruits and vegetables (81.0%), weekly intake of fish (75.0%), or daily intake of whole grains (95.9%).

5.4.2.3 Model

The interaction terms for individual-level linguistic acculturation and CT-level percent Puerto Rican were individually tested in a model adjusted for potential confounders and subsequently deleted since neither was significant at \( \alpha=0.10 \), as were DAG-identified potential confounding variables since none of them produced a 10% or larger change in the effect estimate between models with and without controlling for the variable.\(^{251}\) In the final, unadjusted model, mRFEI score was not associated with a significant change in the odds of a poor LS7 healthy diet score at \( \alpha=0.05 \) (OR: 0.99, 95% CI: 0.96-1.03, p-value: 0.63).

5.4.2.4 Model Diagnostics

Model diagnostics included investigating (1) a sensitivity analysis excluding potentially influential points, (2) whether linearly modeled mRFEI score is approximately
linear in the logit and (3) a sensitivity analysis excluding the four largest clusters. The first sensitivity analysis was conducted after excluding 15 participants living in CTs with mRFEI scores above 20. In the final, unadjusted model, a 1 point increase in mRFEI score was not associated with a significant change in the odds of a poor LS7 healthy diet score at $\alpha=0.05$ at $\alpha=0.05$ (OR: 0.98, 95% CI: 0.94-1.02, p-value: 0.32).

Fractional polynomials indicated that the best form for mRFEI among 44 models fit were squared (2) and cubed (3). However, when using a squared mRFEI score as the exposure of interest, 1 point increase in mRFEI score squared was not associated with a significant change in the odds of a poor LS7 healthy diet score at $\alpha=0.05$ (OR: 1.00, 95% CI: 1.00-1.00, p-value: 0.90). Similarly, when using cubed mRFEI score as the exposure of interest, 1 point increase in mRFEI score cubed was not associated with a significant change in the odds of a poor LS7 healthy diet score at $\alpha=0.05$ (OR: 1.00, 95% CI: 1.00-1.00, p-value: 0.88). Although the number of clusters (i.e., 134 CTs) was sufficient for the model to detect a non-linear trend, the majority of BPRHS participants lived in CTs with mRFEI scores between 0 and 13. mRFEI was therefore treated linearly in the final model.

CT clusters ranged in size from 1 to 60 with an average cluster size of 5.1 participants. The second sensitivity analysis was conducted after excluding the four largest CT clusters, which were 60, 51, 34 and 26 participants in size. As in the primary analysis, no interaction terms were significant at 0.10 and no confounders changed the unadjusted estimate (OR: 1.00, 95% CI: 0.96-1.03, p-value: 0.87) by 10% or more.
5.5 Discussion

This study used publicly-available measures to quantify the association between CT-level geographic food access and individual-level diet among BPRHS participants. Primary analyses pertained to an USDA low access (for urban CTs), an absolute CT-level geographic food access metric, and secondary analyses pertained to the CDC’s mRFEI score, a relative CT-level geographic food access metric. Both metrics were used as a proxy for individual-level geographic food access given the absence of that data among BPRHS participants. Results indicate that neither CT-level geographic food access metric was significantly associated with LS7 healthy diet score nor was there confounding by individual- and CT-level sociodemographic predictors or EMM by CT-level immigrant enclave or individual-level linguistic acculturation status.

Despite the availability of area-level geographic food access metrics from the USDA and CDC, geographic food access and diet are measured differently across studies, making it challenging to compare the results of this study with that of previous research. A review of the literature found the results of studies accessing store availability and accessibility using GIS-based methods lacked consistency in relation to diet. Other studies have focused on informational food access and diet, but not in the context of the geographic food access. An analysis of MESA data from 1,191 urban Hispanic Americans, of whom 117 were mainland Puerto Ricans, found living in CTs with higher proportions of immigrants born in Latin America to be associated with lower high-fat food consumption, controlling for age, gender, income, education, neighborhood poverty and acculturation. A similar study of 345 Hispanic women in New York City, of whom
were mainland Puerto Ricans, found adherence to a healthy dietary pattern consisting of high consumption of vegetables, legumes, potatoes, fish and other seafood was positively associated with neighborhood linguistic isolation and negatively associated with neighborhood poverty, adjusting for individual and neighborhood sociodemographic characteristics. The current study explored sources of informational food access culturally relevant to mainland Puerto Ricans as EMMs of the geographic food access and diet relationship using publicly-available geographic food access measures, and can, therefore, be easily replicated.

This study also shed light on diet quality among BPRHS participants, as captured by LS7 healthy diet score. Only 0.5% of BPRHS participants had an ideal LS7 healthy diet score, a finding that is consistent with age-stratified 2003-2008 NHANES estimate of the prevalence of ideal LS7 healthy diet score among U.S. adults (range: 0.2% to 2.6%).

5.5.1 Strengths

This study possesses several strengths. This study used both an absolute and relative food access metric, which is uncommon in the literature. A systematic comparison the USDA food desert and CDC mRFEI measures used data from a 2009 food environment validation study conducted in 8 South Carolina counties compared the measures with respect to the number of CTs classified as low access by the USDA and the number of CTs with a mRFEI score of 0. The current study separately examined the relationship between diet and USDA low access among urban CTs using a 0.5-mile distance and mRFEI score. By doing so, it highlighted the limitations of using mRFEI
score, which is based on 2000 US CT definitions and 2009 NAICS business codes, and has not been updated. Conversely, USDA measures have been updated using current decennial US Census CT definition and recent population and U.S. food retail industry data.

Also, by using a DAG, this study clearly specified plausible causal relationships between potential confounders, the exposure of interest and the outcome of interest based on the extant literature and biological plausibility. Although none of the potential confounders explored by this study were found to produce a 10% or larger change in the effect estimate between models with and without controlling for the variable, the DAG used by this study may nevertheless inform the inclusion of potential confounding variables in future studies since the geographic food access and diet relationship is a growing area of research. Similarly, the reported effect estimates can be used to inform power calculations for future studies examining the relationship between geographic food access and diet among research samples of understudied populations like mainland Puerto Ricans, so that these studies enroll the minimum number of participants needed to have sufficient statistical power to detect differences in diet behaviors.

Additionally, this study was novel due to its exploration of culturally relevant forms of informational food access identified by the literature as having the potential to modify the relationship between geographic food access and diet. Geographic food access and informational food access have typically been researched separately in relation to diet.
5.5.2 Limitations

Given the geographic location and age range of BPRHS participants, the results of this study may not be generalizable to other groups of mainland Puerto Ricans. Although recruited using approaches such as door-to-door enumeration of randomly selected census blocks with 10 or more Hispanics within the target age range, BPRHS participants were 45 to 75 years of age, resided in the Greater Boston area, and participated in a research study, thus limiting generalizability.

Additionally, the use of cross sectional data does not allow for the deduction of causal relationships and fails to address the importance of temporality and duration of exposure in the geographic food access and diet relationship. For example, with store openings and closings, the geographic food environment is constantly changing, and this fluctuation is not reflected by the USDA and CDC measures. Consequently, analyses relying on these measures cannot speak to the amount of time it takes for a change in geographic food access to lead to a difference in diet.\textsuperscript{260} By using a CT-level geographic food access metric as a proxy for individual-level geographic food access, the exposure of interest may not reflect where participants actually shop for food. This disconnect may underlie the null associations obtained in this study, including the lack of EMM by culturally relevant sources of informational food access, since BPRHS participants may venture outside of their CT to purchase food, in addition bodegas (i.e., small Hispanic/Latino convenience stores) and other sources of culturally appropriate food may not be captured by the USDA low access metric or mRFEI score.\textsuperscript{261}
Furthermore, by using a composite score as the outcome of interest, this study did not examine whether the relationship between the CT-level geographic food access metrics and individual dietary components. However, a sensitivity analysis using USDA low access as the exposure of interest and meeting recommended LS7 daily fruit and vegetable intake as the outcome of interest found no confounding, EMM or significant relationship at \( \alpha=0.05 \) (OR: 0.82, 95% CI: 0.62-1.09, p-value: 0.18).

As evidenced by this study, the utility of a multi-level approach is often hampered by the current understanding of a construct measured at different levels, such as geographic food access at the area- and individual-level. When multi-level data assumed to capture the same construct are not easily accessible or readily available, data from one level is used as a proxy for data from another, which may unintentionally lead to forms of bias common among ecological and individual-level studies, like the ecological and sociologic fallacies. Given the null findings of this study, it is therefore critical that future research examine individual-level geographic food access data and compare results to those obtained by using proxy measures.

5.6 Conclusion

The disparate burdens of CVD and cancer experienced by mainland Puerto Ricans are influenced by a multitude of social determinants of health, including individual-level risk behaviors, such as diet, that have the potential to be modified (Figure A.4). A social determinant of health receiving increased empirical attention that may contribute to CVD and cancer health disparities through its impact on diet is geographic food access. Using
publicly-available exposure measures and LS7 healthy diet score, the current study did not find a significant relationship between geographic food access, measured at the CT-level, and healthy diet among BPRHS participants. No confounding and EMM was found in this relationship, despite using an explicit method to identify individual- and CT-level social determinants of health that had the potential to confound this relationship and exploring individual- and CT-level social determinants of health that captured culturally relevant forms of informational food access that had the potential to modify the geographic food access and diet relationship.

Future studies should expand upon the methods employed by this study to better elucidate the relationship between geographic food access and diet by incorporating individual- as well as area-level geographic food access data and further examining the impact of informational food access. For example, since culturally-appropriate food stores may not be reflected by the CT-level geographic food access metrics employed by this study, future studies conducted among BPRHS participants could collect primary individual-level geographic food access data to better understand where these individuals shop for food, how far they travel to make food purchases and by what means, and what factors influence their food purchasing decisions. Similarly, future studies can explore other indicators of individual- and area-level acculturation as forms of informational food access, like individual-level psychological acculturation and length of time on the U.S. mainland and CT-level U.S. Census data on language use, English-speaking ability and linguistic isolation. To inform intervention and policy efforts, there is a need to better understand how social determinants of health like geographic food access,
acculturation, cultural norms, income, and personal food shopping behaviors impact CVD and cancer health disparities through diet, the county’s leading modifiable chronic disease risk factor.
Chapter 6: The Relationship between Geographic Food Access and Life’s Simple 7 Healthy Diet Score in The Appalachia Community Cancer Network Study

6.1 Abstract

**Background:** County- and individual-level geographic food access have the potential to independently influence diet, and an individual’s geographic food access is influenced by where they live. County-level geographic food access measures based on traditional food retailers continue to be used as proxies of individual-level geographic food access in rural and urban/rural heterogeneous areas. However, food-shopping realities faced by individuals living in rural and urban/rural heterogeneous areas include the closing of traditional food retailers (e.g., grocery stores and supermarkets) and increased patronization of non-traditional food retailers (e.g., dollar stores and big-box stores). This study sought to determine whether an individual-level geographic food access metric adjusted for county-level geographic food access, or a county-level geographic food access was more strongly associated with individual-level diet in a research sample of residents of Appalachia. Appalachia is an urban/rural heterogeneous region of the U.S. that experiences poorer dietary components, the most heavily researched of which is fruit and vegetable intake. For example, the prevalence of inadequate intake of fruit and vegetable is higher in Appalachia compared to the remainder of the U.S. and, among states in the region, the prevalence of inadequate intake of fruit and vegetable intake
tends to be higher among those living in Appalachian counties than those living non-Appalachian counties.\textsuperscript{101, 145 100}

Methods: A directed acyclic graph (DAG) was used to identify individual- and area-level sociodemographic characteristics that had the potential to confound the relationship between each geographic food access metric and diet, as measured by Life’s Simple 7 (LS7) healthy diet score, a composite score based on intake of fruits/vegetables, fish, sodium, sweets/sugar-sweetened beverages and whole grains. Using data from the Appalachia Community Cancer Network (ACCN), generalized estimating equations (GEE) were used to yield nested county- and individual-level geographic food access models adjusted for the DAG-identified potential confounders. The magnitude and statistical significance of the effect estimates were compared across the models as were the quasi-likelihood under the independence model criterion (QIC) values of each model.

Results: Adjusted for potential confounders, county-level geographic food access was not associated with a significant change in the odds of poor LS7 healthy diet score among ACCN study participants (odds ratio (OR): 0.98, 95% confidence interval (CI): 0.96-1.00, p-value: 0.07, QIC: 507.64). Adjusted for potential confounders, including county-level geographic food access, individual-level geographic food access was not associated with a significant change in the odds of poor LS7 healthy diet score among ACCN study participants (OR: 0.99, 95% CI: 0.97-1.02, p-value: 0.64, QIC: 508.90). Based on QIC values, the adjusted county-level model best fit the data, but the effect estimate was weak in magnitude and lacked statistical significance.
Discussion: These findings suggest geographic food access, as measured at the individual- and county-level, does not significantly influence diet in the study population. The primary data collected to derive the individual-level geographic food access measure provided insight into non-geographic factors influencing food purchasing behavior of ACCN study participants and raised questions regarding the interaction between geographic food access and food purchasing behavior in influencing diet. Consequently, future research should assess whether geographic food access is more appropriately treated as an effect measure modifier (EMM) of the relationship between food purchasing behavior and diet rather than as the exposure of interest.
6.2 Introduction

Geographic food access, a measure of distance-based facilitators and barriers of food purchasing and consumption, is a characteristic of the physical context in which an individual lives that has the potential to contribute to CVD and cancer health disparities through diet, a mutual, modifiable risk behavior at the individual-level (Figure A.4).\textsuperscript{15} Due to increasing use of geographic information systems (GIS) and availability of financing and other incentives at the federal, state and local levels to improve access to food in areas currently lacking adequate access, geographic food access is an expanding area of research.\textsuperscript{16, 17} However, the existing food access literature has yielded mixed results on the relationship between geographic food access and diet\textsuperscript{16, 265} Published studies have often relied on area-level measures of geographic food access, either publicly-available metrics or metrics derived from primary and/or secondary data for the purpose of an individual study, to serve as proxies for individual-level geographic food access.\textsuperscript{16, 265}

Although imperfect, publicly-available geographic food access metrics that can readily be used across the county as a proxy for individual-level geographic food access are more cost-effective than deriving individual-level geographic food access metrics for a specific research population by collecting primary data.\textsuperscript{266} Studies typically use one level of data and do not choose to nest multi-level data. Consequently, there have been few opportunities to explore whether an improved understanding of geographic food access and diet in rural and urban/rural heterogeneous areas can be achieved by focusing
on the relationship between individual-level geographic food access and diet, accounting for county-level geographic food access. This type of approach is needed to better reflect the complexity of geographic food access as a social determinant of health (Figure A.4).

Whereas census tracts (CTs) are a common unit of analysis used to proxy geographic food access in urban areas of the U.S., county is a commonly used geographic unit of analysis used to proxy geographic food access in regions of the U.S. that are more rural or heterogeneous in terms of urbanity/rurality.\(^{16, 265}\) Despite using different geographic units of analysis, CT- and county-level geographic food access measures that are publicly-available through the U.S. Department of Agriculture (USDA) utilize a threshold distance of 1 mile in urban areas such that individuals living more than 1 mile from a supermarket or large grocery store contain all the major food departments found in a traditional supermarket (e.g. fresh meat and poultry, dairy, dry and packaged foods and frozen foods) are considered to have low geographic food access, whereas a threshold distance of 10 miles is used in rural areas such that individuals that live more than 10 miles from a supermarket or large grocery store are considered to have low food access.\(^{193, 195, 196}\)

Large grocery stores and supermarkets are used by the USDA as a proxy for the availability of healthy food since these retailers offer a wider range of nutritious and affordable foods compared to other food retailers.\(^{265}\) However, this approach may not reflect shopping realities, particularly in rural areas, due to changes in the geographic food environment. It has been documented that these traditional food retailers are slowly disappearing in rural areas, possibly due to factors such as population loss, commuting
patterns of residents to areas outside of local towns and counties and increased connectivity between rural and urban areas due to increased car ownership and roadways/highways that jeopardize the financial viability of large supermarkets and grocery stores. Additionally, for over a decade, traditional food retailers have also received more competition from non-traditional food retailers, such as big-box or supercenter stores like Wal-Mart, in terms of both customer base and price. Furthermore, a national study found 418 rural counties (20% percent of all rural counties) in which all residents lived more than 10 miles from a supermarket or supercenter, which suggested the traditional distance threshold used in rural areas may be need to be increased to better reflect how far residents typically travel to purchase food in certain rural areas.

To date, the relationship between geographic food access and diet in the Appalachian region has received little empirical attention. Appalachia is a federally-designated region of the U.S. along the Appalachian mountains, that is home to over 25 million residents. Of Appalachia’s population, 60% live in metropolitan counties, 25% live in counties adjacent to metropolitan counties and the remainder live in more remote, rural locations. In addition to geographic diversity, the region is characterized by economic, sociodemographic and resource-based diversity. Although cardiovascular disease (CVD) is the leading cause of mortality in the U.S. and contributes to health disparities within Appalachia, cancer is the region’s leading cause of mortality. Differences in SES, sociodemographic characteristics and the availability of health-promoting resources likely contribute to differences between Appalachia and the
remainder of the U.S. in the burden of CVD, cancer and mutual, modifiable risk behaviors and factors, including diet.\textsuperscript{13} Fruit and vegetable consumption is the most heavily researched dietary component among Appalachian residents. Compared to the rest of the country and residents of non-Appalachian counties within the 13-state region, residents of Appalachia are known to experience a higher prevalence of inadequate fruit and vegetable consumption.\textsuperscript{101, 145, 100} Given noted data limitations, such as the lack of public health surveillance efforts spanning the Appalachian region in its entirety, it is difficult to speak to difference between Appalachia and the rest of the county, or between Appalachian and non-Appalachian counties within the 13-state region, in terms of other components of diet, such as intake of sodium, fish, sweets/sugar sweetened beverages and whole grains.\textsuperscript{34, 51}

6.2.1 Purpose

County- and individual-level geographic food access has the potential to independently influence diet and an individual’s geographic food access is influenced by where they live. Consequently, this study sought to compare nested models adjusted for individual- and county-level sociodemographic characteristics to determine whether a publicly-available county-level geographic food access metric or an individual-level geographic food access metric derived from primary data on the store(s) where participants preferred to shop for food was more strongly associated with individual-level diet in a research sample of residents of Appalachia.
6.3 Methods

6.3.1 Study Population

The ACCN, a joint effort of the University of Kentucky, The Ohio State University, Pennsylvania State University, Virginia Tech University, and West Virginia University, is currently conducting a group randomized trial in a five-state area to test a previously piloted ACCN faith-based intervention focused on individual and environmental level changes to increase physical activity and improve healthy food choices using eHealth technology, which has been described elsewhere. In brief, using county as the unit of randomization, 13 churches were randomly assigned to receive Walk by Faith, the dietary and physical activity faith-based intervention program, and 15 were randomly assigned to receive Ribbons of Faith, a comparison program focused on cancer screening.227

The study was promoted in church bulletins and announcements from the pulpit. Group recruitment events and individual appointments were held at participating churches by research staff, at which time an individual could be screened to determine eligibility to participate. Eligibility criteria included: being at least 18 years of age, attending services at least four times in the past two months at a participating church, able to understand and read English, cognitively able to provide informed consent, resident of an Appalachian county, not residing in a nursing facility or residential home, not planning to move away from the study area, willing to use a computer, no dietary restrictions prescribed for weight-loss or part of a formal weight-loss program, less than
400 pounds, having a body mass index (BMI) of at least 25, and, if female, not pregnant, breastfeeding or less than 9 months post-partum or planning to become pregnant during the study period. Questions modified from the physical activity readiness questionnaire (PAR-Q) were used to determine whether medical clearance from a physician was required to participate. The ACCN protocol was approved through the institutional review boards (IRBs) of the Ohio State University, University of Kentucky, Pennsylvania State University, Virginia Polytechnic Institute and State University and West Virginia University. This study was approved by the IRB of the Ohio State University.

Pennsylvania and Ohio church attendees were recruited between January and September 2012 and Kentucky, Virginia and West Virginia church attendees were recruited between November 2012 and October 2013. Of 866 individuals screened for eligibility, 663 enrolled into the study and completed baseline measurements prior to the start of the intervention, of whom 427 were Walk by Faith participants and 237 were Ribbons of Faith participants. The measures examined in this study were collected via a phone interview (age, gender, marital status, occupation, household income), whereas dietary intake was collected at baseline of the ACCN study using a computer-based visual food frequency questionnaire (FFQ) developed by Viocare, Inc., a private health and wellness systems company.

6.3.2 Outcome of Interest

The outcome of interest was a healthy diet score, as defined by the American Heart Association’s (AHA’s) LS7 of cardiovascular health (CVH) metric, which includes
well-defined poor, intermediate and ideal levels of four health behaviors (smoking, BMI, physical activity, and diet) and three health factors (total cholesterol, blood pressure and plasma glucose).\textsuperscript{5} LS7’s diet component is a composite score of five components worth one point each (≥4.5 servings of fruits/vegetables per day, ≥2 servings of fish per week, <1500 mg of sodium per day, ≤450 kcal (36 oz) per week of added sugars (sweets/sugar-sweetened beverages) and ≥3 servings per day of whole grains).\textsuperscript{5} An ideal diet was defined as a score ≥4, whereas intermediate diet was defined as a score between 3 and 2, and a score less than 2 was considered poor.

Given the data available for ACCN participants, as collected by a computer-based visual FFQ, several LS7 healthy diet score component definitions had to be modified. Daily fruit and vegetable servings were based on standard portion sizes rather than cups, 4 ounces was considered a standard serving of fish and ideal rather than 3.5 ounces, intermediate and poor sweets/sugar-sweetened beverages were derived from a variable of gram equivalents of added sugars rather than kcal/ounces per week of sweets and sugar sweetened beverages.\textsuperscript{233} Since previous research found the prevalence of poor LS7 healthy diet score among all ACCN study participants to be 59.5%, LS7 healthy diet score was dichotomized into poor (<2) or not poor (≥2) for the purpose of this study.\textsuperscript{259}

6.3.3 Exposures of Interest

6.3.3.1 County-Level Indicator of Geographic Food Access

A USDA county-level variable capturing the percentage of the county’s population living more than 1 mile from a supermarket or large grocery store if in an
urban area, or more than 10 miles from a supermarket or large grocery store if in a rural area served as the county-level geographic food access exposure of interest. The USDA derived this variable from a directory of supermarkets and large grocery stores within the U.S. created by merging a 2010 directory of stores authorized to accept Supplemental Nutrition Assistance Program (SNAP) benefits and the 2010 Trade Dimensions TDLinx directory of stores. To be included, supermarkets and large grocery stores must have reported at least $2 million in annual sales and contain all the major food departments found in a traditional supermarket. The Census Bureau’s Urban Area definition was used to designate rural and urban status. Per the Census Bureau, there were two types of urban areas: (1) urbanized areas with 50,000 or more residents; and (2) urban clusters with 2,500 to 50,000 residents. All areas that did not meet either urban area definition were considered rural. Census block-level population data from the 2010 Census of Population and Housing were aerially allocated to ½-kilometer-square grids across the U.S. Distance was determined by calculating the distance from the geographic center of each ½-kilometer-square grid cell to the center of the grid cell containing the nearest supermarket or large grocery store. Data were aggregated to the county-level to determine the number of individuals living more than 1 mile from a supermarket or large grocery store in urban areas and more than 10 miles from a supermarket or large grocery store in rural areas and was divided by the county population size.

6.3.3.2 Individual-Level Indicator of Geographic Food Access

Self-reported, individual-level home-to-preferred food store distance was obtained from a survey that contained food shopping questions adapted from food-related
questions included in the Boehmer/Brownson physical environment survey (Appendix A). The survey was part of a larger take-home survey completed by participants in Ohio, Kentucky, Virginia and West Virginia. For the two stores they patronized most frequently to shop for food, participants were asked to list the name and location of the store, the store-to-home distance, their main method of transportation and the main reason for patronizing. Reasons for patronizing could be geographic (e.g., location) or non-geographic (e.g., price). Since not all participants listed information for two stores, individual-level home-to-preferred store distance was calculated as either the average of home-to-preferred store distance for two stores among participants who provided information on the two stores they patronized most frequently to shop for food or as the home-to-preferred store distance for one store among participants who only provided information on one store they patronized most frequently to shop for food.

6.3.4 Potential Confounders

A DAG was used to determine whether a number of covariates available at the individual- and county-level from available ACCN, Appalachian Regional Commission (ARC) and U.S. Census data met criteria for confounding (Figure A.7). Plausible causal relationships (represented by unidirectional arrows) were determined on the basis of extant literature and biological plausibility. To be considered as a potential confounder of the relationship between an county- and individual-level geographic food access and diet, a variable must (1) have had an arrow pointing in either direction between it and LS7 healthy diet score, and (2) have had an arrow pointing from it to the county-level USDA percent low food access variable and/or individual-level home-to-preferred store distance.
Arrows pointing from either exposure suggest the covariate was on the causal pathway between an exposure and LS7 healthy diet score and treating the covariate as a confounder may induce bias. A potential confounder was removed from the DAG if controlling for another potential confounder would block the direct pathway between it and LS7 healthy diet score and/or the exposure of interest.

Since individual-level food security may be on the causal pathway between individual-level home-to-preferred store distance and LS7 healthy diet score, it was not included in the final DAG (Figure A.8). Although they met the requirements to be considered as a potential confounder, individual-level education and insurance status were also left out of the final DAG since controlling for individual-level household income blocked the direct pathways between these variables and the exposures of interest. As shown in Figure A.8, variables considered as potential confounders of the relationship between the county-level USDA percent low food access variable and LS7 healthy diet score variable were limited to age, sex, marital status and annual household income at the individual-level, and economic strength was considered a potential confounder at the county-level. For the relationship between individual-level, home-to-preferred food store distance and LS7 healthy diet score, potential confounders were limited to age, sex marital status and annual household income at the individual-level and, at the county-level, ARC economic strength and the USDA percent low food access.

Sex was treated as a binary variable, and age at baseline was treated as a continuous variable. Marital status was treated as a binary variable: married/living with partner or not. Annual household income was treated as a categorical variable with 4
levels: <$20,000, $20,000-$59,999, $60,000-$99,999, and $\geq$100,000. ARC FY2012 county economic status was treated as a binary variable: economically distressed or not economically distressed.\textsuperscript{267}

6.3.5 Analyses

Descriptive statistics were used to provide overall characteristics of ACCN participants, their diet and their food purchasing behavior. GEE yielded population average models of the association between the exposures of interest and LS7 healthy diet score with robust standard errors (SEs). Bernoulli was specified as the family (the assumed distribution of LS7 healthy diet score), logit was specified as the link function and a working exchangeable correlation structure was specified to reflect the average dependence among correlated observations. Since the adjusted model using the USDA variable as the exposure of interest included fewer parameters than the adjusted model using individual-level, preferred home-to-food store distance as the exposure of interest and was thus nested within the adjusted model using individual-level, preferred home-to-food store distance as the exposure of interest. Given the nested nature of the two models (Figure A.8), all DAG-identified potential confounders were included in the GEE models. Model diagnostics included the assessment of whether linearly modeled variables were approximately linear in the logit.\textsuperscript{252} QIC was used for model selection.\textsuperscript{255} For the effect estimate of the exposure of interest, p<0.05 was considered statistically significant and 95% CIs were calculated. The magnitude and statistical significance of the effect estimate of the exposures of interest and QIC value were compared across the two models. As with AIC, the model with the smaller QIC value was preferred, meaning it fit the data.
better, taking model complexity into account. All analyses were performed using SAS version 9.3 (SAS Institute, Inc., Cary, NC) and Stata SE version 12 (StataCorp, College Station, Texas).

6.4 Results

6.4.1 Participant Characteristics

Of 663 baseline participants, 155 Pennsylvania participants were excluded because they were not asked to complete the survey from which self-reported, individual-level home-to-preferred food store distance was derived, 11 were excluded due to missing dietary information, 95 were excluded because of incomplete annual household income data, and 23 were excluded due to incomplete individual-level home-to-preferred food store distance data, resulting in a study population of 379 participants (57.2% of original study population). Among these 379 ACCN participants, the average age was 55.8 years (SD: 12.9), the majority were female (71.5%), and 79.7% were married or living as married with a spouse in the household (Table B.12). Nearly half of all ACCN participants reported an annual household income between $20,000 and $59,999 whereas 5% reported an annual household income of less than $20,000 and nearly 15% reported an annual household income of $100,000 or more. Average individual-level home-to-preferred store distance was 10.5 miles (SD: 9.6, median: 9, interquartile range (IQR) (i.e. 75th percentile minus the 25th percentile): 12.5). Nearly half lived in counties categorized as economically distressed in fiscal year (FY) 2012 by the ARC. Per the USDA, participants lived in counties where an average of 16.9% of the population lived
more than 1 mile from a supermarket or large grocery store (if in an urban area) or more than 10 miles from a supermarket or large grocery store (if in a rural area).

6.4.2 Diet Characteristics

As shown in Table B.13, out of a possible score of 5, the average overall LS7 healthy diet score was 1.5 (standard deviation (SD): 0.7). The majority had a poor LS7 healthy diet score (62.3%) and only 1.1% of participants had an ideal LS7 healthy diet score. Although all participants met the recommendation for sweets/sugar-sweetened beverages, the majority fell below recommended daily fruit and vegetable intake (84.2%), weekly fish intake (88.4%), and daily whole grain intake (84.7%). Additionally, the majority went over recommended daily sodium intake (95.5%).

6.4.3 Food Purchasing Behaviors

Participants were asked to report how frequently they went food shopping across a variety of retail locations and allowed to list other locations from which they shopped for food. Of those who responded, the highest percentage categories for each food store were as follows: over 67% of participants reported that they often food shop at a large grocery store or supermarket, 72.8% occasionally/sometimes food shop at a small grocery store or supermarket, 55.4% never food shop at a convenience store or gas station, 66.2% occasionally/sometimes food shop at a general store or dollar store and 73.8% occasionally/sometimes food shop at a farmers market or fruit and vegetable stand. Participants commonly listed Sam’s Club/Wal-Mart, health food stores and personal garden or garden of family/friends when asked to name other locations at which they shopped for food.
For the primary store where they went food shopping most often, the average distance from home was 10.4 miles (SD: 10.9, median: 5, IQR: 15) and ranged from less than 1 mile to 60 miles. Low prices (27.7%), good selection (26.1%) and location close to home (27.9%) were main reasons they chose to patronize these stores. Nearly all participants said that they usually got to the store using their own car (98.4%). Of the five LS7 healthy diet score components, only fruit and vegetable purchasing was explicitly asked about. Overall, 86.8% of participants agreed that it was easy to buy fresh fruits and vegetables at these stores, 88.7% thought these stores carried many types of fruits and vegetables, but only 43.5% felt that the fresh fruits and vegetables carried by these stores were of high quality. In terms of cost, 5.3% of participants felt that the fresh fruits and vegetables were extremely expensive, whereas 70.8% thought they were somewhat expensive. Over 61.0% of participants said that the cost of fresh fruits and vegetables had kept them from purchasing these items.

The distance between the secondary store where they went food shopping most often and their home was estimated by 355 participants. The average distance was 10.9 miles (SD: 14.0, median: 6, IQR: 16) and ranged from less than 1 mile to 160 miles. Low prices (26.0%), good selection (21.9%) and location close to home (26.3%) were main reasons ACCN participants chose to shop at these stores. Nearly all participants said that they usually got to the store using their own car (98.1%). In terms of fresh fruit and vegetable purchasing, 80.1% participants agreed that it was easy to buy fresh fruits and vegetables at these stores, 73.9% thought these stores carried many types of fruits and vegetables, but only 66.1% felt that the fresh fruits and vegetables carried by these stores
were of high quality. In terms of cost, 12.4% of participants felt that the fresh fruits and vegetables were extremely expensive, whereas 63.1% thought they were somewhat expensive. Nearly 55.0% (54.9%) of participants said that the cost of fresh fruits and vegetables had kept them from purchasing these items.

Participants were asked to indicate changes that could be made to help them eat healthier in their community. From most to least helpful, changes included more healthy choices at sit-down restaurants (42.0%), more healthy choices at fast food places (35.4%), better quality and variety where they currently shop (31.7%), an improved grocery store or supermarket nearby (24.8%), a local farmer’s market (24.8%), coupons and price discounts (22.0%), a new grocery store or supermarket nearby (16.6%), a community supported agriculture (CSA) program where people buy a membership in order to get produce directly from farmers (14.8%), and a food-buying co-op where food is bought in bulk by a group of people for a discount (12.4%). Only 2 participants thought free- or low-cost rides to local places that sell food would be helpful, only one thought more places that accept SNAP benefits would be helpful, and no participants felt that more places that accept WIC vouchers would be helpful. Participants commonly listed changes in personal attitudes and behaviors related to eating and lifestyle choices as other changes that would be helpful to eat healthier in their community.

6.4.4 Models

GEE was used to yield a population averaged model of the association between USDA county-level percent low food access and binary LS7 healthy diet score adjusted for the DAG-identified potential confounders. Adjusting for potential confounders, a one-
point increase in the percent of county residents with low food access was associated with a 2% reduction in the odds of a poor LS7 healthy diet score (OR: 0.98, 95% CI: 0.96-1.00, p-value: 0.066). This association was not significant at $\alpha=0.05$ and was in the opposite direction that what was expected.

GEE was also used to yield a population averaged model of the association between individual-level, home-to-preferred food store distance in miles and binary LS7 healthy diet score adjusted for potential confounders, including county-level USDA percent low food access. Treated linearly, individual-level, home-to-preferred food store distance was not associated with a significant change in the odds of a poor LS7 healthy diet score at $\alpha=0.05$, adjusted for potential confounders (OR: 0.99, 95% CI: 0.97-1.02, p-value: 0.64).

Taking model complexity into account, the fit of the model to the data is represented by the QIC value and a lower value suggests better fit. The QIC value of the adjusted model of the association between USDA county-level percent low food access and binary LS7 healthy diet score (507.64) was lower than the QIC value of the adjusted model of the association between individual-level, home-to-preferred food store distance in miles and binary LS7 healthy diet score into account (508.90). Consequently, the adjusted county-level model was the preferred model on the basis of QIC value.

### 6.4.5 Model Diagnostics

Fractional polynomials were used to assess whether county-level USDA percent low food access and individual-level, home-to-preferred food store distance were linear in the logit. Fractional polynomials indicated that the best form for county-level USDA
percent low food access among 44 models fit was squared \((x^2)\). In a model adjusted for DAG-identified confounders, county-level USDA percent low food access squared was associated with a significant change in the odds of poor LS7 healthy diet score at \(\alpha=0.05\), adjusting for potential confounders (OR: 0.9992, 95% CI: 0.9985-0.9999, p-value: 0.02). For example, an increase in county population with low food access from 10% squared to 20% squared was associated with a 21.3% reduction in the odds of poor LS7 healthy diet score. The QIC value of this model was 505.7.

Fractional polynomials indicated that the best form for individual-level, home-to-preferred food store distance among 44 models fit were \(1/\text{individual-level, home-to-preferred food store distance}^2\) and \(1/\text{individual-level, home-to-preferred food store distance}\). When using \(1/\text{individual-level, home-to-preferred food store distance}^2\) in a model adjusted for potential confounders, including county-level USDA percent low food access squared, the exposure of interest was not associated with a significant change in the odds of poor LS7 healthy diet score at \(\alpha=0.05\), adjusting for potential confounders (OR: 1.21, 95% CI: 0.81-1.81, p-value: 0.35). The QIC value was 506.07. Similarly, when using \(1/\text{individual-level, home-to-preferred food store distance}\) in a model adjusted for potential confounders, including county-level USDA percent low food access squared, the exposure of interest was not associated with a significant change in the odds of poor LS7 healthy diet score at \(\alpha=0.05\), adjusting for potential confounders (OR: 1.04, 95% CI: 0.74-1.44, p-value: 0.84). The QIC value was 506.37.

Given the relatively small number of clusters (i.e., 19 counties) included in the sample, which impacts the ability of the model to detect a non-linear trend, and that the
majority of participants lived in counties with low access percentages between 3.96% and 25.87%, county-level USDA percent low food access was treated as linear in the final models for which the effect estimates and QIC values were compared.\textsuperscript{268} Given the lack of statistical significance for the model informed by fractional polynomials, individual-level food distance in miles was treated as linear in the final model for which the effect estimate and QIC values were compared.

6.5 Discussion

County-level proxy measures of geographic food access are often used to capture the geographic food access of residents of urban/rural heterogeneous areas because individual-level geographic food access data are not available. However, county- and individual-level geographic food access have the potential to independently influence diet, and an individual’s geographic food access is influenced by place of residence. To date, there have been limited opportunities to explore whether an improved understanding of geographic food access and diet in rural and urban/rural heterogeneous areas can be achieved by focusing on the relationship between individual-level geographic food access and diet, accounting for county-level geographic food access.

This study found the association between a county-level geographic food access proxy variable and odds of poor LS7 diet score to be nearly equal in strength to the association between the individual-level geographic food access variable and odds of poor LS7 diet score. However, across the two models, the effect estimate of the exposure of interest was weak in magnitude and lacked statistical significance. These findings suggest that geographic food access, measured at either level, does not greatly influence
diet, as measured by the LS7 metric, in the study population. One reason for this may have been the high prevalence of car ownership in nonmetropolitan areas. Car ownership may mitigate the importance of distance as a barrier to food shopping among ACCN study participants, so diet could be driven more by what people buy and consume rather than how far they travel to purchase food, especially since large grocery stores and supermarkets carry a plethora of foods that contribute to healthy diets and foods that contribute to unhealthy diet. Another contributing factor may be the small number of county-level clusters included in the analyses. Future studies that include a greater number of counties should to explore the non-linear trend between USDA county-level percent low food access and LS7 diet score suggested by this study’s model diagnostics.

It is difficult to compare and contrast the findings of this study to previous research conducted in rural and urban/rural heterogeneous areas since geographic food access and diet are measured differently across studies. Given the use of varying methods, geographic food accessibility studies lack consistency in relation to diet. There is also a lack of consistency on the primary collection of distance data across studies focused on food purchasing behavior. For example, a random dial telephone survey conducted among 635 adults responsible for food shopping across eight rural Appalachian Kentucky counties found distance to be a significant predictor of store selection, but distance was captured in minutes rather than miles. There is a need for use of similar measures across studies to foster comparison and move this line of inquiry forward.
This study provided insight into individual-level food purchasing behaviors, which were often in conflict with the assumptions underlying the county-level geographic food access measure. For example, food can be purchased and acquired outside of traditional food retailers. The majority of participants reported that they often food shop at a large grocery store or supermarket and occasionally/sometimes food shop at a small grocery store or supermarket, a general store or dollar store, or a farmers market/fruit and vegetable stand. However, when given the opportunity to provide details on other locations they go to food shop, participants commonly named Sam’s Club/Wal-Mart, health food stores and personal garden or garden of family/friend. The variety of stores at which ACCN study participants reported shopping for food is enlightening given that the USDA county-level geographic food access measure only reflected access to supermarkets and large grocery, and thus neglected nontraditional retail food outlets frequented by participants, such as big-box and supercenter stores like Sam’s Club and Wal-Mart, which also offer a full range of grocery items. Similarly, the thresholds used for home-to-supermarket or large grocery store distance used by the USDA county-level geographic food access measure were 1 mile and 10 miles for urban and rural areas, respectively. Although analyses could not be stratified by residence in an urban or rural area given the data available from the USDA, the respective average distances between the primary and secondary stores where ACCN study participants go food shopping most often and their home were 10.4 miles (SD: 10.9, range: <1 to 60 miles) and 10.9 miles (SD: 14.0, range: <1 to 160 miles), suggesting standard distance thresholds used by
publicly-available geographic food access measures may not be appropriate in the Appalachian region.

It is important to note that an assumption underlying the county- and individual-level geographic food access measures employed by this study was that trips to purchase food are made directly between an individual’s place of residence and a food retailer. The impact of commuting patterns on geographic measures that are static in terms of space and time is a growing area of research in urban areas that needs to be extended to rural and urban/rural heterogeneous regions like Appalachia as commuting patterns may have the ability to modify the impact residence-based geographic food access has on diet.270, 271

As a measure of geographic food access, the USDA county-level geographic food access measure employed by this study inherently prioritized distance as the factor determining where one shops for food. However, low prices, good selection and location close to home were roughly equal in explaining why ACCN study participants chose to patronize these stores. Despite only exploring food purchasing behaviors related to fruits and vegetables, quality and cost raise concern over the perceived quality (and subsequent purchase and consumption) of food that is geographically accessible, as well as the disconnect between geographic food access and affordability. Future research is needed to provide insight into the interplay between quality, cost and geographic food accessibility in terms of diet.
6.5.1 Strengths

Strengths of this study include the use of county- and individual-level geographic food access metrics and a DAG, which are uncommon in the literature. Compared to publicly-available area-level geographic food access measures, individual-level geographic food access measures and the raw data from which they are derived are more costly to obtain and thus infrequently reported. Using the extant literature and biological plausibility as a basis, plausible causal relationships between potential confounders, the county- and individual-level exposures of interest and the outcome of interest were specified in the DAG. Although neither exposure of interest was found to be significantly associated with the odds of poor LS7 healthy diet score in this study, the DAG and reported effect estimates for the county- and individual-level geographic food access metrics can be used to inform future studies on the relationship between geographic food access and diet. Additionally, this study was novel due to its exploration of individual-level food purchasing behavior among a research sample of Appalachian residents.

Lastly, by using the LS7 metric to derive the outcome of interest, this study quantified poor, intermediate and ideal levels of diet among ACCN study participants and fosters comparison between the diet quality of ACCN study participants and other populations. Out of a possible continuous LS7 healthy diet score of 5, the average score was low (1.5, SD: 0.7). Whereas 1.1% of ACCN study participants had an ideal LS7 healthy diet score, 62.3% of participants had a poor LS7 diet score. In terms of individual LS7 healthy diet score components, the majority did not meet recommended
daily fruit and vegetable intake, weekly fish intake, daily whole grain intake or daily sodium intake. Although all ACCN study participants met the recommendation for weekly intake of sweets/sugar-sweetened beverages, this result may be attributed to the modified definition. The variable used to derive this component represented average daily grams of added sugar whereas a variable representing servings of sweets could not be used since it could not be converted to kilocalories/ounces per the AHA’s definition.

The prevalence of ideal LS7 healthy diets score observed in this study is consistent with age-stratified 2003-2008 National Health and Nutrition Examination Survey (NHANES) estimates among U.S. adults (range: 0.2% to 2.6%).\textsuperscript{13, 14} LS7 healthy diet score has not been reported among other rural populations, but the fruit and vegetable component of the LS7 healthy diet score has. A study of 4,754 rural Minnesota adults screened for cardiovascular disease through the Heart of New Ulm (HONU) Project found that 16.4% met the daily recommendation.\textsuperscript{12} This is similar to proportion of ACCN study participants that met the daily recommendation for fruits and vegetables (15.8%) observed by this study.

\textit{6.5.2 Limitations}

Given eligibility criteria to participate in the ACCN study, the results of this study may not be generalizable. ACCN study participants agreed to take part in a research study, regularly attended churches located in select Appalachian counties, and had to have a BMI of at least 25 to participate and, as previously reported, study participants had higher education and income levels and a lower prevalence of current smokers than what has among the region’s population as a whole.\textsuperscript{258} Therefore, future studies may wish to
utilize data from research samples that are more representative of the Appalachian region or rural and urban/rural heterogeneous areas in terms of biometric and sociodemographic characteristics.

Additionally, differences by urban/rural status were not explored. However, since the county-level exposure variable took urban/rural status into account, the utility of stratified models by urban/rural status may have been limited. Although the nested models accounted for ARC county economic strength, the nested models did not control for state. ACCN participants spanned multiple states that are part of different sub-regions of Appalachia known to differ in terms of education, poverty, race/ethnicity and access to healthcare.\textsuperscript{30, 31} Sensitivity analyses were conducted using the nested models to see whether stratifying the models by state impacted the results. The models did not converge for participants who lived in Kentucky and West Virginia, likely due to small sample sizes relative to the complexity of the models. The results of the nested models for participants who lived in Ohio and Virginia did not differ from the primary analyses. Additionally, this study only included two levels of geographic food access: county and individual. Given the null findings of this study, it is unclear whether the inclusion of a third level of geographic food access, such as CT, would have impacted the results.

This study was also limited by its use of cross sectional data. Causal relationships cannot be deduced and the study could not account for important factors such as temporality and duration of exposure that are inherently part of the relationship between geographic food access and diet. The employed USDA measure does not reflect constant changes occurring in a county’s geographic food environment due to store openings and
closings. Consequently, this study is unable to speak to the amount of time it takes for a change in geographic food access result in a change in diet behavior.260

Furthermore, by using a composite score as the outcome of interest, the study did not assess the relationship between geographic food access measures at the county- and individual-level and individual facets of diet, such as fruit and vegetable consumption which has an established association with reduced chronic disease risk.12 A sensitivity analysis using meeting the recommended LS7 daily fruit and vegetable intake guideline as the outcome of interest found neither the county-level USDA percent low food access nor the individual-level, home-to-preferred food store distance model to be significant in adjusted models. Adjusted for potential confounders, the county-level geographic food access metric was not associated with a significant change in the odds of meeting the recommended LS7 daily fruit and vegetable intake guideline (OR: 0.99, 95% CI: 0.96-1.03, p-value: 0.78, QIC: 338.26). Adjusted for potential confounders, including county-level geographic food access metric, the individual-level geographic food access metric was not associated with a significant change in the odds of meeting the recommended LS7 daily fruit and vegetable intake guideline (OR: 1.00, 95% CI: 0.97-1.04, p-value: 0.83, QIC: 339.91). This sensitivity analysis suggests it may be unnecessary to examine the relationship between geographic food access measures at the county- and individual-level and individual facets of diet in addition to a composite diet score.
6.6 Conclusion

The CVD and cancer disparities experienced by residents of Appalachia are influenced by wide-ranging social determinants of health (Figure A.4). Geographic food access is a social determinant of health that has the potential to contribute to CVD and cancer health disparities through diet, a modifiable, individual-level risk factor driving the country’s chronic disease burden. Research does not typically reflect the potential of county- and individual-level geographic food access to independently influence diet and that an individual’s geographic food access is influenced by where they live. In the current study, neither a county- nor an individual-level geographic food access metric significantly impacted the odds of poor LS7 diet score among ACCN study participants and the strength of the association was nearly equal across the two models. Based on QIC values, the adjusted model with USDA county-level percent low food access as the sole covariate fit the data better than the adjusted model with individual-level food-distance in miles as the sole covariate, but this model lacked practical significance since the effect estimate was weak in magnitude and not statistically significant.

These findings suggest geographic food access does not significantly influence diet in the study population. The primary data collected to derive the individual-level geographic food access measure provided insight into non-geographic factors influencing food purchasing behavior of ACCN study participants and raised questions regarding the interaction between geographic food access and food purchasing behavior in influencing diet. Consequently, future research should to assess whether geographic food access is
more appropriately treated as an EMM of relationship between food purchasing behavior and diet rather than as the exposure of interest.
Chapter 7: Conclusions and Implications

Although CVD is currently the leading cause of mortality in the U.S., cancer is expected to become the leading cause of mortality within two decades.\textsuperscript{1, 2} These diseases are important public health issues for the general U.S. population as well as among subgroups in the U.S. that experience health disparities.\textsuperscript{3} Smoking, overweight and obesity, inadequate physical activity, poor diet, high blood pressure, high total cholesterol and high plasma glucose are modifiable risk behaviors and factors influencing CVD and cancer outcomes through similar inflammatory pathways. These mutual risk behaviors and factors highlight the potential for cost-effective interventions to prevent CVD and cancer outcomes in the general population, as well as reduce related health disparities to reach U.S. public health benchmarks.\textsuperscript{20, 21}

When specifically seeking to reduce and ultimately eliminate CVD and cancer health disparities, the high-risk prevention strategy, with emphasis on the social determinants of health, is the approach supported by the literature. By addressing the social of determinants of health underlying disparate CVD and cancer burdens among a given population of high-risk individuals, there is potential to shift the distribution of CVD and cancer among a given population that experiences disparities to a more favorable direction.\textsuperscript{93, 94} However, a better understanding of the status of overlapping CVD and cancer risk behaviors and factors among populations experiencing disparate
CVD and cancer outcomes, and the impact of social determinants on their CVD and cancer health, must first be understood.\textsuperscript{4, 5}

The current study used LS7 to assess the distribution of mutual CVD and cancer risk behaviors and factors, and explore geographic food access as a social determinant of diet, the leading cause of the overall disease burden and the least prevalent ideal LS7 component in the U.S., among research samples from two relatively understudied populations facing health disparities: mainland Puerto Ricans and residents of Appalachia.\textsuperscript{13, 14} The conclusions and implications of each aim are summarized below, with emphasis on broad implications for health disparities research and intervention, and specific implications for future studies directly related to the scope of this research.

7.1 Aim 1 conclusions and implications

The first aim of this study quantified the distribution of each LS7 component and an overall LS7 score within BPRHS and ACCN study participants and compared the distributions to one another and NHANES data. This approach was used to identify precisely where these research samples differed in terms of CVD and cancer risk so that modifiable CVD and cancer risk factors in each study population could be prioritized for future research and intervention. Since BRPHS participants fared worse than ACCN study participants and NHW NHANES participants in terms of smoking, BMI, and plasma glucose and ACCN study participants fared worse than BPRHS and NHW NHANES participants in terms of BMI and blood pressure, these behaviors and factors should be prioritized for future research and intervention. Diet should also be prioritized.
in both study populations given the universal low prevalence of ideal diet observed and since diet is the leading cause of the overall U.S. disease burden.\textsuperscript{13}

### 7.1.1 Strengths and limitations of Life’s Simple 7 (LS7)

As noted in the literature, strengths of the AHA’s LS7 of CVH metric include well-defined poor, intermediate and ideal levels of these healthy behaviors and factors and overall LS7 score is associated with both CVD and cancer outcomes.\textsuperscript{1, 5, 7, 8} By allowing for gradients in risk, the LS7 metric is well-suited to track progress toward ideal status and it even takes changes in medication to control blood pressure, total cholesterol and plasma glucose into account. However, as demonstrated by this study, ideal status may be defined for a given component so strictly that is not attainable by the majority of the population, such as LS7 healthy diet score. Although the strict definition of ideal LS7 healthy diet score may impact research given limited variability, it is important to remember that the metric has been defined in terms of what is ideal from a prevention perspective.

Similarly, as demonstrated by this study, it is not always possible to quantify all seven components to calculate an overall LS7 score when conducting secondary data analysis. Furthermore, studies collecting primary data may be unable to research and intervene on each individual source of risk given financial, time and other logistical constraints. Fortunately, these risk behaviors and factors share similar inflammatory pathways, and are causally-related to a degree such that intervening on one source of risk will likely lead to changes other sources of risk. For example, if the goal of a study is to decrease BMI, there will likely be immediate increases in physical activity and changes
to improve diet quality, which in turn will likely lead to eventual reductions in blood pressure, cholesterol and plasma glucose. However, it is still important to identify the sources of risk for which a population facing CVD and cancer health disparities has unfavorable distributions so that researchers have an idea of where to invest their finite resources. Thus, despite these aforementioned limitations, the LS7 metric is a useful tool to determine the modifiable risk behaviors and factors that should be prioritized for additional research and future intervention among one or more populations that suffer from health disparities.

7.2 Aim 2 and 3 justification

Just as few studies have explored LS7 in more than one research sample, few have applied a multi-level perspective when quantifying the distribution of the LS7 metric among community samples from populations facing CVD and cancer disparities. Diet has been documented as the leading cause of the overall disease burden in the U.S. and the least prevalent ideal LS7 component among U.S. adults. The findings of Aim 1 that ideal healthy diet score was the least prevalent LS7 component across both research samples and healthy diet scores were low among both groups further supports the importance of limiting the scope of Aims 2 and 3 to the exploration of geographic food access as a social determinant of health with the potential to impact CVD and cancer health disparities through diet, as measured by LS7 healthy diet score, among BPRHS and ACCN study participants, respectively.

Although imperfect, publicly-available area-level geographic food access metrics are commonly used as a proxy for individual-level geographic food access since they are
less costly than collecting primary data. These measures were created at the national level and may have limited internal validity when used among specific subpopulations, such as the populations facing health disparities of interest in this study. For example, misclassification of the geographic food access of mainland Puerto Ricans may occur since bodegas and other food retailers that carry culturally-appropriate foods may not be reflected by these measures. Similarly, misclassification of the geographic food access of residents of Appalachia may occur since the distance thresholds applied to the general U.S. population may not reflect how far residents of the Appalachian region are accustomed to traveling to make food purchases. It is equally important to understand these issues from methodological and intervention standpoints, since the objectives of studying the relationship between geographic food access and diet are to ultimately be able to model the relationship as accurately as possible and identify where changes can be made to improve dietary health outcomes.

7.2.1 Aim 2 conclusions and implications

The second aim of this study was to quantify the association between a CT-level food access metric and individual-level LS7 healthy diet score in a research sample of mainland Puerto Ricans, controlling for potential confounding by individual- and CT-sociodemographics, and assessing for EMM by individual- and CT-level social determinants of health indicative of acculturation. When studying geographic food access in urban areas, there is at least consistency across the literature to use the CT as the unit of analysis. However, while economic access is normally controlled for using individual- and/or area-level SES measures, geographic food access studies typically do
not also consider informational food access, a tendency that oversimplifies the relationship between geographic food access and diet among ethnic minorities for whom culture is known to influence diet. This study did not find either publicly-available CT-level geographic food access measure to be significantly associated with diet, and these associations were not modified by culturally relevant individual- and area-level indicators of informational food access. However, further exploration of the relationship between geographic food access and diet among ethnic minority populations that assesses the impact of informational food access is warranted, since such an integrated approach has been absent from the literature.

7.2.1.1 Suggestions for future research

Given the limited internal validity of publicly available CT-level geographic food access measures, future studies conducted among BPRHS participants could collect primary individual-level geographic food access data to better understand where these individuals shop for food, how far they travel to make food purchases and by what means, and what factors influence their food purchasing decisions. Similarly, there are additional indicators of individual- and area-level acculturation that could be explored as forms of informational food access, such as individual-level psychological acculturation and length of time in the U.S. mainland and CT-level data collection by the census on language use, English-speaking ability and linguistic isolation.

In addition to a better understanding of the limitations of using one CT-level geographic food access measure over another, this research also offers insight for future
studies into potential confounders, and the anticipated magnitude and statistical significance of effect estimates. At first glance, the relative measure, mRFEI score, appears to be a preferable metric since it considers the balance of healthy and unhealthy food retailers in a CT, whereas USDA measures, such as the CT-level low access variable used by this study, only considers access in relation to large grocery stores and supermarkets. However, mRFEI is based on 2000 U.S. CT definitions and 2009 NAICS business codes. mRFEI is outdated compared to USDA measures, which have been updated using the current decennial U.S. Census CT definition and recent population and U.S. food retail industry data.

Another issue to be aware of is that ethnic minority study participants living in dense urban areas likely venture outside of their CT to purchase culturally-appropriate food, and that these food retailers may not be captured by publicly-available CT-level geographic food access measures, a limitation of the current study that may have contributed to the null findings and that could be explored in future research. Additionally, studies to better elucidate the geographic food access and diet relationship, and how it is influenced by culture can use the employed DAG as a methodological tool to inform the inclusion of potential confounding variables, whereas the reported effect estimates can be used to inform power calculations.

It is important to acknowledge alternative analytic approaches that could be used in future research. This study used GEE to yield a population-averaged model while accounting for LS7 healthy diet scores among participants in the same CT. A random effects model could have been used to account for correlated LS7 healthy scores, while
allowing for subject-specific effects that differ from the population average. Similarly, DAG-identified potential confounders could be included in a model for the purpose of predicting poor LS7 healthy diet score or as exposures of interest to quantify their association with LS7 healthy diet score.

Individual-level linguistic acculturation and CT-level percent Hispanic/Latino and percent Puerto Rican were suggested by the literature to have a protective effect on diet, but that was never confirmed by the current study, which focused on these variables in the context of EMM. Consequently, the association between these variables and LS7 healthy diet score could be further explored among BPRHS participants.

**7.2.2 Aim 3 conclusions and implications**

The third aim of this study was to compare nested models adjusted for individual- and area-level sociodemographic characteristics to determine whether using county-level food access metric or individual-level home-to-preferred food store distance as the exposure of interest is more strongly associated with individual-level LS7 healthy diet score in a research sample of residents of Appalachia, a region that is urban/rural heterogeneous. In the final adjusted models, neither county-level USDA percent low food access nor individual-level, home-to-preferred food store distance were significantly associated with the odds of having a poor LS7 healthy diet score. As noted in the literature, when studying of geographic food access in an urban/rural heterogeneous area, county-level proxy measures may not reflect food shopping realities like the loss traditional food retailers (e.g., grocery stores and supermarkets) and increased
patronization of non-traditional food retailers (e.g., dollar stores and big-box stores).\textsuperscript{262-264}

Despite the known limitations of county-level proxy measures based on distance to traditional food retailers, they are still widely used in place of individual-level geographic food access, even though individual- and county-level geographic food access have the ability to impact diet independently and individual-level geographic food access is influenced by where someone lives. This study found the relationship between a county-level geographic food access proxy variable and LS7 diet score to be similar to the individual-level geographic food access variable and diet, which, at first glance, suggests investment in the collection of costly and time-consuming individual-level geographic food access data is not necessary. However, findings also suggested that geographic food access, measured at either level, does not greatly influence diet as measured by the LS7 healthy diet score among ACCN study participants, suggesting there is utility in collecting individual-level data to better understand how non-geographic factors impact food purchasing behavior among individuals living in rural and urban/rural heterogeneous areas.

7.2.1.1 Suggestions for future research

The primary data collected to derive the individual-level geographic food access metric highlighted a number of factors that may contribute to null associations between county- and individual-level measures of geographic food access, such as the variety of retail and non-retail locations from which participants obtain food, average travel distances that exceed standard distance thresholds used by publicly-available geographic food access measures, relatively equal impact of low prices, good selection and location
on store patronization, and how quality and cost impact purchase and consumption of food that is geographically accessible.

The impact of commuting patterns spatiotemporal static geographic measures should also be explored in rural and urban/rural heterogeneous regions like Appalachia since commuting patterns have the potential to modify the impact of residence-based geographic food access on diet.\textsuperscript{270, 271} There are a number of ways in which future research among rural and more urban/heterogeneous areas could examine these factors to better reflect the complexity of the geographic food access and diet relationship. For example, a future study could assess whether trip-linking to make food purchases (e.g., buying groceries on the way to or from work or another destination) is associated with diet, controlling for relevant confounders. County-level geographic food access could be treated as a confounding variable, since an individual’s diet and decision to link trips to make food shopping are likely influenced by where they live, or as an EMM to determine whether the relationship between trip linking and diet differs between people living in counties with and without a high proportion of residents that have low access.

Research that includes a greater number of counties may wish to further explore the non-linear trend between USDA county-level percent low food access and LS7 diet score suggested by this study’s model diagnostics using nested GEE models.\textsuperscript{251} Lastly, studies could expand upon the employed DAG to inform the inclusion of potential confounding variables and use the reported effect estimates to inform power calculations. As with Aim 2, there are alternative analytic approaches that could be used in future research. Like GEE, a random-effects model could be used to account for correlated LS7
healthy diet scores among ACCN study participants in the same county. However, unlike GEE, a random effects model would allow for subject-specific effects that differ from the population average. Similarly, DAG-identified potential confounders could be used for the purpose of predicting poor LS7 healthy diet score or as exposures of interest. Exploration of differences between participants’ individual-level and county-level geographic food access fell outside of the scope of the current study, but could be pursued by future research among ACCN study participants.

7.3 Broader conclusions and implications

7.3.1 Advancing the geographic food access literature to inform policy

The lack of consistency regarding the relationship between GIS-based geographic food access measures and diet across the literature suggests this relationship is not straightforward and that it may be premature to intervene on the geographic food environment if the goal is to improve diet. An important implication related to Aim 2 and Aim 3 is the need to be able to better compare and contrast findings across studies of the relationship between geographic food access and diet. Although it is critical to develop new measures and modify existing geographic food access measures, it is also important to come to a consensus as to what measures are considered standard, especially in an area of research for which findings have been null or in mixed directions. The use of standard geographic food access measures in addition to new and/or modified geographic food access measures and explicitly stated methods would allow results to be
replicated and compared across studies focused on different sociodemographic groups, such as specific cultures or geographic locations.

In addition to methodological reasons, there are practical reasons for fostering more unity in this line of inquiry such as addressing diet as the leading modifiable factor contributing to country’s overall disease burden and informing expanding policy-driven efforts to improve geographic food accessibility. Current policy is focused on access to supermarkets and large grocery stores, a narrow definition of access that is limited in its ability to impact diet quality since these traditional food retailers offer both unhealthy and healthy options. Having more consistency in terms of the measures used would allow this field of study to build a stronger body of evidence to determine how geographic food access impacts diet and determine whether intervening to improve geographic food access, through the addition of traditional food retailers like large grocery stores and supermarkets or through more novel means, will lead to improved diet.

To inform geographic food access policies to produce intended health benefits, future research must also reflect the complexity of the geographic food access and diet relationship, and collect longitudinal data to support a causal relationship rather than relying on cross-sectional data. Researchers may collect data before and after implementing an intervention to modify the geographic food environment or take advantage of changes in the geographic food environment unrelated to research. For example, when ACCN study participants in Vinton County completed baseline measurements in Spring/Summer of 2012, including the survey pertaining to food purchasing behavior and the FFQ, there was one grocery store serving the county’s
13,200 residents. However, in September of that same year, the grocery store closed.\textsuperscript{273} To date, Vinton County still lacks a grocery store. The financial implications of the county’s sparse and aging population have deterred businesses from opening a new location.\textsuperscript{273} However, Ohio lawmakers have recently set HFFI funding at 2 million dollars, which means that Vinton County can take part in a competitive application process to secure the necessary funding for a new store, which is estimated to cost 1.5 million dollars.\textsuperscript{274} These events in Vinton County represent a natural experiment that could be conducted among ACCN study participants who still live in the county. Food purchasing behaviors and diet assessed prior to the closing of the county’s grocery store could be compared to measures obtained while the county lacked a grocery store. Additionally, baseline measurements could potentially be compared to measurements after the receipt and use of HFFI funds to improve the county’s geographic food access by adding a new grocery store.

7.3.2 Improving diet in the interim

The relationship between geographic food access and diet does not need to be perfectly delineated in order for action to be taken to improve the diet-related health outcomes of BPRHS and ACCN study participants. This study did not find lack of access to be strongly associated diet among BPRHS and ACCN study participants, but consumption of fruits and vegetables, fish, sodium, sweets/sugar sweetened beverages and whole grains needs to be improved in both research samples. Among other factors not explored by this study, poor diet may be partially due to limited knowledge about what constitutes a healthy diet and how diet impacts health; the cost, quality and cultural
appropriateness of food that is geographically accessible; culturally-relevant norms related to food prepared; and reliance on processed food due to limited time to plan, shop for or cook meals. By directly asking study participants why they make the food purchasing and consumption choices they do, researchers and public health officials can provide education and outreach to improve diet within cultural, financial and time constraints; identify non-geographic determinants of diet for future research and intervention; and advocate for changes to facilitate healthier food purchasing and consumption behaviors.

7.3.2 Collecting more representative data on populations facing health disparities

The conclusions and implications of this study cannot be discussed without acknowledging the need for more representative data on the health status of mainland Puerto Ricans and residents of Appalachia. As noted in the literature, publicly available health data on the health status of mainland Puerto Ricans and residents of Appalachia are limited, and so exploration of specific research questions may only be possible by using data collected from studies focused on a very specific subset of these populations that employ eligibility criteria that further restricted the study population’s representativeness. For example, BPRHS participants were 45 to 75 years of age and were recruited using approaches such as door-to-door enumeration of randomly selected census blocks with 10 or more Hispanics within the target age range. ACCN study participants were adults recruited from churches located in select Appalachian counties, who agreed to participate and had a BMI of at least 25 to participate; this is not necessarily representative of the Appalachian population.
Given their age, BPRHS participants are well suited for the study of allostatic load and the accumulation of modifiable CVD and cancer risk factors and behaviors over time. However, in terms of CVD and cancer prevention, a younger cohort of mainland Puerto Ricans would have been preferred since early onset of disease has been noted as contributing factor to health disparities between ethnic minority populations and non-Hispanic whites, and intervention earlier in life has more potential to reduce disparate impacts of CVD and cancer. Similarly, a cohort of Appalachian residents more heterogeneous in terms of BMI may have yielded more informative results related to mutual, modifiable CVD and cancer risk factors. To more appropriately study the health disparities among these under researched populations, state and national surveillance efforts likes BRFSS and NHANES should be adapted to include sufficient numbers of Hispanic subgroups and an indicator of Appalachian status. Such a sample would produce reliable estimates among mainland Puerto Ricans and foster timely comparisons between residents of the Appalachian counties in 13 states and their non-Appalachian counterparts. Researchers and organizations that rely on BRFSS and NHANES data to make policy and funding decisions that impact the health of populations facing health disparities are negatively impacted by the aforementioned data limitations, such as the ARC, and thus, are uniquely positioned to advocate for these changes.

7.4 Closing remarks

LS7 can be used to identify modifiable sources of CVD and cancer risk for which a population has unfavorable distributions. Ideally, LS7 should be quantified using data collected from research samples that are representative of the population(s) facing health
disparities of interest. However, as demonstrated by this study, obtaining these data is not always possible which means results cannot be generalized to the broader health disparity population(s). Consequently, the results of Aim 1 apply to the research samples, not to all mainland Puerto Ricans and residents of Appalachia. Modifiable risk behaviors and factors to prioritize among BRPHS participants include smoking, BMI, and plasma glucose. Modifiable risk factors and behaviors to prioritize among ACCN study participants include BMI and blood pressure. Given the universal low prevalence of ideal diet observed, and that diet is the leading cause of the overall U.S. chronic disease burden, diet should also be prioritized among BPRHS and ACCN study participants.13

Differences in LS7 components between populations do not explain why CVD and cancer health disparities exist. There are a wide range of multi-level social determinants of health that have the potential to influence disparate CVD and cancer health outcomes independently or through interaction across levels (Figure A.4). Consequently, the LS7 metric must be integrated with data from electronic health records, U.S. Census and NAICS business codes (used to determine access to health promoting resources such as grocery stores, hospitals, health clubs and recreational sports centers) to identify social determinants of health contributing to CVD and cancer health disparities among mainland Puerto Ricans and residents of Appalachia.18, 19 Using this information, intervention efforts to improve health outcomes in these study populations can focus on individual-level behavior change in the context of the social determinants of health as well as modify the environment in which participants live.
Exploration of the social determinants of health and the LS7 components identified in Aim 1 as warranting priority for research and intervention among BPRHS and ACCN study participants was limited for the purpose of this study. Aims 2 and 3 pertained to the relationship between geographic food access and LS7 healthy diet score among BPRHS and ACCN study participants, and the results do not suggest geographic food access, as measured, is a social determinant of health on which to intervene to improve diet among the study populations of interest.

As suggested by the conceptual model (Figure A.4) and literature on the health of mainland Puerto Ricans and Appalachian residents, there are many other social determinants of health to explore in the context of diet. Research pertaining to the other prioritized risk factors and behaviors can target an equally expansive number of social determinants of health. For example, relevant social determinants of health explored among BPRHS study participants include social isolation, networks and support; neighborhood characteristics like crime, building conditions, green space, and density; and environmental exposures due to roadway proximity and traffic density. Similarly, social determinants of health relevant to ACCN and other research samples of Appalachian residents include geographic isolation, access to healthcare, physical activity resources, and the faith community.

Although limited in its scope, the current study used the LS7 metric to shed light on the social determinants of health influencing the CVD- and cancer-related health of BPRHS and ACCN study participants, two research samples derived from understudied populations facing health disparities. However, to address the disparate CVD and cancer
health outcomes experienced by mainland Puerto Ricans and residents of Appalachia, future studies need to be conducted using research samples and data that are more representative of these populations. More generalizable results and improved understanding of the multi-level social determinants of health that contribute to disparate CVD and cancer health outcomes in these populations are needed to inform the efforts of researchers, public health officials and policy makers seeking to reduce and ultimately eliminate these disparities.
Appendix A: Figures
Figure A.1: Appalachian counties by FY2012 Economic Status
Figure A.2: ARC-defined Appalachian sub-regions
Figure A.3: Relationship between modifiable CVD and cancer risk factors and behaviors
Figure A.4: CPHHD model for analysis of population health and health disparities

"Fundamental Causes"

"Social & Physical Context"

"Individual Demographic & Risk Factors"

"Individual Demographics"

"Biologic Responses"

"Biologic/Genetic Pathways"

"Distal Factors"

"Institutional Context"

"Social Context"

"Social Relationships"

"Physical Context"

"Intermediate Factors"

"Proximal Factors"
Figure A.5: Initial DAG for Aim 2
Figure A.6: Final DAG for Aim 2
Figure A.7: Initial DAG for Aim 3
Figure A.8: Final DAG for Aim 3
Figure A.9: Unadjusted prevalence of poor, intermediate and ideal LS7 metric components of BPRHS participants
*For plasma glucose, the red category represents not ideal, rather than poor.

Figure A.10: Unadjusted prevalence of poor, intermediate and ideal LS7 metric components of ACCN study participants
Figure A.11: Unadjusted prevalence of poor, intermediate and ideal LS7 metric components of NHW 2007-2010 NHANES participants.
For plasma glucose, the red category represents not ideal, rather than poor.

Figure A.12: Comparison of age- and sex-adjusted LS7 metric component prevalence among BPRHS and ACCN study participants

*For plasma glucose, the red category represents not ideal, rather than poor.
Figure A.13: Comparison of age- and sex-adjusted LS7 metric component prevalence among BPRHS and NHW 2007-2010 NHANES participants
*For plasma glucose, the red category represents not ideal, rather than poor.

Figure A.14: Comparison of age- and sex-adjusted LS7 metric component prevalence among ACCN study and NHW 2007-2010 NHANES participants
Appendix B: Tables
<table>
<thead>
<tr>
<th>Health Behaviors</th>
<th>Poor</th>
<th>Intermediate</th>
<th>Ideal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking</td>
<td>Current</td>
<td>Former (quit for ≤1 year)</td>
<td>Never or quit for &gt;1 year</td>
</tr>
<tr>
<td>BMI</td>
<td>≥30 kg/m²</td>
<td>25-&lt;30 kg/m²</td>
<td>&lt;25 kg/m²</td>
</tr>
<tr>
<td>Physical Activity</td>
<td>no moderate or vigorous activity</td>
<td>1-74 min/wk vigorous, 1-149 min/wk moderate or equivalent combination</td>
<td>≥75 min/wk vigorous, ≥150 min/wk moderate, or equivalent combination</td>
</tr>
<tr>
<td>Diet*</td>
<td>Healthy diet score 0-1</td>
<td>Healthy diet score 2-3</td>
<td>Healthy diet score 4-5</td>
</tr>
<tr>
<td>Blood Pressure</td>
<td>≥140/90 mm Hg</td>
<td>treated to &lt;120/&lt;80 mm Hg or 120-139/80-89 mm Hg</td>
<td>untreated &amp; &lt;120/&lt;80 mm Hg</td>
</tr>
<tr>
<td>Total Cholesterol</td>
<td>≥240 mg/dL</td>
<td>treated to &lt;200 mg/dL or 200-239 mg/dL</td>
<td>untreated &amp; &lt;200 mg/dL</td>
</tr>
<tr>
<td>Plasma Glucose</td>
<td>≥126 mg/dL</td>
<td>treated to &lt;100 mg/dL or 100-125 mg/dL</td>
<td>untreated &amp; &lt;100 mg/dL</td>
</tr>
</tbody>
</table>

*Healthy diet score is comprised of 5 components worth 1 point each: ≥4.5 cups per day of fruits/vegetables, ≥2 servings of fish per week (3.5-oz servings), <1500 mg/d of sodium, ≤450 kcal (36 oz) per week of sweets/sugar-sweetened beverages, and ≥3 servings per day of whole grains (1.1 g of fiber in 10 g of carbohydrate; 1-oz equivalent servings).
Avoid tobacco use and environmental tobacco smoke.

Achieve and maintain a healthy weight throughout life.

- Be as lean as possible throughout life without being underweight.
- Avoid excess weight gain at all ages. For those who are currently overweight or obese, losing even a small amount of weight has health benefits and is a good place to start. A BMI > 25 is suggested.
- Engage in regular physical activity and limit consumption of high-calorie foods and beverages as key strategies for maintaining a healthy weight.

Adopt a physically active lifestyle.

- Adults should engage in at least 150 minutes of moderate intensity or 75 minutes of vigorous intensity activity each week, or an equivalent combination, preferably spread throughout the week.
- Children and adolescents should engage in at least 1 hour of moderate or vigorous intensity activity each day, with vigorous intensity activity occurring at least 3 days each week.
- Limit sedentary behavior such as sitting, lying down, watching television, or other forms of screen-based entertainment.
- Doing some physical activity above usual activities, no matter what one’s level of activity, can have many health benefits.

Consume a healthy diet, with an emphasis on plant foods.

- Choose foods and beverages in amounts that help achieve and maintain a healthy weight.
- Limit consumption of processed meat and red meat.
- Eat at least 2.5 cups of vegetables and fruits each day for cancer risk reduction. For overall health, eat at least 2 cups of vegetables and 1.5 cups of fruit daily.
- Choose whole grains instead of refined grain products.

If you drink alcoholic beverages, limit consumption.

- Drink no more than 1 drink per day for women or 2 per day for men.
Table B.3: Summary of relationship between data sources and specific aims

<table>
<thead>
<tr>
<th>Data Sources</th>
<th>BPRHS</th>
<th>ACCN</th>
<th>External Data</th>
<th>Geographic food access</th>
<th>U.S. Census</th>
<th>ARC</th>
<th>NHANES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CDC</td>
<td>USDA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aim 1</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Aim 2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Aim 3</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Table B.4: LS7 components derivable from BPRHS and ACCN datasets

<table>
<thead>
<tr>
<th>LS7 Metric</th>
<th>BPRHS</th>
<th>ACCN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>BMI</td>
<td>X</td>
<td>Poor &amp; Intermediate Only</td>
</tr>
<tr>
<td>Physical Activity</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Diet</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Blood Pressure</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Total Cholesterol</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Plasma Glucose</td>
<td>X</td>
<td>Ideal &amp; Not Ideal Only</td>
</tr>
<tr>
<td>Overall LS7 Score Range</td>
<td>0-7</td>
<td>0-5</td>
</tr>
</tbody>
</table>
Table B.5: Comparison of age- and sex-adjusted LS7 metric component prevalence among BPRHS and ACCN study participants

<table>
<thead>
<tr>
<th></th>
<th>BPRHS†</th>
<th>ACCN‡</th>
<th>Wald-statistic</th>
<th>DF</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants, n</td>
<td>934</td>
<td>313</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS7 Metric Prevalence,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% (99% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>44.3 (39.6-49.0)</td>
<td>68.9 (61.7-76.2)</td>
<td>54.34</td>
<td>1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Intermediate</td>
<td>32.3 (27.8-36.8)</td>
<td>27.3 (20.2-34.4)</td>
<td>2.32</td>
<td>1</td>
<td>0.1282</td>
</tr>
<tr>
<td>Poor</td>
<td>23.4 (19.3-27.5)</td>
<td>3.7 (0.5-6.9)</td>
<td>95.86</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td>36.8 (32.2-41.5)</td>
<td>37.9 (29.9-46.0)</td>
<td>0.10</td>
<td>1</td>
<td>0.7551</td>
</tr>
<tr>
<td>Poor</td>
<td>63.2 (58.5-67.8)</td>
<td>62.1 (54.0-70.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weekly Physical Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>73.8 (69.8-77.8)</td>
<td>73.8 (66.6-81.1)</td>
<td>0.00</td>
<td>1</td>
<td>0.9873</td>
</tr>
<tr>
<td>Intermediate</td>
<td>2.1 (0.8-3.5)</td>
<td>24.8 (17.6-31.9)</td>
<td>64.79</td>
<td>1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Poor</td>
<td>24.1 (20.1-28.0)</td>
<td>1.4 (0.0-3.3)</td>
<td>178.73</td>
<td>1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Diet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>0.4 (0.0-1.0)</td>
<td>1.3 (0.0-3.3)</td>
<td>1.08</td>
<td>1</td>
<td>0.0007</td>
</tr>
<tr>
<td>Intermediate</td>
<td>29.6 (25.4-33.8)</td>
<td>40.9 (32.8-49.1)</td>
<td>10.16</td>
<td>1</td>
<td>0.0015</td>
</tr>
<tr>
<td>Poor</td>
<td>70.0 (65.8-74.3)</td>
<td>57.8 (49.6-66.0)</td>
<td>11.65</td>
<td>1</td>
<td>0.2983</td>
</tr>
<tr>
<td>Blood Pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>44.9 (40.4-49.5)</td>
<td>5.2 (1.7-8.6)</td>
<td>322.40</td>
<td>1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Intermediate</td>
<td>27.3 (23.0-31.6)</td>
<td>77.4 (70.5-84.3)</td>
<td>254.11</td>
<td>1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Poor</td>
<td>27.8 (23.6-31.9)</td>
<td>17.4 (11.1-23.7)</td>
<td>12.47</td>
<td>1</td>
<td>0.004</td>
</tr>
<tr>
<td>Plasma Glucose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>27.4 (23.4-31.6)</td>
<td>92.5 (88.8-96.2)</td>
<td>918.66</td>
<td>1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Not Ideal</td>
<td>72.6 (68.4-76.7)</td>
<td>7.5 (3.8-11.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Categorical Ideal LS7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score, % (99% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6.5 (4.3-8.8)</td>
<td>3.3 (0.4-6.2)</td>
<td>5.17</td>
<td>1</td>
<td>0.0232</td>
</tr>
<tr>
<td>2 to 3</td>
<td>57.1 (52.3-61.8)</td>
<td>85.4 (79.5-91.4)</td>
<td>92.47</td>
<td>1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>≤1</td>
<td>36.4 (31.9-40.9)</td>
<td>11.2 (5.9-16.6)</td>
<td>85.83</td>
<td>1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Continuous Ideal LS7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score, mean (SD)</td>
<td>1.9 (0.04)</td>
<td>2.4 (0.04)</td>
<td>7.38</td>
<td>1245</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

† Data has been age- and sex-standardized to the 2000 U.S. standard population
Table B.6: Comparison of age- and sex-adjusted LS7 metric component prevalence among BPRHS and NHW 2007-2010 NHANES participants

<table>
<thead>
<tr>
<th>Participants, n (population estimate)</th>
<th>BPRHS</th>
<th>NHANES 2007-2010 Non-Hispanic Whites†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1067</td>
<td>587 (37,662,918)</td>
</tr>
<tr>
<td><strong>LS7 Metric Prevalence, % (99% CI)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Smoking</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>43.0</td>
<td>(38.8-47.3) 82.6 (76.1-89.2)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>31.0</td>
<td>(26.9-35.0) 3.2 (0.4-6.0)</td>
</tr>
<tr>
<td>Poor</td>
<td>26.0</td>
<td>(22.2-29.9) 14.2 (8.5-19.8)</td>
</tr>
<tr>
<td><strong>Body Mass Index</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>14.6</td>
<td>(11.3-17.9) 28.4 (20.8-36.0)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>31.1</td>
<td>(27.0-35.2) 36.2 (30.7-41.8)</td>
</tr>
<tr>
<td>Poor</td>
<td>54.2</td>
<td>(50.0-58.6) 35.4 (28.5-42.3)</td>
</tr>
<tr>
<td><strong>Weekly Physical Activity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>74.2</td>
<td>(70.5-77.9) 48.0 (40.7-55.1)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>2.2</td>
<td>(1.0-3.4) 24.7 (18.5-30.8)</td>
</tr>
<tr>
<td>Poor</td>
<td>23.6</td>
<td>(20.0-27.2) 27.4 (19.8-35.1)</td>
</tr>
<tr>
<td><strong>Diet</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>0.4</td>
<td>(0.0-0.9) 0.4 (0.0-0.9)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>29.0</td>
<td>(25.1-32.8) 36.1 (30.9-41.3)</td>
</tr>
<tr>
<td>Poor</td>
<td>70.6</td>
<td>(66.8-74.5) 63.5 (58.2-68.8)</td>
</tr>
<tr>
<td><strong>Blood Pressure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>48.0</td>
<td>(43.9-52.1) 49.5 (41.5-57.5)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>25.8</td>
<td>(22.0-29.6) 28.7 (22.4-35.0)</td>
</tr>
<tr>
<td>Poor</td>
<td>26.2</td>
<td>(22.5-29.9) 21.8 (17.1-26.6)</td>
</tr>
<tr>
<td><strong>Total Cholesterol</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>38.5</td>
<td>(34.3-42.7) 26.1 (19.9-32.3)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>53.8</td>
<td>(49.5-58.0) 54.8 (49.0-60.6)</td>
</tr>
<tr>
<td>Poor</td>
<td>7.7</td>
<td>(5.5-10.0) 19.1 (15.2-23.1)</td>
</tr>
<tr>
<td><strong>Plasma Glucose</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>30.4</td>
<td>(26.5-34.4) 43.7 (36.8-50.6)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>45.6</td>
<td>(41.3-50.0) 45.4 (38.5-52.4)</td>
</tr>
<tr>
<td>Poor</td>
<td>23.9</td>
<td>(20.3-27.6) 10.9 (8.1-13.7)</td>
</tr>
<tr>
<td><strong>Categorical Ideal LS7 Score, % (99% CI)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥5</td>
<td>7.3</td>
<td>(5.0-9.6) 9.6 (5.2-14.0)</td>
</tr>
<tr>
<td>3 to 4</td>
<td>41.1</td>
<td>(36.8-45.4) 47.0 (40.6-53.3)</td>
</tr>
<tr>
<td>≤2</td>
<td>51.6</td>
<td>(47.3-55.9) 43.3 (35.6-51.2)</td>
</tr>
<tr>
<td><strong>Continuous Ideal LS7 Score, mean (99% CI)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>(2.4-2.6) 2.8 (2.5-3.0)</td>
</tr>
</tbody>
</table>

† Data has been age- and sex-standardized to the 2000 U.S. standard population

230
Table B.7: Comparison of age- and sex-adjusted LS7 metric component prevalence among ACCN study and NHW 2007-2010 NHANES participants

<table>
<thead>
<tr>
<th>Participants, n (population estimate)</th>
<th>ACCN\textsuperscript{†}</th>
<th>NHANES 2007-2010 Non-Hispanic Whites\textsuperscript{†}</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS7 Metric Prevalence, % (99% CI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>74.4 (67.7-81.0)</td>
<td>76.0 (71.5-80.5)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>23.7 (17.1-30.3)</td>
<td>4.2 (1.7-6.6)</td>
</tr>
<tr>
<td>Poor</td>
<td>1.9 (0.3-3.5)</td>
<td>19.9 (15.0-24.7)</td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td>28.6 (22.8-34.3)</td>
<td>55.2 (49.7-60.7)</td>
</tr>
<tr>
<td>Poor</td>
<td>71.4 (65.7-77.2)</td>
<td>44.8 (39.3-50.3)</td>
</tr>
<tr>
<td>Weekly Physical Activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>70.3 (63.3-77.4)</td>
<td>49.7 (45.0-54.3)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>29.1 (22.1-36.1)</td>
<td>24.8 (19.4-30.3)</td>
</tr>
<tr>
<td>Poor</td>
<td>0.5 (0.0-1.3)</td>
<td>25.5 (20.7-30.3)</td>
</tr>
<tr>
<td>Diet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>0.5 (0.0-1.4)</td>
<td>0.4 (0.0-10.4)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>36.2 (29.3-43.2)</td>
<td>26.3 (21.8-30.8)</td>
</tr>
<tr>
<td>Poor</td>
<td>63.3 (56.3-70.2)</td>
<td>73.3 (68.8-77.8)</td>
</tr>
<tr>
<td>Blood Pressure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>7.5 (2.8-12.2)</td>
<td>60.0 (55.1-64.8)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>76.9 (68.8-85.0)</td>
<td>22.1 (18.6-25.6)</td>
</tr>
<tr>
<td>Poor</td>
<td>15.6 (8.5-22.8)</td>
<td>17.9 (15.1-20.7)</td>
</tr>
<tr>
<td>Plasma Glucose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>90.4 (86.3-94.6)</td>
<td>49.5 (42.9-56.1)</td>
</tr>
<tr>
<td>Not Ideal</td>
<td>9.6 (5.4-13.7)</td>
<td>50.5 (43.9-57.1)</td>
</tr>
<tr>
<td>Categorical Ideal LS7 Score, % (99% CI)</td>
<td>4 4.1 (1.0-7.2)</td>
<td>14.1 (10.3-17.9)</td>
</tr>
<tr>
<td></td>
<td>2 to 3 85.7 (80.0-91.4)</td>
<td>65.2 (59.5-71.0)</td>
</tr>
<tr>
<td></td>
<td>≤1 10.2 (5.4-15.0)</td>
<td>20.6 (16.4-24.9)</td>
</tr>
<tr>
<td>Continuous Ideal LS7 Score, mean (SD)</td>
<td>2.4 (2.3-2.6)</td>
<td>2.4 (2.3-2.4)</td>
</tr>
</tbody>
</table>

\textsuperscript{†} Data has been age- and sex-standardized to the 2000 U.S. standard population
Table B.8: Baseline BPRHS participant sociodemographic characteristics (Primary Analysis)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total (N=1011)</th>
<th>Living in Low Access CT (N=439)</th>
<th>Not Living in Low Access CT (N=572)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual-level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age [mean (SD)]</td>
<td>56.2 (7.6)</td>
<td>57.3 (7.8)</td>
<td>57.1 (7.5)</td>
</tr>
<tr>
<td>Gender [n (%)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>284 (28.1)</td>
<td>120 (27.3)</td>
<td>164 (28.7)</td>
</tr>
<tr>
<td>Female</td>
<td>727 (71.9)</td>
<td>319 (72.7)</td>
<td>408 (71.3)</td>
</tr>
<tr>
<td>Marital Status [n (%)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married/living as married</td>
<td>324 (30.3)</td>
<td>131 (29.8)</td>
<td>174 (30.4)</td>
</tr>
<tr>
<td>Not married/living as married</td>
<td>747 (70.7)</td>
<td>308 (70.2)</td>
<td>398 (69.6)</td>
</tr>
<tr>
<td>Poverty Status [n (%)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At or Below Poverty Line</td>
<td>588 (58.2)</td>
<td>258 (58.8)</td>
<td>330 (57.7)</td>
</tr>
<tr>
<td>Above Poverty Line</td>
<td>423 (41.8)</td>
<td>181 (41.2)</td>
<td>242 (42.3)</td>
</tr>
<tr>
<td>Language acculturation score [n (%)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 50%</td>
<td>810 (80.1)</td>
<td>354 (80.6)</td>
<td>456 (79.7)</td>
</tr>
<tr>
<td>&lt; 50%</td>
<td>201 (19.9)</td>
<td>85 (19.4)</td>
<td>116 (20.3)</td>
</tr>
<tr>
<td><strong>CT-level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent PIR Poor [mean (SD)]</td>
<td>30.2 (13.9)</td>
<td>24.9 (11.9)</td>
<td>34.2 (14.0)</td>
</tr>
<tr>
<td>Percent HL [mean (SD)]</td>
<td>36.2 (24.7)</td>
<td>38.9 (24.6)</td>
<td>34.2 (24.5)</td>
</tr>
<tr>
<td>Percent PR [mean (SD)]</td>
<td>11.6 (7.7)</td>
<td>11.2 (7.6)</td>
<td>11.9 (7.8)</td>
</tr>
</tbody>
</table>
Table B.9: Baseline BPRHS participant diet characteristics (Primary Analysis)

<table>
<thead>
<tr>
<th></th>
<th>Total (N=1011)</th>
<th>Living in Low Access CT (N=439)</th>
<th>Not Living in Low Access CT (N=572)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous Diet Score [mean (SD)]</td>
<td>1.21 (0.8)</td>
<td>1.22 (0.8)</td>
<td>1.22 (0.8)</td>
</tr>
<tr>
<td>Categorical Diet Score [n (%)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>5(0.5)</td>
<td>2 (0.5)</td>
<td>3 (0.5)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>317 (31.4)</td>
<td>134 (30.5)</td>
<td>183 (32.0)</td>
</tr>
<tr>
<td>Poor</td>
<td>689 (68.2)</td>
<td>303 (69.0)</td>
<td>386 (67.5)</td>
</tr>
<tr>
<td>Binary Diet Score [n (%)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Poor</td>
<td>322 (31.9)</td>
<td>136 (31.0)</td>
<td>186 (32.5)</td>
</tr>
<tr>
<td>Poor</td>
<td>689 (68.15)</td>
<td>303 (69.0)</td>
<td>386 (67.5)</td>
</tr>
<tr>
<td>Fruit and Vegetable [n (%)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>191 (18.89)</td>
<td>73 (16.6)</td>
<td>118 (20.6)</td>
</tr>
<tr>
<td>0</td>
<td>820 (81.11)</td>
<td>366 (83.3)</td>
<td>454 (79.4)</td>
</tr>
<tr>
<td>Fish [n (%)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>253 (25.02)</td>
<td>118 (26.9)</td>
<td>135 (23.6)</td>
</tr>
<tr>
<td>0</td>
<td>758 (74.98)</td>
<td>321 (73.1)</td>
<td>437 (76.4)</td>
</tr>
<tr>
<td>Whole Grains [n (%)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>42 (4.15)</td>
<td>14 (3.2)</td>
<td>28 (4.9)</td>
</tr>
<tr>
<td>0</td>
<td>969 (95.85)</td>
<td>425 (96.8)</td>
<td>544 (95.1)</td>
</tr>
<tr>
<td>Sodium [n (%)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0 (0.0)</td>
<td>0 (0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>0</td>
<td>1,011 (100.0)</td>
<td>439 (100.0)</td>
<td>572 (100.0)</td>
</tr>
<tr>
<td>Sweets/Sugar Sweetened Beverage [n (%)]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>746 (73.79)</td>
<td>331 (75.4)</td>
<td>415 (72.6)</td>
</tr>
<tr>
<td>0</td>
<td>265 (26.21)</td>
<td>108 (24.6)</td>
<td>157 (27.4)</td>
</tr>
</tbody>
</table>
Table B.10: Baseline BPRHS participant sociodemographic characteristics (Secondary Analysis)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total (N=680)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age [mean (SD)]</td>
<td>57.1 (7.6)</td>
</tr>
<tr>
<td>Gender [n (%)]</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>194 (28.5)</td>
</tr>
<tr>
<td>Female</td>
<td>486 (71.5)</td>
</tr>
<tr>
<td>Marital Status [n (%)]</td>
<td></td>
</tr>
<tr>
<td>Married/living as married, spouse in HH</td>
<td>200 (29.4)</td>
</tr>
<tr>
<td>Married, spouse not in HH</td>
<td>20 (2.9)</td>
</tr>
<tr>
<td>Divorced/Separated</td>
<td>275 (40.4)</td>
</tr>
<tr>
<td>Widowed</td>
<td>75 (11.0)</td>
</tr>
<tr>
<td>Never Married</td>
<td>110 (16.2)</td>
</tr>
<tr>
<td>Poverty Status [n (%)]</td>
<td></td>
</tr>
<tr>
<td>At or Below Poverty Line</td>
<td>394 (57.9)</td>
</tr>
<tr>
<td>Above Poverty Line</td>
<td>286 (42.1)</td>
</tr>
<tr>
<td>Language acculturation score [n (%)]</td>
<td></td>
</tr>
<tr>
<td>≥ 50%</td>
<td>537 (79.0)</td>
</tr>
<tr>
<td>&lt;50%</td>
<td>201 (31.0)</td>
</tr>
<tr>
<td>CT-level</td>
<td></td>
</tr>
<tr>
<td>Percent PIR Poor</td>
<td>28.8 (11.5)</td>
</tr>
<tr>
<td>Percent HL [n (%)]</td>
<td>37.2 (26.8)</td>
</tr>
<tr>
<td>Percent PR [n (%)]</td>
<td>11.6 (8.4)</td>
</tr>
<tr>
<td>mRFEI score [mean(SD)]</td>
<td>6.8 (5.1)</td>
</tr>
</tbody>
</table>
Table B.11: Baseline BPRHS participant diet characteristics (Secondary Analysis)

<table>
<thead>
<tr>
<th></th>
<th>Total (N=680)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous LS7 Healthy Diet Score [mean (SD)]</td>
<td>1.20 (0.8)</td>
</tr>
<tr>
<td>Categorical LS7 Healthy Diet Score [n(%)]</td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>3 (0.44)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>209 (30.7)</td>
</tr>
<tr>
<td>Poor</td>
<td>468 (68.8)</td>
</tr>
<tr>
<td>Binary LS7 Healthy Diet Score [n(%)]</td>
<td></td>
</tr>
<tr>
<td>Not Poor</td>
<td>212 (31.2)</td>
</tr>
<tr>
<td>Poor</td>
<td>468 (68.8)</td>
</tr>
<tr>
<td>Fruit and Vegetable [n(%)]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>129 (18.97)</td>
</tr>
<tr>
<td>0</td>
<td>551 (81.0)</td>
</tr>
<tr>
<td>Fish [n(%)]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>253 (25.02)</td>
</tr>
<tr>
<td>0</td>
<td>758 (74.98)</td>
</tr>
<tr>
<td>Whole Grains [n(%)]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>42 (4.15)</td>
</tr>
<tr>
<td>0</td>
<td>969 (95.85)</td>
</tr>
<tr>
<td>Sodium [n(%)]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0 (0)</td>
</tr>
<tr>
<td>0</td>
<td>1,011 (100)</td>
</tr>
<tr>
<td>Sweets/Sugar Sweetened Beverage [n(%)]</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>746 (73.79)</td>
</tr>
<tr>
<td>0</td>
<td>265 (26.21)</td>
</tr>
</tbody>
</table>
Table B.12: Baseline ACCN study participant sociodemographic characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total (N=379)</th>
<th>Percent county low access ≤17 (N=199)</th>
<th>Percent county low access &gt;17 (N=180)</th>
<th>Preferred food store distance ≤10 miles (N=210)</th>
<th>Preferred food store distance &gt;10 miles (N=169)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age [mean (SD)]</td>
<td>55.8 (12.9)</td>
<td>55.8 (12.3)</td>
<td>55.7 (13.5)</td>
<td>56.4 (12.8)</td>
<td>55.0 (12.9)</td>
</tr>
<tr>
<td>Gender [n (%)]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>108 (28.5)</td>
<td>57 (28.6)</td>
<td>51 (28.3)</td>
<td>60 (28.6)</td>
<td>48 (28.4)</td>
</tr>
<tr>
<td>Female</td>
<td>271 (71.5)</td>
<td>142 (71.4)</td>
<td>129 (71.7)</td>
<td>150 (71.4)</td>
<td>121 (71.6)</td>
</tr>
<tr>
<td>Marital Status [n (%)]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married/living with partner</td>
<td>302 (79.7)</td>
<td>161 (80.9)</td>
<td>141 (78.3)</td>
<td>154 (73.3)</td>
<td>148 (87.6)</td>
</tr>
<tr>
<td>Not married/living with partner</td>
<td>77 (20.3)</td>
<td>38 (19.1)</td>
<td>39 (21.7)</td>
<td>56 (26.7)</td>
<td>21 (12.4)</td>
</tr>
<tr>
<td>Annual Household Income [n (%)]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;$20,000</td>
<td>19 (5.0)</td>
<td>8 (4.0)</td>
<td>11 (6.1)</td>
<td>11 (5.2)</td>
<td>8 (4.7)</td>
</tr>
<tr>
<td>$20,000 - $59,999</td>
<td>184 (48.6)</td>
<td>102 (51.3)</td>
<td>82 (45.6)</td>
<td>96 (45.7)</td>
<td>88 (52.1)</td>
</tr>
<tr>
<td>$60,000 - $99,999</td>
<td>120 (31.7)</td>
<td>54 (27.1)</td>
<td>66 (36.7)</td>
<td>71 (33.8)</td>
<td>49 (29.0)</td>
</tr>
<tr>
<td>≥$100,000</td>
<td>56 (14.8)</td>
<td>35 (17.6)</td>
<td>21 (11.7)</td>
<td>32 (15.2)</td>
<td>24 (14.2)</td>
</tr>
<tr>
<td>Preferred food store in miles distance [mean (SD)]</td>
<td>10.5 (9.6)</td>
<td>8.3 (10.0)</td>
<td>13.0 (8.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>County-level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARC County Economic Status FY2012 [n (%)]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economically Distressed</td>
<td>178 (46.9)</td>
<td>89 (44.7)</td>
<td>89 (49.4)</td>
<td>129 (61.4)</td>
<td>72 (42.6)</td>
</tr>
<tr>
<td>Not Economically Distressed</td>
<td>201 (53.0)</td>
<td>110 (55.3)</td>
<td>91 (50.6)</td>
<td>81 (38.6)</td>
<td>97 (57.4)</td>
</tr>
<tr>
<td>Percent Low Access [mean (SD)]</td>
<td>16.9 (8.1)</td>
<td>13.8 (7.7)</td>
<td>20.8 (6.9)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table B.13: Baseline ACCN study participant diet characteristics

<table>
<thead>
<tr>
<th></th>
<th>Total (N=379)</th>
<th>Percent county low access ≤ 17 (N=199)</th>
<th>Percent county low access &gt; 17 (N=180)</th>
<th>Preferred food store distance ≤ 10 miles (N=210)</th>
<th>Preferred food store distance &gt; 10 miles (N=169)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous Diet Score</td>
<td>1.5 (0.7)</td>
<td>1.5 (0.7)</td>
<td>1.4 (0.7)</td>
<td>1.5 (0.7)</td>
<td>1.4 (0.6)</td>
</tr>
<tr>
<td>Categorical Diet Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal</td>
<td>4 (1.1)</td>
<td>2 (1.0)</td>
<td>2 (1.1)</td>
<td>2 (0.9)</td>
<td>2 (1.2)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>139 (36.7)</td>
<td>81 (40.7)</td>
<td>58 (32.2)</td>
<td>86 (41.0)</td>
<td>53 (31.4)</td>
</tr>
<tr>
<td>Poor</td>
<td>236 (62.7)</td>
<td>116 (58.3)</td>
<td>120 (66.7)</td>
<td>122 (58.1)</td>
<td>114 (67.5)</td>
</tr>
<tr>
<td>Binary Diet Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥1</td>
<td>143 (37.7)</td>
<td>83 (41.7)</td>
<td>60 (33.3)</td>
<td>88 (41.9)</td>
<td>55 (32.5)</td>
</tr>
<tr>
<td>0</td>
<td>236 (62.3)</td>
<td>116 (58.3)</td>
<td>120 (66.7)</td>
<td>122 (58.1)</td>
<td>114 (67.5)</td>
</tr>
<tr>
<td>Fruit and Vegetable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>60 (15.8)</td>
<td>32 (16.1)</td>
<td>28 (15.6)</td>
<td>34 (16.2)</td>
<td>26 (15.4)</td>
</tr>
<tr>
<td>0</td>
<td>319 (84.2)</td>
<td>167 (83.9)</td>
<td>152 (84.4)</td>
<td>176 (83.8)</td>
<td>143 (84.6)</td>
</tr>
<tr>
<td>Fish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>44 (11.6)</td>
<td>23 (11.6)</td>
<td>21 (11.7)</td>
<td>31 (14.8)</td>
<td>13 (7.7)</td>
</tr>
<tr>
<td>0</td>
<td>335 (88.4)</td>
<td>176 (88.4)</td>
<td>159 (88.3)</td>
<td>179 (85.2)</td>
<td>156 (92.3)</td>
</tr>
<tr>
<td>Whole Grains</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>58 (15.3)</td>
<td>34 (17.1)</td>
<td>24 (13.3)</td>
<td>33 (15.7)</td>
<td>25 (14.8)</td>
</tr>
<tr>
<td>0</td>
<td>321 (84.7)</td>
<td>165 (82.9)</td>
<td>156 (86.7)</td>
<td>177 (84.3)</td>
<td>144 (85.2)</td>
</tr>
<tr>
<td>Sodium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>17 (4.5)</td>
<td>13 (6.5)</td>
<td>4 (2.2)</td>
<td>13 (6.2)</td>
<td>4 (2.4)</td>
</tr>
<tr>
<td>0</td>
<td>362 (95.5)</td>
<td>186 (93.5)</td>
<td>176 (97.8)</td>
<td>197 (93.8)</td>
<td>165 (97.6)</td>
</tr>
<tr>
<td>Sweets/Sugar Sweetened Beverage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>379 (100.0)</td>
<td>199 (100.0)</td>
<td>180 (100.0)</td>
<td>210 (100.0)</td>
<td>169 (100.0)</td>
</tr>
<tr>
<td>0</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>
Appendix C: ACCN Food Access Questionnaire
We would like to know more about where, why and how you buy food in your community.

1. How often do you go food shopping at the following places?

<table>
<thead>
<tr>
<th>Place</th>
<th>Never</th>
<th>Occasionally</th>
<th>Sometimes</th>
<th>Often</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large Grocery Store or Supermarket</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Small Grocery Store or Market</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Convenience Store or Gas Station</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>General Store or Dollar Store</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Farmer’s Market or Fruit and Vegetable Stand</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Other, please tell us:</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

2. Of the places in Question 1 where you said you go food shopping, where do you shop most often?

a. Tell us about the place you shop for food most often. Let’s call this Place #1.

<table>
<thead>
<tr>
<th>Question</th>
<th>Don't know</th>
<th>I prefer not to answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the name of Place #1?</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>What is its address?</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>What city is it in?</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>About how far is it from your home?</td>
<td>[ ] miles</td>
<td>O</td>
</tr>
</tbody>
</table>

Participant ID: [ ] [ ] [ ] [ ] [ ] [ ]
b. What is the main reason you go food shopping at Place #1?

**PLEASE MARK ONLY ONE**
- Low prices.
- Good selection.
- Good quality.
- It's close to home.
- They treat me well there.
- They accept food stamps.
- They accept WIC vouchers.
- Other, please tell us: [ ]
- Don't know
- I prefer not to answer

c. When you go shopping, how do you usually get to Place #1?

**PLEASE MARK ONLY ONE**
- My car.
- I pay someone to drive me there.
- I ride for free in someone else's car.
- Public transportation.
- Taxi
- I ride my bike.
- I walk.
- Other, please tell us: [ ]
- Don't know
- I prefer not to answer

Participant ID: [ ] [ ] [ ] [ ]
d. Please tell us how much you agree or disagree with the following statements about food shopping at Place #1.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neither Agree or Disagree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is easy to buy fresh fruits and vegetables there.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>There are many types of fruits and vegetables there.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>The fruits and vegetables there are of high quality.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>It is easy to buy low-fat products (such as low-fat milk or lean meats) there.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>There are many types of low-fat products there.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>The low-fat products there are of high quality.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

e. How would you rate the cost of fresh fruits and vegetables at Place #1?

**PLEASE MARK ONLY ONE**

- Very expensive
- Somewhat expensive
- Not expensive
- Don’t know
- I prefer not to answer

f. Has the cost of fresh fruits and vegetables at Place #1 ever kept you from buying them?

**PLEASE MARK ONLY ONE**

- Yes
- No
- Don’t know
- I prefer not to answer
g. Now tell us about the second place you shop for food most often. Let’s call this Place #2.

<table>
<thead>
<tr>
<th></th>
<th>Don’t know</th>
<th>I prefer not to answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the name of Place #2?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What is its address?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What city is it in?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>About how far is it from your home?</td>
<td>3 3</td>
<td>miles</td>
</tr>
</tbody>
</table>

h. What is the main reason you go food shopping at Place #2?

**PLEASE MARK ONLY ONE**

- Low prices.
- Good selection.
- Good quality.
- It’s close to home.
- They treat me well there.
- They accept food stamps.
- They accept WIC vouchers.
- Other, please tell us: 
- Don’t know
- I prefer not to answer

i. When you go shopping, how do you usually get to Place #2?

**PLEASE MARK ONLY ONE**

- My car.
- I pay someone to drive me there.
- I ride for free in someone else’s car.
- Public transportation.
- Taxi
- I ride my bike.
- I walk.
- Other, please tell us: 
- Don’t know
- I prefer not to answer
6. What would help you eat healthier in your community?

PLEASE MARK ALL THAT APPLY

- Free or low cost rides to local places that sell food.
- New supermarket or grocery store nearby.
- Improved supermarket or grocery store nearby.
- Better variety and quality where you shop.
- More healthy choices at sit-down restaurants.
- More healthy choices at fast food places.
- Local farmer's market.
- Local community supported agriculture (CSA) program where people buy a membership in order to get produce directly from farmers.
- Food buying co-op where food is bought in bulk by a group of people for a discount.
- More places that accept food stamps.
- More places that accept WIC vouchers.
- Coupons and price discounts.
- Other, please tell us: [ ]
- Don't know [ ]
- I prefer not to answer [ ]
Bibliography


84. United States Preventive Task Force. Recommendations for Adults 2014 [cited 2014 June 20]. Available from:


270


173. Centers for Population Health and Health Disparities. NIH Announces the Funding of the Centers for Population Health and Health Disparities [cited 2014 June 24].
180. Roux AD. The examination of neighborhood effects on health: conceptual and methodological issues related to the presence of multiple levels of organization. Neighborhoods and health. 2003:45-64.


188. Centers for Disease Control and Prevention. Census Tract Level State Maps of the Modified Retail Food Environment Index (mRFEI).

189. U.S. Department of Agriculture ERS. USDA Food Desert Locator Documentation


220. Kwan LL, Bermudez OI, Tucker KL. Low vitamin B-12 intake and status are more prevalent in Hispanic older adults of Caribbean origin than in neighborhood-matched non-Hispanic whites. The Journal of nutrition. 2002;132(7):2059-64.


252. Fitzmaurice GM, Laird NM, Ware JH. Applied longitudinal analysis: John Wiley & Sons; 2012.


258. Baltic RD, Weier RC, Katz ML, Kennedy SK, Lengerich EJ, Lesko SM, et al. Study design, intervention, and baseline characteristics of a group randomized trial involving a faith-based healthy eating and physical activity intervention (Walk by Faith) to reduce overweight and obesity and cancer risk among Appalachian adults 2015


269. Swanson MA, Studts C, Crosby R, editors. Food shopping patterns in rural Appalachia: The environmental context of dietary behavior. 141st APHA Annual Meeting and Exposition (November 2-November 6, 2013); 2013: APHA.


273. Lane MB. Loss of grocery store hurts rural Vinton County. The Columbus Dispatch 2014

274. Buchanan T. 'Food desert' money could help Vinton County, officials say 2015