Heart Failure in the Southern Community Cohort Study: Incidence, Mortality and role of Anthropometric and Socioeconomic Factors

By

Elvis Abang Akwo

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Epidemiology

May, 2017

Nashville, Tennessee

Approved:

Loren Lipworth, ScD.

William J. Blot, PhD.

Edmond K. Kabagambe, DVM., PhD.

Frank E. Harrell Jr., PhD.

Deepak K. Gupta, M.D.

ACKNOWLEDGEMENTS

There are several persons who contributed significantly to the success of this project and I would like to seize this opportunity to recognize them.

First, I would like to thank the members of my dissertation committee: Drs Loren Lipworth; William Blot; Edmond Kabagambe; Frank Harrell and Deepak Gupta for their insightful contributions to this dissertation project.

Second, I would like to thank the leadership (Dr Katherine Hartmann and Dr Todd Edwards) and faculty of the Vanderbilt Epidemiology program; Vanderbilt Epidemiology Center; Institute of Medicine and Public Health and Department of Biostatistics for the coaching I received and for providing a propitious environment for my doctoral training. I would also like to recognize the contribution of the Vanderbilt Translational and Clinical Cardiovascular Research Center and the Division of Cardiovascular Medicine to my research work.

Third, I would like to acknowledge the invaluable contributions of the research team of the Southern Community Cohort Study (SCCS), the International Epidemiology Institute and the SCCS research participants, to the success of this project.

I owe enormous gratitude to my research mentors, Drs. Loren Lipworth and William Blot as well as my academic advisor Dr Edmond Kabagambe for their tutelage and support throughout the entire duration of my training. Dr Lipworth's guidance and leadership were invaluable in the elaboration and completion of this dissertation.

Finally, I am very grateful to Vanderbilt University for providing the opportunity, environment and resources required to complete my training.

TABLE CONTENTS

Page

LIST OF TABLES

PART ONE Page

PART TWO

LIST OF FIGURES

PART ONE Page

PART TWO

LIST OF ABBREVIATIONS

ACE	Angiotensin Converting Enzyme
ACCF	American College of Cardiology Association
ADA	American Diabetes Association
AHA	American Heart Association
AIC	Akaike Information Criteria
ARIC	Atherosclerosis Risk In Communities
BMI	Body Mass Index
BIA	Bioimpedance Absorptiometry
CABG	Coronary Artery Bypass Graft
CAD	Coronary Artery Disease
CAPI	Computer Assisted Interview
CCW	Chronic Conditions Data Warehouse
CHC	Community Health Centers
CHD	Coronary Heart Disease
CHS	Cardiovascular Health Study
CMS	Centers for Medicare and Medicaid Services
CVD	Cardiovascular Disease
CT	Computed Tomography
DEXA	Dual-Energy X-ray Absorptiometry
DME	Durable Medical Equipment
DRG	Diagnosis Related Group

EF	Ejection Fraction
ESRD	End-Stage Renal Disease
ESRI	Environmental Systems Research Institute
FFM	Fat Free Mass
FHS	Framingham Heart Study
GIS	Geographic Information System
HbA1c	Glycated hemoglobin
HBP	High Blood Pressure
HDL	High Density Lipoprotein
HF	Heart Failure
HF*p*EF	Heart Failure with Preserved Ejection Fraction
HF*r*EF	Heart Failure with Reduced Ejection Fraction
hsTNT	High Sensitivity Troponin T
IAAT	Intra-Abdominal Adipose Tissue
ICD – 9	International Classification of Diseases, Ninth Revision
IDF	International Diabetes Federation
LDL	Low Density Lipoprotein
LVH	Left Ventricular Hypertrophy
MAGGIC	Meta-Analysis Group in Global Chronic Heart Failure
MAX	Medicaid Analysis Extract
MBSF	Master Beneficiary Summary File
MEDPAR	Medical Provider Analysis and Review

MESA	Multi-Ethnic Study of Atherosclerosis
MET-hrs	Metabolic Equivalent Hours
MI	Myocardial Infarction
MRI	Magnetic Resonance Imaging
NCEP ATP	National Cholesterol Program Adult Treatment Panel
NDI	National Death Index
NHANES	National Health and Nutritional Examination Survey
NTproBNP	N-Terminal pro Brain Natriuretic Peptide
NYHA	New York Heart Association
PI	Ponderal Index
PC	Principal Components
RV	Right Ventricular
RAAS	Renin Angiotensin Aldosterone System
SAAT	Subcutaneous Abdominal Adipose Tissue
SCCS	Southern Community Cohort Study
SD	Standard Deviation
SES	Socioeconomic Status
SNF	Skilled Nursing Facility
SSA	Social Security Administration
TNF	Tumor Necrosis Factor
TVC	Time varying Covariate
WHO	World Health Organization

WC	Waist Circumference
W/H^n	Weight-Height Index
ZCTA	Zip Code Territorial Area

PART ONE

INTRODUCTION – LITERATURE REVIEW – METHODS

1. Introduction

1.1 Précis: Heart Failure in a low-income multiethnic population.

Heart failure (HF) is a major public health problem, particularly in the southeastern United States (US), which has been described as the "heart failure belt". However, most data informing the current understanding of the risk factors for, incidence of, and survival from HF were derived from cohorts outside of the southeast and primarily comprised of white individuals or multi-ethnic populations with high proportions of middle-class participants. Consequently, there are limited data regarding HF among individuals in the region of US with the highest prevalence of HF and those with limited resources; and potential differences in the patterns of HF incidence and post-HF survival by race and sex in low-income multiethnic populations are not well characterized.

In addition, while there is increased recognition that individual socioeconomic factors contribute significantly to HF risk among middle-class persons in the US, recent evidence also suggests that neighborhood factors may in fact predict HF readmissions independently of individual-level factors in middle-class populations. However it remains uncertain whether such neighborhood factors are independent predictors of HF incidence and post-HF survival particularly in low-income populations.

Finally, in contrast to international trends, the existing data from US counties suggest that poverty-dense counties have high levels of obesity which are paralleled by high prevalence of cardio-metabolic conditions including cardiovascular disease (CVD) and HF. Thus, it is important to critically examine the link between obesity and HF particularly among populations with scant resources. Prior epidemiologic evidence is suggestive of an independent association between excess body weight and increased risk of HF; as well as a contrasting decrease in the risk of post-HF mortality – a phenomenon coined as the obesity paradox. However, most studies investigating these relationships utilized categories of body mass index (BMI) or assumed linearity of effects thereby limiting the elucidation of the natural dose-response relationship between measures of obesity and HF risk as well as post-HF survival.

More importantly, less thought has been given to the suitability of BMI – weight (W)/height $(H)^2$ – in investigating the link between obesity and both HF risk and post-HF survival despite differences in the performance of various weight-height indices across population groups defined by race and sex. Additionally, predictors with "pleiotropic" effects usually have differential functional relationships with varying outcomes suggesting the need to use the data to empirically derive an appropriate weight-height index for each outcome. Such approaches may be utilized to adequately model the intricacies in these data and reveal novel insights that may improve our understanding of anthropometry and general obesity in relation to HF risk and post-HF survival.

We propose to leverage the Southern Community Cohort Study (SCCS) which comprises a large number of black and white participants, predominantly low-income, living in a region with the highest rates of CVD to investigate disparities in HF incidence and mortality by race and sex as well as differential effects of anthropometric and neighborhood socio-economic factors.

Our specific aims are to investigate:

1. Differences in the incidence of HF as well as post-HF survival between groups defined by race and sex: white women, black women, white men and black men.

2. Whether neighborhood characteristics (defined by a composite deprivation index) predict the risk of a) incident HF and b) post-HF survival in the SCCS beyond individual-level socioeconomic status (defined by household income and highest level of education attained).

3a. The appropriate functional form of a data-derived weight-height index (W/H^n) for the association with a) incident HF and b) post-HF survival, and compare its performance in the prediction of either outcome with that of BMI based on model fit and "informativeness".

3b. The dose-response relationship between W/H^n (as a surrogate measure of total body fat) and both incident HF and post-HF survival by race. We would specifically investigate departures from linearity and additivity of effects.

3c. The dose-response relationship between waist circumference (as a surrogate measure of visceral fat) and the risk of incident HF by race and sex and contrast these findings with those obtained using W/H^n.

Heart failure events will be ascertained via linkage of the SCCS cohort with the Centers for Medicare and Medicaid Services (CMS) Research Identifiable Files. For the proposed analyses, we would include SCCS participants ≥ 65years at cohort enrollment or participants < 65 years who: a) reported CMS coverage at baseline; or b) did not report Medicare or Medicaid on the baseline questionnaire but had a CMS claim within 90 days of being enrolled in SCCS. The restriction to these groups would maximize the likelihood of participants having continuous coverage in Medicare and/or Medicaid from the time of SCCS enrollment to the end of the follow-up period (December 31st, 2010), for the ascertainment of incident HF events.

Incident HF will be defined as the first occurrence of a medical claim with ICD-9 code 428.x within the Medicare institutional (Medicare Provider Analysis and Review, MEDPAR), Part B carrier, or outpatient-based claims files or the Medicaid Analytic Extract (MAX) Inpatient and Other Services claims files, from the date of SCCS enrollment through December 31st, 2010. All-cause mortality following a HF diagnosis will be ascertained via linkage of the SCCS cohort with both the Social Security Administration (SSA) vital status service for epidemiologic researchers and the National Death Index (NDI) through December 31st, 2010.

The relevance of the proposed study can be articulated across several axes:

a) It is essential to quantify the burden (incidence and mortality) of HF and any potential differential patterns of HF risk and post-HF survival – by race and sex – in a low-income population with high proportions of African-Americans and women living in a part of the country which is at particularly high risk for CVD. While the current understanding (based on data from cohorts enrolling mostly middle-class participants) suggests higher HF risk among African-Americans (and men), it would be interesting

to see if the patterns persist in a setting with more comparable socioeconomic status between racial groups.

b) In addition to estimating the scope of the HF epidemic in low-income populations in the "HF belt", it is equally important to provide valuable information on the relative contributions of individual and neighborhood socioeconomic factors that likely influence CVD outcomes – HF risk and mortality in particular – in this population. There are data suggesting greater rates of HF hospitalizations in neighborhoods with less resources. Is this trend similar for other HF outcomes like incidence and mortality? The AHA, the Canadian Heart Health plan and other cardiovascular societies recognize that improvements in heart health would require strategies that target the entire spectrum of the healthcare system: public policy, prevention, acute care, chronic care and rehabilitation, and end-of-life planning and care. However, the more "upstream measures" which focus on public policy and prevention may have the greatest potential to mitigate the burden of CVD and improve human health. Areas with the most acute socioeconomic deprivation are most likely at the highest risk for CVD (including HF) and CVD mortality and hence may benefit most from such improvements in public health policies including (but not limited to): improvements in community-level resources (healthy food outlets, physical activity resources, smoking cessation programs etc.).

c) One of the major contributors to the elevated burden of HF in areas with limited resources may be the concomitant increase in the levels of obesity in these settings. Thus, it is important to critically examine the link between adiposity and HF risk (and mortality). With most cohorts relying on surrogate measures of adiposity collected in routine clinical practice, robust approaches are needed to adequately model these data and reveal novel insights that would potentially refine the strategies used to risk-stratify persons in clinical and/or public health settings. Our proposed analysis accomplishes the following: It utilizes the data to empirically derive an appropriate weight-index to investigate the flexible dose response between HF risk (and post-HF survival) and a surrogate marker of total adiposity without making assumptions related to a) using BMI as the "de facto index of choice" regardless of the

demographic make-up of the population under study; b) using pre-specified cut-points that may be ill-suited to the study sample c) linearity and/or additivity of effects across sex and racial groups. Most cardiovascular societies (including the AHA and the ACCF) as well as the WHO make recommendations about specific cut-points for BMI as targets to be utilized in routine clinical practice, risk-stratification, preventive care and public health programs in order to mitigate CVD and HF risk. These recommendations regarding specific BMI cut-points implicitly assume a discontinuity in the dose-response association between BMI and HF (or CVD) risk or the existence of definite inflexion points in a continuum of risk. If the dose-response relationship between the weight-height index (as a surrogate measures of total adiposity) and HF risk (as well as post-HF mortality) is continuous and non-linear (and is modified by race and sex), that would suggest that the use of cut-points for decision making is seemingly counter-intuitive. An alternative decision-making paradigm would be to develop a prediction model for HF risk (and HF mortality) using the relevant anthropometric, lifestyle, clinical and demographic factors in this population; and using the estimates of predicted risk for individuals and cost-effectiveness ratios of efficacious interventions to make decisions in clinical or public heath settings. In such models, for example the HF risk calculator developed by the Meta-Analysis Global Group in Chronic Heart Failure (MAGGIC), it would be interesting to explore differences in model fit (as well as discriminant and calibration properties) when using the established anthropometric surrogates or composite weight-height indices derived for specific race-sex groups.

d) The results of the proposed analyses could provide preliminary data for additional studies of the association between data-derived weight-height indices – as proxies for total body fat – and measures of subclinical CVD like cardiac troponin T as well as N-Terminal pro Brain Natriuretic Peptide (NT-proBNP). The relationship between anthropometric measures and cardiac troponin T is particularly apropos as the latter is thought to be one of the potential links between increased adiposity and incident HF. Cross-sectional data from the ARIC cohort suggest that there is a positive association between BMI and high cardiac troponin T (using a cut-point of 14ng/l) measured with a new high-sensitivity assay (hs-

cTnT). In future studies it would be interesting to leverage the SCCS cohort and flexible modeling approaches accounting for potential nonlinearity of effects and interactions with race and/or sex to investigate the prospective association between proxy-indicators of total and visceral adiposity and hs-cTnT amongst persons free of CVD at baseline. An improved understanding of the association between weight-height indices (used in routine clinical practice and public health) and subclinical CVD could potentially improve risk stratification and prediction.

1.2 Specific Aims

There are over 26 million persons living with heart failure (HF) worldwide.[1] In the US, over 5.7 million adults (≈ 2.5% of the US adult population) are estimated to have HF.[2] About half of persons diagnosed with HF die within 5 years and the estimated total costs of HF in the US exceeded $30 billion in 2012.[3],[4]

Several established cardiovascular disease (CVD) cohorts have investigated HF incidence and mortality, including the Framingham Heart Study (FHS), Cardiovascular Health Study (CHS), Multi-ethnic Study of Atherosclerosis (MESA) and Atherosclerosis Risk in Communities (ARIC).[5-8] The FHS included predominantly white individuals. Other cohorts, including CHS, MESA, and ARIC, enrolled multi-ethnic middle-class populations from select communities and their relatively small sample sizes limited assessment of differential risk patterns between demographic groups defined by both race and sex.

Whilst the data from these previous cohorts are suggestive of differences in incidence rates of HF as well as post-HF mortality between population subgroups, knowledge gaps persist regarding the magnitude and direction of these differences in multi-ethnic low income populations (comprising larger numbers of blacks and women) with high burden of CVD risk factors.

Also, while there is evidence suggesting that individual socioeconomic status (SES) contributes to HF risk among middle-class persons in the US [9, 10], recent data also suggests that neighborhood factors may in fact predict HF readmissions independently of individual-level SES in middle-class populations [11]. However, it is not known whether such neighborhood factors are independent predictors of other HF outcomes like HF incidence and post-HF mortality particularly in low-income populations. It would be of interest to investigate whether among persons with very limited resources, a dearth of community-level resources i.e. neighborhood deprivation, compounds the risk of HF and post-HF mortality above and beyond what is contributed by scant resources and reduced literacy at the individual-level.

Data from varying communities across the US suggest that poverty-dense counties have high levels of obesity which are paralleled by high prevalence of chronic conditions including CVD and heart failure [12]. It is ever so important to critically examine the link between adiposity and HF risk (and survival following a diagnosis of HF) particularly among populations with scant resources. Most cohorts investigating the obesity-HF link have relied on body mass index – weight (W)/height $(H)^2$ – as a proxy for total body fat and increased body mass index has been found to be positively correlated with HF incidence and survival [13-15] but there is little data on potential effect modification by race. Less thought has been given to the suitability of W/H^2 in investigating the obesity-HF link despite differences in the performance of various weight-height indices across population groups defined by race and sex [16, 17]. Additionally, predictors with "pleiotropic" effects usually have differential functional relationships with varying outcomes suggesting the need to use the data to derive the appropriate weight-height index for each outcome. Hypothetically, for any weight-height index given by the general form W/H^n, regressing the log-hazard of HF on the natural logs of both weight and height, would yield n as the absolute value of the ratio of the coefficients of log height and log weight; this could be used to generate a W-H index suitable for a given setting. Also, any potential departures from additivity of effects by race and sex could be investigated using interaction terms with the appropriate weight-height index. Such robust approaches are needed to adequately model these data and reveal novel insights that would potentially refine the strategies used to risk-stratify persons in clinical and/or public health settings.

We therefore propose to use data from the Southern Community Cohort Study (SCCS) to investigate disparities in the risk of incident HF (and post-HF survival) between sex and racial groups and the contribution of neighborhood socioeconomic factors and anthropometric measures to HF risk and mortality. The SCCS is a large, prospective cohort study that enrolled approximately 86,000 adults (over two-thirds black) aged 40 and 79 living in the southeastern region of the US between 2002 and 2009 [18]. Data on personal medical history, demographic, socioeconomic, neighborhood factors, lifestyle, and anthropometric characteristics were ascertained at cohort enrollment.

Our specific aims are to investigate:

1. Differences in the incidence of HF as well as post-HF survival between groups defined by race and sex: white women, black women, white men and black men.

2. Whether neighborhood characteristics (defined by a composite deprivation index) predict the risk of a) incident HF and b) post-HF survival in the SCCS beyond individual-level socioeconomic status (defined by household income and highest level of education attained).

3a. The appropriate functional form of a data-derived weight-height index (W/H^n) for the association with a) incident HF and b) post-HF survival, and compare its performance in the prediction of either outcome with that of BMI based on model fit and "informativeness".

3b. The dose-response relationship between W/H^n (as a surrogate measure of total body fat) and both incident HF and post-HF survival by race. We would specifically investigate departures from linearity and additivity of effects.

3c. The dose-response relationship between waist circumference (as a surrogate measure of visceral fat) and the risk of incident HF by race and sex and contrast these findings with those obtained using W/H^n.

Heart failure events would be ascertained via linkage of the SCCS cohort with Centers for Medicare and Medicaid Services (CMS) Research Identifiable Files. SCCS participants (n = 27,078) who meet the following inclusion criteria would be included in our analyses:

- ≥ 65years (n = 7001) at cohort enrollment
- < 65 years (n = 20,077) at enrollment and either:
    i. reported being covered by Medicaid on the baseline questionnaire;
    ii. reported being covered by Medicare on the baseline questionnaire;
    iii. did not report Medicare or Medicaid on the baseline questionnaire but had a CMS claim within 90 days of being enrolled in SCCS.

The restriction to these groups maximizes the likelihood of participants having continuous coverage in Medicare and/or Medicaid from the time of SCCS enrollment to the end of the follow-up period (December 31st, 2010), for the ascertainment of incident HF events.

Incident HF will be defined as the first occurrence of a medical claim with ICD-9 code 428.x within the Medicare institutional (Medicare Provider Analysis and Review, MEDPAR), Part B carrier, or outpatient-based claims files or the Medicaid Analytic Extract (MAX) Inpatient and Other Services claims files, from the date of SCCS enrollment through December 31st, 2010. Death from any cause will be ascertained via linkage of the SCCS cohort with both the Social Security Administration (SSA) vital status service for epidemiologic researchers and the National Death Index (NDI) through December 31, 2010.

For aim 1, we will investigate differential patterns of HF incidence and post-HF survival by race and sex (and contrast the findings with those from previous CVD cohorts investigating HF risk and survival) using nonparametric methods and multivariable Cox Models. To obtain HF incidence rates, duration of follow-up would be computed from date of entry into the SCCS until the date of the first diagnosis of HF, date of death, or December 31, 2010, whichever occurred first. Incidence rates (IR) of heart failure would be calculated for white women, black women, white men and black men by dividing the number of HF cases by person-time of follow-up, and the rates would be presented per 1,000 person-years. For analyses of post-HF survival among those with a diagnosis of incident HF, follow-up time will be defined as time from HF diagnosis to death or December 31st 2010 whichever occurred first. Kaplan-Meier survival curves would be plotted by race and sex and the Wald tests in Cox models would be used to test for differences in survival between the four groups. In the multivariable models for both outcomes (HF incidence and post-HF survival), we would include indicator variables for white men, black women and black men, with white women as the reference group. The covariates would include: BMI, history of diabetes, hypertension, high cholesterol, MI/CABG or stroke, household income, education, smoking, alcohol intake, marital status and enrollment source (community health centers vs general population).

For aim 2, the data will be considered to have a hierarchical structure with study participants (level-1units) nested within census tracks (level-2 units). The social and economic characteristics of the latter will used to compute neighborhood deprivation index for all SCCS participants. We will test for the effects of neighborhood deprivation index (the level-2 predictor) on the risk of incident HF after adjustment for individual-level factors including SES (annual household income and education); demographics (age, sex and race); lifestyle (smoking and alcohol use) and clinical factors (BMI and history of diabetes, hypertension, high cholesterol, MI/CABG or stroke). Given the correlation of the data points within each census track, and the limitations of a multilevel modelling approach in this setting (related to unbalanced data between clusters with the potential of biasing group-level variances and fixed effects) our primary approach for aim 2 will be a Cox proportional hazards model that takes into account non-independence using the Huber-White cluster Sandwich estimator of variance.

For aim 3, we would first derive the appropriate weight-height index for the current cohort based on the coefficient of the log weight and log height variables in a bivariate Cox model for the log hazard of HF. Second, in separate Cox models, we would regress the restricted cubic splines of the natural log of the data-derived weight-height index and that of BMI on the log hazard of HF. Then, model fit statistics (LR chi square, χ^2 and AIC) would be used to compare the performance of the data-derived weight-index versus that of BMI in relation to a model utilizing restricted cubic splines of log weight and log height. Third, we would use multivariable Cox models which take into account nonlinearity and non-additivity to model a flexible dose-response association between the better performing weight-height index (W/H^n) (modelled using restricted cubic splines with 5 evenly spaced knots) and HF risk adjusting for relevant covariates namely: demographics (age, race, sex); lifestyle factors (smoking and alcohol use); socioeconomic status (household income and education), clinical history (history of diabetes, hypertension, high cholesterol, MI/CABG or stroke) and total physical activity. Interactions between W/H^n and race as well as sex would be tested. These analyses would be repeated for the association

between (W/H^n) and post-HF survival. Similar multivariable models will be utilized to investigate the relationship between waist circumference and HF.

The SCCS cohort comprises a large number of low-income participants (over 52% with annual household income < $15,000), a high proportion of blacks (two-thirds) and women (60%), and overall represents a segment of the US population that has a high burden of CVD risk factors. In addition, data from previous SCCS studies suggests that there is a high prevalence of obesity (defined as body mass index >30 kg/m^2); the prevalence of hypertension is > 50% overall [19] and the prevalence of diabetes is over 21% [20] (compared to a national average of 11% [21]). These characteristics of the SCCS participants make it a unique cohort for the examination of the inter-relationship between anthropometric, demographic and socioeconomic (individual and neighborhood) factors in the etiology of HF and post-HF mortality in a region of the country with the highest rates of CVD.

The results of the proposed study will a) quantify the burden of HF (incidence and mortality) in a low-income population that has been under-represented in previous cardiovascular cohorts b) improve the understanding of the dose-response relationship between weight-height indices (as surrogates of total adiposity) and HF risk and the potential differential influences of race and sex in these relationships c) provide information on the independent contrasting and/or synergistic relationships between individual-level and neighborhood-level effects which may inform individual and community-level interventions aimed at reducing the burden of the HF epidemic.

2. Literature Review and Rationale for Specific Aims

2.1 Background for Specific Aim 1

2.1.1 Heart Failure – definition, etiology and classification.

Heart Failure (HF) is a complex clinical syndrome resulting from the inability of the heart to provide sufficient blood flow at normal filling pressures to meet the metabolic demands of the body [22]. The cardinal manifestations of HF are dyspnea and fatigue and fluid retention, which may lead to pulmonary and/or splanchnic congestion and/or peripheral edema [22].

Heart failure usually results from a variety of conditions affecting the myocardium, heart valves, pericardium, or electrical conduction system, and may be due to hypertension, coronary artery disease, toxins, inflammatory/infectious diseases, genetic or metabolic disorders. Most patients with HF have some form of cardiomyopathy and impaired left ventricular (LV) function; and the latter is usually responsible for the symptoms they present. Heart Failure is often associated with a variety of LV functional abnormalities, ranging from normal LV size and preserved ejection fraction (EF) to severe dilatation and/or markedly reduced EF [22]. However, regardless of EF, varying degrees of diastolic and systolic dysfunction co-occur in most persons with HF. Ejection fraction is important for classifying patients – as having heart failure with reduced ejection fraction or HF with preserved ejection fraction – because of differences in patient prognosis, therapeutic response, comorbidities and demographics [23].

Different cut-points for EF have been proposed for defining Heart Failure with reduced ejection fraction (HF*r*EF) including ≤35% and ≤40% [24, 25]; the ACCF/AHA guidelines recommends using ≤40%. Coronary artery disease (CAD) with antecedent myocardial infarction remains the main etiology for HF*r*EF which occurs in conjunction with varying degrees of LV enlargement [26, 27].

There are several criteria that have been proposed to define HF with preserved EF (HF*p*EF). These include: i) clinical signs or symptoms of HF; ii) evidence of preserved or normal LVEF; and iii) evidence of abnormal LV diastolic dysfunction that can be determined by Doppler echocardiography or

cardiac catheterization [28]. Just as for HF*r*EF, different investigators have proposed variable EF cut-points for defining HF*p*EF ranging from >40% to ≥55% [29]. There appears to be a trend towards increasing proportions of persons with of HFpEF [30] which is paralleled by the increasing prevalence of some of its major causal factors including diabetes and obesity. Other factors incriminated in the occurrence of HFpEF include dyslipidemia, CAD and atrial fibrillation but hypertension remains the principal etiology of HFpEF at the population level [31, 32].

The ACCF/AHA and the NYHA classifications have been used to stratify persons with heart failure with respect to HF stage and functional classification respectively [33, 34].

Table 1: **Comparison of ACCF/AHA Stages of HF and NYHA Functional Classifications**

ACCF/AHA Stages of HF [34]		NYHA Functional Classification [33]	
A	At high risk for HF but without structural heart disease or symptoms of HF	None	
B	Structural heart disease but without signs or symptoms of HF	I	No limitation of physical activity. Ordinary physical activity does not cause symptoms of HF.
C	Structural heart disease with prior or current symptoms of HF	I	No limitation of physical activity. Ordinary physical activity does not cause symptoms of HF.
		II	Slight limitation of physical activity.
		III	Marked limitation of physical activity. Comfortable at rest, but less than ordinary activity causes symptoms of HF.
		IV	Unable to carry on any physical activity without symptoms of HF, or symptoms of HF at rest.
D	Refractory HF requiring specialized interventions	IV	Unable to carry on any physical activity without symptoms of HF, or symptoms of HF at rest.

ACCF: American College of Cardiology Foundation; AHA: American Heart Association.
NYHA: New York Heart Association.

2.1.2 The Burden and risk factors of Heart Failure in the United States.

2.1.2.1 Prevalence, costs and risk factors

It is difficult to overestimate the impact of heart failure (HF) in the US; experts have long recognized it as an important cause of morbidity and mortality and its prevalence continues to rise [35-37]. With the aging of the US population and increasing prevalence of major risk factors for HF (especially diabetes), it has become a serious health concern particularly for elderly Americans [38-41]. Currently, over 5.5 million Americans (2.5%) are estimated to have heart failure [42]. In 2007, the estimated direct and indirect costs of heart failure in the U.S. were in excess of $33.2 billion [42].

Previous studies have suggested that the main risk factors for heart failure include older age, high blood pressure, diabetes, coronary artery disease, valvular heart disease, and atrial fibrillation [6, 43-49]. In addition, other investigators found evidence suggesting that race, body mass index and socioeconomic status also contribute to disparities in heart failure risk in US adults. Using data from the Multi-Ethnic Study of Atherosclerosis, Bahrami et al found that African-Americans had the highest incidence rate of HF, followed by Hispanic, white, and Chinese-American participants. Although the risk of developing HF was higher among black compared with white participants (HR, 1.8; 95% CI, 1.1-3.1) in the more parsimonious models, adding hypertension or diabetes mellitus to the models attenuated racial/ethnic differences in the risk of incident HF[6].

2.1.2.2 Heart Failure Incidence and Mortality

Several established cardiovascular disease (CVD) cohorts have investigated HF incidence and mortality – the Framingham Heart study, Cardiovascular Health Study (CHS), Multi-Ethnic Study of Atherosclerosis (MESA) and Atherosclerosis Risk In Communities (ARIC) among others (Tables 2, 3 and 4).

The earlier cohorts – particularly Framingham – included mostly white participants. Other cohorts like CHS enrolled elderly participants and a small proportion of African-Americans. Subsequent

cohorts like MESA and ARIC enrolled multi-ethnic middle-class populations from restricted communities and had relatively small sample sizes to adequately explore differential risk patterns between demographic groups defined by both race and sex.

Table 2: Incidence rates for HF in other CVD cohorts.				
ARIC [8]		**CHS [5]**		
Age Group	**Incidence rate per 1000 PY***	**Age Group**	**5yr Incidence rate per 1000 PY**	**10yr Incidence rate per 1000 PY**
Overall	5.7	Overall	-	-
White Men	6.0	White Men	25.3	29.7
45-49	2.4	65-69	14.8	18.7
50-54	5.6	70-74	20.1	23.9
55-59	8.4	75-79	33.3	41.4
60-64	14.3			
White Women	3.4	White Women	13.6	17.8
45-49	1.7	65-69	7.1	10.7
50-54	3.1	70-74	11.8	15.7
55-59	4.4	75-79	19.3	26.6
60-64	7.7			
Black Men	9.1	Black Men	22.1	25.5
45-49	5.2	65-69	13.6	18.4
50-54	7.2	70-74	17.3	20.8
55-59	14.0	75-79	28.2	34.6
60-64	13.4			
Black Women	8.1	Black Women	19.5	22.1
45-49	3.8	65-69	13.4	13.4
50-54	7.6	70-74	20.1	22.7
55-59	10.1	75-79	20.1	30.0
60-64	17.4			

MESA: Overall incidence = 3.1/1000 person-years [50]. **Framingham:** Overall in persons ≥45 years, age-standardized incidence rate in men = 7.2/1000 P-Y; in women = 4.2/1000 P-Y. Trend of increasing IR with age from 3/1000 in 50-59 to 27/1000 P-Y amongst 80-89yr males and from 2/1000 in 50-59 to 22/1000 amongst 80-89yr females [7].

*PY: person-years.

Table 3: Post-HF mortality in CHS and ARIC.

Mortality rates among persons with HF in the Cardiovascular Health Study [51].

	Person-Years At risk	**Number of Deaths**	**All-cause Mortality rate per 100 P-Y (95% CI)**
Overall	2690	1020	37.9 (35.8-40.0)
Women	1399	492	35.2 (32.5-37.9)
Men	1291	528	40.9 (37.7-44.1)
Caucasians	2340	889	38.0 (35.7-40.2)
African-Americans	350	131	37.5 (31.9-43.1)
Caucasian Women	1173	416	35.5 (32.5-38.5)
African-American Women	226	76	33.6 (27.2-40.0)
African-American Men	124	55	44.4 (33.7-55.1)
Caucasian Men	1168	473	40.5 (37.1-43.9)

Case Fatality rates among hospitalized HF patients in the ARIC study [8].

	Persons at risk	**Age-adjusted 1yr Case fatality (95% CI)**	**Age-adjusted 5yr Case fatality (95% CI)**
Overall	1198	22.0	42.3
Caucasian Men	495	19.6 (16.2-23.5)	41.2 (36.9-45.6)
African-American Men	164	23.9 (17.8-31.2)	51.8 (44.1-59.4)
Caucasian Women	301	20.8 (16.6-25.8)	35.8 (30.6-41.4)
African-American Women	238	23.5 (18.5-29.4)	46.1 (39.8-52.5)

While these studies suggest differences in incidence rates of HF as well as post-HF all-cause mortality between population subgroups, the studies were limited by small numbers and a few knowledge gaps persist regarding the magnitude and direction of these differences in low income populations (with larger numbers of African-Americans and women) (and better adjustment for SES by design) with high burden of CVD risk factors.

More importantly it remains uncertain whether these racial and sex differences persist across the age spectrum and whether there are differential risk factor associations with HF across categories defined by these demographic variables. The Southern Community Cohort Study (SCCS) could be a valuable tool to investigate these hypotheses.

We therefore propose to use data from the SCCS to investigate differences in the cumulative incidence and incidence rates of heart failure as well as differences in post-HF survival/mortality between groups defined by sex and race in univariate and multivariable models. The SCCS is a large, prospective cohort study that enrolled approximately 86,000 adults (over two-thirds black) aged 40 to 79 living in the southeastern US between 2002 and 2009 [18]. Data on personal medical history, demographic, socioeconomic, neighborhood factors, lifestyle, and anthropometric characteristics were ascertained at cohort enrolment.

Our findings could provide valuable information for risk stratification and prediction in populations with an enormous CVD risk burden. This data could also provide the preliminary data for investigating metabolic correlates of HF and particularly biomarkers (including hs-cTnT) for subclinical CVD.

2.2 Background and rationale for Specific aim 2

2.2.1 The relationship between demographic, socioeconomic and neighborhood factors in the prediction of HF risk in low-income communities.

Current evidence suggests that the highest rates of heart failure have been reported for inhabitants of the southeastern United States. This may in part be related to a higher prevalence of established risk factors (including CVD, obesity, diabetes and high blood pressure) in the southeast [52] which, in turn, could be attributed in part to socioeconomic characteristics (including education and income) that influence health outcomes. Data from ARIC and CHS suggests that individual socioeconomic status (SES) contributes to HF risk among middle-class persons in the US [9, 10]; recent data also suggests that neighborhood factors may in fact predict HF readmissions independently of individual-level SES in middle-class populations [11]. However, it is not known whether such neighborhood factors are independent predictors of other HF outcomes like HF incidence and post-HF mortality particularly in low-income populations. It would be of interest to investigate whether among persons with very limited resources, a dearth of community-level resources i.e. neighborhood deprivation, compounds the risk of HF and post-HF mortality above and beyond what is contributed by scant resources and reduced literacy at the individual-level. Additionally the potential "moderating" effects of race in the form of cross-level interactions could be relevant as well.

The potential relationships between demographic, anthropometric, individual and neighborhood socioeconomic factors are depicted in figure 2.

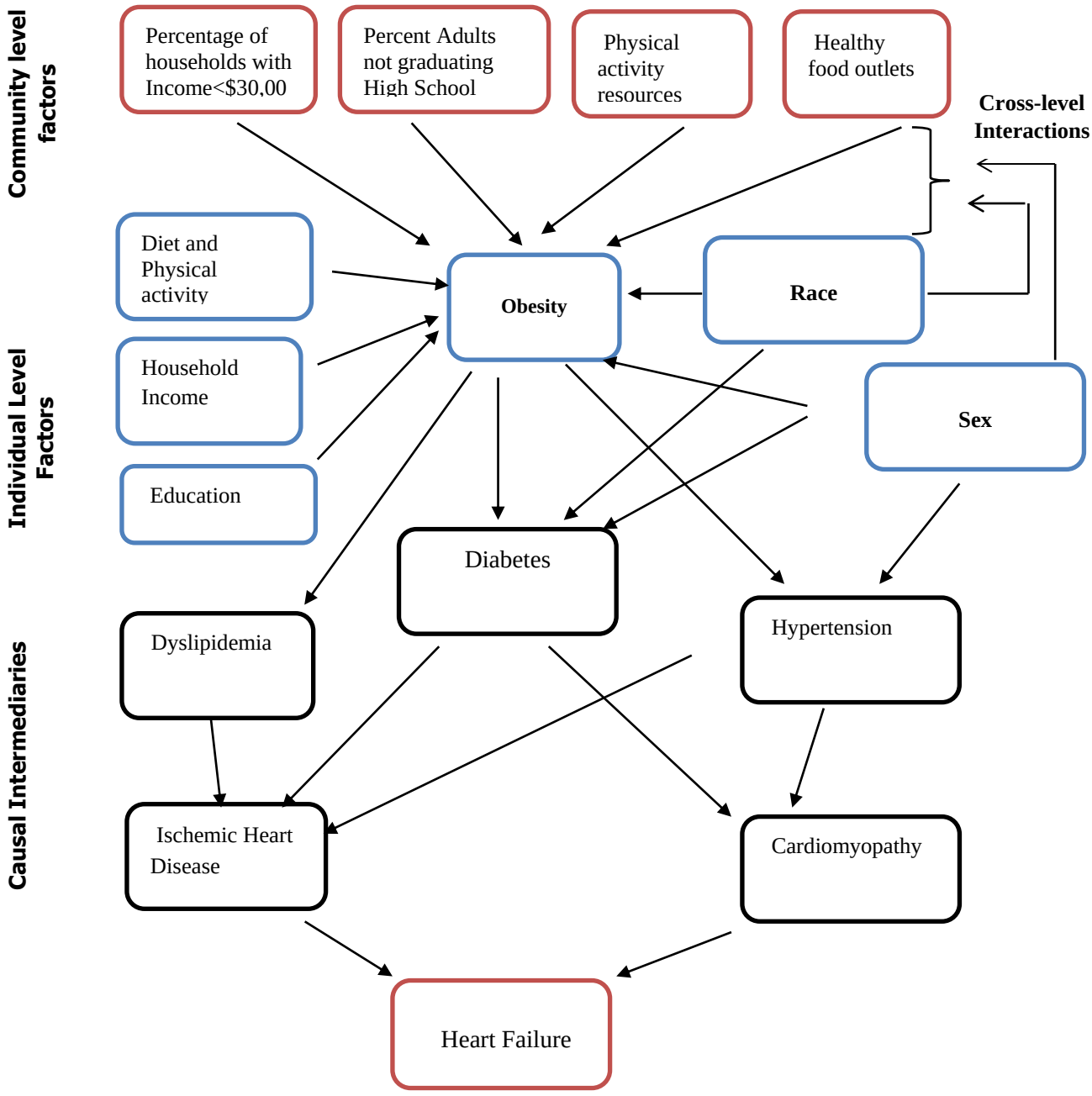


Figure 1: **Causal Model showing the relationship between community-level factors, individual-level predictors, causal intermediaries and heart failure**. We hypothesize that neighborhoods with high proportions of persons without a high school diploma are likely to be those with the lowest density of physical activity resources and healthy food outlets. This, in concert with individual level factors (gender, race, income, employment, education etc) may influence an individual's obesity status. Increased adiposity is expected to increase the likelihood of Ischemic and non-Ischemic heart failure by increasing the risk of causal intermediaries namely diabetes, hypertension and dyslipidemia. In our study, the community-level factors are summarized in the neighborhood deprivation index. We hypothesize that the neighborhood effects may vary for persons of different race and gender, just as the nefarious effects of increased BMI could be mitigated by favorable neighborhood characteristics - "cross-level interactions"

We therefore propose to use data from the Southern Community Cohort Study (SCCS) to investigate whether neighborhood characteristics (defined by a composite deprivation index) predict the risk of incident heart failure beyond individual-level socioeconomic status (defined by household income and highest level of education attained).

For these analyses, study participants (level-1units) are nested within census tracts (level-2 units) and the socio-economic characteristics of the latter were used to compute neighborhood deprivation. We will test for the effects of neighborhood deprivation index (the level-2 predictor) on the risk of incident HF after adjustment for individual-level factors including SES (annual household income and education); demographics (age, sex and race); lifestyle (smoking and alcohol use) and clinical factors (BMI and history of diabetes, hypertension, high cholesterol, MI/CABG or stroke).

We hypothesize that there exists a positive association between neighborhood deprivation and heart failure risk even after adjusting for individual-level socioeconomic factors like income and education.

2.3 Background for specific aim 3

2.3.1 Measuring Obesity/Adiposity

2.3.1.1 Direct and indirect measures of adiposity in biomedical research and clinical practice.

In routine clinical practice and public health, body mass index is often used as a surrogate measure of general obesity or "overweightness" even though it is not ideal to distinguish lean mass vs. fat mass [53]. Body mass index (BMI) is defined simply as weight (Kg)/height^2 (m^2) and is customarily used to classify persons as overweight or obese based on the WHO classification - normal weight: 18.9-24.9 Kg/m^2; overweight: 25.0 to 29.9 Kg/m^2; class I obesity: 30.0 to 34.9 Kg/m^2; class II: 35.0 to 39.9 Kg/m^2 and class III: ≥40.0 Kg/m^2 [54].

Waist circumference (and waist-hip ratio) is also frequently utilized in clinical practice particularly in NCEP ATP II guidelines for defining cut-points for the "metabolic syndrome" (>102 cm/40 inches for men and >90 cm/35 inches for women), in the American Diabetes Association (ADA) standards of care for diabetes and International Diabetes Federation (IDF) clinical practice guidelines for diabetes management [55, 56] [57] [58] [59] .

Table 4 shows the correlation between BMI, WC and several adipose tissue measurements. The data suggests that overall, both BMI and WC have strong positive correlations with total body fat mass but there are some differential patterns by sex [60]. Body mass index appears to be a better surrogate measure for total fat mass and percent body fat amongst women while the reverse is true among men. However, WC is a better predictor of intra-abdominal adipose tissue (IAAT) i.e. visceral fat [61] in both men and women and some studies have found WC to be more predictive of type 2 diabetes and some cardio-renal events including chronic kidney disease and CHD [62-64]. Additionally, WC is better correlated with measures of abdominal adipose tissue (both subcutaneous abdominal adipose tissue, SAAT and IAAT) from gold standard techniques like Computed Tomography (CT) scans and magnetic

resonance imaging (MRI) [60]. That notwithstanding, both BMI and WC have been shown to be significantly associated with cardio-metabolic risk factors (including hypertension, hyperglycemia and dyslipidemia) and hard endpoints including CHD and cardiovascular death [65-69].

Table 4: Correlation, ρ between BMI, WC and adipose tissue compartments in men and women [70].

	Men		**Women**	
	BMI	WC	BMI	WC
Total adipose tissue	0.82	0.87	0.91	0.87
Percent body fat	0.70	0.79	0.86	0.82
Total subcutaneous adipose tissue	0.82	0.83	0.91	0.86
Total intra-abdominal adipose tissue (IAAT)	0.59	0.79	0.69	0.77

Most adipose tissue (~85%) is in fact situated subcutaneously and distributed throughout the body in discrete homogenous pockets and adjacent to body tissues [71]. The contribution of intra-abdominal or visceral fat to total adipose tissue varies with demographic factors – sex, race/ethnicity, age – physical activity and total fat mass [60]. Several methods are currently available for measuring total body fat and assessing body composition either directly or using surrogate measurements. Importantly, the assessment of body composition relies on assumptions regarding the density of body tissues, concentration of water and electrolytes; and /or relationships between body components, body tissues and the distribution of the estimates of these measures among healthy subjects [72].

In addition to BMI and WC (and waist-hip ratio), skinfold thickness is another anthropometric surrogate used as a proxy for assessing adiposity. Measurements of Skin fold thickness (which represent the thickness of the subcutaneous tissue in that area of the body) can be obtained from a variety of body sites including the subscapular and triceps area which are the skinfolds for which we have the most national reference tables available [73]. Their utility is rather limited in overweight and obese individuals and there are significant variations by sex [73].

Bio-impedance absorptiometry (BIA) is another indirect method used to provide estimates of fat mass (FM) by measuring the body's resistance (using sensors/electrodes) to a small amount of alternating current. Bioelectric impedance analyzers produce a measure of impedance that is proportional to total body water (TBW) and is used as predictor variable in a regression equation to provide estimates of FM. The equations used by the analyzers are derived using data from a given "reference population" and these measures are only valid for subjects with the same body type and shape as the subjects from the reference in question [73]. More importantly, the data for the use of BIA in overweight and obese persons is scant and its accuracy in these populations is rather limited [74].

Direct methods like the measurement of total body water (fat free mass is estimated from total body water and FM can then be obtained), total body counting and neutron activation produce somewhat more accurate results [73].

There are also criterion methods that include a) measurements of body density b) dual x-ray absorptiometry (DEXA) and c) CT and MRI. Hydro-densitometry and air displacement plethysmography (ADP) are the 2 main body density measurement techniques currently in use. For the former, body density is measured via underwater weighing and multi-compartment models are used to combine body density with measures of bone density and total body water to calculate body fatness [75, 76]. ADP works follows similar underlying principles without the drawback of subjects having to hold their breath during the measurement to ensure compliance and accuracy [77]. DEXA is the most widely used technique for assessing lean mass, FM and bone density. The two low-energy levels used in DEXA and their differential attenuation through the body allow the discrimination between tissues types – soft tissues (FFM and FM) and bone tissue. Mathematical algorithms allow calculation of the separation components using various physical and biological models. DEXA measurement make assumptions (which vary by manufacturer) related to hydration, potassium content or tissue density. However, DEXA is considered to be 1 of the most reliable and valid techniques for estimating FFM, FM and bone density in much of the population and is currently included in the ongoing National Health and Nutrition Examination Survey

(NHANES) [73, 78]. Computed Tomography and MRI are gold standard techniques for body composition measurements. The former is rarely considered for "whole-body" assessments due to the high level of radiation exposure and thus is mainly used to measure abdominal fat. CT can also provide measurements of intrahepatic and intramyocellular fat which are known to be correlated with the risk of type 2 diabetes [79, 80]. Magnetic resonance imaging is often not able to accommodate large body sizes but can provide accurate whole-body measurements for normal or "moderately overweight" persons [73].

2.3.1.2 Weight-height indices as surrogate measures of total adiposity

Given their relative ease of collection and potential biologic relevance to the occurrence of chronic diseases, anthropometric data like weight (W) and height (H) are often utilized in epidemiologic studies to compute 'weight-height indices' which serve for the most part as proxy measures of obesity or body "fatness".

Almost 2 centuries ago, Quetelet noted that compared to W/H and W/H^3, W/H^2 – later renamed as body mass index by Keys et al [81] – was more stable with increasing height in young adults but he never actively advocated for the latter to be used as a 'measure of adiposity'. However, in the biomedical/epidemiologic research community, the need for a surrogate measure of body fatness that could be readily obtained from anthropometric surrogates – like W and H – routinely measured in clinical settings led to the investigation of several weight-height indices and measures of relative weight as potential proxy indicators of obesity.

Two main criteria have been used to assess these indices. First, such an index would have to be relatively uncorrelated with height [16, 17]. The goal here was to obtain an index that was summarily "a measure of weight-corrected-for height" such that head-to-head comparisons of persons of different heights could be performed using this index regardless of the (modest) correlation between weight and height. Differences in such an index could therefore be attributed primarily to differences in fat (or fat-free) mass. Second, the index had to be highly correlated with a robust measure of body fat [16, 17] like

body density (measured via underwater weighing) or other measures of body fat. However, several studies have found that the performance of these indices on these two criteria varies across population groups defined according race, sex and age.

One of such studies was conducted by Keys et al using data from 7424 healthy men, aged 18-60 in 12 cohorts from 5 countries (US, Finland, Italy, South Africa and Japan) to compare 4 indices – the ponderal index (PI), W/H, W/H^2 and relative weight [17]. The relative weight of the subjects was expressed as a percentage of the average weight (from life insurance industry tables) of a sample of persons of the same height, age and sex in the population to which they belonged. The PI was defined as: $H/\sqrt[3]{W}$. Among white men, W/H^2 had the weakest correlation with height: in particular, among the 249 male executives from Minnesota, aged 49-59, the Pearson correlation coefficient, r of W/H^2 with height was 0.015 while the values for relative weight, W/H and PI were 0.102, 0.181 and 0.304 respectively[17]. Thus, the PI had the strongest correlation with height and hence was the least suitable index based on the first criteria. In contrast, among Bantu men (n = 116, age range: 31-60years) in South Africa, the values of the correlation with height for these indices (W/H^2, W/H and PI) were: 0.249, 0.509 and 0.102 respectively[17]; thus PI had the lowest correlation with height. This suggests that the finding of which index was least correlated with height varied by race.

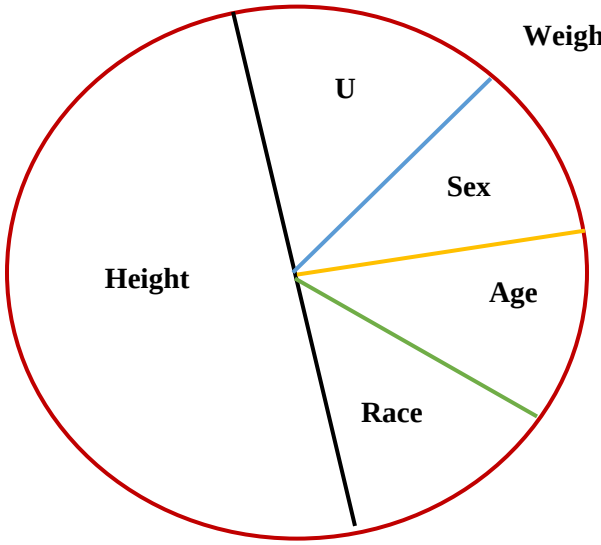
In addition, in the study by Florey et al – which was based on data from the fourth examination of the Framingham Heart Study – there were sex differences in the correlation of the aforementioned indices (W/H^2, W/H and PI) with height[16]. The Pearson correlation coefficient for W/H^2, W/H and PI among women (n = 2519) were -0.20, 0.03 and 0.41 respectively. In men (n = 2003), the corresponding values were: 0.08, 0.22 and 0.36[16]. Hence based on the first criteria (relative independence with height) W/H would be the best choice among women while among men it would be W/H^2.

Several studies have reported on the correlation between these weight-height indices and measures of body fat (skin fold thickness or body density). Overall, most investigators suggested that

W/H^2 (BMI) was: a) more strongly correlated with measures of body fat compared to PI[16, 17] and b) more or at least as correlated as W/H and W/H^3 with body fat measures[16, 17, 82]; with substantive variations related to the demographic characteristics (age, sex and race) of the population under investigation and the attending differences in the anthropometry of these population subgroups. In the study by Keys et al, among 180 young adults (aged 18-24) from the University of Minnesota, the correlation between body density (measured via underwater weighing) and W/H^2, W/H and PI were: -0.850, -0.833 and -0.791 respectively[17]. These values were -0.666, 0.658 and 0.657 among the middle-aged Minnesota executives (n = 249, age range: 49-59years) suggesting differential patterns by age[17]. In other words, while W/H^2 performed better than the other indices among young white males, there were no differences in the correlation of the said indices with body fat among middle-aged adults. Among the middle-aged black men in South Africa, the values of the correlations with body density were not available but the correlation with skin fold thickness for W/H^2, W/H and PI were: 0.732, 0.756 and 0.629 suggesting W/H performed slightly better; a trend different from that observed among middle-aged white men[17]. Also, Florey et al suggested sex differences in the correlation between the indices and infra-scapular skin fold thickness among white women in the Framingham Heart study but the observed differences were minimal – the values of r for W/H^2, W/H and PI were: 0.65, 0.66 and -0.64 respectively[16].

In summary, it would be safe to say that on both criteria – independence from height and strong correlation with body fat measures – there is significant variation in the performance of the indices depending on the demographic characteristics of the population under study. Using simulations, Florey showed that the correlation of the indices with height depended particularly on the slope (and the intercept) of the height variable in a regression equation where weight was regressed on height [16]. Incidentally, the slope parameter (i.e. the mean change/increase in weight per unit increase in height) appears to be vary based on the demographic characteristics of the study population. For instance, in the National Health Survey population, investigators found a slope of 4.0 among white men; among white

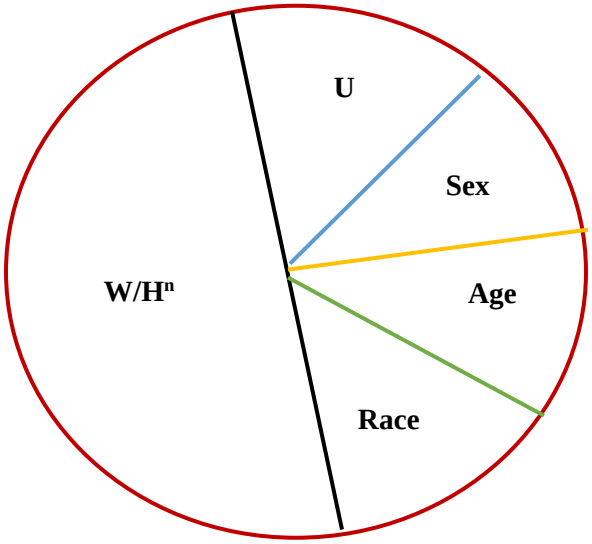
women the slope was 2.3 [83]. Similar values among men and women have been reported in other studies[16]. This difference in the association between weight and height across sex groups could be viewed across the lens of the fundamental axioms of Rothman's sufficient-component cause model. According to this model, component causes in a sufficient cause interact causally in varying patterns to influence outcomes [84]. As such, the association between a component cause and an outcome is influenced by its causal complement in the sufficient cause. Height may be independently correlated with weight but the slope of the height variable when weight is regressed on height may vary based on age, sex, race or ancestry and other unknown/unmeasured (U) covariates as shown in the causal pie below. And working on the premise of the simulations of Florey et al, this would influence the correlation between height and a given weight-height index; which may explain the differences seen across studies.



With such potential variations, it would seem intuitive to use a data-driven approach to determine the appropriate weight-height index to be utilized in any given study. In theory, for any weight-height index given by the general form W/H^n, a regression of body fat on the natural logs of both weight and height, would yield an absolute value of n (i.e. the exponent of the height variable in the 'composite' weight-height index) for the ratio of the coefficient of log height to that of log weight. Ergo, if the appropriate composite index of weight and height was one of W/H, W/H^2 or W/H^3, then by regressing body fat on the natural logs of both weight and height, the absolute value of the regression coefficient for

the log height variable would be 1, 2 and 3 respectively (i.e. the exponent of the height variable in the 'composite' weight-height index)

Working on this premise, one could obtain the appropriate exponent of the height variable for the weight-height index by regressing body fat on log weight and log height in a given dataset obtained from the population under study. The same could be done for any potential outcome e.g. CVD (including HF) that is thought to associated with the said weight-height-index (used here as a surrogate measure of body fat). Hence, the exponent of the height variable in the composite index would likely vary depending on the outcome under study. Theoretically, one could view this heterogeneity of the functional form of the composite weight-height index via prisms afforded by an "adaptation" or "extension" of the Rothman's model. One could postulate that the functional form of the weight-height index (i.e. the value of n) could vary based on a) the characteristics of the population under study and b) the outcome being investigated.



Separate indices may have to be constructed to study the association of interest within subgroups defined by race and/or sex. Otherwise, using a common composite index, interaction terms could be included in multivariable models to investigate differential associations of the weight-height index with CVD including HF.

2.3.2 The Burden of Obesity in the United States.

Every year in the US, overweight and obesity contribute to substantial morbidity and mortality[85, 86] and are responsible for billions of dollars in medical costs and lost productivity [87] . Their prevalence has increased to the extent that a majority of adult Americans are now considered overweight or obese (body mass index ≥ 25 kg/m^2) [88]. Obesity is associated with a myriad of adverse health outcomes including metabolic abnormalities such as dyslipidemia, type 2 diabetes, hypertension, as well as chronic kidney disease and cardiovascular disease [89].

2.3.3 Obesity and the Heart – Metabolic, Hemodynamic, Structural and Functional changes.

Individuals with elevated body mass index usually have both increased adiposity and fat-free mass and these are associated with metabolic, cellular and hemodynamic changes which are in turn associated with alterations in myocardial structure and performance [14].

2.3.3.1 Hemodynamic changes

In persons with excess body weight – both adipose tissue and fat-free mass – there is a proportionate increase in cardiac output which results primarily from an increased left ventricular (LV) stroke volume (figure 1) [90]. In addition, compared to persons with normal weight, persons with class II or III obesity also have higher right ventricular (RV) end-diastolic pressure, mean pulmonary artery pressure, pulmonary vascular resistance, and mean arterial pressure [90].

2.3.3.2 Structural Changes

Increases in adiposity and lean body mass are also associated with obesity-related myocardial changes which are correlated with the extent and duration of obesity [91], systolic blood pressure, LV end-systolic wall stress and LV chamber size in diastole [92, 93].

The most common structural changes observed include: hypertrophy of cardiac myocytes, concentric remodeling, eccentric and concentric LV hypertrophy, increased LV mass, dilated cardiomyopathy and

increased RV wall thickness which occur in varying degrees depending on the presence or absence of hypertension and obesity class [90, 92-99].

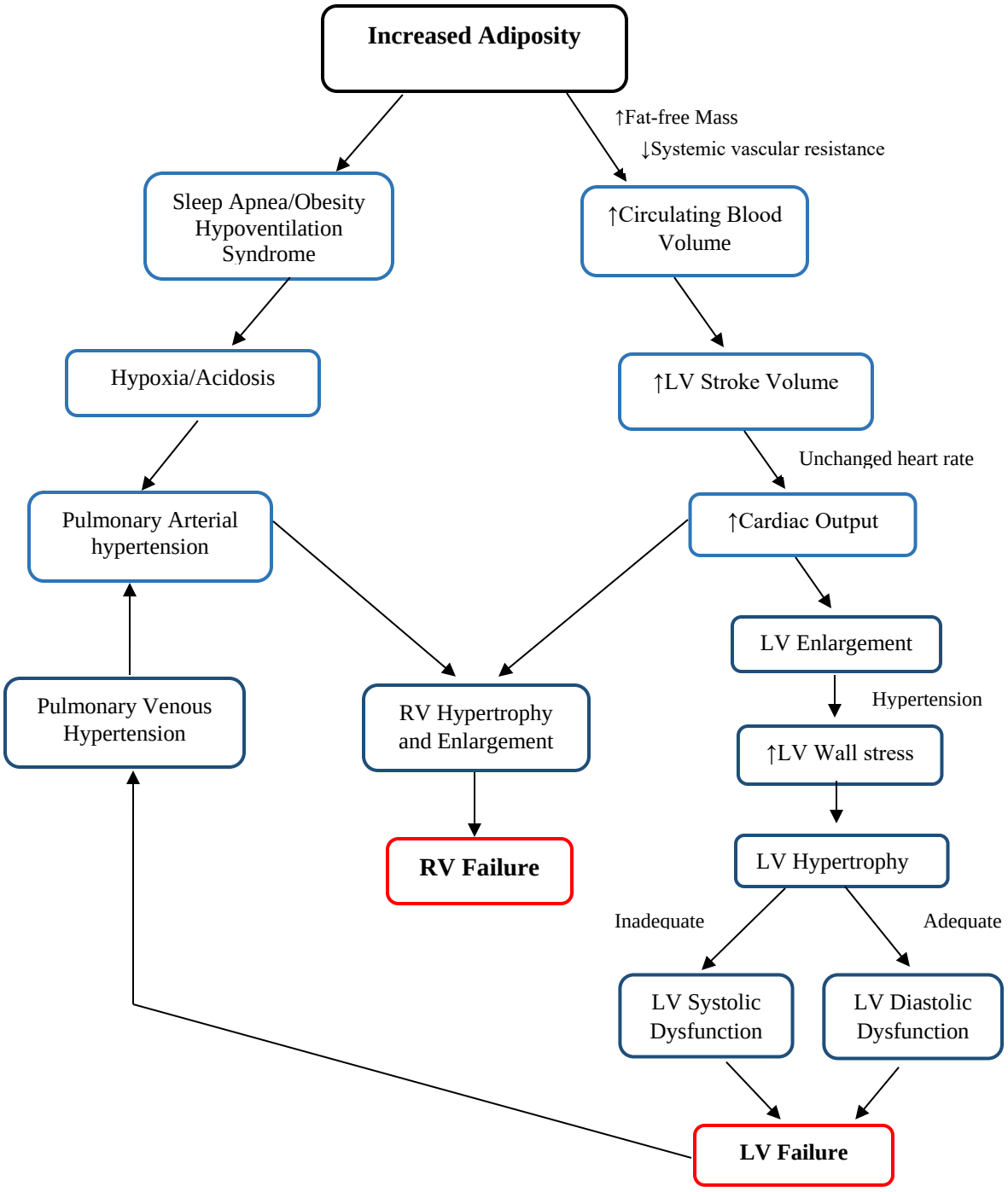


Figure 2: Pathophysiology of Obesity Cardiomyopathy

This figure shows the central hemodynamic, cardiac structural abnormalities, and alterations in ventricular function that may occur in severely obese patients and predispose them to heart failure. Left ventricular (LV) hypertrophy in severe obesity may be eccentric or concentric. In uncomplicated (normotensive) severe obesity, eccentric LV hypertrophy predominates. In severely obese patients with long-standing systemic hypertension, concentric LV hypertrophy is frequently observed and may occur more commonly than eccentric LV hypertrophy. Whether and to what extent metabolic disturbances such as lipotoxicity, insulin resistance, leptin resistance, and alterations of the renin-angiotensin-aldosterone system contribute to obesity cardiomyopathy in humans is uncertain. RV = right ventricular. *(Adapted from Lavie CJ et al: Impact of obesity and the obesity paradox on prevalence and prognosis in heart failure. JACC Heart failure 2013, 1(2):93-102).*

2.3.3.3 Functional Changes – LV Diastolic and Systolic dysfunction

The presence of LV hypertrophy predisposes obese persons to diastolic dysfunction. Several hemodynamic studies found increased LV end-diastolic pressure [90, 92] and other studies utilizing echocardiography and radionuclide techniques reported abnormal diastolic filling pressures suggestive of LV diastolic dysfunction in obese persons, which worsened with increasing levels of obesity [92, 100-103].

Depression of LV systolic myocardial performance is uncommon in obese persons in the absence of concomitant cardiovascular disease. In some obese subjects there was some subclinical LV dysfunction characterized by abnormal myocardial strain which was load-independent [103].

2.3.3.4 Metabolic, cellular and neuro-hormonal changes

Based on data from animal models, it has been hypothesized that the structural and functional changes seen in obese individuals are related to several abnormalities in biochemical and metabolic pathways. These include: decreased insulin sensitivity, hyperinsulinemia, leptin resistance and hyperleptinemia, decreased serum adiponectin levels, increased sympathetic tone, and activation of the Renin Angiotensin Aldosterone System (RAAS), low-grade systemic inflammation (with increased C-reactive protein and tumor necrosis factor). Cellular insults accompanying these processes may engender fibrosis, apoptosis and hypertrophy of cardiac myocytes which ultimately impact diastolic function [14].

2.4 Previous epidemiologic Studies on the association between Obesity and Heart Failure

Previous epidemiologic studies have suggested a strong independent association between elevated body mass index and incident heart failure (Table 5). However, in a community-based study of 550 diabetes-free men and women in Greece, elevated BMI alone was not independently associated with HF risk, whilst metabolic syndrome was associated with a 2.5-fold (95% CI: 1.64-3.58) higher HF risk [104]. Metabolic syndrome was defined using the National Cholesterol Education Program Adult Treatment Panel (NCEP-ATP) III criteria i.e. the presence of 3 or more of the following conditions: abdominal

obesity given as waist circumference (>102 cm in men and >88 cm in women), serum triglycerides >150 mg/dl, high-density lipoprotein cholesterol <40 mg/dl in men and <50 mg/dl in women, BP ≥130/85 mm Hg or use of antihypertensive medications, fasting glucose ≥100 mg/dl.

Table 5: Previous studies investigating the association between measures of obesity and heart failure risk				
Authors	Study Description	Covariates	Findings	Limitations
Kenchaiah et al [13].	Framingham Heart Study (n= 5881, mostly Caucasian, 54% female, mean age = 55.4 years).	Age, alcohol consumption, smoking, total cholesterol, diabetes, hypertension, valvular heart disease, myocardial infarction, LVH (on ECG)	496 (258 women, 238 men) incident HF cases. 1. Continuous BMI HR (95% CI) per 1Kg/m^2 increase: Men: 1.05 (1.02-1.09). Women: 1.07 (1.04-1.10) 2. BMI categories - Overweight vs Normal Men: 1.20 (0.87-1.24). Women: 1.50 (1.12-2.02) - Obese vs. Normal Men: 1.90 (1.30-2.79). Women: 2.12 (1.51-2.97)	1. Assumptions of linearity for continuous BMI analysis. 2. Categorization: loss of precision and misrepresentation of the nature of the dose-response relationship. 3. Couldn't investigate potential differential effects by race.
Ndumele et al [15].	ARIC study (n= 9507, 58.2 female, 21.6% black, mean age = 62.4 years).	Age, sex, race, smoking, diabetes, hypertension, LDL cholesterol, HDL cholesterol, triglycerides, alcohol intake, NT-proBNP, and estimated GFR.	868 incident HF events. 1. Continuous BMI HR per 5Kg/m^2 increase: 1.32. 2. Categorical BMI BMI≥35Kg/m^2 vs Normal (18.5 – 24.9Kg/m^2) Age-adjusted: 3.39 (2.74 - 4.19) Full-adjustment: 2.39 (95% CI: 1.89 -3.01).	1. Assumptions of linearity for continuous BMI analysis. 2. Categorization: loss of precision and misrepresentation of the nature of the dose-response relationship. 3. Potential differential effects by race and sex weren't investigated.
He et al [105].	NHANES (n= 13,643; 59.4% female; 14.8% black, mean age = 49.8 years).	Sex, education, physical activity, cigarette smoking, diabetes, hypertension, valvular heart disease, and coronary heart disease.	1382 CHF events. Categorical BMI Overweight vs Normal: RR (95% CI): 1.30 (1.12-1.52).	1. Categorization: loss of precision and misrepresentation of the nature of the dose-response relationship. 2. Potential differential effects by race and sex weren't investigated.
Chen et al [106].	Established Populations for Epidemiologic studies of the Elderly, New Haven cohort (n= 1749, 58.9% female, 18.4%, black, mean age=74.2 years)	Age categories, type of housing, sex, pulse pressure and diabetes.	173 incident HF events. Categorical BMI BMI≥28Kg/m^2 vs <24.0Kg/m^2) RR (95% CI): 1.6 (1.0-2.4).	1. Categorization: loss of precision and misrepresentation of the nature of the dose-response relationship. 2. Potential differential effects by race and sex weren't investigated.

2.5 Rationale for investigating the association between a weight-height index (as a proxy for total body fat) and HF risk in SCCS.

2.5.1 Rationale for considering a data-derived weight-index (W/H^n)

Data from varying communities across the US suggest that poverty-dense counties have high levels of obesity which are paralleled by high prevalence of chronic conditions including CVD and heart failure [12]. It is ever so important to critically examine the link between adiposity and HF risk (and survival following a diagnosis of HF) particularly among populations with scant resources. Most cohorts investigating the obesity-HF link have relied on body mass index – weight (W)/height $(H)^2$ – as a proxy for total body fat and increased body mass index has been found to be positively correlated with HF incidence and survival [13-15] but there is little data on potential effect modification by race. Less thought has been given to the suitability of W/H^2 in investigating the obesity-HF link despite differences in the performance of various weight-height indices across population groups defined by race and sex [16, 17]. Additionally, predictors with "pleiotropic" effects usually have differential functional relationships with varying outcomes suggesting the need to use the data to derive the appropriate weight-height index for each outcome. Hypothetically, for any weight-height index given by the general form W/H^n, regressing the log-hazard of HF on the natural logs of both weight and height, would yield n as the absolute value of the ratio of the coefficients of log height and log weight; this could be used to a W-H index suitable for a given setting. Also, any potential departures from additivity of effects by race and sex could be investigate using interaction terms with the appropriate weight-height index. Such robust approaches are needed to adequately model these data and reveal novel insights that would potentially refine the strategies used to risk-stratify persons in clinical and/or public health settings.

2.5.2 Rationale for investigating departures from additivity of effects by race in SCCS.

Most of the previous studies enrolled predominantly white populations or small numbers of middle class multi-ethnic populations from small communities; thus the obesity-HF association in blacks was understudied. Importantly, some studies have found that increased BMI have less detrimental health

effects in blacks. Cohen et al and Calle et al reported stronger associations between BMI and cardiovascular outcomes in whites compared to blacks [66, 107]. In the SCCS cohort, we have previously shown that whites with BMI $\geq 40 Kg/m^2$ had a greater than 2-fold increased risk of cardiovascular death compared to those with normal weight, while in blacks the increase in risk was modest (17-40%) and non-significant [107]. Similarly, using data from the SCCS cohort, Lipworth et al found that in whites the risk of atrial fibrillation was 49% (HR: 1.49; 95%CI: 1.11, 2.01) higher in obese persons compared to those of normal weight, while among blacks the corresponding HR was 0.90 (95%CI: 0.69, 1.16) [108]. This suggests that the association between BMI (and potentially other anthropometric surrogates of total body fat) and cardiovascular outcomes including HF needs to be further contrasted between blacks and whites. Additionally, a more flexible modeling approach take into account potential nonlinearity of effects may reveal additional insights into the causal relationships between W/H^n and HF.

Hence, we plan to use the unique opportunity afforded by the SCCS data to investigate potential differential patterns in the dose-response relationship between W/H^n and the risk of incident heart failure by race and sex.

We will investigate these associations using data from the linkage of the SCCS cohort with CMS Research Identifiable Files to ascertain incident HF events. We would use data for body mass index at enrollment based on self-reported weight and height (and validated in a sub-cohort of SCCS) to investigate the W/H^n-HF association.

For all these analyses we would use data for SCCS participants (n = 27,078) who meet the following inclusion criteria would be included in our analyses:

- $\geq$ 65years (n = 7001) at cohort enrollment
- < 65 years (n = 20,077) at enrollment and either:
    i. reported being covered by Medicaid on the baseline questionnaire;
    ii. reported being covered by Medicare on the baseline questionnaire;

iii. did not report Medicare or Medicaid on the baseline questionnaire but had a CMS claim within 90 days of being enrolled in SCCS.

We hypothesize that a) there is a positive non-linear dose-response association between W/H^n and incident HF which is stronger in whites compared to blacks. Cox regression analyses would be used to model the association between W/H^n and the instantaneous risk of incident HF while adjusting for all relevant covariates (age, history of MI/CABG, stroke, hypertension, diabetes, dyslipidemia, income, education, smoking, alcohol use, and total physical activity) and modelling interactions with race and sex.

We would repeat these analyses using waist circumference data which is available for a subset (n ≈ 3300) of our study sample.

2.5.6 Obesity and post-HF survival

2.5.6.1 The Obesity Paradox

Whilst the biologic evidence for the nefarious effects of excess body weight on the structure and functioning of the myocardium and the epidemiologic evidence linking obesity to increased HF incidence are both compelling, several studies (table 6) have suggested that overweight and obese persons with heart failure have a demonstrably higher survival compared to leaner subjects – a phenomenon coined as the obesity paradox.

Table 6: Previous studies investigating the association between measures of Obesity and post-HF mortality.				
Authors	Study Description	Covariates	Findings	Limitations
Horwich et al [109].	1203 patients (23.4% female, mean age = 52 years) with advanced HF (mean LVEF = 22%) referred for care.	Age, sex, medications, hypertension, diabetes, LVEF, peak VO_2, serum sodium, serum lipids, serum creatinine, mitral and tricuspid regurgitation.	- 537 deaths - HR for 1Kg/m2 increase in BMI: 0.952 (0.915-0.991). - Higher Cumulative survival at 2 years in overweight (BMI= 27.8-31Kg/m^2) and obese (BMI >31Kg/m^2) groups compared to normal weight.	1. Small sample size 2. Categorization: loss of precision and misrepresentation of the nature of the dose-response relationship. 3. Potential differential effects by race wasn't investigated.
Lavie et al [110].	209 ambulatory HF patients (mean age = 53.9yrs) with chronic systolic HF referred for care at UCLA cardiomyopathy center.	Age, sex, ischemic vs non-ischemic cardiomyopathy, NYHA class, LVEF and peak VO_2.	28 major events (urgent transplant or CV death). 1. Percent body fat: For every 1% increase in percent body fat there was a 13% reduction in major CV events. 2. Categorical BMI Kaplan-Meier plot showed higher survival for 5th BMI quintile compared to 1st quintile.	1. Small sample size 2. Categorization: loss of precision and potential misrepresentation of the dose-response relationship.
Clark et al [111].	3187 patients (25.2% female, mean age = 53.0 years, mean LVEF = 22.9%) referred for HF management or transplant evaluation at UCLA.	Age, diabetes, LVEF, peak VO_2, NYHA class, HF cause (ischemic vs non-ischemic)	988 Major events (deaths, heart transplants, ventricular assist device placements) 1. Body Mass Index HR (95%CI) for normal vs High BMI (≥25 Kg/m^2): - Men: 1.34 (1.13-1.58); Women:1.38 (1.02-1.89) 2. Waist circumference (WC) - Men HR (95%CI) for normal vs High WC (≥102cm): 2.02 (1.18-3.45) - Women HR (95%CI) for normal vs High WC (≥88cm): 2.99 (0.90-4.8).	1. Categorization: loss of precision and misrepresentation of the nature of the dose-response relationship. 2. Potential differential effects by race wasn't investigated.
Oreopoulos et al [112].	Meta-Analysis of 9 observational HF studies (n= 28,209) with average follow-up ≈ 2.7 years.		BMI categories - HR (95% CI) for Overweight vs Normal CV death: 0.81 (0.72 – 0.92) All-cause Mortality: 0.84 (0.79 – 0.90) - HR (95% CI) Obese vs. Normal CV death: 0.60 (0.53 – 0.69) All-cause Mortality: 0.67 (0.62 – 0.73)	Categorization: loss of precision and misrepresentation of the nature of the dose-response relationship

Some authors have suggested a few underlying biological reasons to explain this phenomenon.

- Higher metabolic reserve among obese persons in addition to leaner HF patients having a higher likelihood of being cachectic and thus at a higher risk of death [113].
- The specificities of the neuro-humoral profile and hemodynamics of obese persons with heart failure such as the secretion of soluble TNF alpha receptors which have beneficial neutralizing effects; higher circulating lipoproteins which may bind and detoxify lipopolysaccharides that play a role in stimulating the release of inflammatory cytokines and elevated blood pressure which raises the tolerance of obese individuals to higher doses of cardio-protective agents including beta blockers, aldosterone antagonists and RAAS inhibitors [114, 115].

Other investigators have advanced epidemiologic reasons related to:

1) Study participants who experienced unintended weight loss before study enrollment and who may have had higher mortality risk [114].
2) The fact that compared to normal weight persons, overweight and obese persons were found to have lower atrial natriuretic peptide levels which was correlated with having higher muscle mass and muscle strength [116]. The latter have been associated with improved survival in other patient populations and this may be analogous to patients with advanced HF as well [117].

2.5.6.2 Rationale for investigating the association between anthropometric surrogates and all-cause mortality among HF patients in SCCS.

There is a wealth of data suggesting that in persons with HF, intentional weight loss mitigates some of the hemodynamic abnormalities and reduces LV mass and chamber size [90, 92, 93, 100, 114, 118]. However, in light of the epidemiologic evidence suggesting an obesity paradox, some cardiovascular societies – whilst still advocating weight loss – have been conservative, recommending intentional weight reduction only at higher cut-points for BMI. For example, the AHA recommends intentional weight loss in HF only for persons with BMI >40 kg/m^2, the Heart Failure Society of America

for individuals with BMI >35 kg/m^2, both the European Society of Cardiology and the Canadian Cardiovascular Society recommend weight loss above a BMI cut-point of 30 kg/m^2 and none of the major societies recommend weight loss for overweight patients with HF.

These recommendations seem to make the assumption that there exists a discontinuous relationship between the survival probability of persons with heart failure and body mass index; and the discontinuities (or sharp inflexion points) are found at the recommended cut points. If the dose-response relationship between body mass index and survival probability in heart failure is in fact a smooth function, making recommendations about specific cut-points may be counter-intuitive.

We plan to use data from SCCS to investigate the dose-response relationship between W/H^n and mortality risk in persons with incident heart failure while accommodating potential nonlinearity of effects and effect modification by race.

[However, many experts acknowledge that in patients with heart failure more data on the effects of intentional weight loss on morbidity (hospitalizations) and long-term prognosis including hard endpoints like cardiovascular and all-cause mortality would provide better insights.]

3. Materials and Methods

The current study would leverage existing data from the Southern Community Cohort Study to a) quantify the burden of heart failure (incidence and mortality) in a low-income population that has been under-represented in previous cardiovascular cohorts b) improve the understanding of the dose-response relationship between weight-height indices (as surrogates of total adiposity) and HF risk and the potential differential influences of race and sex in these relationships c) elucidate the independent contrasting and/or synergistic relationships between individual-level and neighborhood-level effects which may inform individual and community-level interventions aimed at reducing the burden of the heart failure epidemic.

3.1 Study Population: The Southern Community Cohort Study

The SCCS enrolled approximately 86,000 adults, age 40-79, living in rural and urban settings in a 12-state area of the southeastern United States (Tennessee, Arkansas, Louisiana, Missouri, Alabama, Georgia, Florida, South Carolina, North Carolina, Virginia, West Virginia and Kentucky) between 2002 and 2009 [18]. The majority of participants (~85%) were recruited primarily at community health centers (CHCs), which are federally funded healthcare facilities primarily servicing low-income individuals and medically underserved populations. Thus, most SCCS participants are of similar (usually low) socioeconomic and at especially elevated risk of adverse health outcomes including obesity, hypertension (~56%), diabetes (~21%) and cardiovascular disease. The characteristics of participants enrolled in the SCCS are summarized in Table 7.

Table 7. Baseline characteristics of 76,614 SCCS Participants by Sex and Race, 2002–2009								
	Black Men		**White Men**		**Black Women**		**White Women**	
	No.	%	No.	%	No.	%	No.	%
Age, years								
40–49	11,179	51	3,407	37	14,59	47	5,486	38
50–59	7,526	34	3,222	34	10,47	34	5,176	36
≥60	3,309	15	2,714	29	5,737	19	3,785	26
Mean	51.3		54.7		52.		54.	
Education								
<9 years	1,913	9	777	8	2,192	7	1,060	7
9–11 years	5,422	25	1,261	14	6,954	23	2,388	17
High school	8,996	41	3,214	34	11,95	39	5,625	39
Some college	3,766	17	1,810	19	6,314	21	2,975	21
College or postgraduate	1,899	9	2,270	24	3,366	11	2,392	17
Income								
<$15,000	12,925	59	3,730	41	17,94	59	6,955	49
$15,000–$24,999	4,693	22	1,565	17	7,064	23	2,778	20
$25,000–$49,999	2,799	13	1,611	18	3,887	13	2,344	17
$50,000 or more	1,351	6	2,291	25	1,491	5	2,125	15
Smoking								
Never	4,863	22	2,389	26	14,57	48	5,573	39
Former	4,553	21	3,137	34	5,978	20	3,628	25
Current, <1 pack/day	8,983	41	1,200	13	7,684	25	2,268	16
Current, ≥1 packs/day	3,411	16	2,467	27	2,299	8	2,875	20
Alcohol consumption								
None	6,607	30	3,631	40	16,68	55	7,999	56
<1 drink/day	6,284	29	2,964	32	9,610	32	4,944	35
≥1 drinks/day	8,792	41	2,534	28	3,971	13	1,231	9
Body mass index at cohort entry								
<18.5	278	1	82	1	311	1	254	2
18.5–24.9	7639	35	2544	27	4777	16	3638	25
25–29.9	737	35	3542	38	7860	26	3831	26
≥ 30	5,535	29	2,705	34	17,95	58	6,824	47

3.2 Study population for our proposed analyses

SCCS participants (n = 27,078) who meet the following inclusion criteria would be included in our analyses: ≥ 65years (n = 7001) at cohort enrollment or < 65 years (n = 20,077) at enrollment and either: a) reported being covered by Medicaid on the baseline questionnaire; b) reported being covered by Medicare on the baseline questionnaire or c) did not report Medicare or Medicaid on the baseline questionnaire but had a CMS claim within 90 days of being enrolled in SCCS.

The restriction to these groups increases the likelihood of participants having continuous coverage in Medicare and/or Medicaid from the time of SCCS enrollment to the end of the follow-up period (December 31st, 2010), for the ascertainment of incident HF events.

Based on these criteria we had n = 33,018 participants. We further excluded the following groups of persons successively (given these groups were not mutually exclusive) (fig. 4):

a) Persons (n = 1571) who did not self-report as "Black" or "White" given the sample sizes for these groups were too small to make significant inferences.
b) Persons (n = 4312) with a known history of HF before cohort entry (enrollment) i.e. prevalent HF.
c) Persons (n = 51) whose date of HF diagnosis was coincident with date of enrollment (i.e. follow-up time from cohort entry to HF diagnosis = 0) as these persons may have had prevalent HF. Given HF is a chronic condition whose onset is often insidious, these participants may have been symptomatic for HF before enrollment.
d) Persons (n = 6) whose diagnosis of HF was later confirmed to have occurred after end of follow-up (December 31st 2010).

In summary, a further 5940 subjects were excluded leaving a total sample size of 27,078.

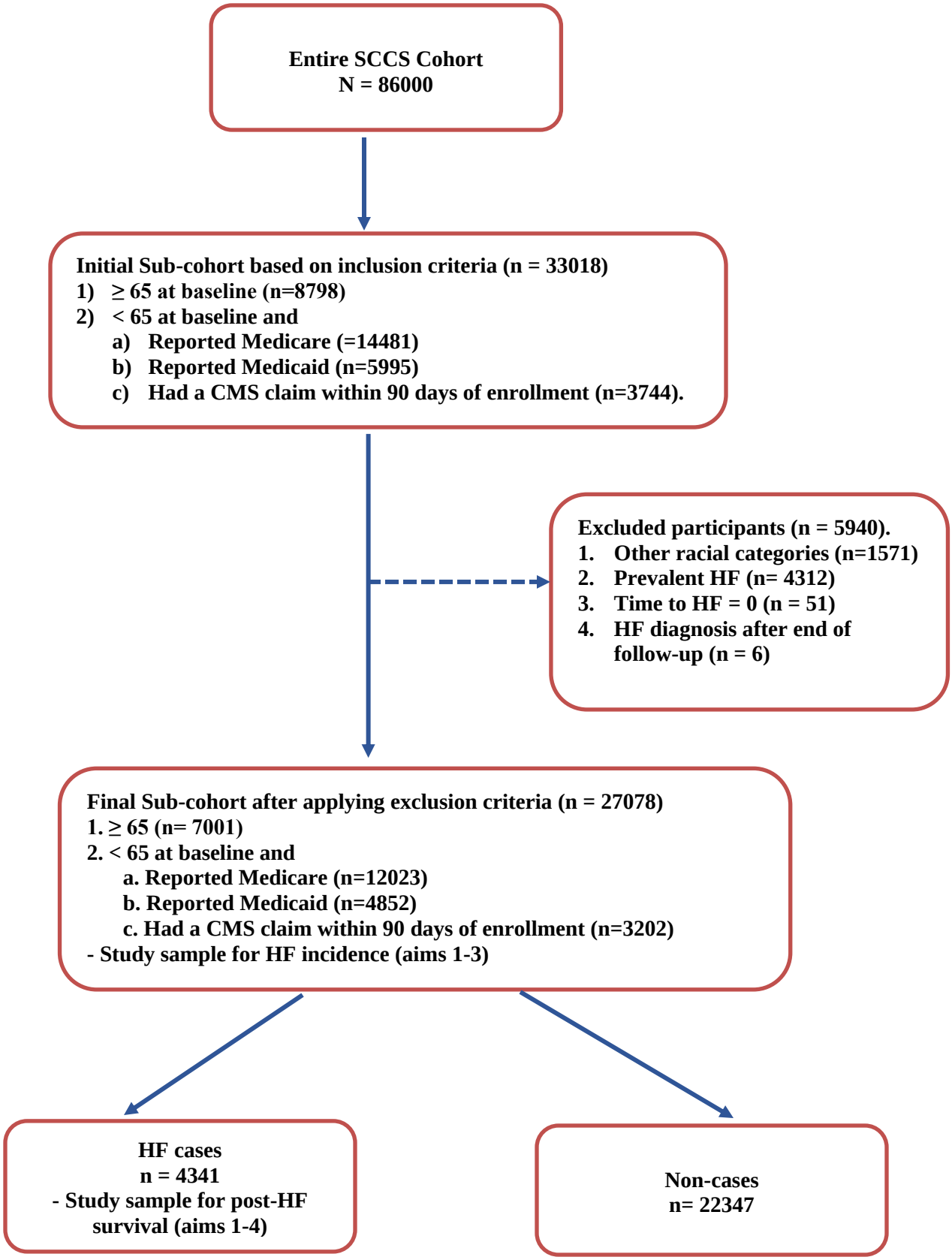


Fig 3: Flow diagram indicating participant eligibility for the proposed study.

3.3 Comparison between SCCS and other cohorts investigating HF

Table 8: Baseline characteristics of SCCS, ARIC, CHS, MESA and FHS participants

	SCCS (n= 27,078)	**ARIC[8] (n= 14,993)**	**CHS[5] (n= 5,888)**	**MESA[6] (n = 5,923)†**	**FHS[7] (n=9,405)‡**
Age (years)*	55.5 (10.4)	54 (6)	72.8 (5.6)	61.8 (10.3)	41 (10)
Women (%)	62.6	54	57.6	53	53
Blacks (%)	68.8	27	15.7	26.1	≈ 0
Education (<high school) (%)	38.4	24	29.5	16	56§
BMI (kg/m^2)*	30.4 (7.8)	27.3 (5.1)	26.7 (4.7)	28.0 (5.4)	24.9 (3.8)‖
Obese (BMI > 30) %	44.8	26	19	32	10.8‖
Diabetes (%)	26.5	11	16.4	11.6	4.1
Hypertension (%)	62.5	33	57.7	42	7
Myocardial infarction (%)	8.6	4	9.6	n/a	1.6#
Stroke (%)	9.6	1.4	4.2	n/a	0.5**
Ever smoked cigarettes (%)	65.3	58.2	53.5	49	57.3§

Compared to ARIC, SCCS participants included in the current study had higher proportions of persons reporting a history of myocardial infarction at baseline as well as hypertension, BMI≥30Kg/m^2 and proportion of black participants.

Compared to CHS and MESA, the SCCS population was younger at baseline, had higher proportions of women and African-Americans; and equally had higher mean BMI and prevalence of hypertension and diabetes.

3.4 Data Collection in SCCS

At enrollment into the SCCS, participants completed a questionnaire which ascertained information about demographics (date of birth, sex, race), lifestyle factors (including smoking and alcohol use), personal and family medical history, anthropometric factors, education, occupation, income, and physical activity data [18]. For CHC participants, the questionnaire was administered in person via a

computer-assisted personal interview (CAPI) with logic-checking and skip pattern technology. Trained interviewers administered the questionnaires with the assistance of handheld cards to facilitate responses, and abstracted anthropometric and clinical information from participants' medical records if those measurements were also collected on the day of enrollment. General population participants (~15%) completed and mailed in a paper version of the study questionnaire. Recruitment began in March 2002 and was completed in September 2009. Institutional review boards at Vanderbilt University Medical Center and Meharry Medical College approved all SCCS research protocols and participants provided informed consent and signed HIPAA authorization forms.

3.5 Quality Control for Measurements and Questionnaires

Most of the questions in the SCCS questionnaire were adapted from questionnaires used and validated in other settings. However the SCCS investigators performed a series of validation studies to assess the reliability of the questionnaires used in data collection. Samples of approximately 150-800 SCCS participants were selected to validate physical activity patterns via repeat interviews and use of accelerometers, tobacco use status via measurement of serum nicotine, and disease occurrence for self-reported diabetes via confirmation in medical records and by measurement of HbA1c [18]. Over 96% of self-reported diabetes diagnoses were validated with data from confirmatory medical records or elevated HbAlc measurements. Self-reported height and weight were compared with contemporaneous clinic recorded measurements for over 20% of the participants (correlation was >95%) [18]. Repeat blood samples were also collected and banked from over 650 participants and repeat urine samples from over 240 participants for future comparability assays to check for concordance of biomarker levels over time periods spanning one to three years.

3.6 Outcomes and Outcome ascertainment

The main outcomes for our study would be incident heart failure and survival following a first diagnosis of heart failure. Heart failure events would be ascertained via linkage of the SCCS cohort (using Social Security Number, date of birth, and sex) with national Centers for Medicare and Medicaid Services

(CMS) Research Identifiable Files. Given that Medicare coverage is almost universal for persons who are aged 65 or older, and the low-income status of SCCS participants under age 65 results in many being eligible for Medicaid, the CMS linkage affords the opportunity for ascertainment of heart failure diagnoses for a significant proportion of subjects (n = 27,078) in the SCCS.

3.6.1 Incident Heart Failure

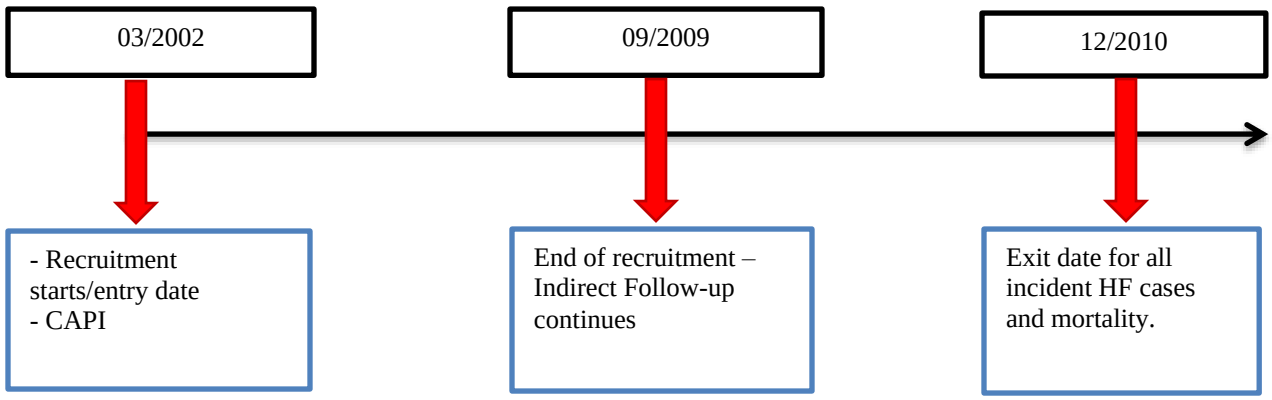
Incident HF would be defined as the first occurrence of a medical claim with ICD-9 code 428.x within the Medicare institutional (Medicare Provider Analysis and Review, MEDPAR), Part B carrier, or outpatient-based claims files or the Medicaid Analytic Extract (MAX) Inpatient and Other Services claims files, from the date of SCCS enrollment through December 31st, 2010.

The diagnosis code (ICD-9 428x) [119] algorithms for identification of HF in the proposed study have been previously validated. A review of the detection of heart failure in administrative claims data included eight studies conducted among Medicare beneficiaries reported positive predictive values (PPVs) between 76% and 99%, with the majority of the studies reporting PPVs over 90% [119]. These codes have also been used with high specificity in a number of studies [120, 121].

3.6.2 All-cause mortality/Post-HF Survival

Deaths, including dates and causes of death, would be ascertained via linkage of the SCCS cohort with both the Social Security Administration (SSA) vital status service for epidemiologic researchers and the National Death Index (NDI) through December 31, 2010. Both NDI and SSA are well-established and reliable means of identifying deaths in the US, and are expected to capture nearly all deaths [51, 122, 123].

Figure 4: Timeline for recruitment, data collection and ascertainment of outcomes



3.6.3 Understanding the linkage with CMS research Files for HF ascertainment

3.6.3.1 Medicare

3.6.3.1.1 Types of Medicare coverage

Medicare is the primary health insurance program for persons aged ≥ 65; hence Medicare data is a valuable resource for making inferences about medical care of older adults. The program also provides insurance coverage for persons < 65 with certain disabilities, and individuals of all ages with ESRD.

Almost all Medicare beneficiaries receive Part A hospital insurance benefits, which helps cover inpatient hospital care, skilled nursing facility stays, home health and hospice care. The majority of beneficiaries equally subscribe to Part B medical insurance benefits, which help to cover physician services, outpatient care, and durable medical equipment (DME) and some home health care. Also, many beneficiaries elect to purchase Medicare Part D prescription drug coverage.

Beneficiaries may elect to receive fee-for-service (FFS) Medicare. In an FFS plan, treatment and diagnostic procedure decisions are under the control of the physician/hospital based on what they consider to be best practice and they're reimbursed for every service offered to the patient or client. As an alternative to FFS Medicare, beneficiaries can enroll in Medicare Part C (Medicare Advantage). These are

private plans similar to manage care organizations which provide Medicare Part A and Part B services. Managed care organizations supervise the financing of medical care delivered to their members. They include health maintenance organizations (HMOs), preferred provider organizations (PPOs) and points of service and the flexibility afforded to the clients in (e.g. for specialist consultation options) vary across these plans.

3.6.3.1.2 The Chronic Condition Ware house (CCW) Medicare research data files

The Chronic Condition Data Warehouse (CCW) Medicare data are extracted from the Centers for Medicare & Medicaid Services (CMS) enrollment files and fee-for service administrative claims submitted for payment to CMS. Data for all beneficiaries enrolled in Medicare is available from the CCW. The CMS Institutional and Non-institutional data files found in the CCW generally represent Medicare FFS claims only (i.e., managed care encounter information is not available). A few exceptions exist, including coverage of Hospice services.

The CCW data files were designed in such a way as to facilitate research across the continuum of care, using data files that could be easily merged and analyzed by beneficiary. Each beneficiary in the CCW is assigned a unique, unidentifiable link key, which allows researchers to easily merge data files and perform relevant analyses across different claim types and enrollment files.

3.6.3.1.3 The Master Beneficiary Summary File (MBSF)

The Master Beneficiary Summary File is created annually and contains demographic and enrollment data for all beneficiaries enrolled in Medicare for any part of the year. This annual person level summary file can be used to determine whether a beneficiary has a sufficient surveillance period for inclusion in the analytic file being created. Variables contained in this file include: the number of months of Medicare Part A, B, C, and D coverage; whether the beneficiary died during the year and other beneficiary demographic and geographic information.

3.6.3.1.4 Claim Types

In general, all CMS administrative data files contain variables which can be used to join the CCW files. When medical services provided to a beneficiary are the focus, the primary linkage will be at the person level, after aggregation of the claim level files.

a) Medicare Institutional Claims

Claims from institutional providers which are covered by the Medicare Part A benefit appear in the Institutional claims file and are processed by Medicare Administrative Contractors (MAC). In addition, claims for institutional-based services covered by the Medicare Part B benefit (e.g., home health, hospital outpatient) appear in the Institutional claims file. For each setting, there is a base file and a revenue center file. These files include the following types of claims:

- Inpatient - This file includes ICD-9 diagnosis and procedure codes, Diagnosis Related Group (DRG) information, dates of service, reimbursement amount, hospital provider, and beneficiary demographic information.
- Outpatient - Contains claims data for outpatient services submitted by institutional outpatient providers (e.g., hospital outpatient departments, Rural Health Centers [RHCs], Federally Qualified Health Centers [FQHCs], renal dialysis facilities, outpatient rehabilitation facilities, comprehensive outpatient rehabilitation facilities, and community mental health centers). This file includes ICD-9 diagnosis codes and CMS Common Procedure Coding System (HCPCS) codes.
- Skilled Nursing Facility (SNF), Hospice, Home Health Agency (HHA) which provide base files for claims submitted by the SNF, Hospice and HHA respectively and all include ICD-9 diagnosis codes.

Institutional claims have base files (contains beneficiary ID, claim type, admitting diagnosis, primary diagnosis, and up to ten additional diagnosis code fields and six procedure codes with associated dates, as well as the DRG; and beginning with claims files from 2009, the claims allow for 25 diagnosis codes and 25 procedure codes) and revenue center files.

b) Medicare Non-Institutional Claims

The Medicare Non-Institutional claims include services covered by the Part B benefit, and consist largely of professional services and DME. All claims are processed by MACs; also known as Carriers. These files include the following types of claims:

- Carrier – Contains claims data for non- institutional providers (e.g., physicians, physician assistants, nurse practitioners, independent clinical laboratories, ambulance providers, and freestanding ambulatory surgical centers).
- Durable Medical Equipment (DME) – Contains non-institutional claims for the Durable Medical Equipment Regional Carrier submitted by DME suppliers and providers.
- Non-institutional claims have base files (contains the beneficiary ID, claim type, referring physician, carrier number and up 8-12 diagnosis codes) and line files (contains the HCPCS and ICD-9 diagnosis codes as well).

3.6.3.1.5 Investigating chronic disease rates in a sample of Medicare Beneficiaries

Ideally, in order to capture all cases of disease and comorbid conditions in a sample of Medicare beneficiaries, it may be best practice to restrict analyses to enrollees who have FFS Medicare A and B coverage. However many beneficiaries with part A coverage have a state buy-in via Medicaid to cover the part B premium. Those enrolled in Medicare Part C have managed care coverage, and the transactional data regarding services received are not included in the claims data files. The MBSF indicates the type of Medicare coverage obtained. The Medicare state buy-in variable appears 12 times in the MBSF to represent each month of coverage. The values within this variable indicate whether the beneficiary had Medicare Part A and/or B coverage for the month, and whether there was state buy- in (i.e., Medicaid) for the Part B premium. A limitation of this state buy-in variable is that it does not provide information regarding whether the beneficiary was entitled to full or only partial Medicaid benefits. To determine whether the beneficiary had Medicare FFS or managed care coverage, the HMO indicator variable

appears 12 times to represent each month of coverage. The MBSF also contains information regarding whether the person was dually eligible for both Medicare and Medicaid services.

Not all Medicare enrolled beneficiaries will have used Medicare-paid services in a particular time frame. Some do not use any medical services at all, while others may use services that are paid by a third party (e.g., the Veteran's Administration). Third party claims do not appear in the Medicare data files.

3.6.3.1.6 Patterns of Health Care Utilization

Institutional claims cover both inpatient and outpatient settings. Non-Institutional claims cover a variety of settings including physician office, laboratory, imaging, procedures, and others. Visits with health care professionals in the ambulatory setting may take place in an outpatient facility or a provider office setting. Outpatient care is identified from the Institutional Outpatient claims files. Physician Office Care, is a small portion of the claims found in the Part B Carrier files.

Hospital Outpatient (HOP) claims are considered Institutional data files due to the type of claim used with Medicare, even though these claims are generally paid through the Medicare Part B benefit. Part B non-institutional – "carrier" – claims include: evaluation and management (physician office, specialist, consultation…), procedures (anesthesia, major cardiac procedure…), Imaging (echography, standard imaging), labs, DME etc.

For some services there could be an Institutional or Non- Institutional claim, or both claim types. In general, the professional component of a service (e.g., the physician or therapist care) appears in the Non-Institutional file (Part B Carrier), whereas the facility claim for an associated service, when applicable, appears in the Institutional file (Part A Inpatient or Skilled Nursing Facility [SNF]).

One of the most common examples of a type of service which may appear in either the Institutional or Non-Institutional claims is outpatient clinic-type services for physician/other provider care. Understanding the ambulatory care provided to a patient (e.g., a physician/clinic visit for a service),

requires examining the hospital outpatient file (hospital-based clinics, RHCs and FQHCs) in addition to the Part B Carrier files.

3.6.3.1.7 Diagnosing Disease in Medicare

Medicare claims use the ninth version of the International Classification of Diseases (ICD-9) to classify all diagnoses, which identify the condition(s) for which a patient is receiving care. Claims data generally allow providers to specify numerous diagnosis codes (up to 25 codes for Part A claims and up to 12 codes for Part B claims beginning with claims files from 2009, when the version J data file layout was implemented), with one diagnosis identified on the claim as the principal or primary diagnosis. The diagnosis codes appear on the base claims.

Medicare assigns hospital discharges to diagnosis-related groups (DRGs), a classification system that groups similar clinical conditions and procedures. The beneficiary's principal diagnosis and secondary diagnoses, as well as any procedures performed during the stay, are used to determine the appropriate DRG. Medicare switched to a modified system, called Medicare Severity Diagnosis Related Groups (MS-DRGs) on October 1, 2007; both DRGs and MS-DRGs appear in the base portion of the claims.

Sometimes, to reduce the risk of false positives during outcome ascertainment, it may be prudent to require the presence of more than one claim to rule-in the condition of interest. In the CCW, for example, some chronic conditions in the Medicare population are pre-coded and appear in the MBSF – Chronic Conditions segment. The algorithms for these conditions are very precise regarding the number of claims, the specific types of services, and the number of years of data which must be examined to make a CCW determination regarding whether a person was likely receiving care for a particular condition. Incidentally, the use of this information assumes that if a claim was processed with a particular diagnosis code, the patient was receiving care for that particular condition.

3.6.3.1.8 Limitations of using Medicare data for computing rates for persons < 65.

Medicare beneficiaries aged < 65 may differ from the general Medicare population and the general U.S. population in several ways that may affect disease outcomes. So when the goal is to calculate rates for certain metrics, it may be desirable to consider sensitivity analyses excluding these persons.

3.6.3.2 Medicaid

3.6.3.2.1 The Medicaid Program

The Medicaid Program provides medical benefits to groups of low-income adults, children (State Children's Health Insurance Program), pregnant women, elderly adults and people with disabilities, some who may have no medical insurance or inadequate medical insurance. Medicaid is administered by states, according to federal requirements and the program is funded jointly by states and the federal government.

Medicaid key eligibility groups include: a) the categorically needy like pregnant women and children under age 6 whose family income ≤ 133 % of the Federal poverty level; b) medically needy like persons aged ≥65 and disabled persons whose income levels do not allow them to qualify as categorically needy; and c) Special groups like Qualified Medicare Beneficiaries (Medicaid pays Medicare premiums, deductibles and coinsurance for individuals with income ≤100% of the Federal poverty level and resources ≤ 2× the standard allowed under SSI).

Mandatory services provided by State Medicaid plans include Inpatient and Outpatient hospital services including FQHCs and RHCs, Labs and X-rays etc.

3.6.3.2.2 Medicaid data sources

Each state's Medicaid agency collects enrollment and claims data for persons enrolled in Medicaid and the Children's Health Insurance Program (CHIP). These data are collected in the state's Medicaid Management Information System (MMIS) which is the basic source of state-submitted eligibility and claims data on the Medicaid population, their characteristics, utilization, and payments. Because the Medicaid program varies by state, the data in the MMIS are converted into a national standard and submitted to CMS via the Medicaid and CHIP Statistical Information System (MSIS).

The MSIS enrollment and claims data are reported to CMS on a quarterly basis. The enrollment data identify Medicaid and CHIP enrollees in each month of that quarter and whose enrollment in a prior period should be revised (due to a correction or retroactive enrollment). The fee-for-service (FFS) claims data identify persons who received service. The FFS claims data are submitted based on the quarter in which the claim was adjudicated, not when the service was performed. The managed care encounter records identify who received what service under which managed care organization and from which provider. The encounter records are submitted based on the quarter in which the encounter was processed, not when the service was performed.

The MSIS data are challenging to use for research because the data represent a mixture of time periods. CMS developed the Medicaid Analytic Extract (MAX) files, which is a more research-friendly set of Medicaid administrative files. The enrollment information in MAX identifies monthly enrollment after the retroactive/correction records have been applied and after certain state-specific data elements are transformed into a consistent, national format. The claims in MAX identify the services rendered. The enrollment pertains to people enrolled in the given calendar year and claims pertain to the services rendered in that same time period, thereby making a consistent—and more meaningful—time period for analyses of enrollment and service utilization.

3.6.3.2.3 Medicaid Analytic eXtract (MAX) Data for Research Purposes

Medicaid analytic extract (MAX) data are person-level data files on Medicaid eligibility, service utilization and payment information for all Medicaid enrollees – whether they have received one or more Medicaid service in a given calendar year or none. The purpose of MAX is to produce data to support research and policy analysis on Medicaid populations.

MAX is produced from 7 Fiscal Year quarters of MSIS data from all 50 states plus the District of Columbia. All 7 quarters of MSIS data needs to be approved (due to lags in obtaining updated eligibility information and adjudicating claims) in order to create MAX data.

a) MAX Production

MAX transforms data from the 7 fiscal year quarters to a calendar year; MAX data is event based. For every service rendered for every Medicaid beneficiary, MAX combines initial claims voids and other adjustments to create a "final action event".

b) MAX Datasets

MAX consists of 1 Person Summary File and 4 Claims Files inpatient, long term care, prescription drug and other services). The former includes person level data on eligibility, demographics, managed care enrollment, a summary of utilization and Medicaid payment by type of service. The files include fee for service (FFS) claims as well as managed care encounter data.

c) MAX Enhancements to MSIS Data

MAX provides more detail on Medicaid Eligibility: Improved identification of unique enrollees; retroactive eligibility in proper chronology; verified eligibility with eligibility data added to each claim.

Data on Dual Medicare and Medicaid Status:

- Qualified Medicare Beneficiaries. (QMB's)
- Specified Low –Income Medicare Beneficiaries. (SLMB's)
- Identifies and links Medicare HIC number and Medicaid Case Number.

MAX provides detail on service claims including types of service and detailed data on conditions and treatments such as ICD-9-CM Diagnostic Codes, Procedure Codes and National Drug Codes, Improved Coding for Services and therapeutic usage data added to each prescription drug record.

d) Enhancements to MAX Data

Some key novel enhancements to the data include: improved verification of SSNs, expanded detail on enrollee race and ethnicity, *Monthly* dual Medicare and Medicaid enrollment (earlier years were

reported quarterly); link to Medicare Enrollment Data Base (best way to identify dual eligible) and capability for other linkages (e.g. SSA)

e) Challenges in Developing Consistent Medicaid Data

There are challenges related to a) Medicaid differences (including eligibility and types of services provided) across states, over time and when Fiscal Agents change; b) Eligibility and Type of services provided

f) Limitations of Medicaid Data

These are related to a) Eligibility (e.g. minimal information on other insurance coverage) and b) Services (provided only during "spells of eligibility", varying service coverage varies by state, incomplete data for "dual eligibles") c) data are not always available in a timely fashion.

3.6.3.3 Preliminary data and challenges in ascertaining complete and continuous CMS coverage for SCCS Sub-cohort

1. Inclusion Criteria for the proposed study

As the SCCS uniquely captures a population of underserved, underrepresented, and often uninsured participants, in order to ascertain incident HF, use of Medicaid and Medicare claims databases will be used. To increase the likelihood of participants having complete and continuous coverage in Medicare and/or Medicaid from the time of SCCS enrollment to the end of the follow-up period (December 31st, 2010), for the ascertainment of incident HF events we are restricting the analyses to:

- Persons ≥ 65years (n = 7001) at cohort enrollment, or persons < 65 years (n = 20,077) at enrollment who: a) reported being covered by Medicaid on the baseline questionnaire; or b) reported being covered by Medicare on the baseline questionnaire; or c) did not report Medicare or Medicaid on the baseline questionnaire but had a CMS claim within 3 months of being enrolled in SCCS.

The numbers for the mutually exclusive groups are given in Table 9 below:

Table 9: Exclusive participant categories based on the inclusion criteria for the current study

Criteria	n	%
Aged ≥ 65 at enrollment	7001	25.9
Aged ≤ 65 and reported Medicaid at enrollment	12023	44.4
Aged ≤ 65 and reported Medicare at enrollment	4852	17.9
Aged ≤ 65 and Any CMS Claim within 90 days	3202	11.8
Total	27078	100

2. Ascertaining CMS coverage for participants aged ≥ 65 for proposed study

Table 10: Self-reported CMS categories among participants aged ≥ 65

	n	%	Any CMS Claim ≤90 days, %
Medicaid Only	537	7.7	76.7
Medicare Only	4264	60.9	75.1
Both	1513	21.6	82.2
None	394	5.6	42.1
Missing	293	4.2	62.8
Total	7001	100	74.4

Overall, over 82% of persons in the current SCCS sub-cohort filed a claim in CMS within 90 days of enrollment regardless of self-reported CMS coverage at baseline; that figure is ~74% among persons aged ≥65. For participants 65 years and older, there may be near-complete capture of Inpatient claims through Medicare; but with data suggesting over 80% of the general Medicare population having part B coverage, it would be important to investigate the proportion of participants in our study with part B coverage to ensure we're capturing events reported via physician encounters. However, with over 21% of persons reporting both Medicare and Medicaid, this could suggest there is a sizable proportion of persons with Medicaid state buy-ins to cover part B carrier benefits.

3. Ascertaining CMS coverage for participants aged < 65 in the proposed study

Table 11. Self-reported CMS categories among participants aged < 65

	n	%	Any CMS Claim ≤90 days, %
Medicaid Only	8910	44.4	81.6
Medicare Only	4852	24.2	78.5
Both	3113	15.5	88.3
None	739	3.7	100
Missing	2463	12.2	100
Total	20077	100	84.8

Among participants < 65, over 84% had a CMS claim within 90 days of enrollment; importantly, even among those who reported no CMS coverage, 100% of them filed a CMS claim probably due to the fact that States allow eligible persons to enroll in Medicaid at the points-of-care when service is needed (and many do in fact enroll at the time care is needed at their coverage status is updated retroactively in MAX). Eligibility for Medicaid does change over time so follow-up data (based on SCCS follow-up interviews) is needed to document the proportion of study participants claiming Medicaid coverage at baseline who still reported Medicaid at follow-up. However, as this preliminary findings suggests, eligible persons could still obtain care via Medicaid (and hence considered as having "continued CMS coverage") when care is needed regardless of self-reported Medicaid coverage or lack thereof. Also, while a participant's enrollment in Medicaid may vary over time, it is likely in the low socioeconomic status population of the SCCS that participants in Medicaid at enrollment will remain in Medicaid throughout the follow up period.

We would consider additional analyses among participants with documented CMS encounters; with over 81% having a Medicaid claim within 90 days of enrollment, we may have higher numbers of persons with documented Medicaid encounters throughout the study.

3.6.3.4 Assessing the validity of International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes for HF

The diagnosis code (ICD-9 428x) [119] algorithms for identification of HF in the proposed study have been previously validated. A review of the detection of heart failure in administrative claims data

included eight studies conducted among Medicare beneficiaries reported positive predictive values (PPVs) between 76% and 99%, with the majority of the studies reporting PPVs over 90% [119]. These codes have also been used with high specificity in a number of studies [120, 121].

Goff et al investigated the validity of the use of ICD 9 codes to identify hospitalizations with clinical evidence of CHF using data from the Corpus Christi Heart Project, a population-based surveillance program set up to investigate the natural history of CHD in non-Hispanic whites and Mexican Americans [119]. Eligible cases (n= 5083, mean age = 60.4, 37% female) were subjects admitted for possible acute myocardial infarction, aortocoronary bypass surgery, percutaneous transluminal coronary angioplasty, and related revascularization procedures. Enrollment took place by monitoring admissions to special care units at the 7 hospitals in the Nueces County, TX. For all potential cases, data on sociodemographic factors, medical history, electrocardiograms, and hospital discharge diagnostic codes were collected.

The clinical documentation of CHF was considered the validation standard and was ascertained based on clinical evidence of acute CHF i.e. a composite variable including a) the presence of physician-diagnosed acute CHF in the medical records and b) radiographic evidence of pulmonary edema [119]. The choice of the clinical diagnosis as the validation standard was based on the premise that the sources of the clinical evidence were the primary data in the medical record; whereas, the discharge diagnoses were applied based on a review of the record by the attending physician and/or other members of the health care team. Thus, the discharge diagnoses are a secondary source of data.

The sensitivity, specificity, PPV and NPV were examined 3 ICD discharge diagnosis code–based classification algorithms: (1) the presence of ICD code 428, (2) the presence of either ICD code 428 or 402, and (3) the presence of any of the ICD codes listed in the table below:

The prevalence of a medical record documented episode of acute CHF was 27.1% when defined either by either criteria. The 2 sources of information agreed for 64.4% of the patients classified as having clinical evidence of an episode of acute CHF, 24.4% were classified on the basis of medical record

notations alone, and 11.2% were classified on the basis of data from the chest radiograph alone; the overall agreement of classification was 90.4% ($\kappa = 0.72$, $P < 0.001$).

The ICD code 428.x for HF, assigned as the primary or a secondary discharge diagnosis, was associated with 62.8% sensitivity, 95.4% specificity, 83.5% positive predictive value, 87.4% negative predictive value, and a 24.8% under-estimation of CHF-related hospitalizations. So while the PPV and NPV were good, there was an underestimation of the number of CHF cases. There were differences in performance across the age spectrum and Goff et al suggested that these performance statistics were likely to differ with those observed in a non-CHD population [119].

Fisher et al examined the accuracy of ICD 9 codes as performed in the 1985 National DRG Validation Study, which carefully re-abstracted and reassigned ICD-9-CM diagnosis and procedure codes from a national sample of 7050 medical records, to determine whether coding accuracy had improved since the Institute of Medicine studies of the 1970s and to assess the contemporaneous coding accuracy of specific diagnoses and procedures [124]. For the identification of CHF among Medicare enrollees, the validation standard were ICD 9 codes applied by accredited medical records technicians who were blinded to the diagnoses assigned by the hospitals. Ergo, this study could be viewed as an assessment of agreement between coders. The sensitivity of any of HF-related ICD codes 402.01, 402.11, 402.91, and 428-428.9, was reported to be 85% (with a PPV of 87%) when the principal diagnosis was examined for the ability to detect the primary reason for admission and 89% when any of the diagnoses were examined for the ability to detect a case of HF identified and treated during the admission; specificity was 99% and 95%, respectively. Thus, greater performance statistics were observed in this Medicare population using a different validation strategy.

In summary, the results of the validation studies suggest that reliance on ICD codes may underestimate the proportion of persons with HF particularly if there is over reliance on a primary claim for HF.

In the proposed study, we would use both primary (i.e. code in the first position) and non-primary (i.e. code in any position) diagnoses of HF to boost our capture of HF cases. However, we do recognize that given there may be some misclassification of cases as non-cases; less so in the reverse direction since most studies have shown very high specificity. This may lead to an under-estimation of our incidence rates given the reduction of the numerator in the incidence rate calculation.

Also, we hypothesize that this misclassification would be non-differential across levels of our predictors of interest (race-sex categories; weight-height index and deprivation index). It is known that the presence (or absence thereof) of bias when we have non-differential misclassification of a binary disease outcome depends on whether we have incomplete ascertainment of outcome (<100% sensitivity) or persons without the outcome are misclassified (< 100%specificity) [125]. If the former, the risk ratio estimate remains unbiased but the absolute magnitude of the risk difference is biased towards the null by a factor equal to the probability of false-negatives. On the other hand if we have imperfect specificity but perfect sensitivity the risk ratio is biased towards the null; the risk difference is biased towards the null by a factor equal to the false-positive probability. With near-perfect specificity reported in most studies and lower sensitivity, there may be little bias in our effect estimates.

3.7 Predictors and Covariates

During recruitment, SCCS participants provided information about demographics (race, sex and date of birth/age), socioeconomic characteristics (annual household income and level of education), personal and family medical history, anthropometric parameters (height, weight and waist circumference), lifestyle factors (tobacco and alcohol use history), marital status and total physical activity.

Participants also provided information on their residential address at the time of interview. The full address history for SCCS participants was then geocoded by a multi-stage process incorporating both batch and interactive processes. The Census 2000 area unit (state, county, census tract and ZIP Code Tabulation Areas – ZCTAs) for the geocoded address was determined by a spatial join to TIGER/Line® Shapefiles using ESRI ArcMap 10.0 software (ESRI, Redlands, CA). Geocoding of SCCS participants'

addresses and linkage to geographic information datasets such as census tract data, allowed development of residence-specific characteristics (and environmental measures using information from external data resources) including an SCCS-derived deprivation index.

Census Tracts are small, relatively permanent statistical subdivisions of a county or equivalent entity that are updated by local participants prior to each decennial census as part of the Census Bureau's Participant Statistical Areas Program [126]. Census tracts generally have a population size between 1200-8000 persons, with an optimum size of 4000. A census tract usually covers a contiguous area; however, the spatial size of census tracts varies widely depending on the density of settlement. Census tract boundaries are delineated with the intention of being maintained over a long time so that statistical comparisons can be made from census to census. Census tracts will be used as proxies for neighborhoods in these analyses.

3.7.1 Weight-Height Index (W/H^n)

We would derive the appropriate weight-height index for the current cohort based on the coefficient of the log weight and log height variables in a bivariate Cox model for the log hazard of HF. As previously mentioned, SCCS participants reported their height and weight at baseline. These were validated using data from a random sample ($n \approx 14{,}000$) of SCCS participants for whom measured weight and height were either a) abstracted from contemporaneous CHC medical records or b) obtained via measurements performed by trained interviewers using a SECA 703 digital scale and a stadiometer on the day of the interview [107]. There was a very high correlation ($r > 0.95$) between measured and self-reported weight and height [18].

3.7.2 Deprivation Index

The SCCS-derived deprivation index is a clustering of social and economic indicators which reflect neighborhood deprivation and that have been linked to adverse health outcomes. It was constructed using principal components analysis.

Principal components is a data reduction technique used to create orthogonal (uncorrelated) variables – from a group of possibly correlated predictors – that best explain the variation in the predictors or "x-space". These orthogonal variables or eigenvectors are linear combinations of the original variables; they described how variables "contribute" to each factor axis and the eigenvalues of each orthogonal variable represents how much of the variance of the "X-space" is predicted by the variable in question. The variables are often scaled and centered before the computation of the principal components. The first principal component, PC1, is the linear combination of the standardized variables having maximum variance [127]. The second principal component (PC2) is the linear combination of predictors having the second largest variance such that PC2 is orthogonal to PC1. For a total of p predictors, the first k PCs (where $k < p$), will explain only part of the variance in the whole system of p predictors or the "x-space" unless one or more of the original variables is exactly a linear combination of the remaining variables [127].

To construct the PCs for neighborhood-deprivation index in SCCS, 11 census tract-level variables representing 4 main dimensions were considered:

a) Social indicators – percentage of housing units with ≥ 1 occupant per room and percent female headed households with dependent children.
b) Wealth and income – percentage of households with income < $30,000 per year, percentage of households with public assistance income, percentage of households with no car and median household value, percentage of occupied housing units with renter/owner's costs > 50% of income and percentage of persons with income below the 1999 poverty status.
c) Education – percentage of persons aged ≥ 25 that did not graduate high school
d) Occupation – percentage of males and females who are unemployed and percentage males in professional occupations.

Only the first principal component was retained for the construction of the deprivation index given it explained most of the variability in the component measures [128].

3.7.3 Socioeconomic Status

a. Education: SCCS participants were placed in 8 categories according to their highest level of education attained ranging from less than high school to graduate-level degrees.

b. Income: Participants reported the range of their total household income for the year prior to enrollment and were placed in 5 categories ranging from less than $15 to over $100,000 or more.

3.7.4 Covariates

Participants reported history of tobacco smoking as never, former and current and also in terms of number of cigarettes per day and alcohol use in number of drinks per day. The presence of traditional cardiovascular risk factors at baseline was based on a self-reported history of physician-diagnosed hypertension, diabetes mellitus, high cholesterol, as well as self-reported use of medications for hypertension, diabetes mellitus, or high cholesterol. History of chronic disease occurrence for self-reported conditions was confirmed via confirmation medical records for a small random sample of SCCS participants. Total physical activity was measured as total metabolic equivalent-hours per day spent doing light, moderate and strenuous occupational/household work and moderate and vigorous exercise.

3.7.4.1 Challenges related to covariate measurements

3.7.4.1.1 Determination of covariates at baseline

Demographic, anthropometric, and cardiovascular risk factors will be determined by self-report of a physician diagnosis and use of medications (for traditional cardiovascular risk factors). While self-report may be susceptible to recall and misclassification bias, these methods have been successfully used and validated in large epidemiologic cohorts. Many of the questions on the SCCS questionnaire were adapted from questionnaires used and validated in other settings, and a series of independent validation studies using biomarkers, repeat interviews or medical records have demonstrated the reliability of the questionnaire within the SCCS population for variables such as tobacco use status, self-reported diseases including diabetes, height and weight.

3.7.4.1.2 Handling changes in covariates over time

Another major challenge we're faced with is the temporal variations in the values of certain covariates during the follow-up period particularly smoking status, diabetes status and body weight. Ignoring these secular variations in the values of these key covariates could be a potential source of bias of our effect estimates and the direction of bias can be difficult to predict. One option would be to model these variables as time-varying covariates (TVCs) in our multivariable Cox Models. There is follow-up data for close to 60% of the SCCS cohort. If this proportion is the same for our current sub-cohort (n = 27078, identified based on our inclusion criteria for the current analyses) we could perform time-dependent analyses for close to 16,000 persons.

However this poses another conundrum; namely: how do we guard against any potential selection bias if those with covariate data for these key time-varying covariates differ meaningfully from those without (n ~ 10,000) on other important covariates such demographics and other lifestyle and clinical factors? Also, given our goals of investigating differences in measures of frequency (incidence rates) across groups defined by race and sex as well as examining potential departures from additivity of effects by race and sex, an analyses restricted to this smaller sample may lack sufficient power. Striking an optimum bias-precision trade-off in this approach would be crucial. One way of boosting the power and precision of these supplemental analyses using TVCs would be to perform multiple imputation of covariate values for study participants without covariate data at follow-up, if the data can be considered to be "missing at random". By utilizing multiple imputation in an attempt to "preserve" information we would simultaneously reduce bias in estimates and provide estimates of the variance-covariance estimates of beta-hat penalized for imputation [127].

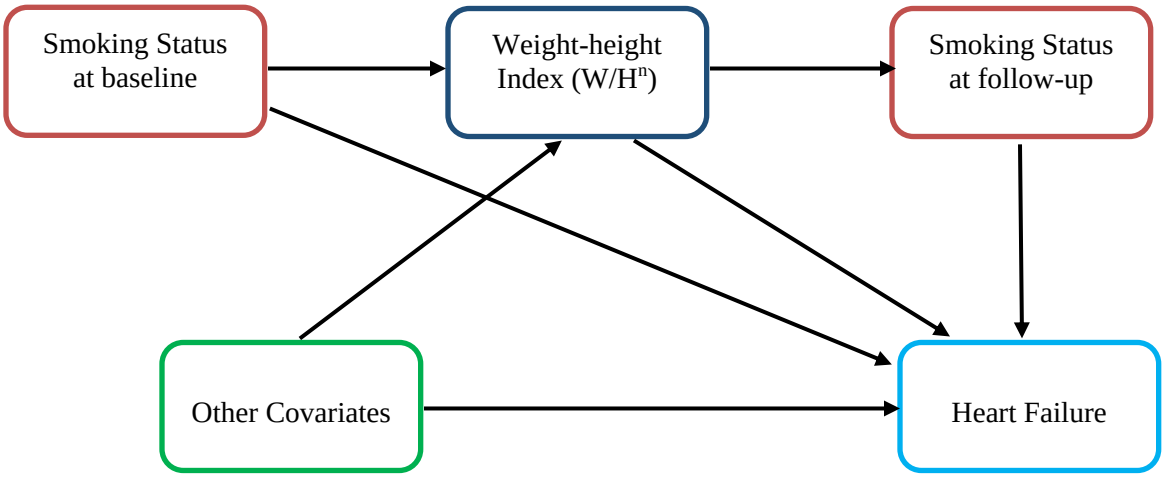
Another important consideration in our supplemental analysis would be the selection of the functional form for modelling smoking status in relation to post-HF survival. As seen with previous studies between smoking status and survival different options e.g. using a step function for "current smoking" that is updated at every time point (1 if yes, 0 if no) or using a time-lagged variable could yield

different results [129]. In the Coronary Artery Surgery Study (CASS) in which patients with mild angina were randomized to medical treatment or CABG, data on smoking status were collected every 6 months, and, for the first analysis, a step function (with 6 months intervals) was used [130]. Much to the surprise of the investigators, whilst not statistically significant, the estimated effect of current smoking on survival was positive i.e. protective. A closer look at the patient's smoking histories revealed that most patients who died were in fact smokers but many had quit smoking at the last follow-up before their death. In several instances this was apparently explainable by hospitalization for a myocardial infarction or congestive heart failure and other compelling health reasons that prompted smoking cessation at last follow-up. So if smokers with a high risk of death quit just before dying we may have ended up with a scenario similar to confounding by indication. By using time-lagged covariates or percentage of time during follow-up for which the subject smoked the investigators found an inverse relationship (as expected) between smoking and survival. In order to adequately model the confounding effect of smoking in our proposed supplemental analyses, similar approaches would have to be employed.

Similarly, study participants could likely experience weight gains (or loss) during follow-up suggesting it may be appropriate to model weight-height index using TVCs vis-a-vis its relationship with post-HF mortality. However, persons who are subsequently diagnosed with HF may experience some unintentional weight loss before HF diagnosis (and lower BMI has been found to be associated with lower post-HF survival – the "obesity paradox"). Hence, would it be appropriate to utilize their weight measurements during follow-up that may be taken just before HF diagnosis? On the other hand if the baseline measure is used, would that be a true measure of the exposure experience for persons whose HF is diagnosed several years later after enrollment? It may be argued that both approaches have trade-offs that have to be weighed-up against each other.

Also we would have to give careful thought to our working hypothesis of the relationships between our time varying covariates as well as their relationship with HF incidence. For example let's examine the hypothesized relationship between smoking, body weight (or weight-height index) and heart

failure risk. First, at baseline there could be an association between smoking status and weight-height index, based on a priori epidemiologic evidence suggesting weight loss among smokers. Second, persons who have lost weight while smoking may subsequently decide to either a) smoke even more in order to foster more weight loss or b) quit; having achieved their goal. The association could be conceived as shown below:



If these assumptions of the associations between these variables are true, then smoking status at baseline could be considered as a confounder of the association between W/H^n and HF risk; while smoking status at follow-up would be considered a mediator hence the term "time-varying confounder-mediator" in some statistics literature. Adjusting for only the baseline or follow-up values of cigarette smoking will lead to biased estimates. One option would be to use marginal structural models to estimate the effect of weight-height index on HF risk by appropriate control for the effects of time-dependent confounders. In these models, the predictor – outcome association is estimated in a regression model that is weighted using the inverse probability of treatment weights [131].

3.7.4.1.3 Residual Confounding and Unmeasured Confounders

The unique design of the SCCS cohort with participants who are of similar (and generally low) SES regardless of race or sex, as well as the well-documented nature of the information on socioeconomic

characteristics that can be additionally controlled for statistically, would mitigate potential confounding by socioeconomic factors of differences in HF risk and post-HF survival by race and sex.

However there is always the possibility of residual confounding due to unmeasured factors. One such unmeasured factor is the treatment for HF; the latter not be available via linkage with the Medicare part A/B carrier and outpatient base files. As an important predictor of heart failure survival, if there are unobserved differential patterns in HF treatment between groups defined by race and sex, these could confound our findings on the patterns of post-HF survival. Importantly, even if HF therapy didn't differ across levels of any of our predictors of interest, being a strong predictor of the outcome, failing to adjust for it moves our hazard ratios towards the null.

3.8 Statistical analyses

3.8.1 Data Management and handling of "missingness"

Data management and quality control for the SCCS data is performed at the International Epidemiology Institute (IEI) which provides de-identified datasets to investigators upon approval of proposal applications.

Before performing substantive data analyses we will screen all variables for abnormal or inconsistent values (e.g. a male subject with an entry for menopausal status) and outliers. Implausible values would be reported to the SCCS data management team at the International Epidemiology Institute for appropriate checks and if unresolved, the observation would be considered for case-wise deletion given that other measurements for that participant may have been incorrectly entered as well.

Proportion of missingness for each variable would be computed and reasons for missingness will be examined (e.g. missing value for number of cigarettes smoked per day because participant was a non-smoker). Logistic models would be used to predict the probability of missingness for each predictor variable given the other covariables and the outcome. Patterns of simultaneously missing variables would be described; we would perform cluster analyses of the missing value status of all variables. Complete

cases would be examined in detail to investigate patterns in the probability of observations being a complete vs non-complete case.

Assuming data are missing at random we will perform multiple imputation for the missing values utilizing the *aregImpute* algorithm in the Hmisc package in R. The aregImpute algorithm takes all aspects of uncertainty in the imputations into account by using the bootstrap to approximate the process of drawing predicted values from a full Bayesian predictive distribution [132]. Different bootstrap resamples are used for each of the multiple imputations. A flexible additive model, is fitted on a sample with replacement from the original data and this model is used to predict all of the original missing and non-missing values for the target variable, then the imputation models are run. By default, linearity is assumed for target variables (variables being imputed) and *nk* = 3 knots are assumed for continuous predictors transformed using restricted cubic splines. *AregImpute* uses predictive mean matching with optional weighted probability sampling of donors rather than using only the closest match [132]. Predictive mean matching works for binary, categorical, and continuous variables without the need for iterative maximum likelihood fitting for categorical variables, and without the need for computing residuals or for curtailing imputed values to be in the range of actual data.

The missing data will be filled-in using these simulations a number of times to create that same number of complete datasets. For each completed dataset, the regression model will be fitted and the regression coefficients will be averaged over multiple imputations.

3.8.2 General analytic approach

Descriptive statistics (means and standard deviations for continuous variables and counts and percentages for categorical variables) would be computed for HF cases and non-cases.

To investigate the incidence of HF, duration of follow-up would be computed from date of entry into the SCCS until the date of the first diagnosis of HF, date of death, or December 31, 2010, whichever occurred first. Incidence rates (IR) of heart failure would be calculated for white women, black women,

white men and black men by dividing the number of HF cases by person-time of follow-up, and the rates would be presented per 1,000 person-years.

For analyses of post-HF survival among those with a diagnosis of incident HF, follow-up time was defined as time from HF diagnosis to death or December 31st 2010 whichever occurred first. Kaplan-Meier survival curves would be plotted by race and sex.

In the multivariable models for both outcomes (HF incidence and post-HF survival), we would include indicator variables for white men, black women and black men, with white women as the reference group. The covariates would include: BMI, history of diabetes, hypertension, high cholesterol, MI/CABG or stroke, household income, education, cigarette smoking, alcohol intake, marital status and enrollment source (community health centers vs general population). Tests for interaction of anthropometric surrogates (W/H^n) and deprivation index with sex and race will be conducted by adding the corresponding cross-product terms (including the linear and non-linear spline terms) to the models.

3.8.3 Selection of Candidate Confounders and spending degrees of freedom for covariates

3.8.3.1. General Approach to selecting confounders for the multivariable models.

1. Directed acyclic graphs.

For each aim we would first develop a conceptual model of the hypothesized relationships between the main predictor of interest (race, sex, weight-height index and deprivation index), the outcome (heart failure incidence or post-HF death) and the covariates that are potentially associated with both the exposure and the outcome based on prior epidemiologic evidence.

Second we would develop directed acyclic graphs for each association of interest in other to ascertain the full model required to properly adjust for confounding effects without opening backdoor paths by adjusting for colliders.

However, in order to avoid compromising precision, we would choose a minimum sufficient set of candidate confounders that minimizes bias and maximizes precision hence optimizing the bias-precision trade-off. Once the minimum sufficient set is fixed models would not be further simplified in order to avoid problems related to "phantom" degrees of freedom or potentially introducing bias into the models resulting from residual confounding or "incomplete conditioning". Also we would choose the number of degrees of freedom to be spent on each variable based on its predictive promise in order to maximize precision.

Colliders would not be adjusted for in the models (unless we can close all open-paths which are opened by adjusting for these) and variables that are considered to be mediators (diabetes, dyslipidemia and hypertension, past history of myocardial infarction and stroke) would only be included subsequently during "mediation" analyses.

We would use this approach for all 4 aims as this ensures the best bias-precision trade-off. Other options are discussed below which we would not consider as each has shortcomings related to residual confounding and a less than optimal bias-precision trade-off.

2. Including covariates in the Full Model if they are found to be both associated with the exposure in the total population and the outcome in the referent group of the predictor of interest (the "unexposed").

Hypothetically, for this approach, first any covariate that has a significant association with the predictor of interest (which varies from aims 1 through 4) and is also associated with the outcome (HF or post-HF death) would be included in the full model. Second, variable removal would be performed by backward elimination based on the variable which results in the least change in log hazard ratio of the exposure of interest when left out of the full model. The model is refitted and the variable with least change in log hazard ratio is removed again until all remaining variables result in a change of more than 10% of the log hazard ratio. This method is not appropriate as it is likely to leave out some important

(albeit weaker) confounders; produce biased estimates of effect; problems related to phantom degrees of freedom subsists and the method is unlikely reproducible making results incomparable across studies.

3. Selecting covariates for inclusion in the full model based on whether they result in a 10% change in estimate (log hazard ratio) when added to a model containing only the predictor of interest.

3.8.3.2 Covariates for the association between race (and sex) and heart failure incidence (and post-HF survival).

Based on the hypothesized relationships between the variables we would use the following approach for adjusting for important covariates when examining the association between race, sex and HF incidence (and post-HF survival):

Three models would be constructed, with white women as the referent category: model 1 would include indicator variables for white men, black women and black men and age (restricted cubic splines with 4 knots); model 2 would be additionally adjusted for body mass index (restricted cubic splines with 4 knots), and history of diabetes, hypertension, high cholesterol, MI/CABG or stroke (all yes/no); model 3 would be additionally adjusted for the following covariates: household income (<$15000, $15000-$24999 & ≥$25000); education (< high school, high school/vocational training/junior college, college degree or higher), smoking status (4 categories: never/former/current <19.5 pack-years/current ≥ 19.5 pack-years), alcohol intake (linear), total physical activity in MET-hours (linear & quadratic terms), marital status (married/living as married with partner, separated/divorced, widowed, single/never married) and enrollment source (community health centers vs general population). The knots for the splines would be equally spaced based on Harrell's recommended percentile distribution.

3.8.3.3 Covariates for the association between neighborhood deprivation and HF incidence

Based on the conceptual model in figure 2 we would use the following approach for adjusting for important covariates when examining the association neighborhood deprivation index and HF incidence

Four models would be constructed, model 1 would include deprivation index (1df), age (restricted cubic splines with 4 knots) race and sex and interaction terms between deprivation index and race and sex. Model 2 would be additionally adjusted for household income (<$15000, $15000-$24999 & ≥$25000); education (< high school, high school/vocational training/junior college, college degree or higher). Model 3 would be additionally adjusted for body mass index (restricted cubic splines with 4 knots), smoking status (4 categories: never/former/current <19.5 pack-years/current ≥ 19.5 pack-years), alcohol intake (linear), total physical activity in MET-hours (linear & quadratic terms) and history of diabetes, hypertension, high cholesterol, MI/CABG or stroke (all yes/no).

3.8.3.4 Covariates for the association between weight-height index and heart failure incidence (and post-HF survival)

Based on the DAG in figure 1 we would use the following approach for adjusting for important covariates when examining the association weight-height index and HF incidence/post-HF survival.

Three models would be constructed; model 1 would include weight-index (restricted cubic splines with 5 knots), age (restricted cubic splines with 4 knots), race and sex. Model 2 would be additionally adjusted for household income (<$15000, $15000-$24999 & ≥$25000); education (< high school, high school/vocational training/junior college, college degree or higher), smoking (4 categories: never/former/current <19.5 pack-years/current ≥ 19.5 pack-years) and alcohol intake (linear) and total physical activity in MET-hours (linear & quadratic terms). Model 3 would include model 2 variables + history of diabetes, hypertension, high cholesterol, MI/CABG or stroke (all yes/no); in order to investigate the effect of mediators. Interaction terms to test for additivity of effects by race and sex would be included.

3.8.3.5. Rationale for the number of degrees of freedom spent on certain covariates

<u>Annual household Income</u>

Over 68.6% of study participants in this SCCS sub-cohort (n= 27,078) have annual household income <$15,000, 17.8% have income between $15,000-24,999, 12.1% have income ≥ $25,000 and 1.5% have missing income data. Thus, using more than 3 categories for income, may raise statistical concerns related to sparse cells and model instability in the multivariable models. Income would therefore be modelled using 3 categories <$15,000, $15,000-$24,999 and ≥$25,000. However, we acknowledge that there may be some residual confounding and potential bias in our effect estimates.

Level of Education

Education would be modelled using 3 categories defined as: < High school, High school/Vocational training/Junior college and College degree or higher. For this variable as well, we are making decisions based on a trade-off between spending degrees of freedom to adequately control for confounding and avoiding sparse cells for the higher educational levels. These categorization of the income and education variables would be used for all multivariable models in aims 1 to 4

Alternatively, we could choose our categories based on the functional form of the variable that best suits the "dose-response" relationship (using all 7 categories) with outcome (HF). The latter approach has obvious shortcomings related to decreased precision and data sparseness as earlier mentioned.

Cigarette Smoking

In light of the fact that a) we're principally interested in the smoking variable as a confounder rather than as a main predictor and b) the literature suggests that the effect of smoking on the occurrence of coronary heart disease (CHD) is much more preponderant than the progression towards HF (and hence it could be considered a "weak" confounder of the obesity-HF link), we would consider approaches which spend the least degrees of freedom in our multivariable models.

In all models, smoking would be modelled using 4 categories – never/former/current <19.5 pack-years/current ≥ 19.5 pack-years; 19.5 the median pack-years among current smokers.

Alternatively, we could consider the following approaches:

a) 3 categories for smoking status: never, former and current.
b) Using these 3 categories and further splitting both former and current categories at specific cut-points (e.g. the median) of pack-years (and the number of cigarettes smoked per day).
c) A more parsimonious approach with 2 categories (never vs ever) + 1 continuous variable for cigarette-years (or pack years) [133]. By using this approach, the effect of pack-years would be estimated by comparing only subjects who had the same value for smoking status, that is, only smokers. In addition, one would center pack-years by subtracting the mean pack-years value from the original value for all smokers, while keeping 0 for never smokers. Such a linear transformation of pack-years does not change its estimated effect, but it allows the effect of ever smoking to compare average smokers with never smokers, since both groups are assigned a value of 0 for centered cigarette-years [133]. Without this transformation of cigarette-years, the estimated hazard ratio for ever smoking would be more difficult to interpret, as it would compare never smokers and hypothetical smokers with 0 cigarette years. Thus, the model using both variables provides interpretable estimates of both the qualitative effect of smoking status and the quantitative effects of smoking exposure.

Sensitivity analyses would be performed to observe any changes in parameter estimates with varying functional forms of the smoking variable.

For all other covariates, in the final models, the degrees of freedom to be spent on each variable would be modified depending on the predictive "promise" of each variable based on the Spearman rank correlations (equivalent to the Kruskall Wallis test for categorical variables) between each covariate and the outcome without performing any hypothesis tests.

3.8.4 Substantive analyses for Aim 1

To investigate differences in a) the incidence of HF and b) post-HF survival between groups defined by race and sex; more specifically: white women, black women, white men and black men.

To investigate the incidence of HF, duration of follow-up would be computed from date of entry into the SCCS until the date of the first diagnosis of HF, date of death, or December 31, 2010, whichever occurred first. Incidence rates (IR) of heart failure were calculated for white women, black women, white men and black men by dividing the number of HF cases by person-time of follow-up, and the rates are presented per 1,000 person-years. The 95% confidence intervals (CI) would be calculated using the quadratic approximation to the Poisson log likelihood for the log-rate parameter [134]. To account for age differences between the demographic categories, age-standardized rates would be computed using the age distribution of the SCCS participants.

Multivariable Cox models would be utilized to test whether differences in crude IRs between categories defined by race and sex persisted after adjustment for baseline covariates. Three models would be constructed, with white women as the referent category: model 1 would include indicator variables for white men, black women and black men and age (restricted cubic splines with 4 knots). The covariates for models 2 and 3 and the cumulative number degrees of freedom spent for each model are tabulated below:

	Covariates	**Functional form**	**Degrees of**
Model 1	Race and Sex	4 categories - white women (ref), white men, black women & black men.	3
	Age	Restricted cubic splines (nk = 4)	3
Model 2	Model 1 variables		
	BMI	Restricted cubic splines (nk = 4)	3
	History of diabetes, hypertension, high cholesterol, MI/CABG, stroke	All (yes/no)	5
Model 3	Model 2 variables		
	Income	(<$15000, $15000-24999 & ≥$25000)	2
	Education	< HS, HS/vocational training/junior college, college degree or higher)	2
	Smoking Status	never/former/current <19.5 pack-years/current ≥ 19.5 pack-years	3
	Alcohol intake	Linear	1
	Physical activity in met-hrs	Linear + quadratic	2
Total			24

The knots for the splines would be equally spaced based on Harrell's recommended percentile distribution [127].

For analyses of post-HF survival among those with a diagnosis of incident HF (n = 4341), follow-up time would be defined as time from HF diagnosis to death or December 31st 2010 whichever occurred first. When date of death is coincident with date of HF diagnosis, follow-up time was set to 0.5 days. We computed cumulative mortality for both HF cases and non-cases using contingency tables. Kaplan-Meier survival curves would be plotted by race and sex. Cox models would be used to investigate differences in cumulative hazard for death (all-cause mortality) using white women as the referent group. Model 1 would comprise indicator variables for white men, black women and black men and age (restricted cubic splines with 4 knots). Variables included in models 2 and 3 would be the same as described previously. P-values for race-by-sex interaction would be computed in models for HF incidence and post-HF survival.

3.8.5. Substantive analyses for Aim 2

To investigate whether neighborhood characteristics (defined by a composite deprivation index) predict the risk of a) incident heart failure b) post-HF survival in SCCS beyond individual-level socioeconomic status (defined by household income and highest level of education attained).

In our data, individuals (level-1 units) are nested within census tracts (level-2 units) and the social and economic characteristics of the latter were used to assess neighborhood deprivation. Several options are available for modelling this data; each having important trade-offs.

3.8.5.1 Multilevel Modelling (MLM)

3.8.5.1.1 Multilevel Cox proportional hazards model

Given the nested structure of the data, a multilevel Cox proportional hazards model [135] could be considered appropriate to model the association between deprivation index measured at the census track level and log hazard of incident HF.

Using MLM, we will test for the effects of neighborhood deprivation index (the level-2 predictor) on the risk of incident HF after adjustment for individual-level factors (level-1 predictors) including household income (and education) and demographics (age, gender and race).

The reduced-form of the multi-level model is as follows [135, 136]:

$$h_{ij}(t) = \exp\left[\gamma_{00}(t) + \sum_{i=1}^{k} \gamma_{0k} W_{kj} + \sum_{i=1}^{p} \gamma_{p} X_{pij} + U_{0j} + R_{ij}\right] h_0(t),$$

Where $h_{ij}(t)$ represents the hazard function, $\gamma_{00}(t)$ represents the intercept, and $h_0(t)$ represents the baseline hazard function whose distribution is unspecified. Time t is defined as the number of days from the participant's entry into the study to the first diagnosis of incident heart failure. W_{kj} represents the neighborhood variable, level-2 predictor (deprivation index), and X_{ij} represents the set of individual-level variables, level-1 predictors (age, gender, race, income and education). U_{0j} represents the intercept random effect, and R_{ij} represents the individual residual. The subscript j represents our level-2 units – the census tracts and the subscript i represents individuals.

The intra-class correlation coefficient would be computed (using the latent variable approach) and used to estimate the proportion of variance explained by differences at the census-tract level. This approach uses the closed form solution of the ICC in the multi-level logistic model to make an approximate estimation of the ICC in the multi-level Cox Proportional hazards model.

$$ICC = \frac{\tau_{00}}{\tau_{00} + \pi^2/3}, \text{ where } \tau_{00} = \text{group-level variance [137].}$$

Neighborhood deprivation index would be modeled as a continuous variable with 1df to preserve parsimony. In separate unadjusted multi-level models, we would also use restricted cubic splines (with 4 knots) and quartiles to explore the functional form of the association between deprivation index and incident heart failure.

In the minimally adjusted models we will estimate the age-, race- and sex-adjusted hazard ratios between neighborhood deprivation and incident heart failure. We would also investigate cross-level interactions between deprivation index and race and sex. Individual-level socioeconomic characteristics (income and education) will then be added.

All multilevel analyses will be performed using Multilevel modeling for Windows (MLwiN) version 1.10.0007, using the macro for the survival models [138, 139]. The first-order marginal quasi-likelihood (MQL) estimation procedure would be used to obtain preliminary estimates then a predictive quasi-likelihood approach (PQL) (combined with a second-order Taylor expansion series) would be utilized to obtain more accurate estimates [136]. Effect estimates would be presented as hazard ratios with 95% confidence intervals.

3.8.5.1.2 Other MLM Approaches

We would consider alternative approaches to fit the baseline hazard and examine differences in the estimates of our coefficients based on certain distributional assumptions.

First, we will consider fitting a multilevel Poisson regression model that assumes a piece-wise constant function for the distribution for the baseline hazard using STATA (version 13, Stata Corp, College Station, Texas, USA) .

Second, we would consider the “stgenreg” package by Crowther et al to build a multilevel exponential proportional hazards model [140]. For both models we would estimate the hazard ratios for deprivation index adjusting for individual level factors.

The advantage of the Poisson and parametric models is that they are more parsimonious, we obtain smooth hazard functions that can be estimated at any point and if our distributional assumptions are correct we may obtain more accurate parameter estimates. In recent years the incidence rate of heart failure in the US has been relatively constant with improved survival being the major driver of the

increased prevalence; so an exponential distribution (which assumes constant hazard throughout the study period) for the baseline hazard may not be inappropriate.

3.8.5.1.3 Pitfalls of the MLM Approaches

While MLM approaches could be an acceptable fit given the hierarchical structure of our data, there are some limitations to their use in light of certain peculiarities of our data.

First, based on our preliminary analyses, there are 4666 census tracts in the data with the number of persons per census track varying between 1 and 21+. Ergo, there are many census tracks or level-2 units with singleton data points and worse some with zero cases of HF. This may lead to a few problems including:

a) In non-linear models, simulations done using multi-level logistic models have suggested that in cases of unbalanced data with a very small group size ($\leqslant 2$), the group level variance components are over estimated by over 30%, with an upwards bias that is most accentuated when dealing with unbalanced data [141].
b) In addition, the fixed effect coefficients are biased up by as much as 16% [141].
c) There may be some issues related to stability of parameter estimates and convergence of the models especially in the light of singleton data points with zero cases in some census tracts.

In addition, based on the evidence from previous studies investigating the contribution of neighborhood effects on other HF and CVD outcomes [137], we do not anticipate a high ICC (> 0.5), hence having some very small clusters is not the ideal scenario for considering multi-level modeling.

MLM would be considered as 1 of the options for sensitivity analyses as this would allow an assessment of the robustness of our findings while using an approach with larger variances and more random components.

We will consider other modelling options that take into account the correlation of the data of the individuals nested within census tracks. However, we would still utilize MLM to compute the ICC as it is

provides information – the proportion of the variance that is explained by differences across neighborhoods – not captured by other methods.

3.8.5.2 Cox Proportional hazards model adjusting for Non-Independence using Huber-White sandwich Estimators.

Given the non-independence of the data points within each census track, and the limitations of a multilevel modelling approach in this setting, our primary statistical analysis approach for aim 2 will be a Cox proportional hazards model that takes into account non-independence using the Huber-White cluster Sandwich estimator of variance, H_c [127] whose general formula for linear and nonlinear models is given below:

$$H_c = I^{-1}(b)[\sum_{i=1}^{c}\{(\sum_{j=1}^{n_i} U_{ij})(\sum_{j=1}^{n_i} U_{ij})'\}]I^{-1}(b),$$

Where I is the information criteria i.e. the second derivative of the log likelihood, log L [127]:

$$\begin{aligned} I(P) &= E\{-\partial^2 \log L/\partial P^2\} \\ &= E\{s/P^2 + (n-s)/(1-P)^2\} \end{aligned}$$

and U is the score statistic i.e. the first derivative of log L [127]

$$U(P) = \partial \log L/\partial P = s/P - (n-s)/(1-P)$$

Log L is computed under the null hypothesis, Ho: equal sample proportions, P_1 (s_1/n_1) = P2 (s_2/n_2) = P as:

$$\log L = s\log(P) + (n-s)\log(1-P)$$

In the specific case of the Cox Model, the cluster sandwich estimator, H_c uses special score residuals for U (the score vector) given there are no per-observation score contributions [127, 142].

We would fit the following models successively:

	Covariates	Functional form	Degrees of freedom
Model 1	Deprivation index	Restricted cubic splines (nk = 4)	3
	Age	Restricted cubic splines (nk = 4)	3
	Race and sex	Race (whites vs blacks); Sex	2
Model 2	Model 1 variables		
	Income	(<$15000, $15000-24999 & ≥$25000)	2
	Education	< HS, HS/vocational training/junior college, college degree or higher)	2
Model 3	Model 2 variables		
	Smoking Status	never/former/current <19.5 pack-years/current ≥ 19.5 pack-years	3
	Alcohol intake	Linear	1
	BMI	Restricted cubic splines (nk = 4)	3
	History of diabetes, hypertension, high cholesterol, MI/CABG, stroke	All yes/no	5
	Physical activity in	Linear + quadratic	2
Total			26

We would equally present the ICC (computed using the latent variable approach) as it is provides information about the proportion of the variance that is explained by differences across neighborhoods which has a greater public health relevance. Hazard ratios for 1 interquartile range increase in deprivation index would equally be presented as these have an intuitive interpretation: i.e. the hazard of the event occurring for a typical person in the middle of the upper half of the distribution to the hazard of the event for a typical person in the middle of the lower half of the distribution.

3.8.6 Substantive analyses for Aim 3

For aim 3, we would first derive the appropriate weight-height index (W/H^n) for the current cohort based on the coefficient of the log weight and log height variables in a bivariate Cox model for the log hazard of HF.

Hypothetically, by regressing a dependent variable, y on the natural logarithm of (W/H^n) i.e. $\log_e W/H^n$ we get the following (natural logs are implied throughout so the e is dropped):

$$y = c + \log W/H^n, \text{ assuming a slope of 1.}$$

$$y = c + \log W + \log H^{-n}$$

$$y = c + \log W - \text{nlog} H$$

So the absolute value of the ratio of the coefficients of log W and log H is n.

In the case specific case of the Cox Model for the log hazard of HF (with no intercept), when log hazard of HF is regressed on log W and log H we get

$$y = \alpha_1 \log W + \alpha_2 \log H$$

Where y = log hazard (HF), α_2 is negative if y and W/H^n are positively correlated and vice versa.

And n would be given by $|\alpha_2/\alpha_1|$

Second, in separate Cox models, we would regress the restricted cubic splines of the natural log of the data-derived weight-height index and that of BMI on the log hazard of HF. Then, model fit statistics (LR chi square, χ^2 and AIC) would be used to compare the performance of the data-derived weight-index versus that of BMI in relation to a model utilizing restricted cubic splines of log weight and log height.

Third, we would run models with log BMI and log height to see if log height is still significant in a model containing BMI. We would also compare the effect size for a 1 interquartile range increase in BMI and the computed W/H^n index.

Fourth, we would use multivariable Cox models which take into account nonlinearity and non-additivity to model a flexible dose-response association between the better performing weight-height index (W/H^n) (modelled using restricted cubic splines with 5 evenly spaced knots) and HF risk adjusting for relevant covariates in a sequential fashion as shown in the table below:

	Covariates	Functional form	Degrees of
Model 1	Weight-height index	Restricted cubic splines (nk = 5)	4
	Age	Restricted cubic splines (nk = 4)	3
	Race and sex	Race (whites vs blacks); Sex (women vs Women)	2
	W/H^n×race terms	Linear + non-linear interaction terms.	4
Model 2	Model 1 variables		
	Income	(<$15000, $15000-24999 & ≥$25000)	2
	Education	< HS, HS/vocational training/junior college, college degree or higher)	2
	Smoking Status	never/former/current <19.5 pack-years/current ≥ 19.5 pack-years	3
	Alcohol intake	Linear	1
	Physical activity in MET-hrs	Linear + quadratic	2
Model 3	Model 3 variables		
	History of diabetes, hypertension, high cholesterol, MI/CABG, stroke	All yes/no	5
Total			28

Interactions between W/H^n and race as well as sex would be tested. Fourth, we would repeat the multivariable models for the relationship between waist circumference and HF. We would present plots of predicted probabilities (or HRs) of incident heart failure versus weight-height index stratified by race and/or sex. These analyses would be repeated using waist circumference.

For the analyses for WC, we have data for 3304 participants and there are 251 cases observed among these participants. Using the rule of thumb of 10-15 cases per df (or parameter to be estimated) that leaves us with 17-25 degrees of freedom allowed in our model. We would reduce the df spent on some less important covariates (based on prior literature). Below is the proposed df to be spent in the multivariable cox model for WC. Formal power calculations for WC are presented in section **4.9**.

	Covariates	Functional form	Degrees of freedom
Model 1	WC	Restricted cubic splines (nk = 4)	3
	Age	Restricted cubic splines (nk = 4)	3
	Race and sex	Race (whites vs blacks); Sex	2
Model 2	Model 1 variables		
	Income	(<$15000, ≥$15000)	1
	Education	< HS, HS/vocational training/junior college, college degree or higher)	2
	Smoking Status	Never, former and current	2
	Alcohol intake	Linear	1
	Physical activity in MET-hrs	Linear + quadratic	2
Model 4	Model 3 variables		
	History of diabetes, hypertension, high cholesterol, MI/CABG, stroke	All yes/no	5
Total			21

For the association between (W/H^n) and post-HF survival we would perform similar analyses as we did for the association with HF risk. Hazard ratios for 1 interquartile range increase in W/H^n, WC and BMI would equally be presented as these have an intuitive interpretation: i.e. the hazard of the event occurring for a typical person in the middle of the upper half of the distribution to the hazard of the event for a typical person in the middle of the lower half of the distribution.

For all our models we would verify the PHM assumption by utilizing Schoenfeld residuals from the Cox Models and log (-log) plots. Martingale residuals and dfbetas would be used to investigate the functional form of predictor variables and influential observations respectively.

3.9 Power Calculations

Based on our preliminary findings, the number of incident HF cases (identified in Medicare between March, 2002 and September, 2010) among SCCS participants in our sub cohort (N= 27078) are 801, 511, 1940 and 1089 among white women (n =5252), white men (n=3202), black women (n=11688) and black men (n = 6936) respectively making a total of 4341 incident HF cases.

3.9.1 Aim 1

We plan to investigate survival and mortality risk among SCCS participants diagnosed with incident HF by race and sex. There are close to 952 deaths amongst the 4341 incident HF cases.

3.9.1.1 Aim 1a

Using white women as the referent group, we can compute the power to investigate differences in log hazard for HF for white men, black men and black women at an alpha of 0.05 as follows:

a. White women vs White men

Cumulative Incidence of CHF among whites = 0.155. Proportion of women among whites = 0.62. With $P_1 = 0.62$, SD $(P_1) = \sqrt{0.62*0.38} = 0.49$. With number of events = 1312 and assuming an R-square of 0.20 for the correlation (to get conservative estimates of power; given that pseudo r-square from logistic regression was 0.10) with candidate covariates we get 97% power to detect a HR of 1.3

Power	Events	HR	SD	Alpha*	R^2
0.77	1312	1.2	0.49	0.05	0.20
0.97	1312	1.3	0.49	0.05	0.20
0.99	1312	1.4	0.49	0.05	0.20
0.99	1312	1.5	0.49	0.05	0.20
0.99	1312	1.6	0.49	0.05	0.20
0.999	1312	1.7	0.49	0.05	0.20
0.999	1312	1.8	0.49	0.05	0.20
0.999	1312	1.9	0.49	0.05	0.20
0.999	1312	2.0	0.49	0.05	0.20

b. White women vs Black men

Cumulative Incidence of CHF among these 2 groups = 0.155. Proportion of women = 0.43. With $P_1 = 0.43$, SD $(P_1) = \sqrt{0.43*0.57} = 0.495$. With number of events = 1890 and assuming an R-square of 0.2 for the correlation with candidate covariates we get 94% power to detect a HR of 1.2

Power	Events	HR	SD	Alpha*	R^2
0.94	1890	1.2	0.495	0.05	0.20
0.99	1890	1.3	0.495	0.05	0.20
0.99	1890	1.4	0.495	0.05	0.20
0.999	1890	1.5	0.495	0.05	0.20
0.999	1890	1.6	0.495	0.05	0.20
0.99	1890	1.7	0.495	0.05	0.20
0.999	1890	1.8	0.495	0.05	0.20
0.999	1890	1.9	0.495	0.05	0.20
0.999	1890	2.0	0.495	0.05	0.20

c. White women vs. Black women

Cumulative Incidence of CHF among these 2 groups = 0.162. Proportion of white women = 0.31. With P_1 = 0.31, SD (P_1) = $\sqrt{}$0.31*0.69 = 0.46. With number of events = 2741 and assuming an R-square of 0.2 for the correlation with candidate covariates we get 94% power to detect a HR of 1.2

Power	Events	HR	SD	Alpha*	R^2
0.98	2741	1.2	0.495	0.05	0.20
0.99	2741	1.3	0.495	0.05	0.20
0.999	2741	1.4	0.495	0.05	0.20
0.999	2741	1.5	0.495	0.05	0.20
0.999	2741	1.6	0.495	0.05	0.20
0.999	2741	1.7	0.495	0.05	0.20
0.999	2741	1.8	0.495	0.05	0.20
0.999	2741	1.9	0.495	0.05	0.20
0.999	2741	2.0	0.495	0.05	0.20

3.9.1.2 Aim 1b

Again, with white women as the referent group, we would compute the power to investigate differences in log hazard for all-cause mortality among HF cases for white men, black men and black women at an alpha of 0.05 as follows:

a. White women vs White men, with HF

Cumulative Incidence of death among white HF cases = 0.226. Proportion of women among white cases = 0.61. With P_1 = 0.61, SD (P_1) = $\sqrt{}$0.61*0.39 = 0.49. With number of deaths = 296 and assuming an R-square of 0.10 for the correlation with candidate covariates we get 90% power to detect a HR of 1.5

Power	Events	HR	SD	Alpha*	R^2
0.31	296	1.2	0.49	0.05	0.10
0.55	296	1.3	0.49	0.05	0.10
0.78	296	1.4	0.49	0.05	0.10
0.90	296	1.5	0.49	0.05	0.10
0.96	296	1.6	0.49	0.05	0.10
0.99	296	1.7	0.49	0.05	0.10
0.997	296	1.8	0.49	0.05	0.10
0.999	296	1.9	0.49	0.05	0.10
0.999	296	2.0	0.49	0.05	0.10

b. White women vs Black men, with HF

Cumulative mortality among these 2 groups = 0.242. Proportion of women = 0.42. With P_1 = 0.42, SD (P_1) = $\sqrt{}$0.42*0.58 = 0.49. With number of events = 457 and assuming an R-square of 0.1 for the correlation with candidate covariates we get 94% power to detect a HR of 1.2

Power	Events	HR	SD	Alpha*	R^2
0.45	457	1.2	0.493	0.05	0.10
0.75	457	1.3	0.493	0.05	0.10
0.92	457	1.4	0.493	0.05	0.10
0.98	457	1.5	0.493	0.05	0.10
0.997	457	1.6	0.493	0.05	0.10
0.99	457	1.7	0.493	0.05	0.10
0.999	457	1.8	0.493	0.05	0.10
0.999	457	1.9	0.493	0.05	0.10
0.999	457	2.0	0.493	0.05	0.10

c. White women vs. Black women, with HF

Cumulative Incidence of death among these 2 groups = 0.178. Proportion of white women = 0.29. With P_1 = 029, SD (P_1) = $\sqrt{}$0.29*0.71 = 0.45. With number of events = 487 and assuming an R-square of 0.1 for the correlation with candidate covariates we get 89% power to detect a HR of 1.4

Power	Events	HR	SD	Alpha*	R^2
0.40	487	1.2	0.45	0.05	0.10
0.70	487	1.3	0.45	0.05	0.10
0.89	487	1.4	0.45	0.05	0.10
0.97	487	1.5	0.45	0.05	0.10
0.993	487	1.6	0.45	0.05	0.10
0.999	487	1.7	0.45	0.05	0.10
0.999	487	1.8	0.45	0.05	0.10
0.999	487	1.9	0.45	0.05	0.10
0.999	487	2.0	0.45	0.05	0.10

3.9.2 Aim2

3.9.2.1 Power for main analyses

a) Association between deprivation index and HF risk

We would be investigating the association between deprivation index and the log hazard of HF. neighborhood deprivation index data was available for 26818 persons and 4300 cases were observed among these participants; hence the probability of an event, Pr (E) = 0.1603. We obtained the r-squared for the association between deprivation index and all the other covariates using a multiple linear regression model; R^2 = 0.2307. We then computed the power to detect a range of hazard ratios for a 1 standard deviation increase in deprivation index for alpha = 0.05 and got over 99% power to detect a HR of 1.1. These estimates may be inflated considering our data are clustered in census tracks and our SEs are larger than would be otherwise.

Power	E	HR	SD	Alpha*	Pr(E)	R2
0.999	4300	1.1	1	0.05	0.1603	0.2307
0.999	4300	1.2	1	0.05	0.1603	0.2307
0.999	4300	1.3	1	0.05	0.1603	0.2307
0.999	4300	1.4	1	0.05	0.1603	0.2307
0.999	4300	1.5	1	0.05	0.1603	0.2307
0.999	4300	1.6	1	0.05	0.1603	0.2307
0.999	4300	1.7	1	0.05	0.1603	0.2307
0.999	4300	1.8	1	0.05	0.1603	0.2307
0.999	4300	1.9	1	0.05	0.1603	0.2307
0.999	4300	2	1	0.05	0.1603	0.2307

b) Association between deprivation index and post-HF all-cause mortality

We would also be investigating the association between deprivation index and the log hazard of all-cause mortality among HF cases. Neighborhood deprivation index data is available for 4300 cases and 940 deaths were observed among these participants; hence the probability of death, Pr (E) = 0.2186. We obtained the r-squared for the association between deprivation index and all the other covariates using a multiple linear regression model; R^2 = 0.2179. We then computed the power to detect a range of hazard ratios for a 1 standard deviation increase in deprivation index for alpha = 0.05 and got over 99% power to detect a HR of 1.1. In truth, our power estimates would be a little more modest given we have clustered data and our SEs are larger than would be otherwise.

Power	E	HR	SD	Alpha*	Pr(E)	R2
0.999	940	1.1	1	0.05	0.2186	0.2179
0.999	940	1.2	1	0.05	0.2186	0.2179
0.999	940	1.3	1	0.05	0.2186	0.2179
0.999	940	1.4	1	0.05	0.2186	0.2179
0.999	940	1.5	1	0.05	0.2186	0.2179
0.999	940	1.6	1	0.05	0.2186	0.2179
0.999	940	1.7	1	0.05	0.2186	0.2179
0.999	940	1.8	1	0.05	0.2186	0.2179
0.999	940	1.9	1	0.05	0.2186	0.2179
0.999	940	2	1	0.05	0.2186	0.2179

3.9.2.2 Power for multilevel modelling

The power for the multi-level analysis was estimated using simulations in MLPowSim. Based on our preliminary analyses we have over 4666 census tracts covered by the 27,078 participants included in our ancillary study. So we estimated that that we have on average 6 persons per census track. For our calculations we used a range of 5-7 persons per census track and 4000-4250 census track in order to be conservative in our power estimates. We used a Poisson distribution for the baseline hazard, standard normal distribution for the deprivation index (mean =0, variance =1) and a beta coefficient of 0.1 per unit change in deprivation index with an intercept of 0.1. We performed 50 simulations at alpha = 0.05 specifying a penalized quasi-likelihood (PQL) approach for estimating our regression coefficients and we got over 99% power for all scenarios.

# of Census tracts	n per census tract	Power for Intercept	Power for Slope
4000	5	0.99	0.99
4000	7	0.99	0.99
4050	5	0.99	0.99
4050	7	0.99	0.99
4100	5	0.99	0.99
4100	7	0.99	0.99
4150	5	0.99	0.99
4150	7	0.99	0.99
4200	5	0.99	0.99
4200	7	0.99	0.99
4250	5	0.99	0.99
4250	7	0.99	0.99

3.9.3 Aim3

3.9.3.1 BMI (or W/H^n) and HF Incidence

For this aim, we're investigating the association between BMI (or W/H^n) and the log hazard of HF. Weight and height data was available for 26713 persons and 4268 cases were observed among these participants; hence the probability of an event, Pr (E) = 0.1598. We obtained the r-squared for the association between BMI and all the other covariates using a multiple linear regression model; R^2 = 0.2088. We then computed the power to detect a range of hazard ratios for a 1 standard deviation increase in BMI for alpha = 0.05 and got over 99% power to detect a HR of 1.1.

Power	**E**	**HR**	**SD**	**Alpha***	**Pr(E)**	**R2**
0.999	4268	1.1	1	0.05	0.1598	0.2088
0.999	4268	1.2	1	0.05	0.1598	0.2088
0.999	4268	1.3	1	0.05	0.1598	0.2088
0.999	4268	1.4	1	0.05	0.1598	0.2088
0.999	4268	1.5	1	0.05	0.1598	0.2088
0.999	4268	1.6	1	0.05	0.1598	0.2088
0.999	4268	1.7	1	0.05	0.1598	0.2088
0.999	4268	1.8	1	0.05	0.1598	0.2088
0.999	4268	1.9	1	0.05	0.1598	0.2088
0.999	4268	2	1	0.05	0.1598	0.2088

3.9.3.2 Waist circumference and HF Incidence

Waist circumference data was available for 3395 persons and 251 cases were observed by the end of follow-up; hence the probability of an event, Pr (E) = 0.0739. We obtained the r-squared for the association between WC and all the other covariates using a multiple linear regression model; R^2 = 0.1253. We then computed the power to detect a range of hazard ratios for a 1 standard deviation increase in WC given E = 251, N= 3395 and alpha = 0.05 and got at least 97% power to detect a HR of 1.3.

Power	E	HR	SD	Alpha*	Pr(E)	R2
0.29	251	1.1	1	0.05	0.0739	0.1253
0.77	251	1.2	1	0.05	0.0739	0.1253
0.97	251	1.3	1	0.05	0.0739	0.1253
0.99	251	1.4	1	0.05	0.0739	0.1253
0.99	251	1.5	1	0.05	0.0739	0.1253
1	251	1.6	1	0.05	0.0739	0.1253
1	251	1.7	1	0.05	0.0739	0.1253
1	251	1.8	1	0.05	0.0739	0.1253
1	251	1.9	1	0.05	0.0739	0.1253
1	251	2	1	0.05	0.0739	0.1253

3.9.3.3 BMI (or W/H^n) and post-HF survival

Weight and height data was available for 4268 HF cases and 934 deaths were recorded among these participants; hence the probability of death, Pr (E) = 0.2188. We obtained the r-squared for the association between BMI and all the other covariates using a multiple linear regression model; R^2 = 0.2085. We then computed the power to detect a range of hazard ratios for a 1 standard deviation increase in BMI for alpha = 0.05 and got over 99% power to detect a HR of 1.2.

Power	E	HR	SD	Alpha*	Pr(E)	R^2
0.736	934	1.1	1	0.05	0.2188	0.2085
0.999	934	1.2	1	0.05	0.2188	0.2085
0.999	934	1.3	1	0.05	0.2188	0.2085
0.999	934	1.4	1	0.05	0.2188	0.2085
0.999	934	1.5	1	0.05	0.2188	0.2085
0.999	934	1.6	1	0.05	0.2188	0.2085
0.999	934	1.7	1	0.05	0.2188	0.2085
0.999	934	1.8	1	0.05	0.2188	0.2085
0.999	934	1.9	1	0.05	0.2188	0.2085
0.999	934	2	1	0.05	0.2188	0.2085

PART TWO

MANUSCRIPTS

Heart Failure Incidence and Mortality in the Southern Community Cohort Study‡

Authors: Elvis A. Akwo, MD, MS[1,5]; Edmond K. Kabagambe, DVM, PhD[1, 5]; Thomas J. Wang, MD[2, 5]; Frank E. Harrell Jr., PhD[3, 5]; William J. Blot, PhD[1, 4]; Michael Mumma, MS[4]; Deepak K. Gupta, MD* [2, 5]; Loren Lipworth, ScD* [1, 5]

[1] Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, 37203.

[2] Division of Cardiovascular Medicine, Vanderbilt Heart & Vascular Institute, Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, 37203.

[3] Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, 37203.

[4] International Epidemiology Institute, Rockville, Maryland, 20850.

[5]Vanderbilt Translational and Clinical Cardiovascular Research Center (V-TRACC), Vanderbilt University School of Medicine, Nashville, Tennessee, 37203.

* Equal contribution

Address correspondence to:
Dr. Loren Lipworth
Division of Epidemiology
Vanderbilt University School of Medicine
2525 West End Avenue, Suite 600
Nashville, TN 37203
Phone: 615-343-0639
Fax: 615-343-5938
loren.lipworth@vanderbilt.edu

Word Count: Manuscript (3228), abstract (250).

Short Title: Heart Failure in low-income blacks and whites in the US

Journal Subject Codes: [8] [110]

‡ Manuscript published in *Circulation: heart failure*. Citation: Akwo EA, Kabagambe EK, Wang TJ, Harrell FE, Jr., Blot WJ, Mumma M, Gupta DK and Lipworth L. Heart Failure Incidence and Mortality in the Southern Community Cohort Study. *Circ Heart Fail*. 2017;10.

Abstract

Background

There is a paucity of data regarding HF incidence among low-income and minority populations. Our objective was to investigate HF incidence and post-HF survival by race and sex among low-income adults in the southeastern US.

Methods

Participants were 27,078 white and black men and women enrolled during 2002-2009 in the Southern Community Cohort Study (SCCS) who had no history of HF and were receiving Centers for Medicare or Medicaid services (CMS). Incident HF diagnoses through December 31, 2010 were ascertained using ICD-9 codes 428.x via linkage with CMS research files.

Results

Most participants were black (68.8%), women (62.6%) and earned < $15,000/year (69.7%); mean age was 55.5 (10.4) years. Risk factors for HF were common: hypertension (62.5%), diabetes (26.5%), myocardial infarction (8.6%) and obesity (44.8%). Over a median follow-up of 5.2 years, 4,341 participants were diagnosed with HF. The age-standardized incidence rates were 34.8, 37.3, 34.9 and 35.6 PY/1000 in white women, white men, black men and black women, respectively, remarkably higher than previously reported. Among HF cases, 952 deaths occurred over a median follow-up 2.3 years. Men had lower survival; hazard ratios and 95% confidence intervals were 1.63 (1.27-2.08), 1.38 (1.11-1.72) and 0.90 (0.73-1.12) for white men, black men and black women compared with white women.

Conclusions

In this low-income population, HF incidence was higher for all race-sex groups than previously reported in other cohorts. The SCCS is a unique resource to investigate determinants of HF risk in a segment of the population underrepresented in other existing cohorts.

Introduction

There are over 26 million persons living with heart failure (HF) worldwide [1]. In the US, over 5.7 million adults (≈ 2.5% of the US adult population) are estimated to have HF [2]. About half of persons diagnosed with HF die within 5 years and the estimated total costs of HF in the US exceeded $30 billion in 2012 [3, 4]. Several established cardiovascular disease (CVD) cohorts have investigated HF incidence and mortality, including the Framingham Heart Study (FHS), Cardiovascular Health Study (CHS), Multi-ethnic Study of Atherosclerosis (MESA) and Atherosclerosis Risk in Communities (ARIC) [5-8]. The FHS included predominantly white individuals. Other cohorts, including CHS, MESA, and ARIC, enrolled multi-ethnic middle-class populations from select communities and their relatively small sample sizes limited assessment of differential risk patterns between demographic groups defined by both race and sex.

While data from these previous cohorts suggest differences in HF incidence rates and post-HF survival between population subgroups, knowledge gaps persist regarding the magnitude and direction of these differences, particularly in multi-ethnic low income populations with high burden of CVD risk factors. The prospective Southern Community Cohort Study (SCCS) provided a valuable opportunity to investigate differences in the incidence of HF as well as in post-HF survival between groups defined by race and sex: white women, black women, white men and black men [18].

Methods

Study sample

The SCCS is a prospective cohort study designed to investigate the incidence of cancer and other chronic diseases, including differential patterns by race and sex, in a low-income under-insured population underrepresented in previous studies. Between 2002 and 2009, the SCCS enrolled approximately 86,000 adults (≈ two-thirds black) aged 40-79 living in 12 southeastern states to investigate various chronic disease outcomes [18]. Approximately 86% of participants were recruited at community health centers (CHC), which provide primary health and preventive care services for low-income populations so that the cohort is made up of a segment of society (minority, poor, rural) seldom included in sizeable numbers in previous cohort studies; particularly those investigating CVD [18, 20]. The remaining 14% were recruited via mail-based general population sampling. Data on socioeconomic, demographic (including self-reported race), lifestyle, and anthropometric characteristics, as well as personal medical history, were ascertained at cohort enrollment via standardized computer-assisted personal interviews for CHC participants, and via self-administered mailed questionnaire for general population participants. Detailed description of SCCS methods has been previously published [18, 20].

SCCS participants (n = 27,078) included in the current analyses were individuals aged ≥ 65 years (n = 7001) at cohort enrollment, or persons < 65 years (n = 20,077) at enrollment who: a) reported being covered by Medicaid (which provides medical benefits to low-income adults and uninsured persons) on the baseline questionnaire; or b) reported being covered by Medicare (the primary health insurance program for persons aged ≥ 65) on the baseline questionnaire; or c) did not report Medicare or Medicaid on the baseline questionnaire but had a Centers for Medicare and Medicaid Services (CMS) claim within 90 days of being enrolled in SCCS. The restriction to these groups ensures that participants would likely have continuous coverage in Medicare and/or Medicaid from the time of SCCS enrollment to the end of the follow-up period (December 31st, 2010), for the ascertainment of incident HF events. Analyses were

restricted to self-reported African American or black and non-Hispanic white SCCS participants, since too few persons in other racial groups were available for stable statistical analysis.

Outcome ascertainment

Heart failure events were ascertained via linkage of the SCCS cohort with CMS Research Identifiable Files (which include Medicare institutional and non-institutional files, and the Medicaid Analytic Extract files). Incident HF was defined as the first occurrence of a medical claim with an International Classification of Diseases, 9th revision, discharge code of 428.x (428.0 to 428.9) within the Medicare institutional (Medicare Provider Analysis and Review, MEDPAR, which includes inpatient, outpatient and skilled nursing facility base files), Part B carrier (includes non-institutional physician services and durable medical equipment), or outpatient-based claims files or the Medicaid Analytic Extract (MAX) Inpatient and Other Services claims files, from the date of SCCS enrollment through December 31st, 2010. Detailed description of the CMS research files are published elsewhere [143].

Deaths, including dates and causes of death, were ascertained via linkage of the SCCS cohort with both the Social Security Administration (SSA) vital status service for epidemiologic researchers and the National Death Index (NDI) through December 31st, 2010. Both NDI and SSA are well-established and reliable means of identifying deaths in the US, and are expected to capture nearly all deaths [51, 122, 123].

Statistical Analysis

Descriptive statistics (means and standard deviations for continuous variables and counts and percentages for categorical variables) were computed for all study participants by race and sex.

To investigate the incidence of HF, duration of follow-up was computed from date of entry into the SCCS until the date of the first diagnosis of HF, date of death, or December 31st, 2010, whichever occurred first. Incidence rates (IR) of heart failure were calculated for white women, black women, white men and black men by dividing the number of HF cases by person-time of follow-up, presented per 1,000

person-years. The 95% confidence intervals (CI) were calculated using the quadratic approximation to the Poisson log likelihood for the log-rate parameter [134]. To account for age differences between the demographic categories, age-standardized rates were computed using the overall age distribution of the SCCS participants.

Multivariable Cox models were utilized to test whether differences in crude IRs between categories defined by race and sex persisted after adjustment for baseline covariates. Three models were constructed, with white women as the referent category: model 1 included indicator variables for white men, black women and black men and age (restricted cubic splines with 4 knots); model 2 additionally adjusted for body mass index (restricted cubic splines with 4 knots), and history of diabetes, hypertension, high cholesterol, MI/CABG or stroke (all yes/no); model 3 additionally adjusted for the following covariates: annual household income (<$15,000; $15,000-$24,999; ≥$25,000); education (< high school, high school/vocational training/junior college, college degree or higher), smoking (never, former, current <19.5 pack-years, current ≥ 19.5 pack-years, 19.5 being the median pack-years among current smokers), alcohol intake (linear and quadratic term), marital status (married/living as married with partner, separated/divorced, widowed, single/never married) and enrollment source (community health centers vs general population). Knots were placed at quantiles of covariate distributions, equally spaced in sample size [144].

For analyses of post-HF survival among those with a diagnosis of incident HF, follow-up time was defined as time from HF diagnosis to death or December 31st 2010 whichever occurred first. When date of death was coincident with date of HF diagnosis, follow-up time was set to 0.5 days. We computed cumulative mortality for both HF cases and non-cases using contingency tables. Age-adjusted estimates of the survivor functions (adjusted to the mean age of SCCS participants diagnosed with HF) were obtained from a stratified Cox model fit and plotted for all race-sex groups. Cox models were used to investigate differences in cumulative hazard for death (all-cause mortality) using white women as the referent group. Model 1 comprised indicator variables for white men, black women and black men and

age (restricted cubic splines with 4 knots). Variables included in models 2 and 3 are the same as described previously. P-values for race-by-sex interaction were computed in models for HF incidence and post-HF survival; and a p-value < 0.05 was considered statistically significant. Model assumptions were verified using Schoenfeld residuals and log (-log) plots.

All analyses were performed using STATA (version 12.1, Stata Corp, College Station, Texas, USA) and the 'rms' package for R version 3.1.1 (R Core Team 2014) [144, 145].

Ethics statement

SCCS participants provided written informed consent, and protocols were approved by the Institutional Review Boards of Vanderbilt University Medical Center and Meharry Medical College.

Results

Among the 27,078 SCCS participants included in this study, 68.8% were black, 62.6% were women, 69.7% had annual household income < $15,000 and 38.4% had less than a high school education. The mean (SD) age at enrollment was 55.5 (10.4) years. At baseline, risk factors for HF were common: hypertension (62.5%); diabetes (26.5%); myocardial infarction (8.6%); and obesity, BMI ≥ 30 kg/m^2 (44.8%) (**Table 1**).

Overall, white men were older and had the highest prevalence of MI and stroke at baseline (**Table 1**). In contrast, black women were more likely to be obese at baseline and report a history of diabetes and hypertension.

Table 1. Comparison* of baseline characteristics of 27, 078 SCCS participants who were receiving Medicare or Medicaid during follow-up between 2002 and 2010, according to race and sex

	Overall N = 27,078	White Women n = 5,252	White Men n = 3,202	Black Women n = 11,688	Black Men n = 6,936
Age (SD), years	55.5 (10.4)	57.7 (10.6)	58.7 (10.5)	54.4 (10.4)	54.3 (9.5)
Age Categories %					
40-54	51.2	41.9	38.0	56.1	56.1
55-64	22.9	23.6	22.7	21.5	24.9
≥ 65	25.9	34.5	39.3	22.4	18.9
BMI (kg/m^2) (SD)	30.4 (7.8)	30.7 (8.2)	28.7 (6.4)	32.4 (8.2)	27.7 (6.1)
BMI Categories %					
< 18.5	1.7	2.2	1.1	1.4	2.0
18.5 - < 25.0	24.0	23.6	28.2	16.5	35.0
25 - < 30.0	29.5	28.0	36.4	25.4	34.3
≥ 30.0	44.8	46.2	34.3	56.7	28.8
History of MI %	8.6	8.6	17.7	6.1	8.6
History of Stroke %	9.6	10.0	10.8	9.1	9.4
Diabetes %	26.5	24.3	23.8	29.8	23.8
Hypertension %	62.5	57.0	56.7	67.6	60.5
High Cholesterol %	39.5	49.5	47.1	38.0	31.0
Education %					
< High school (HS)	38.4	32.0	29.0	40.4	44.2
HS/Junior college/VT	53.1	58.7	53.1	52.1	49.0
≥ College degree	8.5	9.4	17.9	6.5	6.8
Annual Income < $15,000, %	69.7	65.9	53.4	74.5	71.8
Smoking %					
Never	34.7	37.0	21.7	45.1	21.3
Former	25.3	26.8	40.4	20.5	25.3
Current	40.1	36.2	37.9	34.5	53.4
Alcohol Intake %					
0 drink per day	54.9	66.8	48.8	61.0	38.5
>0-2 drinks per day	33.2	29.7	35.6	31.4	37.6
>2 drinks per day	11.9	3.5	15.6	7.5	23.9

* All comparisons between demographic groups were significant (p=0.02 for stroke; p<0.0001 for all other baseline variables).
VT: Vocational Training; MI: Myocardial Infarction; SD: Standard Deviation

HF incidence

Over a median (25^{th}, 75^{th} percentile) follow-up time of 5.2 (3.1, 6.7) years, 4,341 participants (16%) developed incident HF (IR: 32.8/1000 person-years; 95% CI: 31.8-33.8). White men had the highest age-standardized IR, 37.3/1000 PY, compared with 34.8, 34.9 and 35.6 in white women, black men and black women, respectively (**Table 2**).

In models adjusted for age and other risk factors for HF, black women had a significantly lower risk of HF when compared with white women [HR=0.89; 95% CI: 0.82-0.98]. The risk of HF was similar among white men (HR=1.09; 95% CI: 0.97-1.23) and black men (HR=1.04; 95% CI: 0.94-1.15) compared with white women (**Table 2**). There was no evidence of race-by-sex interaction [p = 0.22].

Table 2. Risk of incident heart failure among participants in the Southern Community Cohort Study, overall and stratified by race and sex

	Overall N = 27,078	White Women n = 5,252	White Men n = 3,202	Black Women n = 11,688	Black Men n = 6,936
Incident HF cases (n)	4,341	801	511	1,940	1,089
Person-Years (PY)	132,500	23,339	13,934	60,639	34,589
Cumulative Incidence (%)	16.0	15.3	16.0	16.6	15.7
Incidence Rate/1000PY (95% CI)					
Crude	32.8 (31.8, 33.8)	34.3 (32.0, 36.8)	36.7 (33.6, 40.0)	32.0 (30.6, 33.4)	31.5 (29.7, 33.4)
Age-adjusted	35.1 (34.1, 36.2)	34.8 (32.4, 37.2)	37.3 (34.0, 40.6)	35.6 (33.9, 37.2)	34.9 (32.7, 37.1)
Hazard Ratio (95% CI)					
Model 1		1.00	1.04 (0.93, 1.16)	1.02 (0.94, 1.11)	0.99 (0.91, 1.09)
Model 2		1.00	1.02 (0.91, 1.14)	0.91 (0.83, 0.99)	1.06 (0.97, 1.17)
Model 3		1.00	1.09 (0.97, 1.23)	0.89 (0.82, 0.98)	1.04 (0.94, 1.15)

Model 1: Includes age (restricted cubic splines with 4 knots), race and sex. **Model 2:** Model 1 + BMI (restricted cubic splines with 4 knots), history of diabetes, hypertension, high cholesterol, MI and stroke (all yes/no). **Model 3:** Model 2 + annual household income (<$15000, $15000-$24999 & ≥$25000), education (< high school, high school/vocational training/junior college, college degree or higher), smoking (never, former, current < 19.5 pack-years, current ≥ 19.5 pack-years) and alcohol intake (linear and quadratic term), marital status (married/living as married with partner, separated/divorced, widowed, single/never married) and enrollment source (community health centers vs general population). **P-value for race×sex interaction** = 0.22. CI: Confidence Interval

Post-HF survival

Among the 4,341 individuals who developed incident HF, 952 died (cumulative mortality = 21.9%) over a median (25^{th}, 75^{th} percentile) post-HF follow-up time of 2.3 (0.9, 4.2) years (**Table 3**). Men had higher percent mortality than women (29% vs. 18%), with little difference by race. In persons without HF (n= 22,737), there were 1,929 deaths, corresponding to a percent mortality of 8.5%.

Figure 1 shows age-adjusted survival curves for persons diagnosed with HF stratified by race and sex. The 5-year post-HF survival probability was significantly lower among white men (0.55; 95% CI: 0.49-0.61) and black men (0.64; 95% CI: 0.60-0.67) compared with white women (0.73; 95% CI: 0.69-0.78) and black women (0.77; 95% CI: 0.74-0.79), respectively [$p < 0.0001$]. Racial differences within sex groups were not statistically significant. Similar patterns were observed for 1-year and 3-year survival probabilities.

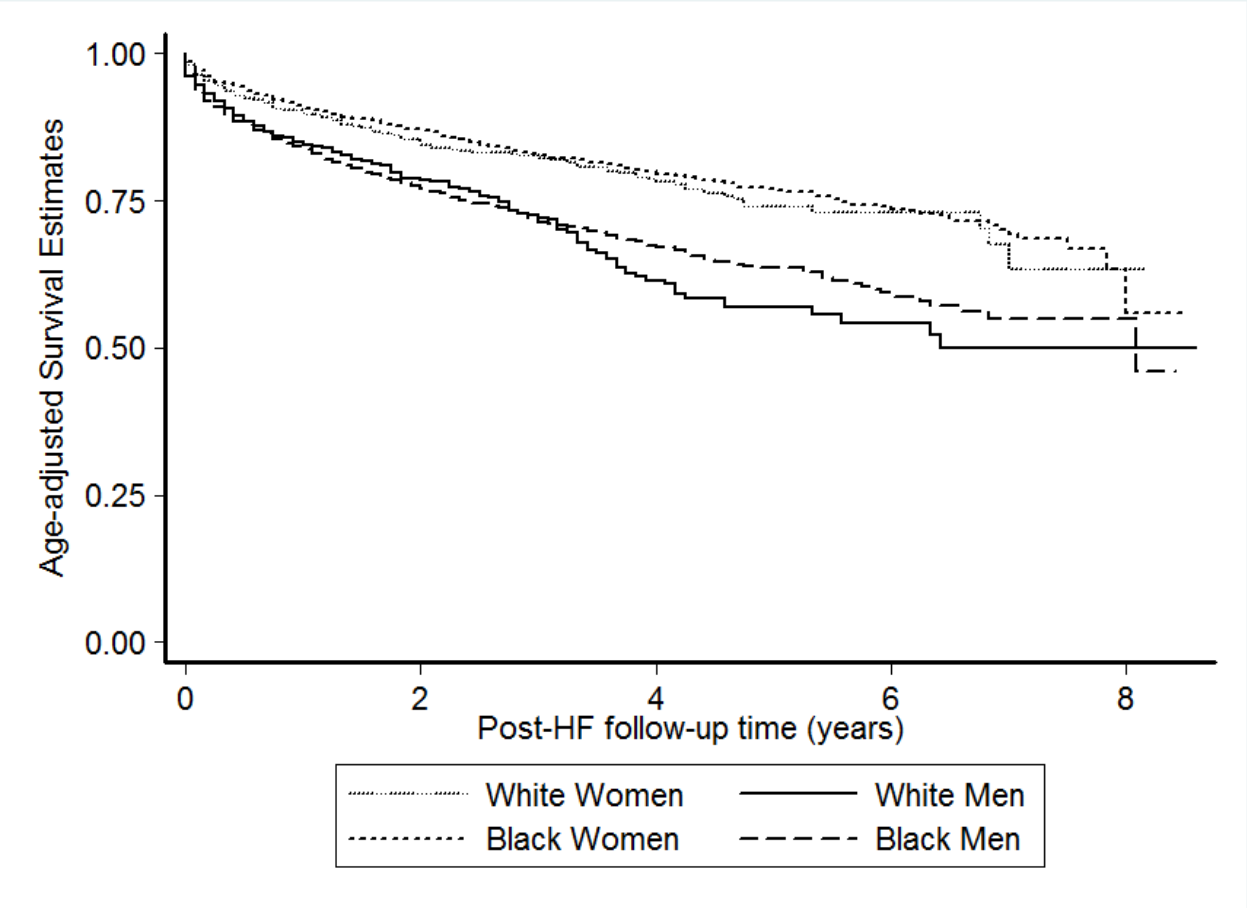
Compared with white women, the risk of death was 60% (95% CI: 27%-202%) higher in white men and 35% (95%CI: 9%-65%) higher in black men in analyses adjusted for age, BMI, hypertension, diabetes, high cholesterol, past history of MI/CABG and stroke (**Table 3**). These findings were robust to further adjustment for lifestyle factors and enrollment source. In contrast, comparisons between black women and white women suggested minimal non-significant relative differences in risk by race in all models. The race-sex interaction term was not statistically significant [$p = 0.92$].

Table 3. Percent mortality of SCCS participants according to heart failure status, overall and stratified by race and sex

	Overall N = 4,341	White Women n = 801	White Men n = 511	Black Women n = 1,940	Black Men n = 1,089
Deaths (n)	952	144	152	343	313
Percent Mortality (%)	21.9	18.0	29.7	17.7	28.7
Risk of death: Hazard ratio (95% CI)					
Model 1		1.00 (ref)	1.73 (1.37, 2.17)	0.91 (0.75, 1.10)	1.61 (1.32, 1.96)
Model 2		1.00 (ref)	1.60 (1.27, 2.02)	0.89 (0.73, 1.09)	1.35 (1.09, 1.65)
Model 3		1.00 (ref)	1.63 (1.27, 2.08)	0.90 (0.73, 1.12)	1.38 (1.11, 1.72)

Model 1: Includes age (restricted cubic splines with 4 knots); race and sex. **Model 2:** Model 1 + BMI (restricted cubic splines with 4 knots), history of diabetes, hypertension, high cholesterol, MI and stroke (all yes/no). **Model 3:** Model 2 + annual household income (<$15000, $15000-$24999 & ≥$25000); education (< high school, high school/vocational training/junior college, college degree or higher), smoking (never, former, current < 19.5 pack-years, current ≥ 19.5 pack-years) and alcohol intake (linear and quadratic term), marital status (married/living as married with partner, separated/divorced, widowed, single/never married) and enrollment source (community health centers vs general population). **P-value for race×sex interaction** = 0.92. CI: Confidence Interval

Figure 1: Age-adjusted Survival curves for SCCS participants diagnosed with HF stratified by race and sex



The adjusted survival estimates were computed at the mean age (58 years) of all participants with HF (n = 4341). Compared with white women and black women, white men and black men had significantly lower survival ($p<0.0001$). The 1-year, 3-year and 5-year age-adjusted survival estimates were (0.91, 0.83, 0.77); (0.90, 0.82, 0.74); (0.84, 0.71, 0.64) and (0.85, 0.73, 0.58) for black women, white women, black men and white men respectively.

Discussion

We investigated heart failure incidence and post-HF mortality in a large multi-ethnic low-income sample from the southeastern United States. Our principal findings are: 1) the incidence rate for HF was remarkably high across all race and sex groups in the SCCS, 2) there was no significant difference in age-adjusted incidence rates across groups defined by race and sex but after full adjustment for socioeconomic status and traditional cardiovascular risk factors black women had the lowest risk of HF, and 3) higher post-HF mortality among men with no significant racial differences.

The incidence rates for heart failure in the SCCS exceeded those previously reported from established CVD cohorts (**Table 4**). In CHS (n = 5,888; age range: ≥ 65 years) for example, the age-standardized HF incidence rates after 10 years of follow-up among white men, black men, white women and black women were 30.2, 19.2, 27.5 and 22.6 per 1000 PY respectively [5]. In ARIC (n = 14,933; age range: 45-64), Loehr et al found IRs of 6.0, 9.1, 3.4 and 3.8 per 1000 PY, respectively, in these race-sex groups [8]. The FHS and MESA reported even lower IRs (7.2 and 4.2 per 1000 PY in men and women in FHS and 3.1 per 1000 PY overall in MESA) [6, 7]. The higher SCCS HF incidence could be explained in part by notably higher prevalence of CVD risk factors (in particular hypertension, diabetes, obesity, prior history of MI) in the SCCS study sample at baseline compared with ARIC, CHS, FHS and MESA (**Table 5**). In addition, SCCS participants were largely of low socioeconomic status, with over two-thirds having annual household income <$15,000. Prior evidence suggests a strong independent association between socioeconomic status and HF risk; [9, 10, 105, 146] thus participants in SCCS may be at higher risk of unfavorable societal stressors and an elevated risk of adverse cardiovascular outcomes including HF.

White men had the highest crude incidence rate of HF in the SCCS, consistent with findings from the CHS. However, minimal differences in age-adjusted incidence rates and HF risk between groups after adjustment for CVD risk factors (except for black women who had significantly lower risk) suggest homogeneity of HF risk profile. Similarly, in ARIC, crude racial and sex differences in incidence density were attenuated by adjustment for CVD risk factors [8].

Table 4: Comparison of heart failure incidence between SCCS, ARIC, MESA and CHS cohorts

	HF Incidence Rates (IR) per 1000 PY							
	SCCS* 45-64 yrs	ARIC[8] 45-64 yrs	SCCS* 65-79 yrs	CHS[5] ** ≥ 65 yrs			SCCS* 40-79 yrs	MESA[6] 45-84 yrs
	n = 15,321	n = 14,933	n = 7,001	n = 5,888			n = 27,078	n = 6,814
				5-yr	10-yr			
Overall	35.8	5.7	39.5	19.3	24.0	Overall	33.7	3.4
White Women	38.3	3.4	34.6	14.5	19.2			
White Men	39.7	6.0	38.3	24.9	30.2	Whites	34.3	2.4
Black Women	35.1	8.1	42.4	19.6	22.6			
Black Men	34.0	9.1	41.9	23.5	27.5	Blacks	33.4	4.6

*Incidence rates computed for SCCS participants aged 45-64, 65-79 and 40-79 for comparability with the ARIC, CHS and MESA cohorts respectively. In addition, the rates are standardized to the age distribution of the SCCS study participants within these age ranges.

**The tabulated values are computed from values presented in Arnold et al and standardized to the age distribution of CHS participants.

ARIC: Atherosclerosis Risk in Communities; CHS: Cardiovascular Health Study; MESA: Multi-ethnic Study of Atherosclerosis; SCCS: Southern Community Cohort Study.

Table 5: Comparison of baseline characteristics of SCCS, MESA, ARIC and CHS participants

	SCCS (n= 27,078)	ARIC[8] (n= 14,993)	CHS[5] (n= 5,888)	MESA[6] (n = 5,923)†	FHS[7] (n=9,405)‡
Age (years)*	55.5 (10.4)	54 (6)	72.8 (5.6)	61.8 (10.3)	41 (10)
Women (%)	62.6	54	57.6	53	53
Blacks (%)	68.8	27	15.7	26.1	≈ 0
Education (<high school) (%)	38.4	24	29.5	16	56§
BMI (kg/m^2)*	30.4 (7.8)	27.3 (5.1)	26.7 (4.7)	28.0 (5.4)	24.9 (3.8)‖
Obese (BMI > 30) %	44.8	26	19	32	10.8‖
Diabetes (%)	26.5	11	16.4	11.6	4.1
Hypertension (%)	62.5	33	57.7	42	7
Myocardial infarction (%)	8.6	4	9.6	n/a	1.6#
Stroke (%)	9.6	1.4	4.2	n/a	0.5**
Ever smoked cigarettes (%)	65.3	58.2	53.5	49	57.3§

*Tabulated values are mean (SD). n/a = not applicable.

†By design, participants enrolled in MESA were free of CVD at baseline so the prevalence of MI and stroke at baseline in this cohort may be best described as not applicable.

‡The baseline data for the FHS pertains whenever available to both the original cohort and the offspring cohort given this larger sample was utilized to compute HF incidence rates referenced in the manuscript. In the absence of such data, we have presented data from the original cohort (or subsamples thereof) as a proxy and indicated so in each case.

§Obtained from a subsample of the parental FHS cohort data (n=1037, 45-62yrs and CHD-free at baseline).

‖Obtained from a subsample of the parental FHS cohort data (n=2922, 30-62yrs and CHD-free at baseline).

#Obtained from the parental FHS cohort data (n=5209) and pertains to the composite of MI, CHD-related sudden death and angina pectoris.

**Obtained from the parental FHS cohort data (n=5209).

ARIC: Atherosclerosis Risk in Communities; CHS: Cardiovascular Health Study; SCCS: Southern Community Cohort Study; MESA: Multi-ethnic Study of Atherosclerosis; FHS: Framingham Heart Study.

Overall, the 5-year post-HF survival in SCCS was higher than the 52% previously reported (data from the Olmsted county study) [3]. This may be due in part to the fact that SCCS participants had shorter post-HF follow-up time, were younger at baseline (55.5 vs 74 years) and temporal trends suggesting improved post-HF survival [3] related to recent improvements in therapeutic options. In addition, participants in the Olmsted county study were mostly non-Hispanic Whites who may be at higher risk of HF with reduced ejection fraction (HF*r*EF), which has a less favorable prognosis compared with heart failure with preserved ejection fraction (HF*p*EF) [22, 23].

The relative patterns of post-HF survival for the four demographic subgroups in SCCS were substantially different from those seen in ARIC and CHS. In ARIC, compared with white men and women, black men and women had the lowest survival probability following admission for HF. The 5-year case fatality for white women, white men, black women and black men were 35.8%, 41.2%, 46.1% and 51.8%, respectively. The racial differences were significant, with black men having the highest all-cause mortality following admission, but the differences by sex were non-significant. In CHS, the mortality rate in white women, white men, black women and black men were 35.5, 40.5, 33.6 and 44.4 per 100 PY respectively. After full adjustment for covariates there were no significant racial differences, but women had a 15% lower risk of all-cause mortality [HR: 0.85; 95% CI: 0.73, 0.99] [51]. In SCCS, white men had the lowest 5-year survival post-HF diagnosis; but after full adjustment, there were mainly sex-differences in post-HF mortality with higher risk of death among men and no significant racial differences. This could be explained in part by the higher prevalence of MI among men. MI is associated with greater risk for the development of HF*r*EF which is known to have a worse prognosis compared with HF*p*EF [22, 23]. However, MI does not fully account for the higher risk of post HF mortality among men, as this risk persisted even after full adjustment for relevant baseline covariates (including history of MI).

Limitations of our study should be noted. Our study sample may not be representative of the background population of the Southeastern states as the recruitment and sampling scheme utilized by the

SCCS was tailored towards low-income, rural and under-insured populations not often included in sizeable numbers in other cohorts investigating chronic disease outcomes. Also, HF was ascertained via linkage with CMS Research Identifiable Files using ICD-9 codes 428.x, rather than independent physician adjudication. However, the diagnosis codes (ICD-9 428.x) algorithm for identification of HF used in this study has been previously validated and utilized in other cohorts [119-121]. A review of the detection of HF in administrative claims data that included eight studies conducted among Medicare beneficiaries reported positive predictive values (PPVs) between 76% and 99%, with the majority of the studies reporting PPVs over 90% [119]. These codes have also been used with high specificity in a number of studies [120, 121] even though no independent validation was conducted by the SCCS investigators. An over-representation of groups with elevated HF risk (persons > 65 and persons < 65 receiving Medicare) in our SCCS sub-cohort compared with the SCCS base population, may have contributed to higher HF incidence rates than would be expected for the total SCCS cohort. However with the mean age of the total cohort being ≈ 52.6 years [107] versus 55.5 years for our sub-cohort, the small age difference between both populations may have had less than dramatic effects on the HF incidence. In addition, with studies suggesting that the sensitivity of ICD-9 code 428.x for HF ascertainment varies between 62.8 and 89% [119, 124], it is plausible that we may have underestimated the incidence rate of HF in our sub-cohort. Also, when contrasting the incidence rates between our study and previous CVD cohorts (like ARIC and CHS) we used data for comparable age groups between studies (Table 4). However, the fact that HF represents a myriad of clinical conditions, the lack of universality in the definition of HF and the heterogeneity in the methods for HF ascertainment between studies makes head-to-head comparisons between studies difficult. Our analyses required assumptions regarding the continuous coverage in CMS of persons less than 65 years, raising the possibility of incomplete capturing of HF events in this age stratum of the SCCS cohort. However, we found that over 81.9% of persons aged < 65 who reported CMS coverage at baseline had a claim for any condition within 90 days of being enrolled in SCCS. This suggests that an even greater proportion of participants included in this study filed at least one claim at some point during follow-up from 2002 to 2010 and thus any HF event would

likely have been captured if it occurred. Data on baseline covariates (including anthropometric and cardiovascular risk factors) were based on self-report of a physician diagnosis and use of medications (diabetes and hypertension). While self-report may be susceptible to recall and misclassification bias, these methods have been successfully used and validated in large epidemiologic cohorts, including the SCCS. Many of the questions on the SCCS questionnaire were adapted from questionnaires used and validated in other settings; and a series of independent validation studies using biomarkers, repeat interviews or medical records have demonstrated the reliability of the questionnaire within the SCCS population for variables such as smoking status, self-reported diseases including diabetes, height and weight [18].

The SCCS cohort is comprised of a substantial number of individuals from minority and low-income populations who are traditionally under-represented in most studies investigating CVD and heart failure in particular. The incidence rates for HF in the SCCS exceeded that of most existing cardiovascular cohorts. Therefore, the SCCS provides an unparalleled opportunity to investigate patterns in HF incidence and mortality among the highest risk individuals. In addition, both black and white participants included in this cohort had minor differences in income and education levels thereby curtailing confounding by socioeconomic differences. The availability of a large sample of participants and HF cases provided the opportunity to adequately explore differential patterns across sex and racial categories. Also, linkage with the NDI and SSA allowed for robust ascertainment of all-cause mortality. In conclusion, in this low-income multiethnic population, we found higher incidence rates for HF in all race-sex groups than previously reported in other CVD cohorts which was paralleled by high prevalence of CVD risk factors at baseline. This suggests that SCCS can be a unique resource to investigate determinants of HF risk in a segment of the population underrepresented in other existing cohorts.

Neighborhood deprivation predicts heart failure risk in a low-income population of blacks and whites in the southeastern United States

Elvis A. Akwo, MD, MS[1,5]; Edmond K. Kabagambe, DVM, PhD[1, 5]; Frank E. Harrell, PhD[2, 5]; William J. Blot, PhD[1, 3]; Justin M. Bachmann, MD, MPH[4,5]; Thomas J. Wang, MD[4,5]; Deepak K. Gupta, MD* [4, 5], Loren Lipworth, ScD* [1, 5]

[1] Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, 37203.

[2] Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, 37203.

[3] International Epidemiology Institute, Rockville, Maryland, 20850.

[4] Division of Cardiovascular Medicine, Vanderbilt Heart & Vascular Institute, Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, 37203.

[5]Vanderbilt Center for Translational and Clinical Cardiovascular Research (V-TRACC), Vanderbilt University School of Medicine, Nashville, Tennessee, 37203.

Address correspondence to:
Dr. Loren Lipworth
Division of Epidemiology
Department of Medicine
Vanderbilt University School of Medicine
2525 West End Avenue, Suite 600
Nashville, TN 37203
Phone : 615-343-0639
Fax : 615-343-5938
loren.lipworth@vanderbilt.edu

Word Count: Manuscript (4966), abstract (346).

Running Title: Neighborhood deprivation and heart failure.

AHA Journal Subject Terms: [Epidemiology] [Heart failure]

Abstract

Background

Recent data suggest that neighborhood socioeconomic environment predicts heart failure (HF) hospital readmissions, yet evidence for the association with HF incidence and post-HF mortality is scant. We sought to investigate whether neighborhood deprivation predicts the risk of incident HF and post-HF survival beyond individual socioeconomic status (SES) in a low-income population.

Methods

Participants included in this study were 27,078 white and black men and women recruited during 2002-2009 in the Southern Community Cohort Study (SCCS), who had no history of HF and were receiving Centers for Medicare or Medicaid services (CMS). Incident HF diagnoses through December 31, 2010 were ascertained using ICD-9 codes 428.x via linkage with CMS research files. Participant residential information was geocoded and the census tract was determined by a spatial join to the US Census Bureau's TIGER/Line Shapefiles using geographic information systems technology. The neighborhood deprivation index was constructed using principal components analysis based on census tract-level socioeconomic variables. Cox models with Huber-White cluster sandwich estimator of variance were utilized to investigate the association between deprivation index and both HF risk and post-HF mortality.

Results

The study sample was predominantly middle-age (mean 55.5 years), black (69%), female (63%), and of low income (70% earned < $15,000/year). Over half of the participants lived in the most deprived neighborhoods. Over a median follow-up of 5.2 years, 4,300 participants were diagnosed with HF. After adjustment for demographic, lifestyle and clinical factors, a 1 interquartile (IQR) increase in deprivation index was associated with a 12% increase in the risk of HF [HR= 1.12; 95% CI: 1.07-1.18] and 4.7% of the variance in HF risk [ICC = 4.8; 95% CI: 3.6-6.4] was explained by neighborhood deprivation. Among

HF cases, 940 deaths occurred over a median follow-up 2.3 years. In multivariable-adjusted models deprivation index was not associated with the risk of post-HF mortality [HR= 1.05; 95% CI: 0.95-1.17].

Conclusions

In this low-income population, scant neighborhood resources compound the risk of HF above and beyond individual socioeconomic status and traditional cardiovascular risk factors. Improvements in community resources may be a significant axis for curbing the burden of HF.

Introduction

Heart failure (HF) is a major public health problem, particularly in the southeastern United States (US) which has been described as the "heart failure belt" [147]. This region of the US has the highest high prevalence of established HF risk factors (including coronary heart disease, high blood pressure, diabetes and obesity) which themselves may be associated with socioeconomic characteristics that influence health outcomes. While evidence from middle-class persons suggests that individual socioeconomic status (SES) contributes to HF risk,[9, 10, 105, 146] recent data support that neighborhood factors may also predict HF readmissions independent of individual-level SES [11]. However, it is not known whether such neighborhood factors are independent predictors of other HF outcomes such as HF incidence and post-HF mortality among persons with already low individual SES.

The Southern Community Cohort Study is a prospective cohort study that recruited persons of low individual SES from 12 states in the southeastern United States. Within this cohort, we tested the hypothesis that neighborhood characteristics (defined by a composite deprivation index) predict the risk of incident HF and post-HF survival beyond individual-level socioeconomic status (defined by annual household income and highest level of education attained).

Methods

Design and study population

The SCCS is a prospective cohort study designed to investigate the incidence of cancer and other chronic diseases, including differential patterns by race and sex, in a resource-limited under-insured population underrepresented in previous studies [18]. A total of 84,797 participants aged 40-79 were enrolled into the SCCS between March 2002 and September 2009 [148]. Approximately 86% of participants were recruited at community health centers (CHC), which provide primary health and preventive care services for resource-limited populations such that the cohort is made up of a segment of society seldom included in substantial numbers in previous cohort studies; particularly those investigating cardiovascular disease (CVD) [18, 20]. The remaining 14% were recruited via mail-based sampling of the general population. Demographic, socioeconomic, lifestyle, and anthropometric data, as well as personal medical history, were ascertained at cohort enrollment via standardized computer-assisted personal interviews for CHC participants, and via self-administered mailed questionnaire for persons recruited from the general population. Detailed description of SCCS methods has been previously published [18, 20, 107].

For the current analyses, we included 27,078 participants who were either ≥ 65 years old at cohort enrollment (n = 7,001), or < 65 years at enrollment (n = 20,077) and: a) reported being covered by Medicaid (which provides medical benefits to low-income adults and uninsured persons) on the baseline questionnaire; or b) reported being covered by Medicare (the primary health insurance program for persons aged ≥ 65) on the baseline questionnaire; or c) did not report Medicare or Medicaid on the baseline questionnaire but had a Centers for Medicare and Medicaid Services (CMS) claim within 90 days of being enrolled in SCCS. The restriction to these groups increases the likelihood of participants having continuous coverage in Medicare and/or Medicaid from the time of SCCS enrollment to the end of the follow-up period (December 31st, 2010), for the ascertainment of incident HF events. Analyses were

restricted to self-reported African American or black and non-Hispanic white SCCS participants, since too few persons in other racial groups were available for stable statistical analyses.

Census tracts

Census tracts were used as proxies for neighborhoods in this study. Census tracts are small, relatively permanent statistical subdivisions of a county or equivalent entity that are updated by local participants prior to each decennial census as part of the US Census Bureau's Participant Statistical Areas Program [126]. Census tract boundaries are delineated with the intention of being maintained over a long time period so that statistical comparisons can be made from census to census. Across the US, census tracts usually cover a contiguous area and generally have a population size between 1200-8000 persons, with an optimum size of 4000. The 27,078 SCCS participants included in the current analyses resided in 4,666 census tracts.

At the time of the SCCS baseline interview, study participants provided information on their residential address which was then geocoded by a multi-stage process incorporating both batch and interactive processes [149]. The census tract for the geocoded address was then determined by a spatial join to the US Census Bureau's Topologically Integrated Geographic Encoding and Referencing (TIGER/Line®) Shapefiles [150] using ESRI ArcMap 10.0 software (ESRI, Redlands, CA) that utilizes GIS (geographic information systems) technology. Geocoding of SCCS participants' addresses and linkage to geographic information datasets such as census tract data allowed development of residence-specific metrics including the SCCS-derived deprivation index.

Neighborhood deprivation Index

The SCCS-derived deprivation index is a clustering of social and economic indicators which reflect neighborhood deprivation and have been linked to adverse health outcomes. It was constructed using principal components analysis based on 11 census tract-level variables representing 4 main dimensions [128, 151]: 1) Social indicators: percentage of housing units with ≥1 occupant per room,

percentage of occupied housing units with renter/owner costs >50% of income and percent female-headed households with dependent children; 2) Wealth and income: percentage of households with income <$30,000 per year, percentage of persons with income below the 1999 poverty status, percentage of households with public assistance income, percentage of households with no car and median value of owner-occupied housing units; 3) Education: percentage of persons aged ≥25 that did not graduate high school; and 4) Occupation: percentage of males and females who are unemployed and percentage males in professional occupations.

In the original description of the neighborhood deprivation index by Messer et al using data from across 8 study areas, the first principal component was retained, as it explained over 67% of the variance (while the second principal component explained less than 10% of the variance) with component loadings ranging between 0.2-0.4, suggesting similar contribution of each of the component variables to the first principal component [128] Signorello et al found that in the SCCS, the first principal component explained most of the variability (over 60%) in the component measures as well, as such it was retained for the construction of the deprivation index in the SCCS [151].

Individual socioeconomic variables and other covariates

SCCS participants reported their highest level of education attained, in 8 categories ranging from less than high school to graduate-level degrees. Participants also reported the range of their total household income for the year prior to enrollment, in 5 categories ranging from less than $15,000 to over $100,000 or more. History of tobacco smoking was self-reported as never, former and current and also in terms of number of cigarettes per day and alcohol use in number of drinks per day. The presence of traditional cardiovascular risk factors at baseline was based on a self-reported history of physician-diagnosed hypertension, diabetes mellitus, high cholesterol, as well as self-reported use of medications for hypertension, diabetes mellitus, or high cholesterol. History of myocardial infarction and stroke was based on self-report and confirmed via medical records for a small random sample of SCCS participants. While self-report of baseline covariates may be susceptible to recall bias, these methods have been

successfully used and validated in large epidemiologic cohorts, including the SCCS.[18] Many of the questions on the SCCS questionnaire were adapted from questionnaires used and validated in other settings; and a series of independent validation studies using biomarkers, repeat interviews or medical records have demonstrated the reliability of the questionnaire within the SCCS population for variables such as smoking status, self-reported diseases including diabetes, height and weight [18] Total amount of moderate and vigorous exercise was measured in metabolic equivalent-hours per day.

Outcome ascertainment

Heart failure events were ascertained via linkage of the SCCS cohort with CMS Research Identifiable Files (which include Medicare institutional and non-institutional files, and the Medicaid Analytic Extract files). Incident HF was defined as the first occurrence of a medical claim with an International Classification of Diseases, 9th revision, discharge code of 428.x within the Medicare institutional (Medicare Provider Analysis and Review, MEDPAR, which includes inpatient, outpatient and skilled nursing facility base files), Part B carrier (includes non-institutional physician services and durable medical equipment), or outpatient-based claims files or the Medicaid Analytic Extract (MAX) Inpatient and Other Services claims files, from the date of SCCS enrollment through December 31st, 2010. Detailed description of the CMS research files are published elsewhere [143].

Deaths, including dates and causes of death, were ascertained via linkage of the SCCS cohort with both the Social Security Administration (SSA) vital status service for epidemiologic researchers and the National Death Index (NDI) through December 31st, 2010. Both NDI and SSA are well-established and reliable means of identifying deaths in the US, and are expected to capture nearly all deaths [51, 122, 123]. When date of death was coincident with date of HF diagnosis, follow-up time was set to 0.5 days.

Statistical Analysis

Descriptive statistics (means and standard deviations for continuous variables and counts and percentages for categorical variables) were computed for all study participants overall and by tertiles of

deprivation index. Tertile cut-points were based on the distribution of the values of deprivation index for census tracts (n = 4666) covered by the population included in the current analyses.

Covariate selection and spending degrees of freedom

The selection of variables to be included in the multivariable models was based on the hypothesized relationships between the baseline covariates in question, deprivation index and the outcomes of interest (HF incidence and post-HF death). The functional form of the covariates (i.e. degrees of freedom spent) and deprivation index was based on *a priori* knowledge and the rank correlations between the covariate and the outcomes.

Modelling hierarchical data

For the current analyses, the data were organized in a hierarchical fashion comprising 2 levels with individual participants (level-1units) nested within census tracts (level-2 units). Given the nested structure of the data, the non-independence of the data points within each census track and the limitations of a multilevel modelling approach in this setting (unbalanced data with many small clusters), we utilized a Cox proportional hazards model that accounts for non-independence using the Huber-White cluster sandwich estimator of variance, H_c [127] whose general formula for linear and nonlinear models is given below:

$$H_c = I^{-1}(b)[\sum_{i=1}^{c}\{(\sum_{j=1}^{n_i} U_{ij})(\sum_{j=1}^{n_i} U_{ij})'\}]I^{-1}(b),$$

where I is the information matrix (the second derivative of the log likelihood, log L) [127] and U is the score statistic – the first derivative of log L. In the specific case of the Cox model, the cluster sandwich estimator, H_c uses special score residuals for U (the score vector) given there are no per-observation score contributions [127, 142].

The proportion of variance explained by differences at the census-tract level was estimated from the intra-class correlation coefficient (ICC), which was computed based on the latent variable approach. This approach uses the closed-form solution of the ICC in the multi-level logistic model to make an approximate estimation of the ICC in the multi-level Cox proportional hazards model.

$$ICC = \frac{\tau_{00}}{\tau_{00} + \pi^2/3}, \text{ where } \tau_{00} = \text{group-level variance [137].}$$

Deprivation Index and Heart Failure Incidence

Multivariable-adjusted Cox analyses were used to model a flexible association between deprivation index (modelled using restricted cubic splines with 4 knots) and HF accounting for nonlinearity and non-additivity of effects by race. The fully adjusted model was then used to dynamically create plots of log relative hazard of incident HF versus deprivation index by race. The covariates included in the full model were: age at enrollment (restricted cubic splines with 4 knots), race (white/black), deprivation index×race interaction terms (linear and nonlinear), sex (men/women), cigarette smoking (never, former, current <19.5 pack-years, current ≥ 19.5 pack-years, 19.5 being the median pack-years among current smokers), alcohol intake (linear), total MET-hours of moderate or greater exercise (linear and quadratic term), body mass index (BMI, restricted cubic splines with 4 knots), history of diabetes, hypertension, high cholesterol, MI/CABG or stroke (all yes/no), annual household income (<$15,000; $15,000-$24,999; ≥$25,000) and education (< high school, high school/vocational training/junior college, college degree or higher).

In further analyses, multivariable Cox models assuming linearity of effects (with deprivation index modelled as a rescaled continuous variable using the interquartile range, IQR) were used to estimate the effect of a 1 IQR increase in deprivation index on HF incidence while adjusting for relevant covariates in a sequential fashion. Hazard ratios (HR) for a 1 interquartile range increase in deprivation index (as a continuous measure) compare the hazard of the event occurring for a typical person in the middle of the upper half (the 75th percentile) of the distribution of deprivation index to the hazard of the event for a

typical person in the middle of the lower half (25th percentile) of the distribution. Model 1 included deprivation index (as a linear IQR-rescaled predictor), age at enrollment (restricted cubic splines with 4 knots), race (white/black), sex (men/women) and the deprivation index×race interaction term. Model 2 additionally adjusted for lifestyle and clinical covariates including: cigarette smoking (never, former, current <19.5 pack-years, current ≥ 19.5 pack-years, 19.5 being the median pack-years among current smokers), alcohol intake (linear), total MET-hours of moderate or greater exercise (linear and quadratic term), body mass index (restricted cubic splines with 4 knots), history of diabetes, hypertension, high cholesterol, MI/CABG or stroke (all yes/no). In model 3, annual household income (<$15,000; $15,000-$24,999; ≥$25,000) and education (< high school, high school/vocational training/junior college, college degree or higher) were added to investigate the potential mediating effects of individual-level SES. Knots were placed at quantiles of covariate distributions, equally spaced in sample size [127].

We conducted a sensitivity analysis (using similar multivariable models) excluding HF cases diagnosed within 2 years of follow-up. The exclusion of incident HF within the first 2 years of follow up limits the possibility that HF cases occurring soon after enrollment in the SCCS may not have been influenced by baseline values of neighborhood deprivation.

Deprivation Index and Post-HF Survival

Multivariable-adjusted Cox analyses as described previously for HF incidence were utilized to model the relationship between deprivation index and post-HF survival. Model assumptions were verified using Schoenfeld residuals and log (-log) plots.

All analyses were performed using STATA (version 12.1, Stata Corp, College Station, Texas, USA) and the 'rms' package for R version 3.1.1 (R Core Team 2014) [144, 145].

Ethical approval

Participants enrolled in SCCS provided written informed consent, and protocols were approved by the Institutional Review Boards of Vanderbilt University Medical Center and Meharry Medical College.

Results

Characteristics of the study population

Among the 27,078 SCCS participants with CMS data, 260 (0.96%) had missing values for deprivation index and were therefore excluded from these analyses. Baseline characteristics of the 26,818 included participants are shown in **Table 1**, overall and by tertiles of deprivation index. Tertile cut-points were based on the distribution of deprivation index at the census tract-level (not individuals). Tertile 1 represents the least deprived census tracts (i.e. the neighborhoods with the most community resources), while tertile 3 represents the most deprived. The inequality in the number of individuals per tertile of deprivation index is explained in part by the wide variation in the number of persons per census tract (mean = 6; range: 1-243) and that census tracts with higher deprivation tended to have more individuals per census tract, i.e. more persons lived in the most disadvantaged neighborhoods.

The mean (SD) age of the study participants at cohort enrollment was 55.5 (10.4) years, 62.7% were women, 69.0% were black, 69.9% had annual household income < $15,000, 38.6% had less than a high school education and 44.8% were obese (BMI $\geq$ 30 kg/m^2). There was a modest negative correlation between neighborhood deprivation and annual household income (Spearman's rank correlation coefficient, ρ = - 0.23) as well as education (ρ = - 0.17). Compared with persons living in the least deprived neighborhoods (tertile 1), participants living in the most deprived neighborhoods (tertile 3) were more likely to be younger, black, and obese, have less than a high school education, earn less than $15,000 a year, and be current smokers. They were also more likely to report a history of diabetes and hypertension at baseline, but they were less likely to report a history of MI or high cholesterol.

Table 1. Baseline characteristics of SCCS participants receiving Medicare or Medicaid during follow-up between 2002 and 2010, overall and by tertile‡ of deprivation index

	Deprivation Index			
	Overall	Tertile 1	Tertile 2	Tertile 3
	0.85 (1.21)	-0.76 (0.34)	0.03 (0.21)	1.61 (0.93)
Census tracts, n	4,666	1, 556	1, 555	1, 555
Participants, n	N = 26, 818	n = 4, 256	n = 6, 478	n = 16, 084
Age, years (SD)	55.5 (10.4)	58.1 (10.7)	56.8 (10.4)	54.2 (10.1)
Age quartiles, %				
40-46	24.8	19.0	20.3	28.1
47-53	26.6	22.0	24.7	28.6
54-64	22.9	21.0	24.8	22.6
≥ 65	25.7	38.0	30.2	20.7
Women, %	62.7	60.0	64.6	62.6
Blacks, %	69.0	37.3	49.8	85.2
Education %				
< High school	38.6	25.5	36.4	42.9
HS/Junior college/VT§	53.1	57.0	55.0	51.3
≥ College degree	8.3	17.5	8.6	5.8
Annual Income, %				
< $15,000	69.9	53.3	65.0	76.2
$15, 000-24, 999	17.9	18.2	20.6	16.8
≥$25, 000	12.2	28.5	14.4	7.0
Smoking, %				
Never	34.6	34.6	36.6	33.9
Former	25.2	32.7	28.4	22.0
Current	19.7	32.7	35.0	44.2
Alcohol Intake, %				
0 drink per day	55.0	53.6	62.1	52.5
>0-2 drinks per day	33.1	36.8	29.8	33.5
>2 drinks per day	11.9	9.6	8.1	14.0
Physical exerciseº, met-hrs/day, (SD)	0.87 (2.3)	1.12 (2.6)	0.77 (2.1)	0.85 (2.3)
BMI, kg/m^2 (SD)	30.4 (7.8)	29.9 (7.4)	30.7 (7.6)	30.4 (7.9)
BMI Categories, %				
Underweight, BMI < 18.5	1.7	1.4	1.6	1.7
Lean, BMI 18.5 - < 25.0	24.0	24.6	21.3	24.9
Overweight, BMI 25 - < 30.0	29.5	32.5	30.4	28.4
Obese, BMI ≥ 30.0	44.8	41.5	46.7	45.0
Diabetes, %	26.5	24.5	28.1	26.4
Hypertension, %	62.5	58.4	63.4	63.3
High Cholesterol, %	39.5	45.4	44.3	35.9
History of MI, %	8.6	10.0	10.1	7.6
History of Stroke, %	9.6	9.4	9.9	9.5

* Other than for physical activity ($p = 0.66$), all comparisons between tertiles of deprivation index were significant ($p < 0.0001$ for all comparisons). ºPhysical Exercise = Total moderate and vigorous exercise in MET-hours.

‡ Tertile cut-points were based on the distribution of deprivation index at the census tract-level (not individuals). The 3rd tertile (with higher mean deprivation index) is the most deprived i.e. represents census tracts with the least community resources while tertile 1 is the most affluent. The inequality in the number of individuals per tertile of deprivation index is explained in part by the significant variation in the number of persons per census tract and the fact that census tracts with the higher deprivation had more individuals per census tract i.e. there were more persons living in the most disadvantaged neighborhoods.

§VT: Vocational training.

Deprivation Index and HF incidence

Over a median (25^{th}, 75^{th} percentile) follow-up time of 5.2 (3.2, 6.8) years, 4,300 participants (16%) developed incident HF. SCCS participants in the 3^{rd} tertile of deprivation index had the highest cumulative incidence of HF, 17%, compared with 13% and 15.7% for persons in tertiles 1 and 2, respectively.

Figure 1 shows a graph of the log relative hazard ($X\beta$) of HF plotted against deprivation index. Among whites, the log relative hazard of HF rises sharply with increasing deprivation index then levels-off after a deprivation index of approximately 0.5. Among blacks on the other hand, the curve has a more gradual slope and it plateaus at higher values of deprivation index (approximately 2.0).

Table 2 shows the risk of incident HF associated with neighborhood deprivation index adjusted for relevant covariates in a sequential fashion (models 1 to 3). Overall, after adjustment for age, sex and race, a 1 IQR increase in deprivation index was associated with a 14% increase in the risk of HF [HR= 1.14; 95% CI: 1.09 to 1.19]. Subsequent adjustment for lifestyle and clinical factors was associated with a minimal change in the point estimate [HR= 1.15; 95% CI: 1.05 to 1.21]. In the full model, further adjustment for the individual level SES factors of income and education (in addition to demographics, lifestyle and clinical factors) showed only a modest attenuation of the strength of the association; namely, a 1 IQR increase in deprivation index was associated with a 12% increase in the risk of HF [HR= 1.12; 95% CI: 1.07 -1.18]. The ICC was 4.8% [95% CI: 3.6-6.4], suggesting that 4.% of the variance in HF risk was explained by neighborhood deprivation.

The race stratified analyses showed similar patterns of increased risk of HF per 1 IQR increase in deprivation index. In the full models, there was a 20% increase in the risk [HR= 1.20; 95% CI: 1.07 - 1.34] of HF per 1 IQR increase in deprivation index among whites and an 11% increase among blacks [HR= 1.11; 95% CI: 1.05 -1.17], *p* for interaction = 0.0005.

In sensitivity analyses excluding HF cases diagnosed within 2 years of follow-up, the effect estimates were similar, with a HR of 1.13 [95% CI: 1.06 -1.20] for the overall cohort in fully adjusted models.

Fig 1. Plot of Log Relative Hazard (Xβ) for incident heart failure versus deprivation index among black and white individuals in the Southern Community Cohort Study.

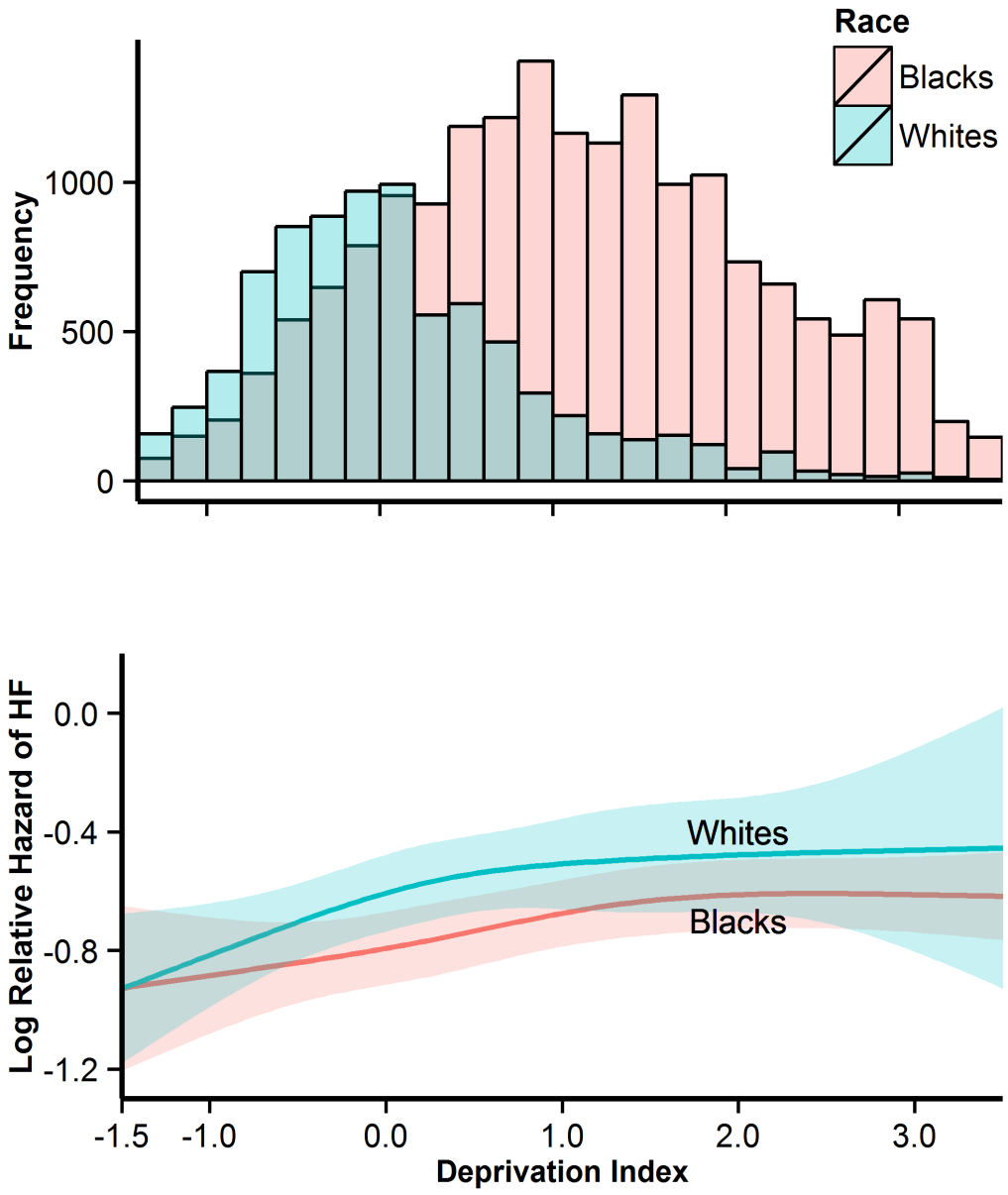


Table 2: Hazard ratios for incident heart failure per 1 interquartile range increase in deprivation index, overall and by race

	Overall N = 26, 818	Whites n = 8, 303	Blacks n = 18, 515
Model 1	1.14 (1.09, 1.19)	1.34 (1.21, 1.47)	1.09 (1.04, 1.15)
Model 2	1.15 (1.05, 1.21)	1.27 (1.14, 1.41)	1.12 (1.06, 1.18)
Model 3	1.12 (1.07, 1.18)	1.20 (1.07, 1.34)	1.11 (1.05, 1.17)

Data presented as hazard ratio (95% confidence interval). **Model 1**: Includes deprivation index, age (restricted cubic splines with 4 knots), race and sex. **Model 2**: Model 1 + smoking (never, former, current < 19.5 pack-years, current ≥ 19.5 pack-years), alcohol intake (linear) and Total moderate and vigorous sports in MET-hours (linear + quadratic) + BMI (restricted cubic splines with 4 knots), diabetes (yes/no), hypertension (yes/no), high cholesterol (yes/no), history of MI (yes/no) and history of stroke (yes/no). **Model 3**: Model 2 + annual household income (<$15000, $15000-$24999 & ≥$25000) and education (< high school, high school/vocational training/junior college, college degree or higher. P-value for deprivation index×race interaction = 0.0005.

Intra-class correlation coefficient = 4.8% (95% CI: 3.6, 6.4)

CI: Confidence interval; HR: Hazard ratio; HF: Heart failure; BMI: Body mass index

Deprivation Index and Post-HF Survival

Among the 4,300 SCCS participants who developed incident HF, 940 died (cumulative mortality = 21.9%) over a post-HF median (25^{th}, 75^{th} percentile) follow-up time of 2.3 (0.9, 4.2) years. Participants in the 3^{rd} tertile of deprivation index had the highest cumulative mortality, 22.9% compared with 21.2% and 19.6 for persons in tertiles 1 and 2, respectively.

Figure 2 shows a graph of the log relative hazard (Xβ) of post-HF mortality plotted against deprivation index. While the initial portions of the plots of log relative hazard of post-HF mortality against deprivation index in blacks and whites were suggestive of a gradual slope, they were mostly flat for the range of observed values of deprivation index.

Overall, in the full model adjusted for demographics, individual SES, lifestyle and clinical factors at baseline, a 1 IQR increase in deprivation index was associated with a nonsignificant 5% increase in the risk of all-cause mortality post-HF [HR= 1.05; 95% CI: 0.95 -1.17] (**Table 3**). A similar finding was

observed among blacks [HR= 1.07; 95% CI: 0.94 -1.21] while the hazard ratio was close to unity among whites [HR= 1.00; 95% CI: 0.79 -1.25].

Fig 2. Plot of Log Relative Hazard (Xβ) of post-heart failure death versus deprivation index among black and white individuals in the Southern Community Cohort Study.

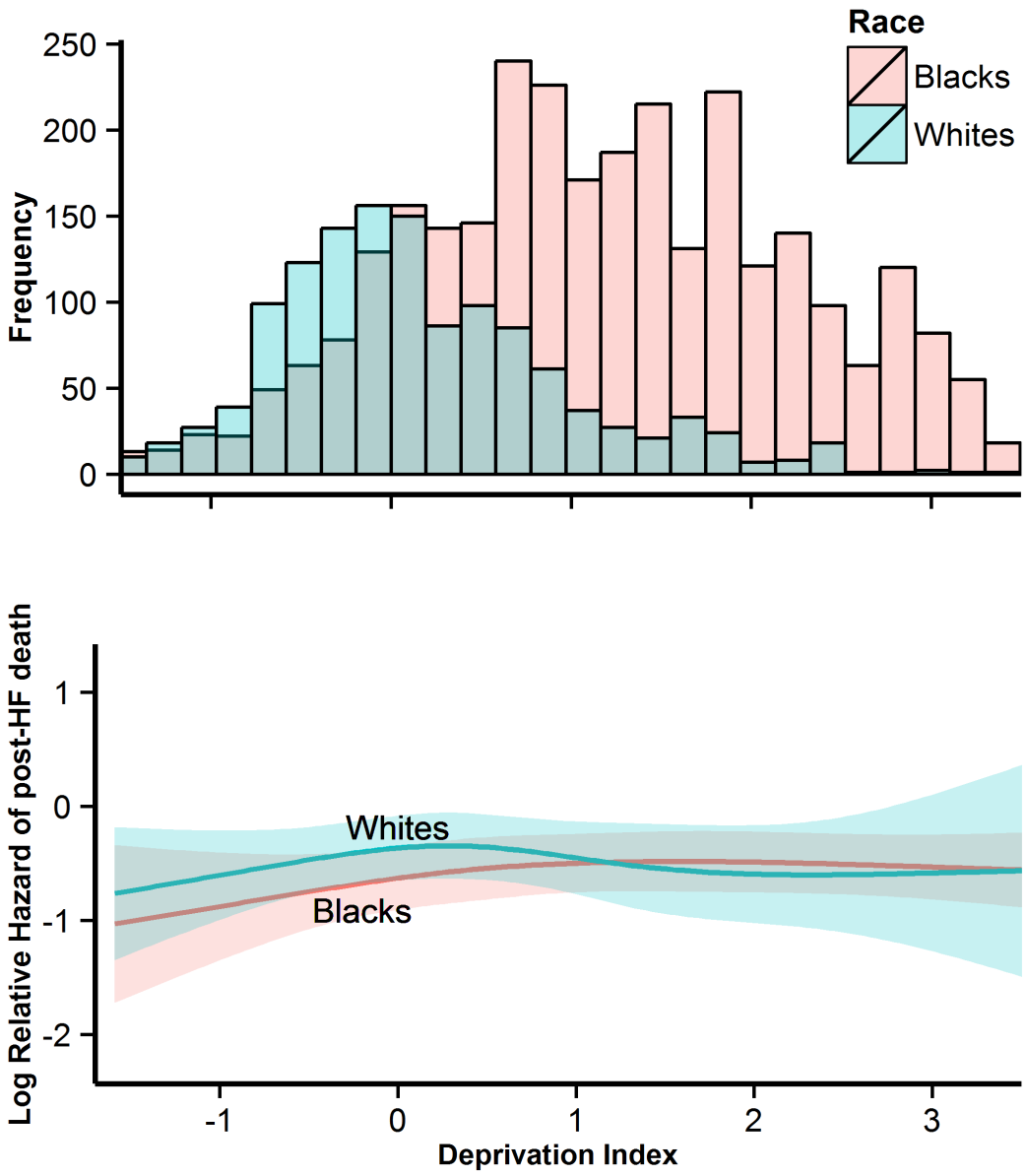


Table 3: Hazard ratios for post-heart failure death per 1 interquartile range increase in deprivation index, overall and by race

	Overall N = 4, 300	Whites n = 1, 290	Blacks n = 3, 010
Model 1	1.11 (1.00, 1.23)	1.11 (0.90, 1.38)	1.10 (0.98, 1.24)
Model 2	1.07 (0.97, 1.19)	1.05 (0.84, 1.31)	1.08 (0.96, 1.22)
Model 3	1.05 (0.95, 1.17)	1.00 (0.79, 1.25)	1.07 (0.94, 1.21)

Data presented as hazard ratio (95% confidence interval). **Model 1**: Includes deprivation index, age (restricted cubic splines with 4 knots), race and sex. **Model 2**: Model 1 + smoking (never, former, current < 19.5 pack-years, current ≥ 19.5 pack-years), alcohol intake (linear) and Total moderate and vigorous sports in MET-hours (linear + quadratic) + BMI (restricted cubic splines with 4 knots), diabetes (yes/no), hypertension (yes/no), high cholesterol (yes/no), history of MI (yes/no) and history of stroke (yes/no). **Model 3**: Model 2 + annual household income (<$15000, $15000-$24999 & ≥$25000) and education (< high school, high school/vocational training/junior college, college degree or higher. P-value for deprivation index×race interaction = 0.96.

Intra-class correlation coefficient = 5.7% (95% CI: 2.6, 12.1)

CI: Confidence interval; HR: Hazard ratio; HF: Heart failure; BMI: Body mass index

Discussion

We investigated the association between neighborhood deprivation index and HF incidence as well as post-HF mortality in a large population of low-income blacks and whites from the southeastern United States. Our main findings were: 1) persons living in the most deprived neighborhoods appear to have a greater burden of CVD risk factors, 2) higher levels of neighborhood deprivation are significantly associated with an increase in HF risk independent of individual SES and traditional CVD risk factors, and 3) there is no strong evidence of an association between neighborhood deprivation and post-HF mortality.

The existing literature provides evidence of a strong independent association between individual socioeconomic characteristics, such as income, education and occupation, and HF risk [9, 10, 105, 146]. In addition, recent data suggests that neighborhood SES also plays a significant role in predicting HF

outcomes including HF admissions and readmissions, [11, 152] but evidence for the association with HF incidence has been scant. The current investigation demonstrates an independent association between increasing neighborhood deprivation and increased risk of HF with a non-negligible proportion of the variance (~ 5%) of HF incidence in this population explained by neighborhood socioeconomic factors. Furthermore, the dose-response curve indicates increasing HF risk with increasing levels of neighborhood deprivation in both blacks and whites before the curve plateaus. Neighborhood factors have been shown to predict the incidence of coronary heart disease (CHD) [136, 137, 153] and we now extend this to heart failure. Further, given that our study was conducted within a population with relatively low individual SES, under-represented in previous studies, it is particularly noteworthy to find that a dearth of community-level resources further compounds the risk of HF in this population.

The neighborhood deprivation index utilized for this study was a composite obtained from 11 components using principal component analysis. Thus, it remains uncertain what specific neighborhood characteristics are responsible for our findings, but several hypotheses may explain the impact of scant community resources on health outcomes. The availability of exercise facilities, healthy food outlets, institutional resources (including healthcare facilities), and tobacco advertising vary considerably across neighborhoods [154-156]. If food deserts are more preponderant in the most deprived neighborhoods, that could reduce access to healthier food choices and potentially increase the consumption of high-calorie foods and foods with high sodium content. Also, the combination of fewer physical activity resources, unaffordable gym memberships, and higher crime rate may predispose persons living in these communities to reduced physical activity and sedentariness. The combined influence of poor nutritional habits and physical inactivity could explain the higher rates of obesity that have been observed in poverty-dense counties in the US [12, 157]. The high rates of obesity are paralleled by high prevalence of obesity-related comorbid conditions including diabetes and high blood pressure (possibly abetted in part by the consumption of foods with high salt content) which further compound the risk of HF in these communities.

In the current cohort, compared with the neighborhoods with the most resources, the more deprived neighborhoods each had higher prevalence of obesity, hypertension and diabetes at baseline and this may have contributed to the observed trend of increasing HF risk with worsening deprivation. But the similar prevalence of these risk factors across the two upper tertiles of neighborhood deprivation may explain in part the plateau observed in the dose-response curve for HF incidence after an initial increase. However, we did observe higher proportions of persons with previous history of MI in the neighborhoods with the most resources which may have been due in part to an overrepresentation of older persons and whites (who have a higher risk of MI) in these areas. Nevertheless, after adjustment for all these factors, a strong independent association between neighborhood deprivation and HF risk persisted, which could not be explained by mediation via CVD risk factors or individual SES. This suggests that some of the correlates of neighborhood deprivation may be less tangible and harder to measure. Some authors have hypothesized that persons living in deprived neighborhoods may be at higher risk of unfavorable societal stressors like noise, air pollution and violence which may culminate in chronic psychological stress and predispose individuals to adverse health outcomes including HF [153, 158]. Marked institutional deficiencies in resource-limited settings are usually mirrored by reduced access to quality education, occupational opportunities and health facilities. These could expose individuals in these communities to reduced scholarship, income and logistics and hence curtail individual ability to seek preventative care, self-management and adherence to recommended treatment guidelines (evidence-based lifestyle strategies and multifactorial medical management approaches) for conditions such as diabetes and hypertension – predisposing them to elevated risk of HF. Finally, reduced social ties and community perceptions about health in low SES communities could mitigate health-seeking behaviors including early screening for diabetes and hypertension, adherence to therapy and holistic management of these chronic conditions which is pivotal to improve outcomes.

While previous data found an association between neighborhood factors and mortality in the general population, [159] among persons with HF, we found no evidence of an association with all-cause

mortality. Meanwhile, significant associations have been observed between neighborhood characteristics and lower adherence to therapy, worse quality of care, and worse outcomes (including rehospitalizations) among patients with HF [160-167]. As these intermediate outcomes may be correlated with death among persons with HF, we may have expected to find a significant association between neighborhood SES and post-HF mortality. However we had a relatively limited post-HF follow-up time (~2.3 years) and it is plausible that a longer time is required to notice the impact of socioeconomic environment on long-term outcomes such as death. In future studies, it may be relevant to re-visit the association with all-cause mortality and investigate the association with cardiovascular death after a more significant post-HF follow-up time.

The current study has a few noteworthy limitations. Some authors consider a 'neighborhood' to be a spatially defined collection of people, infrastructures, and institutions influenced by common environmental, sociocultural, and economic forces [168]. The extent to which a census tract is a rational proxy for 'neighborhood' remains uncertain. However, as a relatively permanent statistical subdivision of a county or equivalent entity, census tracts cover a contiguous area with long term boundaries, and harbor an optimum population close to 4000 persons on average. Thus, using them as proxies for neighborhoods appears reasonable for the purpose of investigating the health outcomes of a population with shared socioeconomic environment. Most of our covariate data (past history of MI, stroke, hypertension, diabetes and high cholesterol) were based on self-report of a physician diagnosis and use of medications. While self-report could be susceptible to recall and misclassification bias, these methods have been validated in the SCCS as well as other epidemiologic studies [18]. Several of the questions on the SCCS questionnaire were adapted from questionnaires that were validated in other settings; and a series of independent validation studies using biomarkers, repeat interviews or medical records have demonstrated the reliability of the questionnaire within the SCCS population for variables such as smoking status and self-reported diseases including diabetes [18]. HF was ascertained by linking our baseline SCCS data with CMS Research Identifiable Files using ICD-9 codes 428.x, rather than independent physician

adjudication. Nonetheless, the diagnosis codes algorithm for identification of HF used in this study has been previously validated and utilized in other cohorts [119-121]. A review of the detection of HF in administrative claims data that included studies conducted among Medicare beneficiaries reported positive predictive values (PPVs) mostly over 90% [119].

Our study leverages data from a large biracial cohort with a sizable number of low-income participants living in resource-limited settings and who are traditionally under-represented in previous cohorts investigating HF outcomes. Thus, it provided the unique opportunity to investigate the role of neighborhood factors on HF risk and post-HF survival in a sample of people already having scant individual resources. The ability to perform geocoding of participant residential information and linkage to geographic information datasets such as census tract data allowed development of residence-specific metrics including the SCCS-derived deprivation index which we were able to utilize as a proxy for neighborhood SES. The large number of level-2 units (census tracts) covered by the study participants allowed for stable statistical analyses and provided some credence to the representativeness of the whole SCCS cohort by the sub-sample utilized for the current analysis.

In conclusion, we found that neighborhood socioeconomic factors significantly predict HF incidence independent of individual income and education level and traditional CVD risk factors, but longer follow-up may be needed to examine the association with post-HF survival. The American Heart Association and other cardiovascular societies recognize that improvements in cardiovascular health requires strategies that target the entire spectrum of the healthcare system including public policy, prevention, acute care, chronic care and rehabilitation. However, the more "upstream measures" which focus on public policy and prevention may have the greatest potential to mitigate the burden of CVD and improve human health. Areas with the most acute socioeconomic deprivation are most likely at the highest risk for CVD (including HF) and CVD mortality and hence may benefit most from such improvements in public health policies including, but not limited to, improvements in community-level resources.

Surrogate measures of obesity predict heart failure risk and post-HF survival among low-income blacks and whites in the Southern Community Cohort Study

Authors: Elvis A. Akwo, MD, MS[1,5]; Frank E. Harrell Jr., PhD[3, 5]; Edmond K. Kabagambe, DVM, PhD[1, 5]; Thomas J. Wang, MD[2, 5]; William J. Blot, PhD[1, 4]; Michael Mumma, MS[4]; Deepak K. Gupta, MD* [2, 5]; Loren Lipworth, ScD* [1, 5]

[1] Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, 37203.

[2] Division of Cardiovascular Medicine, Vanderbilt Heart & Vascular Institute, Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, 37203.

[3] Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, 37203.

[4] International Epidemiology Institute, Rockville, Maryland, 20850.

[5]Vanderbilt Translational and Clinical Cardiovascular Research Center (V-TRACC), Vanderbilt University School of Medicine, Nashville, Tennessee, 37203.

* Equal contribution

Address correspondence to:
Dr. Loren Lipworth
Division of Epidemiology
Vanderbilt University School of Medicine
2525 West End Avenue, Suite 600
Nashville, TN 37203
Phone: 615-343-0639
Fax: 615-343-5938
loren.lipworth@vanderbilt.edu

Word Count: Manuscript (5232), abstract (401).

Running Title: Weight-height indices and heart failure.

AHA Journal Subject Terms: [Obesity] [Heart failure]

Abstract

Background

Previous studies have used body mass index (BMI) as the default weight-height index for the prediction of heart failure (HF) risk and mortality and have assumed linearity of effects or utilized BMI categories. We sought to propose an empirically-derived weight-height index (W/H^n) and investigate the dose-response relationship between weight-height indices (as proxies for total adiposity) and HF risk as well post-HF mortality.

Methods

Study participants were black and white enrollees (n = 27,078) of the Southern Community Cohort Study with no history of HF at baseline and who were receiving Centers for Medicare and Medicaid Services (CMS). Incident HF diagnoses through December 31, 2010 were ascertained using ICD-9 codes 428.x via linkage with CMS research files. For W/H^n, n was defined as the ratio of the coefficients of the log height and log weight parameters from a Cox model for the log relative hazard of HF and post-HF mortality. Cox models were used to investigate the dose-response relationship between surrogate measures of adiposity – W/H^n, BMI and waist circumference (WC) – and both HF risk and post-HF mortality.

Results

For incident HF, n was found to be 1.81 among whites and 1.21 among blacks while for post-HF mortality, the corresponding values were 4.02 and 2.68 respectively. In models for predicting HF risk or post-HF mortality, log W/H^n yielded better model fit (smaller AIC) compared to log BMI especially among blacks. After full adjustment for baseline covariates, a 1 interquartile range (IQR) increase in W/H^n remained associated with a significant 31% increase in the risk of incident HF [HR = 1.31; 95% CI: 1.22-1.39] among whites and a 33% increase in risk [HR= 1.33; 95% CI: 1.27-1.39] among blacks. Similar findings were observed for BMI. For WC, the HR was 1.66 [95% CI: 1.29 -2.14] among whites and 1.46 [95% CI: 1.20 -1.77] among blacks. The plot of log relative hazard of HF vs. BMI was J-shaped;

the nadir for the curve for blacks occurred at a higher BMI (30 versus 25kg/m^2). There was an inverse nonlinear dose-response trend between post-HF mortality risk and BMI which mostly plateaus beyond a BMI ≈ 30kg/m^2.

Conclusions

Empirically-derived weight-height indices offered a better model fit for the prediction of incident HF and post-HF mortality compared to BMI. There were unique nonlinear dose-response patterns observed between BMI and both HF risk and post-HF mortality that could potentially inform current clinical guidelines or recommendations regarding risk stratification.

Introduction

Data from a variety of communities across the US suggest that poverty-dense counties have high levels of obesity which are paralleled by high prevalence of cardio-metabolic conditions including cardiovascular disease (CVD) and heart failure (HF) [12]. Thus, it is important to critically examine the link between obesity and HF particularly among populations with scant resources.

Prior epidemiologic evidence is suggestive of an association between excess body weight and increased risk of HF [13-15] as well as a contrasting decrease in the risk of post-HF mortality – a phenomenon coined as the obesity paradox [109, 111, 112, 114, 117] However, most studies investigating these relationships utilized categories of body mass index (BMI) or assumed linearity of effects thereby limiting the elucidation of the natural dose-response relationship between measures of obesity and HF risk as well as post-HF survival.

More importantly, less thought has been given to the suitability of BMI – weight (W)/height $(H)^2$ – in investigating the link between obesity and both HF risk and post-HF survival despite differences in the performance of various weight-height indices across population groups defined by race and sex [16, 17]. Additionally, predictors with "pleiotropic" effects usually have differential functional relationships with varying outcomes suggesting the need to use the data to empirically derive an appropriate weight-height index for each outcome. Such approaches may be utilized to adequately model the intricacies in these data and reveal novel insights that may improve our understanding of anthropometry and general obesity in relation to HF risk and post-HF survival.

The Southern Community Cohort Study (SCCS) [18] provided the opportunity to investigate: a) the appropriate functional form of a data-derived weight-height index (W/H^n) for the association with incident HF and post-HF survival, and b) the dose-response relationships between W/H^n (as a surrogate measure of general obesity) and waist circumference (WC, a surrogate for visceral fat), with incident HF and post-HF survival in a large sample of blacks and whites drawn from resource-limited settings.

Methods

Design and Study Population

The SCCS is a unique ongoing prospective investigation tracking a population of 84,797 adults, two-thirds black, aged 40-79, recruited in 12 southeastern states between 2002 and 2009 to investigate the incidence of cancer and other chronic diseases [18, 148] Over 86% of the participants in the SCCS were identified from community health centers (CHC), institutions providing primary health and preventative services mainly to low-income populations [18, 20], so that the cohort is made up of a segment of society (minority, poor, rural) seldom included in sizeable numbers in previous cohort studies; particularly those investigating CVD. The remaining 14% were recruited via mail-based sampling of the general population. Data on demographic, socioeconomic, lifestyle, and anthropometric characteristics, as well as personal medical history, were obtained at cohort enrollment using standardized computer-assisted personal interviews for CHC participants, and via self-administered mailed questionnaire for persons recruited from the general population. Detailed description of SCCS methods has been previously published [18, 20, 107].

Participants (n = 27,078) included in the current analyses were SCCS enrollees aged ≥ 65 (n = 7001) at cohort enrollment, or persons < 65 years (n = 20,077) at enrollment who: a) reported being covered by Medicaid (which provides medical benefits to low-income adults and uninsured persons) on the baseline questionnaire; or b) reported being covered by Medicare (the primary health insurance program for persons aged ≥ 65) on the baseline questionnaire; or c) did not report Medicare or Medicaid on the baseline questionnaire but had a Centers for Medicare and Medicaid Services (CMS) claim within 90 days of being enrolled in SCCS. The restriction to these groups maximizes the likelihood that participants would have continuous coverage in Medicare and/or Medicaid from the time of SCCS enrollment to the end of the follow-up period (December 31st, 2010), for the ascertainment of incident HF events. Analyses were restricted to self-reported African American or black and non-Hispanic white

SCCS participants, since too few persons in other racial groups were available for stable statistical analysis.

Assessment of anthropometric data

SCCS participants reported their height and weight at baseline. These were validated using data from a random sample (n ≈ 14,000) of SCCS participants for whom measured weight and height were either a) abstracted from contemporaneous CHC medical records or b) obtained via measurements performed by trained interviewers using a SECA 703 digital scale and a stadiometer on the day of the interview [107] There was a very high correlation ($r > 0.95$) between measured and self-reported weight and height [18]. The latter were used to compute W/H^n and BMI – weight (kg)/height $(m)^2$ – the main predictors for the current analyses.

Waist and hip circumferences were measured for a subset of the cohort using a standardized protocol with a tape measure over a single layer of clothing [107] Of the 27,078 SCCS participants in the current study, 3304 had WC data.

Outcome ascertainment

Ascertainment of HF events was performed via linkage of the SCCS cohort with CMS Research Identifiable Files (which include Medicare institutional and non-institutional files, and the Medicaid Analytic Extract files). Incident HF was defined as the first occurrence of a medical claim with an International Classification of Diseases, 9th revision, discharge code of 428.x (428.0 to 428.9) within the Medicare institutional (Medicare Provider Analysis and Review, MEDPAR, which includes inpatient, outpatient and skilled nursing facility base files), Part B carrier (includes non-institutional physician services and durable medical equipment), or outpatient-based claims files or the Medicaid Analytic Extract (MAX) Inpatient and Other Services claims files, from the date of SCCS enrollment through December 31st, 2010. Detailed description of the CMS research files are published elsewhere.[143]

Deaths, including dates and causes of death, were ascertained via linkage of the SCCS cohort with both the Social Security Administration (SSA) vital status service for epidemiologic researchers and the National Death Index (NDI) through December 31st, 2010. Both NDI and SSA are well-established and reliable means of identifying deaths in the US, and are expected to capture nearly all deaths [51, 122, 123].

Statistical Analysis

Descriptive statistics (means and standard deviations for continuous variables and counts and percentages for categorical variables) were computed for all study participants, overall and by race.

For all participants included in the current study, duration of follow-up for incident HF was computed from date of enrollment into the SCCS until the date of the first diagnosis of HF, date of death, or December 31, 2010, whichever occurred first. For HF cases, post-HF follow-up time was defined as time from HF diagnosis to death or December 31st 2010 whichever occurred first. When date of death was coincident with date of HF diagnosis, follow-up time was set to 0.5 days.

To investigate the association between W/H^n and incident HF, the following analyses were performed in a step-wise fashion.

a. Deriving the value of n and computing W/H^n.

The log relative hazard of incident HF, *Y*, regressed on the logarithm of a generic weight-height index (W/H^n) in a Cox model (with no intercept) could be expressed as follows:

$$Y = \log W/H^n, \text{ assuming a slope of 1.}$$

$$Y = \log W + \log H^{-n}$$

$$Y = \log W - n\log H$$

So the absolute value of the ratio of the coefficients of log H and log W is n.

Thus, if a bivariate Cox model for the log relative hazard of incident HF, with the latter regressed on log W and log H yields α_1 and α_2 as their regression coefficients:

$$Y = \alpha_1 \log W + \alpha_2 \log H ,$$

where Y = log relative hazard (incident HF), then n is given by the absolute value of the ratio of the coefficients:

$$n = |\alpha_2 / \alpha_1 |$$

This approach was used to obtain the value of n and compute W/H^n for all participants included in the current analyses.

b. Comparing the performance of W/H^n and BMI in HF risk prediction

First, we constructed a bivariate Cox model for the log relative hazard of HF comprising the natural log of BMI and log height, both modelled as restricted cubic splines with 5 knots. A Chunk test was performed for the linear and nonlinear terms of log height at alpha of 0.05 to test if the height variable contributes significant additional information to the single-predictor model containing BMI.

Second, in separate Cox models, we regressed the restricted cubic splines of the natural log of W/H^n and that of BMI on the log relative hazard of HF. Then, model fit statistics (LR chi square, χ^2 and AIC) were computed and used to compare the performance of W/H^n versus that of BMI vis-à-vis a model utilizing restricted cubic splines of both log weight and log height as the latter was expected to preserve the most information.

c. Investigating the association between W/H^n and HF risk

Multivariable Cox models accounting for nonlinearity and non-additivity were utilized to model a flexible dose-response association between W/H^n (modelled using restricted cubic splines with 5 evenly spaced knots) and HF risk adjusting for relevant covariates. Model 1 included W/H^n, age at enrollment (restricted cubic splines with 4 knots), race (white/black), sex (men/women) and W/H^n×race interaction

terms. Model 2 additionally adjusted for annual household income (<$15,000; $15,000-$24,999; ≥$25,000); education (< high school, high school/vocational training/junior college, college degree or higher), smoking (never, former, current <19.5 pack-years, current ≥ 19.5 pack-years, 19.5 being the median pack-years among current smokers), alcohol intake (linear) and total physical activity in met-hours (linear + quadratic term). Model 3 additionally adjusted for the following covariates: history of diabetes, hypertension, high cholesterol, MI/CABG or stroke (all yes/no). Knots were placed at quantiles of covariate distributions, equally spaced in sample size.[127] Plots of log relative hazard of incident HF versus W/H^n stratified by race were constructed based on data from the variable-rich model.

Similar multivariable Cox models (models 1 through 3) were used to estimate the effect of an interquartile range (IQR) increase in W/H^n (modelled as a linear IQR-rescaled predictor) on incident HF. The hazard ratio (HR) and 95% confidence interval (CI) for a 1 interquartile range increase in W/H^n (as a continuous measure) compares the hazard of the event occurring for a typical person in the middle of the upper half (the 75^{th} percentile) of the distribution of W/H^n to the hazard of the event for a typical person in the middle of the lower half of the distribution. P-values for the interactions between weight-height indices and race were computed.

Steps a through c were repeated for the association between W/H^n and post-HF mortality.

d. Investigating the association between WC and HF risk

Similar analyses (models, effect estimation and plots) were performed for the association between WC and HF risk. However, given the reduced amount of data available for the analyses involving WC, models 1 through 3 were simplified to reduce the degrees of freedom spent and preserve power.

Model 1 included WC (restricted cubic splines with 4 knots), age (restricted cubic splines with 4 knots), race (white/black), sex (men/women) and WC×race interaction terms. Model 2 additionally adjusted for annual household income (<$15,000/≥$15,000); education (< high school, high school/vocational training/junior college, college degree or higher), smoking (never, former, current),

alcohol intake (linear) and total physical activity in met-hours (linear). Model 3 additionally adjusted for the following covariates: history of diabetes, hypertension, high cholesterol, MI/CABG or stroke (all yes/no).

All analyses were performed using STATA (version 12.1, Stata Corp, College Station, Texas, USA) and the 'rms' package for R version 3.1.1 (R Core Team 2014) [144, 145].

Ethics statement

Participants enrolled in SCCS provided written informed consent, and protocols were approved by the Institutional Review Boards of Vanderbilt University Medical Center and Meharry Medical College.

Results

Characteristics of the Study Population

Table 1 shows the distribution of baseline characteristics of the 27, 078 participants included in the current study, overall and by race. The mean (SD) age of the study participants at cohort enrollment was 55.5 (10.4) years; 68.8 % were black, 62.6% were women, 69.7% had annual household income < $15,000, 38.4% had less than a high school education and 44.8% were obese (BMI $\geq 30 Kg/m^2$). Overall, black SCCS participants were younger, more likely to be obese and to have less than a high school education, less than $15,000 annual income, a past history of high blood pressure and diabetes at baseline. White SCCS participants on the other hand were more likely to be heavy smokers and report a past history of high cholesterol, myocardial infarction and stroke at baseline.

Table 1. Comparison* of baseline characteristics of 27, 078 SCCS participants who were receiving Medicare or Medicaid during follow-up between 2002 and 2010, by race

	Overall	Whites	Blacks
	n = 27,078	n = 8,454	n = 18,624
Age (SD), years	55.5 (10.4)	58.0 (10.6)	54.3 (10.1)
Age quartiles %			
40-46	24.6	18.2	27.5
47-53	23.7	19.9	25.4
54-64	25.9	25.6	26.0
≥ 65	25.8	36.3	21.1
Women %	62.6	62.1	62.8
Education %			
< High school (HS)	38.4	30.9	41.8
HS/Junior college/VT**	53.1	56.5	51.6
≥ College degree	8.5	12.6	6.6
Annual Income < $15,000, %	69.7	61.1	73.5
Smoking %			
Never	34.7	31.3	36.2
Former	25.3	31.9	22.3
Current & pack-years < 19.5	19.6	9.3	24.3
Current & pack-years ≥ 19.5	20.4	27.6	17.2
Alcohol Intake %			
0 drink per day	54.9	60.0	52.7
>0-2 drinks per day	33.2	31.9	33.7
>2 drinks per day	11.9	8.1	13.6
Total Physical Activity (SD), met-hrs	17.2 (15.6)	16.6 (14.9)	17.4 (15.9)
BMI (kg/m^2) (SD)	30.4 (7.8)	30.0 (7.6)	30.6 (7.8)
BMI Categories %			
Underweight, BMI < 18.5	1.7	1.8	1.6
Lean, BMI 18.5 - < 25.0	24.0	25.3	23.4
Overweight, BMI 25 - < 30.0	29.5	31.2	28.7
Obese, BMI ≥ 30.0	44.8	41.7	46.3
History of MI %	8.6	12.1	7.0
History of Stroke %	9.6	10.3	9.2
Diabetes %	26.5	24.1	27.6
Hypertension %	62.5	56.9	65.0
High Cholesterol %	39.5	48.6	35.4

* Other than for sex (p = 0.32), all comparisons between racial groups were statistically significant (p = 0.004 for stroke, p = 0.0001 for total physical activity and p<0.0001 for all other baseline variables).

**VT: Vocational training.

Empiric weight-height indices used to investigate HF risk and Post-HF mortality

For incident HF, the calculated exponent of height, n, in the empiric weight-height index (W/H^n) was found to be 1.81 among whites and 1.21 among blacks while for post-HF mortality, the corresponding values were 4.02 and 2.68 respectively (**Table 2**). Table 2 shows the distribution of the empiric weight-height indices and their correlations with BMI and WC. The W/H^n index for the association with HF risk was strongly correlated with WC (and BMI) in both whites and blacks while BMI appeared weakly correlated with WC among blacks. Similar patterns (correlations with BMI) were observed for empiric weight-height indices used to investigate post-HF mortality but there were limited data to compute reliable correlation coefficients with WC.

Table 2: Empiric weight-height indices derived for both Incident HF and post-HF mortality in SCCS

	Incident HF				Post-HF mortality			
	Whites n* = 8,454		Blacks n* = 18,624		Whites nº = 1,312		Blacks nº = 3,029	
	BMI	$W/H^{1.81}$	BMI	$W/H^{1.21}$	BMI	$W/H^{4.07}$	BMI	$W/H^{2.68}$
Range	13.2 – 78.3	14.5 – 86.8	11.4 – 80.7	15.3 – 121.5	14.0 – 78.3	4.4 – 29.0	14.5 – 80.7	9.9 – 56.7
Mean (SD)	30.0 (7.6)	33.1 (8.4)	30.6 (7.8)	46.3 (11.4)	32.1 (8.5)	11.4 (3.6)	32.6 (8.8)	22.9 (6.4)
Median (IQR)	28.5 (9.1)	31.6 (9.9)	29.3 (10.0)	44.5 (14.5)	30.6 (10.8)	10.6 (4.5)	31.2 (11.6)	22.0 (8.6)
Pearson's correlations (ρ)								
W/H^n §	0.99	1.00	0.98	1.00	0.91	1.00	0.99	1.00
WC♯	0.85	0.85	0.78	0.80				

* Out of the 27, 078 whites and blacks included in these analyses, 365 (1.35%) had missing weight and/or height data at baseline.

º Out of the 4,341 persons who were diagnosed with HF on follow-up, 73 (1.7%) had missing weight and/or height data.

♯ WC: waist circumference; data were available for 1406 White and 1962 Black participants. There were insufficient WC data among HF cases to compute correlations with empiric weight-height indices derived for the association with post-HF mortality.

§ For Incident HF, the exponent of height in W/H^n = 1.81 among whites and 1.21 among blacks. For post-HF mortality, the exponent of height in W/H^n = 4.02 among whites and 2.68 among blacks.

Model fit for models utilizing W/H^n to investigate HF risk and post-HF mortality

Model fit statistics for separate Cox models utilizing log weight + log height, log W/H^n or log BMI to predict HF risk and post-HF mortality are presented in **Table 3**. For predicting HF risk, using log W/H^n yielded better model fit statistics (smaller AIC and higher LR χ^2) compared to using log BMI especially among blacks. A similar pattern was observed for models predicting post-HF mortality in blacks and whites.

Table 3: Model fit statistics for models utilizing weight-height indices for predicting HF risk and post-HF mortality

	Incident HF		Post-HF Mortality	
	Whites		Whites	
Model*	LR χ^2	AIC	LR χ^2	AIC
Log W + Log	141.08	22274.60	57.55	3839.73
Log W/H^n	125.29	22282.40	45.51	3843.77
Log BMI	125.12	22282.56	38.91	3850.37
	Blacks		Blacks	
Model*	LR χ^2	AIC	LR χ^2	AIC
Log W + Log	224.46	55959.06	121.46	9488.58
Log W/H^n	218.01	55957.51	110.63	9491.42
Log BMI	207.75	55967.77	108.94	9493.11

*All variables were modelled using restricted cubic splines with 5 knots.

§ For Incident HF, the exponent of height in W/H^n = 1.81 among whites and 1.21 among blacks. For post-HF mortality, the exponent of height in W/H^n = 4.02 among whites and 2.68 among blacks.

Association between surrogate measures of total and visceral adiposity with HF risk

Over a median follow-up time of 5.2 (range: 0.1 – 8.9) years, 4,341 incident cases of HF were ascertained corresponding to a cumulative incidence of 16%. **Table 4** shows the HR (and 95% CI) for incident HF per 1 IQR increase in W/H^n and BMI adjusted for demographic, lifestyle and clinical factors in a sequential fashion (models 1 to 3). After adjustment for age and sex, a 1 IQR increase in W/H^n was associated with a 49% increase in the risk of HF [HR= 1.49; 95% CI: 1.41-1.58] among whites and a 43% increase [HR= 1.43; 95% CI: 1.37-1.49] in risk among blacks. For BMI, similar effect estimates were observed with HRs of 1.51 [95% CI: 1.42 -1.60] and 1.44 [95% CI: 1.38-1.51] among whites and blacks respectively. After full adjustment for relevant covariates, a 1 IQR increase in W/H^n remained associated with a significant 31% increase in the risk of incident HF [HR = 1.31; 95% CI: 1.22-1.39] among whites and a 33% increase in risk [HR= 1.33; 95% CI: 1.27-1.39] among blacks. Similar findings were observed for BMI.

For WC, the association with incident HF appeared stronger with HRs of 1.74 [95% CI: 1.38 - 2.18] among whites and 1.60 [95% CI: 1.36 -1.89] among blacks in the minimally-adjusted model and HRs of 1.66 [95% CI: 1.29 -2.14] among whites and 1.46 [95% CI: 1.20 -1.77] among blacks in the fully adjusted model (**Table 5**).

The plot of log hazard of HF versus W/H^n (on a standardized scale) appears to have a J-shape with lower values associated with increased risk of HF in both blacks and whites then both curves rise sharply after a nadir – which occurs earlier in the curve for whites compared to that of blacks (**Figure 1a**). The plot for BMI is almost identical to the former with the troughs found at $\approx 25 Kg/m^2$ among whites and $\approx 30\ Kg/m^2$ among blacks (**Figure 1b**). **Figure 1c** shows the plot for WC. Among blacks, the plot is suggestive of an ostensibly linear increase in log hazard of HF with increasing WC. Among whites, whilst the curve appears nonlinear with a trough at ≈ 90cm, in the region where we have the most data (and tighter confidence bands), the risk of HF increases in a seemingly linear fashion.

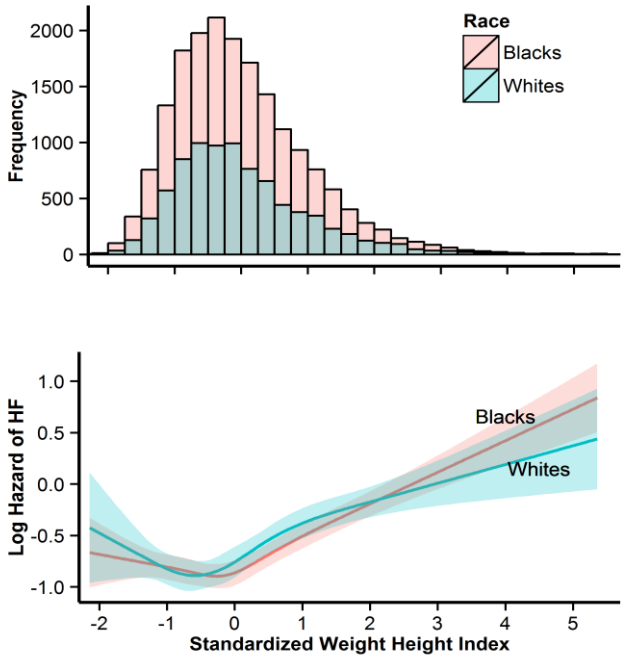


Fig 1a. The log relative hazard of HF plotted against a standardized weight-height index among Blacks and Whites in SCCS.

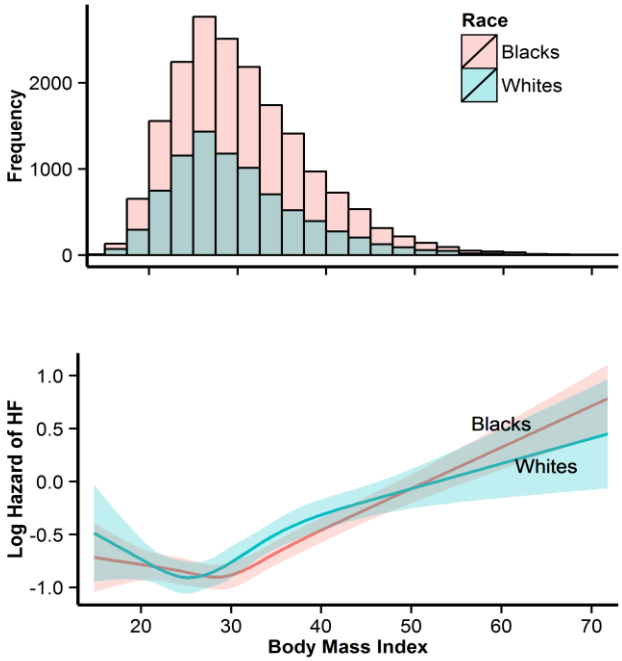


Fig 1b. The log relative hazard of HF plotted against BMI among Blacks and Whites in SCCS.

Fig 1c. The log relative hazard of HF plotted against waist circumference among Blacks and Whites in SCCS.

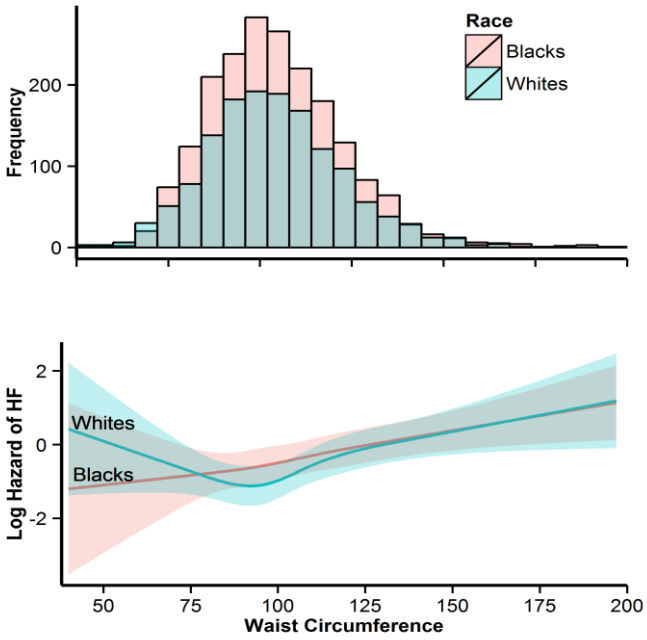


Fig 1. The log relative hazard of HF plotted against 3 surrogate measures of obesity among Blacks and Whites

Association between surrogate measures of total adiposity and post-HF mortality

Among the 4,341 SCCS participants who developed incident HF, 952 died (a cumulative mortality of 21.9%) over a median follow-up time of 2.3 (range: 0 – 8.8) years. The HR (and 95% CI) for post-HF mortality per 1 IQR increase in W/H^n and BMI are shown in **Table 4**. After adjustment for age and sex, a 1 IQR increase in W/H^n was associated with a 27% decrease in the risk of all-cause mortality after a HF diagnosis [HR= 0.73; 95% CI: 0.61 – 0.87] among whites and a 37% decrease [HR= 0.63; 95% CI: 0.55 – 0.71] in risk among blacks. Similar effect estimates were observed for BMI with HRs of 0.73 [95% CI: 0.62 – 0.87] and 0.63 [95% CI: 0.56 – 0.72] among whites and blacks respectively. After full adjustment for relevant covariates, a 1 IQR increase in W/H^n remained associated with a significant 30% decrease in the risk of post-HF mortality [HR = 0.70; 95% CI: 0.58 – 0.85] among whites and a 33% decrease in risk [HR= 0.67; 95% CI: 0.58 – 0.77] among blacks. Comparable estimates were obtained for BMI.

The plot of log relative hazard of post-HF mortality versus W/H^n (on a standardized scale) was suggestive of a negative nonlinear relationship between W/H^n and the risk of all-cause mortality following a HF diagnosis (**Figure 2a**). Among whites the risk decreased sharply with increasing values of W/H^n then plateaued, whereas among blacks, the initial sharp decrease was followed by a more gentle decrease throughout the range of values of W/H^n. An identical pattern was observed for the plot of BMI with the plateau observed among whites occurring at a BMI $\approx$ 30 Kg/m^2 (**Figure 2b**).

Table 4: Hazard ratios for incident HF and post-HF mortality per 1 IQR increase in weight-height indices

	Incident HF			
	Whites		Blacks	
	BMI	W/H^n	BMI	W/H^n
Model 1	1.51 (1.42, 1.60)	1.49 (1.41, 1.58)	1.44 (1.38, 1.51)	1.43 (1.37, 1.49)
Model 2	1.49 91.40, 1.58)	1.48 (1.39, 1.57)	1.46 (1.39, 1.53)	1.45 (1.39, 1.52)
Model 3	1.32 (1.23, 1.40)	1.31 (1.22, 1.39)	1.33 (1.27, 1.40)	1.33 (1.27, 1.39)
	Post-HF Mortality			
	Whites		Blacks	
	BMI	W/H^n	BMI	W/H^n
Model 1	0.73 (0.62, 0.87)	0.73 (0.61, 0.87)	0.63 (0.56, 0.72)	0.63 (0.55, 0.71)
Model 2	0.76 (0.64, 0.91)	0.75 (0.62, 0.90)	0.71 (0.62, 0.81)	0.70 (0.61, 0.80)
Model 3	0.70 (0.58, 0.84)	0.70 (0.58, 0.85)	0.68 (0.59, 0.78)	0.67 (0.58, 0.77)

Model 1: Includes age (restricted cubic splines with 4 knots), race and sex. **Model 2**: Model 1 + annual household income (<$15000, $15000-$24999 & ≥$25000) and education (< high school, high school/vocational training/junior college, college degree or higher, smoking (never, former, current < 19.5 pack-years, current ≥ 19.5 pack-years), alcohol intake (linear) and total physical activity in MET-hours (linear + quadratic). **Model 3:** Model 2 + diabetes (yes/no), hypertension (yes/no), high cholesterol (yes/no), history of MI (yes/no) and history of stroke (yes/no). **P-value for BMI×race interaction** = 0.24. **P-value for W/H^n ×race interaction** = 0.24. CI: Confidence Interval

Table 5: Hazard ratios for incident HF per 1 IQR increase in WC, overall and by race

	Overall	**Whites**	**Blacks**
Model 1	1.65 (1.45, 1.89)	1.74 (1.38, 2.18)	1.60 (1.36, 1.89)
Model 2	1.64 (1.43, 1.88)	1.80 (1.42, 2.28)	1.58 (1.33, 1.88)
Model 3	1.53 (1.32, 1.78)	1.66 (1.29, 2.14)	1.46 (1.20, 1.77)

Model 1: Includes age (restricted cubic splines with 4 knots), race and sex. **Model 2**: Model 1 + annual household income (<$15000, $15000-$24999 & ≥$25000) and education (< high school, high school/vocational training/junior college, college degree or higher, smoking (never, former, current < 19.5 pack-years, current ≥ 19.5 pack-years), alcohol intake (linear) and total physical activity in MET-hours (linear + quadratic). **Model 3:** Model 2 + diabetes (yes/no), hypertension (yes/no), high cholesterol (yes/no), history of MI (yes/no) and history of stroke (yes/no).

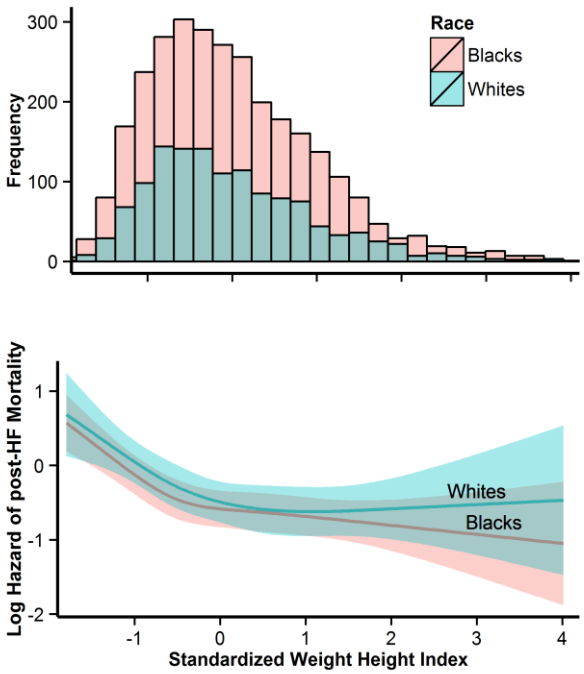


Fig 2a. The log relative hazard of post-HF mortality plotted against a standardized weight-height index among Blacks and Whites in SCCS.

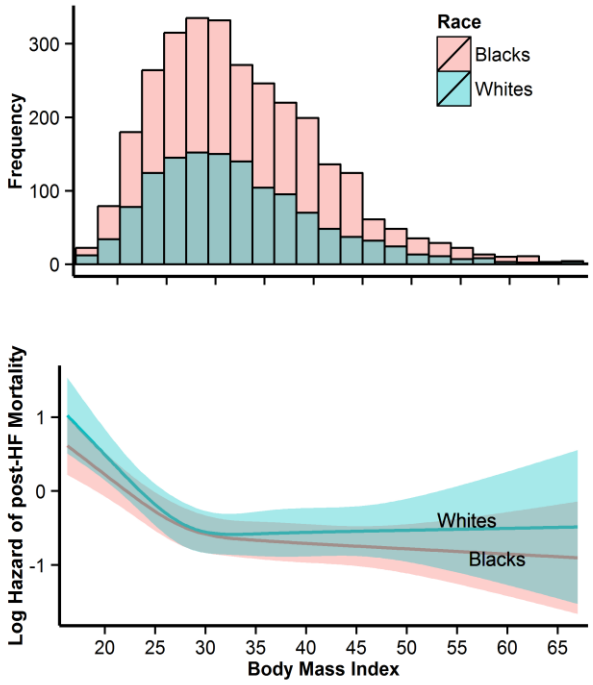


Fig 2b. The log relative hazard of post-HF mortality plotted against BMI among Blacks and Whites in SCCS.

Fig 2. The log relative hazard of post-HF mortality plotted against weight-height indices among Blacks and Whites in SCCS.

Discussion

In this large sample of low-income blacks and whites from the southeastern US, we investigated the association between surrogate measures of obesity and incident HF as well as post-HF mortality. Our principal findings were: 1) Compared to BMI, empirically-derived weight-height indices offer a better model fit when investigating the association with incident HF and post-HF mortality; 2) There exists an independent positive nonlinear association between surrogate measures of general (weight-height indices) as well as visceral obesity (waist circumference) and incident HF; and 3) There is an independent negative nonlinear dose-response relationship between surrogate measures of general obesity (weight-height indices) and post-HF mortality.

Most previous studies investigating the association between obesity and both HF risk and post-HF survival have utilized BMI (W/H^2) as a surrogate measure of total adiposity [105, 169] Meanwhile, some investigators have suggested that other weight-height indices including the ponderal index, W/H and W/H^3 could be equally appropriate depending on the demographic characteristics of the study population [16, 17, 82]. In the current study we propose a novel empirically-derived weight-height index (W/H^n) that utilizes information from the outcome measure to inform its functional form. We obtained better model fit statistics using the empiric index in models used to predict both HF risk and post-HF survival. This suggests that for the purpose of predicting HF outcomes it may be more apropos to use empirical indices rather than BMI which has been the default proxy for total adiposity utilized in prediction models including the HF risk calculator (C index = 0.741) developed by the Meta-Analysis Global Group in Chronic Heart Failure (MAGGIC) [169]. More so, in multivariable logit models for post-HF mortality in SCCS we obtained slightly higher C-statistics utilizing W/H^n compared to BMI (data not shown).

However, when investigating dose-response patterns and computing summary effect estimates for the obesity-HF association, we did observe similar patterns for either index – BMI or W/H^n. Overall, there was a significant independent association of comparable magnitude between both indices and HF

risk (with no evidence of heterogeneity of effects by race) which is corroborated by findings from previous studies. For example, using data from the Framingham Heart Study, Kenchaiah et al found HRs (95%CI) of 1.05 (1.02-1.09) and 1.07 (1.04-1.10) per kg/m^2 among men and women respectively after full adjustment for demographic variables, traditional cardiovascular (CV) risk factors (including smoking, diabetes, hypertension, past history of MI, total cholesterol) and left ventricular (LV) hypertrophy [13]. Using data from the Atherosclerosis Risk in Communities study, Ndumele et al found a HR of 1.32 per 5kg/m^2 after adjusting for traditional CV risk factors, N-terminal pro b-type natriuretic peptide (NT-BNP) and estimated glomerular filtration rate (eGFR) [15]. Ndumele et al also found a strong independent association between BMI and cardiac troponin T, a biomarker of subclinical myocardial injury which has been linked with incident HF [15]. This could explain in part the independent association between BMI and HF after adjusting for causal intermediaries like diabetes and hypertension.

In addition to that, obesity has been shown to be independently associated with impaired myocardial contractile function and relaxation as well as structural myocardial abnormalities [170, 171]. The most common structural changes observed include concentric remodeling, LV hypertrophy (which includes concurrent cardiac myocyte hypertrophy, fibrosis and apoptosis at the cellular level), dilated cardiomyopathy and increased right ventricular (RV) wall thickness which occur in varying degrees depending on obesity class and the presence or absence of hypertension [90, 92-99]. The presence of LV hypertrophy predisposes obese persons to abnormalities in diastolic function including increased LV end-diastolic pressure [90, 92] and abnormal load-dependent myocardial strain [103]. Data from animal models suggest that the structural and functional changes seen in obese individuals are related to metabolic derangements including decreased insulin sensitivity, decreased serum adiponectin levels, increased sympathetic tone, activation of the Renin Angiotensin Aldosterone System (RAAS) and low-grade systemic inflammation (increased C - reactive protein and Tumor Necrosis Factor, TNF) [14].

Still, in contrast to previous investigations on the obesity-HF association, which have either assumed a linear dose-response or reported effect estimates based on pre-specified categories (including those defined by the World Health Organization, WHO), we found a nonlinear J-shaped dose-response curve between BMI and HF risk. This J-shape pattern (suggestive of higher HF risk at very low or high values of BMI or W/H^n) mirrors patterns observed for the association between BMI and other outcomes including end-stage renal disease and all-cause mortality [172-174]. Elevated HF risk with increasing BMI maybe explained by abnormalities associated with increasing levels of adiposity including chronic myocardial injury and myocardial dysfunction [14]. This hypothesis is bolstered by the quasi-linear dose-response curve (in the BMI range where we have the most data) observed for the association between WC (a proxy for visceral adiposity) and HF risk. On the other hand, the underlying reasons behind higher risk at very low BMI remain uncertain and could be related to frailty and reduced lean mass. Importantly, while the nadir for the dose-response curve for blacks occurred at a BMI $\approx 30 kg/m^2$, that for whites occurred at a lower BMI, $\approx 25 kg/m^2$. If we utilized the WHO BMI categories as a referent framework, this would imply that among whites, the risk of HF rises with increasing BMI close to the cut-point for persons considered 'overweight'; meanwhile, among blacks, HF risk surges closer to the threshold for the 'obese' category. This would be consistent with previous data suggesting differential body composition by race whereby on average, whites have higher visceral fat and percent body fat (as well as higher risk of adverse outcomes including ESRD [175], atrial fibrillation [108] and cardiovascular mortality [107]) than blacks at similar BMI [176, 177] – another argument against the ubiquitous utilization of pre-specified BMI categories for risk stratification.

While the epidemiologic data linking obesity to increased HF risk and the biologic evidence for the nefarious effects of excess body weight on the myocardium appear to be compelling, several studies have suggested that overweight and obese persons with HF have a demonstrably higher survival compared to leaner subjects – a phenomenon coined as the obesity paradox [109, 111, 112, 114, 117]. For example, in a meta-analysis of 9 observational studies, Oreopoulos et al found a 33% lower risk [HR

= 0.67; 95% CI: 0.62 - 0.73] of all-cause mortality among obese persons with HF compared to persons with 'normal' weight [112]. Overall, our findings appear to corroborate this paradox for either index (BMI and W/H^n). Several hypotheses have been suggested to explain this phenomenon. Obese persons appear to have a higher metabolic reserve while leaner patients with HF may have a higher likelihood of being cachectic and thus could be at a higher risk of death [113]. Also, compared to persons with 'normal' weight, overweight and obese persons were found to have lower atrial natriuretic peptide levels which was correlated with having higher muscle mass and muscle strength [116]. The latter have been associated with improved survival in other patient populations and this may be analogous to patients with congestive HF as well [117]. Finally, certain underlying biological characteristics of obese persons with HF may play a role including the secretion of soluble TNF-α receptors which have beneficial neutralizing effects and elevated blood pressure which raises the tolerance of obese individuals to higher doses of cardio-protective agents including beta blockers, aldosterone antagonists and RAAS inhibitors [114, 115].

In light of existing data suggesting that in persons with HF, intentional weight loss mitigates some of the concurrent hemodynamic abnormalities and reduces left ventricular (LV) mass [90, 92, 93, 100, 114, 118], some cardiovascular societies recommend intentional weight reduction, albeit at higher cut-points for BMI [114]. The American Heart Association, the Heart Failure Society and the European Society of Cardiology recommend intentional weight loss in persons with HF at BMI cut-points of 40, 35 and 30 kg/m^2, respectively [114]. Importantly, one of the unique aspects of the nonlinear dose-response relationship observed between BMI at cohort entry and post-HF mortality in the current investigation of the SCCS data was that the initial inverse trend mostly plateaus beyond a BMI $\approx 30 kg/m^2$. While this is not clinical trial data, this observation appears to lend some credence to the above recommendations. Nevertheless, the ubiquitous utilization of BMI cut-points to guide patient care remains somewhat contentious as it seemingly assumes discontinuities in risk. An alternative decision-making paradigm may be to develop a highly discriminant prediction model for HF-related outcomes (including post-HF mortality and hospitalizations) using the relevant anthropometric, lifestyle, clinical and demographic

factors and using the estimates of predicted risk for individuals and cost-effectiveness ratios of efficacious interventions to make decisions in clinical or public heath settings. In addition, for any such models including the HF risk calculator developed by the MAGGIC (C index = 0.741), it would be interesting to explore differences in model fit (as well as discriminant and calibration properties) when utilizing BMI versus empiric weight-height indices derived for specific race-sex groups or more robust measures of adiposity whenever available.

Limitations of the study include the utilization of self-reported height and weight for calculation of weight-height indices but a prior SCCS validation study showed very high correlation ($r > 0.95$) between self-reported and measured weight and height [18]. Data from the National Health and Nutrition Examination Survey (NHANES) also suggest that BMI based on self-report has good concordance with BMI from measured values.[178] Data on baseline covariates (including CVD risk factors) were based on self-report of a physician diagnosis and use of medications. While self-report could be susceptible to recall and misclassification bias, these methods have been successfully utilized and validated in large epidemiologic cohorts, including the SCCS. Several of the questions on the SCCS questionnaire were adapted from questionnaires that were validated in other settings; and a series of independent validation studies using biomarkers, repeat interviews or medical records have demonstrated the reliability of the questionnaire within the SCCS population for variables such as smoking status and self-reported diseases including diabetes [18]. Another potential drawback is that for the analyses of weight-height indices in relation to post-HF mortality, we relied exclusively on the anthropometric and covariate data at cohort entry and we lacked these data at the time of HF diagnosis and these may have changed over time. Also, HF was ascertained via linkage with CMS Research Identifiable Files using ICD-9 codes 428.x (428.0 – 428.9), rather than independent physician adjudication. Nonetheless, the diagnosis codes algorithm for identification of HF used in this study has been previously validated and utilized in other cohorts [119-121]. A review of the detection of HF in administrative claims data that included studies conducted among Medicare beneficiaries reported positive predictive values (PPVs) mostly over 90% [119]. These

codes have also been used with high specificity in a number of studies [120, 121] even though no independent validation was conducted by the SCCS investigators. Data on the utilization of medication (ACE Inhibitors, beta blockers, statins etc), devices, investigations and procedures would have been useful in informing the analyses and these were unavailable from the linkages performed with the administrative claims data. In addition information on important predictors like LV ejection fraction (LVEF), NT proBNP, eGFR, or New York Heart Association class may have influenced the observed dose-response curves and summary effect estimates for post-HF mortality. Similarly additional clinic data on blood pressure values, HbA1c and serum lipid levels may have informed the analyses for the association with HF risk but we did adjust for past history of diabetes and hypertension and high cholesterol at baseline.

The current investigation leverages data from a large multiethnic cohort with a sizable number of participants from minority and low-income populations who are traditionally under-represented in previous cohorts investigating CVD and HF in particular. With a large burden of CVD risk factors at baseline including obesity, this cohort provides a unique opportunity to investigate dose-response between surrogate measures of obesity and HF-related outcomes in a population that is at a particularly high risk of adverse CVD outcomes. The availability of a large sample of participants and HF cases afforded the opportunity to adequately explore flexible dose-response patterns across racial categories.

In this low-income biracial cohort, empirically-derived weight-height indices (utilized as a surrogate of general obesity) offered a better model fit when investigating the association with incident HF and post-HF mortality compared to the default index – BMI. However, weight appears to be the more predominant component of the composite weight-height index (W/H^n) hence small changes in n – the exponent of the height variable – based on information provided by the outcome data may have modest effects on model fit statistics and predictions but do not change summary effect estimates and dose-response curves. More importantly, there were unique nonlinear dose-response patterns observed between BMI and both HF risk and post-HF mortality that could potentially inform current clinical

guidelines or add to the knowledge base required to improve the existing recommendations regarding risk stratification and the holistic management of the nutritional status and energy balance of persons with HF.

REFERENCES

1. Ambrosy AP, Fonarow GC, Butler J, Chioncel O, Greene SJ, Vaduganathan M, Nodari S, Lam CS, Sato N, Shah AN *et al*: **The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries**. *J Am Coll Cardiol* 2014, **63**(12):1123-1133.

2. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, Das SR, de Ferranti S, Despres JP, Fullerton HJ *et al*: **Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association**. *Circulation* 2015.

3. Roger VL, Weston SA, Redfield MM, Hellermann-Homan JP, Killian J, Yawn BP, Jacobsen SJ: **Trends in heart failure incidence and survival in a community-based population**. *JAMA* 2004, **292**(3):344-350.

4. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, Ikonomidis JS, Khavjou O, Konstam MA, Maddox TM *et al*: **Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association**. *Circ Heart Fail* 2013, **6**(3):606-619.

5. Arnold AM, Psaty BM, Kuller LH, Burke GL, Manolio TA, Fried LP, Robbins JA, Kronmal RA: **Incidence of cardiovascular disease in older Americans: the cardiovascular health study**. *J Am Geriatr Soc* 2005, **53**(2):211-218.

6. Bahrami H, Kronmal R, Bluemke DA, Olson J, Shea S, Liu K, Burke GL, Lima JA: **Differences in the incidence of congestive heart failure by ethnicity: the multi-ethnic study of atherosclerosis**. *Arch Intern Med* 2008, **168**(19):2138-2145.

7. Ho KK, Pinsky JL, Kannel WB, Levy D: **The epidemiology of heart failure: the Framingham Study**. *J Am Coll Cardiol* 1993, **22**(4 Suppl A):6A-13A.

8. Loehr LR, Rosamond WD, Chang PP, Folsom AR, Chambless LE: **Heart failure incidence and survival (from the Atherosclerosis Risk in Communities study)**. *Am J Cardiol* 2008, **101**(7):1016-1022.

9. Gottdiener JS, Arnold AM, Aurigemma GP, Polak JF, Tracy RP, Kitzman DW, Gardin JM, Rutledge JE, Boineau RC: **Predictors of congestive heart failure in the elderly: the Cardiovascular Health Study**. *J Am Coll Cardiol* 2000, **35**(6):1628-1637.

10. Roberts CB, Couper DJ, Chang PP, James SA, Rosamond WD, Heiss G: **Influence of life-course socioeconomic position on incident heart failure in blacks and whites: the Atherosclerosis Risk in Communities Study**. *Am J Epidemiol* 2010, **172**(6):717-727.

11. Bikdeli B, Wayda B, Bao H, Ross JS, Xu X, Chaudhry SI, Spertus JA, Bernheim SM, Lindenauer PK, Krumholz HM: **Place of residence and outcomes of patients with heart failure: analysis from the telemonitoring to improve heart failure outcomes trial**. *Circulation Cardiovascular quality and outcomes* 2014, **7**(5):749-756.

12. Levine JA: **Poverty and obesity in the U.S**. *Diabetes* 2011, **60**(11):2667-2668.

13. Kenchaiah S, Evans JC, Levy D, Wilson PW, Benjamin EJ, Larson MG, Kannel WB, Vasan RS: **Obesity and the risk of heart failure**. *N Engl J Med* 2002, **347**(5):305-313.

14. Lavie CJ, Alpert MA, Arena R, Mehra MR, Milani RV, Ventura HO: **Impact of obesity and the obesity paradox on prevalence and prognosis in heart failure**. *JACC Heart failure* 2013, **1**(2):93-102.

15. Ndumele CE, Coresh J, Lazo M, Hoogeveen RC, Blumenthal RS, Folsom AR, Selvin E, Ballantyne CM, Nambi V: **Obesity, subclinical myocardial injury, and incident heart failure**. *JACC Heart failure* 2014, **2**(6):600-607.

16. Florey Cdu V: **The use and interpretation of ponderal index and other weight-height ratios in epidemiological studies**. *Journal of chronic diseases* 1970, **23**(2):93-103.

17. Keys A, Fidanza F, Karvonen MJ, Kimura N, Taylor HL: **Indices of relative weight and obesity**. *Int J Epidemiol* 2014, **43**(3):655-665.

18. Signorello LB, Hargreaves MK, Blot WJ: **The Southern Community Cohort Study: investigating health disparities**. *J Health Care Poor Underserved* 2010, **21**(1 Suppl):26-37.

19. Sampson UK, Edwards TL, Jahangir E, Munro H, Wariboko M, Wassef MG, Fazio S, Mensah GA, Kabagambe EK, Blot WJ *et al*: **Factors associated with the prevalence of hypertension in the southeastern United States: insights from 69,211 blacks and whites in the Southern Community Cohort Study**. *Circ Cardiovasc Qual Outcomes*, **7**(1):33-54.

20. Lipworth L, Mumma MT, Cavanaugh KL, Edwards TL, Ikizler TA, Tarone RE, McLaughlin JK, Blot WJ: **Incidence and predictors of end stage renal disease among low-income blacks and whites**. *PLoS One* 2012, **7**(10):e48407.

21. Cowie CC, Rust KF, Ford ES, Eberhardt MS, Byrd-Holt DD, Li C, Williams DE, Gregg EW, Bainbridge KE, Saydah SH *et al*: **Full accounting of diabetes and pre-diabetes in the U.S. population in 1988-1994 and 2005-2006**. *Diabetes care* 2009, **32**(2):287-294.

22. Writing Committee M, Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Jr., Drazner MH, Fonarow GC, Geraci SA, Horwich T *et al*: **2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines**. *Circulation* 2013, **128**(16):e240-327.

23. Fonarow GC, Stough WG, Abraham WT, Albert NM, Gheorghiade M, Greenberg BH, O'Connor CM, Sun JL, Yancy CW, Young JB *et al*: **Characteristics, treatments, and outcomes of patients with preserved systolic function hospitalized for heart failure: a report from the OPTIMIZE-HF Registry**. *J Am Coll Cardiol* 2007, **50**(8):768-777.

24. Heart Failure Society of A, Lindenfeld J, Albert NM, Boehmer JP, Collins SP, Ezekowitz JA, Givertz MM, Katz SD, Klapholz M, Moser DK *et al*: **HFSA 2010 Comprehensive Heart Failure Practice Guideline**. *J Card Fail* 2010, **16**(6):e1-194.

25. McMurray JJ, Adamopoulos S, Anker SD, Auricchio A, Bohm M, Dickstein K, Falk V, Filippatos G, Fonseca C, Gomez-Sanchez MA *et al*: **ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC**. *Eur Heart J* 2012, **33**(14):1787-1847.

26. Cleland JG, Torabi A, Khan NK: **Epidemiology and management of heart failure and left ventricular systolic dysfunction in the aftermath of a myocardial infarction**. *Heart* 2005, **91 Suppl 2**:ii7-13; discussion ii31, ii43-18.

27. Kannel WB: **Incidence and epidemiology of heart failure**. *Heart failure reviews* 2000, **5**(2):167-173.

28. Vasan RS, Levy D: **Defining diastolic heart failure: a call for standardized diagnostic criteria**. *Circulation* 2000, **101**(17):2118-2121.

29. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Jr., Drazner MH, Fonarow GC, Geraci SA, Horwich T, Januzzi JL *et al*: **2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines**. *J Am Coll Cardiol* 2013, **62**(16):e147-239.

30. Steinberg BA, Zhao X, Heidenreich PA, Peterson ED, Bhatt DL, Cannon CP, Hernandez AF, Fonarow GC, Get With the Guidelines Scientific Advisory C, Investigators: **Trends in patients hospitalized with heart failure and preserved left ventricular ejection fraction: prevalence, therapies, and outcomes**. *Circulation* 2012, **126**(1):65-75.

31. Lee DS, Gona P, Vasan RS, Larson MG, Benjamin EJ, Wang TJ, Tu JV, Levy D: **Relation of disease pathogenesis and risk factors to heart failure with preserved or reduced ejection fraction: insights from the framingham heart study of the national heart, lung, and blood institute**. *Circulation* 2009, **119**(24):3070-3077.

32. Owan TE, Hodge DO, Herges RM, Jacobsen SJ, Roger VL, Redfield MM: **Trends in prevalence and outcome of heart failure with preserved ejection fraction**. *N Engl J Med* 2006, **355**(3):251-259.

33. **The Criteria Committee of the New York Heart Association. Nomenclature and criteria for diagnosis of diseases of the heart and great vessels.** , 9th edn. Boston, Mass: Little & Brown; 1994.

34. Hunt SA, Abraham WT, Chin MH, Feldman AM, Francis GS, Ganiats TG, Jessup M, Konstam MA, Mancini DM, Michl K *et al*: **2009 Focused update incorporated into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Heart Failure in Adults A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines Developed in Collaboration With the International Society for Heart and Lung Transplantation**. *J Am Coll Cardiol* 2009, **53**(15):e1-e90.

35. Khavandi K, Khavandi A, Asghar O, Greenstein A, Withers S, Heagerty AM, Malik RA: **Diabetic cardiomyopathy--a distinct disease?** *Best Pract Res Clin Endocrinol Metab* 2009, **23**(3):347-360.

36. Massie BM, Shah NB: **Evolving trends in the epidemiologic factors of heart failure: rationale for preventive strategies and comprehensive disease management**. *Am Heart J* 1997, **133**(6):703-712.

37. Tavazzi L, Opasich C: **Clinical epidemiology of heart failure**. *Congest Heart Fail* 1999, **5**(6):260-269.

38. Braunwald E: **Shattuck lecture--cardiovascular medicine at the turn of the millennium: triumphs, concerns, and opportunities**. *N Engl J Med* 1997, **337**(19):1360-1369.

39. Croft JB, Giles WH, Pollard RA, Keenan NL, Casper ML, Anda RF: **Heart failure survival among older adults in the United States: a poor prognosis for an emerging epidemic in the Medicare population**. *Arch Intern Med* 1999, **159**(5):505-510.

40. O'Connell AM, Crawford MH, Abrams J: **Heart failure disease management in an indigent population**. *Am Heart J* 2001, **141**(2):254-258.

41. Rich MW: **Epidemiology, pathophysiology, and etiology of congestive heart failure in older adults**. *J Am Geriatr Soc* 1997, **45**(8):968-974.

42. Rosamond W, Flegal K, Furie K, Go A, Greenlund K, Haase N, Hailpern SM, Ho M, Howard V, Kissela B *et al*: **Heart disease and stroke statistics--2008 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee**. *Circulation* 2008, **117**(4):e25-146.

43. Alexander M, Grumbach K, Remy L, Rowell R, Massie BM: **Congestive heart failure hospitalizations and survival in California: patterns according to race/ethnicity**. *Am Heart J* 1999, **137**(5):919-927.

44. Cowie MR, Wood DA, Coats AJ, Thompson SG, Poole-Wilson PA, Suresh V, Sutton GC: **Incidence and aetiology of heart failure; a population-based study**. *Eur Heart J* 1999, **20**(6):421-428.

45. Kannel WB: **Vital epidemiologic clues in heart failure**. *J Clin Epidemiol* 2000, **53**(3):229-235.

46. Kannel WB, Hjortland M, Castelli WP: **Role of diabetes in congestive heart failure: the Framingham study**. *Am J Cardiol* 1974, **34**(1):29-34.

47. Kenchaiah S, Narula J, Vasan RS: **Risk factors for heart failure**. *Med Clin North Am* 2004, **88**(5):1145-1172.

48. Senni M, Tribouilloy CM, Rodeheffer RJ, Jacobsen SJ, Evans JM, Bailey KR, Redfield MM: **Congestive heart failure in the community: trends in incidence and survival in a 10-year period**. *Arch Intern Med* 1999, **159**(1):29-34.

49. Westert GP, Lagoe RJ, Keskimaki I, Leyland A, Murphy M: **An international study of hospital readmissions and related utilization in Europe and the USA**. *Health Policy* 2002, **61**(3):269-278.

50. Bahrami H, Bluemke DA, Kronmal R, Bertoni AG, Lloyd-Jones DM, Shahar E, Szklo M, Lima JA: **Novel metabolic risk factors for incident heart failure and their relationship with obesity: the MESA (Multi-Ethnic Study of Atherosclerosis) study**. *J Am Coll Cardiol* 2008, **51**(18):1775-1783.

51. Parashar S, Katz R, Smith NL, Arnold AM, Vaccarino V, Wenger NK, Gottdiener JS: **Race, gender, and mortality in adults > or =65 years of age with incident heart failure (from the Cardiovascular Health Study)**. *Am J Cardiol* 2009, **103**(8):1120-1127.

52. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ, Dai S, Ford ES, Fox CS, Franco S *et al*: **Heart disease and stroke statistics--2014 update: a report from the American Heart Association**. *Circulation*, **129**(3):e28-e292.

53. Wells JC, Fewtrell MS: **Measuring body composition**. *Archives of disease in childhood* 2006, **91**(7):612-617.

54. James PT, Leach R, Kalamara E, Shayeghi M: **The worldwide obesity epidemic**. *Obes Res* 2001, **9 Suppl 4**:228S-233S.

55. Alberti KG, Zimmet P, Shaw J, Group IDFETFC: **The metabolic syndrome--a new worldwide definition**. *Lancet* 2005, **366**(9491):1059-1062.

56. Program NCE: **ATP III Guidelines At-a-Glance Quick Desk ReferenceUS Dept of Health and Human Services, Public Health Service, Bethesda, MD (2001) NIH Publication No. 01-3305. Available at: http://www.nhlbi.nih.gov/guidelines/cholesterol/atglance.pdf. Accessed Spetember 17, 2015.**

.

57. National Cholesterol Education Program Expert Panel on Detection E, Treatment of High Blood Cholesterol in A: **Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report**. *Circulation* 2002, **106**(25):3143-3421.

58. American Diabetes A: **Standards of medical care in diabetes--2012**. *Diabetes Care* 2012, **35 Suppl 1**:S11-63.

59. Force IDFCGT: **Global Guideline for Type 2 Diabetes: recommendations for standard, comprehensive, and minimal care**. *Diabet Med* 2006, **23**(6):579-593.

60. Klein S, Allison DB, Heymsfield SB, Kelley DE, Leibel RL, Nonas C, Kahn R, Association for Weight M, Obesity P, Naaso *et al*: **Waist circumference and cardiometabolic risk: a consensus statement from shaping America's health: Association for Weight Management and Obesity Prevention; NAASO, the Obesity Society; the American Society for Nutrition; and the American Diabetes Association**. *Diabetes Care* 2007, **30**(6):1647-1652.

61. Klein S, Allison DB, Heymsfield SB, Kelley DE, Leibel RL, Nonas C, Kahn R, Association for Weight M, Obesity P, Naaso TOS *et al*: **Waist circumference and cardiometabolic risk: a consensus statement from Shaping America's Health: Association for Weight Management and Obesity Prevention; NAASO, The Obesity Society; the American Society for Nutrition; and the American Diabetes Association**. *Am J Clin Nutr* 2007, **85**(5):1197-1202.

62. Elsayed EF, Tighiouart H, Weiner DE, Griffith J, Salem D, Levey AS, Sarnak MJ: **Waist-to-hip ratio and body mass index as risk factors for cardiovascular events in CKD**. *Am J Kidney Dis* 2008, **52**(1):49-57.

63. Snijder MB, van Dam RM, Visser M, Seidell JC: **What aspects of body fat are particularly hazardous and how do we measure them?** *Int J Epidemiol* 2006, **35**(1):83-92.

64. Lofgren I, Herron K, Zern T, West K, Patalay M, Shachter NS, Koo SI, Fernandez ML: **Waist circumference is a better predictor than body mass index of coronary heart disease risk in overweight premenopausal women**. *The Journal of nutrition* 2004, **134(**5):1071-1076.

65. Colditz GA, Willett WC, Rotnitzky A, Manson JE: **Weight gain as a risk factor for clinical diabetes mellitus in women**. *Ann Intern Med* 1995, **122(**7):481-486.

66. Calle EE, Thun MJ, Petrelli JM, Rodriguez C, Heath CW, Jr.: **Body-mass index and mortality in a prospective cohort of U.S. adults**. *N Engl J Med* 1999, **341**(15):1097-1105.

67. Wang Y, Rimm EB, Stampfer MJ, Willett WC, Hu FB: **Comparison of abdominal adiposity and overall obesity in predicting risk of type 2 diabetes among men**. *Am J Clin Nutr* 2005, **81**(3):555-563.

68. Hu G, Tuomilehto J, Silventoinen K, Barengo N, Jousilahti P: **Joint effects of physical activity, body mass index, waist circumference and waist-to-hip ratio with the risk of cardiovascular disease among middle-aged Finnish men and women**. *Eur Heart J* 2004, **25**(24):2212-2219.

69. Yusuf S, Hawken S, Ounpuu S, Bautista L, Franzosi MG, Commerford P, Lang CC, Rumboldt Z, Onen CL, Lisheng L *et al*: **Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study**. *Lancet* 2005, **366**(9497):1640-1649.

70. Nordhamn K, Sodergren E, Olsson E, Karlstrom B, Vessby B, Berglund L: **Reliability of anthropometric measurements in overweight and lean subjects: consequences for**

correlations between anthropometric and other variables. *Int J Obes Relat Metab Disord* 2000, **24**(5):652-657.

71. Abate N, Garg A, Peshock RM, Stray-Gundersen J, Grundy SM: **Relationships of generalized and regional adiposity to insulin sensitivity in men**. *J Clin Invest* 1995, **96**(1):88-98.

72. Moore FD: **The body cell mass and its supporting environment. London: W.B. Saunders Company; 1963**: London: W.B. Saunders Company; 1963.

73. Duren DL, Sherwood RJ, Czerwinski SA, Lee M, Choh AC, Siervogel RM, Cameron Chumlea W: **Body composition methods: comparisons and interpretation**. *Journal of diabetes science and technology* 2008, **2**(6):1139-1146.

74. Gray DS, Bray GA, Gemayel N, Kaplan K: **Effect of obesity on bioelectrical impedance**. *Am J Clin Nutr* 1989, **50**(2):255-260.

75. Brozek J, Grande F, Anderson JT, Keys A: **Densitometric Analysis of Body Composition: Revision of Some Quantitative Assumptions**. *Annals of the New York Academy of Sciences* 1963, **110**:113-140.

76. Sun SS, Chumlea WC, Heymsfield SB, Lukaski HC, Schoeller D, Friedl K, Kuczmarski RJ, Flegal KM, Johnson CL, Hubbard VS: **Development of bioelectrical impedance analysis prediction equations for body composition with the use of a multicomponent model for use in epidemiologic surveys**. *Am J Clin Nutr* 2003, **77**(2):331-340.

77. McCrory MA, Gomez TD, Bernauer EM, Mole PA: **Evaluation of a new air displacement plethysmograph for measuring human body composition**. *Med Sci Sports Exerc* 1995, **27**(12):1686-1691.

78. Roubenoff R, Kehayias JJ, Dawson-Hughes B, Heymsfield SB: **Use of dual-energy x-ray absorptiometry in body-composition studies: not yet a "gold standard"**. *Am J Clin Nutr* 1993, **58**(5):589-591.

79. Jocken JW, Blaak EE: **Catecholamine-induced lipolysis in adipose tissue and skeletal muscle in obesity**. *Physiology & behavior* 2008, **94**(2):219-230.

80. Goodpaster BH, Thaete FL, Kelley DE: **Composition of skeletal muscle evaluated with computed tomography**. *Annals of the New York Academy of Sciences* 2000, **904**:18-24.

81. Keys A, Fidanza F, Karvonen MJ, Kimura N, Taylor HL: **Indices of relative weight and obesity**. *Journal of chronic diseases* 1972, **25**(6):329-343.

82. Evans JG, Prior IA: **Indices of obesity derived from height and weight in two Polynesian populations**. *British journal of preventive & social medicine* 1969, **23**(1):56-59.

83. Roberts J: **Weight by height and age of adults: United States, 1960–1962. National Center for Health Statistics. Vital Health Stat 11(14). Available from: http://www.cdc.gov/nchs/data/series/sr_11/sr11_014acc.pdf.** In.; 1966.

84. Rothman KJ GS, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams and Wilkins. 2008.

85. Allison DB, Fontaine KR, Manson JE, Stevens J, VanItallie TB: **Annual deaths attributable to obesity in the United States**. *JAMA* 1999, **282**(16):1530-1538.

86. **Overweight, obesity, and health risk**. *Arch Intern Med* 2000, **160**(7):898-904.

87. Wolf AM, Colditz GA: **Current estimates of the economic cost of obesity in the United States**. *Obes Res* 1998, **6**(2):97-106.

88. Flegal KM, Carroll MD, Kuczmarski RJ, Johnson CL: **Overweight and obesity in the United States: prevalence and trends, 1960-1994**. *Int J Obes Relat Metab Disord* 1998, **22**(1):39-47.

89. Guh DP, Zhang W, Bansback N, Amarsi Z, Birmingham CL, Anis AH: **The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis**. *BMC Public Health* 2009, **9**:88.

90. Alpert MA: **Obesity cardiomyopathy: pathophysiology and evolution of the clinical syndrome**. *Am J Med Sci* 2001, **321**(4):225-236.

91. Nakajima T, Fujioka S, Tokunaga K, Hirobe K, Matsuzawa Y, Tarui S: **Noninvasive study of left ventricular performance in obese patients: influence of duration of obesity**. *Circulation* 1985, **71**(3):481-486.

92. Alpert M.A. AJK: **The Heart and Lung in Obesity**. New York: Futura Publishing Company; 1998.

93. Alpert MA, Terry BE, Kelly DL: **Effect of weight loss on cardiac chamber size, wall thickness and left ventricular function in morbid obesity**. *Am J Cardiol* 1985, **55**(6):783-786.

94. Alpert MA, Terry BE, Mulekar M, Cohen MV, Massey CV, Fan TM, Panayiotou H, Mukerji V: **Cardiac morphology and left ventricular function in normotensive morbidly obese patients with and without congestive heart failure, and effect of weight loss**. *Am J Cardiol* 1997, **80**(6):736-740.

95. Kasper EK, Hruban RH, Baughman KL: **Cardiomyopathy of obesity: a clinicopathologic evaluation of 43 obese patients with heart failure**. *Am J Cardiol* 1992, **70**(9):921-924.

96. Lauer MS, Anderson KM, Kannel WB, Levy D: **The impact of obesity on left ventricular mass and geometry. The Framingham Heart Study**. *JAMA* 1991, **266**(2):231-236.

97. Warnes CA, Roberts WC: **The heart in massive (more than 300 pounds or 136 kilograms) obesity: analysis of 12 patients studied at necropsy**. *Am J Cardiol* 1984, **54**(8):1087-1091.

98. Messerli FH, Sundgaard-Riise K, Reisin E, Dreslinski G, Dunn FG, Frohlich E: **Disparate cardiovascular effects of obesity and arterial hypertension**. *Am J Med* 1983, **74**(5):808-812.

99. Peterson LR, Waggoner AD, Schechtman KB, Meyer T, Gropler RJ, Barzilai B, Davila-Roman VG: **Alterations in left ventricular structure and function in young healthy obese women:**

assessment by echocardiography and tissue Doppler imaging. *J Am Coll Cardiol* 2004, **43**(8):1399-1404.

100. Alpert MA, Lambert CR, Terry BE, Cohen MV, Mulekar M, Massey CV, Hashimi MW, Panayiotou H, Mukerji V: **Effect of weight loss on left ventricular diastolic filling in morbid obesity**. *Am J Cardiol* 1995, **76**(16):1198-1201.

101. Pascual M, Pascual DA, Soria F, Vicente T, Hernandez AM, Tebar FJ, Valdes M: **Effects of isolated obesity on systolic and diastolic left ventricular function**. *Heart* 2003, **89**(10):1152-1156.

102. Russo C, Jin Z, Homma S, Rundek T, Elkind MS, Sacco RL, Di Tullio MR: **Effect of obesity and overweight on left ventricular diastolic function: a community-based study in an elderly cohort**. *J Am Coll Cardiol* 2011, **57**(12):1368-1374.

103. Wong CY, O'Moore-Sullivan T, Leano R, Byrne N, Beller E, Marwick TH: **Alterations of left ventricular myocardial characteristics associated with obesity**. *Circulation* 2004, **110**(19):3081-3087.

104. Voulgari C, Tentolouris N, Dilaveris P, Tousoulis D, Katsilambros N, Stefanadis C: **Increased heart failure risk in normal-weight people with metabolic syndrome compared with metabolically healthy obese individuals**. *J Am Coll Cardiol* 2011, **58**(13):1343-1350.

105. He J, Ogden LG, Bazzano LA, Vupputuri S, Loria C, Whelton PK: **Risk factors for congestive heart failure in US men and women: NHANES I epidemiologic follow-up study**. *Arch Intern Med* 2001, **161**(7):996-1002.

106. Chen YT, Vaccarino V, Williams CS, Butler J, Berkman LF, Krumholz HM: **Risk factors for heart failure in the elderly: a prospective community-based study**. *Am J Med* 1999, **106**(6):605-612.

107. Cohen SS, Signorello LB, Cope EL, McLaughlin JK, Hargreaves MK, Zheng W, Blot WJ: **Obesity and all-cause mortality among black adults and white adults**. *Am J Epidemiol* 2012, **176**(5):431-442.

108. Lipworth L, Okafor H, Mumma MT, Edwards TL, Roden DM, Blot WJ, Darbar D: **Race-specific impact of atrial fibrillation risk factors in blacks and whites in the southern community cohort study**. *Am J Cardiol* 2012, **110**(11):1637-1642.

109. Horwich TB, Fonarow GC, Hamilton MA, MacLellan WR, Woo MA, Tillisch JH: **The relationship between obesity and mortality in patients with heart failure**. *J Am Coll Cardiol* 2001, **38**(3):789-795.

110. Lavie CJ, Osman AF, Milani RV, Mehra MR: **Body composition and prognosis in chronic systolic heart failure: the obesity paradox**. *Am J Cardiol* 2003, **91**(7):891-894.

111. Clark AL, Chyu J, Horwich TB: **The obesity paradox in men versus women with systolic heart failure**. *Am J Cardiol* 2012, **110**(1):77-82.

112. Oreopoulos A, Padwal R, Kalantar-Zadeh K, Fonarow GC, Norris CM, McAlister FA: **Body mass index and mortality in heart failure: a meta-analysis**. *Am Heart J* 2008, **156**(1):13-22.

113. Anker SD, Negassa A, Coats AJ, Afzal R, Poole-Wilson PA, Cohn JN, Yusuf S: **Prognostic importance of weight loss in chronic heart failure and the effect of treatment with angiotensin-converting-enzyme inhibitors: an observational study**. *Lancet* 2003, **361**(9363):1077-1083.

114. Lavie CJ, Milani RV, Ventura HO: **Obesity and cardiovascular disease: risk factor, paradox, and impact of weight loss**. *J Am Coll Cardiol* 2009, **53**(21):1925-1932.

115. Mohamed-Ali V, Goodrick S, Bulmer K, Holly JM, Yudkin JS, Coppack SW: **Production of soluble tumor necrosis factor receptors by human subcutaneous adipose tissue in vivo**. *Am J Physiol* 1999, **277**(6 Pt 1):E971-975.

116. Zavin A, Daniels K, Arena R, Allsup K, Lazzari A, Joseph J, Schulze PC, Lecker SH, Forman DE: **Adiposity facilitates increased strength capacity in heart failure patients with reduced ejection fraction**. *International journal of cardiology* 2013, **167**(6):2468-2471.

117. Lavie CJ, De Schutter A, Patel DA, Romero-Corral A, Artham SM, Milani RV: **Body composition and survival in stable coronary heart disease: impact of lean mass index and body fat in the "obesity paradox"**. *J Am Coll Cardiol* 2012, **60**(15):1374-1380.

118. Alexander JK: **Obesity and Cardiac Performance**. *Am J Cardiol* 1964, **14**:860-865.

119. Goff DC, Jr., Pandey DK, Chan FA, Ortiz C, Nichaman MZ: **Congestive heart failure in the United States: is there more than meets the I(CD code)? The Corpus Christi Heart Project**. *Archives of internal medicine* 2000, **160**(2):197-202.

120. Rector TS, Wickstrom SL, Shah M, Thomas Greeenlee N, Rheault P, Rogowski J, Freedman V, Adams J, Escarce JJ: **Specificity and sensitivity of claims-based algorithms for identifying members of Medicare+Choice health plans that have chronic medical conditions**. *Health Serv Res* 2004, **39**(6 Pt 1):1839-1857.

121. Saczynski JS, Andrade SE, Harrold LR, Tjia J, Cutrona SL, Dodd KS, Goldberg RJ, Gurwitz JH: **A systematic review of validated methods for identifying heart failure using administrative data**. *Pharmacoepidemiol Drug Saf* 2012, **21 Suppl 1**:129-140.

122. Agarwal SK, Chambless LE, Ballantyne CM, Astor B, Bertoni AG, Chang PP, Folsom AR, He M, Hoogeveen RC, Ni H *et al*: **Prediction of incident heart failure in general practice: the Atherosclerosis Risk in Communities (ARIC) Study**. *Circ Heart Fail* 2012, **5**(4):422-429.

123. Sathiakumar N, Delzell E, Abdalla O: **Using the National Death Index to obtain underlying cause of death codes**. *J Occup Environ Med* 1998, **40**(9):808-813.

124. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, Hsia DC: **The accuracy of Medicare's hospital claims data: progress has been made, but problems remain**. *American journal of public health* 1992, **82**(2):243-248.

125. Rothman KJ, Greenland S, Lash TL: **Modern Epidemiology**, 3rd edn. Philadelphia: Lippincott Williams and Wilkins

2008.

126. **United States Census Bureau. Geographic Terms and Concepts - Census Tract. Available online at https://www.census.gov/geo/reference/gtc/gtc_ct.html. Accessed September 18th 2015**

127. Harrell FE, Jr. : **Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.** New York: Springer; 2001.

128. Messer LC, Laraia BA, Kaufman JS, Eyster J, Holzman C, Culhane J, Elo I, Burke JG, O'Campo P: **The development of a standardized neighborhood deprivation index**. *J Urban Health* 2006, **83**(6):1041-1062.

129. Fisher LD, Lin DY: **Time-dependent covariates in the Cox proportional-hazards regression model**. *Annual review of public health* 1999, **20**:145-157.

130. Cavender JB, Rogers WJ, Fisher LD, Gersh BJ, Coggin CJ, Myers WO: **Effects of smoking on survival and morbidity in patients randomized to medical or surgical therapy in the Coronary Artery Surgery Study (CASS): 10-year follow-up. CASS Investigators**. *J Am Coll Cardiol* 1992, **20**(2):287-294.

131. Fewell Z, Hernán MA, Wolfe F, Tilling K, Choi H, Sterne JAC, Tien PC: **Controlling for time–dependent confounding using marginal structural models**. *The Stata Journal Stata Press - College Station, TX* 2004, **4**(4):402-420.

132. Foster MC, Hwang SJ, Larson MG, Lichtman JH, Parikh NI, Vasan RS, Levy D, Fox CS: **Overweight, obesity, and the development of stage 3 CKD: the Framingham Heart Study**. *Am J Kidney Dis* 2008, **52**(1):39-48.

133. Leffondré K, Abrahamowicz M, Siemiatycki J, Rachet B: **Modeling Smoking History: A Comparison of Different Approaches**. *American Journal of Epidemiology* 2002, **156**(9):813-823.

134. **StataCorp. 2013. Stata Survival Analysis and Epidemiological Tables Reference Manual. Release 13. College Station, TX: Stata Press**. In.

135. Goldstein H: **Multilevel statistical models.** . London, United Kingdom.

: Edward Arnold; 1995.

136. Sundquist K, Winkleby M, Ahlen H, Johansson SE: **Neighborhood socioeconomic environment and incidence of coronary heart disease: a follow-up study of 25,319 women and men in Sweden**. *Am J Epidemiol* 2004, **159**(7):655-662.

137. Sundquist K, Malmstrom M, Johansson SE: **Neighbourhood deprivation and incidence of coronary heart disease: a multilevel study of 2.6 million women and men in Sweden**. *Journal of epidemiology and community health* 2004, **58**(1):71-77.

138. Institute of Education UoL: **MLwiN: Multilevel Models Project.** . In. London, United Kingdom: University of London; 2000.

139. Yang M, Rasbash J, Goldstein H: **MLwiN macros for advanced multilevel modeling, version 2.0a. Multilevel Models Project.** . In. London, United Kingdom: Institute of Education, University of London.; 2001.

140. Crowther MJ, Lambert PC: **stgenreg: A Stata package for the general parametric analysis of survival data.** . *Journal of Statistical Software* 2013, **53**(13).

141. Clarke P: **When can group level clustering be ignored? Multilevel models versus single-level models with sparse data**. *Journal of epidemiology and community health* 2008, **62**(8):752-758.

142. Lin DY: **Cox regression analysis of multivariate failure time data: the marginal approach**. *Stat Med* 1994, **13**(21):2233-2247.

143. **Chronic Conditions Data Warehouse. Centers for Medicare and Medicaid Services (CMS) research data. Available online at https://www.ccwdata.org/web/guest/technical-guidance-documentation. Accessed November 15, 2016.** [https://www.ccwdata.org/web/guest/technical-guidance-documentation]

144. Harrell FE, Jr.: **Regression Modelling Strategies with Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis.**, Second edn. New York: Springer; 2015.

145. **R Development Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.** In.

146. Ingelsson E, Lind L, Arnlov J, Sundstrom J: **Socioeconomic factors as predictors of incident heart failure**. *J Card Fail* 2006, **12**(7):540-545.

147. Mujib M, Zhang Y, Feller MA, Ahmed A: **Evidence of a "heart failure belt" in the southeastern United States**. *Am J Cardiol* 2011, **107**(6):935-937.

148. **Southern Community Cohort Study. The Southern Community Cohort Study: Cohort description. Available online at http://www.southerncommunitystudy.org/cohort-description.html. Accessed February 1, 2017** [http://www.southerncommunitystudy.org/cohort-description.html]

149. Sonderman JS, Mumma MT, Cohen SS, Cope EL, Blot WJ, Signorello LB: **A multi-stage approach to maximizing geocoding success in a large population-based cohort study through automated and interactive processes**. *Geospatial health* 2012, **6**(2):273-284.

150. **United States Census Bureau. 2008 TIGER/Line® Shapefiles (machine-readable data files). Available online at https://www.census.gov/geo/maps-data/data/tiger-line.html. Accessed on September 21, 2016**

151. Signorello LB, Cohen SS, Williams DR, Munro HM, Hargreaves MK, Blot WJ: **Socioeconomic status, race, and mortality: a prospective cohort study**. *American journal of public health* 2014, **104**(12):e98-e107.

152. Stewart S, Murphy NF, McMurray JJ, Jhund P, Hart CL, Hole D: **Effect of socioeconomic deprivation on the population risk of incident heart failure hospitalisation: an analysis of the Renfrew/Paisley Study**. *Eur J Heart Fail* 2006, **8**(8):856-863.

153. Diez Roux AV, Merkin SS, Arnett D, Chambless L, Massing M, Nieto FJ, Sorlie P, Szklo M, Tyroler HA, Watson RL: **Neighborhood of residence and incidence of coronary heart disease**. *N Engl J Med* 2001, **345**(2):99-106.

154. Diez Roux AV: **Neighborhoods and Health: What Do We Know? What Should We Do?** *American journal of public health* 2016, **106**(3):430-431.

155. Diez Roux AV, Mair C: **Neighborhoods and health**. *Annals of the New York Academy of Sciences* 2010, **1186**:125-145.

156. Diez Roux AV: **Neighborhoods and health: where are we and were do we go from here?** *Rev Epidemiol Sante Publique* 2007, **55**(1):13-21.

157. Levine JA: **Solving Obesity without Addressing Poverty: Fat Chance**. *Journal of hepatology* 2015.

158. Krieger N, Williams DR, Moss NE: **Measuring social class in US public health research: concepts, methodologies, and guidelines**. *Annual review of public health* 1997, **18**:341-378.

159. Borrell LN, Diez Roux AV, Rose K, Catellier D, Clark BL: **Neighbourhood characteristics and mortality in the Atherosclerosis Risk in Communities Study**. *Int J Epidemiol* 2004, **33**(2):398-407.

160. Havranek EP, Wolfe P, Masoudi FA, Rathore SS, Krumholz HM, Ordin DL: **Provider and hospital characteristics associated with geographic variation in the evaluation and management of elderly patients with heart failure**. *Arch Intern Med* 2004, **164**(11):1186-1191.

161. McAlister FA, Murphy NF, Simpson CR, Stewart S, MacIntyre K, Kirkpatrick M, Chalmers J, Redpath A, Capewell S, McMurray JJ: **Influence of socioeconomic deprivation on the primary care burden and treatment of patients with a diagnosis of heart failure in general practice in Scotland: population based study**. *Bmj* 2004, **328**(7448):1110.

162. Foraker RE, Rose KM, Suchindran CM, Chang PP, McNeill AM, Rosamond WD: **Socioeconomic status, Medicaid coverage, clinical comorbidity, and rehospitalization or death after an incident heart failure hospitalization: Atherosclerosis Risk in Communities cohort (1987 to 2004)**. *Circ Heart Fail* 2011, **4**(3):308-316.

163. Menec VH, Shooshtari S, Nowicki S, Fournier S: **Does the relationship between neighborhood socioeconomic status and health outcomes persist into very old age? A population-based study**. *Journal of aging and health* 2010, **22**(1):27-47.

164. Struthers AD, Anderson G, Donnan PT, MacDonald T: **Social deprivation increases cardiac hospitalisations in chronic heart failure independent of disease severity and diuretic non-adherence**. *Heart* 2000, **83**(1):12-16.

165. Rathore SS, Masoudi FA, Wang Y, Curtis JP, Foody JM, Havranek EP, Krumholz HM: **Socioeconomic status, treatment, and outcomes among elderly patients hospitalized with heart failure: findings from the National Heart Failure Project**. *Am Heart J* 2006, **152**(2):371-378.

166. Blackledge HM, Tomlinson J, Squire IB: **Prognosis for patients newly admitted to hospital with heart failure: survival trends in 12 220 index admissions in Leicestershire 1993-2001**. *Heart* 2003, **89**(6):615-620.

167. Antonelli-Incalzi R, Ancona C, Forastiere F, Belleudi V, Corsonello A, Perucci CA: **Socioeconomic status and hospitalization in the very old: a retrospective study**. *BMC Public Health* 2007, **7**:227.

168. Robert J. Sampson, Jeffrey D. Morenoff, Gannon-Rowley T: **Assessing "Neighborhood Effects": Social Processes and New Directions in Research**. *Annual Review of Sociology* 2002, **28**(1):443-478.

169. Sartipy U, Dahlstrom U, Edner M, Lund LH: **Predicting survival in heart failure: validation of the MAGGIC heart failure risk score in 51,043 patients from the Swedish heart failure registry**. *Eur J Heart Fail* 2014, **16**(2):173-179.

170. Abel ED, Litwin SE, Sweeney G: **Cardiac remodeling in obesity**. *Physiol Rev* 2008, **88**(2):389-419.

171. Turkbey EB, McClelland RL, Kronmal RA, Burke GL, Bild DE, Tracy RP, Arai AE, Lima JA, Bluemke DA: **The impact of obesity on the left ventricle: the Multi-Ethnic Study of Atherosclerosis (MESA)**. *JACC Cardiovasc Imaging* 2010, **3**(3):266-274.

172. de Gonzalez AB, Hartge P, Cerhan JR, Flint AJ, Hannan L, MacInnis RJ, Moore SC, Tobias GS, Anton-Culver H, Freeman LB *et al*: **Body-Mass Index and Mortality among 1.46 Million White Adults**. *The New England journal of medicine* 2010, **363**(23):2211-2219.

173. Akwo EA, Cavanaugh KL, Ikizler TA, Blot WJ, Lipworth L: **Increased body mass index may be associated with greater risk of end-stage renal disease in whites compared to blacks: A nested case–control study**. *BMC nutrition* 2015, **1**:24.

174. Calle EE, Thun MJ, Petrelli JM, Rodriguez C, Heath CWJ: **Body-Mass Index and Mortality in a Prospective Cohort of U.S. Adults**. *New England Journal of Medicine* 1999, **341**(15):1097-1105.

175. Akwo EA, Cavanaugh KL, Ikizler TA, Blot WJ, Lipworth L: **Increased body mass index may be associated with greater risk of end-stage renal disease in whites compared to blacks: A nested case-control study**. *BMC Nutr* 2015, **1**.

176. Deurenberg P, Yap M, van Staveren WA: **Body mass index and percent body fat: a meta analysis among different ethnic groups**. *Int J Obes Relat Metab Disord* 1998, **22**(12):1164-1171.

177. Lovejoy JC, de la Bretonne JA, Klemperer M, Tulley R: **Abdominal fat distribution and metabolic risk factors: effects of race**. *Metabolism* 1996, **45**(9):1119-1124.

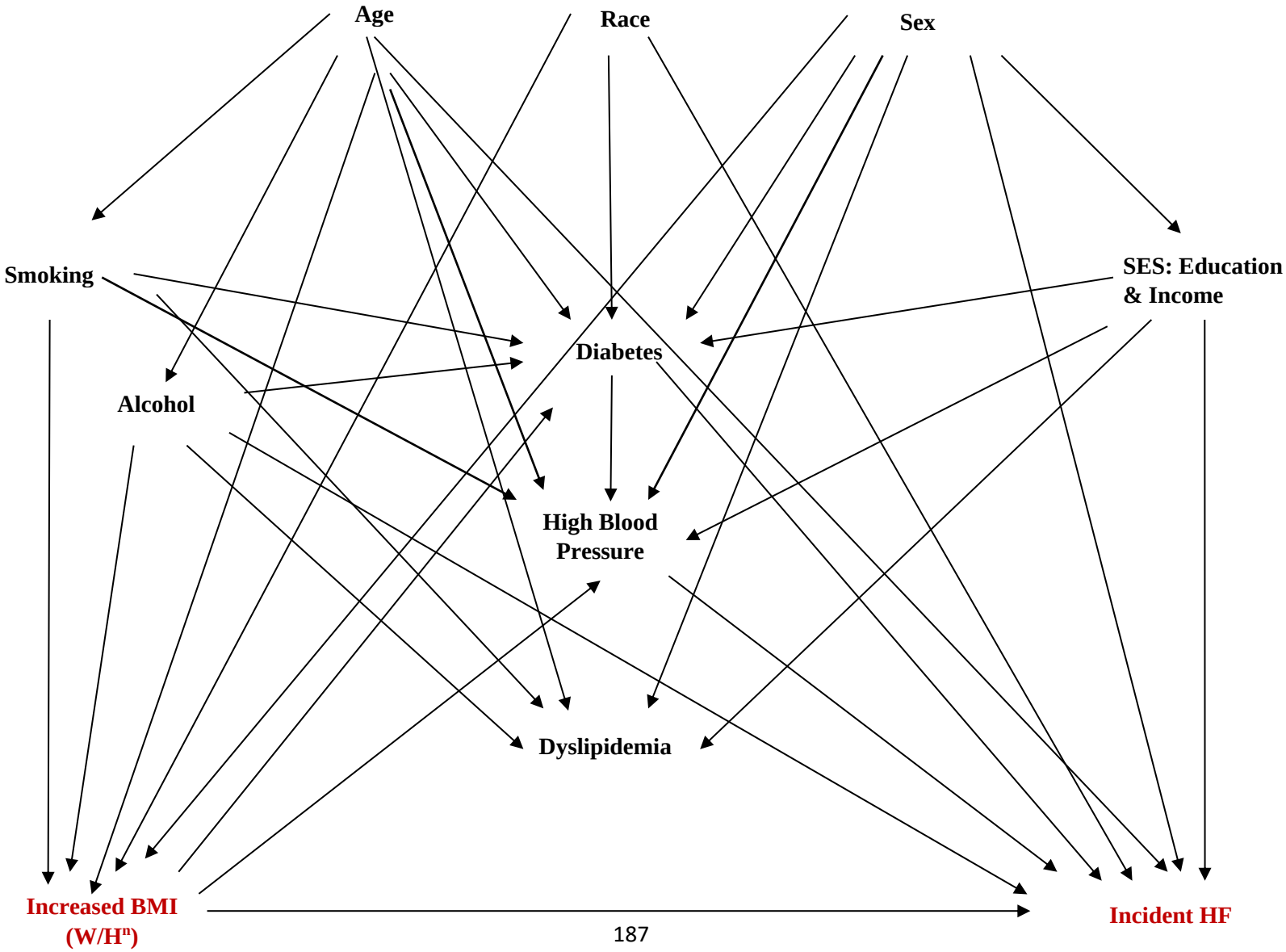
178. Craig BM, Adams AK: **Accuracy of body mass index categories based on self-reported height and weight among women in the United States**. *Matern Child Health J* 2009, **13**(4):489-496.

Appendix

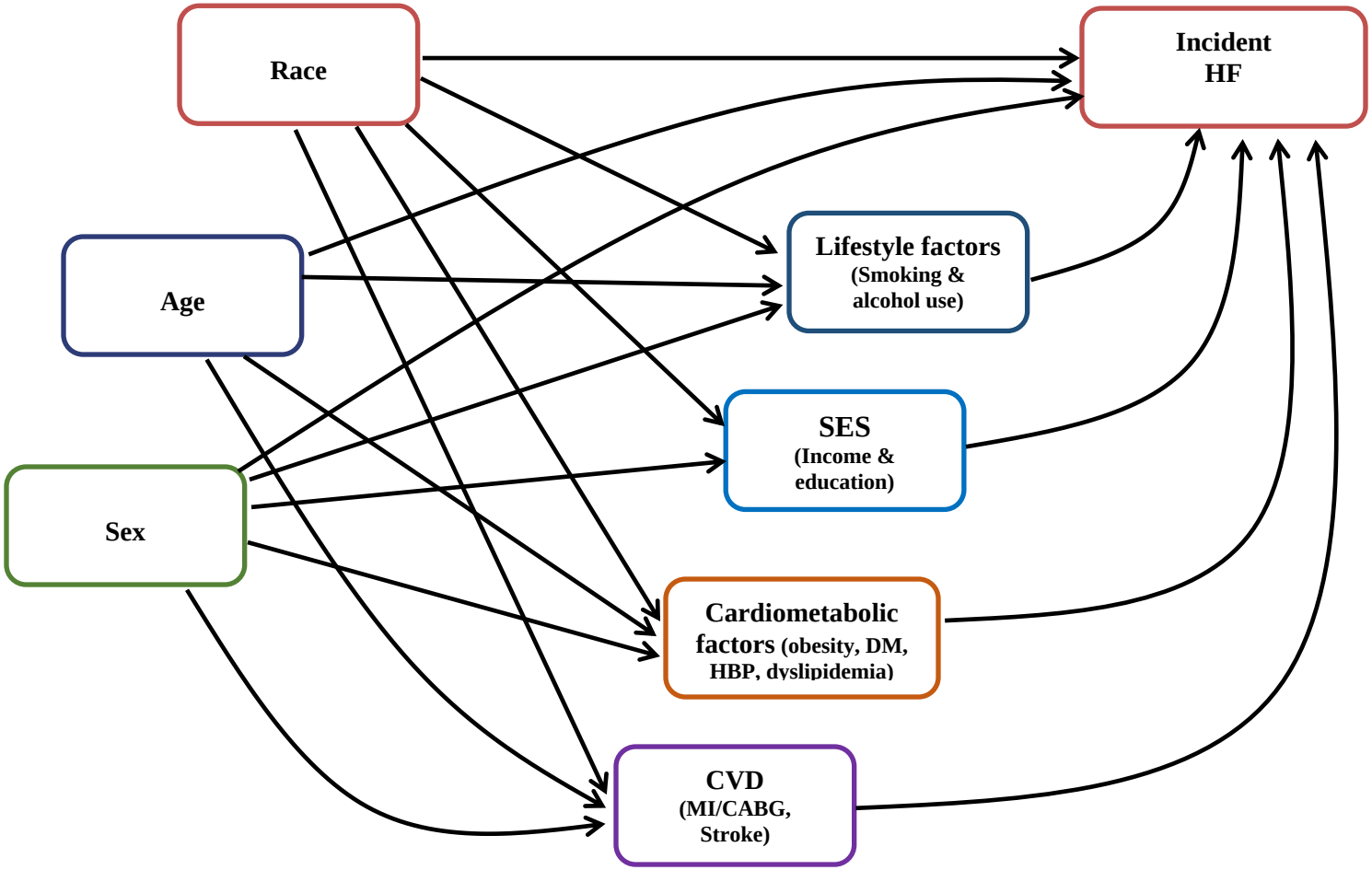
Appendix 1 shows a DAG for the association between weight-height index (as a surrogate for total body fat) and HF risk

Appendix 2 – 5: Causal diagrams showing the hypothesized association between clusters of variables representing anthropometric, socioeconomic, neighborhood, lifestyle & cardio-metabolic factors, CVD and incident HF (and post-HF survival).

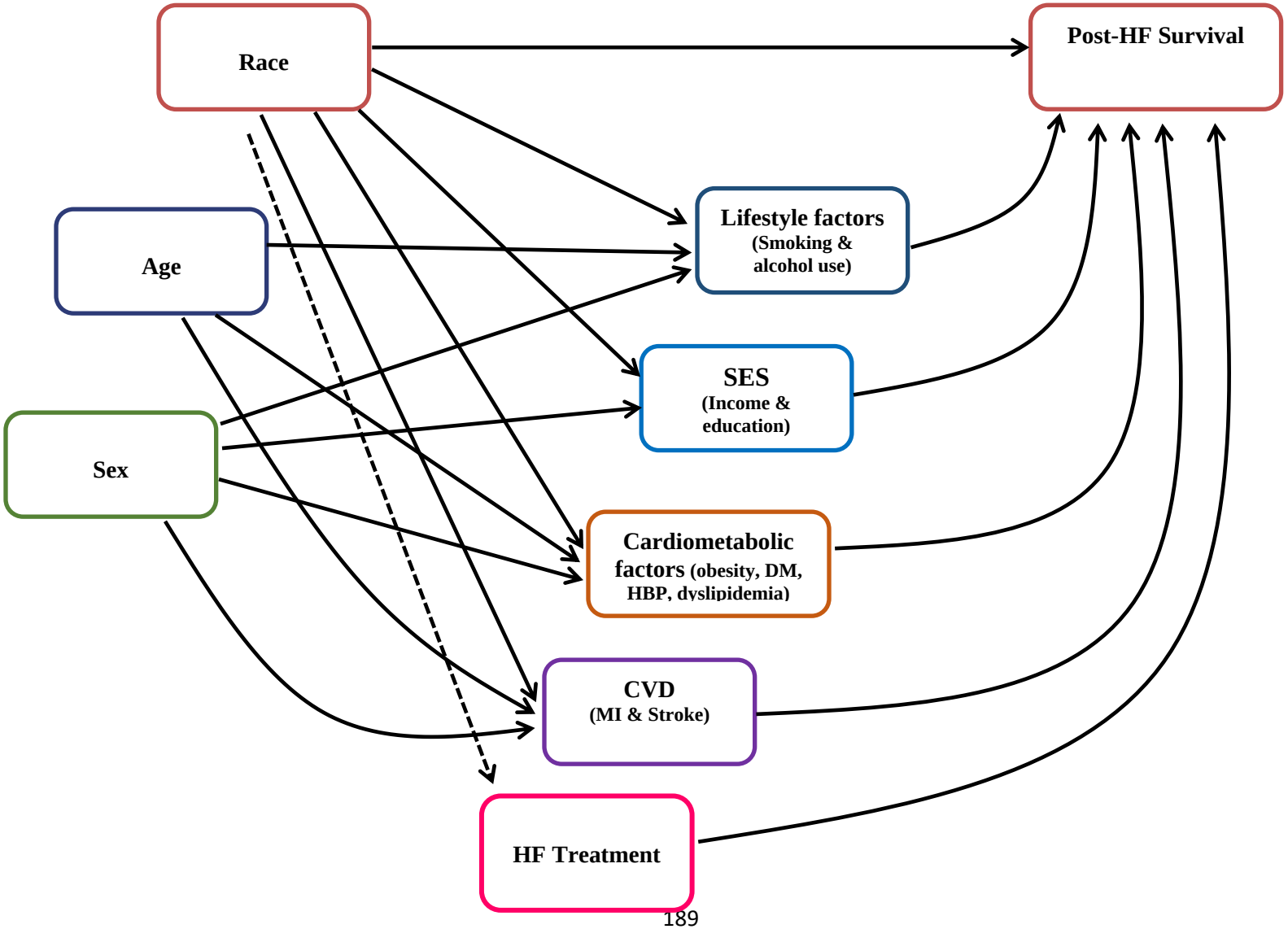
Appendix 1: Directed acyclic graph showing the hypothesized relationships between anthropometric, socioeconomic, lifestyle, cardiometabolic factors and Incident HF.



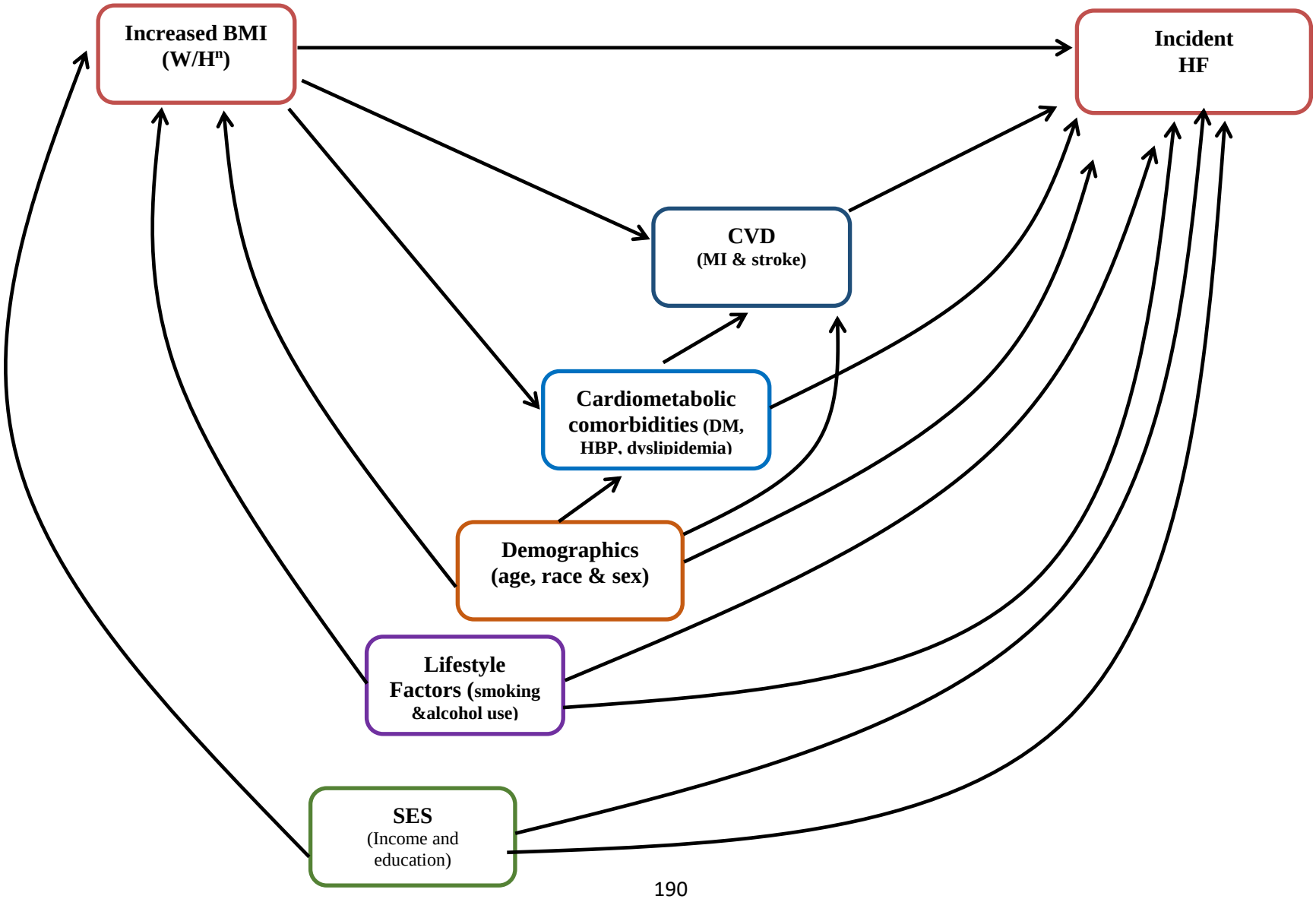
Appendix 2: Causal Diagram showing a summarized version of the hypothesized association between clusters of anthropometric, socioeconomic, lifestyle & cardiometabolic factors, CVD and Incident HF



Appendix 3: Causal Diagram showing a summarized version of the hypothesized association between clusters of anthropometric, socioeconomic, lifestyle & cardiometabolic factors, CVD and Post-HF Survival



Appendix 4: Causal Diagram showing a summarized version of the hypothesized association between clusters of anthropometric, socioeconomic, lifestyle & cardiometabolic factors, CVD and Incident HF (or post-HF survival)



Appendix 5: Causal Diagram showing a summarized version of the hypothesized association between clusters of neighborhood, socioeconomic, lifestyle & cardiometabolic factors, CVD and Incident HF (or post-HF survival).