Socioeconomic Related Health Inequalities in South Africa

by

David Wanyama Khaoya

Thesis Presented for the Degree of

Doctor of Philosophy

in the School of Economics
UNIVERSITY OF CAPE TOWN

August 2015

Supervisor:
Prof. Murray LEIBBRANDT

Co-Supervisor:
Prof. Ingrid WOOLARD

Declaration

I declare that this thesis is my own work, except where acknowledged in the text. I further declare that this thesis has not been submitted for a degree at any other university.

David Wanyama Khaoya

_____

Signature

_____

Date

Abstract

This thesis uses the National Income Dynamics Study (NIDS) data to estimate the extent of, and the factors correlated with, socioeconomic related health inequalities in South Africa. We extend our analysis by investigating whether income has a causal effect on health outcomes. The thesis is divided into four separate, but related chapters.

In chapter two we describe the data and the variables used in the study. We then check the quality of health related data in the NIDS by analyzing attrition trends and establishing whether attrition affects the representativeness of the data in subsequent waves. We use three health outcomes, self assessed health, body mass index and depression, to test for the potential effects of attrition bias on parameter estimates. We test using the attrition probit and Becketti, Gould, Lillard and Welch (BGLW) tests, which are two well known tests for attrition bias in panel data. We find that although the attrition rates of individuals from the sample are high in wave 2 and 3 (21% and 20% respectively), their attrition is random with respect to the health outcomes we use.

In chapter three we establish the socioeconomic factors correlated with health outcomes in South Africa. We use bivariate and panel data approaches. We find significant correlations between health outcomes and socioeconomic factors (income, educational attainment, and demographic factors). Income is positively correlated with self assessed health and body mass index, and it is negatively correlated with depressive symptoms.

In chapter four we build on the findings discussed in chapter three to estimate the extent of Income Related Health Inequality (IRHI). We estimate the index of inequality using a health concentration index. We then decompose the concentration index to establish the extent to which the correlates of health outcome drive the IRHI. The panel nature of the data allows us to investigate whether IRHI is narrowing or widening. We find a positive health concentration index. This implies that better health is concentrated among the rich. The decomposition of the index reveals that these differences are explained by disparities in income and educational attainment. We also find that the IRHI has narrowed from 2008 to 2012. Most of the narrowing is unexplained but about 21% and 20% of the decrease is correlated with the changes in the distribution and response to covariates respectively.

One of the socioeconomic determinants identified from the previous chapters to be corre-

lated with health is income. In the last part of this thesis we extend the analysis to investigate whether this relationship is causal. To do so, we use the Old Age Pension (OAP) programme as a natural experiment. The OAP is based on age eligibility. Therefore, we use this age eligibility as an exogenous income shock to isolate the effect of income on health. We apply a Regression Discontinuity Design on the NIDS data to identify this effect. We do not find any contemporaneous effect of income on three health outcomes considered, namely; self assessed health (SAH), body mass index (BMI), and depression.

Acknowledgements

I am most grateful to my family for the unwavering support I received while undertaking my PhD course. In particular, I would like to acknowledge my wife Dr Elvin Nyukuri, for agreeing to take care of the children while I study. My children Michael Khaoya and Chloe Khaoya for being patient with me. I thank my entire family, mother Mary Nabangala, brothers Moses, Douglas, Ken and sisters Margaret, Florence, Gladys, Loyce, Carren and Stella for the moral and material support I received.

My deepest appreciation goes to my supervisors, Professor Murray Leibbrandt and Professor Ingrid Woolard, for their invaluable support and guidance in the writing of this thesis. I would not have managed without their thorough and precise input, as well as their thoughtful and timely feedback. I must also mention that they gave me space to think.

Many thanks to all the lecturers who taught me course work, Prof Martin Wittenberg, Prof Justine Burnes, Prof Malcolm Keswell, Prof Haim Abraham, Dr Mare Sarr, Mr Andre Hofmeyr, A/Prof. Clara Delavallade, Mr Kevin Kotze, Dr Amos Peters, Dr Vimal Ranchhod, and Dr Asha Sundaram.They lessened my burden during data analysis.

My gratitude also goes to Prof. Eddy Van Doorslaer of Erasmus University Rotterdam who, while on short sabbatical at the University of Cape Town (UCT), helped me shape my thoughts in the formative stage of my thesis.

I acknowledge suggestions from my fellow PhD students in the brown bag seminars and the convener, Professor Paul Dunne, for his guidance both in the formative stages and later in my thesis. I also acknowledge comments received during the African Economic Research Consortium (AERC) bi-annual conferences which gave me different insights.

Finally, I am grateful for the funding I received from the African Economic Research Consortium (AERC), for the first four years, without which my dream would not have materialized. My sincere thanks also goes to the School of Economics at UCT for awarding me the Carnegie Foundation scholarship during my final year to assist me to complete my thesis.

Dedication

To my wife Dr Elvin Nyukuri and children Michael Khaoya and Chloe Khaoya.

List of Acronyms

ACE	Average Causal Effect
ADL	Activities of Daily Living
AERC	African Economic Research Consortium
AIDS	Acquired Immuno Deficiency Syndrome
BGLW	Becketti Gould Lillard and Welch
BHPS	British Household Panel Survey
BMI	Body Mass Index
CACE	Complier Average Causal Effect
CAPS	Cape Area Panel Study
CC	Corrected Concentration
CDC	Centre for Disease Control
CDF	Cummulative Density Function
CES-D	Centre for Epidemiological Studies - Depression
CI	Concentration Index
CPI	Consumer Price Index
CSDH	Commission on Social Determinants of Health
CSM	Continuing Sample Members
DHS	Demographic and Household Survey
EAs	Enumeration Areas
ECHP	European Community Household Panel
EU-SILC	European Union Statistics on Income and Living Conditions
FET	Further Education and Training
GDP	Gross Domestic Product
GHS	General Household Survey
HIV	Human Immunodeficiency Virus
HUI	Health Utility Index
IOTF	International Obesity Task Force
KIDS	KwaZulu-Natal Income Dynamics Study
KMPS	Khayelitsha Mitchell's Plain Study

LFS	Labour Force Survey
LPM	Linear Probability Model
MIC	Middle Income Country
NDoE	National Department of Education
NGO	Non-Governmental Organization
NHS	National Health Service
NIDS	National Income Dynamics Study
NTC	National Training Certificate
OHS	October Household Survey
OLS	Ordinary Least Squares
OOP	Out Of Pocket
PDF	Probability Density Function
QALY	Quality Adjusted Life Years
RCT	Randomized Control Trial
RDD	Regression Discontinuity Design
SAH	Self Assessed Health
SALDRU	Southern Africa Labour and Development Research Unit
SASSA	South African Social Security Service
SES	Socio-Economic Status
SF-36	Short Form 36
T2A	Transition to Adulthood
US	United States
WHO	World Health Organization
WLS	Weighted Least Squares

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Until the early 2000s, inequalities in health did not receive much attention from Economists globally (Wagstaff and van Doorslaer, 2000). The earlier literature on health inequalities was mainly from public health and this focused on population averages such as life expectancy, infant and under-five mortality (Wilkinson and Pickett, 2011). These statistics were used by international development agencies to monitor progress-especially in developing countries-to determine aid allocation. This has changed since the development of tools to estimate inequalities in the health care sector as well as the need to focus on the health of the poor (Van Doorslaer and O'Donnell, 2008; Van de Poel et al., 2012).

The interest of economists in analyzing health inequality is driven by the fact that ill health affects people's productivity and, by extension, their earnings. There is also a trend away from the narrow focus of economic inequality to a more encompassing analysis of inequality in well-being and health as a core determinant of welfare. Most analyses of inequality in health have, however, taken the approach of socioeconomic related inequalities in health rather than inequalities related to other dimensions such as age, gender or race because socioeconomic factors are deemed to be the major source of health inequity[1] (Fleurbaey et al., 2011; Fleurbaey and Schokkaert, 2009).

The majority of empirical measurement applications have focused on socioeconomic related health inequalities in developed countries rather than developing countries (Van de Poel et al., 2012; Van Doorslaer and Jones, 2004). These empirical studies (from developed countries) conclude that health is unevenly distributed in the population, with degrees varying

[1]Inequity is more of a normative concept which requires value judgment while inequality is an empirical concept (Chang, 2002). The international society of equity in health defined equity in health as the absence of potential remedial, systematic differences in one or more aspects of health across socially, economically, demographically or geographically defined population groups (Macinko and Starfield, 2002).

from country to another. This is because health outcomes are determined by factors which may differ across countries. Apart from socioeconomic factors, health outcomes are influenced by factors such as demographic (age and gender) and environmental factors which cannot be influenced easily by policy (Fleurbaey et al., 2011). The studies therefore distinguish between just and unjust health inequalities. Those inequalities in health that are associated with demographic variables are considered just, while those associated with socioeconomic (or well-being) factors and behavioural choice variables are considered unjust (Fleurbaey et al., 2011). In developing countries socioeconomic inequalities translate in uneven distribution of health outcomes[2] which are unjust.

In looking at well-being, socioeconomic variables commonly used in the literature include income, education and wealth. In the case of South Africa, these are unevenly distributed and this is partly attributed to the country's Apartheid history. Inequality in access to basic social services among population groups, and provinces are extensive and help to exacerbate inequalities in health outcomes (Ataguba et al., 2011). Although there is evidence to show that average real incomes have been increasing, aggregate inequality measures have also shown an increase in inequality over the post apartheid years. The country has registered one of the highest levels of income inequality, with an average Gini[3] coefficient of 0.69 for the years 1993, 2000, 2005 and 2008 (Leibbrandt et al., 2015). Despite this high level of income inequality, and disparities in other socioeconomic factors, little research has been undertaken in South Africa to establish the extent to which these inequalities drive health inequalities. In this thesis we seek to establish the extent of inequality in health outcomes and the role of socioeconomic inequalities in driving health outcomes.

Empirical studies on socioeconomic related health inequalities are concerned with either equalizing health outcomes or equalizing opportunities for better health. A study of opportunities for better health requires long-run, panel data for us to observe circumstances that prevail in early life and long-term health related choices (e.g smoking). For this reason we restrict this study to the perspective of equalizing health outcomes.

The measurement of inequality in health outcomes is divided into non rank and rank dependent strands. Non rank dependent strands use comparison of means or ranges of health outcomes between defined groups such as men/women or rural/urban (that is, they measure absolute health differences). On the other hand, rank dependent measures are able to reflect the experience of an entire population (Wagstaff and van Doorslaer, 2000). Rank dependent measures are the most common in the literature. Therefore, to ensure our results are

[2] These are considered unjust because they can be remedied by policy interventions.

[3]Gini coefficient is an index used to measure inequality usually in income. It ranges from 0 when there is perfect equality to 1 when there is maximum inequality.

comparable with previous studies we use this type of measure in this study (Ataguba et al., 2011).

Unequal socioeconomic situations does not however guarantee socioeconomic related inequality in health, especially when there are other interventions from the supply side that may dampen the effect of these factors. To this end we highlight briefly some of the interventions by the South African government aimed at improving health care access (Harrison, 2009; NDoH, 2013). These include policies put in place such as free primary health care, the essential drugs programme[4], the choice of termination of pregnancy, anti-tobacco legislation, and community service for graduating health professionals. Other strategies undertaken by the government include better health systems management through greater parity in district-based expenditure, clinic expansion and improvements, hospital revitalization programmes, and improved immunization and malaria control programmes.

Some of these policies were implemented soon after the transition to democracy in 1994. Despite these policy initiatives, socioeconomic related inequalities in health outcomes continue to be reported partly because the mitigating effects of these policies take time (Ataguba, 2013; McIntyre and J.E., 2012)). In particular, inequalities have been reported between regions, and between genders, and race, and socioeconomic groups, which leads us to ask the following questions: What is the extent of socioeconomic related health inequality in South Africa? What socioeconomic factors drive health inequalities in South Africa? What is the trend in health inequality?

Attempts have been made to analyze socioeconomic related health inequalities and answer the above questions. However, the findings do not provide the relative contribution of the different factors that drive inequalities in health outcomes. This is important for policy-makers in targeting strategies to correct the uneven distribution of good health. In addition previous studies have used the concentration curve and the related index to estimate health inequalities. It has recently been found that this index is sensitive to whether health or ill health is used as the equalisand[5] especially if the health outcome variable is bounded (Erreygers, 2009b,a). Most of the studies also focus on a single cross sectional survey or repeated cross sectional surveys. However, these do not properly capture the dynamics of adjustment concerning socioeconomic related health inequalities at the individual level. In this study we employ the corrected concentration index and exploit a unique dataset, the National Income Dynamics Study (NIDS) which is a nationally representative panel survey, to estimate socioeconomic related health inequalities and their contributing factors. The salient feature of NIDS is that it has a rich set of socioeconomic and health outcome variables.

[4]The policy introduced in 1996 which aims to develop and maintain a medicines list for all South Africans, provide equal access to medicines, increase adequate supply of limited medications and improve prices.

[5]Equalisand is the variable that we seek to equalize, in this case health outcomes. We can account for health outcome in terms of ill health or health if the health indicator is bounded.

We start with an investigation of the quality of health related data in the NIDS before we identify the determinants of health outcomes and the extent of socioeconomic related health inequality. We test for the effect of attrition bias on parameter estimates using attrition probit and the pooled or Becketti Gould Lillard and Welch (BGLW) test (Becketti et al., 1988). We find that, despite losing over 20% of the individuals through attrition in the NIDS dataset in waves 2 and 3, this does not bias the parameter estimates. This is because the attrition is not non-random and therefore does not affect the representativeness of the health related data. This implies that with regard to health outcome variables in NIDS one can estimate parameters in wave 2 and wave 3 separately without worrying about attrition bias. This study also finds that income, educational attainment, as well as demographic and geographical variables are significant correlates of health outcomes in South Africa. We estimate the extent of income related health inequality using the corrected concentration index and find a positive index across all the waves of the survey. This means that better health is concentrated among the well off. A decomposition of the index reveals that income inequality and differences in educational attainment account for the highest contribution of health inequality in South Africa.

In undertaking this analysis we deal with the challenge of health measurement. According to the World Health Organization (WHO), health is a state of complete physical, mental and social well-being, and not just the absence of disease or infirmity (WHO 1995). Health is therefore multidimensional and the challenge with survey data, such as the data we have, is that there is no single indicator that satisfies all the dimensions of health envisaged by the WHO definition. We therefore follow other studies by measuring one dimension at a time (O'Donnell et al., 2008). We explain how we measure health outcomes in the next chapter and in chapter four we discuss in detail the index we use to measure income related health inequalities and contrast it with competing indexes that have been used to measure health inequality in the past.

1.2 Objectives of the Thesis

In view of the above discussion, the purpose of this thesis is to analyze socioeconomic related health inequalities in South Africa by pursuing the following specific research objectives:

1. to assess the quality of health related data in NIDS

2. to establish the key socioeconomic determinants of health for South Africa

3. to estimate income related health inequality and the factors that contribute to this inequality as well as the factors associated with the changes in this inequality over time in South Africa

4. to establish the causal effect of income on health outcomes in South Africa

Addressing these research objectives is fundamental in correcting unjust health inequalities and also poverty and income inequality because of the relationship that exists between income and health. While lack of income can cause ill health, ill health too can cause incomes to decrease. By establishing factors associated with socioeconomic related health inequalities and how these change over time, this study contributes to the efforts to eradicate poverty through targeted policies in the health sector.

We also contribute to the literature on health inequalities in South Africa, which has been studied less than income inequality, although, as we have mentioned, health is a core determinant of welfare.

1.3 The structure of the Thesis

The outline of the thesis is as follows: Chapter one presents the introduction which establishes the area of focus and the research objectives. Each research objective is then addressed separately in subsequent chapters. The first objective, addressed in chapter two, is to assess the quality of health related data in the National Income Dynamics Study (NIDS). This is important because we use the data in our research. In this chapter we describe the dataset as well as the key variables. Since this is a panel dataset, quality checks involve checking whether using the variables (dependent variables) to estimate parameters in a structural model in subsequent waves would be affected by attrition bias.

Chapter three discusses the utilization of the variables described in chapter two, to estimate empirical relationships between health outcomes on the one hand and socioeconomic determinants of health on the other. We apply different econometric techniques on the three waves of NIDS to establish this relationship for South Africa. These data include the standard demographic and socioeconomic information. However NIDS also includes exceptionally rich health outcome data at the individual level for several other health dimensions. We select from these based on their suitability for our analysis.

In chapter four we show how we build on the analysis in chapter three to estimate income related health inequality in South Africa. This entails using income as the ranking

variable and estimating inequality in health outcomes using the Corrected Concentration Index. This index is recommended to the normal concentration index because it unambiguously ranks individuals whether ill health or good health is used as the indicator of health outcome (Van Doorslaer and O'Donnell, 2008). We then decompose the index to establish socioeconomic factors associated with inequalities in health. We apply the Oaxaca type decomposition to identify socioeconomic factors associated with changes in income related health inequalities over time.

We then move to establish the causal effect of income on health outcome, as discussed in chapter five. Chapter six concludes the study with a summary of the findings and the modeling techniques applied, highlighting challenges encountered and limitations. Important policy implications from the study are also presented in this chapter, as well as areas for further research.

1.4 Data

The data used in this study comes from the National Income Dynamics Study (SALDRU, 2009, 2014a,b). NIDS was conducted by the Southern Africa Labour and Development Research Unit (SALDRU) on behalf of the South African Presidency. SALDRU is based at the School of Economics of the University of Cape Town (UCT). The data is publicly available through DataFirst, a data service also based at the School of Economics of the University of Cape Town. This survey involved a nationally representative sample.

The objectives of the survey as well as description of the variables used in this study are discussed in detail in the next chapter where we examine the quality of health related data in NIDS.

Chapter 2

Health related Data in NIDS

2.1 Introduction

This chapter assesses the quality of health related data in NIDS (SALDRU, 2009, 2014a,b). NIDS is the first nationally representative panel survey in South Africa. Other South African panel surveys had limited geographical coverage. These include the KwaZulu-Natal Income Dynamics Study (KIDS), Birth-to-Twenty, Transitions to Adulthood (T2A) survey of South Africa, Cape Area Panel Study (CAPS) (Lee, 2003) and Khayelitsha Mitchell's Plain Study (KMPS) (Baigrie and Eyal, 2014).

NIDS is an initiative of the South African Presidency[1] aimed at establishing a national panel study to provide an information base to benchmark progress and assist in assessing the effectiveness of policies to promote positive social mobility (SALDRU, 2009, 2014a,b). The Southern Africa Labour and Development Research Unit (SALDRU), based at the School of Economics of the University of Cape Town, undertook the first three waves on behalf of the Presidency (SALDRU, 2009, 2014a,b).

The objective of NIDS is to track and explain changes in the well-being of South Africans over time (SALDRU, 2009, 2014a,b; Leibbrandt et al., 2009). These changes are tracked for individuals (28, 247) as they move out of their original 7,301 households (Brown et al., 2012). The first wave of NIDS was conducted in 2008 with the second and third in 2010-2011 and 2012 respectively.[2] Data was collected on all household residents in wave 1 (2008) which formed the base sample of individuals. These are called Continuing Sample Members (CSMs) and these individuals are followed in subsequent waves. In the subsequent waves, data is also collected on children born of female CSMs. Our focus is limited to the evaluation of one aspect of well-being; health, outcomes and its relationship with socioeconomic indicators. We use this relationship to analyze the disparities in health outcomes in South Africa.

[1] Office of the President

[2] These surveys are henceforth referred to as wave 1, wave 2 and wave 3 respectively.

The salient feature of NIDS is that it contains useful longitudinal data not only on health outcomes but also on other indicators of well-being, such as income, education and employment status, as well as individual characteristics of respondents. Such rich information is useful in the analysis of dynamic interrelationships between health and socioeconomic factors.

Despite the rich data collected in NIDS, there is the threat of obtaining inconsistent parameter estimates from the analysis of such data in subsequent waves due to biases introduced by measurement errors and attrition. While attrition can be observed, measurement error cannot.

Error in measurement of an independent variable produces down ward-biased (attenuated) and inconsistent parameter estimates (Bound et al., 2001). At the same time, it inadequately controls for confounding effects of this variable on well measured variables. Assessment of the extent and magnitude of measurement error requires validation studies. Unfortunately, we do not have true values of the same variables to estimate measurement error of our independent variables. We therefore interpret our results with this in mind.

Attrition is also a potential threat because of the challenge of interviewing the same individuals in every wave and the fact that the process through which attrition bias operates is incidental to the analysis and exogenous to the particular structural equations (Baigrie and Eyal, 2014). Because of this, our analysis in this chapter focuses only on the extent, effect and potential solutions to attrition bias with regard to health related models in NIDS. As previous studies have shown, biases arising from attrition are model and setting specific. We therefore consider a variety of health outcomes (Fitzgerald et al., 1998; Maluccio, 2004; Baigrie and Eyal, 2014).

Only a few studies have analyzed the quality of health related data in NIDS and the analyses either do not cover all the health outcome variables that we use in this study or are largely descriptive and fall short of applying some well known tests to tease out the effects of attrition bias (Baigrie and Eyal, 2014) and (Nwosu, 2014). In other cases the analysis does not cover all three waves of the survey (Baigrie and Eyal, 2014). We therefore extend our analysis in cases where all the waves were not considered as well as employ econometric methods to detect attrition bias on a selected list of health outcome variables that we use, as discussed in subsequent chapters.

Attrition is the failure to successfully re-interview targeted households or individuals (Maluccio, 2004). This failure can arise from fieldwork errors, non-contact, refusal or death of household members. We restrict our focus to resident household members who were continuing sample members (CSM) in wave 1 because they have observations in the base wave. We analyze adult and child responses separately unless the variable of interest makes

it relevant to examine these jointly. Throughout this study adults are defined as individuals aged 15 years and above in line with the definition adopted during the survey (Brown et al., 2012).

The analysis in this chapter begins with a description of health outcomes and the key socioeconomic determinants of health, and is followed by an analysis of attrition. The test for attrition entails the use of attrition probit and the pooled or Becketti Gould Lillard and Welch (BGLW) test (Becketti et al., 1988).

2.2 Health indicators in the NIDS

NIDS collected data on several variables related to health outcomes. One of these is self assessed health status (SAH). For this individuals are asked: "How would you describe your health at present?" This question has five possible responses: Excellent, very good, good, fair, and poor. Another health outcome variable is depression. This is collected through the 10 scale questions from the Center for Epidemiological Studies-Depression (CES-D). Responses to these questions are used to evaluate people's emotional or mental health. Other health indicators include prevalence of chronic illnesses, presence of acute illnesses such as Influenza, or fever, activities of daily living (ADLs), physical disorders (of, for example, sight, hearing), as well as anthropometric measurements which are used to compute health risk conditions such as body mass index (BMI), and blood pressure levels (hypertension).

For children, the SAH question was answered by the mother or the caregiver of the child or another household member who has knowledge about the child. Other health outcome data collected for the child include birth weight, data on serious illnesses or disabilities, and anthropometric measures that include height, weight and waist circumference.

Anthropometric measurements for adults include weight, height, systolic blood pressure, diastolic blood pressure, and pulse rate. From the weight and height measurements we compute BMI[3]. We use systolic and diastolic blood pressure measurements to deduce prevalence of hypertension (high blood pressure) which is a risk factor for cardiovascular disease. For children, besides BMI, we can also compute stunting and wasting[4] prevalence rates.

While SAH and chronic and acute illnesses can be interpreted directly, other health outcomes must be computed from a combination of responses. Health outcomes that are computed from a combination of responses include BMI and blood pressure. BMI is a measure of the nutritional status. It is known that having a BMI that is either too low or too high is associated with increased risk of ill health. The WHO classifies individuals as underweight, normal, overweight or obese based on their BMI score. If the BMI score is less than 18.5

[3]Body mass index computed as weight in Kg divided by the square of height in Centimeters

[4]Stunting is the case of having lower weight than usual while wasting is the case of having lower height than the usual for a given age.

the individual is said to be underweight, between 18.5 and 25 is considered normal weight, between 25 and 30 is overweight, while above 30 is obese. We exclude BMI scores that are implausible, such as BMI scores of less than 10 or greater than 45 (see for example (Wittenberg, 2013; Araar et al., 2009)).

BMI scores for preschool children and adolescents are adjusted for age and sex because comparison of their BMI scores is complicated by the fact that they are still growing. Also patterns of growth differ between boys and girls. In this study we follow guidelines (growth charts) provided by the WHO to categorize the BMI of adolescents[5]. In these guidelines, the BMI score for an adolescent is compared with a reference group of boys and girls at different percentiles of BMI (Onis et al., 2007). If the BMI score is less than the 5th percentile of the BMI of the reference group (growth chart), the adolescent is underweight, between 5th and 85th percentile is normal weight, 85 to 95 percentile is considered overweight and above 95th percentile is categorized as obese.

BMI as a health indicator has advantages over other indicators of health because it is easy to calculate and it uses anthropometric measurements which can be collected in a non-intrusive manner. Because of these it is considered as a more objective measure of health outcome. Its analysis is however complicated by the fact that both low (<18.5) and high (>25) values of BMI are a health risk.

Information collected on emotional (mental) health is used to estimate levels of depressive symptoms. Respondents self-report on a depression scale from the short (10 question) version of the CES-D scale which is one of the major international scales of depression used for general populations (Strauss et al., 2010). The questions are asked about how one felt in the past one week on ten items of the CES-D. The responses include, "I was bothered, I had trouble keeping my mind on what I was doing, I felt depressed, I felt everything was an effort, hopeful about the future and fearful." The other responses include "my sleep was restless, I was happy, I felt lonely and could not get going,"

Each of the ten responses in the CES-D scale could be answered as "rarely or none of the time" scored as 0, "some or little of the time" scored as 1, "Occasionally or a moderate amount of the time" scored as 2 and "All of the time" scored as 3. Higher scores are associated with depressive symptoms if the question is a negative one. In the literature if the question is about a positive event, such as being happy in the previous week, it is either reversed or excluded from the analysis (Radloff, 1977). If it is reversed the score would be 0 instead of 3 if the respondent reported being happy "all of the time" because being happy is associated with less depression. In the case of NIDS, two of the ten questions are about positive experiences. These questions are "I was happy" and "I felt hopeful about the future".

[5]Other growth guidelines in use include International Obesity Task Force (IOTF) and the Centres for Disease Control (CDC) guidelines

We recode the NIDS depression scores to be in line with the CES-D 10 score for each question: We adapt them to range from 0 to 3 instead of 1 to 4 as in the dataset. The scores for all the questions are summed up to generate a depression score for an individual that ranges from a minimum of 0 to a maximum of 30 (or 24 if only 8 questions are used instead of 10). The sum of the scores is 0 if eight negative questions are answered "Rarely or none of the time" and 2 positive ones answered "All of the time". The guidelines allow for scoring even when some questions have missing responses. However, if more than four questions are left unanswered, it is not advisable to use the depression score because an individual may be wrongly categorized as having no depressive symptoms because they answered fewer questions (Radloff, 1977). In our case we exclude 5 individuals in wave 1, 48 and 27 individuals in wave 2 and wave 3 respectively because they did not answer between 4 and 9 questions. This is in addition to non responders to this question. The higher the score for this measure, the higher the likelihood of the respondent having depressive symptoms.[6] As a convention in population research, total scores above 10 indicate prevalence of depressive symptoms while scores of 10 and below indicate lack of depressive symptoms (Radloff, 1977).

Data on hypertension (high blood pressure) was collected in the NIDS in two ways. First individuals were asked if they had been told by a doctor, nurse or health care professional that they had high blood pressure (with yes/no response categories). Second, through blood pressure systolic and diastolic measurements. This establishes the measurement objectively. Blood pressure is typically recorded as two numbers written as a ratio with blood pressure (BP) systolic measurement in the numerator and blood pressure diastolic in the denominator (Krisela and Jean-Fourie, 2006). BP systolic measures pressure in arteries when the heart is pumping while BP diastolic measures pressure in arteries when the heart is resting, that is, waiting to fill up with blood. The units of measurement are millimeters of mercury (mm Hg). The normal range for BP systolic is less than 120 mm Hg while BP diastolic is less than 80 mm Hg. BP Systolic of between 120-139 mm Hg and BP diastolic of between 80-90 mm Hg are considered prehypertension, BP systolic of between 140-159 mm Hg and BP diastolic of between 90-99 mm Hg is considered as High blood pressure stage 1, BP Systolic of between 160-180 mm Hg and BP diastolic of 100-110 mm Hg is considered as high blood pressure stage 2. Hypertensive crisis occurs when BP systolic is above 180 and BP diastolic is above 110 mm Hg (Krisela and Jean-Fourie, 2006). The internationally accepted cut off for identifying people with hypertension is 140/90 mm Hg because the lowest threshold for High blood pressure stage 1 for BP systolic and BP diastolic is 140 and 90 mm Hg respectively (Krisela and Jean-Fourie, 2006). We add a caveat here that some individuals are on hypertension medication and could thus falsely be categorized as having normal blood pressure. For this

[6]This explains why responses to positive questions should be scored less if they occur all of the time because they signify less depression.

reason, when we use this health indicator we control for medication status.

BP Systolic is normally given more attention because it is an indicator of a major risk factor for cardiovascular disease. For most people BP systolic rises with age due to increasing stiffness of large arteries, long term build-up of plague and increased incidence of cardiac and vascular disease. Measurements that are implausible such as BP systolic less than 50 and more than 200 mm Hg and BP diastolic less than 40 and greater than 180 mm Hg are excluded from the calculation of hypertension.

The other health outcome variable we analyze in this study is prevalence of chronic illness. In NIDS respondents were asked if they had ever been told by a doctor, nurse or health care professional that they had a particular chronic illness, with responses "yes" or "no". These chronic illnesses include tuberculosis, high blood pressure, diabetes or high blood sugar, stroke, asthma, heart problems and cancer. As noted by Ardington and Gasealahwe (2012) and Nwosu (2014), there are a lot of inconsistencies across waves in the NIDS where individuals who affirmed having been diagnosed with a given chronic condition answered in the negative for the same illness in a subsequent wave. This was mostly a problem for measures of high blood pressure, heart problems, diabetes and tuberculosis. Respondents also frequently changed the year of diagnosis across waves. Because of this, we do not include these health variables in our subsequent regression analysis.

The final health variable is Activities of Daily Living (ADL). Respondents were asked about the level of difficulty they had in carrying out certain specified daily activities by themselves. Responses are on a Likert scale, with lower values indicating no difficulty and higher values indicating increasing levels of difficulty. This question was however only asked in waves 1 and 2 and not in wave 3. For this reason we do not analyze this variable further.

This study focuses on three health outcomes, namely SAH, BMI and depression. These health outcomes are selected because they capture the multidimensional nature of health. They are also better measured in the data compared to other health outcomes. Below we provide a detailed review of each of these health outcomes.

2.2.1 Self Assessed Health

The World Health Organization (WHO) defines health as a state of complete physical, mental and social well-being and not just the absence of disease and infirmity. It is clear from this definition that health is a multidimensional phenomena. It is not easy to have one indicator of health that captures the multidimensional nature of health from a survey. The closest measure of health as envisaged by the WHO is Self assessed health (SAH).

It is believed that when respondents answer SAH question, they consider their physical, mental and social well-being before declaring their health status. SAH as a measure of health

is relatively easy to collect and in most surveys response rates for the question is always high. In the case of NIDS the response rates were close to 100 percent. This measure of health is therefore an increasingly common measure in empirical research (Crossley and Kennedy, 2002). It is further supported by international literature that shows that it predicts mortality and morbidity (Benyamini et al., 1999; Idler and Benyamini, 1997).

In collecting information about SAH, respondents are usually asked a single question regarding how they rate their health status. The responses are qualitative and they range from poor, fair, good, very good and excellent. Benyamini et al. (1999) discusses in detail the psychological processes underlying how individuals make global assessments of their health. One view is that individuals knowledge of their history of medical conditions contribute to how they judge their health. For example knowledge or a history of a chronic illness such as cancer has a greater impact in reporting worse SAH compared with knowledge of a non life threatening condition like flu.

The degree to which individual knowledge of their history of medical conditions influence their self assessment of health depends on the degree to which that knowledge is activated and made salient during the judgment process. Two views exist in the literature, the first is where the knowledge of the disease affects the SAH judgments directly and the second is where SAH are influenced by concrete experience of the disease. Ongoing manifestations of the disease may imply the presence of a serious illness and will influence how individuals make global assessments of their health.

This analysis shows that when individuals self assess themselves in different health categories, they reveal the levels of severity of their disease (health outcomes). Those who self assess in poor health status are worse off than those who self assess in other health status categories. On the basis of this, we can make certain conclusions about the differences in health outcomes between individuals.

Despite its popularity in empirical research, evidence has shown that SAH may be prone to measurement error (Jones et al., 2013). The measurement error could be due to reporting heterogeneity that occurs when subgroups of the population use systematically different cut point levels when reporting their SAH. This may result in different self reports for individuals in the same level of health. Reporting heterogeneity has been shown to vary with respondent characteristics such as age, income and race (Jones et al., 2013). This biases analyses that rely on interpersonal comparability of SAH.

SAH is captured in ordinal scale and is usually modeled using ordered probit model. In this way reporting heterogeneity is ignored or ruled out. Reporting heterogeneity can however be investigated and purged by making cut points dependent on covariates through a generalized ordered regression model. This however requires that we separate a priori the

variables that affect health and those that affect reporting. This is because if they are put together it is not possible to separately identify the effect of the variable on reporting and health at the same time.

One way to solve this problem is to condition the analysis on additional information on true health such as health utility index. Any variations of SAH from the objective health is due to reporting heterogeneity and is purged before any interpersonal comparisons is done. Alternatively, additional information is used on reporting behaviour itself to examine variation in the evaluation of a given health state represented by vignettes. The vignettes are fixed so any variation is attributed to reporting heterogeneity and is purged before interpersonal comparisons.

In our case we have not analyzed the extent of reporting heterogeneity in our data. This is a limitation because the results are likely to be overestimated or underestimated.

2.2.2 Body Mass Index

BMI is a measure of excess body weight. It is therefore an indicator of physical health. It is calculated from a formula that uses a person's weight and height to calculate a number which is representative of their level of excess weight (Hall and Cole, 2006). BMI is often considered an indicator of fatness, but it is a surrogate measure of body fat since it measures excess weight for height. It is correlated to more direct measures of fat such as underwater weighing and dual energy x-ray absorptiometry. In addition, epidemiological studies have shown a high correlation between BMI and mortality (Lewis et al., 2009). For these reasons, BMI is sometimes used in research as a measure of health outcome.

BMI is not intrusive and is also easy to calculate. This has made it to attain growing popularity in empirical research. It is also preferred because it is considered to be an objective measure of health that results from the ratio of two independently measured variables, weight and height. These variables are part of the anthropometric measurements that are collected in our data along-side pulse rate, systolic and diastolic blood pressure. As we mentioned in section 2.2 above, the WHO classifies individuals as underweight, normal, overweight and obese based on their BMI index. From a health perspective, BMI at the upper extreme (obese) is what posses the greatest health risk, although very low BMI (underweight) equally poses a health risk (Ardington and Gasealahwe, 2012).

There is a strong correlation between excess BMI (Obese) and mortality. This relationship has been found to be either linear, U-shaped or J-shaped (Lewis et al., 2009). A U-shaped relationship implies that the underweight and the overweight are more likely to die than those in between. This is a special characteristic of BMI. The turning point at which an increasing BMI is associated with mortality is not fixed as it varies from population to population. It is

this strong correlation of BMI and mortality that makes BMI to be categorized as a health indicator. High BMI is also correlated with other health outcomes such as diabetes and cardial vascular disease.

Because of its strong positive relationship with fatness, BMI is used as a surrogate measure of body fat. Body fats accumulate from excess food (Rosqvist et al., 2014). For this reason, socioeconomic circumstances have a strong influence on BMI through the nutrition channel. The relationship between BMI and socioeconomic variables is however mixed. This is because excess nutrition may not lead to excess fats if one exercises regularly or due to genetic factors that prevent accumulation of fats. As we will argue later, higher levels of education may also influence an individual's choice of diet by avoiding foods that lead to accumulation of fats.

Although BMI is popular in empirical research, the two components of this index suggest possible limitations of its use. Since height is one component, the BMI may be stature dependent over at least part of the age range because the young are still growing tall and the BMI dependents on their height. The use of weight as a numerator too suggests that BMI may reflect both lean tissue and fat tissue to a comparable degree.

2.2.3 Depression (Emotional health)

Depression is a mood disorder that causes a persistent feeling of sadness and loss of interest. It affects how one feels, thinks and behaves and can lead to emotional problems. This definition fits in one of the dimensions of health as defined by the WHO, the mental health dimension. The problem with this health outcome is that, like the SAH, there is no single indicator that can measure depression because it manifests itself in many symptoms.

Diagnosis of depression for clinical intake and evaluation of severity of illness is done differently from depression in the general population. This implies that one can not use the tool used to collect information about depression in a survey to diagnose people for clinical illness of depression. In order to determine prevalence of depression in the general population, information regarding the symptoms is gathered. This is done through short self assessed questions targeting symptoms associated with depression (Radloff, 1977). The tool for this purpose is called the CES-D scale named after the Center for Epidemiological Studies came up with a depression scale. The items of the scale are symptoms associated with depression which are in form of short self assessed questions.

The CES-D items as tools to assess depression in populations have been validated in the field clinical and psychology literature. The major components of symptomatology were identified from the clinical literature and factor analytic studies (Radloff, 1977). The components include depressed mood, feelings of guilt and worthlessness, feelings of helplessness and hopelessness, psychomotar retardation, loss of appetite, and sleep disturbance. In order

to capture these components, short self assessed questions are asked to respondents. The original CES-D scale has 20 short self assessed questions but other versions exist such as the one in NIDS where 10 short questions are selected to represent the above components.

The depression scale contains 10 symptoms, any of which may be experienced occasionally by healthy people. A seriously depressed person is expected to experience many of these symptoms but not necessarily all of them concurrently. It is also noted that positive and negative affect (behaviours) are expected to co-exist, with a low (negative) correlation. Klein (1974) has suggested that severely depressed patients are characterized by absence of positive as well as negative affect[7] so that positive and negative affect will be highly (negatively) correlated. There is also evidence in the literature that shows that different types of people may manifest different types of symptoms, for example, lower socioeconomic status people report more physical symptoms while higher socioeconomic status people report more affective symptoms (Crandell and Dohrenwend, 1967; Hamilton et al., 2005). This potentially is a source of measurement error in our data.

Although the CES-D is not designed for clinical diagnosis, the scale is based on symptoms of depression as seen in clinical cases. The CES-D scale is expected to correlate well with other scales meant to measure depression, such as nurse-clinician rating or Raskin scale, or Hamilton Clinician scale. We can not verify this in this thesis because no other scales of depression are available in the data for comparison.

The responses collected from the CES-D scale are scored into an index that is used to determine the level of depression in the population. The CES-D is not intended as a clinical diagnostic tool. Interpretations of individual scores should not be made. Even group averages are interpreted as indicating levels of depressive symptoms and not rates of illness. The CES-D tool is useful for

2.3 Explanatory variables

As we mentioned earlier, the salient feature of NIDS is that a rich set of socioeconomic and demographic variables important for our study were collected alongside health outcome variables explained above. In this chapter we focus on the variables that are supported by theory, as highlighted in health human capital theory and have been used in empirical work (Grossman, 2000). We mention them briefly here because we describe them in detail in the next chapter. The most prominent are income, education, and age. Other explanatory variables common in empirical work include location of residence, marital status, and race.

In NIDS there are two sources of income data., One is a 'one-shot' question asked in the

[7]This term is used in psychology to mean emotions.

household questionnaire about the total amount of after tax household income received in the previous month. The second source of income data is individual level income questions across all sources of income (Argent, 2009). We use the household income provided in the data which is the self reported income of all individuals in the household. The one shot source is used when the individual income[8] is missing (Argent, 2009). We obtain per capita income by dividing the household income by the number of people living in the household and adjust for inflation using the consumer price index (CPI) deflator provided by Statistics South Africa. For the purpose of our analysis we use the logarithm of this income variable in all our analyses to reduce the effect of outliers in the sample. The real per capita income was computed as a simple household mean income to allow for comparison between our results and other results that have used NIDS data. In addition, there is no established equivalence scale for South Africa (Woolard and Leibbrandt, 2006).

Table 2.1: Variable Description for Health and Socioeconomic Variables used in the Analysis

Variable	**Description**
Self Assessed Health	Ordinal with five categories, 1= poor, 2= fair, 3=good, 4=very good 5=excellent
Body Mass Index	Continuous score also categorized as , Underweight <18, Normal 18-24.9, over weight 25-29.9, Obese 30+
Depression	Depression score ranges from 0-30 (higher scores indicate depressive symptoms)
Hypertension	Categorical, Normal (sys<120,dia<80), Prehypertension (sys 120-129, dia 80-90) HBP Stage1 (sys 130-140, dia 90-100) HBP stage 2, Hypertensive crisis (180+, 110+)
Education	Categorical with five categories, 0=No education, 1=General education, 2=Further education and training, 3=Matric, 4=Higher education
Income	Real per capita monthly income
Race	Four race categories 1=African 2=Coloured 3=Asian 4=White
Region	Categorical with three categories 1=Traditional area 2=Urban 3=Farms
Marital status	Categorical with five categories 1=married, 2=living with partner, 3=Widow/widower, 4=Divorced/separated, 5=Never married
Gender	Binary (Male=1)
Age	Measured in age bands of 10 years from age 15
Province	Categorical 1=Western Cape, 2=Eastern Cape, 3=Northern Cape, 4=Free State, 5=Kwa-Zulu Natal 6=North West, 7=Gauteng 8=Mpumalanga 9=Limpopo

Source: Own compilation. The table shows the description of variables used in this study. HBP is High Blood Pressure

For our education indicator, we create five categories of educational attainment, with no education as the base category. This variable is constructed from the information regarding the highest grade in school the respondent successfully completed at the time of the survey. Those who have attained grade 1 to grade 9 are categorized as having attained general education. Those with grade 10 and 11, as well as a National Training Certificate (NTC) 1-2

[8]We acknowledge that income is one of the most contaminated variables in most surveys. In NIDS however, we rely on the findings of a technical working paper number 3 to

are categorized as having attained further education and training. Respondents with grade 12 and NTC 3 are categorized as having attained Matric. Those with educational attainment above grade 12 are categorized as having attained Higher education.

This categorization is in line with the national department of education which uses similar categories in their reports (NDoE 2013). Previous work on the analysis of Health using NIDS data, by Ardington (2009), uses comparable categories, making it easier for comparison of results. Besides, the use of education categories rather than years of schooling in the regressions provides a flexible functional form, with attention to the key graduation points in the South African education system.

In all our analyses we include demographic variables such as age, gender, race and marital status, as controls. The race variable is particularly important for South Africa because of its Apartheid history. We also control for geographical type variables, namely traditional area, urban areas, and farms, to cater for the uniqueness of these regions. Regional fixed effects are relevant control variables because of the high regional heterogeneity in health status in South Africa. This is partly due to lack of a uniform health care funding system at provincial level, and differential levels of development that have an influence on the availability of health inputs. For this reason we control for province in all our analysis.

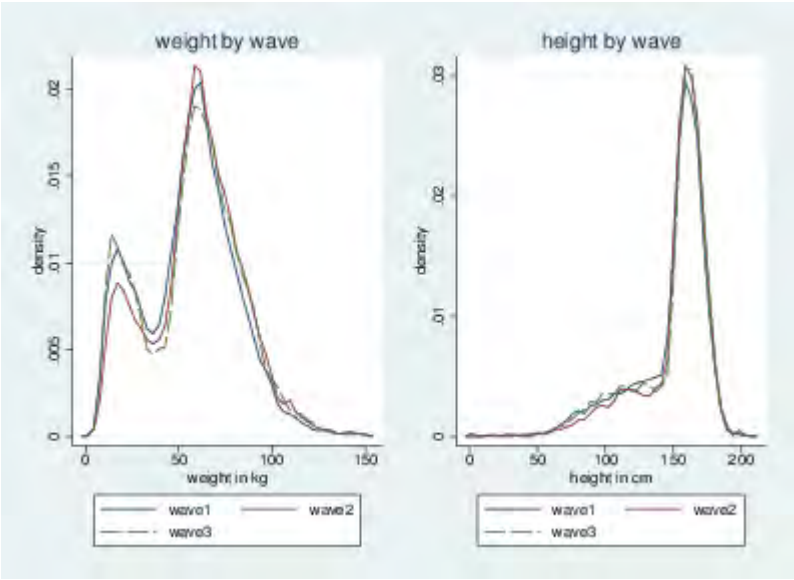
Age is an important explanatory variable because usually health depreciates with age and therefore this variable must be controlled for in the analysis. Age is collected in years in NIDS. We also create 10 year age categories from the continuous age variable in the data. This ensures that we do not impose a functional form a priori but let the data decide. Table 2.1 summarizes the health outcome variables and key explanatory variables used in this study.

In this chapter the main aim is to check whether attrition affects parameter estimates for the health outcome models. The explanatory variables we use in these models are therefore those that potentially affect attrition. We followed the literature in selecting the variables to include in these models. This was also to allow for comparison of of our results with previous studies undertaken in this area for example by Baigrie and Eyal (2014).

2.4 Summary Statistics

Table 2.2 shows summary statistics for the variables we analyze in this study. The mean of categorical variables show weighted proportions of individuals in these categories. We apply frequency weights which indicate how many observations each observation really represents. The weighted proportions are therefore a representation of the total population of South Africa. The proportion of respondents with poor and fair health is lower than those reporting excellent and very good health across all waves. However, the proportion of those reporting poor and fair health declines from wave 1 to wave 2 and wave 3 partly implying that average health improved from wave 1 to wave 3. In wave 3, the proportion of respondents who

Figure 2.1: Kernel density estimates of variation in weight and height across waves



Notes: 1). Panel one shows kernel density of weight of individuals in kg while panel two shows kernel density for their height in cm for the three waves. 2). The kernel density graphs are weighted using cross-sectional post stratification weights provided in each wave. The figures show that there is some variation in weight but not in height of individuals over time.

reported poor health decreased from wave 2 at the same time those who reported excellent health also decreased. The direction of change in average health outcome therefore depends on relative change (decrease/increase) between those with poor health and excellent health.

More than 50% of South Africans are either underweight or overweight, as indicated by a BMI of less than 18.5 or greater than 25 respectively. The proportion of overweight (BMI>25) respondents has dramatically increased over a period of just four years by 5 percentage points (from 47% in wave 1 to 52% in wave 3). At the same time the fraction of respondents who are underweight has declined over the same period from 7.6% in wave 1 to 7.2% and 5.4% in wave 2 and wave 3 respectively. This should be a source of concern since being overweight is a risk factor for many chronic diseases, including cardiovascular diseases. The increase in BMI is associated with an increase in weight because height of adults is fairly fixed. This is evident from the kernel density estimates of weight and height in figure 2.1. The Figure shows that there was more variation in weight relative to height, as expected. The weight of heavier people in wave 1 is higher in wave 2 and 3. This translates to an increase in BMI in wave 2 above the increase in wave 3 for the overweight.

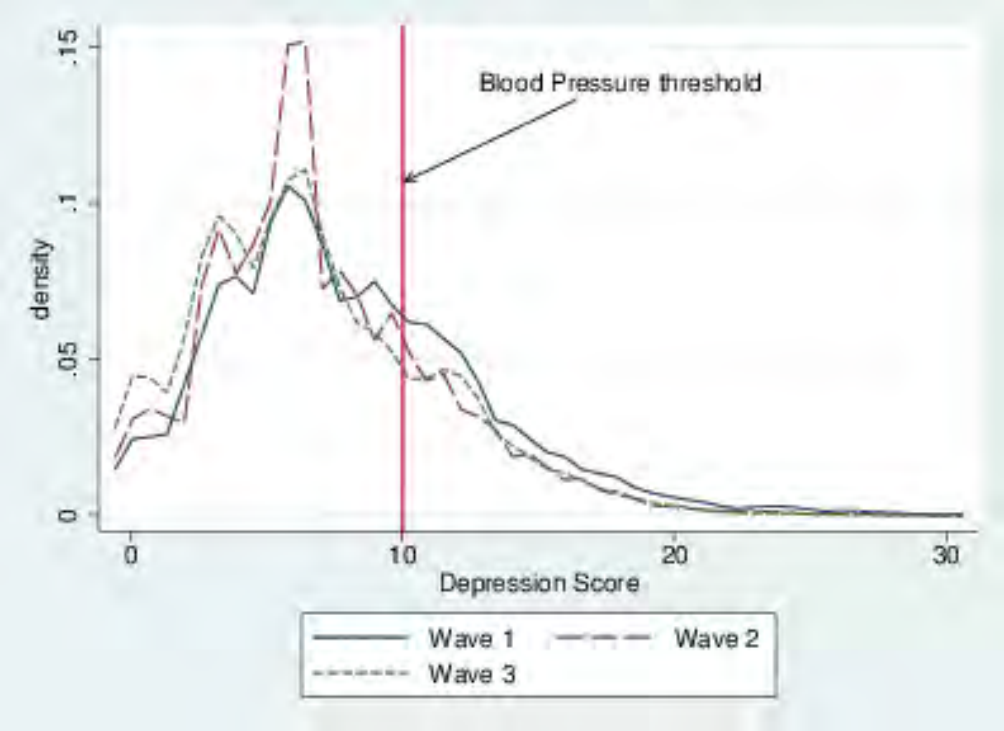
The mean score for emotional (mental health or depression score) health in the population declined from a score of 7.9 in wave 1 to 6.8 and 6.6 in wave 2 and wave 3 respectively. A decline in this score indicates a decrease in depressive symptoms in the population. Figure 2.2 shows the kernel density plot of the depression score for the three waves in NIDS which indicate that there has been a reduction in depressive symptoms from wave 1 to wave 3.

Table 2.2: Summary Statistics

	Wave 1		Wave 2		Wave 3	
Variable	**Mean**	**sd**	**Mean**	**sd**	**Mean**	**sd**
SAH Status						
Poor	0.063	0.243	0.030	0.170	0.030	0.169
Fair	0.114	0.318	0.071	0.256	0.080	0.272
Good	0.230	0.421	0.194	0.395	0.260	0.439
Very good	0.271	0.445	0.297	0.457	0.291	0.454
Excellent	0.321	0.467	0.408	0.492	0.338	0.473
BMI	25.54	6.036	26.14	6.286	26.19	5.927
Emotional health score	7.88	4.74	6.75	4.08	6.60	4.36
Blood pressure Systolic	124.46	22.67	123.89	21.67	124.45	21.15
Blood pressure Diastolic	80.95	14.64	80.97	14.13	81.85	13.68
Per capita monthly real income	1332	1404	1181	1202	1325	1186
Education attained						
No education	0.089	0.284	0.072	0.258	0.064	0.244
General education	0.357	0.479	0.356	0.479	0.352	0.477
Further Education & Training	0.245	0.430	0.252	0.434	0.263	0.440
Matric education	0.201	0.401	0.198	0.399	0.176	0.381
Higher education	0.108	0.310	0.123	0.328	0.146	0.353
Age categories						
15-24	0.281	0.450	0.274	0.446	0.262	0.440
25-34	0.249	0.433	0.250	0.433	0.245	0.430
35-44	0.186	0.389	0.190	0.392	0.197	0.398
45-54	0.132	0.338	0.133	0.339	0.134	0.341
55-64	0.0826	0.275	0.085	0.279	0.088	0.283
65+	0.069	0.253	0.068	0.253	0.074	0.262
Race						
African	0.780	0.414	0.781	0.414	0.787	0.409
Coloured	0.088	0.284	0.089	0.284	0.093	0.290
Asian	0.026	0.158	0.028	0.164	0.026	0.158
White	0.105	0.307	0.102	0.303	0.094	0.292
Province						
Western Cape	0.109	0.312	0.110	0.313	0.114	0.318
Eastern Cape	0.123	0.328	0.119	0.323	0.123	0.329
Northern Cape	0.0222	0.147	0.0228	0.149	0.0218	0.146
Free State	0.0576	0.233	0.0566	0.231	0.0564	0.231
Kwa-Zulu Natal	0.182	0.386	0.193	0.395	0.183	0.386
North West	0.0704	0.256	0.0676	0.251	0.0665	0.249
Gauteng	0.255	0.436	0.256	0.436	0.265	0.441
Mpumalanga	0.077	0.267	0.076	0.265	0.072	0.258
Limpopo	0.103	0.304	0.099	0.299	0.098	0.298
Geographic types						
Traditional area	0.304	0.460	0.319	0.466	0.318	0.466
Urban	0.618	0.486	0.611	0.488	0.625	0.484
Farms	0.079	0.269	0.070	0.255	0.057	0.232
Male	0.461	0.498	0.462	0.499	0.457	0.498
Marital status						
Married	0.315	0.464	0.300	0.458	0.284	0.451
Live with partner	0.088	0.283	0.072	0.258	0.065	0.246
Widow	0.067	0.250	0.062	0.241	0.066	0.248
Divorced	0.031	0.173	0.027	0.161	0.029	0.166
Never married	0.500	0.500	0.540	0.498	0.557	0.497

Notes: 1). Summary statistics are weighted using cross-sectional post stratification weights provided for each wave. 2). Mean of categorical variables in each wave indicate the weighted proportion of individuals in the category.

Figure 2.2: Depression score by wave (unweighted)



Notes: 1). The diagram shows the kernel density of unweighted depression score in each wave obtained by summing the scores of the CES-D questions for each person. 2). The vertical line at depression score 10 shows the threshold depression score beyond which individuals are categorized as depressed.

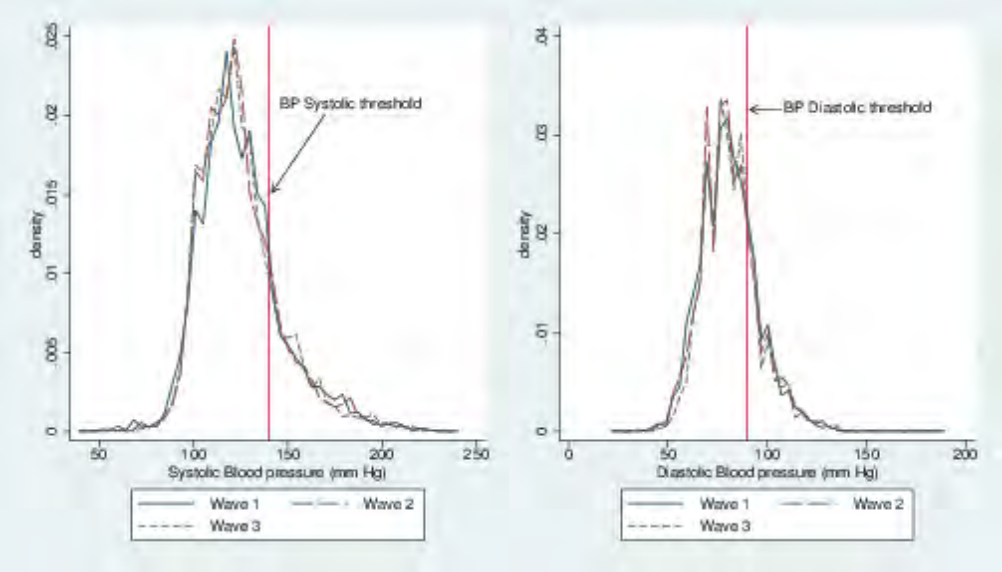
These mean scores were statistically significantly different from each other.

Mean systolic and diastolic blood pressure measurements have remained the same at 124 and 81 mm Hg respectively across the waves of NIDS. This is also confirmed by the kernel densities in figure 2.3 which lie on top of each other for the three waves.

The mean per capita monthly real income in wave 1 was R1666. This dropped to R1479 then rose slightly to R1576 in waves 2 and 3 respectively. Over 35% of the population have general education followed by Further Education and Training qualification (over 24%) and Matric (18%) on average across all waves. About 9% of the population had no education in wave 1 and this is reduced to 7% and 6% in wave 2 and wave 3 respectively. The proportion of individuals in the population with higher education has been steadily increasing from 11% in wave 1 to 12% in wave 2 and 15% in wave 3. This has the implication of increasing mean health outcomes in the population, given the positive correlation between health outcome and education attainment.

53% of respondents were aged between 15 and 34 years but this decreased slightly over the four year period to 52% and 50% in waves 2 and 3 respectively, as people grow older in the sample. This should translate to depreciation in health in line with the life cycle theory of health human capital. Africans constitute the majority of the population in the country, averaging 78% across the waves. This is followed by Whites at 10%, Coloureds at 9% and

Figure 2.3: Systolic and Diastolic blood pressure by wave



Notes: 1).The diagram shows the kernel density of blood pressure systolic and blood pressure diastolic measurements. 2). The vertical lines indicate thresholds of hypertension status for systolic and diastolic measurement respectively.

Asians at 3%. Because of the small numbers of Asians in the sample, results obtained for this group have no power.

The NIDS sample was not drawn to be representative at the provincial level, therefore analysis of the results at province level is not recommended. (Leibbrandt et al., 2009). Nevertheless individuals were drawn from all provinces, with a higher proportion of respondents drawn from KwaZulu-Natal province. Four provinces including KwaZulu-Natal, Western Cape, Eastern Cape and Limpopo had over 10% of the sample each. The other five provinces (Northern Cape, Free State, North West, Gauteng and Mpumalanga) had less than 10% of the sample each. The majority of respondents (over 60%) live in urban areas in these provinces, with the rest in traditional areas (approximately 30%) and on farms (less than 10%).

In this section we have shown that we have a usable set of health outcomes, all of which have wide distributions, including into dangerous areas of ill health. We now move in the next section to show that attrition does not distort this picture.

2.5 Response rates and attrition in NIDS

In table 2.3 transition matrices of the current version[9] of the data show that 26,776 respondents were successfully interviewed in wave 1. However only 21,108 of these were interviewed successfully in wave 2, making up a success rate of 79%. 839 panel individuals (3%) had died by wave 2. About 2% (528) of those who had been interviewed in wave 1

[9]Wave 1 version 5.2, wave 2 version 2.2 and wave 3 version 1.2

Table 2.3: Transition matrix wave 1 to wave 2

	Wave 2 individual outcome					
Wave 1 outcome	Successful	Refused	Hh non response	Moved	Deceased	Total
Successful	21,108	528	4,252	49	839	26,776
	(78.83)	(1.97)	(15.88)	(0.18)	(3.13)	(100)
Refused/NA	950	91	365	2	42	1,450
	(65.52)	(6.28)	(25.17)	(0.14)	(2.90)	(100)
Total	22,058	619	4,617	51	881	28,226
	(78.15)	(2.19)	(16.36)	(0.18)	(3.12)	(100)

Notes: The table shows the unweighted transition matrix between wave 1 and wave 2. Percentages are shown in parentheses and row totals add to 100%. 26,776 were interviewed in wave 1 while 22,058 were interviewed in wave 2 out of 28,226 that were identified in wave 1. 1450 respondents refused or were not available for interview in wave 1. Some of those who had refused to be interviewed (950) in wave 1 were however tracked and interviewed successfully in wave 2.

refused to be interviewed in wave 2. In wave 2 there was no response from households which contained 16% of individuals. 49 respondents had moved out of their original residences by wave 2. Thus the attrition rate was 21% in wave 2. This attrition rate is lower[10] than that reported in earlier panel surveys in South Africa such as the Kwazulu Natal Income Dynamics Study (KIDS) which reported annual[11] attrition rates of 8.37% (Brown et al., 2012; Baigrie and Eyal, 2014). African respondents in NIDS had an attrition rate of 18%, mainly through non contact and death. This group thus had a lower attrition rate than that of all groups combined. However, they contribute more to overall attrition due to their high proportion in the sample.

In table 2.4 the transition matrix between wave 2 and wave 3 shows there was an improvement in re-interview rate. 89% of respondents interviewed in wave 2 were successfully re-interviewed in wave 3, making the attrition rate only 11% (see also Nwosu (2014)). About 68% (423) of those who had refused or were not available in wave 2 were successfully interviewed in wave 3. In addition, 2% (441) respondents died between wave 2 and wave 3, while 8% and 1% were in households that did not respond or refused respectively.

Comparable figures for attrition of Africans in the sample show that the attrition rate for this group was 10%. Non contact and death were the biggest reasons for attrition at 7% and 2% respectively. Refusal rate was very small at less than 1%. Despite this rate being lower than the sample average, attrition by Africans contributed the most in terms of overall attrition because of their high proportion in the sample.

We are also interested to see how many individuals were successfully interviewed four years after having been interviewed in wave 1. In this case we consider the 26,776 individuals

[10]It translates to annual attrition rate of 0.057 for NIDS. See Alderman et al. (2001) for the formula.

[11]The formula for annual attrition by Alderman et al. (2001) is $r = 1-(1-q)^{1/t}$Where q is overall attrition rate and t the number of years covered by the panel.

Table 2.4: Transition matrix wave 2 to wave 3

	Wave 3 individual outcome						
Wave 2 outcome	Successful	Refused	Hh non response	Moved	Deceased w3	Deceased w2	Total
Successful	19,645	208	1,763	1	441	0	22,058
	(89.06)	(0.94)	(7.99)	(0)	(2)	(0)	(100)
Refused/Not available	423	32	152	0	12	0	619
	(68.34)	(5.17)	(24.56)	(0)	(1.94)	(0)	(100)
Hh non response	2,301	77	2,073	13	153	0	4,617
	(49.84)	(1.67)	(44.90)	(0.28)	(3.31)	(0)	(100)
Moved	6	0	3	42	0	0	51
	(11.76)	(0)	(5.88)	(82.35)	(0)	(0)	(100)
Deceased	0	0	0	0	0	881	881
	(0)	(0)	(0)	(0)	(0)	(100)	(100)
Total	22,375	317	3,991	56	606	881	28,226
	(79.27)	(1.12)	(14.14)	(0.20)	(2.15)	(3.12)	(100)

Notes: The table shows the unweighted transition matrix between wave 2 and wave 3. Percentages are shown in parentheses and row totals add to 100%. 22,058 were interviewed in wave 2 while 22,375 were interviewed in wave 3 out of 28,226 that were identified in wave 2, (having been tracked from wave 1). Only 317 respondents refused or were not available for interview in wave 3. Some of those who had refused to be interviewed (423) in wave 2 were, however, tracked and re-interviewed in wave 3.

who were interviewed in wave 1 where 21,384 (80%) were successfully re-interviewed in wave 3, implying that attrition was 20% between wave 1 and wave 3. Nearly 14% of those who attrited in wave 3 were in households that did not respond and 2% and 3% of respondents had died by wave 2 and wave 3 respectively. The remaining 1% were individuals who refused to participate in subsequent rounds of the survey.

Africans still have the lowest attrition rate (18%) between wave 1 and wave 3 .Whites have the highest rate of attrition with more than half (51%) of those targeted not resurveyed. Attrition among Whites is largely due to non contact (44%). Coloureds and Asians have small samples and their attrition rates are 20% and 38% respectively. Attrition of Africans in wave 3 is mainly due to non contact (11%) and death by wave 2 and wave 3 at 2% and 3% respectively. Again because of the high proportion of Africans in the sample they contribute more to overall attrition. Attrition of Asians reduces their sample and substantially affects findings for this group, given that they were only 3% in the base sample.

Besides race, the other variables affecting attrition include age, geographic type of residence (urban, traditional area and farms), income, gender and location of residence (province). For example there is no differential attrition across gender. Western Cape and Gauteng Provinces have the highest per capita attrition rates of 28% and 25% respectively (Brown et al., 2012). Attrition is a potential threat to the parameters we estimate from subsequent waves due to bias that may be introduced especially when such attrition is non random. But as noted by Maluccio (2004), the levels of attrition alone need not necessarily distort

Table 2.5: Transition matrix wave 1 to wave 3

Wave 1 outcome	Wave 3 individual outcome						Total
	Successfully interviewed	Refused/ Not available	HH non-response	Moved	Deceased w2	Deceased w3	
Successful	21,384	267	3,665	55	566	839	26,776
	(79.86)	(1)	(13.69)	(0.21)	(2.11)	(3.13)	(100)
Refused/NA	991	50	326	1	40	42	1,450
	(68.34)	(3.45)	(22.48)	(0.07)	(2.76)	(2.90)	(100)
Total	22,375	317	3,991	56	606	881	28,226
	(79.27)	(1.12)	(14.14)	(0.20)	(2.15)	(3.12)	(100)

Notes: The table shows the unweighted transition matrix between wave 1 and wave 3. Percentages are shown in parentheses and row totals add to 100%. 26,776 were interviewed in wave 1 while 22,375 were interviewed in wave 3 out of 28,226 that were identified in wave 1. 1450 respondents refused or were not available for interview in wave 1. Some of those who had refused to be interviewed (991) in wave 1 were, however, tracked and re-interviewed in wave 3.

inferences made from panel data, as this depends on the model and settings. The processes underlying attrition in the data may differ in such a way that the errors from the structural model and the selection model are not correlated. We therefore apply formal tests to check for the effects of sample attrition in the next section.

2.6 Testing for the effects of sample attrition in NIDS

There is not necessarily a relationship between the size of sample loss from attrition and the existence or magnitude of attrition bias (Fitzgerald et al., 1998; Maluccio, 2004). Even a large amount of attrition causes no bias unless it is non random. In this section we compare the health outcomes of attritors and non-attritors before we apply a formal test to establish if there is attrition bias for the health related data in NIDS. This is necessary because the analysis in this study makes use of the data from all three waves. We focus on the adult sample but exclude adult responses that were collected through proxies. This is necessary because some of the health indicators required personal responses and personalized anthropometric measurements.

Our test is confined to testing the effects of attrition bias as a result of selection on dependent variables[12]. In the attrition probits, health outcome variables are included in the model as explanatory variables. Provided there is enough variation in these health outcome variables, to allow a regression estimation, selection on the basis of independent variables can be tested. Selection on the basis of independent variables is not a serious problem other than that it results in a smaller sample size (Wooldridge, 2002). If selection on the dependent variable is non random, OLS will result in biased and inconsistent parameter estimates.

In table 2.6 the mean of health outcomes and the number of people in the sample present-

[12]Selection on dependent variable is also referred in the literature as endogenous sample selection.

Table 2.6: Summary statistics for adult attritors and non attritors

variable	**Non attritors**			**Attritors**			t-Test for significance	
variable	**mean**	**sd**	**N**	**mean**	**sd**	**N**	of difference in means	
							t	**p- value**
Poor health	0.078	0.268	11927	0.0543	0.227	2964	-8.6069	0.0000
Fair health	0.136	0.343	11927	0.0958	0.294	2964		
Good health	0.245	0.430	11927	0.230	0.421	2964		
Very good health	0.251	0.434	11927	0.280	0.449	2964		
Excellent health	0.289	0.453	11927	0.340	0.474	2964		
BMI	25.57	6.3156	10548	25.239	5.899	2468	-4.5003	0.0000
Depression-CESD_10	8.085	4.755	11936	7.645	4.934	2968	3.1774	0.0015
Depression CESD_8	5.278	4.598	11940	4.936	4.657	2969	1.3933	0.1636
Systolic pressure	126.846	24.376	10937	125.169	22.445	2556	4.4739	0.0000
Diastolic Pressure	81.765	15.219	10935	81.306	14.088	2555	3.6156	0.0000

Notes: The table shows means, standard deviation (sd) and sample size (N) for non attritors and attritors in wave 1 by attrition status in wave 2. The mean is the proportion of the sample that is in each of the health categories in wave 1. The t test for difference in means between non attritors and attritors as well as p-values are also shown. The mean differences between attritors and non attritors for SAH, BMI , Blood pressure systolic and blood pressure diastolic are statistically significant. CESD_10 and CESD_8 indicate the two alternative measures of depression score where 10 and 8 questions are used respectively.

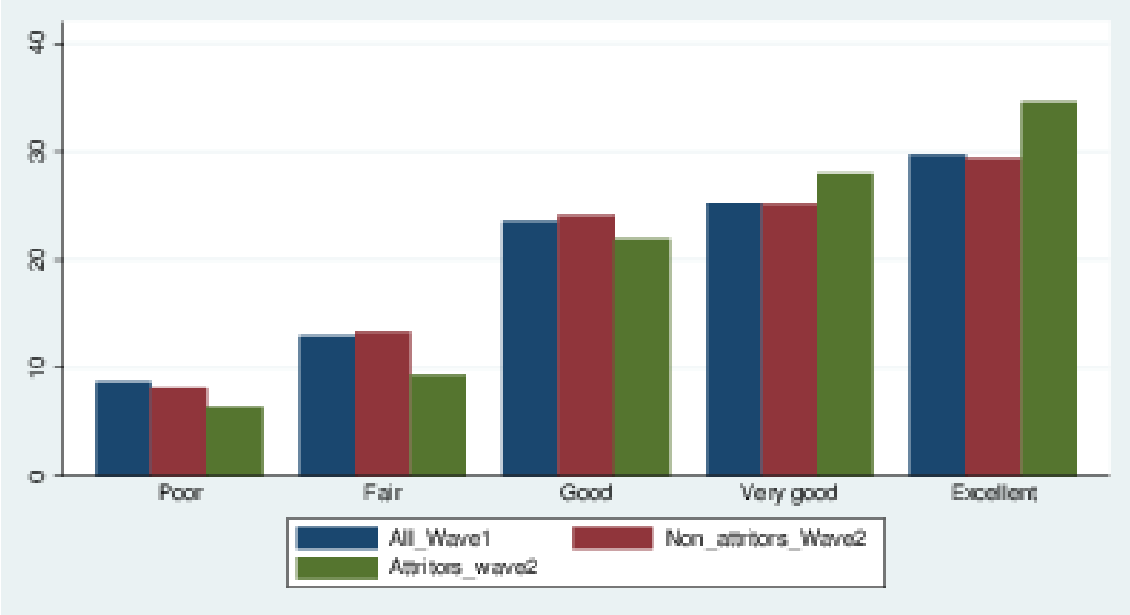
ing with these health outcomes is shown for those who did not attrit and those who attrited in wave 2. For the health outcome variables which are categorical, the mean is the proportion of respondents in these health categories. The proportions in the health category sum to 100 percent. The purpose of this summary table is to show differences in the proportions for attritors and non attritors. The proportions differ for attritors and non attritors and the differences are statistically significant. The mean for depression scores and high blood pressure for non attritors is higher than for attritors.

In figure 2.4 the densities of SAH categories for wave 1, non attritors and attritors in wave 2 respectively are shown. As we can see the densities for non attritors in wave 2 resemble the densities of SAH in wave 1. However the densities for attritors in wave 2 are different from non attritors for all the SAH categories. In particular, the graph shows that a smaller proportion of individuals who had poor health in wave 1 attrited in wave 2. A higher proportion of individuals who had very good and excellent health in wave 1 attrited in wave 2, compared to non attritors in wave 2.

The story for attrition in wave 3 is similar to that for wave 2. Figure 2.5 shows SAH by attrition status in wave 3. The densities of SAH categories in wave 1, and by attrition status in wave 3 are shown. As in the case of attrition in wave 2, individuals who had poor health in wave 1 are less likely to attrit in wave 3 compared to those who had very good and excellent health.

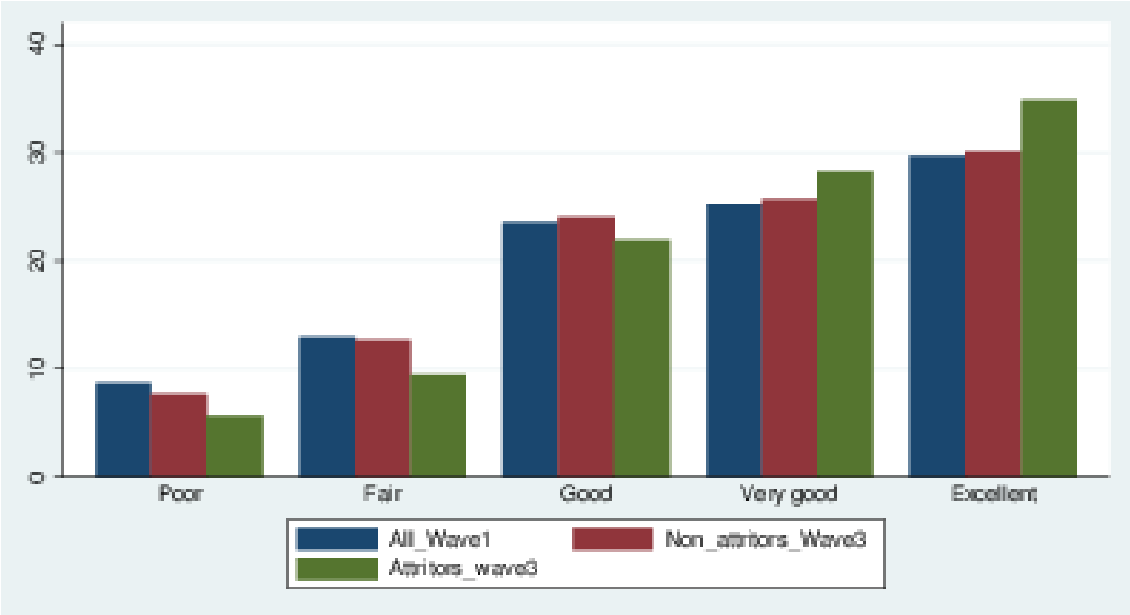
In figure 2.6 we present kernel density plots of the depression score obtained from the sum of scores of the CES-D 10 scale in wave 1 for each individual in the sample and in the

Figure 2.4: SAH by Attrition Status in wave 2



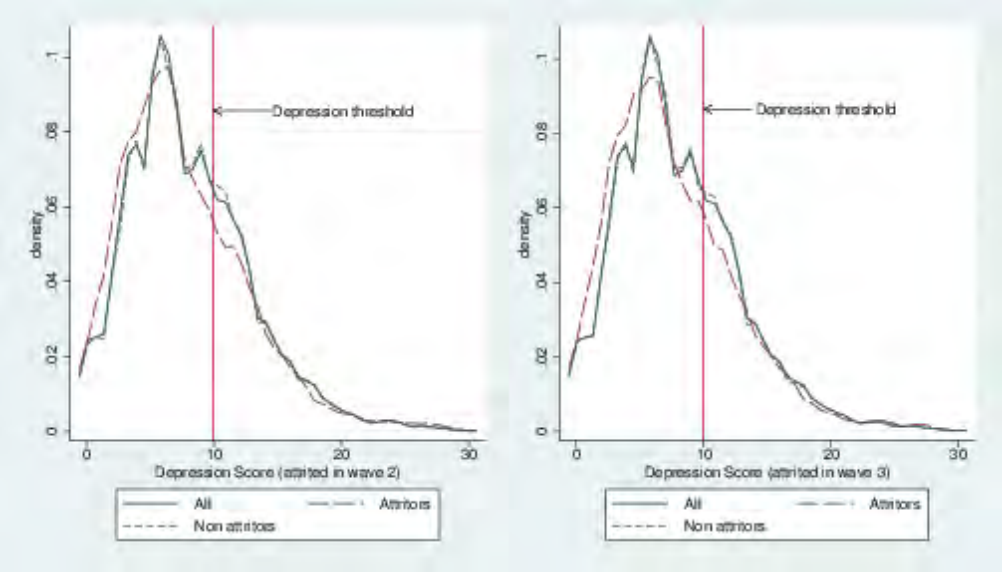
Notes: The bar graph shows un-weighted proportions for SAH for the entire sample in wave 1 and for non-attritors and attritors by wave 2 in each SAH category respectively. The distribution for Non-attritors resembles the entire sample in wave 1 but attritors are distributed differently.

Figure 2.5: SAH by Attrition Status in wave 3



Notes: The bar graph shows unweighted proportions for SAH for the entire sample in wave 1 and for non attritors and attritors by wave 3 in each SAH category respectively. The distribution for Non attritors resembles the entire sample in wave 1 but attritors are distributed differently.

Figure 2.6: Depression by Attrition Status in wave 2 and wave 3



Notes: The figure shows unweighted kernel density plots for the depression (CES-D 10) score for the entire sample in wave 1 and for attritors and non attritors in wave 2 and wave 3 in panel 1 and 2 respectively. The kernel density plot for non attritors traces the entire sample closely, while that for attritors falls to the left of the entire sample. This implies that a higher proportion of attritors have fewer depressive symptoms. Vertical lines indicate thresholds beyond which individuals are categorized as having depressive symptoms.

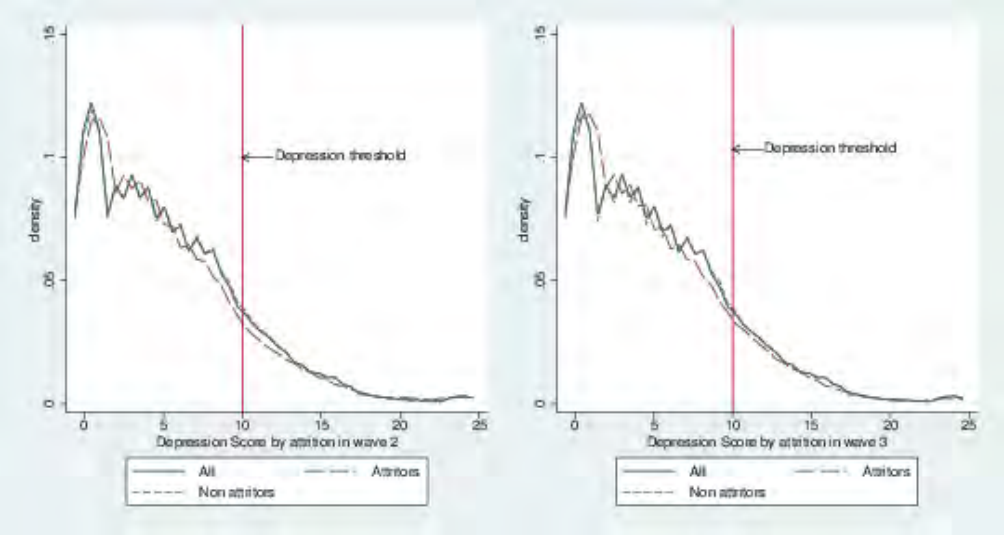
two groups, attritors and non attritors, in wave 2 and wave 3. We categorize respondents with a score of more than 10 on the depression score as depressed and those with a score of less than or equal to 10 as not depressed, using guidelines provided by the WHO (Radloff, 1977). The kernel density for Non attritors in wave 2 and wave 3 is similar to that of the entire sample in wave 1. However the kernel densities for attritors in wave 2 and 3 are to the left of non attritors implying they have a lower mean depression score.

In figure 2.7 we use the depression (CES-D 8) scoring which excludes positive questions in the CES-D list and only considers the eight negative questions to derive the score. Still the attritors and non attritors are slightly different in such away that non attritors mimic the density of the entire sample in wave 1 while attritors tend to fall to the left. This provides enough evidence to make it essential that we test more formally for the effects of attrition.

In figure 2.8 we show kernel density for BMI scores for the whole sample in wave 1 and for the sample split by attrition status in wave 2 and wave 3. The figures show that the attritors had a higher mean BMI score than the entire sample and non attritors. The BMI score for non attritors closely resembles that of the whole sample.

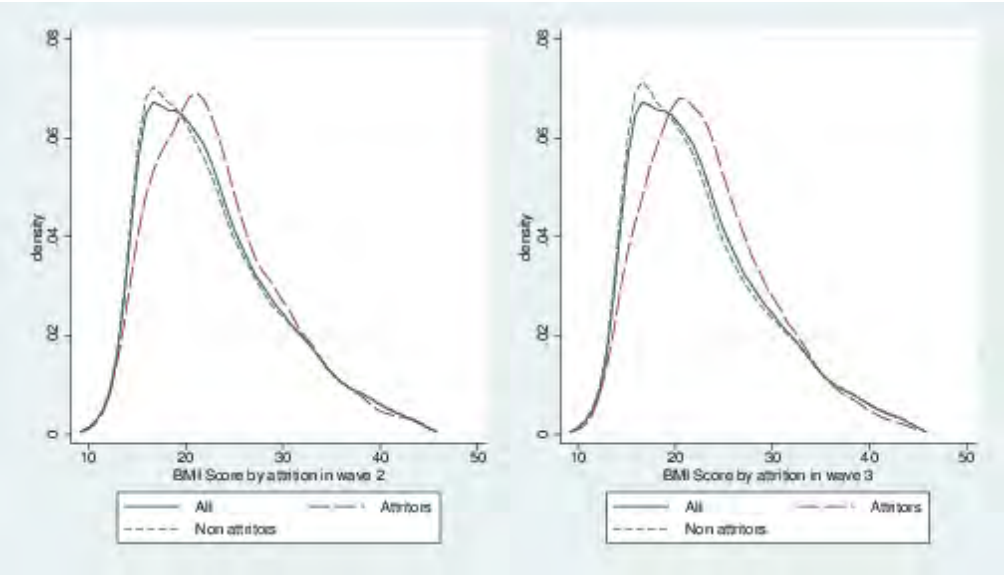
Despite the noticeable differences between attritors and non attritors shown by the kernel densities, the figures do not indicate whether attrition is random. We therefore proceed to formally test if the attrition is random.

Figure 2.7: Depression score (CESD-8) by attrition in wave 2 and wave 3



Notes: The figure shows unweighted kernel density plots for the depression (CES-D 8) score computed from 8 negative questions for the entire sample in wave 1 and for attritors and non attritors in wave 2 and wave 3. The kernel density plot for non attritors traces the entire sample closely, while that for attritors is to the left of the entire sample. This implies that a higher proportion of attritors have fewer depressive symptoms.

Figure 2.8: BMI by attrition status in wave 2 and wave 3



Notes: The figure shows unweighted kernel density plots for BMI scores for the entire sample in wave 1 and for attritors and non attritors by wave 2 and wave 3. The kernel density for non attritors traces the entire sample closely, while that for attritors is to the right of the entire sample. This implies that attritors are more likely to be overweight than non attritors.

2.6.1 Theoretical insights on effects of attrition on parameter estimates

The selection bias model relevant to the problem of attrition in panel data is one where selection is on endogenous observables such as lagged dependent variables. In such a model consistent parameter estimates are obtained with weighted least squares using weights constructed from estimated selection probability functions. In this section we follow the model by Fitzgerald et al. (1998) and Moffit et al. (1999) in exploring attrition bias from selection on observables. The theoretical insight of this model is explained below.

The object of interest is a conditional population density $f(h|x)$ where h is a scalar dependent variable and x is a scalar independent variable. We also define A as a dummy variable indicating that there is attrition when $A = 1$ and 0 if not. Since h is only observed when $A = 0$, we can only estimate the density $g(h|x, A = 0)$. The challenge is how to make inference on $f(.)$ from $g(.)$. To investigate the effects of attrition we use a canonical one period selection model as described in equation 2.1 and 2.2 below.

$$h_i = \mathbf{x_i}'\beta_\mathbf{1} + \varepsilon_i \qquad (h_i \ \ observed \quad only \quad if \quad A_i^* \leq 0) \tag{2.1}$$

$$A_i^* = \mathbf{x_i}'\beta_\mathbf{2} + \mathbf{z_i}'\gamma + v_i \tag{2.2}$$

Equation 2.1 is the model of primary interest which is the structural model. It shows h_i as the outcome variable measured in the second wave of a two period panel dataset. $\mathbf{x_i}$ is a vector of first period measures and z_i is a variable that is endogenous to h and is not in the structural model (such as the lagged value of h). Equation 2.2 represents a selection model that depends on the same independent variables as equation 2.1 as well as additional ones. In other words, x is a subset of z and we represent the extra variable in z as z_i. In practice the latent variable A_i^* is not observed but only an indicator of whether an observation is observed or not. We represent this with an indicator variable $A = 0$ if observed and $A = 1$ if unobserved. If there is correlation between the error terms ε_i and v_i, estimation of equation 2.1 ignoring equation 2.2 leads to biased and inconsistent parameter estimates of $\beta_\mathbf{1}$. In other words, OLS of equation 2.1 on nonattriting sample will generate inconsistent estimates and the estimable density $g(h|x, A = 0)$ will not correspond to $f(h|x)$ because the event $A = 0$ is related to h through z_i. We need additional information or restrictions to infer $f(h|x)$ from $g(h|x, A = 0)$. These can come from the probability of attrition $Pr(A = 0|h, x, z)$, where z is an auxilliary variable (or vector) that is assumed to be observable for all units but not included in x.

In panel data we can correct for sample selection and estimate β consistently even with

attrition using weights provided by the inverse of the attrition probability as in equation 2.3.

$$w(z,x) = \left[\frac{Pr(A=0|z_i,x)}{Pr(A=0|x)}\right]^{-1} \tag{2.3}$$

Where $w(.)$ is the inverse probability weight, $Pr(A=0|z_i,x)$ is the predicted probability from the unrestricted model while $Pr(A=0|x)$ is the predicted probability of the restricted model. Such weights unlike those used in weighted least squares (WLS), are model specific as the z_i variable changes with different choices of h. If there is selection on observables, the critical variable is z, a variable that affects attrition propensities and that is also related to the density of h conditional on x. In this sense z is endogenous to h. We will use a lagged value of h to represent the z_i variable. If z_i is not a determinant of attrition, the inverse probability weight equals one and hence all conditional and unconditional densities are equal and this implies that no attrition bias is present. There is also no attrition bias if h and z_i are independent conditional on x and $A=0$. Two sufficient conditions for the absence of attrition bias due to attrition on observables are either z does not affect A or z is independent of h conditional on x. These two conditions form the basis for two tests for attrition bias from selection on observables; the attrition probit and the pooled test of Becketti Gould Lillard and Welch (BGLW) (Becketti et al., 1988). These tests are complementary to each other and are usually used together (Fitzgerald et al., 1998).

In the attrition probits we establish whether the health outcome[13] variable (h) at the base wave is a determinant of attrition (A). If this variable is not a significant determinant of attrition then the inverse probability weights generated using this variable would equal to one, showing there is no attrition bias on observables when estimating the structural model with this variable as the dependent variable.

The BGLW test entails the reverse of the attrition probits where the health outcome variable (h) in the initial wave of the survey is regressed on the subsequent attrition indicator (A) and other explanatory variables (x). The test for attrition is based on the significance of A in that equation. An alternative BGLW test includes interaction terms between the attrition indicator and explanatory variables and evaluates whether the coefficients of the explanatory variables and the constant differs for those respondents who are subsequently lost to follow-up versus those who are re-interviewed. An F test of joint significance of the attrition dummy and the interaction variables is conducted to determine whether the coefficients from explanatory variables differ between non attritors and attritors from the panel. If the interactions turn out to be significant then estimating the structural model while ignoring attrition biases the results. The two tests (attrition probits and BGLW test) by construction should yield the same conclusion.

[13]This is equivalent to the z_i variable in equation 2.2 because the variable (h) at the base is used

2.7 Results

2.7.1 Attrition Probit results

In tables 2.7 and 2.8 we report the results from the attrition probit regressions. In table 2.7, we show determinants of attrition in wave 2 using an indicator variable for attrition ($A = 1$) as the dependent variable[14]. The indicator of attrition (A) is a binary variable where $A = 1$ if the individual attrited and 0 if the individual was still observed in wave 2. We consider all the health outcomes in turn, where we first estimate a restricted model in a regression with only the health outcome as the explanatory variable, before controlling for the other auxiliary variables at the base wave. The decision on which control variables to include is informed by previous empirical work on determinants of attrition (Baigrie and Eyal, 2014; Brown et al., 2012; Maluccio, 2004; Kasirye and Ssewanyana, 2010).

We consider each health outcome in a separate regression using the same regressors across all models. In model 1, 3, 5 and 7 of table 2.7 we present attrition probit results for the restricted models. The results in model 1 show that the probability of attriting from the panel in wave 2 is higher among individuals with better health status relative to those in poor health. There is also no statistically significant difference in attrition rate between those who reported fair and poor health. In regression 2 we control for a number of auxiliary variables and find that the statistically significant difference between those who have better health and those with poor health disappears. This implies that SAH does not determine attrition in wave 2 of NIDS. The implication of this result is that sample attrition does not bias point estimates in a structural model of SAH in wave 2.

In regression 3 the restricted model includes only BMI scores as the explanatory variable. This model shows that having a higher BMI score is associated with reduced probability of attrition. In regression 4, we control for the same variables as in regression 2 and find that the BMI score no longer determines the probability of attrition in wave 2. This result suggests that it is safe to use BMI in wave 2 in a structural equation without re-weighting with the inverse probability weights because the parameter estimates will still be unbiased.

In regression 5, an increase in depression score which is equivalent to an increase in depressive symptoms is associated with a lower probability of attrition from wave 2. However, when we include control variables in model 6, depressive symptoms do not determine the probability of attrition in wave 2. This means that the depression variable can be used in wave 2 without worrying about attrition bias in parameter estimates.

High blood pressure is included as a binary variable in model 7 and 8. As we can see,

[14]The dependent variable for attrition probits were generated from the individual outcome variable in the data set that indicated if the individual was successfully interviewed in a particular wave.

having high blood pressure does not determine the probability of attrition in wave 2. This is the case even after controlling for other explanatory variables in model 8. This also confirms that parameter estimates obtained by the use of this variable in wave 2 are not affected by sample attrition bias.

We keep the control variables the same across all models in table 2.7 but vary the health outcomes. The partial effects of the control variables remain similar across all models. The exception is the case of educational attainment categories, where there are some inconsistencies. There is no statistically significant difference in the probability of attrition observed between having further education and training and no education when we include high blood pressure as one of the determinants of attrition in the model. However, the difference is statistically significant when we include other health outcomes. Likewise, having Matric relative to no education is not statistically significant when a dummy for high blood pressure is included in the model. Across all the models however, as people grow old they are less likely to attrit from the survey relative to young people, which is expected due to high mobility among the young in search of education or employment. This is also confirmed by the fact that there is no statistically significant difference in attrition between 15-24 year olds, 25-34 year olds and 35-44 year olds because these groups are likely to be similar in many respects including that they may be looking for jobs or be in school. The implication of this is that age should be controlled for in all models of health outcome.

The higher the income the higher the probability of attriting from the sample. This result supports the view noted by Baigrie and Eyal (2014) that there is a positive correlation of income with propensity for refusal. Refusal is more common than non-contact only among the wealthiest 20% of respondents (Baigrie and Eyal, 2014). The high correlation between income and race also implies that Whites are more likely to attrit relative to other race groups because they have higher average incomes. The probability of attriting in wave 2 is higher for Asians and Whites relative to Africans, but it is highest for Whites. We note that the Asian/Indian group is too small in the sample to allow us to make any reliable conclusions from the data. There is no statistically significant difference between Coloureds and Africans when different dimensions of health outcomes are used.

Respondents in other provinces were less likely to attrit relative to those in the Western Cape province. The positive coefficients for some provinces are not statistically significant. Individuals who live in urban areas and on farms have a higher probability of attriting from the survey relative to those who live in traditional areas. In addition, marriage seems to protect people from attriting from the survey, as those who live with a partner (cohabiting) are more likely to attrit compared to those who are married. Those who are divorced too are more likely to attrit in wave 2 compared to those who are married.

The Pseudo R-squared in all the models is less than 10% suggesting that baseline variables

(wave 1 variables included in the regression) explain less than 10% of the panel attrition in wave 2. This leaves about 90% as unexplained (random) attrition. The results in table 2.7 show that all the health outcomes discussed do not determine attrition because they are not statistically significant in the attrition probit regressions. The implication is that inverse probability weights generated from the models in table 2.7 will equal to 1 because of this result.[15] This also implies that conditional density and unconditional density involving these health outcome variables are equal. We can therefore make inference on a structural equation involving these health dimensions from estimable densities of these variables from the sample.

In table 2.8 we report attrition probit regression results for health outcomes in wave 3. We repeat the process shown in table 2.7 but this time the dependent variable is a binary indicating attrition in wave 3. In regression 1 of the table we estimate the restricted model with SAH as the only explanatory variable. Having good health increases the probability of attriting in wave 3 relative to having poor health. The significance disappears with the inclusion of control variables in regression 2, implying that SAH is not a true determinant of attrition in wave 3. The implication of this result is that estimation of parameter estimates in wave 3 using SAH in a structural model will be consistent.

In regression 3 the restricted model includes a continuous variable, the BMI score, as the only explanatory variable. In this model the results show that the probability of attriting from the panel in wave 3 decreases with increasing BMI score but this is not statistically significant. This result does change the sign but remains insignificant when we include control variables in regression 4, showing that this variable is truly not associated with attrition in wave 3.

In regression 5 the results show that the probability of attriting in wave 3 decreases with increasing depressive symptoms, but when auxiliary variables are controlled for in regression 6, the coefficient is no longer significant. In regression 7 having high blood pressure is not significantly correlated with reduced probability of attrition in wave 3. The inclusion of control variables does not make this variable statistically significant, indicating that high blood pressure is not correlated with attrition in wave 3.

The Pseudo R-squared across all the models, shown in table 2.8 is also less than 10%, implying that the variables in these models jointly explain less than 10% of the panel attrition while over 90% is random. The implication of this result is that health dimensions in NIDS are not affected by attrition in wave 3. We can get consistent point estimates by using these health variables in structural equations.

Overall, most of the control variables retain their statistical significance with the use of different health outcomes in the model. For example, the probability of attrition decreases

[15]See equation 2.3 for details. This equation equals one if z (health outcome) does not affect probability of attrition.

with age and educational attainment. There is a statistically significant difference in the probability of attrition between the 25-34 year olds and the 15-24 year olds. In particular, the 25-34 year olds have a higher probability of attriting from the panel relative to the 15-24 year olds. Being White relative to being African increases the probability of attrition. Although the results show significantly higher probability of attriting for Asians relative to Africans, the Asian sample is too small for us to make any inference. Being married reduces the probability of attrition in wave 3, relative to those cohabiting, however there is no difference with other marital status categories. Living in the Western Cape does not significantly reduce the probability of attrition relative to other provinces, except Free State which has statistically significant coefficient albeit at 10% level. Staying in an urban area or on a farm increases the probability of attrition in wave 3, relative to staying in a traditional area.

In table 2.9 we use an alternative CES-D score in the regression. This CES-D score was generated from 8 questions after excluding the two positive questions from the ten questions in the survey, as explained in section 2.2. We use this score as a continuous variable for depression, where a higher score indicates more depressive symptoms. The results are similar to the ones we obtain when we use the CES-D score from 10 questions. The probability of attriting from the survey is not determined by depression in both wave 2 and wave 3. This implies that using this depression indicator in a structural model without weighting with estimated inverse of the probability of attrition would not bias parameter estimates.

2.7.2 The BGLW test for attrition bias

The BGLW test for attrition is implemented by regressing the health outcome at the base wave on the indicator of attrition and the other base level explanatory variables. The test for attrition bias is based on the significance of the attrition indicator in this model. The regression results for attrition in wave 2 and wave 3 are presented in tables 2.10 and 2.11 respectively. The full set of results that include all other explanatory variables are in the appendix to this chapter (A.1 and A.2). Our main focus here is on the statistical significance of the attrition indicator. In the first regression we run an ordered probit model of SAH outcome on all socioeconomic and demographic variables.[16] as well as the attrition indicator. If the attritors and non attritors are similar, then the attrition term in this regression should not be significant.

The results show that attrition in wave 2 is not a significant determinant of SAH. Socioeconomic and demographic variables have the correct signs. For example, older individuals are less likely to report excellent health, while income is positively correlated with reporting

[16]The inclusion of these variables is informed by Grossman theory of health human capital and empirical studies on determinants of health outcomes.

Table 2.7: Attrition Probit Regression Results for Health Outcomes in wave 2

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fair (rel.to Poor)	0.0751	-0.00168						
	(0.117)	(0.119)						
Good	0.203*	0.0456						
	(0.106)	(0.104)						
Very Good	0.357***	0.129						
	(0.109)	(0.107)						
Excellent	0.353***	0.101						
	(0.101)	(0.0971)						
BMI score			-0.00922***	-0.00544				
			(0.00336)	(0.00354)				
Depression					-0.0146**	0.00323		
					(0.00715)	(0.00577)		
High Blood pressure							-0.0591	0.106
							(0.0840)	(0.0909)
25-34 (rel.to 15-24)		0.0879		0.108		0.0679		0.0482
		(0.0655)		(0.0706)		(0.0671)		(0.0854)
35-44		-0.0247		-0.0279		-0.0521		-0.0869
		(0.0669)		(0.0761)		(0.0673)		(0.0942)
45-54		-0.296***		-0.329***		-0.338***		-0.474***
		(0.0915)		(0.102)		(0.0919)		(0.130)
55-64		-0.336***		-0.391***		-0.379***		-0.390**
		(0.107)		(0.117)		(0.108)		(0.155)
65+		-0.323**		-0.390***		-0.370***		-0.380***
		(0.135)		(0.135)		(0.133)		(0.139)
Log real pc monthly income		0.0493*		0.0579*		0.0540*		0.0666**
		(0.0287)		(0.0323)		(0.0286)		(0.0330)
General educ (rel to No educ.)		-0.297***		-0.285***		-0.275***		-0.247**
		(0.0850)		(0.0825)		(0.0833)		(0.105)
FET		-0.236**		-0.250**		-0.212**		-0.186
		(0.102)		(0.0982)		(0.100)		(0.124)
Matric		-0.245**		-0.239**		-0.208*		-0.206
		(0.111)		(0.108)		(0.108)		(0.130)
Higher education		-0.181		-0.157		-0.151		-0.102
		(0.159)		(0.166)		(0.159)		(0.206)
Male		0.164***		0.156***		0.173***		0.190***
		(0.0371)		(0.0368)		(0.0370)		(0.0552)
Coloured (rel. to African)		0.0757		0.0278		0.0775		0.00941
		(0.108)		(0.113)		(0.108)		(0.139)
Asian/Indian		0.417**		0.343*		0.408**		0.392
		(0.194)		(0.203)		(0.194)		(0.272)
White		0.617***		0.517***		0.625***		0.532***
		(0.114)		(0.126)		(0.115)		(0.152)
With Partner		0.234**		0.240**		0.232**		0.317**
(rel. to married)		(0.114)		(0.116)		(0.114)		(0.127)
Widow/widower		0.0293		-0.0884		0.00910		-0.0679
		(0.0905)		(0.104)		(0.0922)		(0.128)
Divorced		0.333***		0.390***		0.316**		0.372**
		(0.128)		(0.147)		(0.123)		(0.160)
Never married		0.0846		0.0829		0.0744		0.128
		(0.0783)		(0.0823)		(0.0782)		(0.0995)
Eastern Cape		0.0593		0.0591		0.0610		0.104
(rel. to Western Cape)		(0.143)		(0.155)		(0.143)		(0.189)
Northern Cape		-0.285**		-0.306**		-0.291**		-0.339**
		(0.125)		(0.130)		(0.126)		(0.157)
Free State		-0.350**		-0.362**		-0.358**		-0.373*
		(0.145)		(0.159)		(0.147)		(0.206)
Kwa-Zulu Natal		-0.377***		-0.374**		-0.388***		-0.369*
		(0.142)		(0.151)		(0.144)		(0.191)
North West		-0.0608		-0.173		-0.0722		-0.135
		(0.202)		(0.200)		(0.204)		(0.240)
Gauteng		-0.139		-0.169		-0.139		-0.246
		(0.128)		(0.128)		(0.131)		(0.156)
Mpumalanga		-0.532***		-0.528***		-0.545***		-0.586***
		(0.147)		(0.168)		(0.147)		(0.206)
Limpopo		-0.211		-0.279*		-0.211		-0.259
		(0.147)		(0.162)		(0.148)		(0.209)
Urban		0.272***		0.261***		0.267***		0.290***
(rel. to Traditional area)		(0.0934)		(0.0986)		(0.0943)		(0.112)
Farms		0.309**		0.321**		0.313**		0.378**
		(0.135)		(0.136)		(0.136)		(0.161)
Constant	-0.964***	-1.082***	-0.486***	-0.879***	-0.577***	-1.052***	-0.729***	-1.158***
	(0.0902)	(0.244)	(0.100)	(0.279)	(0.0757)	(0.245)	(0.0454)	(0.283)
Observations	14,833	14,776	12,962	12,915	14,846	14,791	7,783	7,759
Pseudo R-Squared		0.075		0.066		0.075		0.072

Notes:1) Dependent variable is A=1 if the individual attrited in wave 2 and zero otherwise. 2). Robust standard errors that allow for correlation in un-observables for each individual drawn from the same primary sampling unit are shown in parentheses. 3) In each model we first include only the health outcome before we include the control variables in the next model. In model 1 we include SAH only, 3 we include BMI only, 5 we include depression and model 7 we include hypertension only as the explanatory variable. 4) FET is further education and training 5) *** p<0.01, ** p<0.05, * p<0.1

Table 2.8: Attrition Probit Regression Results for Health Outcomes in wave 3

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fair (rel.to Poor)	0.0753	0.00994						
	(0.116)	(0.112)						
Good	0.193*	0.0197						
	(0.109)	(0.107)						
Very Good	0.290**	0.0557						
	(0.117)	(0.118)						
Excellent	0.326***	0.0859						
	(0.103)	(0.107)						
BMI score			-0.00164	0.00273				
			(0.00342)	(0.00429)				
Depression					-0.0172***	0.00274		
					(0.00589)	(0.00512)		
High Blood Pressure							-0.0914	0.0101
							(0.0868)	(0.104)
25-34 (rel.to 15-24)		0.198***		0.191***		0.185***		0.208**
		(0.0716)		(0.0728)		(0.0708)		(0.0927)
35-44		0.00168		-0.0126		-0.0160		-0.00759
		(0.0703)		(0.0800)		(0.0700)		(0.0955)
45-54		-0.248***		-0.270***		-0.276***		-0.284**
		(0.0939)		(0.101)		(0.0937)		(0.131)
55-64		-0.380***		-0.423***		-0.413***		-0.327**
		(0.106)		(0.128)		(0.106)		(0.153)
65+		-0.326***		-0.301**		-0.366***		-0.172
		(0.123)		(0.128)		(0.122)		(0.159)
Log real pc monthly income		0.0764**		0.0867***		0.0772**		0.134***
		(0.0312)		(0.0334)		(0.0306)		(0.0388)
General educ (rel to No educ.)		-0.253***		-0.195***		-0.241***		-0.130
		(0.0763)		(0.0743)		(0.0742)		(0.114)
FET		-0.232**		-0.195**		-0.218**		-0.117
		(0.0932)		(0.0940)		(0.0924)		(0.136)
Matric		-0.280***		-0.234**		-0.255***		-0.205
		(0.0982)		(0.0938)		(0.0957)		(0.126)
Higher education		-0.120		-0.0576		-0.0932		-0.172
		(0.171)		(0.198)		(0.170)		(0.284)
Male		0.217***		0.239***		0.226***		0.244***
		(0.0411)		(0.0470)		(0.0415)		(0.0521)
Coloured		0.113		0.171		0.112		0.242
(rel. to African)		(0.134)		(0.137)		(0.134)		(0.162)
Asian/Indian		0.351*		0.133		0.350*		0.0186
		(0.210)		(0.189)		(0.212)		(0.157)
White		0.749***		0.640***		0.756***		0.683***
		(0.145)		(0.154)		(0.145)		(0.174)
With Partner		0.267**		0.252**		0.269**		0.206
(rel. to married)		(0.119)		(0.126)		(0.118)		(0.128)
Widow/widower		-0.0200		0.0102		-0.0199		-0.0756
		(0.101)		(0.109)		(0.0992)		(0.144)
Divorced		0.298**		0.249		0.283**		0.245
		(0.140)		(0.174)		(0.138)		(0.167)
Never married		0.0126		0.0155		0.00515		0.0562
		(0.0757)		(0.0864)		(0.0763)		(0.0944)
Eastern Cape		0.113		0.134		0.106		0.193
(rel. to Western Cape)		(0.154)		(0.176)		(0.154)		(0.208)
Northern Cape		-0.0404		-0.109		-0.0530		-0.130
		(0.144)		(0.170)		(0.144)		(0.197)
Free State		-0.347*		-0.292		-0.367*		-0.108
		(0.199)		(0.230)		(0.198)		(0.277)
Kwa-Zulu Natal		-0.189		-0.112		-0.206		0.0203
		(0.146)		(0.165)		(0.147)		(0.194)
North West		0.202		0.169		0.185		0.233
		(0.239)		(0.243)		(0.238)		(0.274)
Gauteng		0.0902		0.144		0.0878		0.0944
		(0.144)		(0.167)		(0.145)		(0.194)
Mpumalanga		-0.0268		0.0180		-0.0368		0.00599
		(0.158)		(0.186)		(0.159)		(0.216)
Limpopo		-0.0896		-0.0599		-0.0970		-0.0185
		(0.163)		(0.183)		(0.163)		(0.216)
Urban		0.224**		0.210**		0.220**		0.201**
(rel. to Traditional area)		(0.0918)		(0.0949)		(0.0930)		(0.102)
Farms		0.333***		0.356***		0.334***		0.351***
		(0.108)		(0.108)		(0.109)		(0.126)
Constant	-0.973***	-1.490***	-0.719***	-1.661***	-0.593***	-1.453***	-0.780***	-2.055***
	(0.100)	(0.273)	(0.0929)	(0.282)	(0.0632)	(0.268)	(0.0453)	(0.319)
Observations	14,364	14,308	12,559	12,513	14,376	14,322	7,576	7,552
Pseudo R-Squared		0.080		0.069		0.080		0.067

Notes: Dependent variable is A=2 if the individual attrited in wave 3 and zero otherwise. 2). Robust standard errors that allow for correlation in un-observables for each individual drawn from the same primary sampling unit are shown in parentheses. 3) In each model we first include only the health outcome before we include the control variables in the next model. In model 1 we include SAH only, 3 we include BMI only, 5 we include depression and model 7 we include hypertension only as the explanatory variable. 4) FET is further education and training 5). *** p<0.01, ** p<0.05, * p<0.1

Table 2.9: Attrition probit results with alternative measurement for depression

VARIABLES	(1)	(2)	(3)	(4)
Depression (CESD8)	-0.0107	0.00286	-0.0160***	-0.00251
	(0.00670)	(0.00562)	(0.00588)	(0.00543)
25-34 (rel.to 15-24)		0.0689		0.191***
		(0.0668)		(0.0706)
35-44		-0.0519		-0.0102
		(0.0671)		(0.0702)
45-54		-0.338***		-0.269***
		(0.0918)		(0.0938)
55-64		-0.379***		-0.407***
		(0.108)		(0.107)
65+		-0.370***		-0.360***
		(0.133)		(0.122)
Log real per capita monthly income		0.0537*		0.0749**
		(0.0283)		(0.0304)
General educ (rel. to No educ.)		-0.277***		-0.244***
		(0.0833)		(0.0744)
Further education and training		-0.214**		-0.223**
		(0.100)		(0.0923)
Matric		-0.212*		-0.265***
		(0.109)		(0.0961)
Higher education		-0.152		-0.105
		(0.158)		(0.170)
Male		0.173***		0.223***
		(0.0371)		(0.0416)
Coloured (rel. to African)		0.0765		0.109
		(0.108)		(0.134)
Asian/Indian		0.406**		0.347
		(0.193)		(0.211)
White		0.622***		0.750***
		(0.115)		(0.145)
With Partner (rel. to married)		0.230**		0.271**
		(0.115)		(0.118)
Widow/widower		0.00957		-0.0143
		(0.0921)		(0.0995)
Divorced		0.318**		0.296**
		(0.124)		(0.139)
Never married		0.0744		0.00757
		(0.0782)		(0.0761)
Eastern Cape (rel. to Western Cape)		0.0625		0.110
		(0.143)		(0.153)
Northern Cape		-0.291**		-0.0491
		(0.126)		(0.143)
Free State		-0.358**		-0.356*
		(0.147)		(0.199)
Kwa-Zulu Natal		-0.387***		-0.199
		(0.144)		(0.146)
North West		-0.0725		0.196
		(0.204)		(0.239)
Gauteng		-0.141		0.0923
		(0.131)		(0.144)
Mpumalanga		-0.547***		-0.0323
		(0.147)		(0.158)
Limpopo		-0.213		-0.0909
		(0.147)		(0.162)
Urban (rel. to Traditional area)		0.267***		0.223**
		(0.0943)		(0.0925)
Farms		0.313**		0.334***
		(0.136)		(0.108)
Constant	-0.637***	-1.037***	-0.646***	-1.409***
	(0.0579)	(0.240)	(0.0515)	(0.263)
Observations	14,851	14,796	14,381	14,327
Pseudo R-Squared		0.074		0.080

Notes: 1) Table shows attrition probit results for four models. In the first two dependent variable A=1 if the individual attrited in wave 2 and zero otherwise 2) In model 3 and 4 dependent variable is A2=1 if the individual attrited in wave 3 and zero otherwise. 4) The depression (CESD) score used in all models is computed from the 8 negative questions of the CESD. 5) Robust standard errors that allow for correlation in un-observables for each individual drawn from the same primary sampling unit are shown in parentheses, 6). *** p<0.01, ** p<0.05, * p<0.1. 7). These results are similar to those obtained when CESD score for all the ten questions was used, i,e depression is a significant determinant of attrition in both the restricted and unrestricted model.

excellent health. The higher the educational attainment the higher the probability of reporting excellent health. These results confirm what we obtained using attrition probit, that is, the parameter estimates obtained when using SAH in a structural equation are unbiased and consistent because they are not affected by attrition bias.

In model 2, the dependent variable is the BMI score which is a continuous variable. Attrition in wave 2 is not a significant determinant of BMI score. The age categories indicate that BMI score increases with age but at a decreasing rate. The BMI score is also associated with an increase in income. Individuals with higher levels of educational attainment are associated with higher BMI scores.

In models 3 and 4, we use two alternative measures of depression. In model 3 the depression score used is obtained from the CES-D 10 questions while in model 4 we use the depression score obtained from CES-D 8 questions, as detailed in section 2.2. Attrition in wave 2 does not determine the presence of depressive symptoms whatever the measure of depression used. Holding other factors constant, the 15-24 year olds have less depressive symptoms and the coefficients indicate that these symptoms increase with age up to 45-54 years but start declining with older age groups. Income is associated with reduced depressive symptoms. Having general education relative to no education is associated with reduced depressive symptoms. Higher levels of education are associated with lower depressive symptoms.

In model 5, high blood pressure is used as the dependent variable. This variable is used here as a binary variable, equal to 1 when an individual has high blood pressure and zero when the individual has normal blood pressure. We therefore run a discrete choice model (probit) to estimate the probability of having high blood pressure. Like the other health outcomes, attrition in wave 2 is not statistically significant. The older groups have higher blood pressure. Income is not associated with the probability of having high blood pressure. Those with higher education have lower probability of having high blood pressure, relative to those with no education. The other education categories are not significantly different from those with no education.

In table 2.11 we present results for the BGLW test in wave 3. These results mimic results already discussed for the BGLW test in wave 2. For example, across all the health dimensions, the attrition dummy in wave 3 is not statistically significant. This implies that attrition in wave 3 is random with regard to these health dimensions and the point estimates from models including these health outcomes are unbiased and consistent.

2.7.3 Sensitivity analysis

For robustness, we compare health outcome regression results we obtain without correcting for sample selection with those we obtain after correcting for sample selection in wave 2 and

Table 2.10: BGLW Test for health attrition in wave 2

VARIABLES	(1) SAH	(2) BMI	(3) DEPR10	(4) DEPR8	(5) PRESSURE
Attrition in wave 2	0.0521	-0.218	0.0972	0.0883	0.0347
	(0.0348)	(0.162)	(0.199)	(0.189)	(0.0974)
25-34 (rel.to 15-24)	-0.282***	2.034***	1.252***	1.090***	0.630***
	(0.0446)	(0.188)	(0.148)	(0.154)	(0.113)
35-44	-0.500***	3.280***	1.543***	1.375***	1.015***
	(0.0529)	(0.233)	(0.195)	(0.193)	(0.109)
45-54	-0.795***	4.131***	1.677***	1.399***	1.555***
	(0.0577)	(0.315)	(0.222)	(0.212)	(0.123)
55-64	-1.037***	4.211***	1.499***	1.360***	1.800***
	(0.0693)	(0.289)	(0.267)	(0.252)	(0.141)
65+	-1.214***	3.906***	1.437***	1.210***	2.029***
	(0.0823)	(0.389)	(0.263)	(0.240)	(0.148)
Log real per capita monthly income	0.0521***	0.230***	-0.512***	-0.388***	-0.0228
	(0.0173)	(0.0860)	(0.0778)	(0.0765)	(0.0282)
General educ (rel. to No educ.)	0.238***	0.937***	-0.540***	-0.407*	0.0109
	(0.0476)	(0.293)	(0.205)	(0.209)	(0.104)
Further education and training	0.342***	1.765***	-0.954***	-0.667***	-0.134
	(0.0587)	(0.307)	(0.218)	(0.223)	(0.114)
Matric	0.605***	1.835***	-1.589***	-1.146***	-0.119
	(0.0633)	(0.307)	(0.262)	(0.252)	(0.133)
Higher education	0.722***	1.625***	-1.412***	-1.260***	-0.401**
	(0.117)	(0.562)	(0.397)	(0.359)	(0.199)
Controls	Y	Y	Y	Y	Y
Observations	14,776	12,915	14,791	14,796	7,759
R-squared		0.240	0.125	0.081	

Notes: The table shows different dimensions of health outcomes regressed on attrition indicator in wave 2 and the other covariates in the model. The controls in this table include marital status, province and location of residence. In all models the coefficient on attrition indicator is not statistically significant. Standard errors reported in parentheses are robust, they are clustered by pid. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.11: BGLW Test for health attrition in wave 3

VARIABLES	(1) SAH	(2) BMI	(3) DEPR10	(4) DEPR8	(5) PRESSURE
Attrition in wave 3	0.0458	0.144	0.0960	-0.0831	0.0347
	(0.0437)	(0.203)	(0.172)	(0.175)	(0.0974)
25-34 (rel.to 15-24)	-0.279***	2.017***	1.241***	1.088***	0.630***
	(0.0450)	(0.188)	(0.148)	(0.156)	(0.113)
35-44	-0.492***	3.297***	1.552***	1.384***	1.015***
	(0.0523)	(0.238)	(0.195)	(0.192)	(0.109)
45-54	-0.771***	4.239***	1.644***	1.342***	1.555***
	(0.0568)	(0.302)	(0.230)	(0.218)	(0.123)
55-64	-1.021***	4.322***	1.476***	1.309***	1.800***
	(0.0706)	(0.301)	(0.268)	(0.250)	(0.141)
65+	-1.210***	3.811***	1.421***	1.169***	2.029***
	(0.0878)	(0.386)	(0.285)	(0.257)	(0.148)
Log real per capita monthly income	0.0521***	0.222***	-0.515***	-0.394***	-0.0228
	(0.0171)	(0.0840)	(0.0776)	(0.0757)	(0.0282)
General educ (rel. to No educ.)	0.250***	0.983***	-0.590***	-0.470**	0.0109
	(0.0486)	(0.286)	(0.216)	(0.224)	(0.104)
Further education and training	0.361***	1.771***	-1.012***	-0.749***	-0.134
	(0.0576)	(0.305)	(0.228)	(0.238)	(0.114)
Matric	0.619***	1.903***	-1.647***	-1.221***	-0.119
	(0.0632)	(0.305)	(0.274)	(0.269)	(0.133)
Higher education	0.718***	1.576***	-1.416***	-1.251***	-0.401**
	(0.118)	(0.542)	(0.394)	(0.356)	(0.199)
Controls	Y	Y	Y	Y	Y
Observations	14,308	12,513	14,322	14,327	7,552
R-squared		0.244	0.126	0.083	

Notes: The table shows different dimensions of health outcomes regressed on attrition indicator in wave 3 and the other covariates in the model. The controls in this table include marital status, province and location of residence. In all models the coefficient on attrition indicator is not statistically significant. Standard errors reported in parentheses are robust, they are clustered by pid. *** p<0.01, ** p<0.05, * p<0.1.

wave 3. The weights to correct for selection bias on the dependent variables are computed as the inverse of the predicted attrition probabilities, as explained in section 2.6.1. These weights are approximately equal to one when the panel attrition is random in which case they do not have any effect on the results. To apply these weights we multiply them with post stratification weights that are provided with the data.

After correcting for sample selection on the dependent variables using the inverse of the predicted attrition probabilities, the parameter estimates across all health models do not change. In fact the parameter estimates obtained after re-weighting are identical to those obtained without re-weighting. These results are shown in table A.3 and A.4 for wave 2 and wave 3 respectively, which are in the appendix to this chapter. These results confirm that the sample loss from attrition is random with respect to the health outcomes and therefore it does not bias the parameter estimates for the health dimensions analyzed in this study.

2.8 Conclusion

Panel studies have benefits over cross-sectional studies in terms of identifying dynamic relationships among variables. In addition panel survey dataset allow us to control for unobserved heterogeneity which we cannot do with cross sectional survey data. But attrition poses a threat to the validity and generalisability of findings based on multiple waves of panel data. This is because of potential bias in structural equation parameter estimates due to non random attrition. In this chapter we began by describing the variables used in the study. We explained how each variable is measured and showed their distributions within the population. We then examined whether there is potential threat of attrition bias with regard to selected health outcomes in NIDS. Our analysis focused on the health outcomes as dependent variables because, as we mentioned earlier, attrition due to independent variables does not pose a threat to parameter estimates other than reducing the sample size (Alderman et al., 2001).

We find average health outcomes of attritors and non attritors to be significantly different which prompts a formal test of the effects of attrition on parameter estimates in these models. This is reinforced by the fact that there is no necessary relationship between the size of the sample loss from attrition and the existence or magnitude of attrition bias (Fitzgerald et al., 1998; Maluccio, 2004).

We followed Fitzgerald et al. (1998) and Moffit et al. (1999), in testing for selective attrition on observables because testing for the effects of attrition on unobservables requires external validation data. Using attrition probit models, we do not find evidence of attrition bias in the health outcomes we study, namely SAH, BMI, mental health (depression) and high blood pressure models. We find observable variables that are correlated with attrition,

but these variables explain only a negligible portion of the attrition in NIDS. We find no compelling evidence that attrition has any effect on the parameter estimates of the health outcome equations we study. We use an alternative test, the BGLW test, and find that these models are not affected by attrition. This confirms that attrition in NIDS was random with respect to the health outcomes.

The implications of the results of this study are twofold. Firstly, we can make inference of population density ($f(h|x)$) by estimating the sample density of those who do not attrit ($g(h|x, A = 0)$). This is made possible because attrition does not affect the representativeness of health outcomes in NIDS. Secondly, given the challenges of interviewing the same individuals in every wave, the findings confirm the high quality of health related data in NIDS.

In this chapter we established that attrition does not affect the representativeness of health related data in NIDS. This paves the way for us to use these health related data to establish determinants of health outcomes in each wave. This is important for us because, as we detail in chapter four, we use the coefficients of the determinants of health outcomes in each wave to estimate income related health inequalities. Before we do that we want to identify determinants of health outcomes in South Africa. To do so we use panel data models to establish these determinants. This is discussed in the next chapter.

Chapter 3

Socioeconomic Determinants of Health Outcomes in South Africa: Evidence from NIDS Data

3.1 Introduction

This study is motivated by the compelling evidence in the literature on the important role of socioeconomic factors in determining health outcomes (Braveman and Gottlieb, 2014; Strauss et al., 2010) and Smith et al., 2010). The evidence asserts that medical care is not the only influence on health. The influence of medical care may be more limited than commonly thought, particularly in determining who becomes sick in the first place (Braveman and Gottlieb, 2014).

The evidence indicates that medical care contributes only up to 40 percent to health outcomes, with the balance of outcomes attributed to non medical factors (Ostry and Frank, 2010). Three factors are advanced in the literature to explain the limits of medical care. First, that the sharp decline in mortality rates in the mid 19th century through to the 1960s, and the subsequent dramatic increase in life expectancy, are normally attributed to non medical factors. Second, the widening regional mortality disparities in countries that have a uniform national health care coverage, for example the United Kingdom. Finally, the disparities between spending on medical care and health outcomes across countries imply that there are limits to medical care (Braveman and Gottlieb, 2014).

Although socioeconomic factors account for about 60 percent of health outcomes, most policies aimed at improvements in health are targeted towards medical care (Bynum, 2008; Link, 2013; Harris, 2004). In addition, even when socioeconomic factors are targeted, their effects are not uniform across all settings. Socioeconomic factors such as income and educa-

tion have inconsistent effects in different settings (Strauss et al., 2010). This inconsistency is clearly not helpful to policy makers interested in improving health outcomes. The inconsistency and the relative importance of socioeconomic factors point to the value of understanding the socioeconomic determinants of health outcomes in a developing country context.

In this chapter we examine the socioeconomic determinants of health outcomes for adult individuals in South Africa. We add to the existing literature on socioeconomic determinants of health outcomes in a developing country context. The results outlined in this chapter form the basis for estimating and decomposing income related health inequality in the next chapter. In particular, we estimate the average contribution of each determinant which we use in the decomposition of income related health inequalities.

First we estimate the determinants of health outcomes using three health dimensions (self assessed health (SAH), body mass index and depression). Second, we use panel data that allows us to control for unobserved heterogeneity. The panel nature of the data is utilized to confirm the suitability of SAH as a measure of the health outcome variable (Ardington and Gasealahwe, 2012). In particular, panel data is used to show that SAH categories are able to predict mortality in South Africa. Thirdly, our sample is restricted to adult individuals. This is because their health outcomes are evaluated differently from those of children. This addresses concerns regarding the health outcomes that are computed from self evaluations such as SAH and depression. Fourthly, our sample has random attrition with regard to health outcomes, as we have demonstrated in the previous chapter, and this addresses potential biases associated with individual level attrition. We aim to provide detailed, and updated evidence on the correlation between socioeconomic factors and health outcomes using South Africa as case study. We examine the nature of the relationships between health outcomes and measures of socioeconomic and demographic factors. We also establish whether socioeconomic factors affect different dimensions of health similarly. Specific emphasis is placed on examining the relationship between health outcomes and socioeconomic and demographic variables.[1] This is important to isolate the factors that can be addressed by policy and those that cannot.

South Africa presents an important opportunity for this study because economic well-being[2] has been on a general upward trend since the transition to democracy in 1994, as indicated by increasing levels of average real incomes (Bhorat et al., 2012; Leibbrandt et al., 2015). At the same time inequality in income and economic resources has also increased and remains high relative to other countries in Sub Saharan Africa (Leibbrandt et al., 2015).

[1]We will not analyze activities of daily living (ADL) because this information was not collected across all waves (not in wave 3).

[2]Although we use income here to measure economic well-being, the human development index that also incorporates longevity and education can be used. We use income because we are interested to see if the increases in income improve health outcomes.

The extent to which growth in average real incomes drives average health outcomes depends in part on the relationship between economic well being and health outcomes. Similarly, the degree to which pervasive inequality in income and other socioeconomic factors such as education are driving similarly large health inequalities requires a better understanding of the relationship between socioeconomic factors and health indicators. It is against this background that we undertake this study to establish these relationships.

While there exists a very large body of literature that examines the relationship between Socioeconomic status (SES) and health measures, they are concentrated in developed countries with little in the developing country context (Correa-Valez et al., 2011; Contoyannis and Forster, 1999). Little research on this topic has been undertaken for South Africa relative to the developed world (Case, 2004; Bradshaw, 2008; Ardington and Gasealahwe, 2012; McCartney et al., 2013a; McIntyre and J.E., 2012). Ardington and Gasealahwe (2012) uses the first two waves of NIDS with a limited number of health dimensions to examine health data and highlights the potential of the NIDS panel for analyzing the relationship between health and socioeconomic status. We build on their findings and present a comprehensive analysis of the relationship between health outcomes and socioeconomic determinants of health in South Africa.

We undertake bivariate as well as panel data econometric analysis to show the relationship between health outcomes and socioeconomic factors using the first three waves of the NIDS data. NIDS has two major advantages over previous surveys in South Africa. Unlike the Demographic and Health Surveys (DHS), the General Household Surveys (GHS) and the Labour Force Surveys (LFS), it is rich in socioeconomic variables and data on health outcomes. Thus using this data we do not need to construct any socioeconomic indices to serve as indicators of socioeconomic status. Secondly, it is nationally representative and is longitudinal, which makes it possible to also bring out the dynamic relationships between socioeconomic variables and health outcomes while controlling for unobserved heterogeneity.

The rest of the chapter is structured as follows: In section 3.2 we give an overview of the health care system in South Africa, in section 3.3 we present the theoretical and empirical literature. Data and our empirical model are discussed in section 3.4. In sections 3.5 to 3.7 we present a bivariate analysis of the relationship between different dimensions of health outcomes and key socioeconomic and demographic factors. Section 3.8 presents the panel estimators of this relationship, and we conclude in section 3.9.

3.2 An Overview of Health Care in South Africa

South Africa had 51.8 million people in 2011 (SA Statistics) with a gross domestic product (GDP) of R1,834,435 million. In 2012, GDP per capita in current dollar terms was USD7352, one of the highest per capita incomes in the region (along with Botswana, Seychelles, Gabon,

and Equatorial Guinea). This also puts South Africa in the league of upper middle income countries, such as Brazil and China (World Bank, 2012). We now move to compare health indicators in these countries with those of South Africa given that they have comparable incomes per capita to that of South Africa.

At a macro level, health indicators in South Africa have been improving over time, however they remain poor compared to other middle income countries (World Bank, 2012). This is despite a higher percentage of the country's GDP being spent on health than many other middle income countries. Table 3.1 compares 2012 figures for health indicators in South Africa with other middle income countries. Following the World Bank classification of defining middle income countries (MICs) through their per capita income, the following countries fall in the category of upper and lower middle-income countries, Brazil, Russia, China, and South Africa are upper middle income while India is lower middle income country (World Bank, 2012). In addition some of these countries belong to a common trading bloc called the BRICS (an acronym for the countries in the group, Brazil, Russia, India, China and South Africa) and this facilitates interaction and opportunities for learning between these countries.

Health expenditure as a percentage of GDP was 8.8 percent compared to China at 5.4 percent, Philippines at 4.6 percent, and Indonesia at 3 percent. Brazil and India spent 9.3 and 4 percent respectively of their GDP on health (World Bank, 2012). These percentages translate to health expenditure per capita in 2012 US dollars of $645 for South Africa, $322 for China, $119 for Philippines and $108 for Indonesia. It is clear that South Africa's per capital expenditure is only exceeded by that of Brazil and Russia. However, health indicators such as infant and under 5 mortality are worse for South Africa than these countries except India.

There is also a large gap between public and private health expenditure in South Africa. In 2012 public health expenditure as a percentage of total expenditure stood at 47.9 percent compared to private health expenditure which was 52.1 percent. Private health insurance accounts for 81.1 percent of private health expenditure but only covers 17 percent of South Africans, with the uninsured mainly utilizing the public health care system (Burger et al., 2012).

The bulk of the public health expenditure is financed through national, provincial and local government budgets. Other private and public health care funding sources are insurance companies, Non Governmental organizations (NGOs), out of pocket expenditure (OOP) and medical aid. At the macro level, South Africa seems to compare well with other middle income countries. We now move to look at the health situation in the country and how the health system works.

South Africa faces a quadruple burden of disease (Mayosi et al., 2009; Bradshaw, 2008).

Table 3.1: Health Indicators for South Africa Compared with other Countries in 2012

Health Indicator	BRICS					OTHERS	
	Brazil	Russia	India	China	SA	Indonesia	Philippines
Health expenditure (% of GDP)	9.3	6.3	4.0	5.4	8.8	3.0	4.6
Per capita health expenditure (US$)	1056	887	61	322	645	108	119
Mortality (Infant /1000 live births)	13	9	43	12	33	26	24
Mortality (Under 5 /1000 live births)	14	11	55	14	45	31	30
TB incidence* (per 100,000 people)	46	91	176	73	1003	185	265

Source: Own compilation from world bank data for 2012. The table shows comparative health indicators for five middle income countries. * indicates that Tuberculosis (TB) incidence is the estimated number of new pulmonary, smear positive, and extra-pulmonary tuberculosis cases per 100,000 people.

This means that there are four major categories of disease burdens namely; diarrhea and malnutrition (Type I causes), chronic diseases such as diabetes and stroke (Type II causes), injury burden (Type III causes) and the HIV/AIDS epidemic. These diseases are unevenly distributed across the population (Bradshaw, 2008). Differentials in health outcomes have been observed between population groups, wealth groups, urban and rural dwellers and at different education levels. Differentials have also been observed in mortality rates (Bradshaw, 2008).

In terms of human resources for health, the national government is responsible for training and recruitment of doctors who work in both public and private hospitals. Private hospitals do not employ doctors directly but apply to be allocated doctors by the government. The private hospitals entice doctors to work in their hospitals by providing them with current technology, equipment and facilities. However, surgeries or clinics can be opened as private businesses and run by private doctors.

Since the transition to democracy in 1994, the health care system in South Africa has continuously been reformed (Mayosi et al., 2009; Harrison, 2009). Much effort has been invested in improving health outcomes by making public health care more accessible to the poor. Accessibility is interpreted in terms of affordability through provision of free care to vulnerable groups such as pregnant mothers and young children and reduction of travel distance through construction of primary health care facilities. The construction of health care facilities was undertaken under the Clinic Upgrading and Building Programme that was initiated as a presidential lead project in 1994. This expansion programme saw about 1300 clinics built or upgraded by 1998 (about 460 new clinics and 810 clinic residential units were built) (Harrison, 2009).

The health care system in South Africa is partially devolved and is managed at both national, provincial and local levels. At the national level, the department of health holds overall responsibility for health care, with specific responsibility for the public sector. Provincial health departments provide and manage comprehensive health services, via a district-based, public health-care model. Local hospital management has delegated authority over opera-

tional issues, such as the budget and human resources, to facilitate quicker responses to local needs (Mayosi et al., 2009).

South Africa has more than 110 registered medical schemes, with around 3.4 million principal members (and 7.8 million beneficiaries). This implies that roughly 11 million South Africans have medical insurance cover. The remaining majority of South Africans rely on out of pocket payments for their medical expenses and this is usually in the public health facilities. This, coupled with the fact that there is a shortage of medical care practitioners implies that the majority poor are not receiving adequate medical care services.

In terms of human resources for health, in 2012, there were 165 371 qualified health practitioners in both public and private sectors who were registered with the Health Professions Council of South Africa, the health practitioner watchdog body. This includes 38 236 doctors and 5 560 dentists. The doctor-to-population ratio is estimated to be 0.77 per 1 000. But because the vast majority of the general practitioners (73%) work in the private sector, there is just one practicing doctor for every 4 219 people (Bradshaw, 2008). This implies that majority of the doctors in South Africa who work in private sector serve a smaller proportion of the population i.e. those with medical insurance.

From this overview, we have compared health indicators of South Africa to countries with similar per capita income, we have also looked at the structure of the health care system, the burden of disease and the financial and human resources for health. It is clear that there is huge variation both between countries and within countries. Some countries spend more on health per capita than others. We have seen that there is no guarantee that higher spending leads to better health outcomes. We have also seen that even within South Africa, the way health care is structured and financed means that health care is not equally accessible among the rich and the poor and this is likely to have implications on health outcomes.

3.3 Literature

3.3.1 Theoretical framework

The work described in this chapter is grounded in human capital[3] theory and in particular the Grossman human capital theory of health. According to human capital theory, increases in a person's stock of knowledge or human capital raises their productivity in the market sector of the economy, where they earn money, and in the non-market sector or household sector where they produce commodities that enter their utility function (Grossman, 1999; Becker, 1975). To realize potential gains in productivity, therefore, individuals have an incentive to invest in formal schooling and on-the-job training (Grossman, 1999; Becker,

[3]Human capital is viewed as a stock of knowledge or characteristics a worker has that contributes to his or her productivity.

1975).

Grossman advanced this theory by arguing that health is a form of human capital which differs from other forms of human capital because it enters the utility function directly and also determines earnings or wealth in a life cycle context (Grossman, 2000, 1972). Grossman constructs a model of demand for health capital itself using a household production function model of consumer behaviour. He uses this model to distinguish between health as an output (health outcome) and medical care as one of the inputs into its production.

Since health is a form of human capital, a person inherits an initial stock of health that depreciates with age (at an increasing rate after some stage in the life cycle), and can be increased by investment in health. Death occurs when the stock of health falls below a certain level. Gross investments are produced by individual production functions that relate output of health (health outcome) to health inputs such as medical care utilization, diet, exercise, cigarette smoking, and alcohol consumption (Grossman, 1999). The production function of health is however, also affected by the efficiency or productivity of a given consumer, as reflected by personal characteristics such as gender, race, age and marital status. Efficiency is defined as the amount of health obtained from a given amount of health inputs. Formal schooling or educational attainment plays a role in this context (Grossman, 1972).

In empirical work, Grossman's human capital framework is used to analyze at least three types of relationships. These include the health production function, the individual demand for medical and behavioural health inputs, and the determinants of individual health status or health outcomes (Schultz, 2004; Zhao, 2008). In this chapter we use Grossman's health human capital model to estimate a reduced form model that explains the relationship between health outcomes and socioeconomic variables in South Africa. Using this theoretical framework, health outcome is expressed as a function of demand side and supply side factors.We therefore estimate a reduced form that has both components of demand and supply side variables in line with other research. Demand side factors include individual characteristics that affect efficiency and productivity of health, such as education, income, age, gender and race. Supply side factors include geographical variables such as province of residence and rural-urban setting. In our case we purposefully omit lifestyle variables such as diet, smoking, and alcohol consumption. This is in line with the current literature that argues that these variables are presumably endogenous[4] to the other variables already included in the model (Van de Poel et al., 2012). Consequently, we formulate our theoretical model as follows:

$$h^*_{it} = h(Demand\ Side,\ Supply\ Side\ factors) \tag{3.1}$$

Where h^*_{it} is health outcome indicator of individual i at time t, demand and supply side

[4]Income is also endogenous therefore any interpretation is associational and not causal.

variables are as explained above (Grossman, 1999). Geographical location, including province of residence and rural-urban setting, control for different health systems that prevail in each province. Other independent demand side variables include marital and employment status (Wilkinson and Pickett, 2011; Marmot and Bell, 2011). We keep determinants similar across all models to achieve our objective of establishing whether they affect health outcomes in the same way.

The supply side determinants of health outcomes include the health care system characteristics. As explained in the overview section, they vary from province to province and from one facility to the other. The variations in health care are therefore from both the supply and demand side factors. We control for provincial effects and urban-rural effects in all our regressions to capture any differences in the delivery of medical care that is unique to these locations. Our general framework is therefore one that includes components of both demand and supply.

3.3.2 Empirical Literature

As mentioned , there are inconsistent results in the literature on the nature of the relationship between health outcomes and socioeconomic variables, especially income and education. While some studies find positive correlation others either find negative or no correlation. The relationship between socioeconomic factors and health can be ascertained directly or indirectly. The direct approach involves investigating the evidence of the relationship between socioeconomic determinants and health, and indirect approach is through the determinants of health inequality.

Hernández-Quevedo and Jiménez-Rubio (2009) use the direct approach to explore the relationship between health and socioeconomic status, using European Union statistics on income and living conditions (EU-SILC) for 13 countries. They find that income is not a statistically significant determinant in some countries while it is in others. They find similar results using a different dataset, the European Community Household Panel (ECHP) (Hernández-Quevedo and Jiménez-Rubio, 2009). They use a pooled probit model and defend its use by arguing that while it does not take into account the panel nature of the data, the estimator is still consistent for the parameter of interest. That said, all the countries in the study are drawn from the developed world and it is difficult to predict the results of a similar study in developing countries.

Strauss et al. (2010) are concerned with measuring the relationship between different dimensions of health outcomes and measures of socioeconomic status (SES) among the elderly in China. They use self reported measures as well as bio markers as indicators of health. They examine correlations between health outcomes and education and the log of per capita expenditure to determine if correlations reported in other countries are replicated in China.

The analysis includes many salient measures of health status, including chronic conditions, general health status, subjective measures of health, body mass index (BMI), hypertension, cognition, and functional measures of disability (ADLs). The SES variables considered include schooling, income and household consumption. They control for age, province, and the rural-urban nature of residence within the provinces.

They use ordinary least squares (OLS) regression for continuous dependent variables and linear probability models (LPM) for binary dependent variables to establish the correlations. Some of their findings are similar to the findings in other parts of the world but some are in contrast. For example they find that BMI rises with education for men. For women, BMI initially rises with schooling but falls at higher levels of schooling. Hypertension is not related to SES in China which is in contrast to results found in studies for US and England. In the Chinese study average ADLs rise with age but decline with higher schooling, but the study does not find education and per capita expenditure significant in determining ADLs. There are more depressive symptoms for women than men and these symptoms are more correlated with schooling and less with income. The male-female differences narrow with education. Regression results show that education is not significantly related to being in poor health but per capita expenditure is negatively related. The researchers argue that while such correlations are not causal, they tell something important about the degree of health differentials by education and per capita expenditure. Many more studies in developed countries report contrasting results. There is therefore a need for country or region specific studies (See also Smith (2009) for details).

Some work has been done in South Africa to establish determinants of health, see for example Case (2004), Case and Deaton (2005), Ardington (2009), and Wittenberg (2013). Case (2004) examines the causal relationship between income and health status using the South African old age pension as an instrument for income. She further investigates pathways through which higher incomes lead to better health. Her analysis uses the Langeberg survey which collected data on three magisterial districts in the Western Cape. She uses an ordered probit of SAH on income while controlling for age, education, and gender to establish the association between income and health status by race. She finds a strong association between income and health status. She also finds that older adults report worse health on average. Whites' health erodes more slowly with age, relative to other races. Education is associated with better health, particularly for whites, but smaller effects are reported for blacks and Coloureds. The problem with this study is that it covers a small area of South Africa making it difficult to convincingly generalize the results.

In another study Case and Deaton (2005) compares Adaipur in India and Khayelitsha in South Africa on measures of economic and health status to test the hypothesis that "wealthier is healthier" using individual level data. This hypothesis asserts that income is the main

determinant of health. This comparison finds that although GDP per capita in South Africa for the year 2000 was higher than that of India, life expectancy was 14 years shorter in South Africa. She further compares well-being using measures of health and living standards and finds that the economically better off in South Africa are healthier in some respects, but not in others. Her conclusion is that the link between health and wealth is far from universally strong, a position that is supported by the contrasting results across many studies. Her study also finds that few South Africans are underweight, but 75% of women are overweight (BMI between 25-30) or obese (BMI over 30). The difference in BMI between men and women in the study is substantial. The generalization of results from Khayelitsha to represent South Africa is problematic as Khayelistha is not representative of the country.

The results are however consistent with results from a nationally representative sample of African South Africans from the DHS of 1998, see for example (Wittenberg, 2013; Puoane et al., 2002; Ardington and Gasealahwe, 2012)

To confirm the strong association between BMI and income in South Africa, Wittenberg (2013) shows that BMI can be used as a marker for well-being because the relationship between BMI and wealth is, he argues, monotonic. This is however not the case for the White population in South Africa. The BMI-wealth relationship is negative for White females. Using 1998 DHS data and the first wave of NIDS (2008), Wittenberg finds that average BMI for South Africa increased from 1998 to 2008. One drawback of this study is that a wealth index is used in the DHS as an indicator of well-being. This makes a comparison of results between 1998 and 2008 problematic because the well being indicator used is not the same in the two surveys.

Ardington (2009) uses the first wave of NIDS data to investigate the joint determinants of economic and health status in South Africa. The health outcomes considered are obesity, hypertension, and malnutrition in children. The study also analyses depression, activities of daily living (ADLs) and self reported illness. It finds BMI for women to be higher than for men at all ages and that it also increases with age. The study uses the 1998 Demographic and Health Survey (DHS) to examine changes in rates of obesity over a period of ten years in South Africa. Findings are that obesity rates for women rose by 3 percentage points (from 30% to 33%) over a decade, while those for men rose by 2 percentage points (from 9% to 11%) over the same period. The BMI is regressed on race, a dummy for location of residence (urban/rural), years of completed education, a count of assets, a score for adults being hungry often or always, and a quadratic for age. The research concludes that in general there is a positive association between socioeconomic variables and BMI for both men and women. This is repeated with obesity as the dependent variable and the results do not change. In the analysis of hypertension, they find that the rates have increased from 1998 (DHS) to 2008 (NIDS). They use a probit regression of an indicator for having hypertension

to examine correlates of hypertension among men and women separately. In addition they show that hypertension and BMI are positively related. This is an important finding in understanding the influence of socioeconomic factors on these two health outcomes. While this study resembles ours in many respects, they use two cross-sectional studies to infer changes in health indicators, although the well-being variables in the two surveys were not similar.

In our case we follow closely the work of Strauss et al. (2010) and Hernández-Quevedo and Jiménez-Rubio (2009) in terms of objectives and method to establish the correlation between health outcomes and socioeconomic variables. We use three waves of NIDS data to investigate these relationships with bivariate and panel data models.

3.4 Method of Analysis

3.4.1 Data and variable description

The data we use in this study is sourced from the three waves of NIDS, as explained in Chapter Two (SALDRU, 2009, 2014a,b). In this section we describe in more detail how we measure the explanatory variables we use in the study. We have not addressed these as adequately as we have addressed our dependent variables in previous chapters. This section also presents panel properties of the data. This is important because we compute panel data estimators later in the chapter.

Our focus is limited to one aspect of well-being, health outcome and its relationship with socioeconomic determinants, as detailed in the theoretical framework. We therefore pay attention to the explanatory variables explicitly mentioned in the theoretical framework, such as income, age, and education. We also highlight variables that appear in most empirical work, and characterize individuals. These variables include race, gender, location of residence, and marital status.

We create five categories of educational attainment, using information on the highest school grade completed by respondents. These categories are; no education, general education, further education and training, Matric, and higher education. These categories have already been explained in the previous chapter.

We control for demographic variables such as age, gender, and race. Age is provided in the data in years. We group this into age categories (15-24, 25-34, 35-44, 45-54, 55-64, 65+) to use in the bivariate analysis (for example table B.2 in the appendix). The categories are in line with other studies that have analyzed health using the NIDS data (Ardington, 2009). For the panel regressions however, we use age in continuous form to control for life cycle changes in health.

An indicator of being male or female, with female as the reference category, is included to account for gender differences in health outcomes. Race is also controlled for in our regressions. As we mentioned earlier, this variable is important for South Africa because of the racial differences that manifest in many measures of well-being. Most countries do not have multiple race groups in sufficient numbers to analyze them separately. We omit highly endogenous (lifestyle) variables such as smoking and alcohol consumption that also characterize the individual because their non consumption could be due to medical advice.

As explained in chapter two, we control for geographical types (traditional areas, urban areas and farms) as well as province of residence. Table 2.1 in chapter two describes the health outcome variables and explanatory variables used in our analysis.

We report panel data summary statistics for adult continuing sample members (CSMs) in the NIDS because we utilize panel data estimators in this chapter. As we saw in chapter two, some individuals are not interviewed in all the waves. The panel is therefore unbalanced. We uniquely identify 34,391 adults across the three waves in NIDS. 13 of these switched gender across waves therefore we exclude them from the analysis. Of the remaining 34,378 individuals 60% are interviewed in all three waves, 15% are interviewed in the first two waves, 8% are interviewed in wave 2 and wave 3 only, 7% are interviewed in wave 1 only, 5% are interviewed only in waves 1 and wave 3, 4% are interviewed only in wave 3 and 1% are interviewed in wave 2 only.

The summary statistics for outcome variables and continuous explanatory variables are presented in table 3.2. We dichotomize the SAH variable such that poor and fair SAH is indicated by 0 and good, very good and excellent health indicated by 1, to be in line with other studies that use panel data fixed effects (Jones and Wildman, 2008). The key reason for dichotomizing SAH is to enable us to use a linear probability model (LPM) which solves the problem of incidental parameters bias that is inherent in probit model fixed effects[5] (Baltagi, 2008). The minimum and maximum for this variable are therefore 0 and 1 because it is binary. The overall mean[6] indicates the mean of the variable that takes into account time and individual variation. For example the overall mean for the binary SAH is 0.852. This is the proportion of the population with better (good, very good and excellent) health over time and between individuals.

The overall mean of the depression score measured by the CES-D scale is 7.2 which is below the threshold score of 10 above which individuals are categorized as having depressive symptoms. This shows that the majority do not have depressive symptoms. The mean takes

[5]Probit model fixed effects introduce bias in the coefficients and standard errors, a problem known in econometrics as incidental parameters bias. The most straight forward solution is to use LPM because these models are not typically subject to incidental parameters bias.

[6]This is given by the formula $\overline{X} = \frac{1}{NT-1}\sum_i \sum_t X_{it}$

into account the variation between individuals and that shown for the same individual over time. The score for depression ranges from a minimum of zero to a maximum of 30. The variation between individuals is higher than the variation for an individual over time. 26,033 individuals in at least one wave had scores that indicated depression.

The BMI score has an overall mean of 26 with more variation between individuals than for the same individual over time. BMI scores range from 10 to 45. Low values of BMI (<18.9) and high values (>25) are not desired as they are both health risk factors. A total of 24,922 individuals have a BMI score in at least one wave.

With regard to independent variables, income and age are summarized together with the health outcomes and the other factor variables are summarized in table 3.3. We measure income in logs to reduce the effect of outliers. The overall mean of log per capita real monthly income is 6.6 with incomes varying more between individuals than for the same individuals over time. A total of 26,840 individuals report income in at least one wave.

The average age is 37 years. This takes into account differences between individual's ages and differences for individuals over time for the adult sample. There is more variation between individuals than for an individual over time because of the short panel of the data, which spans only 4 years. The implication of this is that if fixed effects (FE) estimators are used they are inefficient estimators because they rely on variation for individuals over time, which in this case is smaller than between person variation (between variation).[7]

In table 3.3 we present summary statistics for categorical variables. The time invariant variables do not have within person variation. In our case time invariant variables include race and gender (a person remains in a particular race or gender across all the three waves). For example, overall, 79% of the individuals are Africans, 14% are Coloureds, 2% are Asians and 5% are White. If this was a balanced panel, the overall proportions and between proportions would be equal. Variations only occur when some individuals attrit (leave the survey) from the panel in a disproportionate manner. For example, if more Whites attrit relative to other races the proportion of Whites would drop and the overall percentage of whites would differ from that of between race. The table indicates that there are 27,334 individuals of different race who are uniquely identified and each one of these remains in the race category they originally report, across all the three waves.

With regard to gender, we exclude 13 individuals who switch gender across waves. There are more women (59% overall) than men in the sample of adults. In total, 26,534 adults are uniquely identified across the first three waves of NIDS. The differences in the proportion of between and overall proportions is due to attrition.

[7]The formula for between standard deviation is $s_B = \sqrt{\frac{1}{N-1}\sum_i(\overline{X_i} - \overline{X})^2}$ while for within is $s_W = \sqrt{\frac{1}{NT-1}\sum_i\sum_t(X_{it} - \overline{X}_i)^2}$ some softwares present this as $s_W = \sqrt{\frac{1}{NT-1}\sum_i\sum_t(X_{it} - \overline{X}_i + \overline{X}_O)^2}$

Table 3.2: Summary statistics for overall, between and within variation in NIDS

Variable		Mean	Std.Dev.	Min	Max	Observations		
SAH	overall	0.851	0.356	0	1	N	=	51753
	between		0.293	0	1	n	=	26168
	within		0.219	0.185	1.518	T-bar	=	1.978
Depression	overall	7.292	4.492	0	30	N	=	50876
(CES-D score)	between		3.705	0	30	n	=	26033
	within		2.939	-7.208	24.29	T-bar	=	1.954
BMI	overall	25.87	6.119	10.34	44.99	N	=	46500
	between		5.734	11.53	44.99	n	=	24922
	within		2.290	11.14	43.21	T-bar	=	1.866
ln real income	overall	6.643	1.094	0	13.43	N	=	54194
(per capita monthly)	between		1.019	0	12.01	n	=	26840
	within		0.483	0.897	11.20	T-bar	=	2.019
Age	overall	37.30	17.59	15	105	N	=	60965
	between		17.72	15	105	n	=	27248
	within		1.554	34.30	39.97	T-bar	=	2.237

Notes: Summary statistics for dependent variables and continuous variables in the first three waves of NIDS. For each variable we have the overall mean that takes into account time and individual variation, standard deviation for between and within variation, minimum and maximum as well as the number of observations overall, for between and within person variation. n=34378 and N=85279.

The other factor variables are time variant, such as education, marital status, geographical location type and province of residence. This means that an individual can change these states across waves. With regard to educational attainment, 26,949 individuals are uniquely identified in the panel. Some of these individuals have moved to higher levels of educational attainment in the course of the panel. Every time an individual attains a higher level of education they are counted again (double counting). That is why we have 113% as the total percentage for education. This implies that 13% of the sampled individuals reported a higher (different) level of educational attainment in subsequent waves (they were double counted). This is further shown by only 95% of individuals who reported no education remaining in that educational category at the end of the three waves. The other 5% had changed their educational attainment. Education can only go up but for the other variables individuals can move back and forth. The interpretation for the marital status, geographical location type and province of residence follow a similar pattern.

It is clear from the summary statistics that there are variations in the explanatory variables between individuals and for individuals over time. If these variables are correlated with health outcomes, we expect them to vary across waves to correspond to the variation in explanatory variables.

Table 3.3: Summary statistics for categorical variables

	Overall		Between		Within
	Freq.	Percent	Freq.	Percent	Percent
African	48091	78.68	21787	79.71	100
Coloured	8753	14.32	3838	14.04	100
Asian	930	1.520	391	1.430	100
White	3346	5.470	1318	4.820	100
Total	61120	100	27334	100	100
	(n	=	27334)		
Female	31218	58.77	15046	56.70	100
Male	21905	41.23	11488	43.30	100
Total	53123	100	26534	100	100
	(n	=	26534)		
No schooling	6601	11.59	3230	11.99	94.94
General education	23533	41.31	12378	45.93	92.47
Further education and training	13241	23.24	7360	27.31	83.97
Matric	9071	15.92	5001	18.56	84.49
Higher education	4520	7.930	2480	9.200	81.92
Total	56966	100	30449	113.0	88.51
	(n	=	26949)		
Married	13756	25.96	7244	27.32	87.75
Living with partner	3880	7.320	2733	10.31	73.50
Widow	4313	8.140	2470	9.310	79.75
Divorced	1202	2.270	829	3.130	69.04
Never married	29843	56.31	16513	62.27	94.55
Total	52994	100	29789	112.3	89.03
	(n	=	26520)		
Traditional area	23626	40.36	11627	42.89	96.34
Urban	29116	49.74	13808	50.94	95.94
Farms	5799	9.910	2894	10.68	91.94
Total	58541	100	28329	104.5	95.69
	(n	=	27109)		
Western Cape	7838	13.06	3585	13.13	97.45
Eastern Cape	7223	12.04	3505	12.84	95.65
Northern Cape	4488	7.480	2064	7.560	97.32
Free State	3534	5.890	1575	5.770	96.35
KwaZulu-Natal	16103	26.83	7755	28.41	98.54
Northern West	4081	6.800	1878	6.880	95.71
Gauteng	6996	11.66	3378	12.37	90.30
Mpumalanga	4454	7.420	2100	7.690	95.96
Limpopo	5296	8.820	2556	9.360	94.71
Total	60013	100	28396	104.0	96.13
	(n	=	27298)		

Notes: Summary statistics for factor variables in the first three waves of NIDS. For each variable we have the overall mean that takes into account time and individual variation. For time invariant factor variables the between total and the uniquely identified individuals (n) are the same. For time variant factor variables these totals differ because of double counting as individuals switch from one factor to another over time and are double counted.

3.4.2 Empirical Model Formulation

We use two methods to identify the key socioeconomic determinants of health outcomes in South Africa, the bivariate approach and panel data regression. We use the bivariate approach to show the correlation between health outcomes and a number of explanatory variables. The explanatory variables we consider are income, education, and race because of their importance in empirical work as shown by Grossman (2000).

The use of panel data in this study allows us to control for unobserved heterogeneity by calculating panel estimators for the socioeconomic determinants of health outcomes. We are also likely to obtain more precise estimates because of the large sample obtained from panel observations. Several panel estimators can be estimated, including fixed effects (FE), first difference (FD), random effects (RE), Hausman-Taylor, and the Mundlak specification. In this study we restrict ourselves to the FE and RE estimators because they are the panel estimators used most frequently. We contrast these two with the pooled OLS which involves running an OLS regression on data from the three waves stacked together. The panel health outcome model is specified as an individual specific effects model as follows:

$$h_{it} = \alpha_i + \beta X_{it} + \varepsilon_{it} \tag{3.2}$$

Where h_{it} is the health outcome for individual i in wave t ($t = 1, 2, 3$), α_i is the individual specific effects that do not vary by wave, X_{it} is a vector of explanatory variables for individual i in wave t, and ε_{it} is the idiosyncratic error term. The health outcome variables considered in this study are SAH, BMI and Depression. These health outcomes potentially represent the multidimensional nature of health as earlier defined. The explanatory variables include income, education, age, race, gender, marital status, province and geographical type of residence. The use of these variables is informed by both theory and empirical work (Grossman, 2000). Under the assumption that the regressors are independent of the error term and the individual specific intercept (effects), all the panel estimators we consider in this study would be consistent. The RE estimator is in addition efficient under this assumption and would be preferred. We note that income is unlikely to be exogenous.

A pooled OLS is estimated by stacking the data from the three waves together and estimating an OLS on the pooled data. Alternatively the FE model is estimated. The FE model solves the problem of endogeneity (omitted variable bias) even when the source of the bias is unobserved. This is not possible with cross sectional analysis. In a FE model, α_i is allowed to be correlated with X_{it} in equation 3.2 and we define a composite error (u_{it}) for the equation such that $u_{it} = \alpha_i + \varepsilon_{it}$. This permits X_{it} to be correlated with the time invariant component (α_i) of the composite error while we assume that it is uncorrelated with the idiosyncratic error (ε_{it}) component. The FE model is estimated by demeaning the variables, that is, by subtracting from each variable its mean over time before estimating it with OLS

as in equation 3.3 below.

$$h_{it} - \bar{h_i} = \beta(X_{it} - \bar{X_i}) + (u_{it} - \bar{u_i}) \tag{3.3}$$

Where $\bar{h_i}$ is the mean of h_{it} over time t, u_{it} is a composite error term as explained above and $\bar{u_i}$ its mean over time t, $\bar{X_i}$ is the mean of X_{it} over time t and the other variables are as explained in equation 3.2 above. When the FE model is used such as the one in equation 3.3, all the explanatory variables in X_{it} that are time invariant such as race and gender drop out of the model because they are differenced out.

An alternative model that lies between pooled OLS and FE is the RE model. In this model the time invariant component of the error term is assumed to be random and hence the composite error term is not correlated with the regressors. In order to estimate the RE model, a weighted value of the mean of each variable is subtracted from the variables before estimating using OLS. The weight is provided by lambda $\hat{\lambda} = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + \sigma_\alpha^2}}$. Where σ_ε is the standard deviation of the variance between individuals and σ_α^2 is the variance of the individuals over time (that is, within person variance).

$\hat{\lambda}$ ranges between 0 and 1 because it is the difference between 1 and rho $\left(\frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + \sigma_\alpha^2}}\right)$. Rho is the proportion of the between standard error to the sum of the between and within person standard errors. When lambda approaches 0 (which is equivalent to the variance between individuals dominating the variance within), the RE model approaches the pooled OLS. When lambda approaches 1 (equivalent to the variance within dominating the variance between individuals) the RE model approaches the FE model. The RE estimators are therefore always between the pooled OLS and the FE estimators. The RE model is shown in equation 3.4 below.

$$h_{it} - \hat{\lambda}\bar{h_i} = \beta(X_{it} - \hat{\lambda}\bar{X_i}) + (u_{it} - \hat{\lambda}\bar{u_i}) \tag{3.4}$$

Where h_{it},$\bar{h_i}$,X_{it},$\bar{X_i}$,$\hat{\lambda}$,u_{it} and $\bar{u_i}$ are as explained before. The advantage of the RE model is that all the coefficients are estimated but are inconsistent if the FE model is the appropriate model. If the objective is to estimate coefficients for time invariant explanatory variables then the RE and pooled OLS are preferred. However, the parameter estimates may be inconsistent for the reasons explained above.

The Hausman test is the formal method we use for selecting between the FE and the RE models. In this test the null hypothesis is that there is no correlation between the error term and the explanatory variables in a panel model. A statistically significant difference (meaning there is correlation) is interpreted as evidence against the RE model. This test however assumes that the error terms are independently and identically distributed (iid) which may not be the case. In order to solve this problem, a robust Hausman test is used.

We select between the pooled OLS and the RE model by testing for the presence of unobserved effect (under the assumption that if unobserved effect is present, it is uncorrelated with the regressors). If no unobserved effect is present then the use of pooled OLS is appropriate. If there is an unobserved effect that is not correlated with the regressors then the RE is preferred.

3.5 Health outcomes and Socio-demographic factors in South Africa (Bi-variate analysis)

In this section we present a bivariate analysis of health outcomes and key socioeconomic and demographic variables. We analyze three health outcomes, namely SAH, BMI, and depression and some of the socioeconomic and demographic variables. For each health outcome, we first look at the characteristics in terms of their distributions and trend before we analyze their relationship with key covariates. We concentrate on income, education, and race as the key covariates. The first two because of their importance in theoretical work and the last one because of its importance in the South African context.

3.5.1 Self assessed health (SAH)

Self assessed health (SAH) is our primary variable used to capture health outcome as shown in NIDS. The SAH question has a very high response rate in the NIDS dataset of almost 100% (Ardington and Gasealahwe, 2012). If we view health as a multidimensional[8] phenomenon, then SAH is the closest representation of health from the data. Table 3.4 shows the weighted proportions of individuals reporting different health status across waves.

Table 3.4: Frequency and percentage of respondents in each health status by wave

	Wave 1		**Wave 2**		**Wave 3**	
Health status	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Poor	1247	6.31	672	2.99	623	2.96
Fair	2055	11.42	1438	7.06	1652	8.03
Good	3751	23.00	3684	19.47	5348	26.04
Very Good	3909	27.14	5175	29.68	5445	29.13
Excellent	4515	32.13	6616	40.79	5630	33.85
Total	**15477**	**100**	**17585**	**100**	**18698**	**100**

Source: Own computation from NIDS data (adult data only). The proportions are weighted, therefore the percentages are interpreted as weighted percentages.

As we can see from the table, 8,424 individuals representing 59.27% of the population report having very good or excellent health. 3,302 individuals representing 17.73% of the

[8]Health according to WHO is a multidimensional phenomenon defined as a state of complete physical, mental and social well being and not just the absence of disease or infirmity.

population report poor or fair health in wave 1. In wave 2, comparable proportions are 70.47% (11,791 individuals) reporting excellent or very good health and only 10.05% (2,110 individuals) reporting fair or poor health. In wave 3, a total of 62.98% (11,075 individuals) report excellent or very good health while only 10.99% (2,275 individuals) report fair or poor health. There has been an improvement in wave 2 and wave 3 regarding SAH, because the proportion of individuals who report poor health dropped substantially. The table shows that SAH status has changed over time. We now show, using bivariate analysis, the factors correlated with these changes in health outcomes.

SAH and Income

Household income is divided into five quintiles in each wave of NIDS. This results in disproportionately more people in the poorer quintile in all the waves because poorer households are generally bigger. For example, in wave 1 there are 6400, 7013, 6137, 5275 and 3401 individuals in each income quintile from the poorest to the richest. In general poorer households are bigger because they have more children than richer households.[9] Our analysis focuses on individuals aged at least 15 years. Excluding children aged younger than 15 years has the effect of reducing the individuals in each quintile but poorer quintiles reduce by a higher number because they have the highest number of children. We also reduce the number of individuals in each income quintile by those interviewed by proxy for reasons explained in chapter two, that is, we utilize information on self assessment of health status. The remaining individuals are grouped according to their SAH status in each wave. Because children are excluded richer quintiles end up being better represented than poor quintiles, that is, they have a higher proportion of individuals.

Table 3.5 shows the bivariate relationship between SAH and real per capita monthly income. The real per capita income was computed as a simple household mean income to allow for comparison between our results and other results that have used NIDS data. In addition, there is no established equivalence scale for South Africa (Woolard and Leibbrandt, 2006). Column totals for each wave indicate the proportion (weighted) of individuals in a given income quintile. The aggregate total is the actual number of individuals in the income quintile for the wave. Row totals for the aggregate total add up to the total number of individuals interviewed. Individuals in each health state are divided by income quintile so row totals should add to 100%. Comparison is done across columns which shows that individuals in the richest quintile are approximately twice as likely to report excellent or very good health (1.56 and 1.57 times as likely respectively).

Thus the poorest income quintile, has more individuals with poor health than their proportion in the population. The proportion of those with poor health in the richest income

[9]There are 2920, 2876, 1925, 1201 and 683 children in quintile 1 to 5 respectively in wave 1

quintile is less than their proportion in the population. The proportion of individuals with excellent or very good health increases (relative to their population proportions) as we move to the higher income quintiles.

The above results are similar across waves. For example, in wave 2 those in the richest quintiles are 1.8 times more likely to report excellent health and 1.52 times more likely to report very good health, relative to those in the poorest income quintile. In wave 3 individuals in the richest quintile are 1.65 times more likely to report excellent health and 1.86 times more likely to report very good health. This shows that there is a positive relationship between SAH and income in South Africa. As we have seen, individuals with higher incomes are more likely to have better health. The statistical significance of this relationship is confirmed using econometric methods, as discussed in subsequent sections.

We have used income per capita obtained from simple household mean incomes. The above analysis could therefore change if a different scale is used in calculating income per capita. For the case of South Africa, as we have mentioned, there is no established equivalence scale to use, our results are therefore comparable to other results obtained in South Africa.

SAH and Education

In table 3.6 we show health status (SAH) by educational attainment. The table shows that in wave 1, close to 9% of the population do not have any formal education. Over a third (36.10%) of the population have general education. A quarter of the population have further education and training. 19% have Matric and 11% have a higher education qualification. We expect that if educational attainment does not influence health outcomes, the proportions of the population with different health states should be apportioned similarly. However as we can see from the table, more individuals with no education or general education have poor self-reported health, relative to those with Matric or higher education qualifications. This is true across all waves.

Those with Matric were 3.5 times more likely to report excellent health compared to those with no education. In wave 1 those with no education were 10 times more likely to report poor health than those with Matric. These results are obtained by dividing the proportion of individuals with a given health status and education by their proportion in the population. These are then compared across levels of educational attainment. It is clear from the table that the proportion of individuals with no education and general education who reported poor health was higher than their proportion in the population across all waves.

In wave 2 for example, those with no education were 21 times more likely to report poor health than those with Matric. Those with Matric were 3 times more likely to report excellent health than those with no education. The pattern does not change in wave 3, where those with no education were 8 times more likely to report poor health than those with Matric. This

Table 3.5: SAH and Income Quintiles

Wave 1	**Income Quintiles**					
Health Status	Poorest	Quint 2	Quint 3	Quint 4	Richest	Total
Poor	19.59	22.72	26.55	24.42	6.73	100
Fair	17.58	19.80	23.16	25.76	13.70	100
Good	16.85	19.48	20.37	21.33	21.97	100
Very Good	17.58	18.27	16.97	19.58	27.60	100
Excellent	17.01	14.13	17.78	24.49	26.59	100
Total	**17.36**	**17.67**	**19.32**	**22.57**	**23.07**	**100**
Aggregate total	**3027**	**3483**	**3415**	**3347**	**2205**	**15477**
Wave 2						
Poor	21.69	14.70	18.95	28.89	15.77	100
Fair	19.87	18.99	17.58	22.25	21.31	100
Good	19.38	15.47	18.80	20.60	25.75	100
Very Good	18.24	16.65	17.27	20.03	27.82	100
Excellent	17.43	14.31	15.74	20.43	32.09	100
Total	**18.35**	**15.57**	**17.02**	**20.73**	**28.34**	**100**
Aggregate total	**3545**	**3516**	**3500**	**3530**	**3494**	**17585**
Wave 3						
Poor	23.81	23.76	14.63	21.72	16.08	100
Fair	21.03	18.17	21.81	16.39	22.60	100
Good	19.79	17.97	17.27	18.43	26.55	100
Very Good	16.49	17.11	18.48	17.28	30.64	100
Excellent	17.08	17.44	18.07	19.30	28.12	100
Total	**18.13**	**17.73**	**18.18**	**18.32**	**27.64**	**100**
Aggregate total	**3763**	**3733**	**3763**	**3733**	**3706**	**18698**

Notes: Table shows health status by income quintiles in the three waves of NIDS. The column totals for each wave indicate the weighted proportion of individuals in each income quintile while the other values in the table indicate the weighted proportion of individuals in a specific quintile who report a given health status. Aggregate total is the actual number of individuals in each wave. Row totals add up to one (any difference is due to rounding).

analysis points to the positive relationship between education attainment and SAH which is consistent with the predictions of the Grossman model of health human capital (Grossman, 2000).

SAH and Race

In table 3.7 the African sub group comprises 79% of the total population across all waves of NIDS, while the Asian sub group is very small at 2%. This is confirmed by the aggregate total which shows the total number of individuals in each race. The implication of this is that we may not make inferences with regard to the Asian sub population. The proportion of Whites dropped by 1 percentage point in wave 3 due to attrition.

We compare health outcomes among races by comparing the percentages in the health

Table 3.6: SAH and Educational attainment

Wave 1	**Education categories**					
Health status	**No educ**	**General educ**	**FET**	**Matric**	**Higher educ**	Total
Poor	26.83	50.49	15.63	5.56	1.49	100
Fair	22.06	46.21	17.18	9.42	5.13	100
Good	10.95	35.62	25.95	17.60	9.87	100
Very Good	5.33	35.31	26.69	20.11	12.55	100
Excellent	3.52	30.68	26.10	25.44	14.27	100
Total	**9.31**	**36.10**	**24.55**	**19.11**	**10.94**	**100**
Aggregate total	**2136**	**6498**	**3369**	**2372**	**1102**	**15477**
Wave 2						
Poor	21.65	56.00	17.25	2.70	2.39	100
Fair	18.34	47.23	17.88	11.79	4.75	100
Good	11.34	40.95	23.44	13.89	10.37	100
Very Good	5.91	37.09	24.90	20.53	11.57	100
Excellent	3.35	29.78	27.36	23.48	16.02	100
Total	**7.27**	**36.15**	**24.89**	**19.29**	**12.40**	**100**
Aggregate total	**2133**	**7464**	**4028**	**2720**	**1236**	**17581**
Wave 3						
Poor	16.71	54.91	20.26	5.79	2.33	100
Fair	17.09	42.65	21.80	9.59	8.87	100
Good	9.58	38.55	23.54	14.34	14.00	100
Very Good	4.73	34.42	26.19	18.40	16.26	100
Excellent	2.04	30.23	30.11	20.99	16.64	100
Total	**6.43**	**35.34**	**26.30**	**17.14**	**14.79**	**100**
Aggregate total	**1952**	**7702**	**4613**	**2687**	**1729**	**18683**

Notes: Table shows health status by educational attainment in the three waves of NIDS. The column totals for each wave indicate the weighted proportion of individuals with a given level of education while the other values in the table indicate weighted proportion in a given education category who report a given health status. Aggregate total is the actual number of individuals in each wave. Row totals add up to one (any difference is due to rounding). FET is Further Education and Training.

categories across columns. However this is done after adjusting[10] for the effect of relative size of the race group in the population by comparing the ratios relative to their proportion in the population. For example, using this approach, Africans are twice as likely to report poor health than Whites. As we can see the proportion of Africans who report poor health is higher than their proportion in the population. On the other hand the proportion of Whites who report poor health is smaller than their proportion in the population. This is a pointer that Africans are more likely to be in poor health than Whites. The situation is however different in wave 2. Here although Africans have higher proportions in all categories of health status, the proportion of individuals reporting poor health is lower than their proportion in the population. This pattern is surprising given that the proportion of Whites reporting poor health is higher than their overall proportion in the population in wave 2. This could be due to measurement error in wave 2. The proportions of individuals who report in different health categories in wave 1 and wave 3 are consistent with our expectations.

There was no change in relative advantage in health risk between Africans and Coloureds in wave 3. Africans reported poor health status three times more than Whites. The proportion of Whites who reported poor health was smaller than their proportion in the population. The bivariate analysis points to racial differences in health status but the differences are not so large. We therefore have to test the significance of the differences to make conclusions.

3.5.2 Body mass index

We now direct our focus to another health dimension, BMI, and its relationship with some key covariates. BMI is an indirect health indicator in that high BMI (overweight and obese) is a risk factor associated with diseases such as hypertension, coronary heart disease, diabetes, stroke, and cancer. We focus only on the BMI of adults aged 15 years and older because it is more straight-forward to calculate than that of children.

Figure 3.1 shows kernel density estimates of BMI in South Africa for the total population of adults in the three waves of NIDS. As we can see BMI has increased over time as shown by a shift of the curves to the right. However it is not clear why BMI is higher in wave 2, than wave 3.

We use the Kolmogorov Smirnov (K-S) test to check if the differences we observe in the distribution functions of BMI from one wave to another are statistically significant (Goerg et al., 2009). The K-S tests the difference in the distribution between two groups (waves in our case) at a time. We test the difference between wave 1 and wave 2 and between wave 1 and wave 3 and finally wave 2 and wave 3. Each set of the K-S test output comprises three results based on three hypotheses. The first is that BMI in group 1 (wave 1) contains smaller values

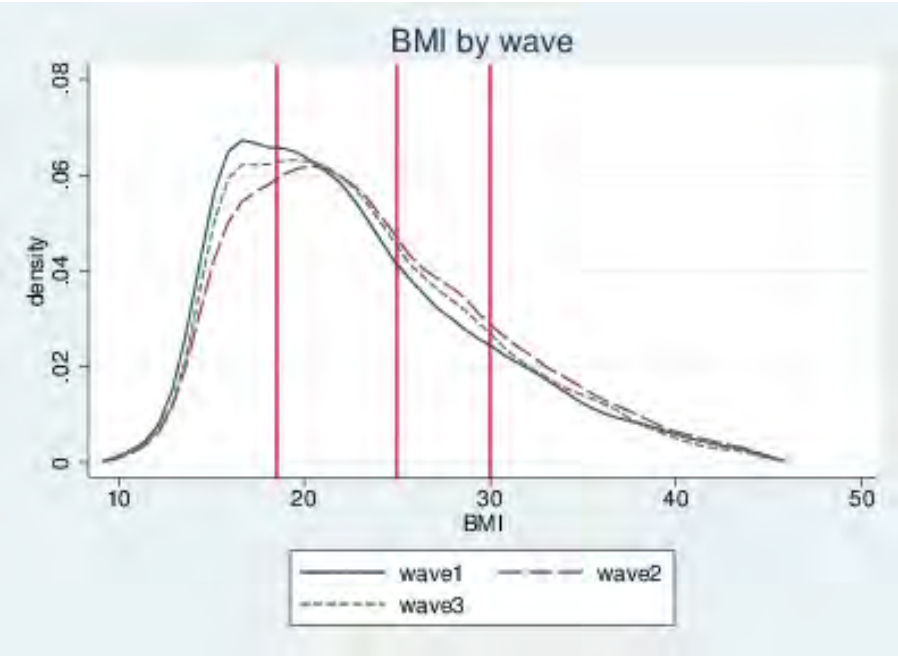
[10]Adjustment entails getting the ratio of the proportion of individuals in a particular health state to their racial total proportion in the population and then comparing this across races.

Table 3.7: Race and SAH by wave

Wave 1	**Race group**				
Health status	**African**	**Coloured**	**Asian**	**White**	Total
Poor	83.65	8.20	1.75	6.41	100
Fair	79.72	9.43	3.61	7.23	100
Good	79.94	7.30	2.78	9.98	100
Very Good	76.10	8.15	2.22	13.53	100
Excellent	79.45	8.37	2.09	10.10	100
Total	**78.95**	**8.17**	**2.44**	**10.4**	**100**
Aggregate total	**12141**	**2207**	**224**	**905**	**15477**
Wave 2					
Poor	71.24	12.89	3.98	11.90	100
Fair	80.88	9.48	3.79	5.85	100
Good	79.29	9.71	2.72	8.28	100
Very Good	81.33	7.39	1.99	9.30	100
Excellent	77.39	8.22	2.69	11.70	100
Total	**78.99**	**8.49**	**2.60**	**9.91**	**100**
Aggregate total	**14602**	**2245**	**193**	**545**	**17585**
Wave 3					
Poor	81.60	9.82	4.85	3.73	100
Fair	76.47	8.65	2.10	1.28	100
Good	80.31	8.28	2.43	8.98	100
Very Good	75.55	10.36	2.51	11.58	100
Excellent	81.76	8.48	2.26	7.50	100
Total	**79.15**	**9.03**	**2.44**	**9.39**	**100**
Aggregate total	**15397**	**2575**	**195**	**531**	**18698**

Notes: The table shows health status by race in the three waves of NIDS. The proportions are weighted by post stratification weights provided for each wave. The column totals for each wave indicate the weighted proportion of individuals in each race group while the other values in the table indicate the weighted proportion in a specific race group by a given health status. Aggregate total is the number of individuals in each race. Row totals add up to 100% (any difference is due to rounding).

Figure 3.1: Body mass index by Wave



Notes: The figure shows kernel density for BMI of adults aged 15 years and above for the three waves of NIDS. The densities for wave 2 and wave 3 are to the right of wave 1 implying that BMI has increased over this period (confirmed as significant by a Kolmogorov Smirnov test). The density for wave 3 is however to the left of wave 2 indicating that BMI is lower in wave 3 than wave 3. Vertical lines in the diagram indicate the cutoff points for underweight, normal weight, and overweight, respectively

than those in group 2 (wave 2). The largest difference between the distribution functions for the two groups forms part of the output of the test results. The second hypothesis is that BMI in group 1 (wave 1) contains larger values than those in group 2. The largest difference in the distribution functions in this direction is also part of the output of the results. The final hypothesis combines the first and second hypotheses and tests the hypothesis that the two distributions are the same. In all cases the P-value is reported which determines whether we reject or do not reject the null hypothesis.

The results of the Kolmogorov Smirnov test for the BMI distributions in wave 1, 2 and 3 are presented in table 3.8. We test for the difference in distribution between wave 1 and wave 2 first and the results show that we reject the null hypothesis and conclude that the distributions are different. In addition, the distribution function of wave 1 BMI has lower values than the BMI distribution function of wave 2. A comparison of the distribution of BMI in wave 1 and wave 3 shows similar results, that the wave 1 distribution function has lower values than the wave 3 distribution function. However when we apply K-S test on the distribution of wave 2 and wave 3, the test reveals that the distribution function of wave 2 does not have larger values than the distribution function of wave 3. The overall conclusion is that the BMI distributions across waves are different.

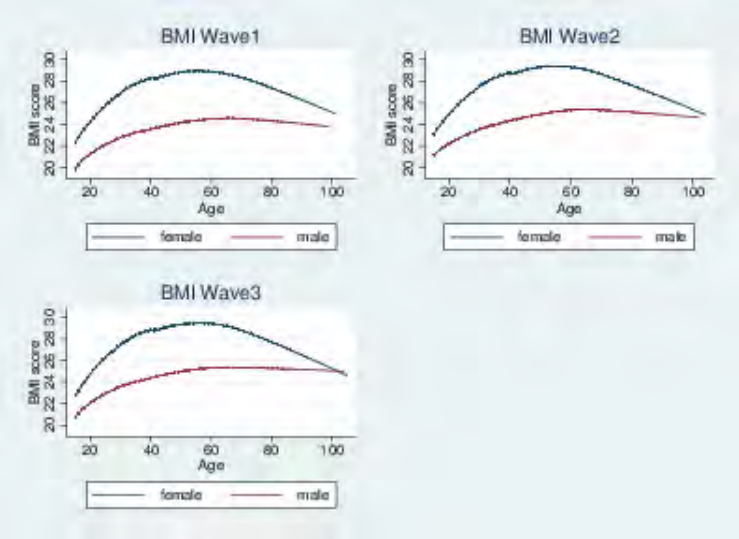
In addition there are differences between BMI for men and women for all ages, with BMI

Table 3.8: Kolmogorov Smirnov test results for difference in BMI distributions

Wave	**D**	**P-value**	**Corrected**
Wave 1	0.0489	0.0000	
Wave 2	-0.0008	0.991	
Combined K-S	15.99	0.0000	0.0000
n	16751		
Wave 1	0.0555	0.0000	
Wave 3	-0.0051	0.665	
Combined K-S	0.0555	0.0000	0.0000
n	26759		
Wave 2	0.0207	0.001	
Wave 3	-0.0101	0.187	
Combined K-S	0.0207	0.002	0.002
n	18994		

Notes: Table shows Kolmogorov Smirnov test results for testing if BMI distributions are different across waves. Each panel tests for the difference between two distributions at a time. The first part test for the difference between wave and wave 2, the second part between wave 1 and wave 3 and the third part between wave 2 and wave 3.

Figure 3.2: BMI by Age and gender for each wave



Notes: The figure shows distribution of BMI by age and gender for the three waves of NIDS. At all ages BMI is higher for women than for men and the difference is significant. Both increase with age but at a decreasing rate.

for women above that of men. This phenomenon is shown in figure 3.2. The figure shows that as people grow old the rate at which BMI increases slows down. We use the Kolmogorov Smirnov test explained above to check if the differences between the distribution functions for men and women are statistically significant and find them to be statistically significant.

BMI and Income

Table 3.9 shows the relationship between BMI and income quintile for each wave in NIDS. Column totals indicate weighted proportion of individuals in income quintiles. Each row indicates the weighted proportion of individuals in a given BMI category and this adds up to 100%. The proportion of individuals in each income quintile ranges from 18% to 27% across all waves. The table shows that BMI in each wave is higher among higher income groups. A higher proportion of individuals are underweight in lower quintiles, relative to their proportion in the population.

On the other hand the proportion of individuals who were obese or overweight is higher in richer quintiles. There is no major difference in weight between income quintiles in wave 1. There is also no major differences in wave 2 between income quintiles. However in waves 2 and 3 individuals in the richest quintile are twice as likely to be obese.

Since BMI is an indicator of fatness (Bailey and Ferro-Luzzi, 1995; Naidu and Rao, 1994; Perry, 2009), it appears that the positive relationship between BMI and income is through increased intake of food afforded by people in higher income quintiles. This is because fatness is associated with food intake among other factors.

BMI and Education attainment

Table 3.10 presents the relationship between BMI and education attainment. The column totals show weighted proportion of individuals with different levels of education. Aggregate totals are the actual number of individuals in the sample with these levels of education. The other values in the table indicate weighted proportions of BMI categories by education. Individuals with higher education are twice as likely to be overweight than those with no education. It is evident across all waves of NIDS that the proportion of underweight individuals decreases with higher levels of education. On the other hand the proportion of those who are overweight also increases with income.

BMI and Race

Table 3.11 shows the relationship between BMI and race. Column totals for each wave indicate the weighted proportion of the population of each race group. The aggregate total for each wave is the actual number of individuals in that race group. The other values in the table indicate the weighted proportion of BMI scores by race. As we can see in the table, there are more Africans (approximately 80%) in the population compared to any other race group. Indians are very few in the population (approximately 2.5%) which means we can not

Table 3.9: BMI categories and Income quintiles

Wave 1	**Income Quintiles**					
BMI category	**Poorest**	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Richest**	Total
Underweight	23.20	21.66	23.26	21.04	10.84	100
Normal	19.78	19.57	19.94	22.12	18.60	100
Overweight	15.99	15.58	16.62	24.40	27.41	100
Obese	14.60	16.05	20.65	22.45	26.25	100
Total	**17.96**	**17.98**	**19.56**	**22.65**	**21.84**	**100**
Aggregate total	**2782**	**3137**	**3054**	**2899**	**1777**	**13649**
Wave 2						
Underweight	24.40	18.52	18.80	19.45	18.84	100
Normal	21.12	16.41	17.86	22.14	22.47	100
Overweight	17.07	16.07	17.17	19.70	29.99	100
Obese	15.05	15.14	15.61	20.53	33.67	100
Total	**18.80**	**16.15**	**17.18**	**20.94**	**26.93**	**100**
Aggregate total	**3130**	**3049**	**2989**	**3023**	**2853**	**15044**
Wave 3						
Underweight	23.16	23.68	18.73	18.60	15.83	100
Normal	20.71	20.32	19.11	19.10	20.76	100
Overweight	16.29	15.22	17.85	17.78	32.85	100
Obese	14.71	15.19	16.75	17.54	35.81	100
Total	**18.17**	**17.87**	**18.17**	**18.34**	**27.45**	**100**
Aggregate total	**3694**	**3661**	**3672**	**3642**	**3598**	**18267**

Notes: Table shows BMI by income quintile in the three waves of NIDS. The column totals for each wave indicate the weighted proportion of individuals in each income quintile while the other values in the table indicate the weighted proportion in an income quintile who are in a given BMI category. Aggregate total is the actual number of individuals in each income quintile. Row totals add up to 100% (any difference is due to rounding).

Table 3.10: BMI and education categories by wave

Wave 1	**Education Categories**					
BMI category	**No educ**	**General educ**	**FET**	**Matric**	**Higher educ**	Total
Underweight	11.65	47.06	25.09	12.43	3.76	100
Normal	8.25	38.29	25.36	20.17	7.93	100
Overweight	8.57	30.74	25.90	19.45	15.34	100
Obese	10.66	37.18	23.24	16.51	12.42	100
Total	**9.14**	**36.91**	**24.99**	**18.57**	**10.40**	**100**
Aggregate total	**1856**	**5840**	**3041**	**2017**	**895**	**13649**
Wave 2						
Underweight	6.44	47.40	23.79	14.69	7.68	100
Normal	6.85	38.49	25.60	19.68	9.38	100
Overweight	7.36	34.97	25.84	19.58	12.24	100
Obese	9.02	33.90	23.05	18.83	15.20	100
Total	**7.51**	**37.09**	**24.86**	**19.08**	**11.46**	**100**
Aggregate total	**1836**	**6411**	**3484**	**2321**	**989**	**15041**
Wave 3						
Underweight	5.20	53.63	25.41	11.64	4.11	100
Normal	5.98	39.71	27.55	16.09	10.67	100
Overweight	6.98	29.08	25.17	20.44	18.32	100
Obese	6.51	30.08	26.16	17.03	20.22	100
Total	**6.34**	**35.24**	**26.45**	**17.24**	**14.72**	**100**
Aggregate total	**1869**	**7522**	**4538**	**2643**	**1680**	**18252**

Notes: Table shows BMI by education in the three waves of NIDS. The column totals for each wave indicate the weighted proportion of individuals in each education category while the other values in the table indicate the weighted proportion in the education categories who are in a given BMI category. Aggregate total is the actual number of individuals in each education category. Row totals add up to 100% (any difference is due to rounding). FET is Further Education and Training.

Table 3.11: BMI and Race by wave

Wave 1	**Race group**				
BMI category	**African**	**Coloured**	**Asian**	**White**	Total
Underweight	83.80	11.45	2.62	2.13	100
Normal	84.94	6.15	1.76	7.15	100
Overweight	76.03	9.23	3.37	11.37	100
Obese	76.39	9.05	2.38	12.18	100
Total	**80.79**	**7.95**	**2.35**	**8.92**	**100**
Aggregate total	**10978**	**1843**	**179**	**649**	**13649**
Wave 2					
Underweight	79.92	13.10	4.32	2.66	100
Normal	84.98	6.87	2.47	5.68	100
Overweight	79.53	8.91	2.62	8.94	100
Obese	77.61	7.89	1.64	12.86	100
Total	**81.39**	**8.07**	**2.42**	**8.12**	**100**
Aggregate total	**12824**	**1734**	**153**	**333**	**15044**
Wave 3					
Underweight	81.65	12.86	2.35	3.14	100
Normal	83.73	9.12	2.13	5.02	100
Overweight	77.04	7.29	3.01	12.66	100
Obese	74.33	8.78	2.47	14.43	100
Total	**79.50**	**8.75**	**2.46**	**9.29**	**100**
Aggregate total	**15087**	**2481**	**189**	**510**	**18267**

Notes: The table shows BMI by race in the three waves of NIDS. The column totals for each wave indicate the weighted proportion of individuals in each race while the other values in the table indicate the weighted proportion in a race who fall into a given BMI category. Aggregate total is the actual number of individuals in each wave in a particular BMI category. Row totals add up to 100% (any difference is due to rounding). There are disproportionately more Africans who are underweight than Whites.

make any meaningful inference for them.

The table shows that disproportionately more Africans are underweight than Whites, and disproportionately more Whites are overweight or obese than Africans. For example relative to Whites, Africans are 4 times more likely to be underweight in wave 1, while they are 3 times more likely to be underweight in waves 2 and wave 3. However, the small number of Whites and Indians in the sample may be driving the results in wave 1.

3.5.3 Depression

In this section we present the relationship between depression (mental health) and three key covariates, namely income, education, and race. These covariates are known in theory to influence health (Grossman, 2000). The other control variables are shown in the panel regression analysis. As we explain in the previous chapter, depression is measured through

a score obtained by summing up the responses from the CES-D questions. A person is categorized as depressed if their CES-D score is greater than ten (Radloff, 1977). In this section therefore we only consider if a person is depressed or not.

Depression and income quintiles

Depression is measured as a binary variable with 1 (CES-D>10) implying the individual is depressed and zero (CES-D<10) implying the individual is not depressed. Table 3.12 indicates the relationship between depression and income. Income quintiles were generated from the household incomes of the whole sample. We however excluded children and those interviewed by proxy. The totals we report are only those adults who have CES-D scores. Column totals indicate the weighted proportion of individuals in the income quintiles. Roughly 16% to 18% of the population on average across waves fall in the poorest income quintile while 22% to 29% fall in the richest income quintiles. These values should not be confused with poverty head counts because they have been derived conditional on age eligibility and having a CESD score.

Across all the waves, relative to the poorest income quintile, individuals in the richest income quintile are 1.4 times more likely not to be depressed. This is reinforced by the fact that the proportion of the depressed in the richest income quintile is lower than their proportion in the population. In addition across all the waves, the proportion of the depressed individuals in the poorest income quintile is higher than their proportion in the population. This implies that depression is negatively correlated with income across all waves of NIDS.

Depression and education

In table 3.13 we present the relationship between depressive symptoms and education. Column totals for each wave indicate weighted proportion of individuals in the education categories. Aggregate total is the actual number of individuals in the education category. Relative to individuals who have attained higher education, those with general education are 1.7 times as likely to be depressed in wave 1, 1.3 times as likely to be depressed in wave 2 and 2 times as likely in wave 3. This implies that depressive symptoms decrease with education. We also notice that as we move from lower levels of education, proportion of those not depressed become higher than their proportion in the population. The reverse is true for depressed status. This further confirms that education insulates people from depressive symptoms. However such bivariate analysis may not show clearly the differences in depression (or any other health status) for groups that are close to each other such as between those with Matric and Higher education.

Depression and Race

Table 3.14 shows the relationship between depression and race. Column totals for each wave indicate the weighted proportion of the population for each race group. The aggregate total

Table 3.12: Depression and income quintiles

Wave 1	Poorest	Quintile 2	Quintile 3	Quintile 4	Richest	Total
Not depressed	14.42	16.79	17.92	23.07	27.80	100
Depressed	20.41	18.86	23.26	22.18	15.29	100
Total	**16.72**	**17.58**	**19.97**	**22.73**	**22.99**	**100**
Aggregate total	**3208**	**3775**	**3750**	**3671**	**2405**	**16809**
Wave 2						
Not depressed	16.70	15.04	16.88	20.68	30.70	100
Depressed	21.64	16.34	18.06	20.16	23.80	100
Total	**18.17**	**15.42**	**17.23**	**20.53**	**28.65**	**100**
Aggregate total	**3655**	**3656**	**3658**	**3650**	**3654**	**18273**
Wave 3						
Not depressed	16.15	16.90	17.61	18.69	30.65	100
Depressed	23.49	20.12	19.77	17.81	18.81	100
Total	**18.01**	**17.71**	**18.16**	**18.47**	**27.65**	**100**
Aggregate total	**3843**	**3813**	**3827**	**3840**	**3810**	**19133**

Notes: The table shows depression by income quintile in the three waves of NIDS. The column totals for each wave indicate the weighted proportion of individuals in each income quintile while the other values in the table indicate the weighted proportion in a specific income quintile who have depressive symptoms. Aggregate total is the actual number of individuals in each income quintile. Row totals add up to 100% (any difference is due to rounding).

Table 3.13: Depression and Education attainment

Wave 1	Education categories					
Depression	No education	General education	FET	Matric	Higher education	Total
Not depressed	7.47	32.92	25.20	21.27	13.14	100
Depressed	11.13	40.29	23.42	18.15	7.01	100
Total	**8.86**	**35.72**	**24.52**	**20.08**	**10.81**	**100**
Aggregate total	**2231**	**6971**	**3650**	**2696**	**1183**	**16731**
Wave 2						
Not depressed	6.45	35.01	25.20	20.05	13.28	100
Depressed	8.78	36.72	25.52	19.16	9.81	100
Total	**7.14**	**35.52**	**25.30**	**19.79**	**12.25**	**100**
Aggregate total	**2270**	**8164**	**4528**	**3198**	**1416**	**19576**
Wave 3						
Not depressed	5.46	33.35	26.25	18.29	16.65	100
Depressed	9.10	40.52	26.30	15.60	8.48	100
Total	**6.38**	**35.16**	**26.26**	**17.61**	**14.59**	**100**
Aggregate total	**2103**	**8403**	**5065**	**3177**	**1921**	**20669**

Notes: The table shows depression by educational attainment in the first three waves of NIDS. The proportions are weighted by post stratification weights provided for each wave. The column totals for each wave indicate the weighted proportion of individuals in each education category while the other values in the table indicate the weighted proportion in specific education categories who report a given depressive symptom. Aggregate total is the actual number of individuals in each education category in the wave. Row totals add up to 100% (any difference is due to rounding).

Table 3.14: Depression and race

Wave 1	Africans	Coloureds	Indians	Whites	Total
Not depressed	75.42	8.52	2.77	13.29	100
Depressed	82.24	9.38	2.27	6.11	100
Total	**78.04**	**8.85**	**2.58**	**10.53**	**100**
Aggregate total	**13061**	**2481**	**263**	**1004**	**16809**
Wave 2					
Not depressed	76.57	9.54	2.60	11.28	100
Depressed	81.41	7.25	3.20	8.14	100
Total	**78.01**	**8.86**	**2.78**	**10.35**	**100**
Aggregate total	**17244**	**3060**	**343**	**1195**	**21842**
Wave 3					
Not depressed	76.55	9.44	2.87	11.13	100
Depressed	85.22	8.74	1.64	4.39	100
Total	**78.74**	**9.27**	**2.56**	**9.43**	**100**
Aggregate total	**17795**	**3213**	**324**	**1147**	**22479**

Notes: The table shows depression by race in the first three waves of NIDS. The proportions are weighted by post stratification weights provided for each wave. The column totals for each wave indicate the weighted proportion of individuals in each race category, while the other values in the table indicate the weighted proportion in specific races who report a given depressive symptom. Aggregate total is the actual number of individuals in each race in the wave. Row totals add up to 100% (any difference is due to rounding).

for each wave is the actual number of individuals in that race group. The other values in the table indicate the weighted proportion by depression status and by race. As we can see in the table, a higher proportion of Africans (approximately 78%) are in the population compared to any other race group. Indians are very few in the population (approximately 2.5%) which means we cannot make any meaningful inference from the data for this group.

The table shows that disproportionately more Africans are depressed relative to Whites and disproportionately more Whites are not depressed across all three waves of NIDS. Africans are roughly twice as likely to be depressed as Whites in wave 1, while they are 1.3 and 2.3 times more likely to be depressed in wave 2 and wave 3 respectively. However, the small number of Whites and Indians in the sample imply that we may not be able to make inferences on the basis of these results.

3.5.4 Blood pressure

In this section we present the relationship between blood pressure and its key correlates, namely income, education and age. As explained in chapter two, we compute blood pressure from the measurements of blood pressure systolic and blood pressure diastolic collected in the survey. Blood pressure is measured as a ratio of systolic blood pressure in the numerator

Table 3.15: Blood Pressure and Income quintiles

Wave 1	Poorest	Quintile 2	Quintile 3	Quintile 4	Richest	Total
Normal BP	19.76	17.91	19.69	22.17	20.47	100
HBP	17.50	20.02	19.47	24.08	18.93	100
Total	**19.45**	**18.20**	**19.66**	**22.44**	**20.25**	**100**
Aggregate total	**1747**	**1875**	**1829**	**1661**	**997**	**8109**
Wave 2						
Normal BP	19.05	15.85	16.92	21.46	26.73	100
HBP	19.04	18.91	18.68	20.54	22.82	100
Total	**19.05**	**16.24**	**17.15**	**22.34**	**26.22**	**100**
Aggregate total	**1744**	**1679**	**1670**	**1721**	**1545**	**8359**
Wave 3						
Normal BP	18.33	17.79	18.07	17.96	27.85	100
HBP	20.12	16.53	17.45	20.61	25.29	100
Total	**18.56**	**17.62**	**17.99**	**18.31**	**27.51**	**100**
Aggregate total	**2127**	**2116**	**2170**	**2100**	**2048**	**10561**

Notes: Table shows blood pressure status by income quintile in each wave of NIDS. The column totals for each wave indicate the weighted proportion of individuals in each income quintile while the other values in the table indicate the weighted proportion in the income quintile who have a given blood pressure state. Aggregate total is the actual number of individuals in each income quintile. Row totals add up to 100% (any difference is due to rounding).

and diastolic blood pressure in the denominator.

Individuals are categorized as falling within a specific category of blood pressure based on thresholds already highlighted in chapter two. We exclude individuals with blood pressure measurements that lie outside the thresholds. The consequence of this is that our sample is greatly reduced. We then divide the sample into those with high blood pressure and those with normal blood pressure. Individuals are categorized as having high blood pressure if blood pressure systolic is greater than 140 mm Hg and blood pressure diastolic is greater than 90 mm Hg.

Blood pressure and Income quintiles

In table 3.15 we present blood pressure and income quintile. In the table column totals indicate weighted proportion of individuals in the income quintile. The aggregate total is the actual number of individuals in the income quintile. The other values in the table indicate weighted proportion of individuals with either normal or high blood pressure who fall within a given income quintile. As we can see there is no clear pattern between blood pressure and income quintiles in wave 1. The weighted proportion of individuals in each income quintile is close to the weighted proportion of individuals with any blood pressure score.. This is also

Table 3.16: Blood pressure and educational attainment

Wave 1	Education categories					
Blood pressure	No education	General education	FET	Matric	Higher education	Total
Normal BP	6.30	35.87	26.93	20.18	10.71	100
HBP	19.51	46.27	17.60	9.46	7.17	100
Total	**8.13**	**37.32**	**25.64**	**18.70**	**10.22**	**100**
Aggregate total	**973**	**3442**	**1907**	**1249**	**538**	**8109**
Wave 2						
Normal BP	4.60	35.25	27.27	21.12	11.76	100
HBP	15.21	46.33	18.63	10.90	8.94	100
Total	**5.98**	**36.69**	**26.15**	**19.79**	**11.39**	**100**
Aggregate total	**827**	**3588**	**2077**	**1333**	**533**	**8358**
Wave 3						
Normal BP	4.32	34.25	28.60	18.15	14.68	100
HBP	12.50	41.96	20.12	13.50	11.93	100
Total	**5.40**	**35.27**	**27.48**	**17.54**	**14.32**	**100**
Aggregate total	**925**	**4303**	**2776**	**1593**	**954**	**10551**

Notes: Table shows blood pressure status by educational attainment in each wave of NIDS. The column totals for each wave indicate the weighted proportion of individuals in each education category while the other values in the table indicate the weighted proportion in the education category who have a given blood pressure. Aggregate total is the actual number of individuals in each education category. Row totals add up to 100% (any difference is due to rounding).

the case in waves 2 and 3.

Blood pressure and educational attainment

In table 3.16 we present the relationship between blood pressure and education. Column totals indicate the weighted proportion of individuals with a given education. Aggregate total is the actual number of individuals with a given level of education. The other values in the table indicate weighted proportion of individuals with different education levels who have normal or high blood pressure. The table shows clearly that high blood pressure is negatively related to educational attainment across all three waves of NIDS. For example, individuals with no education are 3 times more likely to have high blood pressure than individuals with higher education, . Individuals with general education are twice as likely to have high blood pressure as those with higher education, after adjusting[11] for their population size. These results suggest that blood pressure is negatively associated with education.

Blood pressure and race

In table 3.17 we show the relationship between blood pressure and race in NIDS. Column totals for each wave indicate the weighted proportion of the population for each race group. The aggregate total for each wave is the actual number of individuals in that race group. The other values in the table indicate the weighted proportion of blood pressure status (normal or high blood pressure) by race. There is a higher proportion of Africans in the population. Indians are very few in the population (approximately 2%) which means we can not make

[11] Adjustment is done by taking into account their proportions in the population

Table 3.17: Blood Pressure and Race

Wave 1	Africans	Coloureds	Indians	Whites	Total
Normal BP	82.79	7.21	1.97	8.03	100
HBP	78.19	11.84	1.50	8.47	100
Total	**82.15**	**7.85**	**1.91**	**8.09**	**100**
Aggregate total	**6601**	**1066**	**103**	**339**	**8109**
Wave 2					
Normal BP	83.69	6.21	3.01	7.09	100
HBP	78.68	10.84	1.65	8.83	100
Total	**83.04**	**6.81**	**2.83**	**7.31**	**100**
Aggregate total	**7266**	**845**	**85**	**163**	**8359**
Wave 3					
Normal BP	81.31	8.13	2.19	8.37	100
HBP	74.42	11.99	2.19	11.41	100
Total	**80.40**	**8.64**	**2.19**	**8.77**	**100**
Aggregate total	**8891**	**1321**	**101**	**248**	**10561**

Notes: The table shows blood pressure (BP) by race in the first three waves of NIDS. The proportions are weighted by post stratification weights provided for each wave. The column totals for each wave indicate the weighted proportion of individuals in each race category while the other values in the table indicate the weighted proportion of specific races who have normal or high blood pressure (HBP). Aggregate total is the actual number of individuals in each race in the wave. Row totals add up to 100% (any difference is due to rounding).

meaningful inferences for them. In general the sample is small in relation to other health outcomes because we exclude individuals who did not have blood pressure systolic and blood pressure diastolic readings in the prescribed thresholds. There were also a large number of respondents who did not have their blood pressure measured.

For wave 1, there is no clear racial difference in normal blood pressure status. The table shows that the proportion of all races who have normal blood pressure is similar to their proportions in the population. Similarly, there is difference in high blood pressure between Africans and Whites. The same situation replicates itself across all the waves of NIDS. The small number of Whites and Indians in the sample indicates that we may not be able to make inference on the basis of these results.

3.6 Panel Regression Results

We have seen in the bivariate analysis that health outcomes vary with socioeconomic factors. However, in these analyses we only focus on two variables and do not consider other factors which may drive the correlation. For example, the positive correlation between health outcomes (SAH and BMI) and educational attainment could be because those with no education are also old, and age is negatively correlated with these health outcomes. In this section we

present regression results for the panel data models where we estimate partial effects. We present panel regression results for three health outcomes and their correlates. These health outcomes include SAH, BMI and depression (mental health). We run a FE and a RE model for each and we compare with the pooled OLS.

SAH has five ordered categories ranging from poor health (the lowest) to excellent health (the highest). The use of ordered probit regression on panel data introduces incidental parameter bias (Greene et al., 2002). In our case therefore, we dichotomize the SAH categories such that poor and fair health categories are indicated with a zero while good, very good and excellent health are indicated with a one. This is in line with other studies that have used panel data estimators to investigate the socioeconomic determinants of health outcomes (Jones and Wildman, 2008). These argue that a binary variable allows the use of a linear probability model and avoids the problem of incidental parameter bias in panel estimations (Greene et al., 2002).

The results of the panel estimators for SAH are presented in table 3.18. This table presents panel estimators for the FE and the RE and also includes the pooled OLS for SAH. The pooled OLS regression results and the RE results are similar in magnitude while the FE results are quite different. These results imply that the fraction[12] of the variance in the error due to idiosyncratic error is very high (that is, idiosyncratic error dominates individual specific effects). This is confirmed by the summary statistics table which shows that the variation between individuals is higher than the variation for the same individuals over time for all the variables. The pooled OLS results show that the probability of reporting better health is positively correlated with the log of per capita monthly income and negatively correlated with age, holding all other factors constant. These results are consistent with those of the bivariate analysis. The panel summary statistics also show that the within-person variation is small relative to the between person variation for all variables. This means that from the onset we expect the RE estimators to be closer to the pooled OLS estimators than to the FE estimators.

Holding all other factors constant, income increases the probability of having better health in South Africa in the pooled OLS model. The education variable used in our analysis is the level of education classified into five categories with no education as the base. The results from pooled OLS show that the higher the education the higher the probability of having better health relative to the base (No education). This is in line with the predictions of Grossman's model which presents education as an efficiency variable in the production function of health (Grossman, 1972, 2000).

[12]RE estimators are estimated from a model of the form $Y_{it} - \lambda\bar{Y}_i = (1-\lambda)\mu + (X_{it} - \hat{\lambda}\bar{X}_i)\beta + (1-\hat{\lambda})\alpha_i + (e_{it} - \hat{\lambda}\bar{e}_i)$ and $\hat{\lambda} = 1 - \frac{\sigma_e}{\sqrt{\sigma_e^2 + \sigma_\alpha^2}}$

There are gender differences evident in SAH in South Africa. Men have a higher probability of reporting better health. There is also significant differences in the probability of reporting better health for the different race groups. In particular, Whites have higher probability of reporting better health relative to Africans. There is no evidence of health differences between Coloureds and Africans, since the coefficient on Coloureds is not statistically significant. The coefficient on Asians is negative and significant relative to Africans, implying that Asians have a lower probability of reporting better health relative to Africans. However, the Asian sample is too small to allow us to make any inference on the basis of this result.

Marriage seems to protect individuals against poor health. Individuals who are married are more likely to report better health compared to those with another marital status. However, there is no statistically significant difference between those who are divorced and those who are married. This could be working indirectly through the coping mechanisms where individuals who are married are on average more likely to cope with the stresses of lack of income because they can complement each other. This is in contrast to those who do not have spouses to offer this assistance. It is also not surprising that there is no significant health differences between those who are divorced and those in marriage because this could imply that the health protection in marriage does not come from the husband or the wife but from a third variable. Also, individuals in marriage develop networks with relatives who may assist to cope with stresses of poverty even after divorce. In some cases the partners share financial responsibilities even when they are divorced and this may reduce stresses that usually lead to deterioration of emotional health.

There is evidence of SAH difference in South Africa based on where people live. Those who live in urban areas and on farms have a lower probability of reporting better health relative to their counterparts in traditional areas. The coefficient of those who live on farms is however not statistically significantly different from those who live in traditional areas.

Individuals who live in the Northern Cape and Free State provinces have a lower probability of reporting good health relative to those in the Western Cape. Individuals who live in Limpopo have a higher probability of reporting better health relative to those who live in the Western Cape. There is no significant difference between individuals in the Western Cape and the remaining provinces.We include provincial dummies to control for province fixed effects. Province dummies capture everything that differs between provinces, other than the variables included in the model, which might affect health. These may be both from the supply side (such as the way health care is structured) and from the demand side, such as behaviours (dominant cultural practices in some provinces that affect health).

For the FE model, time invariant regressors such as race and gender drop out of the regression because they are differenced out. The within-person variation in this data is small

partly because of the short time period between waves. Thus there is no meaningful variation in the variables of interest (i.e. marital status, geographical type of residence, education, and province of residence). The nature of these variables means we do not expect much variation in these for each individual. For this reason, the FE is not preferred because it relies on variation for the same individual over time. We however report it for comparison purposes. In terms of interpretation of the FE estimators, for example the coefficient of income, each extra unit (in log income) above the average is associated with a decrease in probability of reporting better health. This is surprising because we expect income to be positively correlated with SAH. Education dummies too have the opposite sign to our expectations. The FE model results show that individuals with higher levels of education have a lower probability of reporting better health relative to those with no education. Apart from income, age and education, the explanatory variables do not show any significance in the FE model. These include marital status, province and geographical type of residence.

We turn to the next health outcome, depression, which is measured using CES-D scores. We estimate a pooled OLS, a FE model and a RE model for depression and the results are presented in table 3.19. From the pooled OLS regression results, income negatively determines depressive symptoms in South Africa. Individuals with no education had the highest depressive symptoms compared to individuals with higher education. On the other hand age positively determines depressive symptoms.

The other race groups had fewer depressive symptoms relative to Africans, although the Indian sample is quite small. Married individuals had fewer depressive symptoms relative to those with any other marital status. Individuals who live in traditional areas experience fewer depressive symptoms, compared to those who live in urban areas or on farms. Individuals living in the Western Cape had fewer depressive symptoms than individuals living in other provinces, except the Northern Cape, Gauteng, Mpumalanga and Limpopo which have no significant difference in depressive symptoms with individuals living in the Western Cape.

The FE model for depression shows that income positively determines depressive symptoms while age is a negative determinant of depression. The coefficient on income is however smaller in magnitude compared with the pooled OLS and the RE models. Those who were divorced or separated as well as widows and widowers recorded more depressive symptoms than married people. The coefficient for those who were widowed was significant at 10% level of significance. There was no evidence of any correlation between depression and education, geographical type or province of residence in the FE model.

The RE estimators are as close as possible in terms of magnitude and have similar signs to the pooled OLS estimators, as expected, because the variation between individuals for all variables is higher than variation for an individual over time (within). All the significant

Table 3.18: Regression Results for SAH

VARIABLES	POLS	FE	RE
ln real pc monthly income	0.00669***	-0.0145***	0.00632***
	(0.00193)	(0.00318)	(0.00191)
Age in years	-0.00684***	0.0168***	-0.00681***
	(0.000191)	(0.00159)	(0.000190)
Coloureds (rel. to Africans)	0.0149		0.0142
	(0.00962)		(0.00951)
Asians	-0.0502**		-0.0502**
	(0.0239)		(0.0236)
Whites	0.0469***		0.0472***
	(0.0116)		(0.0115)
Living with partner (rel. to married)	-0.0324***	-0.00605	-0.0322***
	(0.00842)	(0.0141)	(0.00832)
Widow/Widower	-0.0554***	0.00748	-0.0542***
	(0.00986)	(0.0204)	(0.00985)
Divorced or separated	-0.0157	-0.00378	-0.0168
	(0.0146)	(0.0239)	(0.0145)
Never married	-0.0352***	0.0176	-0.0339***
	(0.00533)	(0.0133)	(0.00528)
Urban (rel. to Traditional Areas)	-0.0277***	-0.0178	-0.0270***
	(0.00592)	(0.0126)	(0.00585)
Farms	-0.000641	0.0125	-0.000521
	(0.00857)	(0.0208)	(0.00849)
General_educ (rel. to no education)	0.0413***	-0.111***	0.0419***
	(0.00843)	(0.0335)	(0.00841)
Further educ & Training	0.0864***	-0.137***	0.0870***
	(0.00869)	(0.0336)	(0.00865)
Matric	0.111***	-0.157***	0.111***
	(0.00919)	(0.0345)	(0.00916)
Higher educ	0.143***	-0.188***	0.143***
	(0.00983)	(0.0350)	(0.00978)
Eastern Cape (rel. to Western cape)	-0.00386	-0.0414	-0.00465
	(0.0108)	(0.0309)	(0.0107)
Northern Cape	-0.0276**	-0.00978	-0.0275**
	(0.0127)	(0.0651)	(0.0126)
Free State	-0.0363***	-0.0515	-0.0370***
	(0.0125)	(0.0427)	(0.0124)
KwaZulu-Natal	-0.0149	-0.0648	-0.0151
	(0.00990)	(0.0400)	(0.00981)
North West	-0.0130	-0.0259	-0.0133
	(0.0123)	(0.0466)	(0.0122)
Gauteng	0.0118	-0.0436	0.0117
	(0.0101)	(0.0348)	(0.0100)
Mpumalanga	-0.00780	-0.0785*	-0.00774
	(0.0122)	(0.0405)	(0.0121)
Limpopo	0.0262**	-0.0510	0.0254**
	(0.0126)	(0.0367)	(0.0124)
Male	0.0242***		0.0241***
	(0.00308)		(0.00306)
Constant	1.031***	0.480***	1.031***
	(0.0209)	(0.0741)	(0.0207)
Observations	51,486	51,486	51,486
R-squared	0.159	0.012	
Number of pid		26,047	26,047

Notes: The table reports coefficients of panel data estimators for pooled OLS, fixed effects (FE) and random effects (RE) for the first three waves of NIDS. The dependent variable is binary obtained by dichotomizing SAH where 1 indicates better health (good, very good and excellent) and zero worse health (poor and fair health). All models are restricted to adults aged 15 years and above. Standard errors reported in parentheses are robust, they are clustered by pid. *** p<0.01, ** p<0.05, * p<0.1.

coefficients in the pooled OLS were also significant in the RE model. One difference is that the magnitudes for RE estimators are slightly lower than the pooled OLS estimators. The interpretation is therefore similar to the interpretation already discussed. The FE estimators rely on within-person variation, lack of within-person variation makes the FE estimators inefficient. The RE estimators on the other hand have lower standard errors than the FE estimators. If the individual specific error is random the RE estimators would therefore be the most efficient measure.

The panel estimators of BMI for pooled OLS, FE and RE models are presented in table 3.20. The BMI in this regression is used as a continuous score. From the pooled OLS model, the BMI score is positively correlated with income. Ardington and Gasealahwe (2012) and Wittenberg (2013) have shown in earlier work that this relationship is monotonic for Africans. Africans have the highest average BMI compared with other races. Individuals who are married have a higher BMI on average compared to those with any other marital status. Those who live in urban areas have a higher BMI on average compared to those in traditional areas. On the other hand, there is no significant difference between those who live on farms and in traditional areas. Individuals living in Western Cape have a higher BMI than those living in other provinces, except those in Kwa-Zulu Natal.

The BMI increases with age. As noted in the previous chapter, the height of adults is normally fixed over time so any increase in the BMI is attributable to an increase in weight. The analysis also shows that BMI has shifted over time and the increase has been due to an increase in weight. There are also obvious gender differences in BMI in South Africa, with men having lower BMI than women at all ages. The BMI for men is approximately 4 kg/m^2 below that of women across all ages.

There is evidence in South Africa that relative to those with no education, the educated have higher BMI. These results are a bit surprising given that education is considered one of the efficiency variables in the health production function and that the more educated are able to chose the right food inputs. Viewed in another way, this result is not odd at all. The BMI is a risk factor and the marginal rates of utility for the goods, services and activities that increase BMI might be greater than the marginal utility of the expected improvement in health status associated with the efficiency gains in health production. Education can also increase the efficiency of non-health production as well.

For the FE estimators, the results for income and age are consistent with those of the pooled OLS in terms of the sign and significance, that is, income and age are positively correlated with BMI. Individuals with a general education had higher BMI than those with no education. However, those in the other education categories do not exhibit BMI that is significantly different from individuals with no education. There is also no difference in BMI across provinces, except for Free State province where on average BMI is significantly lower

Table 3.19: Regression Results for Depression (CES-D score)

VARIABLES	POLS	FE	RE
ln real pc monthly income	-0.349***	-0.118**	-0.345***
	(0.0334)	(0.0586)	(0.0336)
Age in years	0.0377***	-0.243***	0.0375***
	(0.00211)	(0.0301)	(0.00211)
Coloureds (rel. to Africans)	-1.470***		-1.467***
	(0.201)		(0.200)
Asians	-2.548***		-2.556***
	(0.241)		(0.242)
Whites	-1.683***		-1.677***
	(0.212)		(0.212)
Living with partner (rel. to married)	0.344***	-0.171	0.335***
	(0.0964)	(0.207)	(0.0965)
Widow/Widower	0.720***	0.408*	0.716***
	(0.0956)	(0.229)	(0.0961)
Divorced or separated	1.294***	0.888***	1.297***
	(0.169)	(0.297)	(0.169)
Never married	0.431***	-0.0257	0.421***
	(0.0693)	(0.173)	(0.0690)
Urban (rel. to traditional area)	0.610***	0.128	0.610***
	(0.125)	(0.254)	(0.124)
Farms	0.495***	0.0748	0.504***
	(0.163)	(0.401)	(0.164)
General educ (rel to no education)	-0.359***	-0.399	-0.373***
	(0.0914)	(0.362)	(0.0917)
Further educ & Training	-0.549***	0.0433	-0.560***
	(0.102)	(0.374)	(0.102)
Matric	-0.690***	0.234	-0.700***
	(0.114)	(0.416)	(0.114)
Higher education	-0.920***	0.223	-0.941***
	(0.126)	(0.451)	(0.126)
Eastern Cape (rel. to Western Cape)	0.707***	-0.638	0.704***
	(0.249)	(0.555)	(0.247)
Northern Cape	0.145	-0.979	0.142
	(0.239)	(0.974)	(0.239)
Free State	0.798***	0.538	0.802***
	(0.280)	(0.916)	(0.279)
KwaZulu-Natal	1.273***	0.660	1.269***
	(0.245)	(0.746)	(0.245)
North West	1.160***	0.721	1.157***
	(0.253)	(0.778)	(0.253)
Gauteng	0.318	-0.0144	0.304
	(0.261)	(0.616)	(0.259)
Mpumalanga	-0.395	-0.677	-0.402*
	(0.244)	(0.845)	(0.243)
Limpopo	0.125	0.828	0.120
	(0.284)	(0.744)	(0.284)
Male	-0.394***		-0.394***
	(0.0415)		(0.0415)
Constant	7.807***	17.00***	7.811***
	(0.359)	(1.310)	(0.359)
Observations	50,622	50,622	50,622
R-squared	0.089	0.016	
Number of pid		25,916	25,916

Notes: The table reports coefficients of panel data estimators for pooled OLS, fixed effects (FE) and random effects (RE) for the first three waves of NIDS. The dependent variable a continuous depression score. All models are restricted to adults aged 15 years and above. Standard errors reported in parentheses are robust, they are clustered by pid. *** p<0.01, ** p<0.05, * p<0.1.

than in Western Cape. This confirms that there is not much variation in this measure across the country.

The results for RE are close to pooled OLS as expected also because between-person variation dominates within-person variation. The coefficients that were significant in the pooled OLS are also significant in the RE model.

The panel estimators for blood pressure are presented in table 3.21. As we can see from the table, the estimation sample is smaller than the other health outcomes already discussed. This is because we excluded individuals with implausible blood pressure measurements . We use blood pressure measurement here as a binary variable, as explained in section 3.8. The first results are for a pooled OLS and they show that income is not correlated with high blood pressure in South Africa. This result is consistent with results obtained in studies in China and the USA which did not show any relationship between income and high blood pressure (Strauss et al., 2010).

We find that high blood pressure is positively correlated with age. As we mentioned earlier this is partly explained by human genetics. There is also evidence in the data that shows racial differences in blood pressure incidence. In particular, Whites and Indians had lower blood pressure relative to Africans, although the Indian sample is too small to allow any inference.

Evidence suggests that education protects people from high blood pressure only if this is higher education, relative to no education. In addition, individuals with general education had higher blood pressure than those with no education, albeit at 10 percent level of significance.

Gender differences are evident in the data, suggesting that men have higher blood pressure than women. There is no difference in high blood pressure due to marital status. As we can see also, there are significant coefficients on province of residence. The data indicates that living in Western Cape rather than the other provinces protects one from high blood pressure.

The FE model for high blood pressure, like the pooled OLS, does not show any evidence of a relationship between income and high blood pressure. In addition the FE model does not show any relationship between high blood pressure and age, marital status and geographical type of residence. Individuals living in Kwa-Zulu Natal, Limpopo, and the Eastern Cape have lower blood pressure than those in the Western Cape province. Again the RE estimators replicate the estimators of pooled OLS with minor differences in magnitude.

Our interest in this chapter is to establish the socioeconomic (education and income) and demographic (race, gender, marital status) determinants of health outcomes. We discuss the relationship with education in detail because of the importance attached to this variable in theoretical work (Grossman, 2000). Later in chapter 4 we discuss the endogenous relation

Table 3.20: Regression Results for BMI

VARIABLES	POLS	FE	RE
ln real pc income	0.394***	0.123***	0.286***
	(0.0439)	(0.0425)	(0.0341)
Age in years	0.0704***	0.221***	0.0826***
	(0.00332)	(0.0138)	(0.00291)
Coloureds (rel. to Africans)	-1.080***		-1.080***
	(0.208)		(0.204)
Asians	-2.907***		-2.604***
	(0.496)		(0.495)
Whites	-1.434***		-1.149***
	(0.307)		(0.292)
Living with partner (rel. to married)	-1.753***	-0.258	-1.264***
	(0.154)	(0.159)	(0.122)
Widow/Widower	-1.630***	-0.297*	-1.292***
	(0.149)	(0.166)	(0.125)
Divorced or separated	-1.138***	-0.0973	-0.847***
	(0.236)	(0.236)	(0.188)
Never married	-2.586***	-0.502***	-1.965***
	(0.114)	(0.144)	(0.0969)
Urban (rel. to traditional area)	0.635***	0.530**	0.614***
	(0.130)	(0.217)	(0.117)
Farms	0.0774	0.0293	0.0859
	(0.180)	(0.277)	(0.160)
General_educ (rel to no education)	1.252***	-0.831***	1.110***
	(0.149)	(0.302)	(0.141)
Further_educ & Training	1.999***	-0.276	1.933***
	(0.169)	(0.333)	(0.157)
Matric	2.429***	0.439	2.531***
	(0.168)	(0.358)	(0.157)
Higher_educ	2.845***	0.607	2.959***
	(0.201)	(0.373)	(0.181)
Eastern Cape (rel to Western Cape)	-0.398	-0.439	-0.519**
	(0.247)	(0.524)	(0.235)
Northern Cape	-1.251***	0.182	-1.314***
	(0.257)	(0.667)	(0.255)
Free State	-0.564**	-1.162*	-0.799***
	(0.274)	(0.596)	(0.253)
Kwa-Zulu Natal	0.660***	-0.448	0.407*
	(0.234)	(0.585)	(0.221)
North West	-0.906***	-0.681	-1.120***
	(0.291)	(0.709)	(0.276)
Gauteng	-0.904***	-0.818	-0.927***
	(0.252)	(0.530)	(0.232)
Mpumalanga	-0.501*	-0.540	-0.613**
	(0.282)	(0.589)	(0.266)
Limpopo	-1.204***	-0.189	-1.234***
	(0.267)	(0.608)	(0.251)
Male	-3.985***		-3.825***
	(0.0801)		(0.0774)
Constant	22.69***	17.71***	22.61***
	(0.442)	(0.775)	(0.384)
Observations	46,304	46,304	46,304
R-squared	0.231	0.030	
Number of pid		24,814	24,814

Notes: The table reports coefficients of panel data estimators for pooled OLS, fixed effects (FE) and random effects (RE) for the first three waves of NIDS. The dependent variable is a continuous BMI score. All models are restricted to adults aged 15 years and above. Standard errors reported in parentheses are robust, they are clustered by pid. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.21: Regression Results for Blood Pressure (BP)

VARIABLES	POLS	FE	RE
ln real pc income	0.00113	-0.00383	-0.000177
	(0.00232)	(0.00380)	(0.00217)
Age in years	0.00865***	0.000169	0.00859***
	(0.000248)	(0.00156)	(0.000239)
Coloureds (rel. to Africans)	0.0305**		0.0341**
	(0.0139)		(0.0139)
Asians	-0.0622***		-0.0550***
	(0.0136)		(0.0137)
Whites	-0.0598***		-0.0551***
	(0.0194)		(0.0190)
Living with partner (rel. to married)	-0.00509	0.0153	-0.00426
	(0.0107)	(0.0222)	(0.0105)
Widow/Widower	0.0184	0.0330	0.0147
	(0.0152)	(0.0274)	(0.0140)
Divorced or separated	-0.0120	0.0367	-0.00219
	(0.0197)	(0.0372)	(0.0191)
Never married	-0.00634	0.0158	-0.00828
	(0.00778)	(0.0199)	(0.00756)
Urban (rel. to traditional area)	0.0131*	-0.00670	0.0138**
	(0.00707)	(0.0184)	(0.00701)
Farms	0.0144	0.00275	0.0165*
	(0.00923)	(0.0296)	(0.00932)
General_educ (rel to no education)	0.0233*	-0.0913	0.0178
	(0.0122)	(0.0577)	(0.0120)
Further_educ & Training	-0.00117	-0.0897	-0.00545
	(0.0121)	(0.0587)	(0.0120)
Matric	-0.0108	-0.0987*	-0.0178
	(0.0122)	(0.0592)	(0.0120)
Higher_educ	-0.0347**	-0.0918	-0.0385***
	(0.0146)	(0.0624)	(0.0144)
Eastern Cape (rel to Western Cape)	-0.0295**	-0.108***	-0.0278*
	(0.0145)	(0.0415)	(0.0144)
Northern Cape	-0.0231	-0.0716	-0.0245
	(0.0154)	(0.0694)	(0.0150)
Free State	-0.0375**	-0.131**	-0.0356**
	(0.0170)	(0.0561)	(0.0167)
Kwa-Zulu Natal	-0.0290**	-0.117**	-0.0279**
	(0.0141)	(0.0514)	(0.0142)
North West	-0.0322**	-0.0537	-0.0264*
	(0.0155)	(0.0548)	(0.0152)
Gauteng	-0.0498***	-0.0400	-0.0454***
	(0.0141)	(0.0444)	(0.0142)
Mpumalanga	-0.0705***	-0.0486	-0.0677***
	(0.0148)	(0.0621)	(0.0148)
Limpopo	-0.0711***	-0.0900*	-0.0674***
	(0.0154)	(0.0503)	(0.0155)
Male	0.0231***		0.0197***
	(0.00440)		(0.00439)
Constant	-0.142***	0.320***	-0.125***
	(0.0269)	(0.0813)	(0.0257)
Observations	26,923	26,923	26,923
R-squared	0.188	0.002	
Number of pid		18,490	18,490

Notes: The table reports coefficients of panel data estimators for pooled OLS, fixed effects (FE) and random effects (RE) for the first three waves of NIDS. The dependent variable binary indicated by 1 (high blood pressure) and zero for normal blood pressure. All models are restricted to adults aged 15 years and above. The sample is small because we excluded blood pressure measurements that do not meet the thresholds. Standard errors reported in parentheses are robust, they are clustered by pid. *** p<0.01, ** p<0.05, * p<0.1.

between health and income. In the literature covering the developed world, years of formal schooling completed is the most important determinant of good health (Grossman, 2000; Cutler and Lleras-Muney, 2006). This finding emerges whether health levels are measured by mortality rates, morbidity rates, self evaluation of health status, or physiological indicators of health, and whether units of observation are individuals or groups. Years of schooling is also more important than income as socioeconomic status variable (Grossman, 2000). A causal relationship between education and health is however difficult to establish because of endogeneity issues.

The schooling variable is endogenous, implying that there are three possible ways in which positive correlations observed between health and schooling can be explained. First, there could be causal relationship that runs from an increase in schooling to better health outcomes. Second, the direction of causality could run from better health outcome to more schooling. Third, there is no causal relationship implied by the correlation, instead differences in one or more "third variable" such as physical and mental ability and parental characteristics, affect both health and schooling in the same direction. If there is high correlation between health and schooling, policy-makers may easily be misled to implement a policy of increasing education budget that may not accomplish the goal of improved health.

Theory explains that the transmission mechanism is that more educated individuals are more efficient producers of health (Grossman, 1972, 2000). The efficiency effect can take two forms, productive and allocative. Productive efficiency arises when more educated individuals obtain a larger health output from given amounts of endogenous[13] inputs. Allocative efficiency pertains to a situation in which schooling increases information about the true effects of health inputs e.g. the more educated will have information about the harmful effects of cigarette smoking. Schooling will therefore increase health to the extent that it leads to selection of a better input mix.

The alternative is that the direction of causality runs from better health to more schooling because healthier students may be more efficient producers of additions to their stock of knowledge via formal schooling. In addition, this causal path may have long lasting effects if past health is an input into current health status, i.e even for non students a positive relationship between health and schooling may reflect reverse causality in the absence of controls for past health.

The third explanation for the positive relationship between health outcomes and schooling covers the large unexplained variation in health outcomes even after controlling for schooling and other determinants. Like BMI scores, the explanatory variables explain only 26% of the variation. The third variable explanation is relevant only if the unaccounted factors

[13]Endogenous inputs are those inputs individuals are able to choose

which affect health outcomes are correlated with schooling. In the case of reverse causality, the omitted variable is identified as past or endowed health i.e biases due to unobserved heterogeneity among individuals. The general conclusion, based on a comprehensive review of the literature by Grossman (1999), is that schooling has a causal impact on good health. In arriving at this conclusion Grossman (1999) acknowledge the difficulty of establishing causality using observational data in social sciences where natural experiments are difficult to perform. Our choice of different dimensions of health outcomes (SAH, BMI, Depression, Hypertension) imply that the significant correlations between educational attainment and health outcomes across all the dimensions can not be by chance but may indicate some causal relationships.

These results are consistent with the bi-variate analysis in the previous section where we showed that health depreciates with age and improves with income and education.

3.7 Conclusion

In this chapter we provide detailed, updated evidence on the socioeconomic and demographic determinants of health outcomes in South Africa. We examine the nature of the determination for different dimensions of health outcomes by measures of socioeconomic and demographic factors by using both bivariate analysis and panel econometric methods. We establish in both cases that income, education, race and other demographic factors are significant determinants of health outcomes in South Africa.

We use a rich set of health outcomes that include SAH, BMI, emotional health (depression) and high blood pressure (hypertension). Together with panel data regressions we find that different dimensions of health outcomes are affected differently by these health determinants. For example, income is positively correlated with reporting better health (improved SAH), as well as negatively correlated with having depressive symptoms. It is however positively correlated with increases in BMI which is a source of other health problems such as diabetes and high blood pressure. Income does not significantly influence high blood pressure directly, as we saw in the panel regressions. Such evidence is important in designing policies aimed at reducing unintended[14] health effects.

The statistically significant determinants of health outcome regressions is testimony to socioeconomic related health disparities in South Africa. As we have seen however, one determinant may be correlated to different health dimensions in different directions (positive

[14]By implementing social and economic policies to improve peoples incomes, positive health effects are realized but at the same time other health outcomes are negatively affected such as BMI because people can afford more food with increased incomes.

for some and negative for others). While it is correlated with improvements in one health dimension it may be associated with worsening of others. The challenge therefore remains how to measure health outcomes to reflect the disparities in these outcomes given that the determinants of health affect these health outcomes differently.

Finally, our analysis in this chapter does not allow us to say much about causality due to endogeneity of our socioeconomic variables (income and education). There is simultaneity between the health outcomes and the socioeconomic explanatory variables. The small R squared in our regressions also indicate that some explanatory variables are omitted. As we point out, certain fundamentals of the health care system are the responsibility of the national government, such as training of health care personnel, while other functions are coordinated at the provincial level. This together with the fact that the public sector serves about 80% of the population means that we must control for supply side factors when analyzing determinants of health outcomes, otherwise we introduce omitted variable bias. This bias is however minimized through the inclusion of the province variables in the regressions. The health care quality differences from the supply side are captured by the differences in health care systems of provincial governments.

We move to an analysis of inequality in health outcomes in the next chapter. In order to do so we need to establish stable partial effects that form the basis of the decomposition of socioeconomic related health inequalities.

Chapter 4

Income Related Health Inequality in South Africa

4.1 Introduction

Interest in the measurement of health inequalities has increased globally. This is because most of the health inequalities across social groups or individuals are considered unjust as they reflect an unfair distribution of the underlying socioeconomic determinants of health (for example access to educational opportunities and income). On the other hand some extreme views exist that deny the role of socioeconomic determinants in the creation of health inequalities and rather point to some examples where health inequalities would not normally be considered unjust, such as life stage differences and random genetic mutation (Kawachi et al., 2002). These extreme positions point to the need to clearly understand the sources of health inequalities in order to assist policy-makers and academics alike to solve the problem of inequalities in health outcomes.

To understand inequalities in health outcomes, we build on the findings in the previous chapter which establishes the determinants of health outcomes in South Africa. We find that health outcomes are correlated with several socioeconomic factors. If these factors are unevenly distributed in the population, we expect health outcomes to be unevenly distributed as well. However, although the presence of a significant determinant is important in explaining its contribution to health inequality, it is not the only consideration. The distribution of socioeconomic determinants also plays a role. In this chapter we estimate the extent of income related health inequality which we decompose into factors that contribute to inequality in health outcomes. In doing this we employ the corrected concentration index and the decomposition proposed by Wagstaff (Wagstaff et al., 1991).

There have been new insights regarding the unsuitability of the standard concentration index, which has been the preferred measure of inequality in health outcomes. When this

index is applied to bounded health outcomes such as self assessed health (SAH), it estimates the health outcome inequality index ambiguously. Depending on whether ill health or good health is used, a different index of health outcome inequality is obtained.[1] Erreygers and Van Ourti (2011) recommend the corrected concentration index as an alternative measure.

Measurement of health inequalities is fairly recent in the literature, compared to investigations of income inequality. Early contributors include Le Grand (1978) who estimated inequalities in the distribution of public health expenditure. After Wagstaff introduced the concentration curve and index in 1989 more work on health inequality was undertaken, particularly in the developed world. The literature has evolved with much attention focused on measuring socioeconomic related health inequality, rather than total inequality, because of the belief that these socioeconomic factors are responsible for unjust inequalities in health (Van Doorslaer 2012).

Interest in health inequality also arises from the notion that poverty is inadequately captured by money metrics, such as income or expenditure. Following the capability theory of Sen, poor health is an intrinsically important deprivation of basic capabilities (Sahn and Younger, 2009). Therefore health is a more direct measure of capability deprivation than income. Better health is one of the foundations for escaping poverty traps and attaining positive socioeconomic trajectories.

Sahn and Younger (2009) have shown that in developing countries, inequalities in health outcomes are associated with increases in poor health outcomes for people at the lower end of the income distribution. An increase in poor health reduces the time available for production among the poor (both at individual and household levels). This effect can be direct when it reduces the time available for production, or indirect when productive time is spend caring for the sick within the household. The effects of poor health can also be long lasting, irreversible and costly especially when illness is chronic or results in disability. Poor health affects cognitive skills of children and into adulthood (Duflo, 2000). Despite these negative effects of poor health, some health inequalities are unjust and preventable (Marmot and Bell, 2011).

The concentration index has dominated as a tool used in the measurement of socioeconomic related health inequalities. It is a rank dependent measure adapted to bivariate analysis. Inherent in rank dependent measures is the bivariate distribution of health and another socioeconomic variable used to rank individuals or households (Van Doorslaer 2012). The most common ranking variable used in bivariate analysis is income. However, a few studies have also used education and consumption expenditure (Cutler and Lleras-Muney, 2006; O'Donnell et al., 2008). The most important property required for a ranking variable

[1]This property is called the mirror property and is violated by standard concentration index.

is ordinal scaling, in order to rank individuals or groups.

We analyze inequalities in health outcomes in relation to the underlying socioeconomic determinants of health but also control for demographic variables. We have shown that some of the socioeconomic and demographic factors determine health outcomes. These factors are unevenly distributed in the population. It is plausible to assume that these factors also drive inequality in health outcomes. In this chapter we proceed to identify which of these factors drive inequality in health outcomes in South Africa.

South Africa presents an appropriate setting for this study because of its high levels of inequality in socioeconomic determinants of health. Some of these inequalities, such as income inequality, is traced to the apartheid legacy, which was characterized by policies that exacerbated inequalities in access to resources as well as opportunities in education and employment. These have persisted despite policies aimed at addressing them. This means that without any remedial health interventions the level of income related health inequality, which is our focus, is also expected to be high (McIntyre and J.E., 2012). Very little is known about the relationship between these pervasive inequalities in socioeconomic determinants and inequalities in health outcomes in South Africa. This makes it interesting to investigate the relative association (contribution) of these factors to income related health inequalities.

Our objective in this chapter is to estimate an index of income related health inequality using SAH as the measure of health outcome. SAH is a non-specific measure of health that is probably the most feasible and inclusive of health status (Jylhä, 2009). It is widely accepted as a strong predictor of morbidity and mortality (Idler and Benyamini, 1997; Jylhä, 2009). The comprehensive nature of the SAH question allows it to capture elements of health (multidimensional) that more guided questions, specific to a particular illness, cannot (Au and Johnston, 2013).

SAH is collected in ordinal scale which poses a challenge in using it directly to estimate inequality as the concentration index requires a cardinal or ratio scale variable. To solve this problem, we use the predicted values of the interval regression to convert it to an interval scale (Van Doorslaer and Jones, 2003; Ziebarth, 2010). This technique has the added advantage, apart from scaling the ordered SAH, of presenting health outcome on the same scale as health utility[2]. This makes the scaled SAH outcome directly comparable to a true measure of health utility which is often used to measure quality of life in epidemiological studies (Ziebarth, 2010). In addition, unlike the dichotomized SAH, scaling of SAH into a continuous variable allows any analysis at different levels of the distribution. Other scaling techniques, such as the use of predictions from probit regression, require normalization before they are used in the concentration index and are therefore not appropriate for our purpose.

[2]Health utility is a unit-less measure of health related quality of life that ranges from a minimum of 0 (death) to a maximum of 1 (perfect health)

We apply the corrected concentration index to the scaled health outcome measure to estimate income related health inequality. This study is among the first to use the corrected concentration index rather than the standard concentration index to compute income related health inequality in South Africa. We contribute to the literature in several other ways. First, we use a nationally representative longitudinal survey to estimate income related health inequality in each wave of the panel survey, and generate an index of income related inequality in health outcomes. Second, we decompose the index of income related health inequality into factors associated with (that contribute to) the index. Third, we make use of the panel nature of our data to explore the factors associated with the change in the index across the waves, using the Oaxaca type decomposition.

Even in the presence of unequal distribution of the socioeconomic determinants of health, it is possible to have fewer inequalities in health outcomes if there are interventions to address them. Since the transition to democracy in 1994, the South African government has undertaken steps aimed at redressing health inequalities particularly for the highly vulnerable groups. These include entrenching health provision in the constitution[3], recruitment of foreign doctors to work in under-served areas, introduction of free health care at all public health facilities for pregnant women and children under six years old (Harrison, 2009) and McIntyre et al., 2003), and currently, the roll out of National Health Insurance (NHI) (NDoH, 2013). Unfortunately, these interventions have concentrated on supply side factors. However, there are huge differences in health care demand patterns in South Africa that may perpetuate inequalities in health outcomes that are correlated with socioeconomic factors (Skordis-Worrall et al., 2011).

Despite government intervention to redress health inequalities, studies have reported persistent health inequalities in South Africa (Ataguba, 2013; Harling et al., 2008; Zere and McIntyre, 2003; Case, 2004; Bradshaw, 2008). Our own exploratory analysis of data from the three waves of NIDS indicates that there are variations in health outcomes. This is the case whether subjective or objective indicators of health are analyzed. The NIDS data shows that about 12% of the sample report poor and fair health while 88% report good health, very good, and excellent health across waves 2 and 3. For wave 1 the figures were 22% and 78% respectively (see table 3.4 in chapter 3). What is not explicit from the literature on inequality in health outcomes in South Africa is that inequality in health is not quantified. Through an index of health inequality we summarize the extent of income related health inequality and this is decomposed to establish factors that contribute to this inequality.

[3]This has the implication of giving health expenditure vote preference over other expenditure votes in the event of competing use for resources thereby ensuring prioritization of health provision.

4.2 Literature Review

4.2.1 Theoretical foundations

In this section we highlight theories that are advanced to explain the sources of health inequalities as well as review some of the empirical evidence. Four theories have been advanced in the literature since the Black[4] Report in 1980 that sought to explain sources of health inequalities. These include the artefact theory, the selection theory, the cultural and behavioural theory and the material/structural theory (McCartney et al., 2013b). The artefact theory suggests that the association between markers of social status and health outcomes is explained by statistical artefacts relating to the way in which social status has been classified over time. This theory has been criticized heavily on the grounds that even where different measures of social status are used (such as income, education, and employment status) inequalities in health outcomes are ubiquitous, making it difficult to believe that health outcomes are unrelated to social status.

The selection theory proposes that health selection might explain inequalities in health outcomes. This theory is essentially a reverse causation one where poor health causes social selection which leads to the observed association between ill health and low social status. The selection can be direct or indirect. Indirect selection occurs when factors in early life (genetics, infections, ante/post natal care) set the foundation and cause one to be in a given social position. These early life factors also influence health related behaviour that determine health (Mel Bartley 2005). In order to test this theory however, we would require longitudinal data that measures pre-morbid social status, and test for an association with subsequent morbidity. In our case we use data that spans only four years therefore it is not feasible to use this theory to explain health inequalities in this study.

The third theory advanced to explain health inequality is the cultural and behavioural theory which suggests that differences in behaviour such as smoking, alcohol consumption, diet, physical activity or differences in the dominant cultures between groups explain the source of inequalities in health outcomes. For health behaviour to cause health inequalities however, socioeconomic factors have to be effect modifiers in the relationship. This theory has been criticized on two grounds. One is the finding that when comparing mortality rates of two groups with the same risk exposure but different socioeconomic status, the group with the lower socioeconomic status has higher mortality rates (Smith et al. 1998). Secondly, a simple focus on behaviour ignores how and why individuals in a particular social group adopt a particular unhealthy behaviour (Nettle D. 2009). Some evidence in the literature has shown that the patterning of adult health behaviour can be explained by their earlier exposure to

[4]Sir Douglas Black chaired a Research Working Group on Inequalities in Health in England and published the findings in a report that is famously called Black Report of 1980.

socioeconomically determined situations (Lynch et al 1997).

Another theory that explains the sources of health inequalities is the material or structural theory. This proposes that differences in socioeconomic circumstances of social groups (including differences in income, wealth, power, environment and access) at all stages of the life course, are associated with differences in the health outcome. According to structural theorists the other three theories are subordinate, that is, behaviour and cultures may theorise potential mechanisms linking structural determinants and health outcomes but do not identify the "causal roots" of health inequalities. This theory has further been supported in the literature which has shown that people with the most resources within any society are always the healthiest, regardless of their behaviour (CSDH, 2008; Wilkinson and Pickett, 2011). In this study we adopt the material or structuralist theory because it provides the dominant frame for analysis in most empirical research (McCartney et al., 2013b).

4.2.2 Empirical literature

Empirical studies assume that health inequalities emanate from differences in socioeconomic determinants of health (income, wealth, education, and demographic characteristics). These studies therefore assume that material or structural theory is the source of inequalities in health outcomes. They therefore first identify the health determinants and evaluate their contribution to health inequality. Many Sub-Saharan African countries have an uneven distribution of socioeconomic factors which translates to uneven distribution of health outcomes. South Africa is one such country which has very high levels of inequalities in income and other socioeconomic factors such educational attainment.

In a study in Uganda, Ssewanyana S. (2012) uses determinants of child nutrition to explain health inequalities. Similar approaches in South Africa include Zere and McIntyre (2003), Charasse-Pouélé and Fournier (2006), Ataguba et al. (2011) and Ataguba (2013). Zere and McIntyre (2003) for example use determinants of self assessed illness and health care to analyze health inequalities in South Africa. They find that the poor reported themselves sick more frequently. Charasse-Pouélé and Fournier (2006) establishes health inequalities based on racial groups while Booysen (2003) establishes health inequalities based on urban rural divide in health care access. Ataguba et al. (2011) and Ataguba (2013) use the standard concentration index to estimate health inequality in self reported acute and chronic illnesses as well as multimorbidity. Inequalities in health outcomes vary in these studies because of differences in the time of study, methods used, and the indicators of health and socioeconomic measures applied.

4.3 Method of Analysis

4.3.1 Data and variable definitions

This study uses the NIDS dataset which we have already described in chapter two. In this chapter we use SAH as our primary variable to compute income related health inequality. SAH is measured as an ordinal scale and the corrected concentration index cannot use ordinal scale. In the following section we describe how we measure the variables used in this chapter.

4.3.2 Measurement of health

The WHO defines health as the state of complete physical, mental and social well-being and not just the absence of disease or infirmity (WHO, 2010). Thus health is a multidimensional phenomenon and this poses a challenge because it is unlikely that we can find an overall measure which collapses the separate dimensions into one construct. Several index-scoring algorithms have been developed to measure different health profiles, such as SF-36, HUI[5], and Euroqol-5D. Although these are the ideal measures to use, their availability is limited because they are collected only in health interview surveys which are rare. They are not in NIDS.

SAH is perhaps the closest variable that captures all health dimensions. The idea is that when individuals evaluate their health, they consider all the dimensions before they report themselves in a given health category. For this reason we use SAH to measure health outcome in this study. SAH has also been supported by international literature because of its power to predict mortality and health care utilization (Idler and Benyamini, 1997). The former ability has also been confirmed by Ardington (2009) using our dataset.

We used SAH as a dichotomized variable in the previous chapter, we now utilize all its categories in this chapter. This inconsistency does not affect our income related inequality of health estimates because the use of SAH as dichotomized variable in the previous chapter was to avoid incidental parameters bias.

The SAH variable is collected as an ordinal scale variable, but underlying it is the assumption of a latent health variable that informs the observed responses. We denote the latent health variable by h^* and assume that the SAH categories and the latent health variable (h^*) are related as follows:

[5]We use health utilities here as is applied in health economics. According to Tolley (2009), "In health economics, utilities are cardinal values that reflect an individual's preferences for different health outcomes. They are measured on an interval scale with zero reflecting states of health equivalent to death and one reflecting perfect health. This concept is different from utility as we know it from consumer theory. In the literature there are three generic approaches of valuing health as mentioned in the text above. We refer to values obtained from HUI in this thesis as health utility. Viewed this way health utility is equivalent to health outcome.

$$SAH = i \quad if\ \alpha_{i-1} < h^* \leq \alpha_i,\ for\ i = 1, 2, 3, 4, 5$$
$$Where \quad \alpha_1 = -\infty, \quad and \quad \alpha_5 = +\infty \tag{4.1}$$

SAH is self assessed health, α_i are the thresholds for the health categories. These thresholds are unobservable but have been estimated in various ways in the literature as we will see shortly.

The index of inequality that we use to estimate income related health inequality can not be applied directly to an ordinal scale variable such as SAH. Many previous researchers have dichotomized or converted this to a cardinal scale before using this measure to estimate income related health inequality (Van Doorslaer and Jones, 2003, 2004; Van Doorslaer et al., 2004). Unfortunately, dichotomization leads to loss of information between categories that are merged. In addition the choice of cut offs for dichotomization is normally arbitrary and influences the levels of inequality obtained. When different countries or regions are compared dichotomizing SAH categories can lead to rank reversals. Wagstaff and Van Doorslaer (1994) and Van Doorslaer and Jones (2004) have demonstrated rank reversals when comparing health inequalities in the Netherlands and countries in Europe using different dichotomization. We therefore avoid dichotomising the SAH.

In the literature, a number of approaches have been proposed to convert variables captured in ordinal scale such as SAH to cardinal or ratio scales (Van Doorslaer and Jones, 2003; O'Donnell et al., 2008). These include the use of predicted values from ordinary least squares (OLS) regression, predicted (linear predictions) values from probit or logit regression applied on SAH after dichotomising it to a binary variable, predictions from ordered probit or logit regression, and predictions from interval regression (Wagstaff and Van Doorslaer, 1994; Van Doorslaer and Jones, 2003).

It has been shown that predictions from an ordered probit approach overestimate inequality (Van Doorslaer and Jones, 2003). The ordered probit predictions also require ex post re-scaling because they are not generated on the same scale as health utility (which is normally bounded between 0 and 1) and hence cannot be used directly as quality weights or utility proxies (Cutler et al., 1997; Van Doorslaer and Jones, 2003). In our case we cannot use predicted values of OLS to convert the SAH to cardinal scale because the SAH variable was collected as an ordered variable with only five categories. We therefore use predicted values from interval regression. This has additional advantages, as this approach predicts health outcomes more accurately than predictions from other models (Van Doorslaer and Jones, 2003). More importantly, predictions from interval regression are generated on the

same scale as health utility[6] making them directly comparable with the health utility index (HUI) which is a true measure of health outcome normally used in health related quality of life studies to measure health states.

The use of interval regression to scale SAH to cardinal scale entails regressing the SAH categories on the covariates of health to predict health outcome and the health outcome variable obtained is therefore conditional on the covariates (Van Doorslaer and Jones, 2003).

To implement interval regression, we use thresholds of SAH categories. As mentioned earlier, these are not observed so we estimate them using external information. We use the Canadian health utility index (HUI) to set these thresholds for SAH categories. This is because there is no health survey within South Africa which has simultaneously collected SAH and a true health indicator such as HUI or short form 36 (SF-36). The Canadian HUI has been widely applied with European data and Chinese data to estimate thresholds for SAH (Van Doorslaer and O'Donnell, 2008).

We use HUI to set thresholds for SAH because there is a stable mapping from the HUI to the latent variable that determines reported SAH for all individuals (Van Doorslaer and Jones, 2003). The implication is that an individual's rank according to HUI corresponds to his rank according to SAH. This is so because SAH is generated in the same scale as HUI, after fixing the lowest and highest bounds of SAH corresponding to those of HUI. This also makes the quantiles in the HUI and SAH correspond to each other, such that the q-th quantile of the distribution of HUI corresponds to the q-th quantile of the distribution of SAH. As mentioned earlier, the lowest bound of HUI is zero which we assign to the lower bound of poor health category and the highest bound of 1 to the higher bound of excellent health. The rest of the thresholds obtained from the Canadian HUI[7] we assign to corresponding thresholds of SAH. We then estimate an interval regression model of SAH categories on the covariates of health outcomes and predict the linear index values as proxy for utility or Quality Adjusted Life Years (QALY) which we use as a continuous indicator of health outcomes. The covariates we include are informed by theory and previous empirical studies, as discussed in previous chapter.

The interval regression model is linear and this is important because the concentration indices (both standard and corrected concentration indices) calculated using the predictions from a linear model are suitable for decomposition analysis (Van Doorslaer et al., 2004).

[6]Health utilities are cardinal values that reflect an individual's preferences for different health outcomes. They are measured on an interval scale with zero reflecting states of health equivalent to death and one reflecting perfect health. These utilities can be based on direct patient experience although clinicians and other health professionals are sometimes used as proxies backed by evidence in the literature (Tolley, 2009).

[7]The Canadian HUI thresholds that correspond to the middle thresholds for SAH are 0.428, 0.756, 0.897 and 0.947. These, together with 0 and 1 form the thresholds of the SAH categories for interval regression purposes.

In effect, the interval regression technique exploits the between SAH category variation to generate some within SAH category variation in HUI, while HUI itself is unobserved.

4.3.3 Measuring inequality in health

After measurement of health outcome, we now turn to measurement of inequality in the health outcome. Health inequalities have traditionally been measured by rank dependent measures such as the concentration index. This index is bivariate in that it measures inequality in health outcomes conditional on some socioeconomic variable such as income or education (Cutler et al., 1997; Cutler and Lleras-Muney, 2006). If we measure inequalities in health per se we would compute a health Gini index. However, because we compute inequalities in health conditional on some other socioeconomic variable (income in our case), the concentration curve and the related concentration index (CI) are the preferred tools in the computation of income related health inequalities (Fleurbaey and Schokkaert, 2009).

While the standard CI is common and also popular, new insights have shown that it is not recommended for measuring inequalities for cardinal health indicators that are bounded, such as SAH or HUI (Erreygers and Van Ourti, 2011). The health outcome we obtain using predicted values of the interval regression is bounded between 0 and 1 because it is estimated on the same scale as HUI. If the CI is used to measure income related health inequality using bounded health outcomes, it violates some desirable properties of a rank-dependent inequality index, namely, the mirror and scale invariance properties (Erreygers and Van Ourti, 2011).

The mirror property requires that for a bounded health outcome like SAH, if h is the good health distribution and s is its associated ill health distribution, then applying the inequality index should yield consistent results in each case i.e. $I(h) = -I(s)$ where I(.) is the inequality index. Unfortunately, the concentration index of health returns a different index from that of ill health.[8] In addition the CI is only meaningful for health variables measured on a ratio scale since the index must remain unchanged under a positive proportional transformation of the health variable. For cardinal variables the value of the CI depends upon the chosen cardinalization since it is not invariant to positive linear transformations (Van Doorslaer and Jones, 2004; Erreygers, 2009a,b; Erreygers and Van Ourti, 2011). Scale invariance property on the other hand requires that the index is invariant to positive affine (linear) transformation.

The bounded nature of the health outcome (SAH) complicates the comparison of health inequalities in the population over time if we apply CI. This is because the bounds of the CI depend upon the lower and upper limit and the mean value of the health outcome[9] in the population (Van Doorslaer and Van Ourti 2012). This means that we cannot compare CI

[8]This is the context in which we argue that the CI is not consistent in ranking socioeconomic related health inequalities when applied to bounded health variables because it depends on whether ill health or health is used.

[9]I.e $-B \leq CI(h) \leq B$ with $B = [(h^{max} - \bar{h})(\bar{h} - h^{min})]/[\bar{h}(h^{max} - h^{min})$

for two populations unless they have equal average population health, as the index will have different bounds.

In light of the above shortcomings, Wagstaff and Van Doorslaer make adjustments on the CI to get the extended concentration index (or Wagstaff index) which unfortunately does not satisfy all the properties (Clarke et al., 2002). Erreygers (2009a) shows that the class of inequality indices that satisfies these two properties (mirror and scale invariance) is given by the general formula:

$$I^{\theta}(h) = \frac{8}{n^2\left[4\mu(h)(1-\mu(h))\right]} \sum_{i=1}^{n} z_i h_i \tag{4.2}$$

Where $I^{\theta}(.)$ is inequality index, θ is inequality aversion parameter, h_i is the health outcome, $\mu(h)$ is the mean of the health outcome and z_i is the fractional rank of the individual. This equation reduces to the Wagstaff index (extended concentration index) if $\theta = 1$ and to Erreygers index (corrected concentration index) if $\theta = 0$.

In addition to the two properties already mentioned (mirror and scale invariance), Erreygers (2009a) and Erreygers and Van Ourti (2011) also impose a convergence property on their index.[10] This condition precludes the suitability of the Wagstaff index from the family of inequality indices shown by equation 4.2 because the convergence condition can only hold if $\theta < 0$.

Following the above explanation, this study adopts the Erreygers index, also called the corrected concentration (CC) index because it satisfies all the desirable properties explained above. This index is given by the formula:

$$CC = \frac{8}{n^2(b_x - a_x)} \sum_{i=1}^{n} z_i x_i \tag{4.3}$$

Where x_i is an indicator of either good health or ill-health, b_x is the upper bound of the health indicator, a_x is the lower bound of the health indicator, z_i is the socioeconomic fractional rank of the individual (income rank in our case). In our case the upper and lower bounds are 1 and 0 respectively which simplifies the above formula even further.

4.3.4 The Factor Decomposition Method

In this section we describe how we determine the contributing factors of the CC index. The CC index, like the CI index, is useful in comparing income related health inequality over

[10]This property requires that $lim_{r\to 0} I(rh) = 0$. This condition examines what happens when a given unequal distribution is gradually reduced to perfectly equal distribution i.e all individual levels reduced to zero by means of a proportional reduction r. When all individuals have zero health, the distribution is equal (even) and so the value of the index should tend to zero.

time or across countries but on its own, it does not tell us much about the factors associated with these inequalities until we decompose it (Van Doorslaer and van Ourti 2010).

The technique to decompose the CC index into contributing factors is proposed by Wagstaff et al. (2003). The technique is meant for decomposing the standard concentration index (CI) but we adapt it to decomposition of the CC index. As we have explained before the CC index is a modification of the standard CI. The modification is aimed at obtaining an index that satisfies the desired properties of a rank dependent index.

In order to decompose the income related health inequality index into contributing factors, we first establish the covariates of the health outcome, as discussed in chapter three. The decomposition of income related health inequality relies on the assumption that the underlying health demand/production function is linear. We therefore re-estimate the health demand model using a linear function. To do this we use the parameter estimates of the interval regression model used to estimate conditional health outcome. The intuition is that ceteris paribus, if there are disparities in covariates of health, this is reflected in disparities in health outcome.

We also evaluate the dynamics of income related health inequality. The changes in income related health inequalities are explained by heterogeneous responses of health to the covariates, changes in inequality of covariates, as well as interaction between the two (Jann, 2008). Decomposition of income related health inequalities is important because it uncovers the contribution of each factor in CC. It is possible that overall health inequality may be negligible because the contributing factors have opposing effects. In addition decomposition of the CC is useful in addressing why there are differences in income related health inequality in the population at a given time and changes over time (Sahn and Younger, 2009).

To implement the decomposition we express the health outcome variable (SAH) as a linear function of socioeconomic, demographic and geographical variables as follows:

$$h_i = \alpha + \sum x_{ki}\beta_k + v_i \tag{4.4}$$

Where h_i is the health outcome for individual i in this case the intervals of SAH categories, x_{ki} is a vector of covariates of health (socioeconomic, demographic and geographical variables) as in chapter three, α is the intercept, β_k is a vector of slope coefficients and v_i is the error term. Following Wagstaff et al. (2003), the standard health concentration index (CI) for h_i is written as:

$$CI(h) = \sum_k \frac{\beta_{xk}\bar{x_k}}{\mu} CI_{xi} + \frac{GV_{\varepsilon k}}{\mu} \tag{4.5}$$

Where $CI(h)$ is the standard concentration index, β_{xk} the coefficient of x_{ki} and $\bar{x_k}$ the mean of x_{ki}. Since the corrected concentration has a close relationship with the standard concentration index, the decomposition of the two indices is identical in proportionate terms (Van Doorslaer and Jones, 2003). Consequently, given the linear health function in equation 4.4, the decomposition of the corrected concentration index is the weighted sum of the concentration index for each covariate with the weights being the partial effects. This is written as:

$$CC(h) = \sum_k \beta_{xi} CC_{xi} \tag{4.6}$$

Where $CC(h)$ is the health corrected concentration index, CC_{xi} is corrected concentration index for variable x_i, and β_{xi} is coefficient of variable x_i. This analysis implies that a factor can only contribute to inequality if it satisfies two conditions, first it should be associated with health, as indicated by significant coefficients in the linear regression and secondly the factor should be unequally distributed across socioeconomic status (Van Doorslaer and Van Ourti 2010). This approach therefore allows identification of the importance of each of these two components within each factor's total contribution (Van Doorslaer and Jones, 2003).

We extend the decomposition approach to explain sources of differences in the income related health inequality over time as in the Oaxaca (1973) and Blinder (1973) as follows:

$$CC^{w3} - CC^{w1} = \sum_k \beta_{xi}^{w1}[CC_{xi}^{w3} - CC_{xi}^{w1}] + \sum_k CC_{xi}^{w1}[\beta_{xi}^{w3} - \beta_{xi}^{w1}] \tag{4.7}$$

Where CC^{w3} is corrected concentration index in wave 3, CC^{w1} is corrected concentration index in wave 1. β_{xi}^{w3} are coefficients of the health covariates in wave 3, β_{xi}^{w1} are coefficients in wave 1. CC_{xi}^{w1} and CC_{xi}^{w3} are the corrected concentration index for variable x_i in wave 1 and wave 3 respectively. The left hand side of the equation 4.7 is the change in health concentration index between wave 1 and wave 3. The first term on the right hand side is the sum of the change in the CC index attributed to the changes in the distribution of the covariates of health weighted by their partial effects while the second term is the change attributed to heterogeneous responses to covariates (change in coefficients) weighted by the distribution of covariates at the base wave.

While the factor decomposition is enlightening in identifying contributing factors to the health inequality index, we make two caveats. First, decomposition of the index holds only if equation 4.4 is linear. Second, the factor decomposition is descriptive and cannot be interpreted in terms of causal relationships unless equation 4.4 has a causal interpretation (Van Doorslaer and Van Ourti 2010).

4.4 Results

4.4.1 Interval regression results

We estimate conditional mean health of the population from linear index predictions of interval regression, as explained earlier in section 4.3.2. These predictions have the same scale as HUI which is unit-less and ranges from 0 to 1 with a score of 0 indicating the worst possible health state (death) and 1 the best possible health state (Tolley, 2009). The predictions are conditional on a set of covariates. The results show that conditional mean health was 0.83 in wave 1 and 0.88 and 0.87 in waves 2 and 3 respectively. These results coincide with the good SAH category in the HUI.[11] In other words, on average, South Africans have good self assessed health.

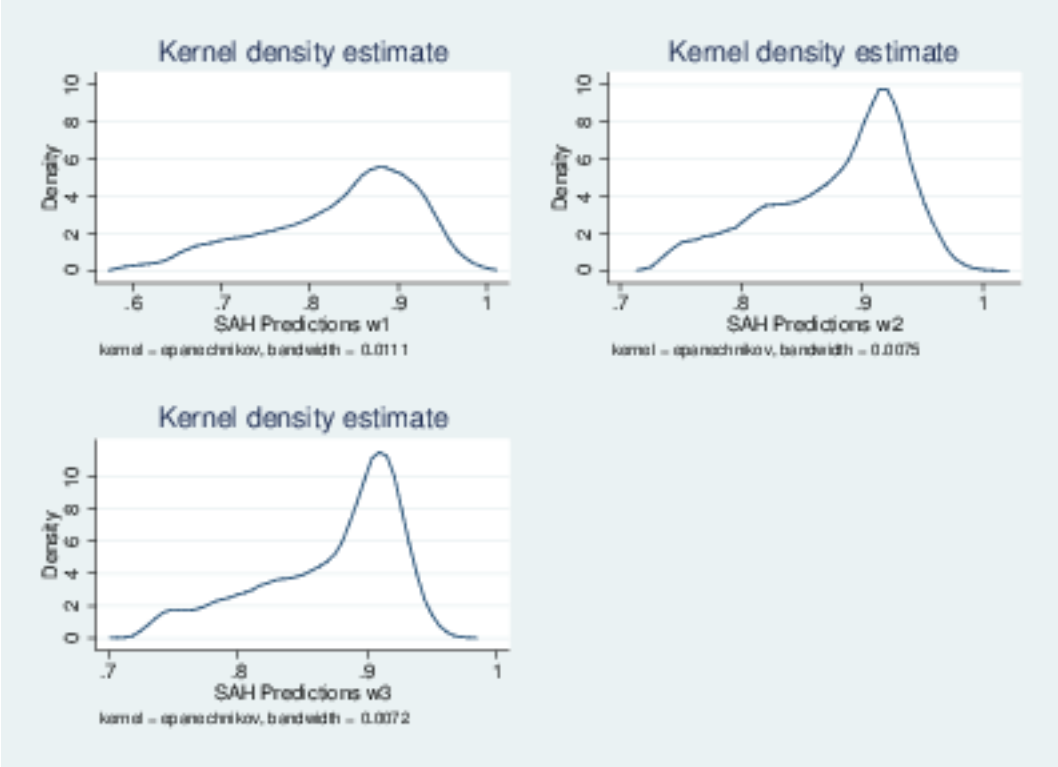
The distribution of the conditional health outcome is presented in figure 4.1. This figure plots the density (frequency) of SAH on the vertical axis and conditional predictions of SAH from an interval regression on the horizontal axis. As we can see the kernel density plots of the conditional health outcomes in each wave are skewed to the left, implying that the majority of South Africans enjoy good health rather than poor health. This is consistent with the histograms we obtain in chapter two from the categories of SAH that show disproportionately more people reporting good to excellent SAH relative to poor and fair SAH. The conditional health outcome variable (SAH) is bounded between 0 and 1 with a similar scale as that of HUI since it is generated on the scale of HUI.

The interval regression results are presented in table 4.1. Unlike in chapter three where we use panel data estimators, these results are separate for each wave. This is because we use the predicted values from these regressions to scale SAH so that we estimate inequality in health for each wave. We therefore discuss these again below. From the results, income is positively correlated with higher values of health utility (better health). The relationship is stable across waves (approximately a significant coefficient of 0.005 across all waves) with the exception of wave 2 where the coefficient is 0.003 but not significant. For wave 1 these results mean that a 1% increase in income is associated with a 0.005 increase in health utility. Put differently, doubling incomes is associated with a 0.5 increase on the health utility index.

In wave 1, respondents with general education (attained some primary school education) have better health than those with no education. The differences in health increases progressively with increasing levels of education relative to no education. This trend continues in subsequent waves, however there is no significant difference in health outcomes between those with general education and those with no education in wave 2 and only a difference at

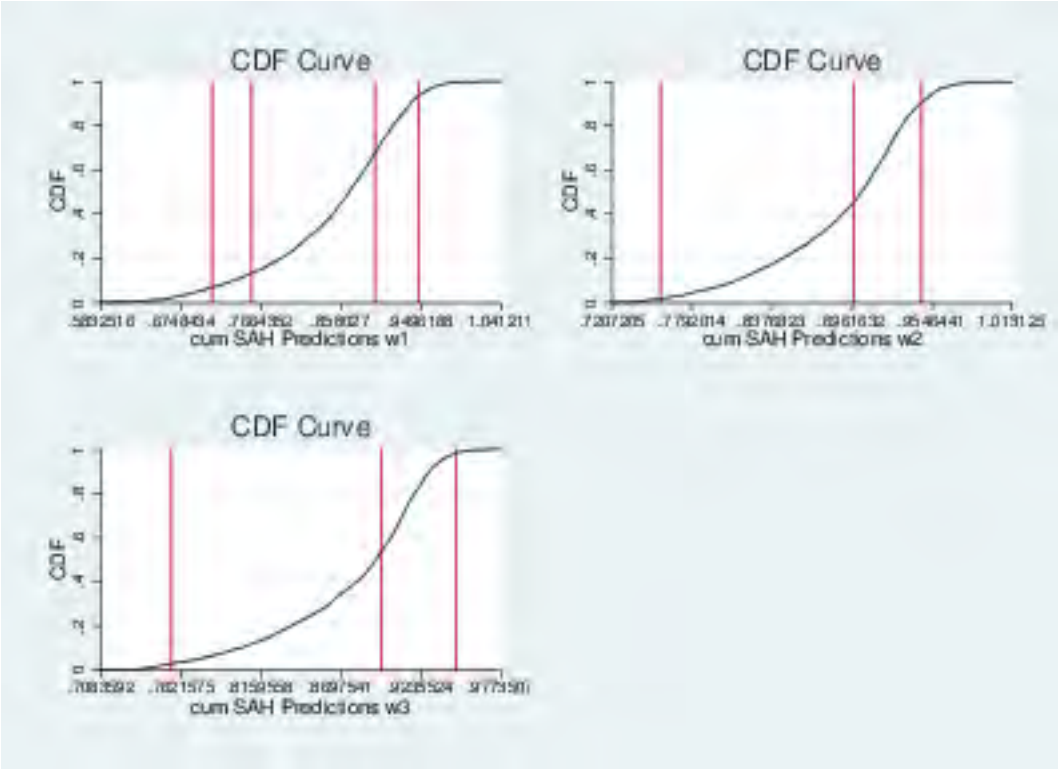
[11]The HUI thresholds are 0-0.428 Poor, 0.428-0.756 Fair, 0.756-0.897 Good, 0.897-0.947 Very Good and 0.947-1 excellent (Van Doorslaer and Jones, 2003).

Figure 4.1: Kernel density of SAH health predictions from interval regression



Notes: Kernel density plots for linear predictions of health from interval regressions for waves 1-3. The density plots are skewed to the left implying that the majority of people have good health relative to poor health.

Figure 4.2: Cumulative density plots for conditional SAH



Notes: The figure shows cumulative densities for linear predictions of health from interval regressions for waves 1-3. The vertical lines indicate the thresholds for the health categories according to Health utility that corresponds to SAH. The CDFs show that no one falls in the lowest category of health utility. The convex shape of the the CDF implies that the majority of the individuals have good health.

10% level of significance in wave 3. In general, these results are consistent with the results of chapter 3 although in this chapter we analyze each wave separately. In addition, the results are consistent with predictions of the theory of health demand by Grossman which postulates that the more educated are more efficient producers of health. The efficiency effect can take two forms, productive and allocative efficiency. Productive efficiency arises when the more educated obtain a larger health output from given amounts of endogenous inputs. Allocative efficiency on the other hand pertains to a situation in which education increases information about the true effects of health inputs. For example the more educated will on average have accurate information about the harmful effects of what they consume.

Health outcome depreciates over the life cycle, and the young have better health . 25-34 year olds have lower health status in terms of health utility than 15-24 year olds. The differences (decreases) in health status proxied by health utilities progressively increase with increasing age. The pattern is the same across all three waves. However the differences from one age category to another reduces with age. This confirms non linearity of age in explaining health outcomes. We find similar results when we use an ordered probit in the previous chapter.

There is evidence of racial differences in health outcomes in wave 1 and 2 where Whites have a higher health utility score than Africans. This difference is significant at 5% in wave 1 and 10% level of significance in wave 2. There is no statistically significant health outcome difference between Coloureds and Asians relative to Africans in wave 1 and wave 2. However in wave 3, Coloureds have higher health utility score while Asians report lower health utility score than Africans. This is significant at 5% and 10% level respectively.

Widows and widowers are less likely to report a high health utility score than married people. However there is no statistically significant differential in health outcomes between divorcees and married individuals. Those who had never married at the time of the survey reported lower health utility scores than those who were married. There was however no evidence of differential health outcomes between different marital status in wave 2 and only those never married reported lower health outcomes relative to married in wave 3.

In wave 1 respondents in Western Cape reported better health utility scores relative to all the other provinces except Eastern Cape and Limpopo, where there was no statistically significant difference. However these results were not stable across waves. Respondents in Gauteng and Limpopo provinces reported statistically significant better health utility scores than those in the Western Cape province in wave 2. In wave 3, respondents in the Western Cape province reported poorer health than those of the other provinces. Living in different area types (traditional areas, urban areas or on farms) also resulted in health differences. Those in urban areas reported poorer health outcomes than those in traditional areas in

waves 1 and 2 but no differential health outcomes were reported in wave 3. The reversal of signs for province dummies is surprising, with possible explanation being the roll-out of anti retro-viral treatment across the provinces in South Africa that may have changed peoples' perception about their health. This however remains speculative for now.

Gender differences in health outcomes were significant. Men had higher health utility scores across the waves and these differences were statistically significant at 1% level of significance. These results confirm that there are health differences in South Africa based on socioeconomic and demographic and on geography. The next challenge is to quantify the differences in health outcomes highlighted in the regression as an index. This has the advantage of showing who is favoured by these health outcomes.

4.4.2 The Factor Decomposition Results

Table 4.2 presents the results of the income related health inequality index (CC) and its decomposition. From the table the CC index is 0.0359 in wave 1, 0.0144 and 0.0194 in wave 2 and wave 3 respectively. Since the index is positive, this means that better health is distributed in favour of the rich across the waves of NIDS in South Africa. In addition the index has standard errors of 0.0036 in wave 1, 0.0023 in wave 2 and 0.0018 in wave 3. This implies that income related health inequality actually decreases in later waves. We note however that the standard errors are quite small and this could be partly because the predictions from interval regression smooth out some of the variation we would expect from health outcomes. In addition, the range of health utility predictions is small (0 to 1) with majority clustered towards the higher levels of health utility.

In addition to summarizing the extent of income related health inequalities, we also decompose the index into factors that drive the inequalities. A positive (negative) contribution for a variable implies that it increases (decreases) income-related health inequalities favouring the rich. The decomposition results show that disparities in income have a positive contribution to income related health inequality across the waves. The contribution of the inequality in income to the income related health inequality is 0.0172. This contribution decreases to 0.0148 in wave 3.

Disparities in educational attainment increase income related health inequality favouring the rich because the overall contribution is positive across all waves. This is despite some categories of educational attainment, such as general education and further education and training, decreasing the income related health inequality. This means that having general educational levels among the population serves to reduce inequalities in health outcomes. On the other hand, having Matric level of education or higher education serves to increase income related health inequalities favouring the rich. The aggregate contribution of categorical variables is invariant to the base category although contribution of specific categories may

Table 4.1: Interval Regression of SAH on its covariates.

VARIABLES	(1)	(2)	(3)
Log of real pc income	0.00581***	0.00260	0.00548***
	(0.00185)	(0.00166)	(0.00197)
General Education (rel. to No educ.)	0.0437***	0.0128	0.0160*
	(0.00881)	(0.00807)	(0.00903)
Further Education and Training	0.0651***	0.0289***	0.0318***
	(0.00934)	(0.00873)	(0.00835)
Matric	0.0856***	0.0437***	0.0433***
	(0.00984)	(0.00857)	(0.00911)
Higher education	0.108***	0.0564***	0.0486***
	(0.0112)	(0.00948)	(0.00894)
25-34 (rel. to 15-24)	-0.0359***	-0.0238***	-0.0209***
	(0.00493)	(0.00294)	(0.00455)
35-44	-0.0681***	-0.0442***	-0.0464***
	(0.00687)	(0.00469)	(0.00554)
45-54	-0.109***	-0.0747***	-0.0698***
	(0.00855)	(0.00595)	(0.00513)
55-64	-0.149***	-0.103***	-0.0958***
	(0.0109)	(0.00913)	(0.00855)
65+	-0.180***	-0.138***	-0.134***
	(0.0128)	(0.0115)	(0.00966)
Coloureds (rel to Africans)	-0.00532	0.00226	0.0174**
	(0.00835)	(0.0108)	(0.00724)
Asians	0.000448	-0.0127	-0.0253*
	(0.0128)	(0.0151)	(0.0146)
Whites	0.0212**	0.0187*	0.00288
	(0.00905)	(0.00981)	(0.00832)
With partner (rel. to married)	-0.0263***	-0.0104	-0.00699
	(0.00880)	(0.00653)	(0.00701)
Widow	-0.0409***	-0.0154	-0.0142*
	(0.00912)	(0.00975)	(0.00841)
Divorced	-0.0128	0.000325	-0.000338
	(0.0155)	(0.0155)	(0.0102)
Never married	-0.0121**	-0.00761*	-0.00966*
	(0.00538)	(0.00426)	(0.00496)
Eastern Cape (rel. to Western Cape)	0.00765	0.0102	0.0197**
	(0.0106)	(0.0103)	(0.00947)
Northern Cape	-0.0172*	-0.000448	-0.00308
	(0.00958)	(0.00847)	(0.0102)
Free State	-0.0327***	0.0109	0.00922
	(0.0108)	(0.0120)	(0.0103)
Kwa-Zulu Natal	-0.0593***	0.00577	0.0220***
	(0.00917)	(0.0110)	(0.00805)
North West	-0.0271**	0.0120	0.0222**
	(0.0115)	(0.0108)	(0.00946)
Gauteng	-0.0168**	0.0258**	0.0257***
	(0.00822)	(0.0105)	(0.00797)
Mpumalanga	-0.0306***	0.00910	0.0170*
	(0.0113)	(0.0101)	(0.00891)
Limpopo	0.00403	0.0362***	0.0306***
	(0.0107)	(0.0112)	(0.00919)
Male	0.0222***	0.0133***	0.0138***
	(0.00363)	(0.00320)	(0.00251)
Urban (rel. to Traditional area)	-0.0155**	-0.0112**	-3.11e-05
	(0.00674)	(0.00517)	(0.00444)
Farms	0.00907	-0.0102	0.00539
	(0.00982)	(0.00883)	(0.00828)
Constant	0.842***	0.881***	0.835***
	(0.0179)	(0.0164)	(0.0180)
Observations	15,446	17,323	18,673

Notes: The dependent variable is self assessed health (intervals). Robust standard errors that allow for correlation in un-observables for each individual drawn from the same primary sampling unit are shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

vary with the base category.

Age differences decrease income related health inequalities favouring the rich across all waves. In addition, differences due to race, province of residence and in marital status increase income related health inequalities favouring the rich except in wave 3 when differences in provinces of residence decrease income related health inequalities favouring the rich. Gender differences in health outcome also increase the income related health inequality in favour of the rich. Differences in health outcomes due to geographical type of residence (Traditional area, in urban areas or on farms) reduce income related health inequality favouring the rich in wave 1 and 2 but do not have any influence in wave 3.

Overall, disparities in income and educational outcomes have the highest positive contribution to the income related health inequalities favouring the rich in South Africa.

The index is enlightening and is used to compare levels of inequality in health outcomes from one time period to another. However, it is more useful when it is decomposed over time to identify the factors associated with changes in the index. To do so, we use Oaxaca type decomposition of the income related inequality index. We note, however, that the dataset we use represents only four years between the first and the third wave. This may not show noticeable differences in income related health inequalities. Our objective here is to address the fundamental question of why income related health inequality decreased over this short time period. Table 4.3 presents the Oaxaca type decomposition results for this exercise.

Following Jann (2008), we decompose the change in income related health inequality into three components. The first is a change in inequality due to a change in the distribution of socioeconomic determinants. The second is a change in inequality due to changes in the coefficients. Finally, a change due to the interaction or change in both coefficients and distribution of covariates. The results in table 4.3 show that the biggest contribution to the change in income related health inequality is the interaction (in some literature this is also referred to as, unexplained) (Jann, 2008). The change in income related health inequality due to the distribution of covariates is higher than the change due to coefficients, although the difference between the two is small.

4.4.3 Sensitivity Analysis

The interval regression we use to scale SAH requires cut points (thresholds) for each SAH category. We use a generic measure of health to determine these thresholds. In the absence of such a measure in the data, we applied the Canadian HUI which is widely used in the literature(Van Doorslaer and O'Donnell, 2008). In the Canadian data, both HUI and SAH were collected at the same time hence the HUI thresholds are used to set SAH thresholds. These thresholds are used where only SAH is collected. We check for robustness of our

Table 4.2: Decomposition of the corrected concentration index

		Wave 1		Wave 2		Wave 3	
Variable		**CC**	**Aggregate CC**	**CC**	**Aggregate CC**	**CC**	**Aggregate CC**
Income		0.0172	0.0172	0.0076	0.0076	0.0148	0.0148
Education	General Education	-0.0123		-0.0035		-0.0045	
	FET	-0,0026		-0.0013		-0.0024	
	Matric	0.0134		0.0062		0.0054	
	Higher education	0.0300	0.0285	0.0149	0.0163	0.0153	0.0138
Age	25-34	-0.0019		-0.0003		-0.0006	
	35-44	-0.0047		-0.0038		-0.0039	
	45-54	-0.0035		-0.0035		-0.0027	
	55-64	-0.0049		-0.0039		-0.0035	
	65+	-0.0030	-0.0180	-0.0031	-0.0146	-0.0022	-0.0129
Race	Coloureds	-0.0003		0.0002		0.0006	
	Asians	0		-0.0007		-0.0012	
	Whites	0.0065	0.0062	0.0054	0.0049	0.0008	0.0002
Marital status	With partner	0		0.0001		0.0001	
	Widow	0.0008		0.0002		0.0001	
	Divorced	-0.0005		0		0	
	Never married	0.0031	0.0034	0.0020	0.0023	0.0022	0.024
Province	Eastern Cape	-0.0011		-0.0008		-0.0020	
	Northern Cape	0		0		0	
	Free State	0		0		-0.0001	
	Kwa-Zulu Natal	0.0080		-0.0007		-0.0023	
	North West	-0.0003		0.0001		0	
	Gauteng	-0.0039		0.0054		0.0060	
	Mpumalanga	-0.0001		-0.0001		-0.0001	
	Limpopo	-0.0003	0.0023	-0.0032	0.0007	-0.0025	-0.0010
Male		0.0032	0.0032	0.0019	0.0019	0.0021	0.0021
Geographical types	Urban	-0.0070		-0.0048		0	
	Farms	0.0001	-0.0069	0.0001	-0.0047	0	0
Total (CC)		**0.0359**		**0.0144**		**0.0194**	

Notes: The table shows magnitudes and direction of contribution of each factor in the income related health inequality index (corrected concentration index). The aggregate contribution of categorical variables is invariant to the base category although contribution of specific categories such as specific provinces may vary with the base category. Positive (negative) values indicate that the factor increases (decrease) income related health inequality favouring the rich.

Table 4.3: Factor Decomposition over time (Oaxaca Type decomposition)

		1	2	3	4	5	6	7	8	9
Variable		CC_{w1}	CC_{w3}	β_{w1}	CC_{w3}-CC_{w1}	β_{w3}	$\beta_{w3}-\beta_{w1}$	$CC_{w1}(\beta_{w3}-\beta_{w1})$	$\beta_{w1}(CC_{w3}$-$CC_{w1})$	Interaction
Income		0.0172	0.0148	0.006	-0.002	0.006	0.000	0.0000	0.0000	-0.0024
Education	General Education	-0.0123	-0.0045	0.044	0.008	0.016	-0.028	0.0003	0.0003	0.0071
	FET	-0,0026	-0.0024	0.065	0.000	0.032	-0.033	0.0001	0.0000	0.0001
	Matric	0.0134	0.0054	0.086	-0.008	0.043	-0.042	-0.0006	-0.0007	-0.0067
	Higher education	0.0300	0.0153	0.108	-0.015	0.048	-0.060	-0.0018	-0.0016	-0.0113
Age	25-34	-0.0019	-0.0006	-0.036	0.001	-0.021	0.015	0.0000	0.0000	0.0014
	35-44	-0.0047	-0.0039	-0.068	0.001	-0.046	0.022	-0.0001	-0.0001	0.0010
	45-54	-0.0035	-0.0027	-0.109	0.001	-0.070	0.039	-0.0001	-0.0001	0.0010
	55-64	-0.0049	-0.0035	-0.149	0.001	-0.096	0.053	-0.0003	-0.0002	0.0019
	65+	-0.0030	-0.0022	-0.180	0.001	-0.134	0.046	-0.0001	-0.0001	0.0011
Race	Coloureds	-0.0003	0.0006	-0.005	0.001	0.017	0.023	0.0000	0.0000	0.0009
	Asians	0	-0.0012	0.000	-0.001	-0.025	-0.026	0.0000	0.0000	-0.0012
	Whites	0.0065	0.0008	0.021	-0.006	0.003	-0.018	-0.0001	-0.0001	-0.0055
Marital status	With partner	0	0.0001	-0.026	0.000	-0.007	0.019	0.0000	0.0000	0.0001
	Widow	0.0008	0.0001	-0.041	-0.001	-0.014	0.027	0.0000	0.0000	-0.0024
	Divorced	-0.0005	0	-0.013	0.001	0.000	0.013	0.0000	0.0000	0.0005
	Never married	0.0031	0.0022	-0.012	-0.001	-0.010	0.002	0.0000	0.0000	-0.0009
Province	Eastern Cape	-0.0011	-0.0020	0.008	-0.001	0.020	0.012	0.0000	0.0000	-0.0009
	Northern Cape	0	0	-0.017	0.000	-0.003	0.014	0.0000	0.0000	0.0000
	Free State	0	-0.0001	0.033	0.000	0.009	0.042	0.0000	0.0000	-0.0001
	Kwa-Zulu Natal	0.0080	-0.0023	0.059	-0.010	0.022	-0.037	-0.0003	-0.0006	-0.0094
	North West	-0.0003	0	-0.027	0.000	0.022	0.049	0.0000	0.0000	0.0003
	Gauteng	-0.0039	0.0060	-0.017	0.010	0.026	0.043	-0.0002	-0.0002	0.0102
	Mpumalanga	-0.0001	-0.0001	-0.031	0.000	0,017	0.048	0.0000	0.0000	0.0000
	Limpopo	-0.0003	-0.0025	0.004	-0.002	0.031	0.027	0.0000	0.0000	-0.0022
Male		0.0032	0.0021	0.022	-0.001	0.014	-0.008	0.0000	0.0000	-0.0010
Regions of residence	Urban	-0.0070	0	-0.016	0.007	0.000	0.016	-0.0001	-0.0001	0.0072
	Farms	0.0001	0	0.009	0.000	0.005	-0.004	0.0000	0.0000	-0.0001
Total (CC)		**0.0359**	**0.0194**					-0.0033	-0.0035	-0.0096

Notes: The table shows magnitudes and direction of contribution of each factor to the corrected concentration index in wave 1 and wave 3. Positive values indicate that the factor increases income related health inequalities favouring the rich and vice versa. Column 7 shows inequality attributed to change in coefficients while column 8 shows inequality attributed to change in covariates. The last column captures inequality due to interaction.

results using alternative generic[12] measures of health in setting the thresholds for the SAH categories.

The alternative generic measures we use to set the thresholds are the short form 12 (SF-12) and the EuroQo-15D. These are generic measures of health which are collected in continuous scale like the HUI. They differ from HUI in the scoring algorithm used. The algorithm used to score responses in these alternative generic measures is developed using econometric modeling while for HUI a multiplicative, multi-attribute utility function is used. In practice they give fairly comparable values (Tolley, 2009). We use these alternative measures to set thresholds for SAH and then generate predicted values from the interval regressions in the same way as we generated the cardinalized SAH health using HUI. We also keep the regressors the same to make the predicted values comparable.

We find that the patterns of kernel densities for predicted interval regressions using these alternative generic health measures do not deviate from the ones we obtained from the HUI. For example, figure 4.4 shows kernel densities of predicted values from interval regression where we use SF-12 index to set thresholds for the three waves in NIDS. Figure 4.5 shows predicted values from interval regression where we use EuroQo-15D index to set thresholds for the three waves in NIDS. In all cases the densities are skewed to the left implying that the majority of the individuals have better health outcomes according to these measures. These results confirm that whichever generic health measure we use to set thresholds for SAH, our initial results do not change.

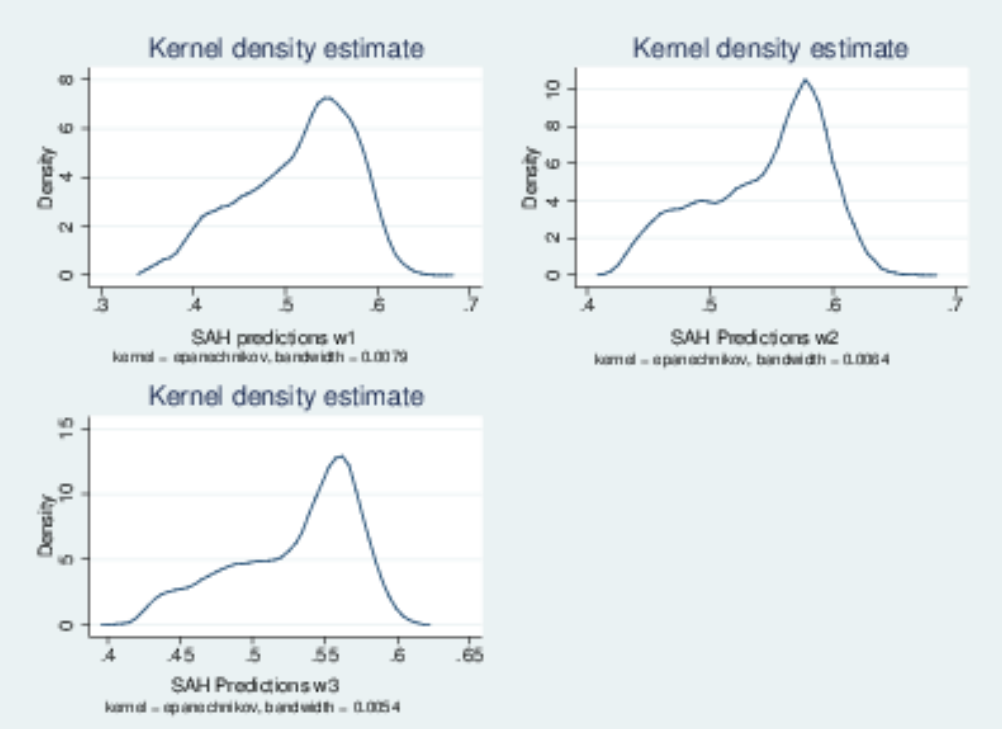
We also estimate income related health inequality using the corrected concentration index applied on these cardinalized SAH measures. We find that the index is positive implying that better health is concentrated among the well off. This is consistent with the results we find with the predictions from interval regression where we used the HUI index to set SAH category thresholds. The results of these indices together with their standard errors are shown in table 4.6 as well as the index we obtained earlier using HUI to set SAH category thresholds. As we can see income related health inequality decreased from wave 1 to wave 3 irrespective of the generic health we use to set thresholds. The results are therefore robust to the use of alternative generic measures to set SAH category thresholds.

4.5 Discussion of the transmission mechanisms

In this section, we discuss the transmission mechanisms responsible for the endogenous relationship between income and health outcomes. We have already discussed in detail the transmission mechanism between education and health in the previous chapter. We start

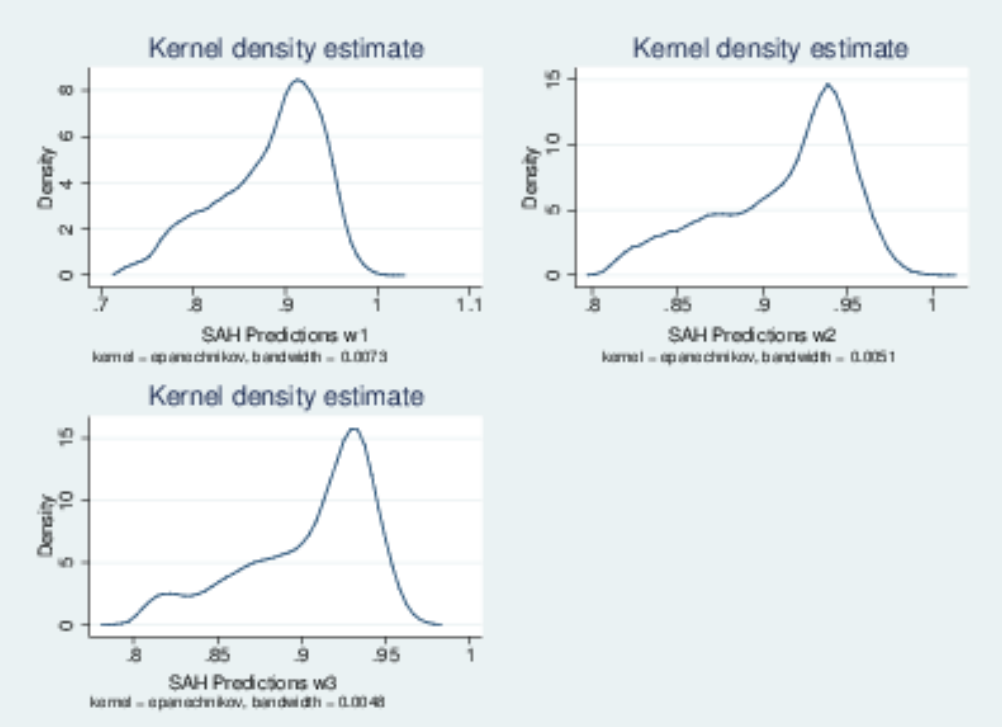
[12]Health measures can be classified as i) Subjective health (e.g. SAH) ii) Generic health (e.g. HUI, SF12, EQ-15D) iii)Vignettes based health iv) Objective health (e.g. height for age, weight for height).

Table 4.4: Kernel density of SAH health predictions using SF-12



Notes: Kernel density plots for linear predictions of health from interval regressions for waves 1-3. Thresholds for SAH categories are set by thresholds from SF-12 index. The kernel density plots are skewed to the left implying that majority of the people have better health.

Table 4.5: Kernel density of SAH predictions using EQ-15D



Notes: Kernel density plots for linear predictions of health from interval regressions for waves 1-3. Thresholds for SAH categories are set by thresholds from EuroQo-15D (EQ-15D) index. The kernel density plots are skewed to the left, implying that the majority of people have better health.

Table 4.6: Corrected concentration indices for health

	Corrected concentration index	
Generic measures of health	**Wave 1**	**Wave 3**
Health Utility Index (HUI)	0.0359	0.0194
	(0.00365)	(0.00193)
Short Form (SF)-12	0.026	0.016
	(0.0025)	(0.00161)
EuroQo 15D	0.023	0.013
	(0.00238)	(0.00143)

Notes: The table shows the corrected health concentration index for SAH when the SAH category thresholds are fixed using three different generic health. The standard deviations are shown in parentheses. The health concentration index decreases from wave 1 to wave 3 in all cases showing that our results are robust to the different ways of setting SAH thresholds.

by looking at how income affects health before we look at the transmission from health to income. The order in which we discuss the transmissions does not imply that one transmission is more important than the other. Throughout our analysis, we have assumed that the transmission is unidirectional, flowing from income to health. For this reason we have included income as one of the explanatory variables in all the health models.

There are four potential channels of transmission identified in the literature that explain how income affects health, namely; the medical care channel, water and sanitation channel, nutrition channel and psycho-social channel (Case, 2002; Halla and Zweimüller, 2013; Adler and Newman, 2002). One way in which income might affect health is through its effects on health care. Access to, use of, and quality of health care vary with income levels (Adler and Newman, 2002). Higher incomes may allow people to spend more time and money seeking out health services, or they may spend more on goods and services associated with better health such as more nutritious foods. Higher incomes too can afford individuals medical insurance that allows them to access good quality health care. Individuals may also derive health benefits directly from income by relieving stress (or depressive symptoms) occasioned by poverty.

The medical care transmission channel can be tested empirically in the data by modeling health care utilization where the dependent variable is health care utilization and the explanatory variable of interest is the exogenous income such as old age pension. Health care utilization options include provider ownership such as private, public, or mission owned. A significant coefficient on income in this model would confirm the existence of the medical care transmission channel.

An alternative way in which income may affect health is through better water and sanitation. Adler and Newman (2002) refers to this channel as the environmental exposure channel. Higher incomes may be used to upgrade dwelling facilities and the improvements made have positive health consequences. Some of these improvements include installation of

a flush toilet, installation of running water and general improvement in the type of dwelling. The income may also enable one to live in a better neighbourhood, free from exposure to health damaging agents in the environment such as industrial waste.

Another example that is directly linked to environmental quality is overcrowding (or crowding). Overcrowding occurs when there are too many people in the available space and facilities. This can be at household level, measured in persons per area of living space, or at community level. Crowding may also be defined at room level when there are too many people per room or at building level when there are too many people per building. The number of rooms available per person plays a critical role in determining the nature of interactions in the household, and is related to poor mental and physical health. These channels imply that the poor are directly exposed to disease causing pathogens and risks of contracting diseases associated with alcoholism such as liver cirrhosis (Falkingham and Namazie, 2002). Behaviours such as smoking and drinking of alcohol also act as pathways through which income affects health.

The transmission channel discussed above can be tested empirically by estimating the probability of improving the dwelling facility or changing environmental exposure as a function of exogenous income receipt. A significant coefficient would imply the existence of this channel of transmission from income to health.

A third way through which income may influence health outcomes is through its effects on nutrition. The likelihood of skipping a meal due to poverty is reduced in higher income households compared to lower income households. Poor households also have lower consumption of fiber and fresh fruits. Given the direct effects of nutrition on health, households that skip meals are more likely to have poorer health outcomes.

The three channels discussed above are indirect pathways through which income affects health. We now turn to the fourth channel through which income directly affects health. That is the psycho-social channel. In chapter two we showed that depression is measured through an index constructed from the CES-D scale. The items in the CES-D scale are self assessed short questions regarding depressive symptoms. Some of these depressive symptoms are caused by poverty and the state of hopelessness. Coping with very low income is a source of stress. The receipt of income by a depressed person or one who feels hopeless may reduce his stress or depressive symptoms in general and improve his emotional health. This channel can also be tested by regressing the CES-D items on exogenous income with a significant coefficient on income confirming existence of this channel.

There is an endogenous relationship between health and income that potentially biases results obtained from OLS regression of these variables. This will become clearer in the next chapter where we discuss how to disentangle the effect of income on health outcomes. We

now focus on understanding the transmission mechanism from health to income. To do this we note that part of the correlation between income and health outcomes in South Africa is due to the reduced earnings potential of South Africans who have been chronically ill (Case, 2002). Good health is a prerequisite to work hard and to perform on the labour market (Halla and Zweimüller, 2013).

Chronic illnesses have negative and significant effects on health status and are related to labour force participation and hours worked. This effect is direct, if the individual himself is ill, or indirect, if the individual spends productive time caring for the chronically ill. This in turn affects earnings.

The other health-income channel which appears in the macro literature is the incentive channel (Finlay, 2007). This channel assumes that individuals who are healthier and have a greater life expectancy will have the incentive to invest in education as the time horizon over which returns can be earned is extended. Health according to this channel works through education and therefore has an indirect effect on income.

Finally, both health and income may be correlated with a third factor. One possible confounding factor is given by a low rate of time preference. Individuals who discount future benefits relatively little are supposed to invest more in human capital, as well as to engage in a healthier life-style. In this case income and health are correlated, but not in any causal way (Halla and Zweimüller, 2013; Grossman, 1972).

4.6 Conclusion

In this chapter we set out to estimate the extent of income related health inequalities in South Africa and decompose this into contributing factors. We do this by computing the corrected concentration index because of its desirable properties over other indices (such as the standard concentration index and Wagstaff index). The suitability of the corrected concentration index arises from the boundedness of the health outcome variable we use. We use SAH as our indicator of health outcome because the international literature prefers this indicator due to its predictive power regarding mortality and subsequent use of health services. This variable predicted two year mortality for South Africa using NIDS data, further adding credence to its suitability for our purpose (Ardington and Gasealahwe, 2012).

SAH is a qualitative indicator of health that does not directly lend itself to the calculation of a health concentration index. We scaled it from ordinal to cardinal scale using predicted values of interval regression, a method we borrowed from Van Doorslaer and Jones (2004). The advantage of scaling the SAH using predictions of interval regression is that the health outcome obtained is measured on the same scale as health utilities, which are true measures

of quality of life. Although the health utility index is unit-less it has clear bounds where 0 reflects the worst health status (death) and 1 the best health status and can be interpreted as quality adjusted life years (QALYs).

We also utilized the rich socioeconomic data in NIDS such as income that we used as a socioeconomic ranking variable in the computation of the corrected concentration index. The panel nature of NIDS data allows us to counter check one of the most important characteristic of SAH, that is, whether it predicts mortality. We find a positive corrected concentration index across all waves of NIDS implying that better health in South Africa is concentrated among the rich. This is the case irrespective of the generic health measure we use in setting the threshold for SAH for the purpose of cardinalizing it.

We decompose the index to identify the factors associated with inequalities in health outcomes. It turns out that the most consistent factors associated with income related health inequality in South Africa are disparities in income and educational attainment. These factors are associated with increase in income related health inequality, while disparities in age are associated with a decrease in inequality. These findings are important for policymakers who want to solve the problem of inequalities in health. Inequalities in health outcomes can be solved partly by reducing inequalities in the factors that determine health, such as income and education. For example, increasing the opportunity for education among low income groups, will improve their health outcomes thereby reducing the income related health inequalities.

Chapter 5

Disentangling the Effects of Income on Health in South Africa: A Regression Discontinuity Design

5.1 Introduction

Previous studies that span many disciplines have shown a strong association between income and health status (Case, 2004)(Case 2001, 2004, Smith et al. 2010). This does not by itself demonstrate that income has a causal effect on health. Threads run from income to health, and from health to income, with third factors potentially influencing both of them (Smith 1999, Meer et al 2003, Case 2002). Moreover, studies using panel data to model the dynamics of the relationship between income and health have questioned if a causal effect of income changes on health exists (Adams et al., 2004; Contoyannis et al., 2004). This issue continues to be debated in the literature.

In chapter three of this thesis, we identify income and other socioeconomic variables as significant determinants of health outcomes in South Africa. We adopt the methods applied by others who have studied this topic. The approaches we use, the bivariate and panel data estimators, do not lead to causal interpretation of the parameters. While the panel data approach controls for unobserved heterogeneity, if the heterogeneity is time variant, the bias in the parameter estimates is not completely controlled by panel data estimators. We now extend that work in this chapter by investigating whether the correlation between income and health outcomes is causal.

A clear understanding of the causal effect of income on health outcomes is important for several reasons. First, the question is of direct scientific interest because only when it can be demonstrated that income causes health status does it become important to study the actual mechanisms underlying such relationship (Ettner, 1996) and (Gordon and Miller,

2012). Secondly, the causal effect is vital for policy design aimed at narrowing the health inequalities in society (Frijters et al., 2005). Thirdly, resources devoted to different policy options are scarce and should therefore depend on their relative effectiveness (Case, 2004). Finally, understanding the causal link forms a natural benchmark to evaluate health care interventions based on improvements resulting from income transfers (Case, 2004; Case and Deaton, 2005).

Whether and how income influences health has however proved difficult for researchers to disentangle because of endogeneity[1] of health and income. Thus conflicting results have been reported in the literature with some showing a positive relationship (Pritchett and Summers, 1996; Case, 2004; Strauss et al., 2010) while others showing a negative relationship (Hernández-Quevedo and Jiménez-Rubio, 2009; Snyder and Evans, 2006) or no relationship at all (Adams et al., 2004).

To investigate whether income has a causal effect on health requires identifying a source of income that is not itself determined by the respondent's health status, that is, an exogenous source of income. In the absence of randomized control experiments, which are generally not feasible in this case (Frijters et al., 2005) researchers need to tackle two main issues in order to establish causality (Wooldridge, 2002). In the first place there are likely to be individual heterogeneities that are unobserved by the researcher which jointly determine both income and health (third factor explanation). Examples include genetic endowments, parental socioeconomic status, social factors and discount rates (Frijters et al., 2005). Secondly there is the issue of reverse causality. Increased income may lead to improved health through better lifestyle or diet, fewer monetary worries, better access to medical services and improved input mix (Smith 1999). However, it is also true that people in good health are more likely to be economically active and have higher incomes (Frijters et al., 2005). These issues are not easy to deal with in empirical research because they introduce biases in the parameter estimates.

The use of panel data to identify a causal relationship between income and health is often problematic due to attrition. As we saw in chapter two, over 20 percent of individuals attrited from the sample in wave 2 and wave 3. We are only able to test for attrition bias with respect to observable variables. The bias due to attrition with respect to unobservables could still be a problem. We find that income is a determinant of attrition. Attrition is therefore a problem when those attriting are either the unhealthiest or the rich (Fitzgerald et al., 1998; Contoyannis et al., 2004) and (Van Ourti 2003). This may lead to an upward trend in health outcomes (if only the unhealthy attrit) which may be wrongly attributed to income changes, or a downward trend if only the rich attrit (Contoyannis et al., 2004).

[1] Endogeneity can be as a result of simultaneity, the likelihood that both health and income are jointly determined by a third factor, and finally the measurement errors.

In this chapter we use the exogenous variable of old age grant (old age pension) in South Africa to identify the effect of income on health outcomes. South Africa presents us with an important case study because there exists a twin problem of inequality in income and inequality in health outcomes as we have demonstrated in chapter 4, yet there is no study at the national level that causally links the two. This is only possible if we first establish the causal effect of income on health. In the case of South Africa, there is an old age grant program in place that was available to non-Whites in 1944 that we use as an exogenous source of income. Since it is based on age eligibility we use it to identify the effect of income on health at the threshold age using Regression Discontinuity Design.

Previous studies in South Africa have used the old age pension income to analyze the effect of income on health, but only on localized samples (Case, 2004; Case and Deaton, 2005). Our study complements this research in two ways. First we use a nationally representative dataset to answer a similar research question. Second, the old age pension eligibility rule enables us to apply a robust quasi experimental technique that is able to identify the impact of income on health outcomes if it exists.

The rest of the chapter is structured as follows, in section 5.2 we explain the old age grant in South Africa (that is, the context). We review the related literature in section 5.3, and we explain the method of analysis and the data used in section 5.4. We present the findings in section 5.5 and conclude in section 5.6.

5.2 Brief overview of Old Age Grant in South Africa

The Old Age Grant (which used to be called the old age pension) is a non-contributory public transfer of resources by the South African Social Security Agency (SASSA) to older men and women in South Africa with permanent residence status who satisfy a means test. The pension programme was introduced in 1928 as an income supplement for elderly low income Whites only (Hamoudi and Thomas, 2005). In 1944, eligibility was extended to the Coloured and black African population groups, with payment schedules dependent upon the recipient's race and location of residence. In 1965, differentiation based on residence was abolished but that on racial gaps was maintained. In 1975, for example, eligible Whites received about R1200 (in 2004 prices), while Coloureds received about R600 and Africans received less than R200 (Hamoudi and Thomas, 2005). The racial gaps were reduced during the 1970s and 1980s. With the end of Apartheid payments were set at R370 in 1993 for all races (equivalent to R740 in 2004 prices). This was adjusted to R960 per month from 1st April 2008 and R1060 per month from 1st April 2009.

Age eligibility differed initially by gender with women aged 60 years and older being

eligible while from 1st April 2008, men aged 63 and older were eligible and from 1st April 2009 men aged 61 and older were eligible. From 1st April 2010, the age threshold for men was reduced to 60 years to match women's.

The old age grant, like other social grants is subject to a means test, which implies that SASSA evaluates the income and assets of the person applying for the grant in order to determine whether the person's means are below a stipulated amount. Both the recipients and their spouses are subjected to means tests. The means test is set so that almost all black South Africans are eligible for the old age grant (Case and Deaton 1998). In 2004, eligibility required a monthly income of R1502 or less and assets of less than R266,400 (excluding owner occupied housing).

Although this amount appears small, in relative terms, as noted by Case and Deaton (1998) it is quite substantial especially for Africans (blacks) whose average monthly income in 2008 from the National Income Dynamics Study (NIDS) data set is R6755 . In this chapter we concentrate on teasing out the causal effect of income on health outcomes using women because the pension thresholds for men have changed over time.

5.3 Related Literature

The theoretical underpinning of this chapter is the Grossman human capital model which we have already discussed in chapter three. Below we discuss the empirical literature that focuses on how the exogenous income variation has been used to tease out the causal effect of income on health outcomes.

In the literature, many techniques have been applied to establish causal effects of income on health. Identification strategies have tried to establish exogenous variations in income and use this in the regression. Some authors have argued that regression methods are normally misspecified when non linearity of the relationship between health and income is neglected, which then act, as omitted variables and, if correlated with other variables, leads to biased estimates. Other relatively recent robust techniques such as regression discontinuity design have been used to identify causality.

Among the first studies in this area were those by Ettner (1996) and Pritchett and Summers (1996) who used instrumental variables on cross sectional data from a number of US surveys and cross-country time series data respectively. Ettner (1996) adopted a two stage instrumental variable approach, where a family income was instrumented by respondent's wage rate and non-earnings income on the assumption that these variables impact health through their effect on family income. They find that income remains significant and the coefficient on the instrument is higher. They conclude that this could be due to a mea-

surement error in the income variable. Pritchett and Summers (1996) estimates the effect of income on health using cross-country time series macro data on health and income per capita. The study employs an instrumental variables approach to identify the pure income effect on health performance, and finds that long run income elasticity of infant and child mortality in developing countries lies between -0.2 and -0.4.

Adams et al. (2004) use a panel of older respondents to establish the effect of income on health. Their reasoning is that focusing on people aged above 70 years together with conditioning on previous health outcomes in the econometric model reduces potential bias due to reverse causality, i.e health affecting wealth. They find that for this elderly group, there is very little evidence that changes in wealth lead to health changes.

Contoyannis et al. (2004) use random effects ordered probit models on both balanced and unbalanced samples of the British Household Panel Survey (BHPS) and also condition on previous health states in the model. They also control for the problem of initial conditions arising in dynamic panel data models by including individual specific characteristic averages. They correct for attrition bias using inverse probability weights and find significant association between log-income and better health. However, these could not be interpreted as causal due to the potential correlation between average income and the unobservable individual effect. The results also show a significant relationship between current income and health but the quantitative effect is small and insignificant in most of the models they use. They find current income to be only significant for men in the random effects model and not in the pooled models. For women they do not find any significance at all.

Frijters et al. (2005) use exogenous variation in income brought about by the reunification of East and West Germany on health satisfaction in the years following reunification. Using a fixed effects ordinal estimator and a causal decomposition technique to account for panel attrition, they find evidence of a significant positive effect of income changes on health satisfaction. The quantitative effect of the size of this effect is however small for both current and permanent income.

Snyder and Evans (2006) use a social security notch that existed in the US pension system to introduce exogenous variation in income on mortality. They use a reduced form relationship between the notch and mortality. They also apply regression discontinuity design to tease out this impact. They compare mortality rates among men born in the last quarter of 1916 and those born in the first half of 1917. Their argument is that these men had similar observable characteristics partly because their birth dates were close, but once they reached retirement age, one group was rewarded with substantially higher social security payments than the other. They use this variation to examine the potential effect of income on mortality in an elderly population. They find that the 1916 birth cohort who received more income recorded higher levels of mortality than the lower earning 1917 birth cohort.

We are not the first to use old age pension as an exogenous source of variation in income to tease out the effect of income on health in South Africa. Our study is however different in terms of the wider coverage of the data which assures national representation. We also apply a quasi experimental technique, the Regression Discontinuity Design (RDD) which reveals the impact of income on health outcomes for women at the threshold age of pension grant receipt in case it exists. Duflo (2000) uses the old age pension in South Africa to establish the causal effect of income on health among children. The basic idea of her identification strategy is to compare the differences in height between children in pension eligible and non eligible households and between children exposed to the programme for a fraction of their lives and those exposed their entire lives, and attribute any changes in childrens' height to pension income. The use of household income is meant to bring exogenous variation in income.

Case (2004) uses Langeberg Survey data to identify the impact of money on health status by comparing the self -reported health status of black and Coloured adults in households that pooled resources and those that did not. Her identification strategy is based on the exogenous increase in income associated with changes in the South African old age pension. Her study finds that in households that pooled their income the exogenously determined old age pension protected the health of all household members. They further establish that the main mechanism underlying this relationship is through improvements in nutritional status. However these results only cover a small area in South Africa and therefore their external validity is uncertain.

Hamoudi and Thomas (2005) investigates whether pension eligible adults are more likely to co-reside with other adults who have lower levels of human capital, as measured by height and education. Their argument is that, since height and education are fixed for adults, changes in mean height and mean education in pension households reflects the selection of adults who co-reside with older adults when they become eligible for the pension. Education and height are therefore used as instrumental variables for family composition. They show the potential value of moving beyond theory and data which are bound by the confines of a spatially determined definition of the household.

The difference between our study and the above studies is that we focus on the health outcomes of the recipients of the pension and not the other members of the household. This focus is meant to give extra confirmation on the effect of income on health by complementing the results obtained by Duflo (2000), Case (2004) and Hamoudi and Thomas (2005) who investigated the effect of pension income on other members of the household in South Africa."

Gordon and Miller (2012) apply the Regression Discontinuity Design (RDD) to establish a causal link between income and health using old age pension data from the October Household Surveys (OHS) in South Africa. Their identification strategy focuses only on discontinuities in age and so it is only able to pick up impacts local to changing pension eligibility status.

The impacts are therefore for short term measures of health and long term investment in health. Our study complements these studies and uses a nationally representative data set to answer similar research questions.

Eibich (2014) uses a regression discontinuity design to understand a different set of relationships, that is, the effect of retirement on health, health behaviour and health care utilization.

5.4 Method of Analysis

We utilize the exogenous variation in income from old age grants to tease out the effect of income on health outcomes using a nationally representative data. We use an identification strategy that isolates the effect of income on health outcomes based on the Regression Discontinuity Design (RDD). In the RDD we follow the approach by Gordon and Miller (2012) and look for discontinuities in health outcomes associated with receiving an old age grant. Since the RDD works when there is a clear rule, the old age grant is preferred because it has a prescribed threshold eligibility age in South Africa.

The continuous age (design or forcing) variable that determines eligibility is subject to random variability at the threshold due to measurement error, sampling variability and chance factors. Random variability at the threshold implies that people who are just above the eligibility threshold and those just below it are similar, in expectation, on all observed and unobserved pre-old age grant characteristics, like in a randomized control trial (RCT). In addition, although this is not directly testable, we believe that it is appropriate to assume that there are no other discrete changes to individual health that coincide with becoming old age grant eligible. We are also not aware of any possible causes or confounders that produce changes in health at the old age grant eligibility threshold. In other words we are not aware of any health condition that attacks women at the age of old age grant eligibility threshold. In view of this assumption, we do not expect to see a dramatic discontinuous change in health outcomes at the old age grant eligibility threshold. The only thing that changes discontinuously at that age is that the individual becomes eligible for an old age grant.

We also expect that health outcomes may decline with age but does so in a smooth fashion. The strategy to identify the effect of income then is to look for discontinuous changes in health outcomes associated with old age grant eligibility, and to attribute these as the impacts of exogenous income from the old age grant.

There are two main approaches in estimating RDD, the parametric and non-parametric approaches. In the parametric case we explicitly estimate the effects of age (forcing variable)

on the old age grant by estimating a higher order polynomial of age on health outcomes. We then look for a a jump at the threshold age. To correctly identify the effects of the grant just below the threshold and just above, we need to get the functional form exactly right. This is explained further below:

We assume a given health outcome is related to age a and old age grant eligibility $d = 1[age >= age\_cutoff]$ in the following manner:

$$h = m(a) + d\alpha + \varepsilon \tag{5.1}$$

Where h is health outcome and $m(.)$ is the relationship between health and age. With the assumption that the probability of treatment (receipt of old age grant) changes discontinuously at the treatment age and second that in the absence of old age grant receipt, the relationship between health and age is smooth, the estimate of α identifies the causal impact of the old age grant (income) on health outcome.

The second approach of estimating the RDD is the non parametric approach. In this case we restrict the sample to individuals close to the cutoff. We undertake a strictly local estimation at the threshold. This makes it less sensitive to functional form assumptions but requires enough observations around the threshold to undertake this. Because we do not know the functional form a priori, we adopt the non parametric approach. This works well for us because we have enough observations on both sides of the threshold to permit the analysis.

If we look at a small enough neighbourhood around the threshold age, our estimates of the effect do not depend on correctly specifying age ($m(.)$) in the model. For values of age close enough to the threshold age, every one has the same values of the estimates on age in the model ($m(.)$). This boils down to estimating $limE(h|age)$ from below and above. For this strategy to work, there must be enough observations on either side of the age eligibility threshold.

Secondly, for the impact of income to make intuitive sense we distinguish between health outcomes that respond to income shocks in the short term and long term. Health outcomes that potentially respond to shocks in income in the short term (say within a year or less) that we analyze in our data include self-assessed health and depressive symptoms (emotional/mental health). Adult mortality and chronic illnesses are examples of long term health outcomes in the data. Our analysis is on the contemporaneous effect of income on health outcomes, i.e. effects within one year.

Although the RDD is a quasi-experimental technique that identifies the effect of exogenous income on health, its application in panel data is not fully developed (Fé and Hollingsworth, 2011). The identification is based on the discontinuity at the threshold age. This threshold age occurs to an individual once in the survey. For example, in the first wave individuals

younger than the threshold age do not receive pension and they act as controls, while those older receive pension and form the treatment group. In wave 2, individuals who were younger than the threshold age in wave 1 but have now grown older and crossed the threshold age would be in the treatment group implying that such individuals will be both in the control group and the treatment group at the same time. This makes implementation of the RDD difficult. For this reason we only use wave 1 to investigate the effect of income on health (Fé and Hollingsworth, 2011).

5.4.1 Data, Descriptive statistics and discontinuity checks

This study uses data from the National Income Dynamics Study (NIDS) of South Africa. This has already been described in chapter two. For this chapter, we are only analyzing data for women from wave 1. In this section we describe our estimation sample.

28226 individuals were interviewed in wave 1 of NIDS. 16871 are adults and 9605 children. Information about an extra 1750 individuals was collected indirectly from a third person (proxy) on behalf of these individuals. Since some of the information we use in the analysis is personal assessment and anthropometric measurements, we exclude those interviewed by proxy because the information on their health outcomes may not be accurate.

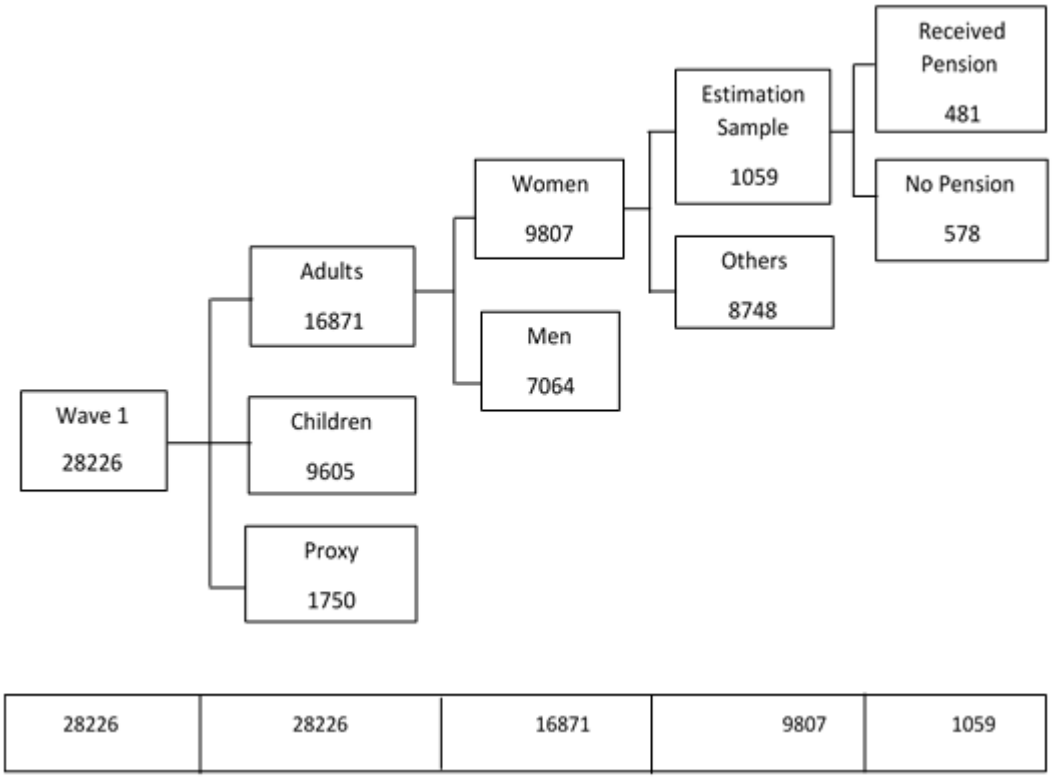
We make further sample restrictions where we restrict our analysis to only adult women aged 55 to 65 years (estimation sample). This is because old age grant eligibility changed for men in the course of the NIDS survey period and we therefore exclude them from the analysis. Their eligibility threshold is also different from that of women. The old age grant eligibility age is 60 years for women and the chosen age range accommodates this threshold (Case, 2004; Duflo, 2000), and (NDoSocial services). The final estimation sample consists of 1059 women, 481 of whom reported receipt of the old age grant while 578 did not receive it. Figure 5.1 shows a schematic representation of the sample we finally analyze.

In the overall sample of wave 1 of NIDS, there are approximately 8% recipients of old age grants in South Africa in 2008. Of these, 77% were Africans, 10% Coloureds, 4% Asian and 9% Whites. Over 72% of the recipients of old age grants were female. This again tells us that although we analyze women only, we have captured a large percentage of the recipients.

In this study we use three health outcome variables to show the effects of income on health, namely SAH, BMI and Depression (mental health). These variables have been described in chapter three where we use them in bivariate and panel data to establish determinants of health outcomes in South Africa. As we have seen income is significantly correlated to these health outcomes. We now want to establish whether the correlations we found can be interpreted as causal.

Table 5.1 shows summary statistics for the estimation sample. As we can see the mean age of the estimation sample is 59 years with 13 years of education. The majority of women

Figure 5.1: Schematic description of sample selection



Notes: This figure shows a schematic description of how the estimation sample was selected from the initial 28226 individuals interviewed in wave 1. The last column shows the estimation sample of 1059 women.

Table 5.1: Summary statistics of estimation sample

Variable	Mean	Number	n
Age	59.30		1059
Years of education	13.39		1058
Marital status			1055
Married		462	
With Partner		45	
Widow		315	
Divorced		62	
Never Married		171	
Race			1059
African		773	
Coloured		153	
Indian		27	
White		106	
Geographical regions			1059
Traditional areas		486	
Urban		470	
Farms		103	

Notes: The table shows summary statistics and breakdown of the estimation sample.

are married or widowed. The majority of the women are black and almost equal proportions live in the traditional and urban areas.

5.4.2 Fuzzy regression discontinuity

We follow Imbens and Lemieux (2008) in the implementation of this design. We think age eligibility matters because it strongly predicts whether someone actually receives an old age grant. We therefore create a "first stage" plot that shows the probability of receiving an old age grant (the treatment) against age for women between the age of 55 and 65 years. Figure 5.2 shows the scatter plot for these probabilities. As is clear from the figure, the probability of receiving an old age grant jumps dramatically from about 0.1 to more than 0.5 at the age of eligibility (threshold) and increases thereafter with age. But as shown by the diagram, eligibility is not a perfect predictor of old age grant receipt, that is, age does not deterministically affect the probability of receiving an old age grant. For example, we note from the data that some women who were not yet age eligible nonetheless reported receiving an old age grant. This could be due to reporting or recording errors in the survey. In addition to early take up, delays in take up are also observed, probably because of delays in the application process or the fact that the respondent does not meet the means test at the time of eligible age.

About 30 percent of our estimation sample (309 out of 1059) report being in employment status, 6 percent of whom also report receipt of pension. This could be the case of reporting

or recording errors that we have alluded to already. We expect those who receive pension to be pension age eligible and are therefore retired from employment as well. Even if this is the case the negative shock due to loss of income from employment seems negligible because there were not many employed women who attained pension eligibility status in our sample. The fuzzy regression discontinuity design accommodates such types of errors by assuming that receipt of pension is probabilistic and proceeds to adjust for them.

Because take up is not universal, it is difficult to infer what the effect of the old age grant would be based solely on the intent to treat analysis[2]. The applicable design in such a case is the fuzzy regression discontinuity. In a fuzzy regression discontinuity design, the treatment assignment indicator (W) determines treatment status, but only probabilistically (Moscoe et al., 2015). The probability of treatment does not switch from 0 to 1 at the threshold. In this design we weight the jump in the regression of the outcome on the covariate with the probability of treatment on each side of the cut-off (Imbens and Lemieux, 2008). The treatment effect in this case is analogous to the complier average causal effect (CACE) in a Randomized Control Trial (RCT). Formally the estimand is

$$ACE = \frac{lim_{x\downarrow c}E[h|A=c] - lim_{x\uparrow c}E[h|A=c]}{lim_{x\downarrow c}E[W|A=c] - lim_{x\uparrow c}E[W|A=c]} \tag{5.2}$$

Where ACE is average causal effect, h is the health outcome, A is the forcing variable (age in this case), c is the threshold age of 60 years, and W is an indicator variable for eligibility for an old age grant. The numerator in the above equation is the intent to treat which is scaled by the probability of treatment (denominantor) on either side of the threshold. Our interest is to investigate what happens at the threshold age. We have therefore restricted the age variable to 5 years on either side of the threshold age, which accommodates the age threshold of 60 for women. This also allows us to use a non parametric approach to estimate the treatment effect. When using the non parametric approach, we do not worry about the functional form of the forcing variable.

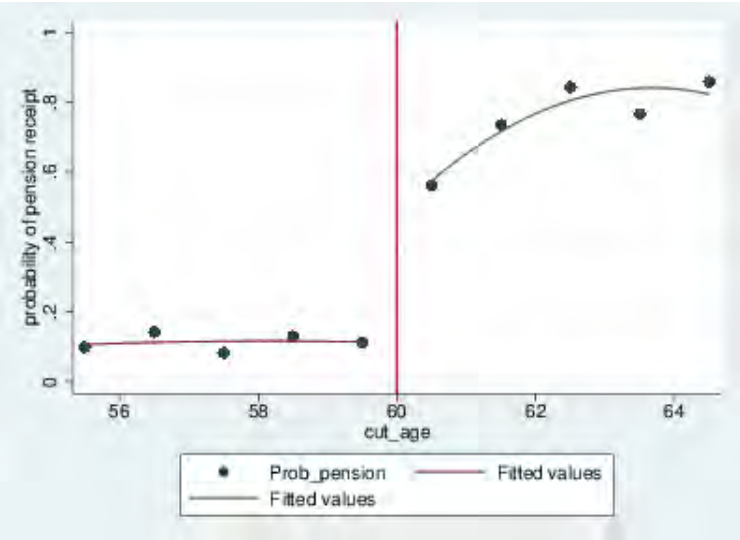
5.4.3 Outcomes by the forcing variable

In this section we present scatter diagrams for different health outcomes with age. Figure 5.3 shows the mean of SAH categories for women in our data set against age. From the figure there is no visible discontinuity at the threshold for mean SAH. On the other hand the mean of the raw SAH categories for women shows a declining[3] trend but as soon as the women receive the old age grant the trend changes direction. This is a pointer that receiving income

[2]Intent to treat measures the effect of treatment on those who are eligible irrespective of whether they receive the treatment or not

[3]Declining SAH implies worsening health condition with age and vice versa.

Figure 5.2: Probability of receiving old age grant for South African women



Notes: This is a first stage plot. The figure plots the probability of treatment on the Y axis and age on the X axis. The probability of receiving an old age grant jumps at age 60 years from roughly 0.1 to 0.5. This suggests the use of fuzzy regression discontinuity design.

may have long term effects as opposed to short term effects on health.

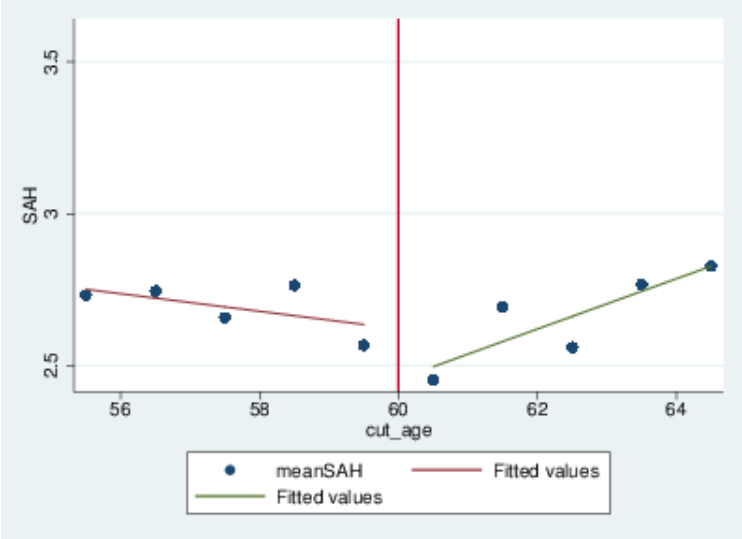
Figure 5.4 presents scatter plots for mental health (depression) measured by the CES-D score against age for women in the age range 55 to 65 years. From the figure, there is evidence of discontinuity at the threshold age of 60 years. Depressive symptoms generally decrease with age for women. Besides the discontinuity the slope of the relationship also increases after the threshold.

Figure 5.5 presents scatter plots for BMI score for women in NIDS by age for the age range 55 to 65 years. As the figure shows there is no visible discontinuity at the threshold age. However the trend for BMI seems to change direction after the threshold age, showing that BMI score increases with age before the threshold age but initially increases and then decreases with age after the threshold age.

5.4.4 Density of the forcing variable

We checked that the outcome is a smooth function of the assignment variable (age). Figure 5.6 shows the proportion of individuals at each age (bin). It appears reasonable to assume that there is no discontinuity in the assignment variable (age) at the eligible age of 60 for women in our sample. In particular there is no discontinuity at age 60 for women. This also confirms that people do not manipulate their ages to become eligible for an old age grant. We also check for discontinuities in other covariates such as education. If there is any discontinuity in covariates at the eligible age, this would cast doubt on the identification strategy since the results could be driven by unobserved confounders.

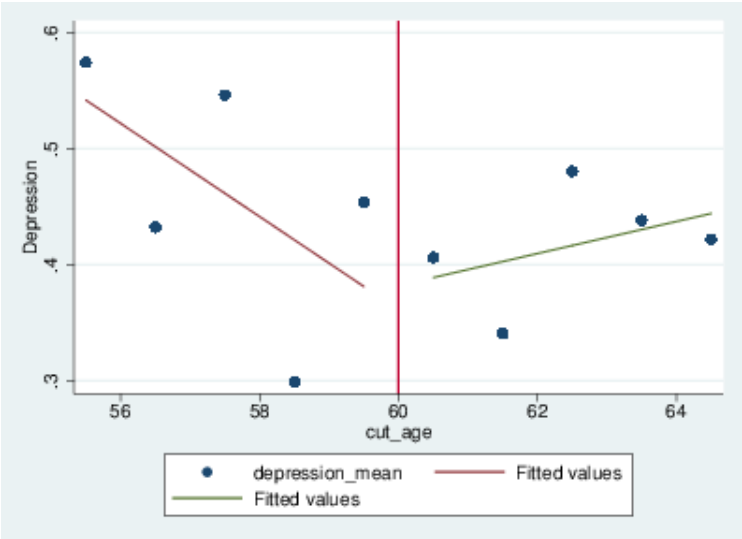
Figure 5.3: Mean SAH by age for Women in NIDS (age 55 to 65 years)



Notes:

Source : Own calculation from NIDS data
Scatter points denote averages for SAH at each age for women. There are 5 bins on each side of the threshold age of 60. The linear fit is estimated separately on both sides of age 60.

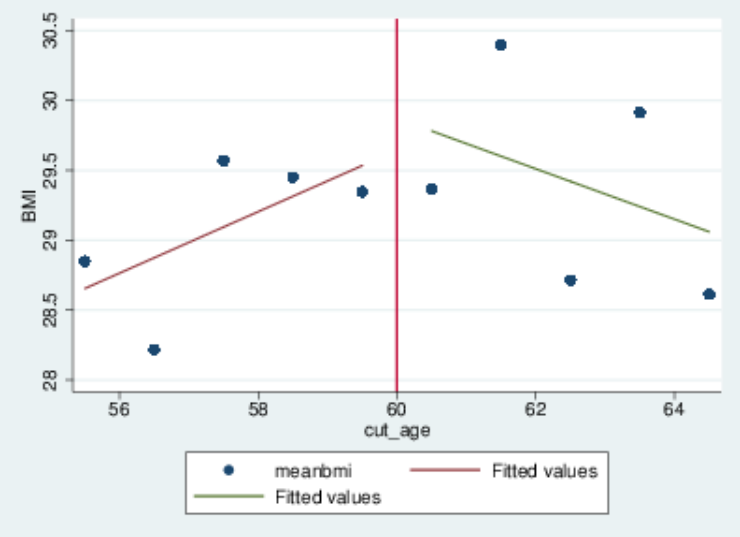
Figure 5.4: Mean CES-D score by age for Women in NIDS (age 55 to 65 years)



Notes:

Source : Own calculation from NIDS data
Scatter points denote averages for CES-D (depression) score at each age for women. There are 5 bins on each side of the threshold age of 60. The linear fit is estimated separately on both sides of age 60.

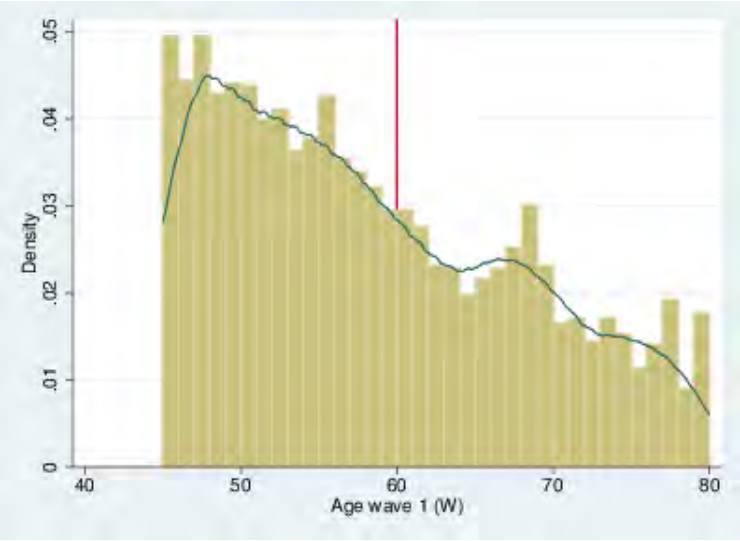
Figure 5.5: BMI score for Women in NIDS (age 55 to 65 years)



Notes:

Source : Own calculation from NIDS data
Scatter points denote averages for BMI score at each age for women. There are 5 bins on each side of the threshold age of 60. The linear fit is estimated separately on both sides of age 60.

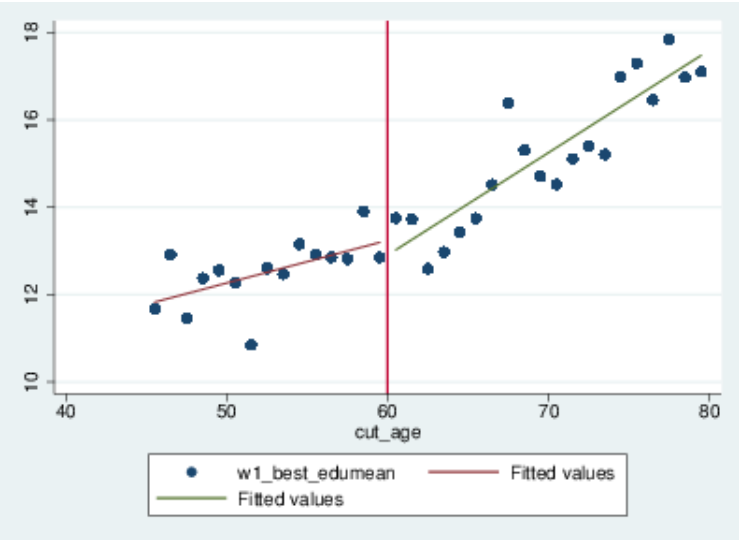
Figure 5.6: Assignment variable (Age) frequency for Women in NIDS



Notes:

Source : Own calculation from NIDS data
The histogram shows population density on the vertical axis and age on the x-axis. The histogram shows that the forcing variable (age) is smooth at the threshold age of 60 for women. There is no humping observed around the threshold.

Figure 5.7: Years of schooling attained by age



Notes:

Source : Own calculation from NIDS data

Scatter points denote average years of schooling attained at different ages. There is a general increase in average years of schooling as age increases. There is no discontinuity at age 60 for years of schooling. The linear fit is estimated separately on both sides of age 60.

5.4.5 Covariates by forcing variable

Some of the covariates of health include education (years of schooling) and marital status. We plot some of the covariates against age to determine whether the discontinuity observed at the threshold could be explained by the covariates. In figure 5.7 we plot average years of schooling against health outcomes at different values of the forcing variable (age). We have improved the visual clarity of the plots by adding a smoothed regression line on both sides of the threshold for women at age 60. The figure confirms that there is no discontinuity in years of schooling at the threshold age of 60 for women. This further confirms that any discontinuity in health outcomes at the threshold age cannot be as a result of differences in years of schooling.

5.4.6 Choice of bandwidth

A crucial decision in RDD is the choice of the bandwidth (Eibich, 2014). It determines which observations are used in the estimation by setting the maximum distance from the discontinuity. A small bandwidth minimizes bias[4] however the variance might be big due to a small sample (n) while a large bandwidth (h) minimizes the variance (Lee and Lemieux, 2009; Lee and Lemieuxa, 2010). Previous studies investigating the effect of retirement on health have considered a bandwidth of 10 years for the main specification (Moreau and

[4] $Bias = \frac{h^2}{2}\mu_2 m''$

Stancanelli 2012). This leads to an estimation sample of age 50 to 70 years. 10 years before the discontinuity of 60 years and 10 years after the discontinuity. We will however use 5 years on either side of the discontinuity, as in Gordon and Miller (2012). As a robustness check however we will try different bandwidths.

5.5 Results

One salient feature of RDD is that even before the regression results are presented we have a feel of the results presented diagrammatically. The regression results therefore serve to give statistical meaning to the scatter diagrams already discussed. We apply this analysis to three dimensions of health outcomes, namely self assessed health (SAH), body mass index (BMI) and Depression (mental health) to tease out the causal effect of income if it exists. The regression results indicate that at the pension threshold age of 60 years for women in South Africa, health outcomes do not show any discontinuity, that is, the health dimensions we analyze are not affected contemporaneously by income.

We present the first stage results in table 5.2. These results show that at the old age grant eligible age of 60 years for women, the probability of receiving the grant (pension) is 45.5 percentage points higher than just below this age. Since we have used a linear probability model, we do not need to exponentiate the coefficients. The constant in this regression (0.09) indicates that just below this old age grant threshold age the probability of receiving an old age grant is 9 percentage points regardless of the eligibility age (also called percent of always-takers).

The other coefficients in the first stage regression are agediff (that is, age – 60) which is the age centered variable, and the interaction between it and the eligibility indicator. The centering is done to make them be interpreted as slopes, while the interaction allows the interpretation on either side of the threshold.

As we can see from table 5.2, for each additional year, the probability of receiving pension is 0.07 percent for women below the cut off point. However, this increase in probability is statistically insignificant. This implies that for these women below the cut off age of 60, an additional year does not increase their chances of getting old age grant, at least not until they cross the cut off age of 60. These results confirm the evidence in figure 5.2. In fact like in figure 5.2, the regression results in table 5.2 show that the probability of receiving old age grant is constant at 9 percentage points until women reach the threshold age of 60.

The coefficients on these variables are interpreted as follows; the coefficients on agediff (that is, age – 60) shows that for women below the old age grant threshold age of 60, the probability of receiving an old age grant for an additional year is zero (this is the slope of the

Table 5.2: First Stage regression

VARIABLES	pension
ageDiff	0.000742
	(0.00867)
1.eligibility	0.455***
	(0.0722)
1.eligibility#c.ageDiff	0.0728***
	(0.0205)
Constant	0.0915***
	(0.0322)
Observations	1,037
R-squared	0.434

Notes: The table shows first stage regression results where receipt of old age grant is the dependent variable and 1.eligibility is the dummy variable indicating that the individual is age 60 and over (eligible). ageDiff is an age-centered variable. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

fitted line when age is less than the threshold age). This implies that although the probability of receiving the old age grant is 9 percentage points for women below the threshold age, an additional year does not increase this probability for these women. The coefficient on the agediff and its interaction with the indicator of eligibility show that for women, any additional year above the age of 60 the probability of receiving old age grant is 7.35 percentage points (0.0728+0.000742).

Table 5.3 presents the reduced form regression results. In the reduced form regressions, the health outcomes are regressed on eligibility, centered age variable and their interactions. We use three health outcomes, namely SAH, depression and BMI in these models. SAH and depression are dichotomized while BMI is used as a continuous variable. The results for all the health outcomes analyzed show that the health outcomes of women who are just above the threshold age (eligible) are not statistically significantly different from those just below the threshold age (ineligible). The eligibility variable is the indicator variable showing that an individual is either eligible or ineligible based on their age. Women are eligible for an old age grant when they are 60 years and older. The coefficient of this variable captures the intent to treat estimate. The women who are just above the threshold age ($eligibility == 1$) have 9 percentage points lower probability of having better SAH, 11.4 percentage points higher probability of having depressive symptoms and one BMI score lower than those just below the threshold age. These effects are however not statistically significant.

The constant terms in the reduced form regressions are interpreted as follows: Women who are just below the threshold age have mean SAH of 0.63 and mean depression of 0.24. This means that 63 percent of the women in our estimation sample who were just below the threshold age had better health while 24 percent of them were depressed. These women (those just below the threshold) also had a mean BMI (weight for height) score of 30.

The probability of having better SAH increases by 0.721 percentage points for every addi-

Table 5.3: Reduced Form Regressions

VARIABLES	(1) SAH	(2) depression	(3) BMI
1.eligibility	-0.0912	0.114	-1.164
	(0.0916)	(0.0885)	(1.257)
ageDiff	0.00721	-0.0611***	0.402
	(0.0213)	(0.0218)	(0.265)
1.eligibility#c.ageDiff	-0.0136	0.0759***	-0.100
	(0.0288)	(0.0282)	(0.410)
Constant	0.634***	0.240***	30.29***
	(0.0734)	(0.0744)	(0.883)
Observations	1,028	1,036	877
R-squared	0.007	0.020	0.012

Notes: The table shows reduced form regression results where health outcomes are the dependent variables. SAH and depression are binary dependent variables while BMI is a continuous score. 1.eligibility is a dummy variable indicating that the individual is aged 60 and over (eligible). ageDiff is an age-centered variable. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

tional year in age just below the threshold age. BMI increases by 0.402 scores for women just below the threshold age for every additional year. The probability decreases by 6 percentage points for depression for every extra year for women just below the threshold age. On the other hand, for women just above the threshold age, every additional year decreases probability of reporting better health by 0.00639 (0.00721-0.0136) and increases the probability of reporting higher BMI by 0.302 (0.402-0.100). The probability of depressive symptoms increases by 0.0148 (0.0759-0.0611) for every additional year for women just above the threshold. Only the slope coefficients on either side of the age threshold for depression are significant.

From the first stage and the reduced form regression results we can derive the Wald estimators (instrumental estimator). This is done by dividing the coefficient of eligibility in the reduced form regression by that of the first stage regression. However, in order to get analytical standard errors for the complier average causal effects of income on health outcomes, we estimate it by two-stage least squares, using eligibility ($1[age_i > 60]$) as an instrument for receipt of old age grant. The point estimates are identical to the ones obtained by dividing intent to treat estimates (reduced form) with first stage coefficients on the eligibility variable. Table 5.4 presents the results for the three health outcomes.

The results show that the treatment (receipt of old age grant), instrumented by eligibility $1[age_i > 60]$, has no causal effect on health outcomes. This is shown by the fact that the coefficients on the old age pension variable (receipt of old age grant) for all three health outcomes are not statistically significant. This implies that women who receive an old age pension do not have better health outcomes than those who do not receive an old age pension. This could also show that the effect of income on health outcomes is not contemporaneous.

Table 5.4: Instrumental Variable Regression Results

VARIABLES	(1) SAH	(2) depression	(3) BMI
Old age pension	-0.203	0.250	-2.357
	(0.196)	(0.191)	(2.662)
1.eligibility#c.ageDiff	0.000725	0.0577**	0.0820
	(0.0293)	(0.0278)	(0.485)
ageDiff	0.00781	-0.0613***	0.398
	(0.0215)	(0.0221)	(0.263)
Constant	0.654***	0.217**	30.50***
	(0.0880)	(0.0905)	(1.074)
Observations	1,028	1,036	877
R-squared	0.031	0.019	

Notes: The table shows instrumental variable regression results where receipt of old age grant is instrumented by a dummy indicating that the individual is aged 60 and over (1.eligibility). ageDiff is an age-centered variable. SAH and depression are binary dependent variables while BMI is a continuous score. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Health is a state variable which depreciates gradually and also responds gradually to income shocks. Individuals need to invest in health over time. Any effect would then occur with a lag. Our identification strategy however is only able to identify short term effects.

We undertake a robustness check on depression (mental health) by using alternative measurements for depression. This is because the results for the slopes of the fitted lines of this health outcome on either side of the eligibility threshold show significant coefficients. For example, women just above the old age grant eligibility threshold have the probability of being depressed of 0.0036 (0.0577-0.0613) lower for every additional year in age. Those just below the threshold age have a 0.0613 lower probability of depression for every additional year. The coefficients on the age centered variables measure slope and they indicate that the probability of being depressed drops faster per year for women just below than the women just above the threshold age. These slope coefficients for depression are significant.

Table 5.5 shows reduced form results for alternative measures of depression, while table 5.6 presents the instrumental variable results for these alternative measures of depression. The first column in each of these tables reproduces the depression results already presented in the previous analysis. In column 2, depression is measured as a continuous score computed as explained in chapter two (CES-D 10 questions are used to derive the score). In column 3, a binary variable is shown which is created from CES-D 8 questions after excluding two positive questions, as explained in chapter two. The last column shows the use of this score (the CES-D 8) as a continuous variable. In all cases income has no causal effect on depressive symptoms for women at the age of 60 years. In the Instrumental regression results the receipt of a pension is instrumented by the eligibility variable $1[age_i > 60]$.

Table 5.5: Reduced form regressions for different measures of depression

VARIABLES	(1) depression Binary	(2) CESD_10 Index	(3) depression_CESD_8 Binary	(4) CESD_8 Index
1.eligibility	0.114	1.117	0.0409	0.674
	(0.0885)	(1.236)	(0.0736)	(0.987)
ageDiff	-0.0611***	-0.687**	-0.0306	-0.489
	(0.0218)	(0.315)	(0.0208)	(0.299)
1.eligibility#c.ageDiff	0.0759***	0.909**	0.0291	0.639*
	(0.0282)	(0.371)	(0.0233)	(0.338)
Constant	0.240***	6.863***	0.143**	4.689***
	(0.0744)	(1.010)	(0.0659)	(0.863)
Observations	1,036	1,036	1,036	1,036
R-squared	0.020	0.022	0.011	0.014

Notes: The table shows reduced form regression results where alternative measures of depression are the dependent variables. 1.eligibility is the dummy variable indicating that the individual is aged 60 and over (eligible). ageDiff is an age-centered variable. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 5.6: Instrumental variable regression results for depression

VARIABLES	(1) depression Binary	(2) CESD_10 Index	(3) depression_CESD_8 Binary	(4) CESD_8 Index
old age pension	0.250	2.453	0.0899	1.480
	(0.191)	(2.630)	(0.161)	(2.138)
1.eligibility#c.ageDiff	0.0577**	0.730**	0.0226	0.531*
	(0.0278)	(0.334)	(0.0186)	(0.288)
ageDiff	-0.0613***	-0.689**	-0.0307	-0.490
	(0.0221)	(0.318)	(0.0209)	(0.301)
Constant	0.217**	6.639***	0.134*	4.554***
	(0.0905)	(1.226)	(0.0796)	(1.042)
Observations	1,036	1,036	1,036	1,036
R-squared	0.019	0.050	0.016	0.029

Notes: The table shows instrumental variable regression results where receipt of old age grant is instrumented by a dummy indicating that the individual is age 60 and over (1.eligible). ageDiff is an age-centered variable. Standard errors in parentheses. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

5.6 Conclusion

In this chapter we examine the effect of income on health outcomes of women aged 60 years. We use, a robust quasi-experimental approach, the Regression Discontinuity Design, applied on data from NIDS. In particular we analyze whether health outcomes change discontinuously as a result of the income shock that comes in the form of receipt of an old age grant that is given to women at the age of 60 in South Africa. We establish that the changes in health outcomes due to income that occur at age 60 for South African women are not statistically significant. The approach we use captures contemporaneous effects if they exist. This implies that income does not have any contemporaneous effect on these health outcomes. Our suspicion that income could potentially affect depressive symptoms due to the expectations that come with such income in a household prompts us to use different measurements of depression. However, the results did not change, they remained insignificant.

This is an important finding for policymakers in South Africa interested in improving women's health using cash transfers. We have shown that cash transfers do not translate into short term health effects. The major limitation with our study is that these results relate only to women at the age of 60 and we cannot generalize to women of other ages or other groups such as children or men. Also, health is a state variable, and like capital, needs to be invested in over time, and therefore we may not witness any discontinuity in the short term. We cannot therefore rule out that income effects on health outcomes are long term as our strategy captures only contemporaneous health effects. Future studies could look at the effect of income on long term health outcomes.

Chapter 6

Summary of Findings, Recommendation for Further Research and Conclusion

6.1 A summary of Findings

The main aim of this thesis is to analyze socioeconomic related inequalities in health outcomes. In particular we seek to identify the socioeconomic variables that determine health outcomes in South Africa. We also establish the extent of income related health inequality and use socioeconomic determinants to isolate the factors that contribute to income related health inequalities. The thesis also seeks to establish whether the correlation between health outcomes and income can be interpreted as causal, at least in the short run. We start by verifying the quality of the data and testing for attrition bias to ensure that the characteristics of the data do not influence the results. The general empirical literature on this subject and specifically in the context of Africa is inconclusive on several vital questions. This inconclusiveness is partly because of the diversity of measurement approaches for health outcomes and the tools used to estimate inequality. We have applied robust approaches on NIDS data, in scaling health outcomes, estimating health inequality, and establishing the causal effect of income on health. The literature favours the analysis of socioeconomic related health inequalities instead of inequalities related to other dimensions because the former is viewed as the source of health inequity.

The main empirical findings are chapter specific and are summarized within the respective empirical chapters. This section summarizes the empirical findings in each of these chapters. In chapter two we examine the pattern of attrition in NIDS, then formally test for its potential effects on parameter estimates of health outcome models. In Chapter three , we establish the socioeconomic determinants of three health outcomes (SAH, BMI and depression) using

bivariate and panel data approaches. In t Chapter four, we estimate an index of income related health inequality and decompose it into the socioeconomic factors contributing to the index as well as to changes in the index over time. Finally, in Chapter five we extend the analysis by investigating whether the significant correlations between health and income are causal.

We describe the pattern of attrition and test for its potential effects on the parameter estimates using two tests, the attrition probits and the Becketti, Gould, Lillard and Welch (BGLW) tests. These are two well known tests for attrition in panel data. In the attrition probit test, we verify if the health variables are significantly correlated with the probability of attrition, after controlling for other covariates. If the health outcome variable is significant in an attrition probit model, estimating a health outcome model while ignoring attrition would bias the parameter estimates. The BGLW test is the reverse of the attrition probit test. Here we check if the attrition indicator is significantly correlated with the health outcome in a health outcome model. Fitzgerald et al. (1998) provide the theoretical insights for these tests and propose that in case there is attrition, consistent and unbiased estimates could still be estimated provided we re-weight the model with the inverse of the probabilities of attrition. We undertake a sensitivity analysis where we re-weight the health outcome models with the inverse of attrition probabilities and find that the parameter estimates do not change. This confirms that attrition is random with respect to the health outcomes we analyze.

We select three dimensions of health in the dataset for this investigation, self assessed health (SAH), body mass index (BMI) and mental health (Depression). These health dimensions represent both subjective and objective measures of health outcomes. SAH is considered a subjective measure of general health while BMI and Depression are considered objective measures even though depression is derived from subjective responses. We find that although the attrition rates are high, 21% in wave 2 and 20% in wave 3, parameter estimates obtained from these health outcomes are not affected by attrition bias in wave 2 and wave 3 of the survey. The attrition is random with respect to these health outcomes. This means that we can use these health outcomes as dependent variables in regressions in waves 2 and 3 without necessarily re-weighting them using inverse probabilities and still obtain unbiased and consistent parameter estimates.

A robustness check confirmed that even after re-weighting using the inverse of the probabilities of attrition, the parameter estimates obtained are not different from those obtained without re-weighting for each health dimension. This implies that attrition bias does not affect the parameter estimates obtained using the three health dimensions in wave 2 and wave 3 of the NIDS.

In Chapter 3 we build on the findings of Chapter 2 to establish the correlates of the health outcomes using bivariate and panel data approaches. Panel data models include the FE and

the RE models. We also used the pooled OLS for comparison. The panel nature of the NIDS dataset allows the use of these models to establish the significant correlates of health outcomes and in the process control for unobserved heterogeneity among individuals. We follow the Grossman model for health that provides a theoretical foundation for the relationship between health outcomes and socioeconomic correlates such as income, education, age, race, and marital status, as well as geography.

The measurement of health outcomes varies from one health dimension to the other. In all cases we follow best practice in the literature in establishing the health variable as well as the appropriate model. For example, SAH was collected as an ordinal variable. We dichotomize this to a binary variable to establish its covariates using panel data estimators. Panel data analysis does not work well with ordered variables because they are prone to incidental parameter bias when they are used in panel models. BMI and mental health on the other hand are measured as continuous scores.

An important finding in this chapter is that income and education are significant correlates of health outcomes in South Africa. In particular, these two factors are positively correlated with SAH and BMI scores and negatively correlated with mental health scores. These factors are unevenly distributed in South Africa. This raises an important question as to what extent the inequality in these factors drive the uneven distribution of health outcomes. This forms the subject of the investigation in Chapter four of this study.

In Chapter four we estimate the extent to which health outcomes differ across income levels. We use SAH to estimate the extent of income related health inequality because it is a measure of general health that has received a lot of support in the international literature. This variable was re-scaled before being used to estimate income related health inequality. We identify and explain the socioeconomic factors associated with income related health inequality. The NIDS data allows an investigation of whether the differences in health outcomes are narrowing or widening over time. This is shown to be the case, despite the four year time period between waves being short for this analysis. We identify socioeconomic factors associated with the change in the distribution of health outcomes notwithstanding the short period between waves. The extent of income related health inequality was ascertained by the use of the corrected concentration index, which is specifically suited for bounded health outcome variables such as SAH.

The corrected concentration index reveals that good health in South Africa is distributed in favour of the rich. The decomposition of the inequality index shows that income related health inequality is explained mainly by income and educational attainment although other factors like age serve to reduce it. The corrected concentration index decreased from wave 1 to wave 3 implying that income related health inequality narrowed over this period. This

is attributed in a large part to unexplained factors (59%), but changes in the distribution of the covariates explain 21% and changes in the coefficients explain 20% of the narrowing of income related health inequality over this period.

In Chapter five we build on the results in chapter 3 and 4, to investigate whether the significant correlations between health outcomes and income are causal. As we noted in chapter three, if the covariates of health outcomes are unevenly distributed, we expect health outcomes to also be unevenly distributed. However, this depends on whether the relationship is causal . To establish the causal effect of income on health outcomes, we use the old age grant for women in South Africa. The old age grant provides a natural experiment that we utilize to identify this effect. This is a social grant given to women in South Africa when they attain the age of 60 years (threshold age) and can satisfy a means test. We apply the Regression Discontinuity Design which identifies the causal effect if it exists at the threshold age. If there is a discontinuity in health outcomes at the threshold age, this must be caused by the old age grant because the women just below and just above are assumed to be similar in all other respects.

We find that income does not have any contemporaneous causal effect on SAH, BMI and depression, for women aged 60 years in South Africa. These results do not change even when we use alternative measurements of depression. We argue that this could be because health is a state variable that depreciates slowly and may also respond to income with a lag. Our identification strategy however is only internally valid and only able to identify contemporaneous effects.

6.2 Recommendation for Further Research

In this section we highlight areas for further research as well as some of the challenges faced and make proposals on how these challenges can be addressed. Wagstaff et al. (2003) argue that it is only when the linear health production function is causal that we can interpret the decomposition of income related health inequality as causal. This means that although income and educational attainment were found to be significant correlates of health outcomes, this relationship cannot be interpreted as causal. The parameter estimates from an OLS estimation of health production function with income and health as explanatory variables cannot be interpreted as causal because income, education and health are endogenous. The results of this study therefore point to the need for further research with an identification strategy that can tease out causal relationships between health outcomes, income and education. We use a quasi experimental approach, the Regression Discontinuity Design, in this study. However, this is not externally valid as the results only apply to individuals at the threshold age. One way in which the problem of endogeneity can be solved is by using instrumental variables

whenever they are available. Another proposal is to use the effect of parental income on the health of their children, with the assumption that childrens' health does not affect the income of their parents.

Secondly, the multidimensional nature of health makes its measurement in surveys difficult. The World Health Organization (WHO) defined health as the state of complete physical, mental and social well-being and not just the absence of disease and infirmity. It is difficult to get an objective measure of health that encompasses all its dimensions, that is, physical, mental and social well-being. The SAH is the closest variable that presumably summarizes an individuals health condition. SAH is always collected qualitatively which makes its use in the analysis of socioeconomic related health inequalities limited because the estimation tools for health inequality require cardinal or ratio scale variables. We overcame this challenge by transforming the ordinal SAH to cardinal scale. This may have affected the standard error of the index computed from the transformed variable. It may be necessary for future studies to utilize cardinal measures of health such as SF-36, Euroqo 15D or Health Utility Index (HUI) which are true measures of health status. Such measures are collected in health surveys.

Thirdly, the choice of a ranking variable has been shown in the literature to influence interpretation of the extent of health inequality. It is therefore important for a study to explicitly state the ranking variable. Perhaps the results of more than one ranking variable should be provided in future studies to describe different aspects of inequality. This could offer a broader comparative perspective.

Fourthly, this thesis identifies socioeconomic factors associated with income related health inequality. However, apart from educational attainment, the thesis does not discuss the transmission mechanisms between the other socioeconomic variables and health outcomes. This is partly because the mechanisms that link socioeconomic variables and health outcomes differ with social status and need close attention. For example, it may be that the poor suffer from malnutrition because they cannot afford a balanced diet, while the rich suffer from obesity because they eat unhealthy foods. Although the BMI of both groups is a product of income, the transmission mechanism could be different. For this reason an investigation requires an in-depth and independent analysis that is outside the scope of this thesis. We recommend that a future study focuses on the transmission mechanism of socioeconomic factors related to health outcomes.

Finally, the use of the regression discontinuity design to estimate the effect of income on health outcomes. This only captures contemporaneous effects, yet the effect of income on health outcomes may be long term. For this reason future studies could explore the long term effects of income on health outcomes.

6.3 Conclusion

In conclusion, determining the presence of health inequality is a first step in identifying inequity in health. Health inequalities can be attributed to many factors. This includes choices, such as choosing to consume alcohol or smoking, biological variations (genetics), the physical environment and factors beyond the control of the individuals concerned. Other factors include age differences that lead to differences in physical capabilities between the elderly and the young. In addition, inequalities also arise due to differences in socioeconomic circumstances. This last source of inequality formed the focus of this thesis partly because some of the socioeconomic conditions are amenable to change through appropriate policy. The contribution of this thesis was to quantify the socioeconomic factors that drive income related health inequalities. This could be a first step for policy-makers to reduce inequalities in health outcomes.

The use of robust methods to estimate socioeconomic related health inequalities, as well as the effect of income on health outcomes ensured that the results are robust to different measurements of health outcomes. The corrected concentration index, for example, was appropriate because it provided an accurate assessment of inequality in health outcome across the distribution of income groups. The index revealed substantial inequalities in health, showing the health disadvantages of the poor in South Africa. One of this index's desirable properties is that, it does not vary depending on whether good health or ill-health is the measure of health outcome. The decomposition reveals that income related health inequality is explained by factors that are avoidable, such as low income and poor educational attainment. As we discussed in the overview of health care in South Africa, the government has identified vulnerable groups, such as women and children, and provides them with free medical care. The findings of this thesis show that proper analysis should be done in identifying vulnerable groups to assist the poor and uneducated enjoy better health.

Although there is clearly a need for further research on socioeconomic related health inequalities, we have contributed to the literature by identifying opportunities for further reduction of income related health inequalities. We have also demonstrated that new and robust techniques can be applied to an analysis of socioeconomic circumstances to provide insight into socioeconomic related health inequalities in the context of low income countries.

Bibliography

Adams, P., Hurd, M. D., McFadden, D. L., Merrill, A., and Ribeiro, T. (2004). Healthy, Wealthy, and Wise? Tests for direct causal paths between health and socioeconomic status. In *Perspectives on the economics of aging*, pages 415–526. University of Chicago Press.

Adler, N. E. and Newman, K. (2002). Socioeconomic disparities in health: pathways and policies. *Health affairs*, 21(2):60–76.

Alderman, H., Behrman, J. R., Kohler, H.-P., Maluccio, J. A., and Watkins, S. C. (2001). Attrition in longitudinal household survey data. *Demographic research*, 5(4):79–124.

Araar, A., Levine, S. A., and Duclos, J.-Y. (2009). Body Mass Index, Poverty and Inequality in Namibia: Paper Prepared for the Central Bureau of Statistics,.

Ardington, C. & Case, A. (2009). Health: Analysis of the NIDS wave 2 dataset. NIDS discussion paper, no. 2.

Ardington, C. and Gasealahwe, B. (2012). Health: Analysis of the NIDS wave 1 and 2 datasets.

Argent, J. (2009). Household income: Report on NIDS wave 1, technical paper no. 3.

Ataguba, J. (2013). Inequalities in multimorbidity in South Africa.

Ataguba, J. E., Akazili, J., and McIntyre, D. (2011). Socioeconomic-related health inequality in South Africa: evidence from General Household Surveys.

Au, N. and Johnston, D. W. (2013). An econometric analysis of self-assessed health: what does it mean and what is it hiding? Technical report, HEDG, c/o Department of Economics, University of York.

Baigrie, N. and Eyal, K. (2014). An evaluation of the determinants and implications of panel attrition in the National Income Dynamics Survey (2008-2010). *South African Journal of Economics*, 82(1):39–65.

Bailey, K. and Ferro-Luzzi, A. (1995). Use of body mass index of adults in assessing individual and community nutritional status. *Bulletin of the World Health Organization*, 73(5):673.

Baltagi, B. (2008). *Econometric analysis of panel data*, volume 1. John Wiley & Sons.

Becker, G. S. (1975). Front matter, human capital: a theoretical and empirical analysis, with special reference to education. In *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, 2nd ed.*, pages 22–0. NBER.

Becketti, S. et al. (1988). The panel study of income dynamics after fourteen years: An evaluatio n. *Journal of Labor Economics*, 6(4):472–92.

Benyamini, Y., Leventhal, E. A., and Leventhal, H. (1999). Self-assessments of health what do people know that predicts their mortality? *Research on aging*, 21(3):477–500.

Bhorat, H., Oosthuizen, M., and Van Der Westhuizen, C. (2012). Estimating a poverty line: An application to free basic municipal services in South Africa. *Development Southern Africa*, 29(1):77–96.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, pages 436–455.

Booysen, F. (2003). Urban rural inequalities in health care delivery in South Africa. *Development Southern Africa, 20:5,659-673.*

Bound, J., Brown, C., and Mathiowetz, N. (2001). Measurement error in survey data. *Handbook of econometrics*, 5:3705–3843.

Bradshaw, D. (2008). Determinants of health and their trends: Primary health care: in context. *South African Health Review*, pages 51–69.

Braveman, P. and Gottlieb, L. (2014). The social determinants of health: It's time to consider the causes of the causes. *Public health reports (Washington, DC: 1974)*, 129:19–31.

Brown, M., Daniels, R. C., De Villiers, L., Leibbrandt, M., and Woolard, I. (2012). National Income Dynamics Study wave 2 Beta Release User Manual. Cape Town.

Burger, R., Bredenkamp, C., Grobler, C., and van der Berg, S. (2012). Have public health spending and access in South Africa become more equitable since the end of apartheid? *Development Southern Africa*, 29(5):681–703.

Bynum, B. (2008). The McKeown thesis. *The Lancet*, 371(9613):644–645.

Case, A. (2002). Health, income and economic development. In *Proceedings of the World Bank Conference on Development Economics, 2001/2002*, pages 221–41.

Case, A. (2004). Does money protect health status? Evidence from South African pensions. In *Perspectives on the Economics of Aging*, pages 287–312. University of Chicago Press.

Case, A. and Deaton, A. (2005). Health and Wealth among the Poor: India and South Africa Compared. *The American Economic Review*, 95(2):229.

Chang, W. (2002). The meaning and goals of equity in health. *Journal of epidemiology and community health*, 56(7):488–491.

Charasse-Pouélé, C. and Fournier, M. (2006). Health disparities between racial groups in South Africa: a decomposition analysis. *Social Science & Medicine*, 62(11):2897–2914.

Clarke, P. M., Gerdtham, U.-G., Johannesson, M., Bingefors, K., and Smith, L. (2002). On the measurement of relative and absolute income-related health inequality. *Social Science & Medicine*, 55(11):1923–1928.

Contoyannis, P. and Forster, M. (1999). The distribution of health and income: a theoretical framework. *Journal of Health Economics*, 18(5):605–622.

Contoyannis, P., Jones, A. M., and Rice, N. (2004). The dynamics of health in the British Household Panel Survey. *Journal of Applied Econometrics*, 19(4):473–503.

Correa-Valez, I., Grace, L., Eliana, J. S., and Brown, R. (2011). Socioeconomic Status and Health Outcomes in a Developing Country.

Crandell, D. L. and Dohrenwend, B. P. (1967). Some relations among psychiatric symptoms, organic illness, and social class. *American Journal of Psychiatry*, 123(12):1527–1538.

Crossley, T. F. and Kennedy, S. (2002). The reliability of self-assessed health status. *Journal of Health Economics*, 21(4):643–658.

CSDH (2008). Closing the gap in a generation: health equity through action on the social determinants of health. Final report of the Commission on Social Determinants of Health. Technical report, Geneva, World Health Organization.

Cutler, D. M. and Lleras-Muney, A. (2006). Education and health: evaluating theories and evidence. Technical report, National Bureau of Economic Research.

Cutler, D. M., Richardson, E., Keeler, T. E., and Staiger, D. (1997). Measuring the health of the US population. *Brookings papers on economic activity. Microeconomics*, pages 217–282.

Duflo, E. (2000). Child health and household resources in South Africa: Evidence from the old age pension program. *American Economic Review*, pages 393–398.

Eibich, P. (2014). Understanding the effect of retirement on health using regression discontinuity design.

Erreygers, G. (2009a). Correcting the concentration index. *Journal of health economics*, 28(2):504–515.

Erreygers, G. (2009b). Correcting the concentration index: a reply to Wagstaff. *Journal of Health Economics*, 28(2):521–524.

Erreygers, G. and Van Ourti, T. (2011). Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: a recipe for good practice. *Journal of Health Economics*, 30(4):685–694.

Ettner, S. L. (1996). New evidence on the relationship between income and health. *Journal of health economics*, 15(1):67–85.

Falkingham, J. and Namazie, C. (2002). Measuring health and poverty: a review of approaches to identifying the poor. *London: DFID Health Systems Resource Centre*.

Fé, E. and Hollingsworth, B. (2011). Estimating the effect of retirement on health via panel discontinuity designs. *Unpublished Manuscript, University of Manchester/University of Lancaster*.

Finlay, J. (2007). The role of health in economic development. *Harvard Initiative for Global Health [Online]*.

Fitzgerald, J., Gottschalk, P., et al. (1998). An Analysis of Sample Attrition in Panel Data. *Journal of Human Resources*, 33(2):251–299.

Fleurbaey, M. and Schokkaert, E. (2009). Unfair inequalities in health and health care. *Journal of health economics*, 28(1):73–90.

Fleurbaey, M., Schokkaert, E., et al. (2011). Equity in health and health care. *Handbook of health economics*, 2:1003–92.

Frijters, P., Haisken-DeNew, J. P., and Shields, M. A. (2005). The causal effect of income on health:Evidence from German reunification. *Journal of Health Economics*, 24(5):997–1017.

Goerg, S. J., Kaiser, J., and Bundesbank, D. (2009). Nonparametric testing of distributions–the epps-singleton two-sample test using the empirical characteristic function. *Stata Journal*, 9(3):454.

Gordon, D. and Miller, D. L. (2012). The south african pension program and the health of the elderly and their families: Regression discontinuity evidence from october household surveys.

Greene, W., Han, C., and Schmidt, P. (2002). The bias of the fixed effects estimator in nonlinear models. *Unpublished Manuscript, Stern School of Business, NYU*.

Grossman, M. (1972). On the concept of health capital and the demand for health. *The Journal of Political Economy*, 80(2):223–255.

Grossman, M. (1999). The human capital model of the demand for health. Technical report, National Bureau of Economic Research.

Grossman, M. (2000). The human capital model. *Handbook of health economics*, 1:347–408.

Hall, D. and Cole, T. (2006). What use is the bmi? *Archives of disease in childhood*, 91(4):283–286.

Halla, M. and Zweimüller, M. (2013). The effect of health on earnings: Quasi-experimental evidence from commuting accidents. *Labour Economics*, 24:23–38.

Hamilton, D., Sutherland, G., and Iacono, T. (2005). Further examination of relationships between life events and psychiatric symptoms in adults with intellectual disability. *Journal of Intellectual Disability Research*, 49(11):839–844.

Hamoudi, A. and Thomas, D. (2005). Pension income and the well-being of children and grandchildren: New evidence from South Africa. *California Center for Population Research*.

Harling, G., Ehrlich, R., and Myer, L. (2008). The social epidemiology of tuberculosis in South Africa: a multilevel analysis. *Social science & medicine*, 66(2):492–505.

Harris, B. (2004). Public health, nutrition, and the decline of mortality: the mckeown thesis revisited. *Social history of medicine*, 17(3):379–407.

Harrison, D. (2009). An overview of health and health care in South Africa 1994–2010: Priorities, progress and prospects for new gains. *Washington, DC: Henry J Kaiser Family Foundation*.

Hernández-Quevedo, C. and Jiménez-Rubio, D. (2009). A comparison of the health status and health care utilization patterns between foreigners and the national population in Spain: new evidence from the spanish national health survey. *Social Science & Medicine*, 69(3):370–378.

Idler, E. L. and Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of health and social behavior*, pages 21–37.

Imbens, G. W. and Lemieux, T. (2008). Regression Discontinuity Designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635.

Jann, B. (2008). A stata implementation of the blinder-oaxaca decomposition. *Stata Journal*, 8(4):453–479.

Jones, A. M., Rice, N., d'Uva, T. B., and Balia, S. (2013). *Applied health economics*. Routledge.

Jones, A. M. and Wildman, J. (2008). Health, income and relative deprivation: Evidence from the BHPS. *Journal of Health Economics*, 27(2):308–324.

Jylhä, M. (2009). What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Social science & medicine*, 69(3):307–316.

Kasirye, I. and Ssewanyana, S. (2010). Impacts and determinants of panel survey attrition: The case of Northern Uganda survey 2004-2008.

Kawachi, I., Subramanian, S., and Almeida-Filho, N. (2002). A glossary for health inequalities. *Journal of epidemiology and community health*, 56(9):647–652.

Klein, D. F. (1974). Endogenomorphic depression: a conceptual and terminological revision. *Archives of General Psychiatry*, 31(4):447–454.

Krisela, S. and Jean-Fourie, N. T. (2006). Chronic diseases of lifestyle in South Africa since 1995-2005 technical report. Technical report, Medical Research Council.

Le Grand, J. (1978). The distribution of public expenditure: The case of health care. *Economica*, 45(178):125–42.

Lee, D. S. and Lemieux, T. (2009). Regression discontinuity designs in economics. Technical report, National Bureau of Economic Research.

Lee, D. S. and Lemieuxa, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.

Lee, U. (2003). *Panel Attrition in Survey Data: A Literature Review*, volume 41. Centre for Social Science Research, University of Cape Town.

Leibbrandt, M., Wegner, E., and Finn, A. (2015). The policies for reducing income inequality and poverty in South Africa.

Leibbrandt, M., Woolard, I., and Villiers, d. L. (2009). Methodology: Report of NIDS wave 1, Technical paper number 1, NIDS.

Lewis, C. E., McTigue, K. M., Burke, L. E., Poirier, P., Eckel, R. H., Howard, B. V., Allison, D. B., Kumanyika, S., and Pi-Sunyer, F. X. (2009). Mortality, health outcomes, and body mass index in the overweight range a science advisory from the american heart association. *Circulation*, 119(25):3263–3271.

Link, B. (2013). Fundamental Cause Theory. *Medical Sociology on the Move: New Directions in Theory*, page 105.

Macinko, J. A. and Starfield, B. (2002). Annotated bibliography on equity in health, 1980-2001. *International Journal for Equity in Health*, 1(1):1.

Maluccio, J. A. (2004). Using quality of interview information to assess nonrandom attrition bias in developing-country panel data. *Review of Development Economics*, 8(1):91–109.

Marmot, M. and Bell, R. (2011). Social determinants and dental health. *Advances in dental research*, 23(2):201–206.

Mayosi, B. M., Flisher, A. J., Lalloo, U. G., Sitas, F., Tollman, S. M., and Bradshaw, D. (2009). The burden of non-communicable diseases in South Africa. *The Lancet*, 374(9693):934–947.

McCartney, G., Collins, C., and Mackenzie, M. (2013a). What (or who) causes health inequalities: Theories, evidence and implications? *Health Policy*, 113(3):221–227.

McCartney, G., Collins, C., and Mackenzie, M. (2013b). What (or who) causes health inequalities: Theories, evidence and implications? *Health Policy*, 113(3):221–227.

McIntyre, D. and J.E., A. (2012). What do we know about poverty, inequality and health in South Africa and what are the health system reforms implications? an overview of recent research. *Conference paper in Carnegie III*.

Moffit, R., Fitzgerald, J., and Gottschalk, P. (1999). Sample attrition in panel data: The role of selection on observables. *Annales d'Économie et de Statistique*, pages 129–152.

Moscoe, E., Bor, J., and Bärnighausen, T. (2015). Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. *Journal of clinical epidemiology*, 68(2):122–133.

Naidu, A. and Rao, N. (1994). Body mass index: a measure of the nutritional status in indian populations. *European journal of clinical nutrition*, 48:S131–40.

NDoH (2013). Annual report 2012/2013.

Nwosu, C. (2014). Impact of Health on Labour Force Participation in South Africa.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, pages 693–709.

O'Donnell, O. A., Wagstaff, A., et al. (2008). *Analyzing health equity using household survey data: a guide to techniques and their implementation*. World Bank Publications.

Onis, M. D., Onyango, A. W., Borghi, E., Siyam, A., Nishida, C., and Siekmann, J. (2007). Development of a who growth reference for school-aged children and adolescents. *Bulletin of the World Health Organization*, 85(9):660–667.

Ostry, A. S. and Frank, J. (2010). Was Thomas McKeown right for the wrong reasons? *Critical Public Health*, 20(2):233–243.

Perry, L. (2009). Is bmi an accurate measure of nutritional status in older people? *Nursing*.

Pritchett, L. and Summers, L. H. (1996). Wealthier is healthier. *Journal of Human Resources*, 31(4):841–868.

Puoane, T., Steyn, K., Bradshaw, D., Laubscher, R., Fourie, J., Lambert, V., and Mbananga, N. (2002). Obesity in South Africa: the South African demographic and health survey. *Obesity Research*, 10(10):1038–1048.

Radloff, L. S. (1977). The CES-D scale: A self-reported depression scale for research in general population. *Centre for Epidemiologic Studies, National Institute of Mental Health*.

Rosqvist, F., Iggman, D., Kullberg, J., Cedernaes, J., Johansson, H.-E., Larsson, A., Johansson, L., Ahlström, H., Arner, P., Dahlman, I., et al. (2014). Overfeeding polyunsaturated and saturated fat causes distinct effects on liver and visceral fat accumulation in humans. *Diabetes*, 63(7):2356–2368.

Sahn, D. E. and Younger, S. D. (2009). Measuring intra-household health inequality: explorations using the body mass index. *Health Economics*, 18(S1):S13–S36.

SALDRU (2009). *National Income Dynamics Study 2008, wave 1 [dataset]. version 5.2. Cape Town:*. Southern Africa Labour and Development Research Unit [producer]: Cape Town DataFirst [distributor].

SALDRU (2014a). *National Income Dynamics Study 2010-2011, wave 2 [dataset]. version 2.2. Cape Town:*. Southern Africa Labour and Development Research Unit [producer]: Cape Town DataFirst [distributor].

SALDRU (2014b). *National Income Dynamics Study 2012, wave 3 [dataset]. version 1.2. Cape Town:*. Southern Africa Labour and Development Research Unit [producer]: Cape Town DataFirst [distributor].

Schultz, T. P. (2004). Health Economics and applications in developing countries. *Journal of health economics*, 23(4):637–641.

Skordis-Worrall, J., Hanson, K., and Mills, A. (2011). Estimating the demand for health services in four poor districts of Cape Town, South Africa. *International health*, 3(1):44–49.

Snyder, S. E. and Evans, W. N. (2006). The effect of income on mortality: evidence from the social security notch. *The Review of Economics and Statistics*, 88(3):482–495.

Ssewanyana S., K. I. (2012). Causes of health inequalities in Uganda: Evidence from the demographic and health surveys: 1995-2006, Economic Policy Research Centre Uganda.

Strauss, J., Lei, X., Park, A., Shen, Y., Smith, J. P., Yang, Z., and Zhao, Y. (2010). Health outcomes and socio-economic status among the elderly in China: Evidence from the CHARLS Pilot. *Journal of population ageing*, 3(3-4):111–142.

Tolley, K. (2009). What are health utilities. *Hayward Medical Communications, London*.

Van de Poel, E., Van Doorslaer, E., and ODonnell, O. (2012). Measurement of inequity in health care with heterogeneous response of use to need. *Journal of health economics*, 31(4):676–689.

Van Doorslaer, E. and Jones, A. M. (2003). Inequalities in self-reported health: Validation of a new approach to measurement. *Journal of health economics*, 22(1):61–87.

Van Doorslaer, E. and Jones, A. M. (2004). Income-related inequality in health and health care in the European Union. *Health Economics*, 13(7):605–608.

Van Doorslaer, E., Koolman, X., and Jones, A. M. (2004). Explaining income-related inequalities in doctor utilisation in Europe. *Health Economics*, 13(7):629–647.

Van Doorslaer, E. and O'Donnell, O. (2008). *Measurement and explanation of inequality in health and health care in low-income settings*. Number 2008/04. WIDER Discussion Papers, World Institute for Development Economics (UNU-WIDER).

Wagstaff, A., Paci, P., and Van Doorslaer, E. (1991). On the measurement of inequalities in health. *Social science & medicine*, 33(5):545–557.

Wagstaff, A. and Van Doorslaer, E. (1994). Measuring inequalities in health in the presence of multiple-category morbidity indicators. *Health Economics*, 3(4):281–291.

Wagstaff, A. and van Doorslaer, E. (2000). Measuring and testing for inequity in the delivery of health care. *Journal of Human Resources*, 35(4).

Wagstaff, A., Van Doorslaer, E., and Watanabe, N. (2003). On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in vietnam. *Journal of Econometrics*, 112(1):207–223.

WHO (2010). *World Health Statistics 2010*. World Health Organization.

Wilkinson, R. G. and Pickett, K. (2011). *The Spirit Level: Why Greater Equality Makes Societies Stronger*. Bloomsbury Press.

Wittenberg, M. (2013). The Weight of Success: The Body Mass Index and Economic Well-being in Southern Africa. *Review of Income and Wealth*, 59(S1):S62–S83.

Woolard, I. and Leibbrandt, M. (2006). Towards a poverty line for south africa: Background note.

Wooldridge, J. (2002). *Introductory Econometric: A Modern Approach*. Cengage Learning.

World Bank (2012). World development indicators 2012.

Zere, E. and McIntyre, D. (2003). Equity in Self-reported Adult Illness and Use of Health Service in South Africa: Inter-temporal Comparison. *Journal of Health, Population and Nutrition*, 21(3):205.

Zhao, Z. (2008). Health demand and health determinants in China. *Journal of Chinese Economic and Business Studies*, 6(1):77–98.

Ziebarth, N. (2010). Measurement of health, health inequality, and reporting heterogeneity. *Social Science & Medicine*, 71(1):116–124.

Appendix A

Table A.1: BGLW test for sample attrition in wave 2

VARIABLES	(1) SAH	(2) BMI	(3) DEPR10	(4) DEPR8	(5) PRESSURE
Attrition in wave 2	0.0521	-0.218	0.0972	0.0883	0.0347
	(0.0348)	(0.162)	(0.199)	(0.189)	(0.0974)
25-34 (rel.to 15-24)	-0.282***	2.034***	1.252***	1.090***	0.630***
	(0.0446)	(0.188)	(0.148)	(0.154)	(0.113)
35-44	-0.500***	3.280***	1.543***	1.375***	1.015***
	(0.0529)	(0.233)	(0.195)	(0.193)	(0.109)
45-54	-0.795***	4.131***	1.677***	1.399***	1.555***
	(0.0577)	(0.315)	(0.222)	(0.212)	(0.123)
55-64	-1.037***	4.211***	1.499***	1.360***	1.800***
	(0.0693)	(0.289)	(0.267)	(0.252)	(0.141)
65+	-1.214***	3.906***	1.437***	1.210***	2.029***
	(0.0823)	(0.389)	(0.263)	(0.240)	(0.148)
Log real per capita monthly income	0.0521***	0.230***	-0.512***	-0.388***	-0.0228
	(0.0173)	(0.0860)	(0.0778)	(0.0765)	(0.0282)
General educ (rel. to No educ.)	0.238***	0.937***	-0.540***	-0.407*	0.0109
	(0.0476)	(0.293)	(0.205)	(0.209)	(0.104)
Further education and training	0.342***	1.765***	-0.954***	-0.667***	-0.134
	(0.0587)	(0.307)	(0.218)	(0.223)	(0.114)
Matric	0.605***	1.835***	-1.589***	-1.146***	-0.119
	(0.0633)	(0.307)	(0.262)	(0.252)	(0.133)
Higher education	0.722***	1.625***	-1.412***	-1.260***	-0.401**
	(0.117)	(0.562)	(0.397)	(0.359)	(0.199)
Male	0.199***	-3.642***	-0.584***	-0.623***	0.189***
	(0.0284)	(0.207)	(0.106)	(0.100)	(0.0649)
Coloured (rel. to African)	-0.0781	-0.743*	-0.657*	-0.600	0.0947
	(0.0727)	(0.385)	(0.362)	(0.365)	(0.116)
Asian/Indian	-0.0996	-1.034	-0.955	-0.113	-0.250
	(0.102)	(0.700)	(0.732)	(0.772)	(0.424)
White	0.0354	-0.991**	-1.668***	-0.969***	-0.371**
	(0.0712)	(0.454)	(0.365)	(0.323)	(0.169)
With Partner (rel. to married)	-0.212***	-1.741***	0.896***	0.660***	0.0535
	(0.0636)	(0.247)	(0.235)	(0.235)	(0.127)
Widow/widower	-0.215***	-0.659*	1.232***	1.109***	-0.00664
	(0.0515)	(0.354)	(0.283)	(0.260)	(0.0978)
Divorced	-0.107	-1.152**	2.109***	1.872***	0.175
	(0.104)	(0.458)	(0.518)	(0.498)	(0.165)
Never married	-0.0518	-1.699***	0.603***	0.512***	0.0171
	(0.0444)	(0.243)	(0.174)	(0.177)	(0.0835)
Eastern Cape (rel. to Western Cape)	-0.0574	-0.373	0.992*	0.218	-0.230
	(0.0884)	(0.530)	(0.534)	(0.447)	(0.173)
Northern Cape	-0.151*	-1.697***	0.591	0.692	0.166
	(0.0770)	(0.453)	(0.536)	(0.457)	(0.157)
Free State	-0.348***	-0.887	1.607***	1.520***	-0.174
	(0.101)	(0.550)	(0.588)	(0.519)	(0.194)
Kwa-Zulu Natal	-0.476***	-0.119	1.198**	0.772	0.0245
	(0.0694)	(0.536)	(0.541)	(0.494)	(0.152)
North West	-0.293***	-1.075*	2.127***	2.134***	0.0191
	(0.0945)	(0.606)	(0.584)	(0.498)	(0.176)
Gauteng	-0.208***	-0.935*	0.738	0.913**	-0.288*
	(0.0702)	(0.499)	(0.528)	(0.450)	(0.150)
Mpumalanga	-0.299***	-1.080*	0.471	0.691	-0.0526
	(0.0890)	(0.558)	(0.537)	(0.464)	(0.175)
Limpopo	-0.0406	-2.056***	1.098**	1.212**	-0.233
	(0.0905)	(0.539)	(0.550)	(0.474)	(0.202)
Urban (rel. to Traditional area)	-0.0991*	0.404*	0.510**	0.378	0.103
	(0.0569)	(0.231)	(0.247)	(0.248)	(0.0821)
Farms	0.102	-0.482	-0.299	-0.200	0.0151
	(0.0762)	(0.308)	(0.298)	(0.301)	(0.131)
Constant		24.00***	9.971***	6.444***	-1.909***
		(0.720)	(0.724)	(0.686)	(0.239)
cut1	-1.768***				
	(0.140)				
cut2	-1.048***				
	(0.135)				
cut3	-0.219				
	(0.139)				
cut4	0.571***				
	(0.143)				
Observations	14,776	12,915	14,791	14,796	7,759
R-squared		0.240	0.125	0.081	

Notes: Table shows different dimensions of health outcomes regressed on attrition indicator in wave 2 and other covariates in the model. In all models the coefficient on attrition indicator is not statistically significant. Standard errors reported in parentheses are robust, they are clustered by pid. *** p<0.01, ** p<0.05, * p<0.1.

Table A.2: BGLW test for sample attrition in wave 3

VARIABLES	(1) SAH	(2) BMI	(3) DEPR10	(4) DEPR8	(5) PRESSURE
Attrition in wave 3	0.0458	0.144	0.0960	-0.0831	0.0347
	(0.0437)	(0.203)	(0.172)	(0.175)	(0.0974)
25-34 (rel.to 15-24)	-0.279***	2.017***	1.241***	1.088***	0.630***
	(0.0450)	(0.188)	(0.148)	(0.156)	(0.113)
35-44	-0.492***	3.297***	1.552***	1.384***	1.015***
	(0.0523)	(0.238)	(0.195)	(0.192)	(0.109)
45-54	-0.771***	4.239***	1.644***	1.342***	1.555***
	(0.0568)	(0.302)	(0.230)	(0.218)	(0.123)
55-64	-1.021***	4.322***	1.476***	1.309***	1.800***
	(0.0706)	(0.301)	(0.268)	(0.250)	(0.141)
65+	-1.210***	3.811***	1.421***	1.169***	2.029***
	(0.0878)	(0.386)	(0.285)	(0.257)	(0.148)
Log real per capita monthly income	0.0521***	0.222***	-0.515***	-0.394***	-0.0228
	(0.0171)	(0.0840)	(0.0776)	(0.0757)	(0.0282)
General educ (rel. to No educ.)	0.250***	0.983***	-0.590***	-0.470**	0.0109
	(0.0486)	(0.286)	(0.216)	(0.224)	(0.104)
Further education and training	0.361***	1.771***	-1.012***	-0.749***	-0.134
	(0.0576)	(0.305)	(0.228)	(0.238)	(0.114)
Matric	0.619***	1.903***	-1.647***	-1.221***	-0.119
	(0.0632)	(0.305)	(0.274)	(0.269)	(0.133)
Higher education	0.718***	1.576***	-1.416***	-1.251***	-0.401**
	(0.118)	(0.542)	(0.394)	(0.356)	(0.199)
Male	0.197***	-3.610***	-0.591***	-0.600***	0.189***
	(0.0287)	(0.209)	(0.109)	(0.101)	(0.0649)
Coloured (rel. to African)	-0.0636	-0.613	-0.666*	-0.611	0.0947
	(0.0727)	(0.402)	(0.370)	(0.378)	(0.116)
Asian/Indian	-0.0951	-0.941	-1.001	-0.137	-0.250
	(0.106)	(0.710)	(0.748)	(0.784)	(0.424)
White	0.0484	-0.971**	-1.703***	-0.956***	-0.371**
	(0.0732)	(0.452)	(0.353)	(0.310)	(0.169)
With Partner (rel. to married)	-0.202***	-1.764***	0.872***	0.659***	0.0535
	(0.0657)	(0.245)	(0.235)	(0.233)	(0.127)
Widow/widower	-0.228***	-0.588	1.189***	1.088***	-0.00664
	(0.0573)	(0.368)	(0.290)	(0.264)	(0.0978)
Divorced	-0.0994	-1.197**	2.146***	1.917***	0.175
	(0.106)	(0.468)	(0.532)	(0.514)	(0.165)
Never married	-0.0474	-1.742***	0.592***	0.493***	0.0171
	(0.0457)	(0.251)	(0.177)	(0.178)	(0.0835)
Eastern Cape (rel. to Western Cape)	-0.0541	-0.333	0.973*	0.211	-0.230
	(0.0907)	(0.532)	(0.539)	(0.449)	(0.173)
Northern Cape	-0.152*	-1.538***	0.486	0.611	0.166
	(0.0773)	(0.441)	(0.541)	(0.467)	(0.157)
Free State	-0.338***	-0.809	1.593***	1.491***	-0.174
	(0.103)	(0.547)	(0.594)	(0.524)	(0.194)
Kwa-Zulu Natal	-0.463***	0.0541	1.148**	0.726	0.0245
	(0.0708)	(0.513)	(0.540)	(0.486)	(0.152)
North West	-0.294***	-0.980	2.111***	2.125***	0.0191
	(0.0949)	(0.598)	(0.587)	(0.498)	(0.176)
Gauteng	-0.209***	-0.892*	0.751	0.926**	-0.288*
	(0.0707)	(0.502)	(0.534)	(0.457)	(0.150)
Mpumalanga	-0.296***	-0.915*	0.441	0.677	-0.0526
	(0.0911)	(0.541)	(0.540)	(0.465)	(0.175)
Limpopo	-0.0375	-1.966***	1.054*	1.168**	-0.233
	(0.0911)	(0.531)	(0.549)	(0.470)	(0.202)
Urban (rel. to Traditional area)	-0.110*	0.372	0.492*	0.380	0.103
	(0.0577)	(0.230)	(0.253)	(0.253)	(0.0821)
Farms	0.0906	-0.489	-0.282	-0.184	0.0151
	(0.0782)	(0.310)	(0.309)	(0.308)	(0.131)
Constant		23.88***	10.09***	6.603***	-1.880***
		(0.722)	(0.725)	(0.687)	(0.245)
cut1	-1.753***				
	(0.141)				
cut2	-1.038***				
	(0.134)				
cut3	-0.205				
	(0.137)				
cut4	0.588***				
	(0.142)				
Observations	14,308	12,513	14,322	14,327	7,552
R-squared		0.244	0.126	0.083	

Notes: Table shows different dimensions of health outcomes regressed on attrition indicator in wave 3 and other covariates in the model. In all models the coefficient on attrition indicator is not statistically significant. Standard errors reported in parentheses are robust, they are clustered by pid. *** p<0.01, ** p<0.05, * p<0.1.

Table A.3: Health outcome regression results with and without correction for sample attrition in w2

	Not corrected for sample selection				Corrected for sample selection			
VARIABLES	SAH0	BMI0	CESD0	BP0	SAH1	BMI1	CESD1	BP1
25-34 (rel.to 15-24)	-0.266***	1.773***	0.879***	0.800***	-0.266***	1.773***	0.879***	0.800***
	(0.0414)	(0.225)	(0.148)	(0.121)	(0.0414)	(0.225)	(0.148)	(0.121)
35-44	-0.481***	2.944***	1.405***	1.055***	-0.481***	2.944***	1.405***	1.055***
	(0.0480)	(0.262)	(0.168)	(0.133)	(0.0480)	(0.262)	(0.168)	(0.133)
45-54	-0.785***	3.536***	1.389***	1.467***	-0.785***	3.536***	1.389***	1.467***
	(0.0547)	(0.310)	(0.189)	(0.129)	(0.0547)	(0.310)	(0.189)	(0.129)
55-64	-1.061***	3.854***	1.675***	1.692***	-1.061***	3.854***	1.675***	1.692***
	(0.0653)	(0.372)	(0.231)	(0.149)	(0.0653)	(0.372)	(0.231)	(0.149)
65+	-1.255***	3.032***	1.409***	1.891***	-1.255***	3.032***	1.409***	1.891***
	(0.0789)	(0.421)	(0.263)	(0.160)	(0.0789)	(0.421)	(0.263)	(0.160)
Log real per capita monthly income	0.0156	0.238**	-0.248***	-0.0244	0.0156	0.238**	-0.248***	-0.0244
	(0.0145)	(0.0933)	(0.0572)	(0.0404)	(0.0145)	(0.0933)	(0.0572)	(0.0404)
General educ (rel. to No educ.)	0.0437	0.775***	-0.334*	-0.113	0.0437	0.775***	-0.334*	-0.113
	(0.0536)	(0.275)	(0.180)	(0.111)	(0.0536)	(0.275)	(0.180)	(0.111)
Further education and training	0.205***	1.599***	-0.858***	-0.291**	0.205***	1.599***	-0.858***	-0.291**
	(0.0624)	(0.321)	(0.204)	(0.138)	(0.0624)	(0.321)	(0.204)	(0.138)
Matric	0.378***	2.140***	-1.043***	-0.441***	0.378***	2.140***	-1.043***	-0.441***
	(0.0652)	(0.342)	(0.216)	(0.159)	(0.0652)	(0.342)	(0.216)	(0.159)
Higher education	0.636***	0.644	-1.161***	-0.234	0.636***	0.644	-1.161***	-0.234
	(0.141)	(0.763)	(0.384)	(0.296)	(0.141)	(0.763)	(0.384)	(0.296)
Male	0.148***	-3.666***	-0.411***	0.194***	0.148***	-3.666***	-0.411***	0.194***
	(0.0305)	(0.165)	(0.100)	(0.0709)	(0.0305)	(0.165)	(0.100)	(0.0709)
Coloured (rel. to African)	0.121*	-0.954**	-1.309***	0.179	0.121*	-0.954**	-1.309***	0.179
	(0.0720)	(0.400)	(0.207)	(0.146)	(0.0720)	(0.400)	(0.207)	(0.146)
Asian/Indian	-0.0308	-3.511***	-2.745***	-0.287	-0.0308	-3.511***	-2.745***	-0.287
	(0.126)	(0.639)	(0.460)	(0.310)	(0.126)	(0.639)	(0.460)	(0.310)
White	0.307***	-0.117	-0.985***	-0.0347	0.307***	-0.117	-0.985***	-0.0347
	(0.0890)	(0.567)	(0.252)	(0.263)	(0.0890)	(0.567)	(0.252)	(0.263)
With Partner (rel. to married)	-0.113*	-1.734***	0.562**	0.245	-0.113*	-1.734***	0.562**	0.245
	(0.0625)	(0.345)	(0.223)	(0.163)	(0.0625)	(0.345)	(0.223)	(0.163)
Widow/widower	-0.104	-0.394	0.730***	0.0183	-0.104	-0.394	0.730***	0.0183
	(0.0639)	(0.391)	(0.253)	(0.121)	(0.0639)	(0.391)	(0.253)	(0.121)
Divorced	0.0509	-0.420	0.895***	-0.272	0.0509	-0.420	0.895***	-0.272
	(0.126)	(0.632)	(0.297)	(0.208)	(0.126)	(0.632)	(0.297)	(0.208)
Never married	-0.0426	-2.198***	0.882***	-0.0436	-0.0426	-2.198***	0.882***	-0.0436
	(0.0444)	(0.250)	(0.151)	(0.104)	(0.0444)	(0.250)	(0.151)	(0.104)
Eastern Cape (rel. to Western Cape)	0.184**	-0.468	0.427*	0.149	0.184**	-0.468	0.427*	0.149
	(0.0721)	(0.403)	(0.255)	(0.171)	(0.0721)	(0.403)	(0.255)	(0.171)
Northern Cape	0.0805	-1.151***	0.605**	0.237*	0.0805	-1.151***	0.605**	0.237*
	(0.0704)	(0.386)	(0.267)	(0.136)	(0.0704)	(0.386)	(0.267)	(0.136)
Free State	0.299***	-0.0489	0.627**	0.194	0.299***	-0.0489	0.627**	0.194
	(0.0862)	(0.466)	(0.299)	(0.174)	(0.0862)	(0.466)	(0.299)	(0.174)
Kwa-Zulu Natal	0.152**	0.921**	1.655***	-0.00365	0.152**	0.921**	1.655***	-0.00365
	(0.0768)	(0.383)	(0.253)	(0.168)	(0.0768)	(0.383)	(0.253)	(0.168)
North West	0.168**	-0.884**	1.173***	0.0212	0.168**	-0.884**	1.173***	0.0212
	(0.0853)	(0.441)	(0.277)	(0.230)	(0.0853)	(0.441)	(0.277)	(0.230)
Gauteng	0.367***	-0.807**	1.273***	0.105	0.367***	-0.807**	1.273***	0.105
	(0.0742)	(0.393)	(0.238)	(0.163)	(0.0742)	(0.393)	(0.238)	(0.163)
Mpumalanga	0.0389	-0.948**	0.540**	-0.234	0.0389	-0.948**	0.540**	-0.234
	(0.0734)	(0.432)	(0.256)	(0.171)	(0.0734)	(0.432)	(0.256)	(0.171)
Limpopo	0.433***	-1.029**	0.349	-0.370**	0.433***	-1.029**	0.349	-0.370**
	(0.0823)	(0.425)	(0.267)	(0.176)	(0.0823)	(0.425)	(0.267)	(0.176)
Urban (rel. to Traditional area)	-0.0858**	0.312	0.359***	0.0644	-0.0858**	0.312	0.359***	0.0644
	(0.0370)	(0.200)	(0.137)	(0.0996)	(0.0370)	(0.200)	(0.137)	(0.0996)
Farms	-0.0788	-0.625**	0.198	-0.159	-0.0788	-0.625**	0.198	-0.159
	(0.0524)	(0.265)	(0.172)	(0.111)	(0.0524)	(0.265)	(0.172)	(0.111)
Constant	-2.033***	24.66***	6.943***	-1.869***	-2.033***	24.66***	6.943***	-1.869***
	(0.140)	(0.796)	(0.457)	(0.351)	(0.140)	(0.796)	(0.457)	(0.351)
Constant	-1.341***				-1.341***			
	(0.136)				(0.136)			
Constant	-0.489***				-0.489***			
	(0.136)				(0.136)			
Constant	0.370***				0.370***			
	(0.136)				(0.136)			
Observations	17,302	14,675	16,445	8,245	17,302	14,675	16,445	8,245
R-squared		0.234	0.092			0.234	0.092	

Notes: The table shows health outcome regressions in wave 2 without and with correction for sample selection. We use the inverse of probabilities from attrition probits to correct for attrition bias. The coefficients do not change implying that attrition is random for these health outcomes. The dependent variables are health outcomes indicated as column titles. *** p<0.01, ** p<0.05, * p<0.1.

Table A.4: Health outcome regression results with and without correction for sample attrition in w3

	Not corrected for sample selection				Corrected for sample selection			
VARIABLES	SAH0	BMI0	CESD0	BP0	SAH1	BMI1	CESD1	BP1
25-34 (rel.to 15-24)	-0.188***	1.726***	0.985***	0.483***	-0.188***	1.726***	0.985***	0.483***
	(0.0434)	(0.186)	(0.142)	(0.101)	(0.0434)	(0.186)	(0.142)	(0.101)
35-44	-0.459***	2.966***	1.202***	1.008***	-0.459***	2.966***	1.202***	1.008***
	(0.0498)	(0.216)	(0.183)	(0.109)	(0.0498)	(0.216)	(0.183)	(0.109)
45-54	-0.693***	3.433***	1.484***	1.433***	-0.693***	3.433***	1.484***	1.433***
	(0.0519)	(0.275)	(0.190)	(0.110)	(0.0519)	(0.275)	(0.190)	(0.110)
55-64	-0.904***	3.998***	1.481***	1.793***	-0.904***	3.998***	1.481***	1.793***
	(0.0578)	(0.296)	(0.226)	(0.129)	(0.0578)	(0.296)	(0.226)	(0.129)
65+	-1.162***	3.224***	1.863***	1.946***	-1.162***	3.224***	1.863***	1.946***
	(0.0677)	(0.358)	(0.244)	(0.138)	(0.0677)	(0.358)	(0.244)	(0.138)
Log real per capita monthly income	0.0453***	0.455***	-0.337***	-0.0146	0.0453***	0.455***	-0.337***	-0.0146
	(0.0160)	(0.0762)	(0.0594)	(0.0344)	(0.0160)	(0.0762)	(0.0594)	(0.0344)
General educ (rel. to No educ.)	0.140***	0.997***	-0.313	0.0330	0.140***	0.997***	-0.313	0.0330
	(0.0493)	(0.256)	(0.190)	(0.0938)	(0.0493)	(0.256)	(0.190)	(0.0938)
Further education and training	0.298***	1.868***	-0.493**	-0.0256	0.298***	1.868***	-0.493**	-0.0256
	(0.0576)	(0.298)	(0.220)	(0.116)	(0.0576)	(0.298)	(0.220)	(0.116)
Matric	0.397***	2.115***	-0.897***	-0.0510	0.397***	2.115***	-0.897***	-0.0510
	(0.0602)	(0.308)	(0.222)	(0.126)	(0.0602)	(0.308)	(0.222)	(0.126)
Higher education	0.674***	1.159*	-1.409***	-0.551**	0.674***	1.159*	-1.409***	-0.551**
	(0.107)	(0.598)	(0.415)	(0.262)	(0.107)	(0.598)	(0.415)	(0.262)
Male	0.149***	-3.716***	-0.0479	0.219***	0.149***	-3.716***	-0.0479	0.219***
	(0.0289)	(0.135)	(0.104)	(0.0618)	(0.0289)	(0.135)	(0.104)	(0.0618)
Coloured (rel. to African)	0.130**	-1.312***	-2.145***	-0.0907	0.130**	-1.312***	-2.145***	-0.0907
	(0.0601)	(0.284)	(0.231)	(0.111)	(0.0601)	(0.284)	(0.231)	(0.111)
Asian/Indian	-0.198*	-1.987***	-2.313***	-0.241	-0.198*	-1.987***	-2.313***	-0.241
	(0.118)	(0.604)	(0.405)	(0.250)	(0.118)	(0.604)	(0.405)	(0.250)
White	-0.111	-0.0641	-2.639***	-0.202	-0.111	-0.0641	-2.639***	-0.202
	(0.0737)	(0.457)	(0.272)	(0.153)	(0.0737)	(0.457)	(0.272)	(0.153)
With Partner (rel. to married)	-0.0677	-1.584***	0.0701	-0.343**	-0.0677	-1.584***	0.0701	-0.343**
	(0.0677)	(0.358)	(0.239)	(0.135)	(0.0677)	(0.358)	(0.239)	(0.135)
Widow/widower	-0.0758	-1.106***	0.869***	-0.0464	-0.0758	-1.106***	0.869***	-0.0464
	(0.0583)	(0.323)	(0.270)	(0.116)	(0.0583)	(0.323)	(0.270)	(0.116)
Divorced	0.000805	-0.959*	1.418***	0.216	0.000805	-0.959*	1.418***	0.216
	(0.0912)	(0.544)	(0.342)	(0.192)	(0.0912)	(0.544)	(0.342)	(0.192)
Never married	-0.0642	-2.127***	0.509***	0.0145	-0.0642	-2.127***	0.509***	0.0145
	(0.0422)	(0.208)	(0.152)	(0.0881)	(0.0422)	(0.208)	(0.152)	(0.0881)
Eastern Cape (rel. to Western Cape)	0.134**	-0.258	-0.241	-0.109	0.134**	-0.258	-0.241	-0.109
	(0.0638)	(0.341)	(0.250)	(0.131)	(0.0638)	(0.341)	(0.250)	(0.131)
Northern Cape	-0.144*	-0.491	-1.294***	-0.0674	-0.144*	-0.491	-1.294***	-0.0674
	(0.0782)	(0.338)	(0.230)	(0.171)	(0.0782)	(0.338)	(0.230)	(0.171)
Free State	0.0815	-0.624*	-1.221***	-0.574***	0.0815	-0.624*	-1.221***	-0.574***
	(0.0707)	(0.354)	(0.266)	(0.132)	(0.0707)	(0.354)	(0.266)	(0.132)
Kwa-Zulu Natal	0.147**	0.396	-0.452*	-0.194	0.147**	0.396	-0.452*	-0.194
	(0.0644)	(0.335)	(0.251)	(0.129)	(0.0644)	(0.335)	(0.251)	(0.129)
North West	0.176**	-0.593*	-1.450***	-0.0881	0.176**	-0.593*	-1.450***	-0.0881
	(0.0715)	(0.353)	(0.283)	(0.163)	(0.0715)	(0.353)	(0.283)	(0.163)
Gauteng	0.239***	-0.905***	-1.844***	-0.349***	0.239***	-0.905***	-1.844***	-0.349***
	(0.0638)	(0.321)	(0.245)	(0.124)	(0.0638)	(0.321)	(0.245)	(0.124)
Mpumalanga	0.122*	-0.362	-2.783***	-0.509***	0.122*	-0.362	-2.783***	-0.509***
	(0.0682)	(0.359)	(0.266)	(0.161)	(0.0682)	(0.359)	(0.266)	(0.161)
Limpopo	0.222***	-1.061***	-1.745***	-0.322**	0.222***	-1.061***	-1.745***	-0.322**
	(0.0712)	(0.361)	(0.279)	(0.149)	(0.0712)	(0.361)	(0.279)	(0.149)
Urban (rel. to Traditional area)	0.0280	0.447**	0.604***	0.226***	0.0280	0.447**	0.604***	0.226***
	(0.0345)	(0.174)	(0.126)	(0.0790)	(0.0345)	(0.174)	(0.126)	(0.0790)
Farms	0.0854	0.151	-0.213	0.193*	0.0854	0.151	-0.213	0.193*
	(0.0558)	(0.257)	(0.229)	(0.0990)	(0.0558)	(0.257)	(0.229)	(0.0990)
Constant	-1.714***	22.84***	9.544***	-1.903***	-1.714***	22.84***	9.544***	-1.903***
	(0.140)	(0.652)	(0.539)	(0.297)	(0.140)	(0.652)	(0.539)	(0.297)
Constant	-0.970***				-0.970***			
	(0.136)				(0.136)			
Constant	0.0372				0.0372			
	(0.135)				(0.135)			
Constant	0.852***				0.852***			
	(0.136)				(0.136)			
Observations	18,665	18,081	18,632	10,538	18,665	18,081	18,632	10,538
R-squared		0.258	0.125			0.258	0.125	

Notes: The table shows health outcome regressions in wave 3 without and with correction for sample selection. We use the inverse of probabilities from attrition probits to correct for attrition bias. The coefficients do not change implying that attrition is random for these health outcomes. The dependent variables are health outcomes as indicated in column titles. *** p<0.01, ** p<0.05, * p<0.1.

Appendix B

B.1.1 Blood pressure and age

In table B.1 we present the relationship between blood pressure and age. Column totals indicate the weighted proportion of individuals in a given age group. Aggregate total is the actual number of individuals in a given age group. The other values in the table indicate weighted proportions of individuals in different age groups who have normal or high blood pressure. The table shows that high blood pressure increases with age. As we move from younger age groups to older ones, the weighted proportion of those with high blood pressure increase relative to the weighted proportion of the age group in the population.

B.1 Bivariate analysis of socioeconomic factors and health outcomes

B.2.1 SAH and Age

Age is measured in years as a continuous variable in the survey. However for our analysis we group age in 10 year age groups, as it is more tractable to compare the relationship between health outcomes and age group. As mentioned before, we only include adult individuals, defined as those aged 15 years and above. Table B.2 shows SAH by age group in each wave. The column totals for each wave show weighted proportion of individuals in a given age group. Aggregate totals indicate the number of individuals in each age group. The other values in the table show weighted proportion of individuals in age categories who report themselves in different health states in each wave of the survey.

We observe for example that in wave 1, 25-34 year olds are approximately 3 times more likely to report their health as excellent than those 55-64, after controlling for population size. On the other hand, 55-64 year olds are 4 times more likely to report poor health status than 25-34 year olds. In general, reporting poor health status increases with age, while reporting excellent health status decreases with age. From the table, health depreciates over the life

Table B.1: Blood pressure and age

Wave 1	Age Groups						
Blood pressure	**15-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65+**	**Total**
Normal BP	37.25	27.38	17.58	9.57	5.15	3.08	100
HBP	5.11	13.22	18.09	26.32	20.79	16.47	100
Total	**32.79**	**25.41**	**17.65**	**11.89**	**7.32**	**4.94**	**100**
Aggregate total	**2919**	**1572**	**1328**	**1035**	**677**	**578**	**8109**
Wave 2							
Normal BP	35.87	26.62	18.77	10.20	5.35	3.18	100
HBP	4.15	16.86	20.21	24.61	18.74	15.43	100
Total	**31.75**	**25.35**	**18.95**	**12.08**	**7.09**	**4.78**	**100**
Aggregate total	**3202**	**1752**	**1231**	**976**	**660**	**538**	**8359**
Wave 3							
Normal BP	35.07	27.23	17.75	10.87	5.91	3.17	100
HBP	5.47	11.23	20.00	24.70	22.61	15.99	100
Total	**31.15**	**25.12**	**18.05**	**12.70**	**8.12**	**4.86**	**100**
Aggregate total	**3961**	**2258**	**1540**	**1271**	**867**	**664**	**10561**

Notes: Table shows blood pressure status by age group in each wave of NIDS. The column totals for each wave indicate the weighted proportion of individuals in each age group while the other values in the table indicate the weighted proportion in a specific age group who have a given blood pressure. Aggregate total is the actual number of individuals in each age group. Row totals add up to 100% (any difference is due to rounding).

cycle, which is consistent with Grossman's health human capital theory[1].

This pattern is the same across all waves. In wave 2, individuals in age category 25-34 are 3 times more likely to report excellent health than 55-64 year olds. This is also the case in wave 3. In general if the age categories were not associated with SAH, we would expect each health status category to be proportionately distributed in the population as the age groups. However we observe that as individuals grow old they are more likely to report poor and fair health.

[1]Grossman life cycle theory postulated that people's health depreciate with age therefore individuals continuously invest in health inputs to better their health (Grossman, 2000).

Table B.2: SAH and Age categories

Wave 1	**Age in categories**						
Health status	**15-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65+**	Total
Poor	5.25	12.83	19.82	21.46	19.63	21.01	100
Fair	9.05	13.96	18.68	22.07	18.18	18.06	100
Good	23.12	24.70	18.99	15.03	9.77	8.39	100
Very Good	32.48	26.64	17.79	11.60	6.93	4.57	100
Excellent	38.87	28.55	18.93	8.93	3.16	1.55	100
Total	**27.99**	**24.49**	**18.66**	**13.35**	**8.46**	**7.05**	**100**
Aggregate total	**4706**	**2923**	**2585**	**2174**	**1533**	**1537**	**15458**
Wave 2							
Poor	3.84	8.85	19.70	20.68	20.12	26.82	100
Fair	5.26	15.02	15.61	24.94	19.10	20.07	100
Good	19.45	20.34	18.71	16.65	13.36	11.49	100
Very Good	28.67	25.46	20.68	12.21	8.00	4.97	100
Excellent	35.64	29.08	18.62	10.10	3.94	2.61	100
Total	**27.32**	**24.71**	**19.07**	**13.37**	**8.54**	**7.00**	**100**
Aggregate total	**5745**	**3497**	**2712**	**2355**	**1697**	**1578**	**17584**
Wave 3							
Poor	5.31	12.45	20.68	20.58	19.33	21.66	100
Fair	6.13	12.31	17.69	20.54	18.67	24.67	100
Good	22.07	20.60	19.36	15.90	11.17	10.90	100
Very Good	28.28	25.25	20.61	13.43	8.36	4.06	100
Excellent	34.45	30.56	19.12	9.43	4.21	2.23	100
Total	**26.30**	**24.42**	**19.55**	**13.50**	**8.84**	**7.39**	**100**
Aggregate total	**5935**	**3874**	**2844**	**2516**	**1839**	**1688**	**18696**

Notes: Table shows health status by age categories in the three waves of NIDS. The column totals for each wave indicate the weighted proportion of individuals in a given age group while the other values in the table indicate the weighted proportion in the age group who report a given health status. Aggregate total is the actual number of individuals in each age group. Row totals add up to 100 (any difference is due to rounding).

B.2.2 SAH and Gender

Table B.3 shows the weighted proportion of SAH categories by gender in the three waves of NIDS. Column totals indicate weighted proportions of men and women in the population while aggregate totals indicate the actual number of individuals in the sample. Row totals for each wave add up to 100%. As we can see from the table, the population in wave 1 is 56% women and 44% men. This stabilized at 54% women and 46% men in waves 2 and 3.

The health-gender relationship is such that disproportionately more women report poor health across all waves. Women make up 56% of the population, 66% reported poor health in wave 1 and 64% in wave 2 and wave 3. Men are more likely to report excellent health. This analysis supports the gender-health gradient established in the epidemiological literature and for this reason gender is controlled for in all regressions that follow.

Table B.3: SAH and Gender by Wave

	Wave 1			Wave 2			Wave 3		
Health Status	**Female**	**Male**	**Total**	**Female**	**Male**	**Total**	**Female**	**Male**	**Total**
Poor	66.03	33.97	100	64.42	35.58	100	64.49	35.51	100
Fair	65.42	34.58	100	65.77	34.23	100	64.21	35.79	100
Good	60.20	39.80	100	57.29	42.71	100	57.31	42.69	100
Very Good	53.41	46.59	100	53.30	46.70	100	54.35	45.65	100
Excellent	49.41	50.59	100	49.86	50.14	100	48.69	51.31	100
Total	**55.86**	**44.14**	**100**	**53.89**	**46.11**	**100**	**54.29**	**45.71**	**100**
Aggregate Total	**9267**	**6217**	**15477**	**10291**	**7285**	**17576**	**11132**	**7566**	**18698**

Notes: Table shows health status by gender in the three waves of NIDS. The column totals for each wave indicate the weighted proportion of individuals of each gender while the other values in the table indicate the weighted proportion in a specific gender who report a given health status. Aggregate total is the actual number of individuals in each gender. Row totals for each wave add up to 100 (any difference is due to rounding).

B.2.3 Province and Geographical Region of Residence

As already mentioned, NIDS is not designed to be representative at provincial level and thus analysis of results at provincial level is not recommended (Leibbrandt et al., 2009). However, individuals from each province were interviewed. 28% of the sample were from Kwa-Zulu Natal province, 12% each were from the Western and Eastern Cape provinces. The sample from each of the rest of the provinces was less than 10%. Weighted proportion indicate that Gauteng had 25.5% of the representation in the sample, followed by Kwa-Zulu Natal with 18%. Western Cape and Eastern Cape provinces had approximately 10% each while all the other provinces had less than 10% each. In all the regressions we control for province fixed effects.

Within the provinces four geographical types were identified initially in the data based on the Statistics SA census of 2001. Enumeration areas in 2001 census were classified as either rural formal, traditional authority area, urban formal or urban informal. Individuals selected from these enumeration areas were therefore classified as such. The majority of the households sampled were either in tribal authority areas (42.9%) or urban formal areas (40.6%). The rest of the households sampled were from rural formal (9.82%) and urban informal (6.65%) areas in wave 1. After the release of the Statistics South African 2011 census data, the geographic type data in NIDS were reclassified to three types, namely urban, traditional area, and farms, in accordance with the latest census classification of enumeration areas to ensure consistency and comparability.

Table B.4 shows weighted proportion of SAH by geographic type of residence as per the new classification in the three waves of NIDS. In wave 1, about 32% of individuals live in traditional areas while 60% and 8% live in urban areas and on farms respectively. The

distribution of individuals with poor health status is almost the same as that of the entire population, where 36% live in traditional areas, 54% in urban areas and 9% on farms. This implies that geographical type of residence does not influence SAH status. The pattern is the same in wave 2 and wave 3 which rules out evidence of rural-urban inequality in SAH.

Table B.4: SAH and Geographical Region of Residence

Wave 1	**Region of residence**			
Health status	**Traditional area**	**Urban**	**Farms**	Total
Poor	36.55	54.52	8.93	**100**
Fair	35.14	57.49	7.37	**100**
Good	30.68	60.39	8.93	**100**
Very Good	33.27	59.44	7.29	**100**
Excellent	29.28	62.11	8.61	**100**
Total	**31.81**	**59.98**	**8.21**	**100**
Aggregate total	**6301**	**7456**	**1720**	**15477**
Wave 2				
Poor	30.15	59.09	10.76	**100**
Fair	32.81	59.56	7.63	**100**
Good	31.50	60.49	8.00	**100**
Very Good	35.99	56.59	7.42	**100**
Excellent	30.66	63.25	6.08	**100**
Total	**32.54**	**60.35**	**7.11**	**100**
Aggregate total	**7685**	**7928**	**1787**	**17400**
Wave 3				
Poor	32.04	59.48	8.49	**100**
Fair	31.93	63.08	4.99	**100**
Good	37.84	56.33	5.82	**100**
Very Good	29.93	63.86	6.22	**100**
Excellent	29.75	65.07	5.18	**100**
Total	**32.15**	**62.12**	**5.73**	**100**
Aggregate total	**8111**	**8996**	**1591**	**18698**

Notes: Table shows health status by enumeration area type of residence in the three waves of NIDS. The column totals for each wave indicate the weighted proportion of individuals in the area of residence while the other values in the table indicate the weighted proportion of individuals in an area who report a given health status. Aggregate total is the actual number of individuals in each regions. Row totals add up to 100% (any difference is due to rounding).

B.2.4 Depression and age

Table B.5 shows the relationship between depressive symptoms and age, measured in age groups. Column totals for each wave indicate the weighted proportion of individuals in each

age group. The other values indicate the weighted proportion with or without depressive symptoms. Aggregate total is the actual number of individuals in the age group.

Individuals aged 55-64 are 1.2 times more likely to be depressed than those in age group 15-24. The 15-24 age group has a smaller proportion of individuals who are depressed than their proportion in the population. 55-64 year olds have a slightly larger proportion who are depressed, compared to their proportion in the population. Those above age 65 have a weighted proportion of depressed also slightly higher than their representation in the population. This trend is the same across all three waves of NIDS, pointing to the fact that age is positively associated with depressive symptoms.

Table B.5: Depression and age groups

Wave 1	Age Groups						
Depression	15-24	25-34	35-44	45-54	55-64	65+	Total
Not depressed	30.12	24.77	18.27	12.19	7.84	6.81	100
Depressed	24.97	25.17	19.11	14.79	8.93	7.03	100
Total	**28.14**	**24.92**	**18.59**	**13.19**	**8.26**	**6.89**	**100**
Aggregate total	**5122**	**3268**	**2802**	**2333**	**1634**	**1650**	**16809**
Wave 2							
Not depressed	29.22	24.45	17.80	13.31	8.52	6.70	100
Depressed	22.95	26.20	21.75	13.20	8.36	7.35	100
Total	**27.35**	**24.97**	**18.98**	**13.28**	**8.47**	**6.95**	**100**
Aggregate total	**6914**	**4612**	**3418**	**2908**	**2029**	**1961**	**21842**
Wave 3							
Not depressed	27.63	24.39	19.74	13.09	8.60	6.55	100
Depressed	22.08	24.68	19.72	14.43	9.21	9.88	100
Total	**26.23**	**24.46**	**19.73**	**13.43**	**8.76**	**7.39**	**100**
Aggregate total	**6818**	**4892**	**3574**	**3001**	**2154**	**2040**	**22479**

Notes: Table shows depression by age category in the three waves of NIDS. The column totals for each wave indicate the weighted proportion of individuals in each age group while the other values in the table indicate the weighted proportion of individuals with a given depressive symptom. Aggregate total is the actual number of individuals in each age group. Row totals add up to 100% (any difference is due to rounding).