EFFECT OF NUTRITIONAL STATUS ON HEALTH OUTCOMES: AN ECONOMIC ANALYSIS

Thesis submitted to Jawaharlal Nehru University

for award of the degree of

DOCTOR OF PHILOSOPHY

SHIVANI GUPTA

Centre for International Trade and Development

School of International Studies

Jawaharlal Nehru University

New Delhi-110067

India

2019

अन्तर्राष्ट्रीय व्यापार एवं विकास केंद्र
Centre for International Trade & Development
School of International Studies
Jawaharlal Nehru University, New Delhi 110067

Date: 12/07/2019

DECLARATION

I declare that the thesis entitled **"Effect of Nutritional Status on Health Outcomes: An Economic Analysis"** submitted by me for the award of the degree of **Doctor of Philosophy** of Jawaharlal Nehru University is my own work. The thesis has not been submitted for any other degree of this University or any other University.

SHIVANI GUPTA

CERTIFICATE

We recommend that this thesis be placed before the examiners for evaluation.

8/7/2019

PROF. MEETA K. MEHRA
Chairperson, CITD

Sangeeta Bansal

PROF. SANGEETA BANSAL
Supervisor

Professor
Centre for International Trade & Development
School of International Studies
Jawaharlal Nehru University
New Delhi - 110 067

Phone: 26704389 / 26704340, Fax: 91-11-26741586, E-mail: chair_citd@mail.jnu.ac.in

Acknowledgement

The work presented in this thesis has been made possible with the help of a number of people who have constantly supported and encouraged me throughout my PhD research work. I express my sincere thankfulness to all of them.

First of all, I would like to express my sincere gratitude and special appreciation to my supervisor Professor Sangeeta Bansal for her invaluable guidance and constant encouragement throughout the period of my research work. Her advice and comments have been extremely helpful in structing and improving my research work. It is because of her continuous support and consistent motivation that I have been able to execute my research work systematically and finally accomplish the completion of my Doctor of Philosophy synopsis and thesis.

I express my thankfulness to all the faculty members at Centre of International Trade and Development, School of International Studies, Jawaharlal Nehru University for their worthy comments and suggestions that facilitated further enrichment of my research work. I am very thankful to Dr. Mandira Sarma and Dr. Rashmi Barua at Centre of International Trade and Development, School of International Studies, Jawaharlal Nehru University for their valuable suggestions and comments on my research work. I am also grateful to the all professors and research students present at the seminars and conferences, where I have presented my research work, for their academic inputs and comments. I also acknowledge my thankfulness to the library and office staff at Jawaharlal Nehru University for their help and cooperation during my research work.

I have a special thanks and deep gratitude towards my family who have been a constant support and source of my strength during my research work. Their love and encouragement have always motivated me to achieve new goals in life.

Shivani Gupta

Dedicated to my Parents and Brother

Contents

List of Tables

List of Figures

Abbreviations

BMI	Body Mass Index
BRFSS	Behavioral Risk Factor Surveillance System
FAO	Food and Agricultural Organization
GMM	Generalized Method of Moments
IDF	International Diabetes Federation
IV-2SLS	Instrument Variable - Two Stage Least Squares
NCDs	Non-Communicable Diseases
NFHS	National Family Health Survey
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
WHO	World Health Organization

Chapter 1

Introduction

1.1 Background and Motivation

Improvement in health and nutritional status is an important goal for the development. The Sustainable Development Goals aim at improving nutrition and ensuring healthy lives across the population worldwide. The recent increase in overnutrition may pose a challenge to the fulfilment of these goals.[1] During the recent decades, obesity prevalence has increased across almost every country of the world and worldwide obesity prevalence has nearly tripled since 1975 (World Health Organization (WHO)).[2] In 2016, globally around 39% of the adult population was overweight and 13% was obese (WHO). Once associated with the high-income countries, overweight and obesity has now become an emerging problem among the low- and middle-income countries.

The United States continues to be amongst the highest obesity countries of the world. More than two-third of the adult population in the United States is either overweight or obese and more than one-third is obese (WHO, 2016). The age standardised obesity prevalence estimates given by WHO show that high-income countries have a considerably higher obesity prevalence as compared to the low- and middle-income countries, however, if the obesity prevalence continues to grow at the current rates then it may become a pandemic problem and may impose a considerable health and financial burden among the low- and middle-income countries.

Many low- and middle-income countries are presently facing the dual burden of inadequate nutrition in the form of widespread undernutrition and emerging overnutrition which is posing a new challenge to the health sector of these countries.

[1] Overnutrion is measured by the proportion of population which is overweight and/or obese.

[2] "At aggregate level, obesity prevalence is defined as the proportion of population having BMI ≥ 30 kg/m^2 and, overweight and obesity prevalence is defined as the proportion of population having BMI $\geq$ 25 kg/m^2. Body Mass Index (BMI) of an individual is defined as person's weight in kilograms divided by the square of his height in meters (kg/m^2). WHO International BMI classification categorises individuals having BMI < 18.5 kg/m^2 as underweight; 18.5 $\leq$ BMI < 25 kg/m^2 as normal weight; 25 $\leq$ BMI < 30 kg/m^2 as overweight and BMI $\geq$ 30 kg/m^2 as obese" (WHO).

Given the recent rise in the overnutrition, Food and Agricultural Organization (FAO) has termed "obesity as developing world's new burden". Available evidence in the Indian context points towards the rapid emergence of overnutrition problem. The overweight and obesity prevalence in India increased from 11.4% to 20.4% during 2005-15 (National Family Health Survey (NFHS)). During the same period, underweight prevalence declined from 35% to 22.5%.[3] These figures indicate that India is currently facing a dual burden of inadequate nutrition and, presently, at national level, prevalence of both overnutrition and undernutrition is high.

The increase in calorie consumption over calorie expenditure is the fundamental cause for the rise in overweight and obesity. This energy imbalance is driven by several factors such as decrease in physical activity levels, increase in consumption of calorie-intense foods and other related factors that influence the net calorie intake. Worldwide, there has been an increase in the consumption of calorie-intense foods and a decline in the physical activity levels resulting from the increase in the sedentary behaviour in the form of less strenuous workstyle, improved modes of transportation, urbanisation, etc. (WHO, 2016). Many studies have documented the effects of an increase in calorie consumption and reduction in physical activity levels on obesity rates. The study by Cutler et al. (2003) shows that calorie intake in the form of increased consumption of snacks contributes to obesity and the study by Spanier et al. (2006) states that the increase in sedentary behaviour can explain the rise in obesity.

Given the substantial rise in overnutrition during the recent decades, it is imperative to examine the consequences of the increase in overweight and obesity prevalence on health outcomes. An increase in overweight and obesity prevalence is a major risk factor for the non-communicable diseases (FAO and World Obesity Federation). These diseases include cardiovascular diseases, hypertension, diabetes, arthritis, certain type of cancers, etc. Most of these diseases are the leading causes of mortality and morbidity. The severity of the health impacts of overnutrition is reflected by WHO findings which state that overweight and obesity cause more deaths worldwide than underweight. Worldwide about 2.8 million deaths each year are caused due to overweight and obesity related factors (WHO, 2016). In India, the share of deaths attributable to non-communicable diseases has surged up during the past decade while the share of deaths

[3] Underweight prevalence is defined as the proportion of population having BMI <18.5 kg/m^2.

attributable to communicable diseases has declined (Cause of Death Statistics, India 2004-13). In the United States, more than four-fifth of the deaths are caused due to non-communicable diseases (World Bank, 2016). These observations motivate the analysis of the impact of overweight and obesity on health outcomes and identification of the major factors contributing to the rise in overweight and obesity. Also, the projections on obesity show a steady increase in the obesity prevalence until at least year 2030 (Organisation for Economic Co-operation and Development (OECD), 2017). The growth in overnutrition is a matter of concern as it adversely affects several health and economic outcomes.

In this thesis, we focus upon overnutrition and its associated health impacts. The broad objective of the thesis is to examine the effects of overweight and obesity on non-communicable diseases and health outcomes, and to identify the major factors that contribute to the rise in the overweight and obesity prevalence. Many studies have examined the impacts of overnutrition on the health outcomes, however, most of these studies are based on the high-income countries and only a little research is available in this regard for the low- and middle-income countries such as India.

Existing evidence on the relationship between overnutrition and non-communicable diseases suggests that overnutrition increases the risk of non-communicable diseases. In addition, a rise in the prevalence of non-communicable diseases as a consequence of increased overweight and obesity prevalence may also worsen the health outcomes such as death rate, longevity, years of life lost, etc. The study by Sikdar et al. (2010) provides an evidence for such effects on life expectancy and health-adjusted life expectancy. Rowley et al. (2017) document the positive association between diabetes and mortality. In the present study, our first research objective is to examine how does a rise in overweight and obesity prevalence affects longevity at a macro level. For this, we conduct a state level empirical analysis for the United States and India.

Diabetes has been growing at alarming rates in the recent years (FAO). India is presently facing considerable health risks from diabetes. About one-fifth of the urban population and one-tenth of the total population is diabetic in India (International Diabetes Federation (IDF), 2017). Recently, India has also been stated as diabetes capital of the world (Diabetes Foundation, India (DFI)). The diabetes prevalence in India has doubled across both urban as well as rural areas during 2005-15 (NFHS). This motivates the need to quantify the effect of overweight and obesity on diabetes in India. In our second research objective, we examine how does an increase in overweight and obesity affects diabetes for the population in India using an empirical framework at micro level.

Examining the potential health effects associated with the overweight and obesity in India, lastly, in our third research objective, we explore the factors that may contribute to the rise in overweight and obesity by examining the effects of urbanisation, sedentary lifestyle and consumption of calorie-intense foods on overweight and obesity in India at micro level.

1.2 Health Burden of Overnutrition

The health effects of overnutrition can be assessed in terms of prevalence of non-communicable diseases and several health indicators. An increase in overnutrition may adversely affect the health status of the population by increasing prevalence of non-communicable diseases and worsening their health outcomes.

Many studies have examined the health effects of overnutrition for the population in high-income countries. Fontaine et al. (2003) show that obesity contributes to the increase in years of life lost. Studies have also documented the adverse effects of obesity on longevity (Brunello et al., 2009; Preston and Stokes, 2011 and Lichtenberg, 2011). Obesity is also expected to lower future longevity (Preston et al., 2014 and Stewart et al., 2009). Based on the available evidence, one may expect overnutrition to have negative health impacts.

In the context of the effects of overnutrition on non-communicable diseases, much of the evidence finds a positive association between the two. Colditz et al. (1995), Geiss et al. (2017) and Rowley et al. (2017) find that the risk of diabetes is increasing in Body Mass Index (BMI) amongst the population in the United States. Sepp et al. (2014) and Malley et al. (2010) provide similar evidence for the European population. The study by Huffman et al. (2011) also shows the similar relation in the Indian context. The health consequences of overnutrition may vary with the underlying nutritional demography of the population in a country. The differences in physique of the population across different regions may affect their susceptibility towards obesity associated health risks. Gray et al. (2011) suggest that Asian population faces a higher risk of non-communicable diseases. This can be explained by higher abdominal obesity among Asian population (Olinto et al., 2017). This raises many concerns about the possible ill-health effects of the recent rise in overnutrition in India.

1.3 Economic Burden of Overnutrition

The growth in overweight and obesity imposes a burden on the healthcare costs. Available literature suggests that the impacts of obesity on disability or morbidity are far greater than the impacts on mortality (Gregg and Guraliak, 2007). Through increased morbidity, overweight and obesity contributes to the economic burden in the form of increased spending on healthcare. There exists evidence which suggests that obese individuals incur a higher health expenditure than the normal weight individuals (Withrow and Alter, 2011). The study by Bhattacharya and Sood (2011) finds that obesity substantially increases lifetime healthcare costs. Cawley and Meyerhoefer (2012) also predict that obesity raises the annual medical care costs and about 20% of the total annual healthcare costs are spent on obesity related illness in the United States. Thorpe et al. (2004) find that 27% of the increase in inflation-adjusted health expenditure is explained by the rise in prevalence and costs of obesity in the United States.

Overnutrition may also may contribute to the economic burden through other indirect ways in the form of reduced productivity in the labour market, increased pressure on the policy makers to deal with overnutrition, etc. Brunello et al. (2009) state that obesity may reduce labour market productivity through poor health status. Productivity in the labour market is directly related to the human capital in the form of health and education. An increase in obesity is expected to have an adverse impact on both health as well as education (or incentive to invest in education) through poor health conditions such as increased morbidity, lower longevity, etc. and, therefore, it is expected to lower the productivity.

Given that the increase in overnutrition may raise the spending on healthcare, it is important to analyse who pays for these increased healthcare costs. It is observed that the distribution of health expenditure burden between the government and private households varies considerably across countries. In high-income countries, a major proportion of the total health expenditure is financed by the government. However, amongst many low- and middle-income countries, the government's contribution in the total health expenditure has remained low. In India, about 30% of the total health expenditure is financed by government indicating that a substantial healthcare cost burden is borne by the private households (National Health Accounts India, 2014-15). Therefore, inadequate nutrition is expected to have both health as well as economic implications in the form of increased prevalence of diseases, poor health outcomes and increased spending on the healthcare.

Recognising the threat posed by overnutrition, several policies have been devised to fight the overweight and obesity problem. Major strategies include price interventions (taxes and/or subsidies), improvements in nutrition labelling, sensitising the population towards a healthy lifestyle through information and awareness, etc. The imposition of 8% tax on the food products containing high sugar in Mexico is one such policy example. Recently, Food Safety and Standards Authority of India (FSSAI) has proposed to display a mandatory red code or label on the packaged food products containing high levels of salt, sugar or fat. Food or nutrition labelling informs the consumer about the attributes of the product and help them in making informed choices. An appropriate and effective labelling on the food products about their health effects can modify the consumers' food choices towards healthier products. However, the effectiveness of these interventions relies on how well the targeted population embraces these policies and most importantly on the self-control issues. There exits evidence stating that taxing fat foods may not be effective and can be regressive as compared to the subsidies on the healthy foods (Cash et al., 2007).

At the global level, "WHO Global Strategy on Diet, Physical Activity and Health" has been adopted by the World Health Assembly in the year 2004 which lays down the strategies that aim at improving the diet and physical activity patterns at global, regional and local levels. The challenge posed by overnutrition in India has also been recognised and taken into the policy consideration. This is evident from the health interventions that took place in the recent years. One such policy response is imposition of 14.5% fat tax on the junk food in Kerala in the year 2016. Another major policy intervention can be seen in the form of assignment of high-sugar content drinks to the highest tax bracket of 28% under Goods and Services Tax in the year 2017. Besides a high tax liability, these goods also attract an additional 12% sin tax. Also, given the rapid emergence of overnutrition problem in India, we are likely to witness many such policies in the coming years.

1.4 Structure of the Thesis

The thesis is organised as follows. The introductory chapter discusses the context of this study along with the motivation and relevance of it. It briefly presents the research objectives of this thesis along with their findings. The second chapter presents a systematic review of the available literature in the field of health economics with a focus on overnutrition and its associated effects on health and economic outcomes. This chapter also identifies the gaps in the existing literature and discusses how the present study contributes to the literature. The third chapter discusses the recent trends in overnutrition and highlights the growing overnutrition problem across the low- and middle-income and high-income countries. The chapter also brings out the recent emergence of overnutrition problem in India at both national and state level. It also documents the widespread overnutrition problem in the United States at both national and state level. The analysis of the three main research questions of the thesis is presented in Chapters 4 to 6.

Chapter 4 examines the effect of overnutrition on the health outcomes as measured by life expectancy at birth. We hypothesise that the relationship between life expectancy at birth and obesity prevalence is concave. We test this relationship for the United States and India using an empirical framework based on state level aggregate data. At aggregate level, the effect of overnutrition on life expectancy may differ based on the demography of nutritional status across the population (that is, the distribution of population across underweight and obese categories). Therefore, to have a better understanding about the effects of obesity prevalence on life expectancy at macro level, we consider one high-income country – the United States, characterised by high overnutrition prevalence, and one low- and middle-income country – India, characterised by the dual burden of inadequate nutrition. This allows us to explore the effects of existing nutritional dynamics on the relationship between the life expectancy at birth and obesity prevalence.

For the analysis in the United States, we construct a panel data set at state level for a fifteen-year period during 2000-14. We estimate a Fixed Effects and a System

Generalized Method of Moments model. Estimation of a System Generalized Method of Moments model, allows us address to the potential endogeneity of the regressors. For the analysis in India, we construct a panel data set at the state level for the years 2005-06 and 2015-16. We estimate a Fixed Effects and a Two-Stage Least Squares model with instrument variable. For this purpose, we consider households' possession of assets as an instrument for the overweight and obesity prevalence. The results suggest that life expectancy at birth has a concave relationship with the overnutrition prevalence across both the countries. More interestingly, it is found that India is presently on the upward sloping segment of this concave curve while the United States lies on the downward sloping segment. This suggests that the extent of overnutrition governs the relationship between longevity and overnutrition.

Chapter 5 examines the micro level relationship between overweight and obesity, and diabetes for the population in India. We estimate the average marginal effect of an additional unit gain in BMI on the diabetes status of an individual. We use two alternative indicators for measuring diabetes across population – self-reported diabetes status and blood glucose levels (ordinal measure). We provide an evidence for the causal effect of overnutrition on diabetes. We address the potential endogeneity in the relationship between BMI and diabetes status of an individual by instrumenting individual's BMI with a non-biologically related household member's BMI.

For the purpose of analysis, we consider a nationally representative data set from the fourth round of National Family Health Survey for the year 2015-16. We apply different econometric specifications to estimate the average marginal effects of BMI on diabetes. Estimating an IV-Probit model, we find that the likelihood of being diabetic is thrice among the overweight or obese individuals as compared to the non-overweight individuals. We also find that the level of risk of being diabetic differs across genders, regions and wealth quintiles and the effects are more severe among population in the urban areas, population belonging to the richest wealth quintile and men.

Chapter 6 examines the effect of urbanisation on the BMI levels of the individuals. The chapter also examines the effect of sedentary lifestyle and consumption of calorie-intense foods on the overweight and obesity status for the population in India. Urbanisation is a potential factor that may lead to overweight and obesity (Garden and Jalaludin, 2009 and Banwell et al., 2009). In this study, urbanisation is measured in terms of the place of residence of the individuals (urban or rural). We estimate the effect of living in the urban areas on the BMI levels of the individuals. We address the endogeneity problem in the relationship between living in urban areas and BMI levels by exploiting an exogenous variation in the form of road network availability at the state level.

For this analysis, we use the same data set as used in Chapter 5. We estimate a Two-Stage Least Squares model in which the place of residence of the individual is instrumented by the total road length per kilometer of state area. In addition, to estimate the effects of sedentary lifestyle and consumption of calorie-intense foods on overweight and obesity status of the individuals, we estimate a logistic regression model. For this purpose, we include proxy measures for sedentary lifestyle such as frequency of television watching and household's ownership of assets which may influence physical activity or promote sedentary lifestyle – private transportation and certain electronic items. Also, consumption of calorie-intense foods is measured by daily or weekly consumption of fried foods and aerated drinks. Our results suggest that population living in the urban areas is at a higher risk of overweight and obesity as compared to the rural population. We find that population indulging in sedentary behaviour is more likely to be overweight or obese. We could not find evidence for the effect of consumption pattern on the overweight and obesity status of the individuals.

Finally, Chapter 7 summarises the key findings and concludes the thesis.

Chapter 2

Literature Review

This chapter presents a systematic review of the studies that have examined the health and economic impacts of overnutrition. We also discuss studies that have analysed the factors that contribute to the rise in overnutrition.

This chapter is divided into four sections. Section 2.1 discusses the effect of overnutrition on the health outcomes such as longevity and mortality. Section 2.2 discusses the link between overnutrition and Non-Communicable Diseases (NCDs). Section 2.3 discusses the factors that may lead to a rise in the overnutrition. Section 2.4 documents how the rise in overnutrition contributes to the economic burden.

2.1 Overnutrition: Effect on Health Outcomes

Available literature that examines the effect of overnutrition on the health outcomes shows that overnutrition adversely affects health outcomes such as longevity, death rate, years of life lost, quality of life, etc.[4] The studies such as Brunello et al. (2009), Lichtenberg (2011) and Bansal and Zilberman (2016) have investigated these effects for longevity. The study by Brunello et al. (2009), using cross-country data for the United States, Japan and ten European countries during the period of 1979-2004, examine the effect of obesity prevalence on life expectancy at birth. To address the potential endogeneity of the explanatory variables, they define life expectancy as a function of one period lagged values of the explanatory variables. The results show a negative relationship between obesity and life expectancy at birth. Another study by Lichtenberg (2011) also provides an evidence on the negative relationship between overweight and obesity prevalence, and life expectancy by estimating a weighted-least squares model using state level data for the United States during the period of 1991-2004. Bansal and Zilberman (2016) have developed an analytical model which defines average life expectancy as a non-linear function of obesity prevalence. They also test

[4] Quality of life can be measured by health-adjusted life expectancy (HALE) and quality-adjusted life years (QALYs).

their model empirically at macro level using cross-country panel data and find a statistically significant concave relationship between obesity prevalence and life expectancy at birth.

Studies have also examined the impact of obesity on years of life lost (YLL) and deaths or mortality attributable to obesity (Fontaine et al., 2003; Mayhew et al., 2009; Preston and Stokes, 2011; Flegal et al., 2004; Allison et al., 1999 and Thornton and Rice, 2008). Fontaine et al. (2003) discuss the adverse effects of obesity on life expectancy across different age groups in the United States. They show that YLL due to obesity increase with body mass index but at higher values of body mass index, YLL due to obesity decline with age. This indicates the heterogeneity in the effects of obesity on life expectancy across age groups. Mayhew et al. (2009) estimate the effect of excess body fat, measured by body mass index and waist-to-height, on YLL extracting data from the Health and Lifestyle Survey for the United Kingdom. They find a J-shaped relationship between these two measures of obesity and YLL across both male and female population.

Preston and Stokes (2011) estimating a population attributable fraction find that obesity contributes to the low life expectancy at 50 years of age in the United States.[5] Flegal et al. (2004) also estimate the deaths attributable to the obesity in the United States using population attributable fraction index and find that relatively higher deaths are caused due to obesity among the older age groups as compared to the younger age groups. Another study by Allison et al. (1999) based on five prospective cohort studies find that about 2.8 lakhs annual deaths are attributable to obesity among US adults. Thornton and Rice (2008) estimate the effect of lifestyle factors on crude death rate in the United States and find a highly statistically significant positive effect of obesity on crude death rate.

Some studies have also investigated the effect of obesity on the future life expectancy. Preston et al. (2014) examine the effect of past and future changes in obesity and smoking on future life expectancy for the United States and find that obesity has a

[5] "$PAF_i = \frac{\sum_j (C_{ij}M_{sj} - C_{ij}^{*}M_{sj})}{\sum_j (C_{ij}M_{sj})}$; where i is an indicator for each country, age, and sex combination; C_{ij}=proportion of population i in BMI category j, M_{sj}=death rate in BMI category j, and C_{ij}^{*}=proportion of population i in BMI category j if all individuals above the optimal BMI were redistributed to the optimal category" (Preston and Stokes, 2011).

negative effect on the future life expectancy, however, this loss in life expectancy is expected to get more than off-set by the gains accruing to reduced smoking. Another study by Stewart et al. (2009) finds a contrary result. They find that the negative effects of increased obesity would outweigh the positive effects of reduced smoking on future life expectancy in the United States. However, the common inference is that obesity is expected to increase and impose a detrimental effect on the future longevity.

From the above discussion, we find that most of the evidence on the effects of obesity on longevity and mortality indicators comes from empirical studies based on the population in the United States and other high-income countries. We contribute to the existing literature by providing an evidence on the effects of overweight and obesity on longevity in India, where high underweight prevalence co-exists with rapidly rising overweight and obesity prevalence. We also conduct a similar analysis for the United States.

2.2 Overnutrition and Non-Communicable Diseases

Overnutrition is linked to the increased risk of NCDs such as diabetes, hypertension, heart diseases, certain type of cancers, etc. (Huffman et al., 2011; Colditz et al., 1995 and Dhana et al., 2016). Huffman et al. (2011), based on a cohort study, show that both obesity and diabetes prevalence have increased among married women in South Delhi, India during 1998-2002. Estimating a logistic regression model, they find that an increase in BMI has a statistically significant positive impact on the probability of having hypertension and diabetes. Colditz et al. (1995) using a prospective cohort study on women in the United States during 1976-90 find that the risk of diabetes is increasing in BMI and every 10 kg rise in weight after 18 years of age doubles risk of diabetes. Geiss et al. (2017) examine the recent trends in obesity and diabetes at county level for the United States during 2004-12 and find that the growth in the incidence of diabetes is slowing down but has not reversed (indicating that diabetes will increase but at comparatively low rates) and the level of risk still remains high. The study by Rowley et al. (2017) using National Census data forecasts the changes in diabetes, and finds that diabetes is expected to increase in the United States during 2015-30. Sepp et al. (2014) find a statistically significant positive relationship between blood glucose levels and BMI for a sample of elderly population (65 years of age or above) in Southern Estonia, Europe. Another study by Malley et al. (2010) conducting a cross-sectional study across young obese children in Europe finds a positive association between BMI and fasting blood glucose levels. All these studies suggest that NCDs prevalence is likely to escalate with the rise in overweight and obesity.

Studies have also examined how the increased prevalence of NCDs affects health outcomes. Sikdar et al. (2010) find that diabetes reduces both life expectancy and health adjusted life expectancy based on a representative data set for a province in Canada. Eliminating diabetes improves life expectancy by more than one year and health adjusted life expectancy by up to two years. The study by Rowley et al. (2017) obtains similar results considering mortality as the health indicator instead of life expectancy. They find that annual deaths associated with diabetes are expected to increase during 2015-30 in the United States. Andreyeva et al. (2007) investigate the effect of obesity on chronic health conditions across ten European countries using a logistic regression model. They find that people who are moderately or severely obese are almost twice more likely to report poor health status, disability or some chronic health conditions. Studies also advocate that the risk of NCDs is higher among Asian population as compared to the population in the European countries, Canada and the United States. (Razak et al., 2007 and Gray et al., 2011). Razak et al. (2007) consider a random sample of 1078 individuals from Asian and European countries and find that for a given BMI, Asian population is more likely to have elevated blood glucose levels and blood pressure levels as compared to the European population. Gray et al. (2011) also obtain similar results. They consider 4688 White Europeans and 1333 South Asians in the age group 40-75 years who are residents of the United Kingdom during 2004-07. They find that for a given BMI, South Asian population faces a higher risk of cardiovascular diseases and diabetes as compared to the European population. Also, diabetes is more prevalent in South Asian countries like India, Pakistan and Bangladesh as compared to the European countries. Patel et al. (2001) compare the perceived and actual obesity among diabetic and non-diabetic women across these two regions and find that awareness about obesity is low among South Asian women than European women. Also, women aware about their diabetes status have more realistic body weight perception.

From the above discussion, it is found that overnutrition increases the risk of NCDs and this risk varies across population in different regions. Much of the evidence on the effects of overweight and obesity on NCDs comes from high-income countries and is based on small population size, therefore, the results obtained from these studies cannot be generalised for the entire population. We fill this gap in the literature by examining the link between BMI and NCDs (with a focus on diabetes) for the population in India.

2.3 Overweight and Obesity: Causes

Available literature identifies several potential factors that may lead to the overnutrition problem. These factors include consumption preferences, reduced food costs (both monetary and time cost), screen time, occupational factors, type of transportation facility used, place of residence, urban sprawl, built environment, etc. All these factors affect calories intake and/or calories expended and thereby affect the overweight and obesity status of the individuals.

Let us first discuss the consumption preferences. Cutler et al. (2003) find that the changes in the food consumption may explain the rise in obesity in the United States during 1977-96. An increase in the intake of calories has been observed primarily in the form of an increase in meals consumed per day, especially consumption of snacks. They state that the technological improvements in the food industry can explain this rise in calorie intake. The division of labour along with improved packaging and preservation methods has led to the mass production of food. This resulted into the reduced food costs and subsequently led to higher consumption of food, especially snacking items, and contributed to the higher obesity in the United States. Lakdawalla et al. (2005) using a theoretical model state that technological change leads to obesity by reducing food prices and physical labour. Schmidhuber and Shetty (2005) analysing the nutritional transition among the developing countries, state that food industry influences the food consumption patterns and the recent growth has made cheap meals more accessible. It is also found that subsidised agriculture contributes to the obesity by lowering food prices and making it less expensive to increase portion sizes of food (Alston et al., 2006).

There exists evidence about the rise in energy or calories intake due to the increase in liquid consumption. Fletcher et al. (2011) state that increased consumption of soda or soft drinks results into higher obesity in the United States. The study by McCrory et al. (2002) states that liquid consumption (sweetened or aerated beverages) has zero compensation effect in the sense that liquid consumption does not reduce the food/calorie intake in the subsequent meals. This suggests that consuming drinks along with food does not reduce tendency to eat or does not compensate for extra calories consumed in the subsequent meals. However, for solid food items, compensation of extra calories takes place by eating less calories in the subsequent meals on same day or next day(s). These findings strongly suggest a restraint on liquid consumption in the form of sweetened or aerated drinks.

Changes in the dietary intake may also influence obesity rates. Traill and Mazzocchi (2005) emphasise the importance of understanding the interaction between dietary choices of people and their health outcomes. Meenakshi (2016) investigates the trends in malnutrition and micronutrients intake in India during 1992-93 and 2015-16, and finds that the overnutrition prevalence has increased, micronutrients intake has remained low and calorie intake in the form of sugar and oils has increased. Variyam et al. (1999) using an Ordinary Least Squares model find that health information influences health behaviour among the population in the United States. They find that individuals who are more aware of the importance of avoiding fat show less fat intake.

Another factor that may cause obesity is restaurant food or eating out behaviour. Many studies find that eating out leads to higher obesity rates as restaurant food has a higher calorie content and portion size (Young and Nestle, 2002; Mello et al., 2006 and Chou et al., 2004). However, Anderson and Matsa (2011) estimating a Two-Stage Least Squares model using interstate highway proximity to instrument restaurant access, find no evidence for the causal relationship between obesity and restaurant food consumption for the population in the United States.

Philipson and Posner (2008) reviewing available literature on obesity suggest that analysis of obesity must take into account changes in physical labour both at home and work. Many studies have examined the link between obesity and energy expenditure (Spanier et al., 2006; Lopes et al., 2014 and Owens et al., 2013). The study by Spanier et al. (2006) analyse the trends in the screen time for a sample comprising of young adults in Canada and find existence of sedentary behaviour. They suggest that the current measurements of energy expenditure such as time spent on television watching, computers, and other physical activities including walking, cycling, etc. should also include low physical intensity activities so as to capture the sedentary behaviour more precisely. Lopes et al. (2014) examine the trends in screen time (television watching and computer or video games) on week days and weekends for Southern Brazilian students for the years 2001 and 2011. They find a significant increase in computer or video game use and reduction in television watching in 2011. The trends on weekends are even worse and substantially higher percentage of students in age group 15-19 years watched television and used computer or played video games. Owens et al. (2013) conduct a prospective longitudinal study to analyse the trends in sedentary behaviour among adolescents transitioning out of compulsory education in the United Kingdom. Findings show that the physical activity declined significantly though the changes in screen time did not show any significant change. Results of binary logistic regression indicate that the decline in physical activity is less likely among females.

Occupational characteristics may affect overweight and obesity through energy expenditure. Griffiths and Bentley (2001) find that agricultural and manual female labourers are less likely to be overweight or obese as compared to the women working at offices on clerical, sales and other related posts in Andhra Pradesh, India. Also, a vast section of empirical literature has examined how obesity affects wage earnings in the labour market (Brunello et al., 2009; Cawley, 2000 and Lindeboom et al., 2009). Brunello et al. (2009) examine the labour market discrimination due to obesity (which is an observable characteristic) and find a negative association between obesity and earnings among European countries. Cawley (2000) goes a step further and examines that whether it is the unobserved factors that are causing both higher obesity and lower earnings or is it the case that higher body weight causes lower wages. The study applies an instrument variable approach among women in the United States. The study instruments mother's body weight using the BMI of one of her children and finds that higher weight does lower wages, however, no such evidence is found on the probability for getting employment. Lindeboom et al. (2009) also attempt to establish a causal link between obesity and lower wages using National Child Development Study for individuals in the Great Britain. They use instrument variable approach and consider parent's obesity status as an instrument. They find a negative but statistically insignificant association between wages and obesity.

Existing literature on obesity and transportation suggests that usage of public transit is associated with lower obesity (Teimann et al., 2008 and She et al., 2017). Teimann et al. (2008) show that areas where more people complete their journey to work by walking, biking or taking public transit have lower overweight population. She et al. (2017) estimate the effect of public transit usage on the obesity prevalence and find that an increase in the usage of public transit significantly reduces the obesity prevalence.[6]

Studies have also investigated for the effects of urbanisation and built environment on obesity (Sengupta et al., 2015; Garden and Jalaludin, 2009; Kostova, 2011 and Popkin et al., 2012). Sengupta et al. (2015) provide an evidence for the effect of the place of residence on overweight and obesity among Indian women using data from NFHS 2005-06. They find that women residing in large cities or capitals are more prone towards overweight and obesity. Garden and Jalaludin (2009) find that the urban sprawl (based on population density) increases the risk of being overweight or obese for the population in Sydney, Australia.[7] Kostova (2011) examines the link between built environment and obesity. The study estimates the impact of residential sprawl and proximity of local parks on the physical activity levels and obesity status using a linear probability and Two-State Least Squares model. They use historical park size growth and park acquisition rate as instruments. The paper finds no evidence for the causal impact of residential sprawl and proximity of local parks on physical activity and obesity. The study by Popkin et al. (2012) summarises the available literature on obesity and discusses the obesity pandemic among the developing countries. The paper documents the rapid shift in diet and physical activity levels as the contributors to obesity, and states that urbanisation could be driving the obesity in low- and middle-income countries.

It can be inferred that factors which affect the net calorie intake have an influence on obesity, however, what exact factors play a dominant role in determining the obesity rates may vary across different populations. Much of the literature on the obesity is based on high-income countries and evidence in respect of low- and middle-income countries is comparatively limited. We fill this gap by providing an evidence on the factors which affect the overweight and obesity status for the population in India.

[6] This study considers county level estimates for the United States from BRFSS and National Household Travel Survey data sets.
[7] They focus on individuals who are staying in or shifting to areas with less population density.

2.4 Overnutrition and Healthcare Costs

Previous sections have documented the health burden associated with overnutrition, however, overnutrition may also impose an economic burden in the form of increased spending on healthcare. Overnutrition is associated with increased risk of NCDs and it may result into a rise in health expenditures through increased incidence of these diseases. That is, besides imposing real costs in the form of poor health conditions, reduced quality of life and high mortality, overweight and obesity also imposes a substantial monetary cost.

There exists ample evidence on the effects of obesity on healthcare spending (Thornton and Rice, 2008; Cawley and Meyerhoefer, 2012; Bhattacharya and Sood, 2011 and Yesudian et al., 2014). Thornton and Rice (2008), using data for 50 states of the United States for year 1998 and estimating both Two-Stage and Three-Stage Least Squares models, find that obesity has a positive effect on the healthcare spending. The study by Cawley and Meyerhoefer (2012) addresses the endogeneity of the relationship between obesity and medical care costs by adopting an instrumental variable approach. They instrument respondent's weight using weight of his biological relative (child) and find that obesity raises annual medical costs in the United States. Bhattacharya and Sood (2011) estimate the lifetime healthcare costs associated with obesity in the United States. They use a Future Elderly Model to estimate these costs and find that till age of 65 years obesity increases lifetime healthcare costs substantially over lifetime medicare costs but after 65 years of age lifetime healthcare costs decreases below medicare costs as the principle costs are now being borne by the other people in the obese individual health insurance.[8] Also, if YLL due to obesity are assigned a value then the marginal costs of becoming obese in terms of reduced life expectancy shows an upward trend. The study indicates a huge cost burden imposed by increased medicare costs especially among the elderly population. Yesudian et al. (2014) review studies that have estimated the costs of diabetes in India and state that the total direct costs per year varies from Rs. 7000-25000 (this amount differs based on the components included in the cost

[8] "Future Elderly Model (FEM) has two components: (1) Health is defined by chronic conditions and functional status is measured by activities of daily life like eating, bathing, dressing etc. Here, future disease acquisition is modelled as a function of baseline health and functional status with help of first-order Markov process. (2) Healthcare cost is modelled as function of current health, functional status, risk factors and demographics" (Bhattacharya and Sood, 2011).

measurements by various studies). Dall et al. (2014) estimate the economic burden associated with diabetes and prediabetes in the United States and find that total burden increased by 48% during 2007 and 2012.

Obesity may also contribute to the increased healthcare costs through expenditure on prescription drugs. The real per capita expenditure on prescription drugs rose by 84% in the United States during 1990-98. Vandegrift and Datta (2006) examine the possible cause for this increase. Results obtained from the estimation of a Fixed Effects model show that obesity explained nearly one-tenth of the increase in the expenditure on the prescription drugs. Berndt (2001) and Reinhardt (2001) state that people tend to use lower priced generic drugs in situations where they bear the costs of drugs in the form of out-of-pocket expenditure but when they are covered under some insurance scheme, they are more likely to go for prescription drugs which are relatively expensive. This adds to healthcare costs burden associated with obesity. Lundin (2000) also finds similar results that people having full insurance coverage have less incentive to look for low-cost generic drugs and may spend more on prescription drugs. Kaufman et al. (2002) find that usage of prescription drugs is high among elderly population due to a higher incidence of diseases and chronic conditions.

Obesity also imposes an external cost on the society (Bhattacharya and Sood, 2011). Insurance has an associated moral hazard as it increases the demand for healthcare by reducing out-of-pocket health expenditures, thus some of the healthcare costs are borne publicly as well (especially by those who face relatively lower health risks in the pool of obese people who have purchased the insurance). However, if health insurance premiums are risk adjusted then the costs are internalised and do not distort decisions regarding body weight (Ehrlich and Becker, 1972 and Bhattacharya and Sood, 2006).

From the above discussion, we find substantial evidence on the effects of obesity on health expenditure, therefore, one may expect a rise in the overnutrition to impose a considerable economic burden by increasing the spending on healthcare.

Chapter 3

Rising Trends in Overnutrition

Inadequate nutrition both in the form of undernutrition and overnutrition has remained an important part of the policy consideration worldwide. High-income countries such as the United States and European nations have been facing a major health concern in the form of overnutrition, while countries like India and many other low- and middle-income countries have been facing a major challenge in the form of undernutrition. The recent growth in the overnutrition has now begun to pose a new upcoming challenge to the health sector of India.

This chapter is divided into five sections. Section 3.1 discusses the trends in obesity prevalence in a global context. Section 3.2 examines the dual burden of inadequate nutrition in India during the recent decade at both national and state level. Section 3.3 documents the persistent and growing overnutrition problem in the United States by examining the data at both national and state level. Section 3.4 discusses the policies to tackle the overweight and obesity problem. Finally, Section 3.5 presents the concluding remarks.

3.1 Overnutrition: Global Scenario

At the global level, the average BMI levels of the population across almost every country of the world have been rising during the recent decades (WHO). The proportion of population which is overweight (having $25 \leq$ BMI <30 kg/m^2) and obese (having BMI ≥ 30 kg/m^2) has also increased. WHO estimates of the age standardised overweight and obesity prevalence show a very clear picture of the rising overweight and obesity problem worldwide. Therefore, one can no longer perceive obesity as high-income countries' health problem, and obesity is now a worldwide phenomenon and an upcoming challenge in many countries.

Figures 3.1 and 3.2 present the age-standardised obesity prevalence for selected low- and middle-income and high-income countries, respectively, during 1980 and 2016.[9] Comparing the obesity prevalence across the two types of countries, we find a stark difference in the obesity prevalence with high-income countries having substantially higher levels of obesity prevalence. The obesity prevalence has been rising persistently across all the countries, however, the rate of increase in the obesity prevalence has been slowing down among the high-income countries and has been rising among the low- and middle-income countries. We also analysed the obesity prevalence for all the countries of the world and observed an increase in the obesity prevalence in almost every country during the same period.

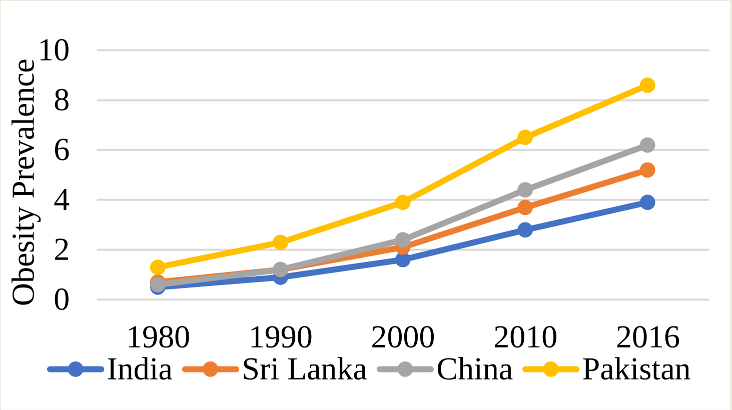


Figure 3.1: Obesity Prevalence for Selected Low- and Middle-Income Countries during 1980-2016

Source: Figure constructed by authors using the data from Global Health Observatory Data Repository, WHO.

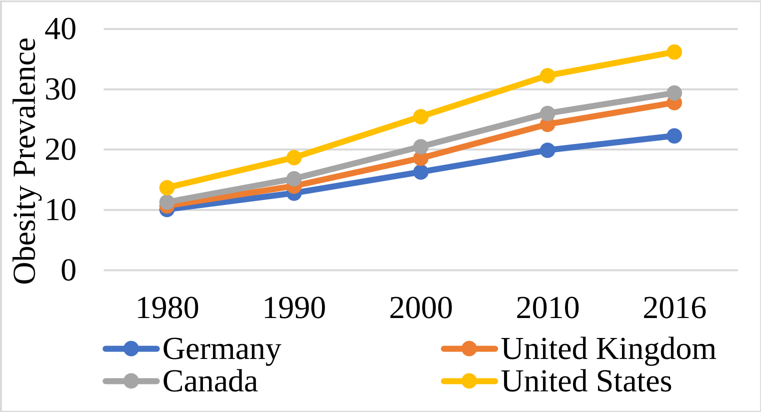


Figure 3.2: Obesity Prevalence for Selected High-Income Countries during 1980-2016

Source: Figure constructed by authors using the data from Global Health Observatory Data Repository, WHO. Note: United Kingdom's obesity estimate also includes Northern Ireland.

[9] The countries are categorised as low- and middle-income or high-income countries according to the World Bank country classification by income (Note: Income in year 2017 is considered for the classification). Here, low- and middle-income countries include - low-income economies, lower-middle-income economies and upper-middle-income economies (World Bank, 2017).

This rise in the obesity prevalence worldwide is expected to contribute to the increase in the prevalence of NCDs thereby worsen the mortality and morbidity indicators. WHO estimates show that more than two-third of the deaths at global level in the year 2016 were due to NCDs. Besides the health burden, obesity is also associated with an additional burden of increased spending on the healthcare. One may expect that a rise in overnutrition is likely to change the distribution of health burden more towards NCDs and also impose additional healthcare costs.

Next, we present country specific analysis of nutritional trends, first, for India in Section 3.2 and then for the United States in Section 3.3.

3.2 Nutritional Transition in India

Available evidence from India points towards the rapid emergence of overweight and obesity prevalence problem. However, India has always been characterised by high undernutrition in the past, but due to the recent increase in the overnutrition, India is now being said to be facing a dual burden of inadequate nutrition – high undernutrition coexisting with overnutrition.

We, first, examine the nutritional status at the all India level, and then examine the heterogeneity across states. We also bring out gender differences in malnutrition. Here, we examine the change in nutritional trends experienced in India during the period of 2005-06 and 2015-16.

3.2.1 Rapid Emergence of Overnutrition in India

Malnutrition at All India Level

In this section, we, first, analyse the average prevalence of overnutrition at all India level and then examine the BMI distribution of the Indian population. Here, we measure overnutrition by overweight and obesity prevalence, and undernutrition is measured by underweight prevalence.

The overweight and obese population in India (having age 18 years or above) increased from 14.2% in 2005 to 19.1% in 2015 (WHO) which amounts to an increase of about 35%. Table 3.1 shows the proportion of total population (age group 15-49 years), which is underweight (BMI < 18.5 kg/m^2) and overweight or obese (BMI ≥ 25 kg/m^2) for the period of 2005-06 and 2015-16. These estimates are extracted from NFHS data. These values show a sharp rise in the overweight and obese population over the period of ten years. The overweight and obesity prevalence increased from 11.4% to 20.4% while underweight prevalence declined from 35% to 22.5% for the total population. Overnutrition among males doubled, and increased by 65% among females. Though the growth in the prevalence of overnutrition is higher among males but it may be noted that females have persistently higher overweight and obesity prevalence as compared to males. We also observe a considerable decline in the underweight population. Examining these trends, we find a nutritional pattern in the form of shift away from undernutrition towards overnutrition. This compels us for further analysis, especially at disaggregated level, to better understand the observed nutritional shift and make more conclusive inferences.

Table 3.1: Underweight, and Overweight and Obesity Prevalence across Total Population (15-49 years age group) for the Period of 2005-06 and 2015-16

	Underweight Prevalence		**Overweight and Obesity Prevalence**	
	2005-06	**2015-16**	**2005-06**	**2015-16**
Total Population	35%	22.5%	11.4%	20.4%
Male	34.2%	20.2%	9.3%	18.9%
Female	35.5%	22.9%	12.6%	20.7%

Source: Compiled by authors using NFHS data.

Table 3.1 documents how the proportion of overweight and obese population has changed over the ten-year period. It provides a macro level picture but does not capture how BMI is changing for individuals. Another way of capturing the change in nutritional status of Indian population is to examine the distribution of BMI. Therefore, we now plot the BMI distribution for the population in India by extracting individual level data from the Demographic and Health Survey for years 2005-06 and 2015-16, and compare these distributions. Demographic and Health Survey of India, namely, NFHS. Figure 3.3 illustrates the BMI distributions. It shows that the BMI distribution has shifted towards right and the mass of the population with higher BMI values has increased (specifically for BMI ≥ 25 kg/m^2) and the mass of the population with lower BMI values (BMI < 18.5 kg/m^2) has declined during 2005 and 2015. This shift of BMI distribution indicates a transition away from undernutrition towards overnutrition in

India. The mean BMI increased by almost one unit, from 20.9 kg/m^2 to 21.7 kg/m^2. In the next section, we analyse the heterogeneity in the nutritional transition across states.

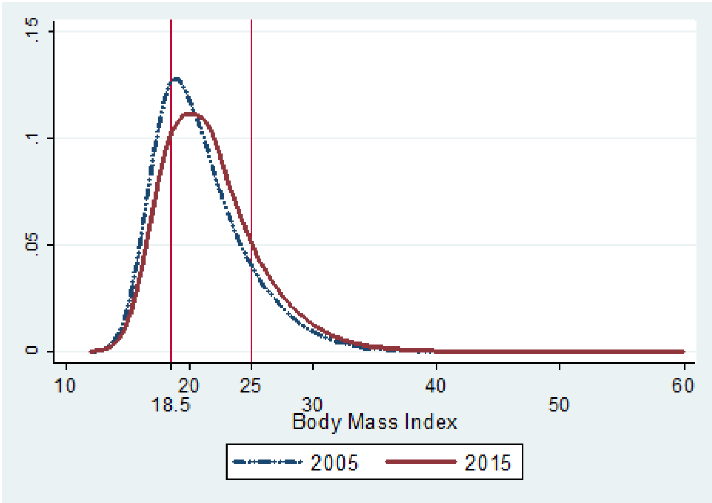


Figure 3.3: BMI Distribution for the Total Population in India for the years 2005 and 2015

Source: Figure constructed by authors using NFHS data.

3.2.2 State level Analysis

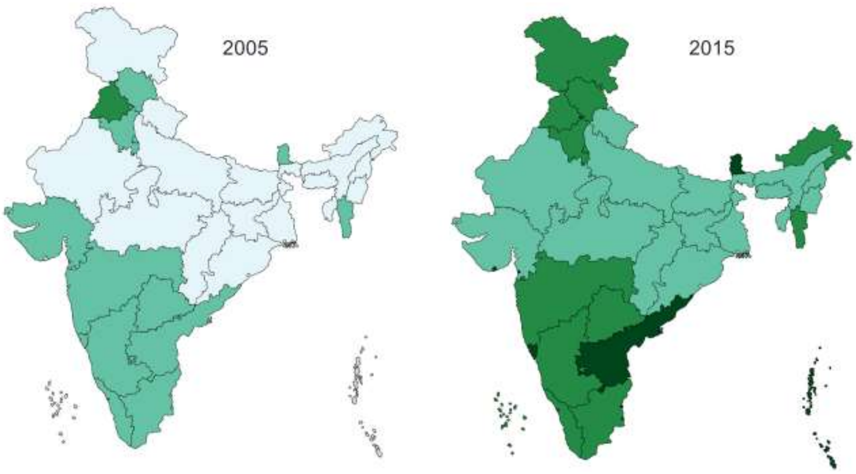
We, first, graphically present the extent of overnutrition across states. For this, on a map of India, we colour coded the states with specific overweight and obesity prevalence using an online map tool (mapchart.net). We divided the overweight and obesity prevalence in four exclusive intervals and assigned a specific colour code to each interval, as shown below the maps. Each interval indicates the percentage of population having BMI $\geq$ 25 kg/m^2 and the intervals considered are 0-9.9, 10-19.9, 20-29.9, and 30 or above. Figure 3.4 presents these maps for the years 2005 and 2015 for male and female population. We find a stark rise in the overweight and obesity prevalence among several states. It can be observed that most states have transited from below 10% or 20% prevalence in 2005 to above 20% or 30% prevalence by 2015 for both male and female population. These results are consistent with the shift in the BMI distribution discussed above since every state is found to have experienced a rise in the overnutrition. Although every state experienced a rise in the overweight and obesity prevalence, the increase has been much higher in some states.

For male population, Punjab was the only state having above 20% overweight and obesity prevalence (22.2%) in year 2005. By 2015, more than 15 states have an overweight and obesity prevalence of above 20% and states like Goa, Andhra Pradesh and Sikkim have overweight and obesity prevalence of above 30%. Further, more than half of the states had a below 10% overweight and obesity prevalence in 2005 but by year 2015 none of the states have an overweight and obesity prevalence of below 10%. This pattern again suggests the shift in nutritional status towards higher overnutrition even at state level. We observe similar trends among female population.

Another interesting observation is in the form of regional patterns. There exists a notable difference in the North, Central and North-Eastern regions (mostly comprising of BIMARU and EAG states) and Southern region, with Southern region having a relatively higher overweight and obesity prevalence.[10] This trend continues in year 2015 as well. This prompts questions on the role played by diet and lifestyle patterns across these regions. The implication of this is that nutritional policies must take into account this heterogeneity and many states in the North, Central and North-Eastern regions do not require policies that aim at reducing calorie consumption.

[10] BIMARU states refer to states having poor economic conditions. These states include Bihar, Madhya Pradesh, Rajasthan and Uttar Pradesh. Empowered Action Group (EAG) states include eight socioeconomically backward states of India. These states include Bihar, Chhattisgarh, Jharkhand, Madhya Pradesh, Orissa, Rajasthan, Uttarakhand and Uttar Pradesh.

Proportion of Overweight or Obese Males



Proportion of Overweight or Obese Females

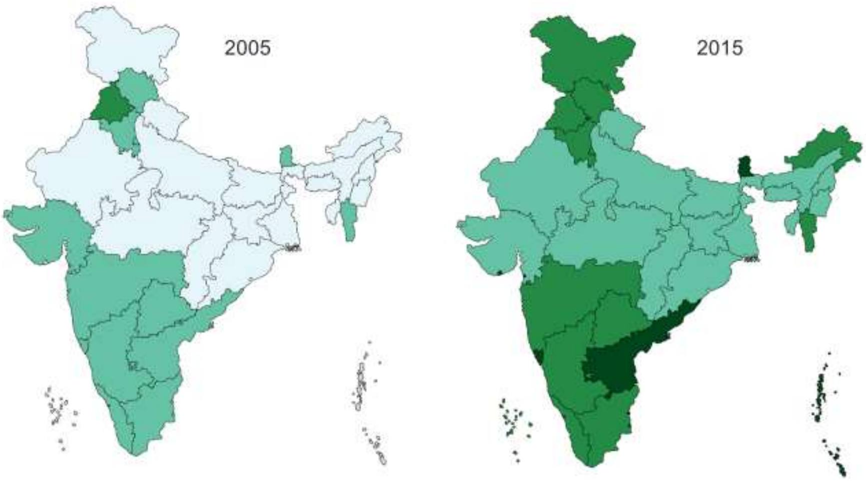
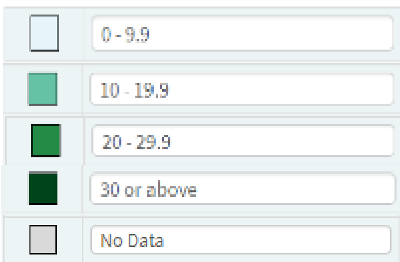


Figure 3.4: Proportion of Overweight or Obese Population in India for the years 2005 and 2015

Source: Figure constructed by authors using NFHS state level data.



Reversal of Trend – The Nutritional Transition

For better understanding about the change in the nutritional status, we now bring in the underweight prevalence, and overweight and obesity prevalence together and conduct a comparative analysis of the changes in underweight, and overweight and obesity prevalence during the recent decade for male and female population. Table 3.2 presents these estimates at state level extracted from NFHS for years 2005-06 and 2015-16. Table 3.2 presents the estimates for the states that had a higher overweight and obesity prevalence than underweight prevalence by year 2015-16. In the table, the first column reports underweight prevalence among males for 2005-06 and this is compared with overweight and obesity prevalence among males for 2005-06 given in the second column. The relationship between the two is indicated by an inequality sign between the two columns for each state. Similarly, the estimates for female population are presented in columns 3 and 4 for year 2005-06. Columns 5 to 8 can be interpreted in similar way for year 2015-16.

A striking pattern is observed in the form of Reversal of Nutritional Trend in India. About 15 states in India which had a higher underweight prevalence than overweight and obesity prevalence in year 2005-06 have witnessed a stark change in this pattern and by year 2015-16 all these states had a higher overweight prevalence as compared to underweight prevalence. This trend is similar across both male and female population. This is indicated by the reversal of sign of inequality across the two years among both males and females. This striking change witnessed in about half of the Indian states during a decade's time is quite intriguing yet alarming. States like Kerala, Punjab and Delhi continue to have a higher overweight and obesity than underweight prevalence since 2005. Therefore, we may conclude that presently more than half of the Indian states, 18 states or 24 states plus union territories, are experiencing greater burden due to overnutrition as compared to the undernutrition. Remaining 11 states continue to have a higher underweight prevalence even in year 2015. The percentage increase in overweight and obese population has varied considerably across states with some states experiencing over 100% increase. States like Andhra Pradesh, Delhi, Goa, Kerala, Tamil Nadu and Punjab are among the states having the highest overnutrition in India in year 2015-16. It is found that the nutritional trends have overturned to a considerable extent, and a sharp and continuing rise in the prevalence of overweight and obesity along with a persistent decline in the underweight prevalence is witnessed in India. If these trends continue, then soon overnutrition would be a dominating burden throughout India.

Table 3.2: Nutritional Transition across States during 2005-06 and 2015-16 among Male and Female Population

States	2005-06						2015-16					
	Males			Females			Males			Females		
	(1)		(2)	(3)		(4)	(5)		(6)	(7)		(8)
	BMI < 18.5		BMI ≥ 25	BMI < 18.5		BMI ≥ 25	BMI < 18.5		BMI ≥ 25	BMI < 18.5		BMI ≥ 25
Andhra Pradesh	30.8	>	13.6	33.5	>	15.6	14.8	<	33.5	17.6	<	33.2
Arunachal Pradesh	15.2	>	7.1	16.4	>	8.8	8.3	<	20.6	8.5	<	18.8
Delhi	15.7	<	16.8	14.8	<	26.4	17.7	<	24.6	14.8	<	33.5
Goa	24.7	>	15.5	27.9	>	20.2	10.8	<	32.6	14.7	<	33.5
Haryana	30.9	>	10.8	31.4	>	17.4	11.3	<	20	15.8	<	21
Himachal Pradesh	29.7	>	10.6	29.9	>	13.5	18	<	22	16.2	<	28.6
Jammu and Kashmir	28	>	6.2	24.6	>	16.7	11.5	<	20.5	12.1	<	29.1
Karnataka	33.9	>	10.9	35.4	>	15.3	16.5	<	22.1	20.7	<	23.3
Kerala	21.5	>	17.8	18	<	28.1	8.5	<	28.5	9.7	<	32.4
Manipur	16.3	>	9.2	14.8	>	13.3	11.1	<	19.8	8.8	<	26
Meghalaya	14.1	>	5.9	14.6	>	5.3	11.6	>	10	12.1	<	12.2
Mizoram	9.2	<	11.4	14.4	>	10.6	7.3	<	20.9	8.4	<	21.1
Nagaland	14.2	>	5.7	17.4	>	6.4	11.4	<	13.9	12.3	<	16.2
Punjab	20.6	<	22.2	18.9	<	29.9	10.9	<	27.8	11.7	<	31.3
Sikkim	12.2	>	11.9	11.2	<	15.4	2.4	<	34.8	6.4	<	26.7
Tamil Nadu	27.1	>	14.5	28.4	>	20.9	12.4	<	28.2	14.6	<	30.9
Tripura	41.7	>	4.8	36.9	>	7.1	15.7	<	15.9	19	>	16
Uttarakhand	28.4	>	7.9	30	>	12.8	16.1	<	17.7	18.4	<	20.5

Source: Compiled by authors using NFHS state level data.
Notes: (1) BMI < 18.5 kg/m^2 indicates the proportion of underweight population.
(2) BMI ≥ 25 kg/m^2 indicates the proportion of overweight or obese population.

In this analysis, the BMI cut-off used for defining overweight and obesity prevalence is 25 kg/m^2, which is based on the WHO International BMI classification. WHO in its report on Asia-Pacific region has redefined obesity by measuring obesity using separate classification or cut-offs for Asia-Pacific region, and acknowledged the differences in the body types (physique, body build, etc.) and genetic composition across the population in the Asia-Pacific region and population in the European countries, Canada and the United States. The new set of cut-offs has been proposed by WHO called Asian BMI classification, which defines overweight and obesity using a cut-off of 23 kg/m^2.

Although a common BMI standard will facilitate an easy comparison across population in different regions or countries but customised and well-designed region-specific cut-offs based on genetic composition and demography will make the understanding of nutritional status and epidemiology more accurate. If Asian BMI classification is used to measure the overweight and obesity prevalence, then its prevalence will be more severe and higher proportion of population will be at risk of overweight and obesity associated adverse health impacts. This difference is captured in Figure 3.5. A difference in the overweight and obesity prevalence is witnessed when measured using the two separate classifications. It can be easily observed that International classification understates the overweight and obesity prevalence by a difference of more than 10% in both the years. In year 2005-06, the proportion of overweight or obese population is 13.9% as per International classification and the same figure is 24.4%, based on Asian classification. For year 2015, these values are 18.1% and 31.3% for each of the respective classification.

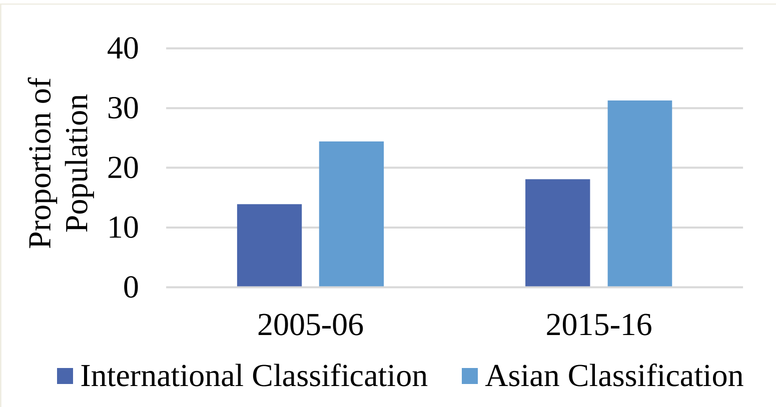


Figure 3.5: Overweight and Obese Population in India based on WHO International and Asian BMI Classification

Source: Figure constructed by authors using NFHS individual level data set for 2005-06 and 2015-16.
Note: WHO International Classification measures overweight and obesity using BMI ≥ 25 kg/m^2 cut-off and Asian Classification applies BMI ≥ 23 kg/m^2 cut-off.

Nutritional Bias across Genders

The data analysis also provides insights about the differences in the malnutrition across genders. We examine malnutrition amongst males and females in 2015-16 across states in India. Overnutrition and undernutrition trends across some selected states for both males and females for year 2015-16 are presented in Table 3.3. We find that both underweight prevalence as well as overweight and obesity prevalence is higher amongst females as compared to males. This implies that a smaller proportion of females lie in the normal weight range. Females have a higher underweight prevalence than males across 24 states and a higher overweight and obesity prevalence across 21 states in year 2015-16. The overnutrition and undernutrition trends bring out an interesting aspect of our socioeconomic scenario. While analysing the nutritional status across different wealth quintiles at national level (all-India average), we find that underweight prevalence is higher among lower (or poorer) wealth quintiles while overweight and obesity is more prevalent among higher (or richer) wealth quintiles (NFHS, 2015-16). Also, across lower wealth quintiles, females have higher underweight prevalence as compared to males whereas across richer wealth quintiles, females have a higher overweight and obesity prevalence than males. This suggests that among households that belong to the lower wealth quintiles allocation of nutrition is biased against females thus females are more likely to be underweight. While higher overweight and obesity for females among higher wealth quintiles could be related to the restrictions on women among the higher wealth quintiles such as restraints on working outside, participating in sports and related physical activities, etc. It could also be related to availability of domestic help in the richer households. Our analysis suggests that women are at a higher risk of both undernutrition and overnutrition in India.

Table 3.3: Across Gender Comparison of Nutritional Trends during 2015-16

States	**Overweight and Obesity Prevalence**			**Underweight Prevalence**		
	Females		**Males**	**Females**		**Males**
Chhattisgarh	11.9	>	10.2	26.7	>	24.1
Madhya Pradesh	13.6	>	10.9	28.4	=	28.4
Uttarakhand	20.4	>	17.7	18.4	>	16.1
Haryana	21	>	20	15.8	>	11.3
Gujarat	23.7	>	19.7	27.2	>	24.7
Tamil Nadu	30.9	>	28.2	14.6	>	12.4
Punjab	31.3	>	27.8	11.7	>	10.9
Kerala	32.4	>	28.5	9.7	>	8.5

Source: Compiled by authors using NFHS state level data.

3.2.3 Overweight, Obesity, Non-Communicable Diseases and Health Outcomes

Following the above discussion in Indian context, one may expect the rise in overweight and obesity prevalence to have an adverse impact on the prevalence of NCDs and health outcomes such as death rate, longevity, etc. thereby worsening the mortality and morbidity conditions. It is observed that the burden associated with NCDs has surged up in India which is evident from the rise in NCDs and mortality attributable to NCDs. WHO as well as NFHS estimates indicate a rise in the prevalence of diabetes, heart disease and hypertension.

The health burden associated with NCDs has been rising not only in the form of increased prevalence but also in the form of rise in the share of deaths caused as a consequence of NCDs. The mortality attributable to NCDs has increased from 47.9% to 51.8% among males and 42.2% to 45.8% among females during 2004-13 (Cause of Death Statistics, India 2004-13). During the same period, mortality caused by communicable diseases has fallen by about 10% among both males and females. This substantial rise in the gap between mortality attributed to NCDs and communicable diseases indicates the shift in health burden more towards NCDs. Thus, the nutritional transition experienced in India can be related to epidemiological transition both in terms of increase in prevalence of NCDs and mortality associated with it.

Economic Burden of Overnutrition

The rise in healthcare expenditure associated with overnutrition reflects the economic burden of overnutrition. This health expenditure can either be borne by government or by private households. In either situation, the rise in total healthcare costs or expenditures are expected to impose an economic burden. The total per capita health expenditure in India increased from Rs.1201 in 2004-05 to Rs.3826 in 2014-15 (National Health Accounts, 2014-15).

Figure 3.6 illustrates the government health expenditure and out-of-pocket expenditure as a proportion of the total health expenditure in India. It can be clearly seen that government's contribution has remained below 30% during the past one decade and showed only marginal increments. Out-of-pocket healthcare expenditure by households constitutes a major proportion of the total health expenditure. Therefore, an increase in the overnutrition prevalence is likely to impose a heavy monetary burden on the households by increasing their healthcare care spending.

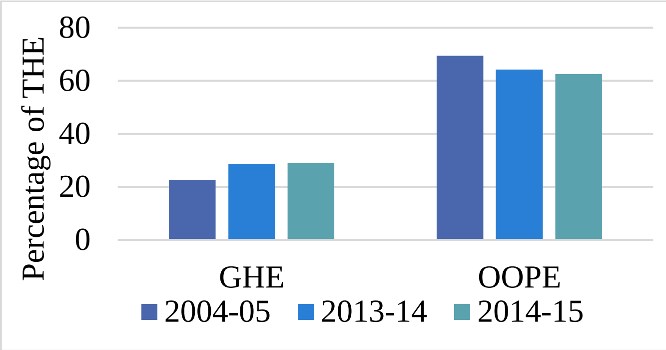


Figure 3.6: Government Health Expenditure and Out-of-Pocket Expenditure as the Proportion of Total Health Expenditure, India

Source: Figure constructed by authors using the data from National Health Accounts, 2014-15.
Note: GHE - Government Health Expenditure, OOPE - Out of Pocket Expenditure and THE - Total Health Expenditure.

3.3 Overnutrition in The United States

The overnutrition problem has been prevalent in the United States for more than three decades. Among high-income countries, the United States continues to be one of the highest obesity countries (OECD, 2017). Table 3.4 presents the overweight and obesity prevalence, and obesity prevalence estimates for the United States during 1980 and 2016 based on WHO data. The proportion of obese population nearly tripled during 1980 and 2016. By year 2016, more than one-third of US adult population was obese (36.2%) and about two-third (67.9%) of adult population was either overweight or obese.

Table 3.4: Overweight and Obesity Prevalence, and Obesity Prevalence for the Total Population (age 18 years or above), The United States

	1980	1990	2000	2010	2016
Overweight and Obesity Prevalence	44	50.9	58.6	64.8	67.9
Obesity Prevalence	13.7	18.7	25.5	32.3	36.2

Source: Compiled by authors using WHO data.

3.3.1 State level Analysis

Analysing the state level estimates of obesity prevalence extracted from Behavioral Risk Factor Surveillance System (BRFSS), we observe a considerably high obesity prevalence in almost every state with a variation of about 6-10% between the states having highest and least obesity prevalence during 2000-14. This variation is quite low as compared to the across state heterogeneity observed in Indian context, as discussed in Section 3.2. In India, some states have high underweight while others have high overweight and obesity, leading to the huge variation in overnutrition prevalence whereas in the United States, underweight prevalence is too low and we observe widespread obesity in every state. Also, underweight prevalence, at national level, was only 1.8% in 2017.

Examining the state level trends in obesity prevalence across genders during 2000-14 using BRFSS data, we find a considerable rise in obese population across all states. However, we find no persistent pattern in the obesity prevalence across genders in the United States.[11] At state level, it is observed that some states have a higher obesity for males while others for females. Also, the trend does not remain same over years even in a specific state. Therefore, we may say that there exists no observable and persistent pattern in obesity prevalence across gender.

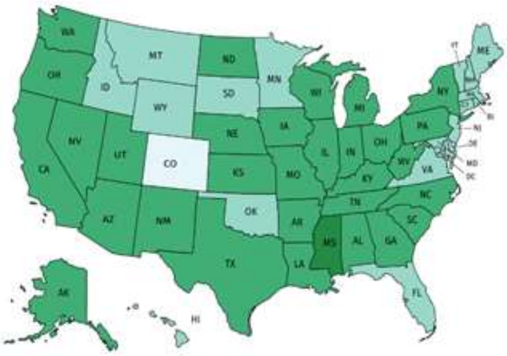
A graphical illustration depicting the rise in obesity prevalence through a map graph has been presented in Figure 3.7. The graph has been plotted using the same online tool as discussed in case of India (in Section 3.2). However, the intervals over which obesity prevalence has been analysed are different. The intervals considered are 0-14.9, 15-19.9, 20-24.9, 25-29.9, and 30 and above percentage of obese population. From Figure 3.7, it can be observed that all states had a below 25% obesity prevalence in year 2000 for both males and females, except Mississippi which had 25.9% obesity prevalence among males. By year 2014, almost every state had an above 25% obesity prevalence for both males and females, except for states like Colorado, Hawaii, Massachusetts, Vermont, Montana and California. This highlights the sharp transition in obesity

[11] To understand the pattern of obesity prevalence across genders, we also examined how obesity rates varied at national level for each year (from 2011 to 2016). We find no clear pattern across genders at national level. For years 2011, 2014 and 2016 males had a higher obesity prevalence while for years 2012, 2013 and 2015 females had a higher obesity prevalence indicating no persistent across gender pattern.

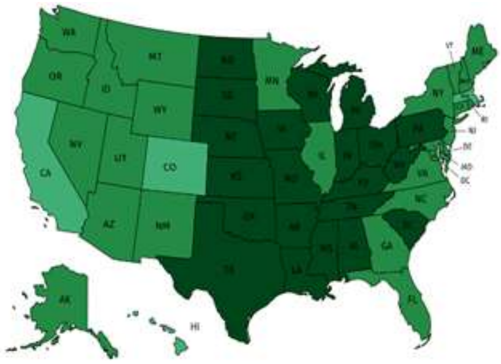
prevalence witnessed in the United States during the period of fifteen years. We find that spread of higher obesity rates has increased tremendously and most states now belong to darker regions. The states which showed the highest obesity prevalence across both male and female population are Indiana, Ohio, Mississippi, Texas, Kentucky, Oklahoma, Louisiana, Alabama, West Virginia and Arkansas.

Proportion of Obese Males

2000

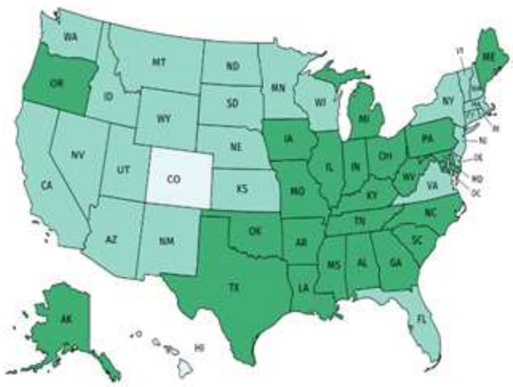


2014



Proportion of Obese Females

2000



2014

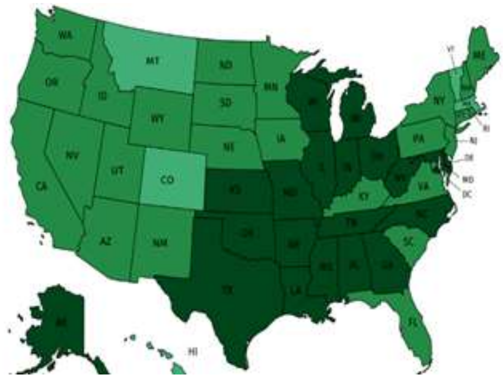
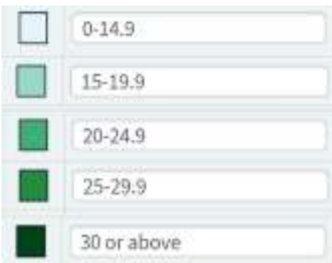


Figure 3.7: Proportion of Obese Population in the United States for the years 2000 and 2014

Source: Figure is constructed by authors using the data from BRFSS.



3.3.2 Obesity, Non-Communicable Diseases and Health Outcomes

As discussed earlier, obesity elevates the risk of morbidity and mortality in the form of rise in NCDs. In the United States, state level prevalence of diabetes and hypertension has increased during 2000 and 2014. Diabetes prevalence increased from 6.02% to 10.17% and hypertension increase from 25.6% to 32.64% (BRFSS). Analysing the deaths by cause estimates, we observe that heart disease and cancer continue to be a major cause of deaths in the United States across both male and female population. In year 2015, the proportion of total deaths caused by heart disease were 24.4% and 22.3% for males and females respectively. While diabetes caused 3.1% and 2.7% of the total deaths among males and females respectively. The same figures for strokes are 4.2% and 6.1% respectively (National Vital Statistics Reports, CDC, 2016). Thus, NCDs have continued to be a major contributor to the deaths in the United States. In year 2000, 88% of the total deaths were caused by NCDs and in year 2016 this figure increased very marginally to 88.3% (World Bank, 2016).

Economic Burden of Obesity

Economic burden associated with high obesity rates can be measured in both monetary and non-monetary terms. It includes healthcare costs, labour market costs, health burden associated with higher NCDs, etc. Here, we focus on healthcare costs. The United States has the highest per capita health expenditure among high-income countries. The per capita current health expenditure doubled during 2000-15 and rose from 4561.9 US Dollars in year 2000 to 9535.9 US Dollars in year 2015.[12] Also, current health expenditure accounts for about 16% of GDP (WHO, 2015). Government's share in total current health expenditure was more than 50% in year 2015 (WHO). These figures suggest that any further increase in the obesity prevalence may aggravate the cost burden associated with obesity.

3.4 Policies So Far

The health and economic consequences of overnutrition are growing worldwide. Also, the rise in overweight and obesity prevalence is expected to have different implications

[12] "Health spending measures the final consumption of healthcare goods and services (i.e. current health expenditure) including personal healthcare (curative care, rehabilitative care, long-term care, ancillary services and medical goods) and collective services (prevention and public health services as well as health administration), but excluding spending on investments" (OECD, 2015).

on health and financial sector of every country based on their development levels, demography of the population, efficiency of healthcare systems, policy system, etc. The persistent rise in overweight and obesity prevalence and its associated health risks justify the need for intervention.

Policies that can potentially bring down the obesity rates include price interventions (taxing unhealthy foods and/or subsidising healthy foods), regulations in the form of ban or restriction on the sale of unhealthy foods, labelling (nutrient and calorie information on products), mandates on products such as minimum nutrient requirement or limiting the content of fat, sugar and salt, creating awareness and providing information through media and other sources about the importance of healthy lifestyle and eating habits, etc.

Many countries have implemented certain policies to fight the obesity problem. First sin tax was implemented in Norway in year 1981. This tax was imposed on sugar or goods containing refined sugar. Many European countries have imposed taxes on goods containing sugar and saturated fat. In 2011, Hungary levied a 4% tax on products containing high sugar and salt. In 2015, Berkeley, California, United States imposed a 10% tax on beverages containing sugar.

India has also recognised the likely adverse impacts associated with overnutrition and has initiated certain steps to control this rising problem. In 2016, Kerala, India imposed a 14.5% fat tax on junk food. A major step taken at national level was the imposition of a 40% tax (sin tax) on aerated and high sugar content drinks under Goods and Services Tax in the year 2017. University Grants Commission, in August, 2018, directed all the Colleges and Universities to ban the sale of junk food in their respective campuses. Delhi government has also suggested for a tax on foods and drinks containing high salt, sugar or saturated fat.

Although many policies have been implemented to reduce obesity problem, however, the effectiveness of these policies may vary across countries or regions and depends on the type of policy implemented, its coverage and responsiveness of the targeted population towards that policy. Some studies have also suggested the need for multipolicy framework to tackle obesity problem (Anderson and Matsa, 2011).

3.5 Conclusion

Overnutrition has been rising persistently during the recent decades. However, there exists substantial heterogeneity in the form as well as extent of inadequate nutrition across countries. Some countries have high overnutrition while some face a major challenge of undernutrition or even dual burden of both undernutrition and overnutrition. Our analysis suggests that India is experiencing a nutritional transition in the form of a shift away from undernutrition towards overnutrition. The available data shows that the gap between the underweight prevalence, and overweight and obesity prevalence, at all India level, has drastically reduced over the past ten years, from 23.6% to about only 2%. If similar trends continue, the prevalence of overweight and obesity would dominate the prevalence of underweight very soon. For the United States, it is found that high obesity prevalence continues to be a major health challenge.

Chapter 4

Examining the Effect of Overweight and Obesity Prevalence on Life Expectancy

4.1 Introduction

Worldwide obesity prevalence has nearly tripled since 1975 (WHO, 2016). Many high-income countries continue to have high obesity prevalence while many low- and middle-income countries have witnessed a rapid emergence of overnutrition and are going through a nutritional transition. The goal of this chapter is to assess the effects of overnutrition on health outcomes. We conduct the analysis for two countries, one, India which is a low- and middle-income country where overnutrition is an emerging issue and, other, the United States which is a high-income country where prevalence of obesity is already high. In the present chapter, we consider life expectancy at birth, denoted by $LEXP$, as the health outcome variable and examine how does an increase in obesity prevalence affects $LEXP$.

The present chapter empirically examines the non-linear relationship between obesity prevalence and $LEXP$. Bansal and Zilberman (2016) develop an analytical model that defines life expectancy of an individual as a function of the distribution of body mass index and aggregates the individual life expectancy function to derive the macro level average life expectancy as a function of obesity prevalence. They empirically test their model at cross-country data. Theirs is the first macro level study that examines the relationship between longevity and obesity prevalence. They find $LEXP$ to be a concave function of obesity prevalence and in countries where obesity prevalence is above 30%, $LEXP$ is decreasing in obesity.

We use their analytical model to estimate the concave relationship between $LEXP$ and obesity prevalence. We do a state level analysis for the United States and India. However, our empirical methodology is stronger in many respects. Given that the countries are quite heterogenous, the analysis at aggregate level may not be sufficient to understand the major processes of the relationship between obesity prevalence and $LEXP$. Hence, it will be useful to look at within country analysis using more disaggregated level data. An important contribution of our study is that we estimate the causal impact of overweight and/or obesity prevalence on the $LEXP$.

Our is a longitudinal study based on state level data. For the United States, the study extracts data from BRFSS for most of the variables used in the analysis. We use state level data across 50 states during the period of 2000-14. For India, the study extracts state level data on health and behavioural factors from the third and fourth rounds of NFHS for the years 2005-06 and 2015-16 respectively. Life expectancy data is taken from life tables published by Sample Registration System. We also consider other data sources (discussed in Sections 4.3 and 4.4).

In this analysis, we include many covariates such as per capita health expenditure, per capita gross domestic product (GDP), educational attainment, tobacco consumption, alcohol consumption and other covariates, however, our main covariate of interest is obesity prevalence. Here, a crucial element that affects $LEXP$ is health expenditure. Due to the existence of obesity problem, individuals may spend more on the medical services to offset the adverse effects of overnutrition, that is, there is a trade-off between keeping weight under control and spending more on healthcare. Health expenditure is a key input in the health (or $LEXP$) production function (Lichtenberg, 2002). Many researchers have advocated the positive effects of health expenditure on health outcomes such as longevity (Lichtenberg, 2002; Lichtenberg, 2011; Brunello et al., 2009; Bhattacharya and Sood, 2011 and Bansal and Zilberman, 2016). However, the marginal effects of a rise in the health expenditure on the health outcomes may differ across countries, genders, age groups, etc. but are expected to be positive. Therefore, we include per capita health expenditure in our empirical model.

To understand the relationship between $LEXP$ and obesity prevalence, we, first, estimate a Fixed Effects model assuming that the unobserved heterogeneity affecting the relationship between obesity prevalence and $LEXP$ is time invariant (we relax this assumption later). However, Fixed Effects estimates may get biased if these unobserved factors are time variant and the unobserved factors are correlated with one or more explanatory variables, that is, the model is likely to suffer from endogeneity problem. We address the potential endogeneity problem by estimating a Generalized Method of

Moments (GMM) model for the analysis in the United States and for India, we estimate an Instrument Variable - Two Stage Least Squares (IV-2SLS) model and establish a causal relationship between overnutrition and $LEXP$.[13] The findings from this study are expected to have implications in structuring the health policies. Based on our results, we also identify the states which are at a higher risk of longevity decline thereby guiding policies that aim at optimising the nutritional intake across regions.

The rest of the chapter is organised as follows. Section 4.2 discusses the conceptual framework and empirical methodology. Sections 4.3 and 4.4 present the data and the estimation results for the United States and India respectively. Section 4.5 concludes the chapter and highlights the important results.

4.2 Conceptual Framework and Methodology

In this section, we present the conceptual framework used to study the relationship between overnutrition and $LEXP$. To briefly state, we hypothesise a concave relationship between obesity prevalence and $LEXP$. We develop our conceptual framework based on the analytical model presented by Bansal and Zilberman (2016). Their study provides a systematic framework that aggregates the micro level relationship between the nutritional status and life expectancy of an individual, and obtain the macro level relationship between the average $LEXP$ and obesity prevalence. We, first discuss the link between BMI and life expectancy at individual level and then aggregate this relationship across all individuals to derive the macro relationship.

At micro level, the life expectancy of an individual is expected to vary with BMI. Say, for an individual, the BMI is initially below 18.5 kg/m^2. A rise in BMI of this individual will offset the adverse effects of undernutrition thereby improving his nutritional status, hence, it is expected to increase his life expectancy and we may continue to have this effect as BMI increases further. However, for high or very high values of BMI, when a person is characterised as being overweight or obese, the health risks associated with overnutrition increases. This may have a detrimental effect on the health status, therefore, life expectancy of the individual is expected to fall. Hence, it can be stated

[13] Due to data limitations, application of GMM estimation is not feasible for the analysis in Indian context.

that life expectancy initially increases with BMI and then later falls for higher values of BMI.

We reproduce the conceptual framework of Bansal and Zilberman (2016). Life expectancy at birth of an individual, L, can be defined as a function of BMI (m), health expenditure (HE), gender (S), and other variables (V):

$$L = f(m, HE, S, V) \tag{4.1}$$

Let $h(m)$ denote the density function, $m_1 \leq m \leq m_2$ and $\int_{m_1}^{m_2} h(m)dm = 1$. It is assumed that life expectancy of an individual is a unimodal function of BMI, and reaches its peak at BMI value $m = \bar{m}$, under ceteris paribus. Thus, it can be stated that life expectancy of an individual is increasing in BMI for $m_1 \leq m \leq \bar{m}$ and decreasing for $\bar{m} \leq m \leq m_2$.

Now, the average life expectancy function at macro level, $LEXP$, is determined by the individual life expectancy function and BMI distribution of the population. $LEXP$ at macro level is the sum of life expectancy at each BMI (m) times the proportion of population having a BMI value m. In the macro level function, individual's life expectancy of micro function is defined as average life expectancy while individual's BMI is defined as the proportion of obese population, denoted by OBS. Linking the micro concept to aggregate level, we expect the average life expectancy of the population to increase in the obesity prevalence at lower values of obesity prevalence and eventually fall in obesity prevalence after reaching a threshold value, OBS^*. That is, at macro level, life expectancy at birth, $LEXP$, is:

1) Increasing for $OBS \leq OBS^*$, i.e., $\frac{\partial LEXP}{\partial OBS} \geq 0$.
2) Declining for $OBS \geq OBS^*$, i.e., $\frac{\partial LEXP}{\partial OBS} \leq 0$.

One may expect a concave relationship between the $LEXP$ and obesity prevalence as illustrated in Figure 4.1. In Figure 4.1, the threshold obesity prevalence beyond which $LEXP$ falls is given by OBS^* at which the $LEXP$ is at its peak, $LEXP^*$. Any rise in the obesity prevalence when $OBS < OBS^*$ will lead to an increase in the $LEXP$ whereas a rise in the obesity prevalence when $OBS > OBS^*$ will lead to a decline in the $LEXP$. Also, at aggregate level, initially, when a significant proportion of the population is underweight and obesity prevalence starts rising, then an increase in obesity prevalence is expected to increase $LEXP$ as the proportion of underweight population is likely to fall, that is, we will have relatively less proportion of people who are underweight. However, any further rise in the obesity prevalence, when the proportion of underweight population is at low levels, is likely to reduce $LEXP$ as a substantial proportion of the population has now become obese and are facing health risks associated with overnutrition.

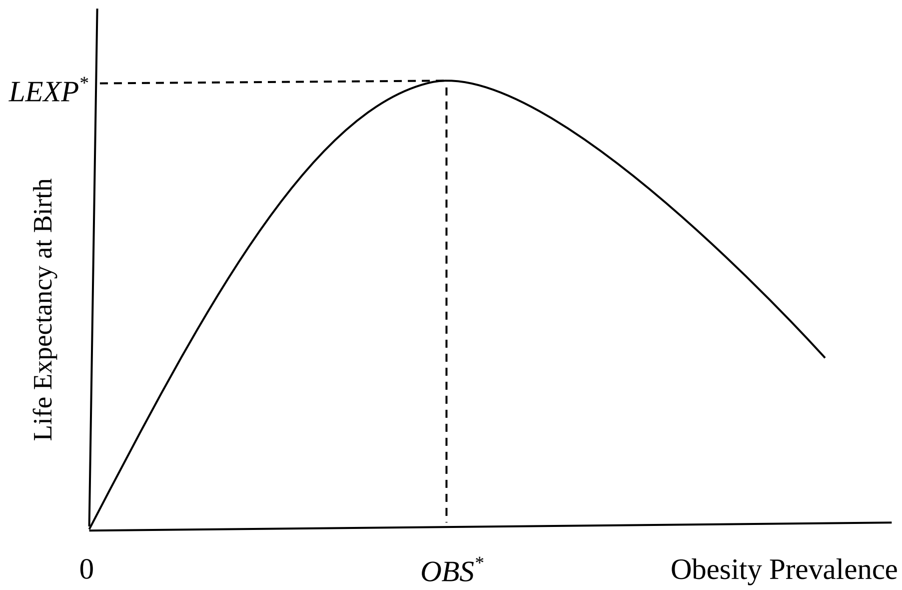


Figure 4.1: Life Expectancy at Birth and Obesity Prevalence

Source: Figure constructed by authors.

The threshold beyond which $LEXP$ falls is expected to vary across countries and is determined by their respective levels of obesity prevalence, socio-economic and demographic factors. Here, we discuss how the risk of longevity decline varies across countries having low vis-à-vis high levels of obesity prevalence. For this, we plot the BMI distribution of the population across these two types of countries, in Figure 4.2. In Figure 4.2, $h_L(m)$ is the BMI distribution of the country having low obesity prevalence and $h_H(m)$ is the BMI distribution of the country having high obesity prevalence. We expect BMI distributions across different countries to have similar shapes, however, the BMI densities may vary, in the sense that, density for high (low) BMI values is expected to be higher for high (low) obesity countries. In Figure 4.2, the proportion of population having high levels of BMI, say m_0 or above, is much higher for high obesity country (represented by area $b\ m_0\ m_2$) as compared to the low obesity country (represented by area $a\ m_0\ m_2$). Similarly, the proportion of population having low levels of BMI is higher for low obesity country as compared to the high obesity country. Relating the proportion of obese (or high BMI levels) population across these two countries to the non-linear concave curve presented in Figure 4.1, we state that a higher proportion of population is at the risk of longevity decline in high obesity country as compared to the low obesity country. A low obesity country is likely to lie to the left of OBS^* in Figure 4.1. Hence, in a low obesity country, we expect $LEXP$ to improve with an increase in obesity prevalence given low levels of obesity. In a high obesity country, the obesity prevalence is likely to lie to the right of OBS^* and $LEXP$ is expected to decline with an increase in obesity prevalence.

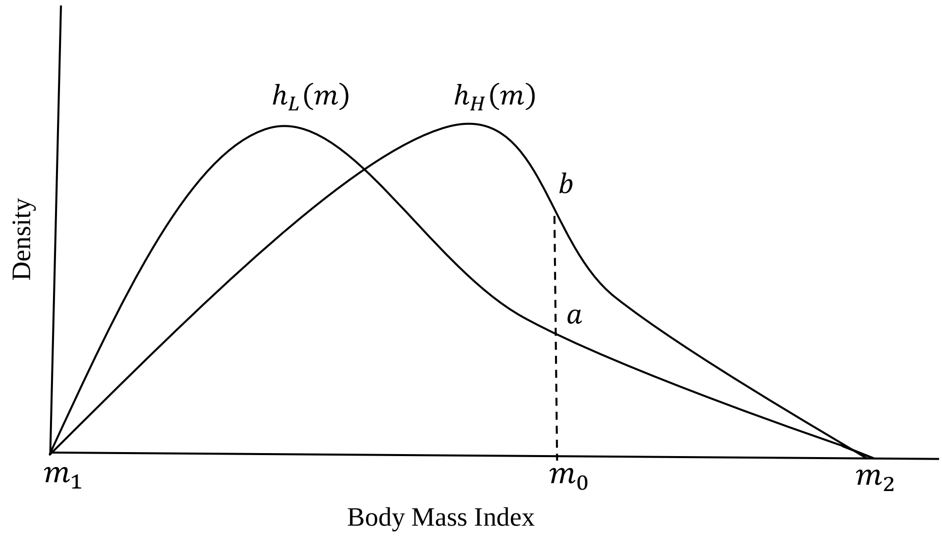


Figure 4.2: BMI Distribution of the Population in a Low and a High Obesity Country

Source: Figure constructed by authors.

Empirical Model

Now, we describe the empirical model estimated in this study based on the above discussed conceptual framework. Here, we estimate the effect of obesity prevalence on $LEXP$ and the effectiveness of per capita health expenditure in countering the adverse effects of obesity prevalence on longevity while controlling for per capita GDP, educational attainment, tobacco consumption, alcohol consumption and other covariates where obesity prevalence takes a quadratic form. Equation (4.2) presents the regression equation estimated in this study:

$$LEXP_{it} = \phi_0 + \phi_1 OBS_{it} + \phi_2 OBS_{it}^2 + \phi_3 HE_{it} + \phi_4 X_{it} + \gamma_i + \tau_t + \varepsilon_{it} \quad (4.2)$$

where, i represents state, $i = 1, 2, \dots, n$, and t represents time period in years, $t = 1, 2, \dots, T$;

$LEXP_{it}$ is life expectancy at birth for i^{th} state in year t;

OBS_{it} is obesity prevalence or proportion of population having BMI ≥ 30 kg/m^2 for i^{th} state in year t;

HE_{it} is per capita health expenditure for i^{th} state in year t;

X_{it} is vector of controls for i^{th} state in year t;

γ_i are state fixed effects;

τ_t are time fixed effects;

ε_{it} is the error term.

Here, X_{it} includes variables such as educational attainment, tobacco consumption, alcohol consumption, per capita GDP, interaction terms of obesity prevalence and its square with gender dummy, and other controls.

In equation (4.2), following the conceptual framework discussed above, we expect $\phi_1 > 0$ and $\phi_2 < 0$, that is, $LEXP$ increases at a decreasing rate with an increase in the obesity prevalence.

To derive the threshold for obesity prevalence beyond which $LEXP$ is expected to fall, we differentiate the equation (4.2) with respect to OBS:

$$\frac{\partial (LEXP)_{it}}{\partial (OBS)_{it}} = \phi_1 + 2\phi_2 OBS_{it} \quad (4.3)$$

The optimum (or threshold) for obesity prevalence is defined as:

$$OBS_{Thrsd} = -\frac{\phi_1}{2\phi_2} \quad (4.4)$$

Above threshold, OBS_{Thrsd}, is the mathematical representation of the threshold, OBS^*, discussed in Figure 4.1.

With an objective to see how the effects of a rise in the obesity prevalence on $LEXP$ varies across genders, we also test for existence of any gender differential effect.

We test the following hypotheses for both the United States and India:

Hypothesis 1: Life expectancy at birth is a concave function of the obesity prevalence, that is, life expectancy at birth initially increases in the obesity prevalence and then declines after reaching a certain threshold, at macro level.

Hypothesis 2: Per capita health expenditure counters the adverse impact of obesity prevalence on health and improves the life expectancy at birth.

Hypothesis 3: The effects of a rise in the obesity prevalence on life expectancy at birth differs across men and women, that is, there exists a gender differential effect.

Empirical Methodology

To test our hypotheses, we empirically estimate regression equation (4.2). We, first, estimate the relationship between obesity prevalence and $LEXP$ using a Fixed Effects model and then estimate a System Generalized Method of Moments (System GMM) model and an IV-2SLS model for the United States and India respectively.

In model the presented in equation (4.2), the obesity prevalence is likely to be correlated with the unobserved factors that determine $LEXP$. These unobserved factors may vary over time. This may lead to the potential endogeneity problem in the form of Omitted Variable Bias (OVB). In our model, obesity prevalence and its square could be potentially endogenous. In presence of unobserved heterogeneity, Ordinary Least Squares estimates may be biased and inconsistent as the explanatory variables in the model are correlated with the error term. To obtain consistent estimates of ϕ_i, we estimate the following models:

1. Fixed Effects Model

Fixed Effects estimates are used to eliminate the potential bias caused by unobserved heterogeneity. The time-invariant state-specific unobserved heterogeneity gets differenced out in the Fixed Effects estimation. Under the assumption that omitted variable bias or unobserved heterogeneity is time-invariant, the Fixed Effects estimates will be consistent (Wooldridge, 2001). We also estimate our model using estimation methods where the assumption of time-invariant unobserved heterogeneity is not required. These estimation methods are described next.

2. Two-Stage Least Squares (2SLS) Model

With an objective, to resolve the potential endogeneity problem arising from OVB, we estimate a Two-Stage Least Squares (2SLS) model. We estimate a Fixed Effects - IV-2SLS model. IV-2SLS can be applied to the models having one or more explanatory variables that are correlated with the error term provided we are able to find at least as many instruments as the number of endogenous variables in the model. In the model given in equation (4.2), we have two endogenous variables, OBS and OBS^2, implying that we need at least two instruments. The instrument must satisfy two conditions, first, it must be correlated with the endogenous explanatory variable, that is, it must be powerful, and, second, it must not be correlated with the error term, that is, exclusion restriction must hold. We estimate the Fixed Effects - IV-2SLS model for India. The instruments used are described in Section 4.4.2.

3. Generalized Method of Moments (GMM)

GMM can be applied to the models having unobserved heterogeneity and containing the lagged dependent variable as a covariate (Wooldridge, 2001 and Roodman, 2009). Existence of the lagged dependent variable gives dynamic nature to the model and therefore it is called as a dynamic panel model. GMM has two variants, Difference GMM and system GMM.

The difference GMM (Arellano and Bond, 1991) estimation applies first-differencing to the equation and uses the lagged levels of the endogenous variables as instruments. The Arellano–Bover/Blundell–Bond (Arellano and Bover, 1995 and Blundell and Bond, 1998) estimator augments the Arellano–Bond model by adding the original equation in levels to the system and uses the lagged first differences of the variables as instruments. This system of two equations, the original equation and the transformed one, is called system GMM.

The dynamic nature of life expectancy function has been recognised by Brunello et al. (2009). They modelled the change in life expectancy at time t as a function of life expectancy in period $t-1$. Lichtenberg (2002) has also modelled life expectancy as a function of its own lagged value. Following this, we now rewrite equation (4.2) as the following dynamic panel model and use a system GMM model to estimate it:

$$LEXP_{it} = \psi_0 + \psi_1 LEXP_{i,\ t-1} + \psi_2 OBS_{it} + \psi_3 OBS_{it}^2 + \psi_4 HE_{it} + \psi_5 X_{it} + \tau_t + \eta_{it} \tag{4.5}$$

where, $LEXP_{i,\ t-1}$ is the life expectancy at birth for i^{th} state in year $t-1$ and $\eta_{it} = \mu_i + \nu_{it}$, which includes both fixed effects, μ_i, and idiosyncratic (time-varying) error component, ν_{it}. Other variables have similar interpretation as discussed earlier.

4.3 Examining the Relationship between Obesity Prevalence and Life Expectancy at Birth in the United States

$LEXP$ increased by 2.1 years in the United States during 2000-14 (WHO). After experiencing a persistent rise during 2000-14, with only a dip of 0.1 year in 2005 and 2013, $LEXP$ declined for two consecutive years from 79 years in 2014 to 78.6 and 78.5 years in 2015 and 2016 respectively. During the same period, the obesity prevalence increased from 25.5% in year 2000 to 34.9% in year 2014.

In this section, we examine the effect of a rise in overnutrition on $LEXP$. Overnutrition is measured in terms of obesity prevalence where WHO International BMI classification is used to define obesity. Here, we present data and estimation results for the United States.

4.3.1 Data

Data Sources and Definitions

Data on nutritional status, demographic characteristics and behavioural risk factors is obtained from Behavioral Risk Factor Surveillance System (BRFSS) established by Centers for Disease Control and Prevention (CDC) in year 1984. It is the largest survey that collects data on health-related risk behaviours, chronic health conditions and use of preventive services through telephone surveys in the United States. Data on remaining variables is collected from other sources, namely, Global Health Data Exchange, Institute for Health Metrics and Evaluation, and Centers for Medicare and Medicaid Services (CMS).

Longitudinal state level data is considered for 50 states over a fifteen-year period from 2000-14. Data includes observations for men as well as women. Our data set contains 1500 observations on each variable.[14] The list of states included in the analysis is provided in Table A.4.1 of Appendix.

List of Variables with Definition:

(i) **Life Expectancy at Birth:** Average number of years an individual is expected to live at the time of birth (in years).

(ii) **Obesity Prevalence:** Percentage of adult population having a body mass index value of 30 kg/m^2 or above (in percentage).

(iii) **Per Capita Total Health Expenditure:** Total health expenditure divided by state population. The estimates are chained at 2012 prices using GDP deflator (in US Dollars).[15]

(iv) **Educational Attainment – High School:** Percentage of adult population having completed post high school or General Education Development (GED) as the highest grade or years of schooling (in percentage).

(v) **Tobacco Consumption or Smoking:** Percentage of adult population who are current smokers (in percentage).

(vi) **Alcohol Consumption:** Percentage of adult population who have consumed at least one drink of any alcohol beverage within past 30 days (in percentage).

(vii) **Per Capita Real GDP:** The monetary value of all goods and services produced within the boundaries of the state during a given period of time divided by state population. The per capita real GDP is chained at 2012 prices (in US Dollars).

[14] For each state, we have 15 annual observations (during 2000-2014) for males and 15 annual observations for females. That is, we have 30 observations for every state in each year. Total number of states considered is 50. Therefore, we have 30 * 50 = 1500 observations for each variable.

[15] Components included in health expenditure are "personal healthcare, hospital care, physician & clinical services, other professional services, dental services, home health care, prescription drugs and other non-durable medical products, durable medical products, nursing home care, and other health, residential, and personal care. The estimates do not include expenditure on or cost of insurance programs, research, and other construction costs" (Health Accounts by State of residence, CMS).

(viii) **Income $50,000 or Above:** Percentage of adult population having annual household income of 50,000 US Dollars or above (in percentage).

(ix) **Activity Status or Exercise:** Percentage of adult population who participated in any physical activity (moderate or vigorous) during the past month (in percentage).[16]

(x) **Age 65 years or Above:** Percentage of adult population having age 65 years or above (in percentage).

(xi) **Cholesterol Checked:** Percentage of adult population who have had their blood cholesterol checked within the last five years (in percentage).

For the variables such as life expectancy at birth, obesity prevalence, educational attainment, tobacco consumption or smoking, alcohol consumption, income $50,000 or above, activity status or exercise, age 65 years or above and cholesterol checked, data is segregated at gender level and for the variables such as per capita total health expenditure and per capita real GDP, complete state level aggregates have been used. A complete list of data sources is provided in Table A.4.2 of Appendix.

Descriptive Statistics

Table 4.1 presents the descriptive statistics for above listed variables. Mean $LEXP$ is 77.89 years. The minimum $LEXP$ is 70.55 years for males in Mississippi for year 2000 and the highest $LEXP$ is 84.01 years for females in Hawaii for year 2010. Mean obesity prevalence is 25.36%. The highest obesity prevalence of 37.9% is for females in Mississippi in 2014 while the least value of 13.9% is for males in Colorado in 2000. We also compared the averages across genders and found that females have higher $LEXP$ and lower obesity prevalence as compared to males, and the differences are statistically significant.

[16] The variable includes physical activities such as "exercises to strengthen muscles, aerobics, walking, etc. Also, physical activity is measured by 30+ minutes of moderate physical activity five or more days per week, or vigorous physical activity for 20+ minutes three or more days per week", as per BRFSS definition.

Table 4.1: Descriptive Statistics, The United States

Variables	Observations	Mean	Standard Deviation	Minimum Value	Maximum Value
Life Expectancy at Birth (in years)	1500	77.89	2.96	70.55	84.01
Obesity Prevalence (in %)	1498	25.36	4.35	13.90	37.90
Per Capita Health Expenditure (in US $)	1500	13897.23	2371.66	7977.89	21197.43
Educational Attainment - High School (in %)	1498	30.69	4.21	19.60	42.90
Smoking (in %)	1498	20.69	4.31	7.70	34.80
Alcohol Consumption (in %)	1398	53.37	11.81	19.30	77.00
Per Capita Real GDP (in US $)	1500	48551.54	9387.05	30564.00	79894.00
Income $50,000 or Above (in %)	1498	43.37	9.36	18.00	71.00
Exercise (in %)	1498	75.77	4.84	57.70	86.40
Age 65 or Above (in %)	1498	17.45	3.10	7.20	26.40
Cholesterol Checked (in %)	700	74.82	5.02	61.30	87.00
Gender	1500	0.50	0.50	0	1
Year	1500	2007	4.32	2000	2014

4.3.2 Estimation Results and Interpretation

Fixed Effects Estimates

Table 4.2 presents the Fixed Effects estimation results. Three model specifications have been estimated which differ in terms of covariates used. All the models include time trend and standard errors are clustered at the state level. In all the models, the coefficient of obesity prevalence is positive and statistically significant, and the coefficient of square of obesity prevalence is negative and highly statistically significant. $LEXP$ is increasing in obesity prevalence at a decreasing rate indicating that the relationship between longevity and obesity prevalence is concave.

We focus on model specification (3) of Table 4.2. The model contains variables such as obesity prevalence, obesity prevalence square, per capita health expenditure, educational attainment (high school), smoking and alcohol consumption. To control for the wealth or income of the population, we also include a variable showing the percentage of adult population having annual household income of 50,000 US Dollars or above.

For model (3), the change in $LEXP$ with an increase in obesity prevalence is given by $\frac{\partial LEXP}{\partial OBS} = 0.065 - 0.002 * 2\ (OBS)$.[17] The threshold beyond which $LEXP$ falls in obesity prevalence is 20.2%, = $\left[\frac{-0.0647857}{2 * (-0.0016044\ )}\right]$, calculated using equation (4.4).[18] $LEXP$ is increasing in obesity prevalence until obesity reaches 20.2% and then declines for higher values of obesity prevalence. As per this model, at mean obesity for year 2014, $LEXP$ is decreasing in obesity prevalence.[19]

Per capita health expenditure affects $LEXP$ positively and is highly statistically significant. We find that a $1000 increase in per capita health expenditure improves average $LEXP$ by about 0.07 years (0.8 months).

The sign of other covariates is as expected. Educational attainment has a positive and highly statistically significant effect. A unit rise in the percentage of population having

---

[17] $\frac{\partial LEXP}{\partial OBS} = \ \phi_1 + \ 2\phi_2\ OBS$, as per equation (4.3).

[18] Complete values of the estimates are used to compute threshold. The thresholds for models (1) and (2) are 20.83% and 21.25% respectively.

[19] $\frac{\partial LEXP}{\partial OBS}\Big|_{OBS=29.4} \ \leq\ 0$, where 29.4% is the mean obesity prevalence for the year 2014.

highest degree as high school or GED improves the average $LEXP$ by 0.03 years. Behavioural risk factors such as smoking and alcohol consumption have a highly statistically significant adverse effect on longevity. A unit increase in the proportion of population drinking alcohol reduces the average $LEXP$ by 0.01 years (0.12 months). Similar observation can be made for smoking as well. Wealth or income of the population has a positive effect on average $LEXP$.

Sensitivity Analysis – Robustness Check

To check for robustness of our results, we include some potential confounding covariates in our model such as per capita real GDP, exercise - percentage of adult population who participated in any physical activity during the past month, and percentage of population having age 65 years or above. To proxy for health awareness and preferences towards health monitoring, we use a variable on the percentage of adult population who have had their blood cholesterol checked within the last five years. Table 4.3 reports these results. The coefficients of obesity prevalence, and its square continue to have expected signs and are statistically significant. Also, the threshold obesity prevalence beyond which $LEXP$ falls is close to our earlier estimate, 21%.[20] Per capita real GDP is negative and weakly statistically significant. Since per capita health expenditure and per capita real GDP are expected to be correlated, this could be the reason for negative coefficient. Percentage of population having done a cholesterol check has a positive and a highly statistically significant effect on $LEXP$. Other additional covariates do not have a statistically significant effect.

Heterogeneity Analysis – Differential Effect across Genders

With an objective to check for the existence of any differential effect of obesity prevalence on longevity across genders, we introduce interaction of obesity prevalence and its square with female dummy.[21] Table 4.4 reports these results. The coefficients on obesity prevalence and its square, and both the interaction terms are highly statistically significant. This indicates existence of a highly statistically significant gender differential effect of obesity prevalence on longevity across males and females. We find that both positive and negative effects of obesity prevalence on $LEXP$, as

[20]The threshold for models (1) to (4) are 20.47%, 21.24%, 21.24% and 21% respectively.
[21]Female dummy takes value 1 if gender is female and 0 if male.

shown by the coefficients of obesity prevalence, its square and the two interaction terms, are higher for males as compared to females.

Table 4.2: Effect of Obesity Prevalence on Life Expectancy at Birth, Fixed Effects Estimates

Variables	**(1)**	**(2)**	**(3)**
Obesity Prevalence (in %)	0.073** (0.030)	0.071** (0.028)	0.065** (0.029)
Obesity Prevalence Square (in %)	-0.002*** (0.001)	-0.002*** (0.001)	-0.002*** (0.001)
Per Capita Health Expenditure (in US $)	0.00007*** (0.000)	0.00007*** (0.000)	0.00007*** (0.000)
High School (in %)		0.028*** (0.004)	0.031*** (0.004)
Smoking (in %)		-0.028*** (0.004)	-0.025*** (0.005)
Alcohol Consumption (in %)		-0.011*** (0.004)	-0.011*** (0.003)
Income $50,000 or Above (in %)			0.004* (0.002)
Time Trend	0.142*** (0.008)	0.139*** (0.007)	0.141*** (0.007)
Constant	75.194*** (0.333)	75.503*** (0.499)	75.331*** (0.470)
Observations	1,498	1,398	1,398
R^2	0.917	0.929	0.930
F Statistic	480.04	362.82	300.99
Prob > F	0.0000	0.0000	0.0000

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses. Standard errors are clustered at the state level. Within R square is reported.

Table 4.3: Effect of Obesity Prevalence on Life Expectancy at Birth, Robustness Check - Fixed Effects Estimates

Variables	(1)	(2)	(3)	(4)
Obesity Prevalence (in %)	0.073** (0.029)	0.071** (0.028)	0.071** (0.029)	0.065** (0.030)
Obesity Prevalence Square (in %)	-0.002*** (0.001)	-0.002*** (0.001)	-0.002*** (0.001)	-0.002*** (0.001)
Per Capita Health Expenditure (in US $)	0. 00008*** (0.000)	0.00007*** (0.000)	0.00007*** (0.000)	0.00007*** (0.000)
High School (in %)	0.027*** (0.004)	0.028*** (0.004)	0.028*** (0.004)	0.035*** (0.006)
Smoking (in %)	-0.027*** (0.004)	-0.029*** (0.004)	-0.028*** (0.004)	-0.013** (0.005)
Alcohol Consumption (in %)	-0.012*** (0.004)	-0.011*** (0.004)	-0.011*** (0.003)	-0.014*** (0.004)
Per Capita Real GDP (in US $)	-7.00e-06* (0.000)			
Exercise (in %)		-0.002 (0.004)		
Age 65 years or above (in %)			-0.0002 (0.014)	
Cholesterol Checked (in %)				0.014*** (0.005)
Time Trend	0.138*** (0.006)	0.138*** (0.007)	0.139*** (0.008)	0.149*** (0.006)
Constant	75.720*** (0.510)	75.671*** (0.534)	75.508*** (0.607)	74.213*** (0.726)
Observations	1,398	1,398	1,398	700
R^2	0.930	0.929	0.929	0.943
F Statistic	329.41	344.89	318.97	316.78
Prob > F	0.0000	0.0000	0.0000	0.0000

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses. Standard errors are clustered at the state level. Within R square is reported.

Table 4.4: Effect Obesity Prevalence on Life Expectancy at Birth, Heterogeneity Analysis by Gender - Fixed Effects Estimates

Variables	**(1)**	**(2)**	**(3)**
Obesity Prevalence (in %)	0.194***	0.208***	0.208***
	(0.036)	(0.036)	(0.036)
Obesity Prevalence Square (in %)	-0.003***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)
Obesity Prevalence * Female Dummy (in %)	-0.147***	-0.160***	-0.154***
	(0.027)	(0.030)	(0.030)
Obesity Prevalence Square * Female Dummy (in %)	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)
Per Capita Health Expenditure (in US $)	0.00005***	0.00005***	0.00007***
	(0.000)	(0.000)	(0.000)
High School (in %)		0.012***	0.010**
		(0.004)	(0.004)
Smoking (in %)		-0.024***	-0.022***
		(0.004)	(0.005)
Alcohol Consumption (in %)		-0.006	-0.006*
		(0.004)	(0.004)
Per Capita Real GDP (in US $)			-8.77e-06**
			(0.000)
Time Trend	0.149***	0.143***	0.142***
	(0.008)	(0.007)	(0.007)
Constant	74.795***	75.077***	75.354***
	(0.290)	(0.468)	(0.455)
Observations	1,498	1,398	1,398
R^2	0.939	0.944	0.945
F Statistic	541.88	528.98	466.00
Prob > F	0.0000	0.0000	0.0000

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses. Standard errors are clustered at the state level. Within R square is reported.
Note: Female dummy takes value 1 for females and 0 for males.

Generalized Method of Moments Estimates

Table 4.5 presents estimates obtained from the system GMM model. Based on GMM estimates, we find stronger effects of obesity prevalence on the $LEXP$ as compared to the Fixed Effects estimates, that is, coefficients of both obesity prevalence and its square have a higher value in the present model.

In model (1), the coefficients of obesity prevalence and its square have expected sings and are highly statistically significant indicating that the relationship between obesity prevalence and longevity is concave. The threshold obesity prevalence is found to be considerably higher than the threshold given by the corresponding Fixed Effects model. The threshold is 25.7%.[22] This threshold is about 5% higher than the threshold given by the fixed effects model. As per this model, at mean obesity for year 2014, $LEXP$ is decreasing in obesity prevalence.[23]

One period lagged value of $LEXP$ is highly statistically significant. This result is suggestive of the dynamic nature of average life expectancy function. Per capita health expenditure has a positive and highly statistically significant effect on longevity. A \$1000 increase in per capita health expenditure can improve $LEXP$ by about 0.24 months. This effect is lower than what we found in the Fixed Effects model. Coefficient on high school does not have an expected sign. Smoking and alcohol consumption have a statistically significant adverse effect on longevity.

For both the models, the null hypothesis that error term is not serially correlated at first order (AR1) is rejected at 1% significance level while the null hypothesis that error term is not serially correlated at second order (AR2) is not rejected at 5% significance level. Both Sargan and Hansen test for overidentifying restrictions having null hypothesis that all instruments are valid and uncorrelated with the error term are not rejected at 10% significance level.

Next, we consider all the variables in logarithmic form. This allows us to estimate the elasticity of life expectancy with respect to obesity prevalence. Table 4.6 presents these results. The concave relationship between obesity prevalence and longevity continues

[22] The threshold is computed using equation (4.4). We consider complete value of estimates for computing the thresholds.

[23] $\frac{\partial LEXP}{\partial OBS}\Big|_{OBS=29.4} \leq 0$, where 29.4% is the mean obesity prevalence for the year 2014.

to hold and is highly statistically significant. The thresholds value for obesity prevalence beyond which $LEXP$ falls are 22.38% and 22.27% for model (1) and (2) respectively. One period lagged value of $LEXP$ is found to be highly statistically significant. Smoking and alcohol consumption have a statistically significant negative effect on $LEXP$.

We also compute elasticity of $LEXP$ with respect to obesity prevalence.[24] We find that 1% increase in obesity prevalence reduces $LEXP$ by 0.013% and 0.014% in models (1) and (2) respectively. Based on elasticity values, it can be said that the relationship between obesity and longevity is inelastic. This result is similar to the estimates given by Brunello et al. (2009). They find that elasticity between obesity prevalence and $LEXP$ is -0.008 among high-income countries.

For both the models, the null hypothesis that error term is not serially correlated at first order (AR1) is rejected at 1% significance level while the null hypothesis that error term is not serially correlated at second order (AR2) is not rejected at 10% significance level. Both Sargan and Hansen test for overidentifying restrictions are not rejected at 10% significance level.

The graphical illustration of the relationship between $LEXP$ and obesity prevalence is presented in Figure 4.3. We have plotted the predicted values of $LEXP$ against the obesity prevalence for model (1) in Table 4.5, while keeping all other variables included in the model at their mean values. We reproduce the same graph in Figure 4.4 for selected years – 2000, 2005, 2010 and 2014, in separate graphs to facilitate better analysis. In each graph, the threshold (25.7%) is labelled and represented by the red coloured vertical reference line. It can be seen that $LEXP$ has a concave relationship with the obesity prevalence. By year 2014, $LEXP$ is declining in the obesity prevalence for many states.

[24] For computing elasticity of $LEXP$ with respect to obesity prevalence, we differentiated the regression equation (in logarthmic form) with respect to OBS at time t and obtained elasticity, $\eta = \frac{\partial LEXP}{\partial OBS} \frac{OBS}{LEXP} = \phi_1 + 2\phi_2 \ln(OBS)$. We substituted coefficient estimates and mean obesity prevalence in this expression to obtain elasticity value.

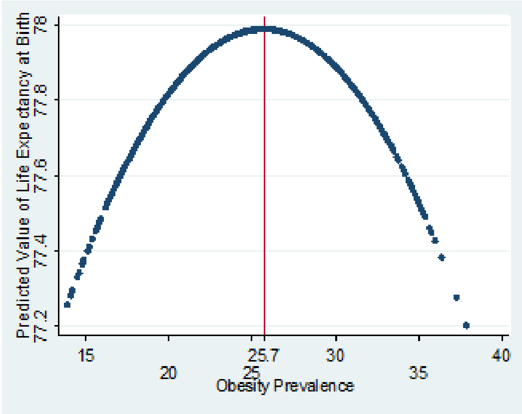


Figure 4.3: Predicted Values of Life Expectancy at Birth, The United States

Source: Author's calculations based on sample data.

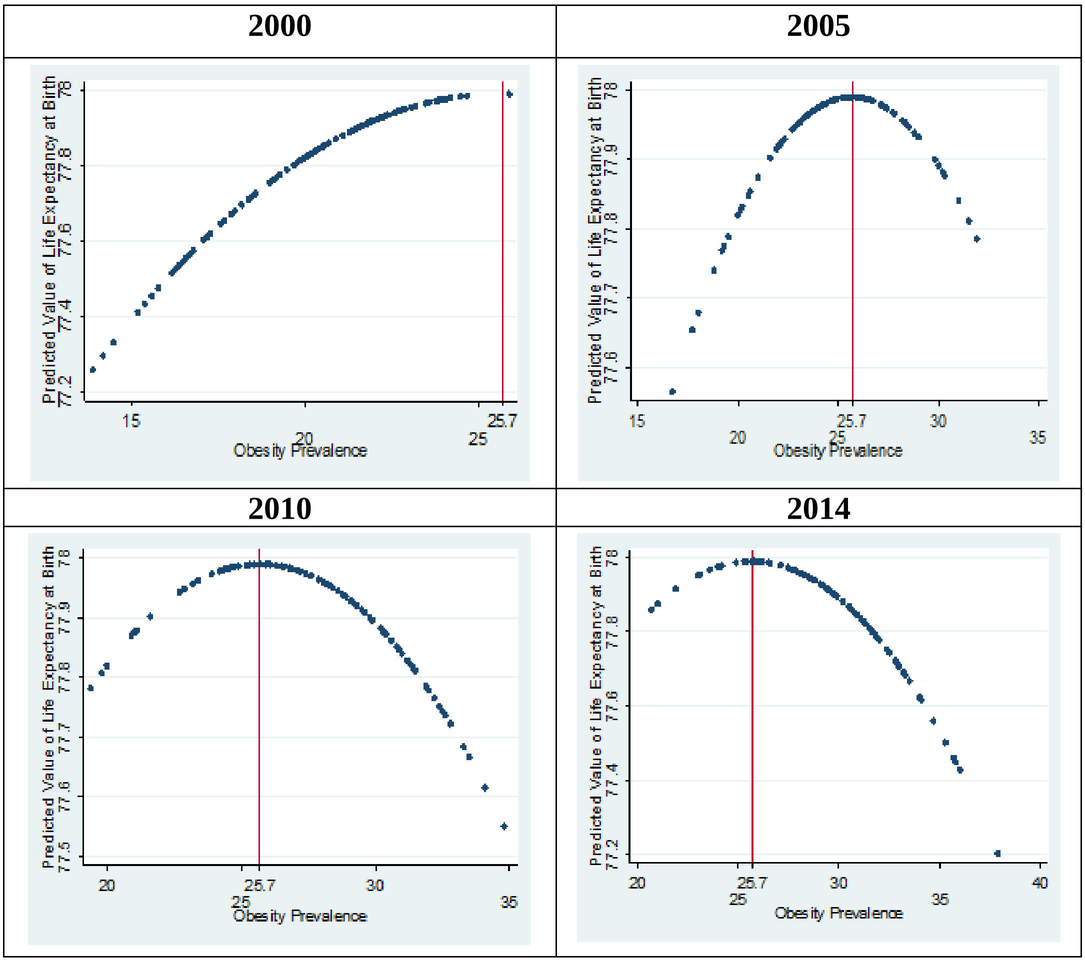


Figure 4.4: Predicted Values of Life Expectancy at Birth for Selected Years, The United States

Source: Author's calculations based on sample data.

Table 4.5: Effect of Obesity Prevalence on Life Expectancy at Birth, Generalized Method of Moments Estimates

Variables	(1)	(2)
Lagged Life Expectancy.1 (in years)	0.956*** (0.015)	0.956*** (0.014)
Obesity Prevalence (in %)	0.271*** (0.038)	0.270*** (0.037)
Obesity Prevalence Square (in %)	-0.005*** (0.0008)	-0.005*** (0.0008)
Per Capita Health Expenditure (in US $)	0.00002*** (8.13e-06)	0.00002*** (7.70e-06)
High School (in %)	-0.005* (0.002)	-0.005* (0.002)
Smoking (in %)	-0.021*** (0.005)	-0.021*** (0.004)
Alcohol Consumption (in %)	-0.003* (0.002)	-0.003* (0.002)
Per Capita Real GDP (in US $)		-6.32e-08 (1.08e-06)
Time Trend	-0.015* (0.008)	-0.015* (0.008)
Constant	0.0691 (1.593)	0.709 (1.505)
Observations	1398	1398
F statistic	15431.53	13745.05
Prob > F	0.0000	0.0000
Arellano-Bond test for AR(1) in first differences, Z value	-7.37	-7.38
Prob > z	0.000	0.000
Arellano-Bond test for AR(2) in first differences, Z value	1.76	1.77
Prob > z	0.078	0.076
Sargan test, chi2 value	2.75	2.75
Prob > chi2	0.432	0.431
Hansen test, chi2 value	6.00	5.99
Prob > chi2	0.112	0.112
Order of lag used	4	4

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses.

Table 4.6: Effect of Obesity Prevalence on Life Expectancy at Birth, Generalized Method of Moments Estimates (all variables defined in logarithmic form)

Variables	**(1)**	**(2)**
ln (Lagged Life Expectancy.1) (in years)	0.947*** (0.014)	0.945*** (0.013)
ln (Obesity Prevalence) (in %)	0.347*** (0.048)	0.347*** (0.048)
[ln (Obesity Prevalence)] Square (in %)	-0.056*** (0.008)	-0.056*** (0.008)
ln (Per Capita Health Expenditure) (in US $)	0.005*** (0.002)	0.006*** (0.002)
ln (High School) (in %)	-0.001 (0.001)	-0.001 (0.001)
ln (Smoking) (in %)	-0.002* (0.001)	-0.003* (0.001)
ln (Alcohol Consumption) (in %)	-0.004*** (0.001)	-0.004*** (0.001)
ln (Per Capita Real GDP) (in US $)		-0.001 (0.001)
Time Trend	0.0002 (0.0001)	0.0002 (0.0001)
Constant	-0.325*** (0.102)	-0.309*** (0.096)
Observations	1398	1398
F Statistic	10639.72	8953.74
Prob > F	0.0000	0.0000
Arellano-Bond test for AR(1) in first differences, Z value	-7.35	-7.37
Prob > z	0.000	0.000
Arellano-Bond test for AR(2) in first differences, Z value	1.29	1.29
Prob > z	0.198	0.197
Sargan test, chi2 value	0.88	0.80
Prob > chi2	0.831	0.850
Hansen test, chi2 value	3.18	3.13
Prob > chi2	0.364	0.372
Order of lag used	4	4

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses.

4.3.3 Discussion

The results suggest that $LEXP$ and obesity prevalence have a concave relationship. The threshold obesity prevalence beyond which $LEXP$ falls is around 26%, based on GMM estimates. $LEXP$ is found to be decreasing in obesity prevalence during the recent years. For year 2014, $\frac{\partial LEXP}{\partial OBS} \leq 0$ at the mean obesity prevalence. The graphical analysis, based on GMM estimates, highlights that many states in the United States have already passed the threshold obesity prevalence and our model predicts that any further rise in obesity prevalence may reduce $LEXP$ in these states. These states include Alabama, Arkansas, Louisiana, Mississippi, Texas, South Carolina and West Virginia. Our finding that longevity is declining in obesity prevalence for the United States is line with the studies by Preston and Stokes (2011), Fontaine et al. (2003) and Mehta and Chang (2009) which state that obesity is reducing life expectancy in the United States.

Another important result is that per capita health expenditure improves longevity. A $1000 increase in per capita health expenditure can improve $LEXP$ by up to 0.24 months. The results suggest that to increase $LEXP$ by one year, an increase of about 50,000 US dollars in per capita health expenditure is required, or, to increase $LEXP$ by one month, an increase of about 4200 US dollars is needed. This indicates that there is a substantial cost associated with obesity and it is expensive for the US economy.

Behavioural risk factors, as measured by the proportion of population which smokes or consumes alcohol, adversely affects $LEXP$, therefore, reducing their consumption may improve longevity.

4.4 Examining the Relationship between Overweight and Obesity Prevalence and Life Expectancy at Birth in India

In this section, we conduct a similar analysis for India measuring overnutrition in terms of overweight and obesity prevalence which is the proportion of population having BMI ≥ 25 kg/m^2. Here, we present data and estimation results.

4.4.1 Data

Data Sources and Definitions

Data on nutrition and health related variables is taken from state reports of the third and fourth rounds of National Family Health Survey for the years 2005-06 and 2015-16 respectively. Data on life expectancy at birth is available only at state level, therefore, we restrict to state level data in this analysis. Data on life expectancy at birth is taken from Sample Registration System.

State level data is considered for 21 states over two periods, 2005-06 and 2015-16. Data includes observations for men and women across rural and urban regions. Our data set contains 168 observations on each variable.[25] The list of states included in the analysis is provided in Table A.4.3 of Appendix. Although the variable heads are similar but BRFSS and NFHS definitions differ slightly, therefore, we again report these variables along with their definitions.

List of Variables with Definition:

i. **Life Expectancy at Birth:** As defined earlier (in years).[26]
ii. **Overweight and Obesity Prevalence:** Percentage of population having a body mass index value of 25 kg/m^2 or above, that is, overweight plus obese population (in percentage).

[25] For each state, we have eight observations over two time periods, that is, four observations for each year. Therefore, the total number of observations are 168 (= 21 states * 2 time periods * 2 genders * 2 regions).
[26] Life expectancy at birth is available for a period of 5 years interval, therefore, for year 2015-16 we use life expectancy at birth values of 2011-15. Similarly, for year 2005-06 we use values for year 2001-05.

iii. **Per Capita Total Health Expenditure:** The total health expenditure divided by state population (in Indian Rupees). These estimates are available at current market prices, therefore, to obtain data at constant price, we divide entire series by GDP deflator to obtain values at 2004-05 prices.[27]

iv. **Literacy Rate:** Percentage of population who can read a whole sentence or part of a sentence and who have completed standard 6 or higher (who are assumed to be literate) (in percentage).

v. **Tobacco Consumption or Smoking:** Percentage of population consuming tobacco in the form of smoking cigarettes and bidis (in percentage).

vi. **Alcohol Consumption:** Percentage of population consuming/drinking alcohol (in percentage).

vii. **Per Capita Net State Domestic Product (NSDP):** The monetary value of all goods and services produced within the boundaries of the state during a given period of time after deducting the consumption of fixed capital divided by state population (in Indian Rupees). This data is available at constant prices but with different base years, therefore, to obtain constant price series with same base year, we divide entire series by GDP deflator to obtain values at 2004-05 prices.

viii. **Gini Coefficient:** It measures inequality of the distribution of income within a state. It takes a value between zero and one; zero indicates perfect equality and one indicates perfect inequality. Gini coefficient of distribution of consumption is considered.

ix. **Head Count Ratio:** The proportion of population that lives below the poverty line. Tendulkar estimates for year 2004-05 and 2011-12 are considered (in percentage).[28]

x. **Monthly Per Capita Consumer Expenditure (MPCE):** Household consumer expenditure divided by household size (in Indian Rupees).[29]

---

[27] "Total health expenditure is the sum of current and capital health expenditure incurred by government and private sources. Current Health Expenditure is defined as final consumption expenditure of resident units on healthcare goods and services net capital expenditures. Capital expenditures include expenditure on building capital assets, renovations and expansions of buildings, purchasing of vehicles, machines, equipment, medical/ AYUSH/ paramedical education, research and development, training (except on the job trainings), major repair work, etc." (National Health Accounts, 2014-15).
For year 2005-06 we have considered 2004-05 data and for year 2015-16 we have used data for 2014-15.

[28] For year 2005-06 we have considered 2004-05 data and for year 2015-16 we have used data for 2011-12 due to data limitations. The most recent data available is for the year 2011-12.

[29] We consider the data for 2011-12 as proxy for year 2015-16 due to unavailability of data. For year 2005-06 we have taken the data for year 2004.

xi. **Total Food Grains Production:** It is the total amount or quantity of food grains produced in a state throughout one-year period (in Thousand Tonnes).

xii. **Television:** Percentage of households having possession of television (black and white or colour) (in percentage).

xiii. **Mobile:** Percentage of households having possession of mobile phone (in percentage).

xiv. **Computer:** Percentage of households having possession of computer (in percentage).

xv. **Car:** Percentage of households having possession of car (in percentage).

For the variables such as life expectancy at birth, overweight and obesity prevalence, tobacco consumption or smoking, alcohol consumption, life expectancy at age one and death rate, data is segregated at gender and residence level and for literacy rate, data is segregated only at gender level. For the variables such as Gini coefficient, head count ratio, monthly per capita consumer expenditure, television, mobile, computer and car, data is segregated at residence level and for variables such as per capita total health expenditure, per capita net state domestic product and total food grains production, complete state level aggregates have been used. Also, variables such as overweight and obesity prevalence, literacy rate, tobacco consumption or smoking and alcohol consumption are defined for the population in the age group 15-49 years. A complete list of data sources is provided in Table A.4.4 of Appendix.

Descriptive Statistics

Table 4.7 presents the descriptive statistics. The mean $LEXP$ is 68.92 years and mean overweight and obesity prevalence is 17.98%, indicating that about one-fifth of the population in our sample data is having a BMI $\geq$ 25 kg/m^2. This figure is comparable to the average overweight and obesity prevalence at national level for years 2005-06 and 2015-16 (NFHS). The highest overweight and obesity prevalence of 42.5% is for the female population across the urban regions of Andhra Pradesh in year 2015 and the minimum value of 1.5% is for the female population across the rural regions of Jharkhand in year 2005.

We also compared the averages across genders and found that females have both higher $LEXP$, and overweight and obesity prevalence as compared to males and the differences are statistically significant. Comparing the averages across regions, we found that urban regions have both higher $LEXP$, and overweight and obesity prevalence as compared to the rural regions and the differences are statistically significant.

Table 4.7: Descriptive Statistics, India

Variables	Observations	Mean	Standard Deviation	Minimum Value	Maximum Value
Life expectancy at Birth (in years)	151	68.92	4.59	56.20	80.20
Overweight and Obesity Prevalence (in %)	167	17.98	9.81	1.50	42.50
Literacy Rate (in %)	168	74.03	14.75	36.20	98.70
Smoking (in %)	164	15.28	16.29	0	58.90
Alcohol Consumption (in %)	164	16.69	17.11	0	54.60
Per Capita NSDP (in Rs.)	152	36611.10	19324.14	7588.00	79077.21
Per Capita Health Expenditure (in Rs.)	160	1508.39	747.98	499.00	3668.05
Head Count Ratio (in %)	168	24.36	13.48	4.30	60.80
Gini Coefficient	168	0.31	0.60	0.20	0.44
Monthly Per Capita Consumer Expenditure (in Rs.)	160	1396.74	757.47	404.78	3253.29
Total Food Grains (in Thousand Tonnes)	168	10771.92	10163.76	111.73	42550.76
Television (in %)	164	64.66	24.34	10.90	96.60
Mobile (in %)	164	57.90	36.08	0.70	98.80
Computer (in %)	164	8.68	8.29	0	32.40
Car (in %)	164	7.15	7.05	0.10	30.80
Residence	168	0.50	0.50	0	1
Gender	168	0.50	0.50	0	1
Year	168	2010	5.01	2005	2015

4.4.2 Estimation Results and Interpretation

Fixed Effects Estimates

Table 4.8 presents the fixed effects estimation results. Five model specifications have been estimated which differ in terms of covariates used. In all the models, standard errors are clustered at state level. In all the model specifications, we find that the coefficient of overweight and obesity prevalence is positive and statistically significant, and the coefficient of square of overweight and obesity prevalence is negative and statistically significant. The coefficient of overweight and obesity prevalence and its square indicate that $LEXP$ is increasing in overweight and obesity prevalence at a decreasing rate, that is, $LEXP$ has a concave relationship with overweight and obesity prevalence.

We focus on model (5) in Table 4.8. The change in $LEXP$ with an increase in overweight and obesity prevalence is given by $\frac{\partial LEXP}{\partial OBS} = 0.239 - 0.004 * 2\ (OBS)$.[30] The threshold beyond which $LEXP$ falls in overweight and obesity prevalence is 31.2%, $= \left[\frac{-0.2385035}{2*(-0.0038257)}\right]$, calculated using equation (4.4).[31] $LEXP$ is increasing in overweight and obesity prevalence until its prevalence reaches 31.2% and then $LEXP$ declines for higher values of overweight and obesity prevalence. As per this model, at the mean overweight and obesity for year 2015, $LEXP$ is increasing in overweight and obesity prevalence.[32]

Per capita health expenditure has a positive and statistically significant effect on the $LEXP$ suggesting that per capita health expenditure mitigates the adverse impacts of overnutrition on longevity. A Rs.1000 increase in per capita health expenditure improves average $LEXP$ by about 8.4 months.

Literacy rate has a positive and highly statistically significant effect on $LEXP$. Smoking does not have any statistically significant effect on longevity. Alcohol consumption has a statistically significant adverse effect on longevity. A unit increase in the proportion

---

[30] OBS denotes overweight and obesity prevalence for Indian analysis.

[31] Complete values of the estimates are used to compute the thresholds. The thresholds for models (1)-(4) are 33.48%, 31.53%, 31.54% and 32.80% respectively.

[32] $\frac{\partial LEXP}{\partial OBS}\Big|_{OBS=22.25} \geq 0$, where 22.25% is the mean obesity prevalence for the year 2015.

of population drinking alcohol reduces the average $LEXP$ by approximately 0.1 years (1.2 months). Gini coefficient which measures the income inequality does not have any statistically significant effect on $LEXP$.[33]

Sensitivity Analysis - Robustness Check

To check for robustness of our results, we include variables such as head count ratio (as a proxy for poverty), monthly per capita expenditure and total food grains production (as a proxy for food availability) in the model. These results are presented in Table 4.9. Estimates for overweight and obesity prevalence, and its square continue to have expected signs and are statistically significant (except for model (1) where the significance is not obtained at conventional levels). Coefficients on head count ratio, monthly per capita consumer expenditure and total food grains production have expected signs. Inclusion of these covariates does not alter our main results.

Heterogeneity Analysis – Differential Effect across Genders

To examine the differential effect of overweight and obesity prevalence on longevity across genders, we introduce interaction of overweight and obesity prevalence and its square with female dummy. Table 4.10 presents these results. Both positive as well as negative effects of overweight and obesity prevalence are higher among females as compared to males. There exists gender differential effect but these differences do not persist when we control for literacy rate, smoking and alcohol consumption since there could be gender differences within these three variables.

[33] If we include time trend in the model, the variables lose significance. It could be due to some relationship between the time effects and the way variables change overtime.

Table 4.8: Effect Overweight and Obesity Prevalence on Life Expectancy at Birth, Fixed Effects Estimates

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Overweight and Obesity Prevalence (in %)	0.335** (0.125)	0.238** (0.100)	0.239** (0.100)	0.252** (0.092)	0.239** (0.101)
Overweight and Obesity Prevalence Square (in %)	-0.005** (0.002)	-0.004** (0.002)	-0.004** (0.002)	-0.004** (0.002)	-0.004** (0.002)
Per Capita Health Expenditure (in Rs.)	0.002*** (0.0004)	0.0007# (0.0004)	0.0006 (0.0007)		0.0007* (0.0004)
Literacy Rate (in %)		0.202*** (0.037)	0.204*** (0.039)	0.219*** (0.034)	0.201*** (0.036)
Smoking (in %)		0.039 (0.034)	0.039 (0.035)	0.042 (0.033)	0.038 (0.033)
Alcohol Consumption (in %)		-0.095** (0.042)	-0.096** (0.044)	-0.103** (0.043)	-0.093** (0.040)
Per Capita NSDP (in Rs.)			3.94e-06 (0.00002)	0.00002 (0.00001)	
Gini Coefficient					1.689 (3.142)
Constant	61.390*** (1.139)	50.851*** (2.769)	50.691*** (2.874)	49.809*** (2.538)	50.370*** (3.000)
Observations	144	144	140	140	140
R^2	0.8066	0.8847	0.8848	0.8832	0.8850
F Statistic	74.65	86.64	98.64	77.88	73.09
Prob > F	0.0000	0.0000	0.0000	0.0000	0.0000

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses. Standard errors are clustered at the state level. Within R square is reported.
Weakly significant at 11% significance level.

Table 4.9: Effect Overweight and Obesity Prevalence on Life Expectancy at Birth, Robustness Check - Fixed Effects Estimates

Variables	(1)	(2)	(3)
Overweight and Obesity Prevalence (in %)	0.197[@] (0.116)	0.235* (0.122)	0.221** (0.094)
Overweight and Obesity Prevalence Square (in %)	-0.003[&] (0.002)	-0.004* (0.002)	-0.003* (0.001)
Per Capita Health Expenditure (in Rs.)	0.0002 (0.0005)	0.001 (0.0007)	0.0007[%] (0.0004)
Literacy Rate (in %)	0.187*** (0.041)	0.203*** (0.036)	0.164*** (0.034)
Smoking (in %)	0.061* (0.033)	0.036 (0.044)	0.052 (0.040)
Alcohol Consumption (in %)	-0.105** (0.044)	-0.093** (0.040)	-0.109** (0.047)
Head Count Ratio (in %)	-0.056 (0.048)		
Monthly Per Capita Consumer Expenditure (in Rs.)		.00003 (.0004)	
Total Food Grains Production (in Thousand Tonnes)			0.0001** (0.00004)
Constant	54.386*** (4.091)	50.830*** (2.584)	52.312*** (2.266)
Observations	144	142	144
R^2	0.8903	0.8837	0.9007
F Statistic	82.44	86.02	66.81
Prob > F	0.0000	0.0000	0.0000

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses. Standard errors are clustered at the state level. Within R square is reported.
[@] Weakly significant at 10.7% significance level.
[&] Weakly significant at 11.4% significance level.
[%] Weakly significant at 10.8% significance level.

Table 4.10: Effect Overweight and Obesity Prevalence on Life Expectancy at Birth, Heterogeneity Analysis by Gender - Fixed Effects Estimates

Variables	(1)	(2)
Overweight and Obesity Prevalence (in %)	0.268** (0.122)	0.199* (0.112)
Overweight and Obesity Prevalence Square (in %)	-0.004* (0.002)	-0.003 (0.002)
Overweight and Obesity Prevalence * Female Dummy (in %)	0.223*** (0.050)	0.026 (0.076)
Overweight and Obesity Prevalence Square * Female Dummy (in %)	-0.003*** (0.001)	-0.002 (0.001)
Per Capita Health Expenditure (in Rs.)	0.002*** (0.0004)	0.0006 (0.0005)
Literacy Rate (in %)		0.220*** (0.052)
Smoking (in %)		0.042 (0.038)
Alcohol Consumption (in %)		-0.088* (0.046)
Constant	60.867*** (1.177)	49.804*** (3.568)
Observations	144	144
R^2	0.8252	0.8868
F Statistic	48.96	85.85
Prob > F	0.0000	0.0000

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses. Standard errors are clustered at the state level. Within R square is reported.
Note: Female dummy takes value 1 for females and 0 for males.

Instrument Variable – Two-Stage Least Squares Estimates

Household's possession of assets - television and mobile – are used to instrument overweight and obesity prevalence. We consider variables on the percentage of households having possession of television (black and white or colour) and mobile. We construct a weighted average index of these two assets by assigning equal weights to each variable, $Index = [(Tevelevison + Mobile)/2]$, and use this index as an instrument.[34] For overweight and obesity prevalence square, we use square of this index as an instrument. Household's possession of these assets is used as a proxy measure for physical inactivity or sedentary behaviour and can be regarded as a measure for access to sedentary technology.

The identification strategy is based on the assumption that households having possession of these assets/electronics are more likely to have higher average BMI or overweight and obesity due to increased sedentary behaviour. In addition, our instrument is likely to be uncorrelated with other determinants of $LEXP$. The only other way our instrument may affect $LEXP$ is through inequality and income levels, and we control for the same. We use Gini coefficient as measure to gauge inequality and for income level, we use two alternative measures, per capita net state domestic product and monthly per capita consumer expenditure.

The IV-2SLS estimates are reported in Table 4.11. In both the models, standard errors are clustered at state level. Both models include controls for behavioural risk factors, alcohol consumption and smoking, and literacy rate. We also control for per capita NSDP (or monthly per capita consumer expenditure) and Gini coefficient.

The concave relationship between $LEXP$, and overweight and obesity prevalence is robust even when we estimate IV-2SLS model, however, the coefficients do change when we estimate the causal effect. We find much stronger effects of overweight and obesity prevalence and its square on $LEXP$ as compared to the Fixed Effects estimates, that is, the coefficient estimates for both overweight and obesity prevalence and its square have a higher magnitude in the present model. The thresholds are 25.9% and 23.1% for models (1) and (2) respectively. These thresholds are about 5-7% lower than

[34] We find a highly statistically significant correlation between overweight and obesity prevalence and these two variables, 0.8172 for the percentage of households having possession of television and 0.6137 for the percentage of households having possession of mobile.

the thresholds given by the Fixed Effects estimates. Also, at the mean overweight and obesity for year 2015, $LEXP$ is increasing in overweight and obesity prevalence.[35]

Per capita health expenditure positively affects life expectancy at birth although controlling for income and Gini coefficient, per capita health expenditure is statistically insignificant, arguably the two are related. Literacy rate has a positive and highly statistically significant effect. Smoking does not have any statistically significant effect. Alcohol consumption has a statistically significant adverse effect on longevity. A unit increase in the proportion of population drinking alcohol reduces the average $LEXP$ by 0.1 years (1.2 months). This result is same as given by the Fixed Effects estimates. The measures for inequality and income levels do not have a statistically significant effect.

The Kleibergen-Paap rk LM statistic is used for under-identification test. A rejection of the null hypothesis indicates that the model is identified. For both models, we have a p-value of less than 0.01, therefore, we reject the null hypothesis indicating that the model is identified. Kleibergen-Paap rk Wald F statistic for weak identification test reports F values which are higher than the conventional value of 10 or 12 indicating that our instruments are not weak.

A graphical illustration of the relationship between $LEXP$, and overweight and obesity prevalence is presented in Figure 4.5. We have plotted the predicted values of $LEXP$ against overweight and obesity prevalence for model (1) while keeping all other variables included in the model at their mean values. We reproduce the same graph in Figure 4.6 for the two years separately, 2005-06 and 2015-16. In each graph, the threshold overweight and obesity prevalence (25.9%) is labelled and represented by the red coloured vertical reference line. Both the figures highlight, the concave relationship between $LEXP$, and overweight and obesity prevalence.

[35] $\frac{\partial LEXP}{\partial OBS}\Big|_{OBS=22.25} \geq 0$, where 22.25% is the mean obesity prevalence for the year 2015.

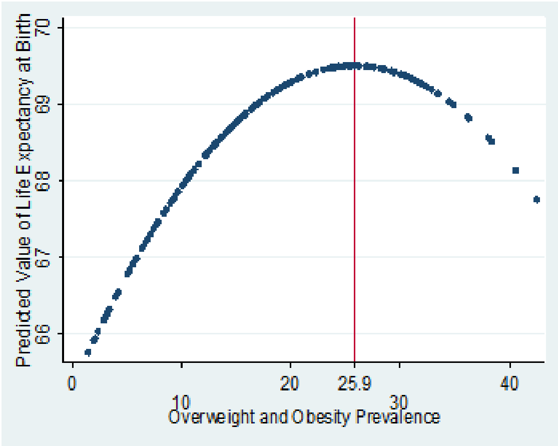


Figure 4.5: Predicted Values of Life Expectancy at Birth, India

Source: Author's calculations based on sample data.

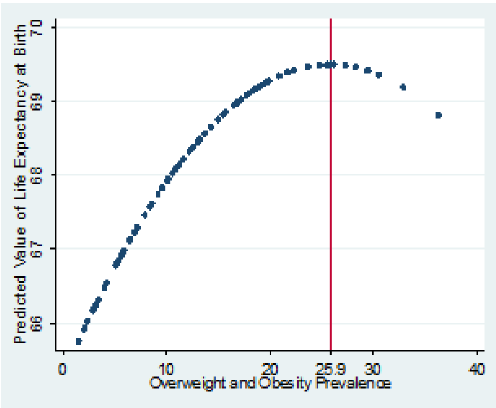
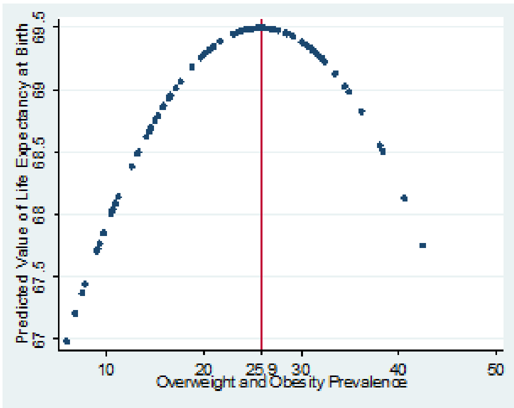
2005-06	2015-16
	

Figure 4.6: Predicted Values of Life Expectancy at Birth for years 2005-06 and 2015-16, India

Source: Author's calculations based on sample data.

Sensitivity Analysis – Robustness Check

To check robustness our IV estimates, we control for variables such as head count ratio and total food grains production. Head count ratio is used as a proxy for poverty and total food grains production is used as a proxy for food availability. These estimates are reported in columns (1) and (2) of Table A.4.5 of Appendix. Based on these results, we find that the concave relationship between overweight and obesity prevalence, and $LEXP$ continues to hold.

We checked sensitivity for our estimates by considering alternative instruments. First, we constructed a new weighted average index by including additional assets such as computer and car possession by households in our index, i.e., $Index^{New} = [(Tevelevison + Mobile + Computer + Car)/4]$. These results are presented in column (3) of Table A.4.5 of Appendix. Our main results continue to hold even when the new index is used as instrument.[36]

Recognising that health expenditure is a key input in the average life expectancy function, we check sensitivity of our estimates by trimming per capita health expenditure by 5%, by removing top 5% and bottom 5% values. This takes care of the outliers' effect. These results are reported in column (4) of Table A.4.5 of Appendix. Our main results are robust to this check.

Heterogeneity Analysis – Differential Effect across Genders

To check for the differential effect of overweight and obesity prevalence on $LEXP$, we introduce interaction terms for overweight and obesity prevalence and its square with female dummy, using the original index as instrument. These results are presented in Table 4.12. The first stage regression results are reported in Tables A.4.6 and A.4.7 of Appendix for models (1) and (2) respectively. We do not find a statistically significant gender differential effect, however, both positive and negative effects of overweight and obesity prevalence are higher for females as compared to males. This result is same as given by the Fixed Effects estimates.

[36] We also considered television possession by households and its square as instruments and find similar results. Using mobile possession by households and its square as instruments also generate similar results.

Table 4.11: Effect of Overweight and Obesity Prevalence on Life Expectancy at Birth, IV-2SLS Estimates with First- Stage Regressions

Variables	(1)			(2)		
	LE	**First Stage - OBS**	**First Stage -OBS Square**	**LE**	**First Stage - OBS**	**First Stage - OBS Square**
Overweight and Obesity Prevalence (in %)	0.326*** (0.109)			0.353*** (0.127)		
Overweight and Obesity Prevalence Square (in %)	-0.006*** (0.002)			-0.008*** (0.003)		
Per Capita Health Expenditure (in Rs.)	0.0007 (0.0008)	0.003 (0.004)	0.146 (0.215)	0.0007 (0.0007)	0.001 (0.003)	0.097 (0.160)
Literacy Rate (in %)	0.196*** (0.034)	-0.133 (0.134)	-2.514 (7.467)	0.183*** (0.037)	0.076 (0.134)	5.786 (6.246)
Smoking (in %)	0.052 (0.038)	-0.027 (0.087)	-2.072 (3.56)	0.051 (0.051)	-0.096 (0.093)	-4.314 (4.548)
Alcohol Consumption (in %)	-0.105** (0.041)	0.052 (0.116)	0.871 (5.997)	-0.096** (0.039)	-0.070 (0.126)	-4.999 (6.104)
Per Capita NSDP (in Rs)	7.81e-06 (4.166)	-0.0001 (0.0001)	-0.006 (0.007)			
Monthly Per Capita Consumer Expenditure (in Rs)				0.0005 (0.0004)	-0.005** (0.002)	-0.242** (0.119)
Gini Coefficient	3.368 (4.166)	20.986 (19.635)	1530.723 (1096.66)	3.621 (4.640)	30.121 (25.679)	1918.276 (1368.74)
Instruments						
Index		0.048 (0.065)	-8.021** (3.475)		-0.067 (0.088)	-13.18*** (4.310)
Index Square		0.002*** (0.0004)	0.133*** (0.024)		0.003*** (0.001)	0.192*** (0.032)
Observations	128	128	128	124	124	124
R^2	0.8766			0.8627		
F Statistic	54.52	89.54%	56.28%	72.22	96.81%	49.08%
Prob > F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Kleibergen-Paap rk LM Statistic	9.885			8.592		
Chi2 p-value	0.0017			0.0034		
Kleibergen-Paap rk Wald F Statistic	13.837			22.770		

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses. Standard errors are clustered at the state level.
LE = Life expectancy at birth and OBS = Overweight and obesity prevalence.
% Sanderson-Windmeijer multivariate F test of excluded instruments reported.

Table 4.12: Effect of Overweight and Obesity Prevalence on Life Expectancy at Birth, Heterogeneity Analysis by Gender - IV-2SLS Estimates

Variables	(1)	(2)
Overweight and Obesity Prevalence (in %)	0.317*** (0.119)	0.333*** (0.128)
Overweight and Obesity Prevalence Square (in %)	-0.006*** (0.002)	-0.007*** (0.003)
Overweight and Obesity Prevalence * Female Dummy (in %)	0.008 (0.101)	0.042 (0.113)
Overweight and Obesity Prevalence Square * Female Dummy (in %)	-0.0005 (0.003)	-0.001 (0.003)
Per Capita Health Expenditure (in Rs.)	0.0007 (0.0008)	0.0007 (0.0007)
Literacy Rate (in %)	0.203*** (0.051)	0.183*** (0.059)
Smoking (in %)	0.054* (0.032)	0.049 (0.043)
Alcohol Consumption (in %)	-0.103** (0.043)	-0.090** (-0.037)
Per Capita NSDP (in Rs)	8.49e-06 (0.00002)	
Monthly Per Capita Consumer Expenditure (in Rs.)		0.0005 (0.0006)
Gini Coefficient	3.406 (4.229)	3.619 (4.715)
Observations	128	124
R^2	0.8765	0.8636
F Statistic	46.47	70.78
Prob > F	0.0000	0.0000
Kleibergen-Paap rk LM Statistic	7.188	9.123
Chi2 p-value	0.0073	0.0025
Kleibergen-Paap rk Wald F Statistic	2.110	10.486

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses. Standard errors are clustered at the state level.
Female Dummy takes value 1 for female gender and 0 for male gender.

4.4.3 Discussion

The results suggest that $LEXP$, and overweight and obesity prevalence have a concave relationship in India. The threshold overweight and obesity prevalence beyond which $LEXP$ falls is around 26%, based on IV-2SLS estimates. $LEXP$ is found to be increasing in overweight and obesity prevalence. For year 2015, $\frac{\partial LEXP}{\partial OBS} \geq 0$ at the mean overweight and obesity prevalence. The graphical analysis shows that by year 2015-16 a higher number of states have passed the threshold overweight and obesity prevalence and are facing a risk of longevity decline due to increased overnutrition. These states include Andhra Pradesh, Delhi, Gujarat, Kerala, Punjab and Tamil Nadu. We suggest for policy attention in these states. More specific analysis among these states can help to better devise more targeted and effective policies.

Per capita health expenditure improves longevity. However, controlling for income and Gini coefficient in IV-2SLS models this effect is loses significance. The magnitude of the effect of per capita health expenditure on $LEXP$ is similar across both Fixed Effects and IV-2SLS models. A Rs.1000 increase in per capita health expenditure can increase $LEXP$ by up to 8.4 months. That is, to increase $LEXP$ by one year, an increase of about Rs. 1400 in per capita health expenditure is required. This rise in the amount required to increase $LEXP$ by one year is equivalent to about 100% increase in the current mean per capita health expenditure. These results suggest that increasing health expenditure may generate substantial gains in $LEXP$ in India.

4.5 Conclusion

This chapter examines the effect of overnutrition, as measured by overweight and/or obesity prevalence, on $LEXP$ using an empirical framework. We empirically test the analytical model presented by Bansal and Zilberman (2016) at state level aggregate data and provide an evidence for existence of a concave relationship between overweight and/or obesity prevalence, and longevity for the United States and India. In addition, we also establish a causal relationship between overnutrition and longevity.

The most important result is that the $LEXP$ and overweight and/or obesity prevalence have a concave relationship, and existing levels of overweight and/or obesity prevalence determine the nature of the effect of overnutrition on longevity. In other words, we may say that $LEXP$ will increase (decrease) in overweight and/or obesity prevalence at lower (higher) levels of overweight and/or obesity prevalence. An interesting finding is that longevity is decreasing in overnutrition for the United States and increasing in overnutrition for India. We show this graphically as well as algebraically. Based on the graphical analysis, it is found that obesity prevalence in most states has passed the threshold level and is placed on the downward sloping segment of the concave curve for the United States by year 2014, as presented in Figure 4.4. However, in India for year 2015, most states have an overweight and obesity prevalence which is less than the threshold level and are placed on the upward sloping segment of the concave curve, as presented in Figure 4.6. Also, based on the coefficient estimates obtained from the estimation of different models, it is found that $\frac{\partial LEXP}{\partial OBS} \leq 0$ at the mean obesity prevalence in the United States for year 2014. For India, $\frac{\partial LEXP}{\partial OBS} \geq 0$ at the mean overweight and obesity prevalence in the year 2015.

Another important result is that per capita health expenditure can generate substantial gains in $LEXP$. This effect is much stronger in low- and middle-income country like India and greater gains in $LEXP$ can be generated by increasing per capita health expenditure among these countries. Using the IV-2SLS coefficient estimates, we find that a Re.1 increase in the per capita health expenditure increases $LEXP$ by 0.0007 years. This implies that to increase $LEXP$ by one year, the additional per capita health expenditure required is Rs.1400. Comparing this to the mean per capita health expenditure in India, amounts to almost doubling of per capita health expenditure. At margin a lot more health expenditure is required to increase $LEXP$ in the United States.

In the United States, the effects of obesity prevalence on $LEXP$ differs across genders with men facing both higher positive as well as negative effects of obesity. However, we do not find enough evidence on the gender differential effect of overweight and obesity prevalence on $LEXP$ for India.

The policy implication of this study is that the overnutrition prevalence must be reduced in the states having a high overweight or obesity prevalence as these states face a higher risk of longevity decline resulting from overnutrition. In addition, a higher budgetary allocation to the health sector can also help in mitigating the adverse effects of overnutrition on longevity.

Chapter 5

Overnutrition and Risk of Diabetes: A Micro Data Analysis for India

5.1 Introduction

In the previous chapter, we have documented the adverse effects of overnutrition on longevity at the macro level. In the present chapter, we examine the micro level relationship between overnutrition and diabetes in the context of India.

The rise in the diabetes prevalence during the past decade has begun to pose a new challenge to the health policy makers in India. In 2017, about 72 million people (8.8% of the total population having age 18 years or above) and 20% of the urban population was diabetic in India (International Diabetes Federation (IDF)). According to Diabetes Foundation of India, "people suffering from diabetes are likely to go up to 80 million by 2025, making India the 'Diabetes Capital' of the world". Analysing NFHS data for the increase in the diabetes prevalence in India (in the age group 15-49 years) over a ten-year period, 2005 to 2015, we find that diabetes prevalence has doubled in both rural as well as urban areas, and there has been a considerable increase in almost every state.

Overnutrition has been found to be a major risk factor for a number of diseases such as diabetes, hypertension, heart diseases, certain type of cancers, etc. (Huffman et al., 2011; Colditz et al., 1995 and Dhana et al., 2016). Overnutrition is one of the potential factors that may generate insulin resistance, which in turn may increase the sugar or glucose content in the blood leading to diabetes (Kahn and Flier, 2000). Colditz et al. (1995) using a prospective cohort study on women in the United States find that the risk of diabetes is increasing in BMI. The study by Huffman et al. (2011) finds similar results among married women in Delhi, India. Other factors that may lead to diabetes include smoking, alcohol consumption, high sugar intake, genetic predisposition, etc. (Fagard and Nilsson, 2009; Carlsson et al., 2003 and Howard et al., 2004).

India is going through a nutritional transition brought about by a rapid emergence of the overnutrition. The rising overnutrition may be associated with the growing diabetes problem in India. Overnutrition is associated with the increased risk of mortality and morbidity (Bhattacharya and Sood, 2011 and Preston and Stokes, 2011). However, Asian population faces this risk even at lower BMI values, ranging from 23-25 kg/m^2 and above, that is, the risk of chronic conditions is higher among Asian population due to the increased susceptibility towards NCDs even at lower BMI levels as compared to the population in the European countries and the United States (Gray et al., 2011; Razak et al., 2007 and Asia-Pacific Perspective Report, WHO, 2000).[37]

In this chapter, we examine the causal effect of an increase in BMI on the likelihood of suffering from diabetes using an individual level nationally representative data set in India. This is the first study to examine the micro level relationship between BMI and diabetes for population across India while addressing the potential endogeneity arising from the omitted variable bias. The BMI of an individual is likely to be correlated with the omitted determinants of his/her diabetes status. These omitted variables are expected to be related to the individual's genetic and non-genetic predisposition towards overweight and obesity as well as diabetes. We address this issue by using an instrumental variable approach and instrument BMI of an individual by BMI of a non-biologically related household member. The BMI of a non-biologically related household member is correlated with the common household environment but there is no reason to believe that it will systematically affect the individual's predisposition towards diabetes. We also control for several covariates on individual characteristics, household characteristics and behavioural risk factors such as tobacco and alcohol consumption, eating habits, etc. We extract individual level data from the fourth round of NFHS for the year 2015-16.

Studies such as Gray et al. (2011) and Sepp et al. (2014) have estimated the relationship between overnutrition and NCDs. Most studies are, however, based on a small sample

[37] We examined the IDF and WHO recent estimates on the diabetes prevalence and obesity rates for adult population and found that high-income countries like United Kingdom, France and Australia had a lower diabetes prevalence than the low- and middle-income countries like India, China and Sri Lanka despite having much higher obesity rates. Asian countries like Bangladesh, Pakistan, Bhutan, Sri Lanka, India and China have an obesity prevalence between 3-6% and the diabetes prevalence is found to be 7% or above among these countries, with India having 8.8% diabetic population. The European country like France has a diabetes prevalence of 7.2% with obesity prevalence of 21.6%. Australia and United Kingdom have a diabetes prevalence of 6.5% and 5.9% respectively.

size, and, the results may not be representative for the entire population. Further, much of the evidence on the link between the overnutrition and NCDs comes from the high-income countries (Rowley et al., 2017; Geiss et al., 2017 and Sikdar et al., 2010). The findings from these studies cannot be generalised for the Indian population due to the regional differences in the body types and distribution of body fat. South Asian population is found to have a higher abdominal obesity as compared to the population in the European regions, therefore, the susceptibility towards certain types of diseases, such as diabetes, may vary across these regions even if the BMI values are comparable (Patel et al., 2001 and Asia-Pacific Perspective Report, WHO, 2000).[38]

The evidence on the effect of BMI on diabetes for Indian population is limited. Huffman et al. (2011) consider a cohort sample of 1100 women in Delhi and show that an increase in BMI has a statistically significant impact on diabetes among married women. The study by Ramachandran et al. (2001) finds a positive association between diabetes and BMI for urban population across six major cities of India. None of the studies, however, have considered WHO Asian BMI classification, which defines an individual having BMI ≥ 23 kg/m^2 as overweight or obese to examine the effect of overnutrition on diabetes for India.

In this study, our main dependent variable is diabetes status of an individual. We use two alternative measures for indicating diabetes status across individuals, one, self-reported diabetes status, and the other, blood glucose levels. This also acts as a robustness check for our estimates. In the self-reported diabetes status measure, individuals report whether or not they suffer from diabetes. For the second measure, NFHS reports blood glucose levels measured at the time of the survey. We convert the reported blood glucose levels into an ordinal measure by dividing it into three categories. The ordinally defined blood glucose levels gives us an advantage of estimating the effect of BMI on prediabetes as well. In addition, it also addresses the issue of any measurement error in the self-reported diabetes status.

[38] There exist data limitations in the comparison of abdominal obesity across the countries since comparable national level estimates are not available although some of the papers based on small population samples across specific regions do provide with a rough estimate on the abdominal obesity. Olinto et al. (2017) review available literature and state that prevalence of abdominal obesity in South Asian population is around 69% which is much greater than the general obesity. “Abdominal obesity is measured by waist circumference, waist–hip ratio and waist–height ratio. A waist circumference of 90 cm or above for men and 80 cm or above for women defines abdominal or central obesity for South Asian population” (IDF).

We aim at estimating the change in the probability of being diabetic with an additional unit gain in BMI. Our interest lies in comparing this effect across non-overweight and overweight or obese population. For this comparison, we apply both WHO International BMI classification, which defines an individual having BMI $\geq$ 25 kg/m^2 as overweight or obese as well as WHO Asian BMI classification, which defines an individual having BMI $\geq$ 23 kg/m^2 as overweight or obese. One may expect the urban population and the population belonging to the higher wealth quintiles to face a higher risk of diabetes due to the lifestyle related factors and increased access to calorie dense foods, therefore, we also examine the heterogeneity in the effect of BMI on diabetes across different subgroups of the population based on gender, region – rural and urban and different wealth quintiles.

The findings of this study have policy implications for several reasons. Diabetes, unlike other NCDs, which mainly affect older age group population, affects younger age group population as well (Colditz et al., 1995 and Huffman et al., 2011). Also, IDF estimates show that in India in the year 2017, of those who died from diabetes, 50.7% of people died before the age of 60 years, that is, 50.7% of deaths due to diabetes are among the individuals under the age of 60 years.[39] These estimates show that not only diabetes causes morbidity and mortality, but also that these effects are being witnessed even among young and medium age group population (below 60 years of age). Diabetes also elevates the risk of other NCDs such as cardiovascular diseases, strokes, etc. (Asia-Pacific Perspective Report, WHO, 2000). Individuals with diabetes are less likely to report having a good health as compared to the non-diabetic individuals. Diabetes reduces health adjusted life expectancy (Sikdar et al., 2010). This provides a strong case for identifying the potential factors that contribute to the rise in diabetes in India. This will inform policy makers to undertake suitable policy interventions to arrest the growing rates of diabetes in India. The relevance of this study in policy making can also be explained by huge monetary cost burden associated with diabetes. Treatment of diabetes is expensive and is expected to impose an economic burden in the form of

[39] At global level in year 2017, of those who died from diabetes, 46.6% of people died before the age of 60, that is, 46.6% of deaths due to diabetes are among the people under the age of 60 years.

increased healthcare spending (Cawley and Meyerhoefer, 2012; Ramachandran et al., 2007 and Yesudian et al., 2014).[40]

The rest of the chapter is organised as follows. Section 5.2 presents the conceptual framework and the methodology applied. Section 5.3 discusses the data set used along with definition of the variables considered in the analysis. This section also presents the descriptive statistics. Section 5.4 presents the estimation results and their interpretation. Section 5.5 presents a discussion on the findings of the study. Finally, Section 5.6 presents the concluding remarks.

5.2 Conceptual Framework and Methodology

IDF has identified physical inactivity, consumption of unhealthy foods and lifestyle changes towards modernisation (characterised by sedentariness) as the factors that influence diabetes. BMI of an individual captures the effect of most of these factors as the changes in any these factors gets reflected in the BMI of an individual. Higher consumption of unhealthy foods and lower physical activity are expected to bring a positive change in the BMI of an individual. A rise in the BMI of an individual is expected to increase his/her susceptibility towards diabetes as well as towards higher blood glucose levels (Malley et al., 2010 and Sepp et al., 2014). It is possible that a higher BMI increases individual's blood glucose levels but the levels are not high enough to be characterised as diabetes, that is, an individual may become prediabetic (defined in the next paragraph) initially and later diabetic with a further rise in the blood glucose levels if adequate measures are not taken to control the rising blood glucose levels. Therefore, we conduct a twofold analysis by estimating the effect of a rise in BMI on both self-reported diabetes status as well as the ordinal blood glucose levels of an individual. As discussed in Section 5.1, the risk of diabetes is expected to increase

[40] Diabetes is associated with huge direct as well indirect costs. Direct costs include hospital, drug, transportation costs, etc. while indirect costs include loss of working days due to absenteeism, loss due to some permanent disability, etc. Although there are some estimates on direct costs of diabetes but only a little is known about the indirect costs of diabetes (Cawley and Meyerhoefer, 2012; Ramachandran et al., 2007 and Yesudian et al., 2014). Also, it is staggering to find that "12% of global health expenditure is spent on diabetes which amounts to about 727 billion dollars" (IDF, 2017).
"In 2017, worldwide 1,736 US dollars per person were spent on population with diabetes (IDF). Countries such as United States, United Kingdom, France and Australia spent more than 5,000 US dollars per person on population with diabetes. While this figure is relatively low for most of the South Asian countries and remained around or below 100 US dollars per person. For India, this figure is 119.4 US dollars per person" (IDF, 2017). These estimates highlight the huge economic burden associated with diabetes in the form of considerably high health care costs.

in overnutrition, therefore, we may expect an overweight or obese individual to face a higher risk of diabetes as compared to a non-overweight individual.

We now define the dependent variable used in the analysis. The main health outcome variable is the diabetes status of an individual. We measure this variable using two alternative indicators – self-reported diabetes status and blood glucose levels (ordinal measure). For the first measure, we use self-reported diabetes status as the outcome variable which takes value 1 if an individual is diabetic and 0 otherwise. For the second approach to measure diabetes, we assign ordinal values (0, 1and 2) to the blood glucose levels of individuals by dividing these values into three mutually exclusive categories. The blood glucose level measures the amount or concentration of the glucose in a blood sample as milligrams per decilitre (mg/dl). Following the random glucose/sugar test, we have the following three categories for the blood glucose levels:[41]

(i) Less than or equal to140 mg/dl corresponds to low or moderate blood glucose levels – *Normal Blood Glucose Levels*
(ii) Between 141 and 200 mg/dl corresponds to high blood glucose levels – *Prediabetes*
(iii) Greater than 200 mg/dl corresponds to very high blood glucose levels – *Diabetes*

In our analysis, the ordinally defined blood glucose levels assign value 0 to normal blood glucose levels, 1 to prediabetes and 2 to diabetes. Using the blood glucose levels of an individual not only allows us to measure diabetes status but also provides with a measure for prediabetes and normal blood glucose levels, which enables us to quantify the effect of a rise in BMI on both diabetes as well as prediabetes. The study by Dall et al. (2014) finds that both diabetes and prediabetes contribute to a rise in the economic burden in terms of higher healthcare costs.

One may also expect that the population living in the urban areas and the population belonging to the higher wealth quintiles to face a higher risk of diabetes. This can be explained by the differences in the consumption and physical activity patterns across

[41] Random glucose/sugar test is a diagnostic test conducted to identify the diabetes status of an individual. Random blood glucose levels are tested, and based on the concentration of glucose in the blood sample, individual's diabetes status is identified as per categories defined above. Oral glucose tolerance test is another test that is used to diagnose diabetes amongst individuals and it is also based on the above defined categories.

different subpopulations. Olinto et al. (2017) reviewing available literature find that socioeconomic status in terms of higher income and wealth are associated with higher obesity among men. The socioeconomic status and urban lifestyle factors may affect diabetes status through higher BMI levels, therefore, we also examine the heterogeneity in the effect of BMI on diabetes across different regions and wealth quintiles.

Based on the two measurements of the outcome variable, we test the following hypotheses:

Hypothesis 1: Being overweight or obese increases the risk of diabetes among Indian population, that is, with an increase in BMI, the likelihood of being diabetic increases more for an overweight or an obese individual as compared to a non-overweight individual.

Hypothesis 2: Being overweight or obese increases the risk of prediabetes among Indian population, that is, with an increase in BMI, the likelihood of being prediabetic increases more for an overweight or an obese individual as compared to a non-overweight individual.

Hypothesis 3: Population belonging to the higher wealth quintiles is more likely to be prediabetic and diabetic as compared to the population among lower wealth quintiles.

Hypothesis 4: Population living in the urban areas is more likely to be prediabetic and diabetic as compared to the population living in the rural areas.

While the ordinal measure of diabetes can test all the above hypotheses, the self-reported diabetes status measure tests all hypotheses except hypothesis 2.[42] We test the third and fourth hypotheses for a sub-sample comprising of overweight or obese population as they are expected to be facing a higher risk of diabetes. We identify an individual as overweight or obese using WHO International BMI classification. Additionally, we also test our hypotheses using WHO Asian BMI classification.[43]

[42] The self-reported diabetes status measure tests hypotheses 3 and 4 only for diabetes.
[43] An advantage of using individual level data is that it allows us to fix different BMI cut-offs for the analysis, which is not possible in an aggregate data set (at district, state or national level) since the cut-offs are predetermined in the data set (commonly at 18.5 kg/m^2, 25 kg/m^2 and 30 kg/m^2).
We apply the two BMI classifications only to test hypotheses 1 and 2.

Our main explanatory variable of interest is the BMI of an individual. We control for a rich set of covariates both at the individual level as well as at household level that are likely to affect the risk of diabetes. Additionally, we control for the state fixed effects. Individual characteristics include age, gender, educational attainment, behavioural risk factors and eating habits. Behavioural risk factors controlled for in our regressions include a comprehensive set of variables that measure tobacco consumption of an individual such as – smoking cigarette, smoking pipe, chewing tobacco, snuffing, smoking cigar, chewing paan, gutkha, paan with tobacco, etc., and alcohol consumption. These risks factors are likely to affect the diabetes status of an individual (and blood glucose levels). Available literature suggests that smoking elevates the risk of diabetes. Smoking generates insulin resistance leading to the increased risk of diabetes (Chang, 2012 and Fagard and Nilsson, 2009). Studies have also suggested for smoking cessation programs. In regard to alcohol consumption and diabetes or blood glucose levels, available literature suggests that moderate consumption may reduce the risk of diabetes (Howard et al., 2004 and Carlsson et al., 2003) while binge drinking may increase this risk (Carlsson et al., 2003 and Kerr et al., 2009). This directs towards the potential effect of alcohol consumption on diabetes status as well as blood glucose levels of an individual.

Another important factor that may have an impact on diabetes status of an individual are the eating habits. We capture the eating habits of an individual by looking at the frequency of consumption for specific food or drink items. We focus on the daily or weekly consumption of fried foods and aerated drinks. These variables also capture individuals' health and consumption preferences. Gulati and Misra (2014) infer that increase in per capita sugar consumption leads to the development of insulin resistance, abdominal adiposity and risk of diabetes. Food habits such as consumption of aerated drinks, fast-foods, fried foods, etc., increase the risk of obesity and insulin resistance (Pereira et al., 2005; Astrup, 2005 and Teufel-Shone et al., 2014).

We also control for household characteristics such as wealth quintile, family structure (nuclear or joint), region (rural or urban), religion, caste, availability of health insurance, whether the household belongs to below poverty line and other covariates (listed in Section 5.3).

Although we control for a large number of covariates, we still expect the unobserved genetic and other related factors to affect the relationship between BMI and diabetes status. Genetic factors may influence both BMI and diabetes status of an individual. An individual with a family history of diabetes is more likely to develop diabetes even without being overweight or obese (Asia-Pacific Perspective Report, WHO, 2000 and Bener et al., 2005).[44] Bener et al. (2005) find that the reported diabetes is higher among population with a family history of diabetes. These unobserved genetic and other related factors may lead to endogeneity resulting from the omitted variable bias. Therefore, we resort to an Instrumental Variable estimation to address the potential endogeneity. The endogeneity issue is elaborated later in this section.

Empirical Framework

I. Body Mass Index and Self-Reported Diabetes Status: Probit and IV-Probit Model

The outcome variable, self-reported diabetes status, is a binary variable, therefore, we estimate a Probit model (Greene and Hensher, 2010). The following model is estimated, having D_i^* as the dependent variable:

$$D_i^* = \beta' X_i + \nu_i \tag{5.1}$$

where,

$$D_i = \begin{cases} 0 \; if \; individual \; is \; non-diabetic, \\ 1 \; if \; individual \; is \; diabetic. \end{cases} \tag{5.2}$$

i = 1, 2, …., n, represents i^{th} individual;

D_i^* represents latent selection variable for self-reported diabetes status of i^{th} individual and is unobserved;

X_i represents vector of controls including BMI for i^{th} individual;

[44] We use self-reported diabetes status. Due to data limitations, we are unable to identify whether the individual has diabetes type-1 or type-2. “Type-1 diabetes is also called juvenile-onset diabetes as it often begins in childhood. This type of diabetes may be caused by a genetic predisposition. Type-2 diabetes is called adult-onset diabetes” (Asia-Pacific Perspective Report, WHO, 2000).

v_i represents the error term and is assumed to be independent of X_i and has a standard normal distribution.

We estimate a binary response model, in which a non-linear function, $\Phi(.)$ which is a standard normal cumulative distribution function in case of Probit models, is applied to the response function. For estimating binary or ordinal response models, Maximum Likelihood Estimation (MLE) is used. MLE maximises the log-likelihood function for a sample of size n, which is a function of parameters and observed values of the dependent and independent variables, and hence determines the parameter estimates of the model. MLE estimates for the random samples are consistent, asymptotically normal and asymptotically efficient (Wooldridge, 2006).

We first estimate the Probit model assuming that there are no unobserved factors that affect both BMI and self-reported diabetes status of an individual, that is, $Cov\ (X_i, v_i\ ) = 0$. We estimate the average marginal effects of BMI on self-reported diabetes status and examine their signs and magnitudes. We are interested in estimating the partial or marginal effect of BMI on the probability of being diabetic, i.e., effect of BMI on $P(D_i = 1|\ X)$, and test for the following relation:

$$\left[\frac{\partial P(D=1|\ X)}{\partial BMI};\ if\ BMI\ \geq 25\ kg/m^2\right] > \left[\frac{\partial P(D=1|\ X)}{\partial BMI};\ if\ BMI < 25\ kg/m^2\right] > 0 \tag{5.3}$$

that is, change (or increase) in the probability of being diabetic with a unit increase in BMI is higher among the overweight or obese individuals as compared to the non-overweight individuals.

IV-Probit Model

In the relationship between BMI and diabetes status, the unobserved genetic and other related factors may play a role. An individual may inherit the risk of developing diabetes from his/her biological parents, and the genetic factors may also influence overweight and obesity status thereby BMI of an individual. We suspect potential endogeneity in the relationship between BMI and diabetes status in the form of omitted variable bias resulting from the unobserved genetic and other related factors. Our sample data provides self-reported values for the diabetes status, which could introduce another source of endogeneity in the form of measurement error in the self-reported diabetes status of the individuals. Although in the case of large dataset, the measurement error in the dependent variable do not bias the estimates (Fearon, 2001). We resort to an Instrumental Variable estimation which addresses the endogeneity problem caused by both omitted variable bias and measurement error.

We instrument BMI of an individual using BMI of a non-biologically related household member. We instrument an individual's BMI with the BMI of his/her spouse, BMI^S. The instrument must fulfil the following two requirements (Wooldridge, 2006):

(i) BMI of a non-biologically related household member, BMI of individual's spouse, must be uncorrelated with the unobserved factors that explain variations in the diabetes status of an individual, i.e., the instrument must be uncorrelated with the error term:

$$Cov\ (BMI^S, \nu\ ) = 0 \tag{5.4}$$

(ii) Instrument must be correlated with the BMI of individual, in other words, instrument must be powerful:

$$Cov\ (BMI^S, BMI\ ) \neq 0 \tag{5.5}$$

Common household factors may affect BMI of all residing individuals in a similar way due to shared family or household environment (Nelson et al., 2006 and Hewitt, 1997). Studies have also documented the similarities in BMI movements among married couples (Cobb et al., 2015; Falba and Sindelar, 2008 and Katzmarzyk et al., 2002). Therefore, we expect the BMI of an individual and BMI of his/her spouse to be correlated.

For BMI of spouse to be a valid instrument it should not have an independent effect on the diabetes status of the individual. BMI of an individual's spouse is likely to be uncorrelated with the unobserved genetic factors that affect the diabetes status of the individual. However, it is possible that the common household factors which affect BMI of the individual may also affect his diabetes status. Therefore, we control for several variables on the household characteristics in our model. Inclusion of these variables should control for the effects of such household factors, if they exist. In addition, it is possible that the common household factors do not necessarily affect the diabetes status of the individual.

With an objective to measure the causal effect of BMI on self-reported diabetes status of an individual and to address the potential endogeneity problem, we consider equation (5.1) and estimate an IV-Probit model. The first stage equation for this model can be written as:

$$BMI_i = \delta_0 + \delta_1 BMI_i^s + \delta_2 x_i + \eta_i \tag{5.6}$$

where, BMI_i represents BMI of the i^{th} individual; BMI_i^s represents BMI of i^{th} individual's spouse (used as an instrument); x_i represents vector of controls (excluding BMI for i^{th} individual, that is, x_i includes all exogenous variables of the second stage regression) and η_i is the error term.

II. Body Mass Index and Blood Glucose Levels: Ordered Probit Model

Since the indicator of diabetes status is a categorical variable, and has more than two ordered categories, we estimate an Ordered Probit model (Becker and Kennedy, 1992; Boes and Winkelmann, 2006 and Chiburis and Lokshin, 2007). We follow the methodology as described above with the dependent variable now being, BG_i^*:

$$BG_i^* = \alpha' X_i + \varepsilon_i \tag{5.7}$$

Blood glucose levels (dependent variable) are sorted into $j + 1$ categories, where j = 0, 1, 2:[45]

$$BG_i = \begin{cases} 0 \ \ if \ \ BG_i^* \le \mu_1 \\ 1 \ \ if \ \mu_1 < BG_i^* \le \mu_2 \\ 2 \ \ if \ \mu_2 < BG_i^* \end{cases} \tag{5.8}$$

where BG_i represents the observed blood glucose levels for i^{th} individual. The μ_j's are the threshold coefficients or cut-off points. We estimate the probability of an individual belonging to one of the j categories:

$$P(\mu_j < BG_i^* \le \mu_{j+1}) = \Phi(\mu_{j+1} - \alpha' X_i) - \Phi(\mu_j - \alpha' X_i) \tag{5.9}$$

We estimate the above defined model using MLE and obtain the average marginal effects of BMI on ordinal blood glucose levels and examine their signs and magnitudes.

[45] Here, total categories are $j + 1 = 2 + 1 = 3$.

We hypothesise that with a rise in BMI an overweight or obese individual is more likely to be prediabetic (diabetic), that is, the increase in the probability of being prediabetic (diabetic) with a unit increase in BMI is higher among the overweight or obese individuals as compared to the non-overweight individuals:

$$\left[\frac{\partial P(BG=j|\,X)}{\partial BMI};\ if\ BMI\ \geq 25\ kg/m^2\right] > \left[\frac{\partial P(BG=j|\,X)}{\partial BMI};\ if\ BMI < 25\ kg/m^2\right] > 0\ ;\ j = 1, 2 \qquad (5.10)$$

5.3 Data and Descriptive Statistics

For the purpose of empirical analysis in the present chapter, we extract individual level data from Demographic and Health Surveys (DHS) of India, namely, NFHS. We consider the fourth round of NFHS for the year 2015-16. NFHS provides representative data for the population in India. This survey has rich information on household characteristics and individual characteristics such as age, education, anthropometry, diseases and related sufferings, etc. The survey reports the measured levels of blood glucose (our health outcome variable) although the diabetes status is self-reported. The survey covers females having age 15-49 years and males having age 15-54 years. We extract individual level data from three different Stata format data files published by DHS and merged these files into one, for only desired variables. These files are Household Member Recode, Individual Recode (Women's Recode) and Men's Recode. Our analysis considers all 36 states and union territories of India. The list of states and union territories included in the analysis is provided in Table A.5.1 of Appendix. The list of variables included in the study along with their definitions is provided in Table 5.1. In our sample, we include all the observations that report BMI, and either self-reported diabetes status or blood glucose levels. This gives us a total sample size of about 0.8 million observations.

In IV-Probit model, we limit our sample to the individuals who are married and are currently living in the same household. Since, for our IV model we need to know the relationship between household members, and NFHS provides the relationship data for each individual in terms of their relationship to the head of the household and not with regard to all household members, therefore, our sample further restricts to married couples living together in the same household of whom either is head of the household.[46] This section presents the descriptive statistics for the full sample data.

Descriptive Statistics

Table 5.2 presents the descriptive statistics. The mean blood glucose level for the total sample is found to be 104.7 mg/dl. About 1.5% of individuals in our sample are diabetic based on the self-reported diabetes status. This is lower than the estimates given by IDF (8.8% for year 2017). This could be due to a couple of reasons, individuals not being aware of their diabetes status, differences in age groups considered for measuring the diabetes prevalence, and also the year of sample data. Our estimate is based on age group 15-49 years for females and 15-54 years for males while IDF estimate for diabetes is for 20-79 years age group. Diabetes prevalence is expected to increase with age. Based on blood glucose levels, 94% individuals have normal blood glucose, about 5% are prediabetic and about 1% are diabetic (in Table 5.4). Blood glucose levels of some diabetic individuals could be regulated via use of medicines. The mean BMI is 21.71 kg/m^2 indicating that on average population belongs to normal weight category. The average age in our sample is 30 years. About 86% individuals are females and 73% individuals are married.

Table 5.3 presents the descriptive statistics grouped by overweight and obesity status. We also report the mean difference across two groups with its statistical significance. It can be seen that both average blood glucose levels (both actual and ordinal values) and average diabetes prevalence (self-reported) are higher among overweight or obese individuals as compared to the non-overweight individuals. Average diabetes prevalence (self-reported) is three times among the overweight or obese individuals as compared to the non-overweight individuals. Mean Blood glucose levels are 10 mg/dl higher among overweight or obese individuals. The mean BMI among non-overweight individuals is 20.25 kg/m^2 and for overweight or obese individuals it is 28.32 kg/m^2. The average age of sample which is overweight or obese is 6 years higher than the non-overweight sample implying that the BMI tends to rise with age. Overweight or obese individuals' sample has higher averages for education, fried food and aerated drinks

[46] For every individual, NFHS provides data on their relationship to the head of the household. Therefore, we consider only those individuals in our sample who are head of the household and use BMI of individual who reported themselves as husband or wife of the household head as an instrument for the BMI of the head of the household. That is, we use $BMI_i^{Husband\ or\ Wife}$ as an instrument for BMI_i^{Head}.

consumption, wealth quintile, and are more likely to be married and belong to urban regions as compared to the non-overweight individuals' sample. Also, overweight or obese individuals are less likely to belong to below poverty line households, schedule caste and schedule tribe.

In Table 5.4, we report the individual and household characteristics (for only selected variables) for the full sample data and two sub-samples based on overweight and obesity status of individuals. These values are reported in terms of the proportion of individuals belonging to each sub-category. We find that a greater proportion of overweight or obese individuals are prediabetic or diabetic as compared to the non-overweight individuals. Overweight or obese individuals are more likely to belong to the higher wealth quintiles and reside in urban areas. The proportion of male and female population across overweight or obese and non-overweight sub-samples does not vary much. The descriptive statistics for the restricted sample are provided in Table 5.5.

Now, we graphically analyse the relationship between self-reported diabetes status and BMI. Figure 5.1 illustrates the BMI distribution by self-reported diabetes status for the full sample. For plotting the distributions, we, first, divide the BMI data for the full sample on the basis of self-reported diabetes status of the individuals and then we plot two separate BMI distributions for diabetic and non-diabetic population. In Figure 5.1, the solid red line represents the BMI distribution for the diabetic population and dash-dotted blue line represents the BMI distribution for the non-diabetic population. It can be observed that the BMI distribution for diabetic population lies to the right of the distribution for the non-diabetic population. This indicates that the diabetic population is more likely to have higher BMI, or in other words, we can say that at lower BMI values an individual is less likely to be diabetic whereas the likelihood of being diabetic is greater at higher BMI values. For BMI $\geq$ 25 kg/m^2, the proportion of diabetic population is considerably higher than the non-diabetic population indicating a positive association between diabetes, and overweight and obesity.

The BMI distribution for diabetic population has a mean BMI of 24.98 kg/m^2 and the proportion of overweight or obese population is 46.27%. The BMI distribution for non-diabetic population has a mean BMI of 21.66 kg/m^2 and the proportion of overweight or obese population is 17.82%. This suggests that the likelihood of being diabetic is considerably greater for overweight or obese individuals as compared to the non-overweight individuals.

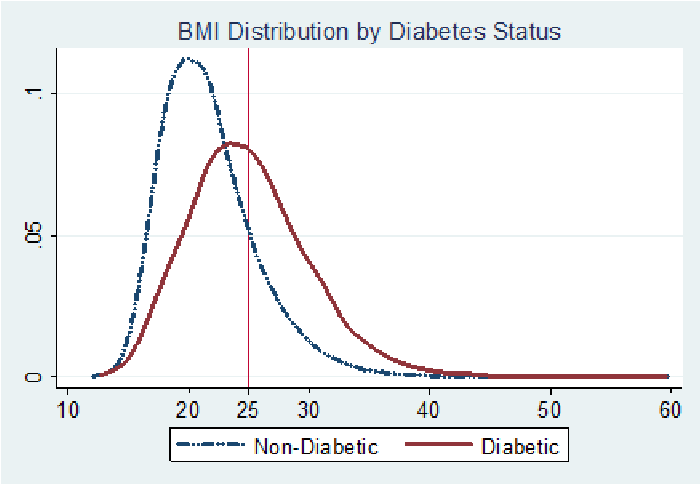


Figure 5.1: BMI Distribution by Self-Reported Diabetes Status

Source: Figure constructed by author based on NFHS data for year 2015-16.

We also analyse the relationship between the blood glucose levels and BMI graphically. Figure 5.2 illustrates the BMI distribution by blood glucose levels for the full sample. For plotting these distributions, we, first, categorise the BMI data for the total sample on the basis of blood glucose levels of the individuals (as per categories defined in Section 5.2) and then we plot three separate BMI distributions for each blood glucose category. In Figure 5.2, the solid green line represents BMI distribution for the diabetic population (having blood glucose levels more than 200 mg/dl), dashed red line represents the BMI distribution for the prediabetic population (having blood glucose levels between 141 and 200 mg/dl) and dash-dotted blue line represents the BMI distribution for the population having normal blood glucose levels (having blood glucose levels less than or equal to 140 mg/dl). It can be observed that the BMI distribution for diabetic population lies to the extreme right and indicates that the mass of the population having very high blood glucose levels is substantially greater among higher BMI values (BMI $\geq$ 25 kg/m^2) while the mass of population having normal blood glucose levels is higher among low BMI values (BMI < 25 kg/m^2). We may infer that amongst the population having BMI $\geq$ 25 kg/m^2, the likelihood of being diabetic is highest followed by prediabetes while the likelihood of having normal blood glucose levels is the least.[47]

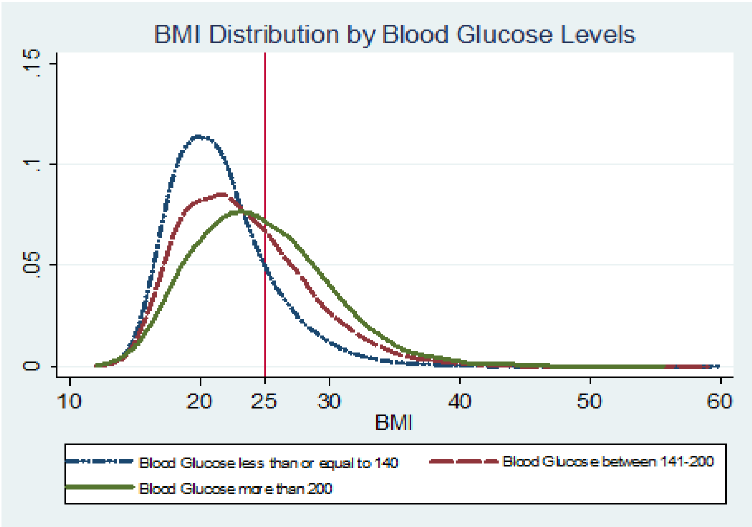


Figure 5.2: BMI Distribution by Blood Glucose Levels

Source: Figure constructed by author based on NFHS data for year 2015-16.
Note: Blood glucose levels are measured in mg/dl.

[47] The mean BMI for diabetic category is 24.71 kg/m^2 and the proportion of overweight or obese population is 44.46%. The mean BMI for prediabetic category is 23.37 kg/m^2 and the proportion of overweight or obese population is 32.95%. And the mean BMI for normal blood glucose category is 21.57 kg/m^2 and the proportion of overweight or obese population is 16.93%.

Table 5.1: List of Variables with Definition and Type

Variable	Definition	Type
Health Outcome Variables:		
Ordinal Blood Glucose Levels	• BG = 0 if blood glucose level is less than or equal to140 mg/dl • BG = 1 if blood glucose level is between 141 and 200 mg/dl • BG = 2 if blood glucose level is higher than 200 mg/dl	Ordinal
Self-Reported Diabetes Status	• D = 0 if individual is non-diabetic • D = 1 if individual is diabetic	Binary
List of Independent Variables:		
Individual Characteristics:		
Body Mass Index	Person's weight is kilograms divided by square of his/her height in meters (in kg/m^2).	Continuous
Age	Age of the individual (in years).	Continuous
Gender	• = 0 if Male@ • = 1 if Female	Binary
Education	• = 0 if no education or preschool@ • = 1 if Primary • = 2 if Secondary • = 3 if Higher	Ordinal
Marital Status	• = 0 if Never married@ [48] • = 1 if Married	Binary
Bank Account	• = 0 if individual does not have bank account@ • = 1 if individual has bank account	Binary
Time since last ate	Time since last ate (in hours). Time is recorded before blood glucose measurements are taken.	Continuous
Time since last drank	Time since last drank (in hours), something other than plain water. Time is recorded before blood glucose measurements are taken.	Continuous
Behavioural Risk Factors[49]	• = 1 if smokes cigarette, 0 otherwise@ • = 1 if smokes pipe, 0 otherwise@ • = 1 if chews tobacco, 0 otherwise@ • = 1 if snuffs, 0 otherwise@ • = 1 if smokes cigar, 0 otherwise@ • = 1 if chews paan or gutkha, 0 otherwise@ • = 1 if chews paan with tobacco, 0 otherwise@ • = 1 if drinks alcohol, 0 otherwise@	Binary

[48] Includes married but gauna not done.
[49] Contains a set of eight dummy variables.

Table 5.1 (Continued)		
Variable	**Definition**	**Type**
Eating Habits[50]	• = 1 if eats fried food daily or weekly, 0 otherwise@ • = 1 if drinks aerated drink daily or weekly, 0 otherwise@	Binary
Household Characteristics:		
Wealth Quintile	• = 0 if poorest@ • = 1 if poorer • = 2 if middle • = 3 if richer • = 4 if richest	Ordinal
Religion	• = 0 if Hindu@ • = 1 if Muslim • = 2 if Christian • = 3 if Sikh • = 4 if Buddhist/neo-Buddhist • = 5 if Jain • = 6 if Jewish • = 7 if Parsi/Zoroastrian • = 8 if no religion • = 9 if some other religion	Ordinal
Caste or Tribe[51]	• = 1 if Scheduled Caste, 0 otherwise@ • = 1 if Scheduled Tribe, 0 otherwise@ • = 1 if Other Backward Classes, 0 otherwise@	Binary
Insurance	• = 0 if any usual member of household is not covered by a health scheme or health insurance@ • = 1 if any usual member of household is covered by a health scheme or health insurance	Binary
Below Poverty Line	• = 0 if household does not have BPL card@ • = 1 if household has BPL card	Binary
Family Structure	• = 0 if nuclear family@ • = 1 if non-nuclear or joint family	Binary
Number of Household Members	Number of total household members in all age groups.	Continuous
Region	• = 0 if Rural@ • = 1 if Urban	Binary

@ Indicates the base category.

[50] Contains a set of two dummy variables.
[51] Contains a set of three dummy variables.

Table 5.2: Descriptive Statistics

Variables	Observations	Mean	Standard Deviation	Minimum Value	Maximum Value
Individual Characteristics					
Self-Reported Diabetes Status	784042	0.015	0.120	0	1
Ordinal Blood Glucose levels	806905	0.070	0.294	0	2
Blood Glucose levels – Actual Values (in mg/dl)	806905	104.689	29.602	20	499
Body Mass Index (in kg/m^2)	811465	21.714	4.094	12.01	59.96
Age (in years)	811465	30.066	9.967	15	54
Gender	811465	0.863	0.344	0	1
Education	809904	1.483	0.994	0	3
Married	782387	0.732	0.443	0	1
Bank Account	810731	0.914	0.281	0	1
Time since last ate (in hours)	805589	3.132	3.543	0	48
Time since last drank (in hours)	800779	5.384	14.049	0	95
Behavioural Risk Factors					
Smokes Cigarette	795856	0.024	0.154	0	1
Smokes Pipe	795856	0.001	0.025	0	1
Chews Tobacco	795856	0.012	0.108	0	1
Snuffs	795856	0.001	0.034	0	1
Smokes Cigar	795856	0.001	0.037	0	1
Chews Paan or Gutkha	795856	0.049	0.216	0	1
Chews Paan with Tobacco	795856	0.043	0.204	0	1
Consumes Alcohol	795856	0.065	0.246	0	1
Eating Habits					
Fried Food	795856	0.455	0.498	0	1
Aerated Drinks	795856	0.242	0.429	0	1
Household Characteristics					
Wealth Quintile	811465	1.983	1.384	0	4
Religion	811465	0.520	1.26	0	9
Scheduled Caste	811465	0.181	0.385	0	1
Scheduled Tribe	811465	0.182	0.386	0	1
Other Backward Classes	811465	0.387	0.487	0	1
Insurance	806832	0.262	0.440	0	1
Below Poverty Line	810055	0.386	0.487	0	1
Family Structure	811465	0.503	0.500	0	1
Number of Household Members	811465	5.772	2.651	1	41
Region	811465	0.292	0.455	0	1

Note: Values are based on full sample.

Table 5.3: Descriptive Statistics by Overweight and Obesity Status

Variables	Overweight or Obese		Non-Overweight		Difference#
	Mean	Standard Deviation	Mean	Standard Deviation	(t-statistic)
Individual Characteristics					
Self-Reported Diabetes Status	0.037	0.189	0.010	0.097	0.027*** (78.578)
Ordinal Blood Glucose levels	0.154	0.443	0.051	0.245	0.103*** (1.2e+02)
Blood Glucose levels – Actual Values (in mg/dl)	113.571	42.203	102.723	25.586	10.848*** (1.3e+02)
Body Mass Index (in kg/m^2)	28.317	3.323	20.249	2.493	8.068*** (1.1e+03)
Age (in years)	35.282	8.705	28.910	9.859	6.372*** (2.3e+02)
Gender	0.867	0.340	0.862	0.345	0.005*** (4.944)
Education	1.625	0.983	1.451	0.994	0.174*** (60.889)
Married	0.897	0.304	0.695	0.460	0.202*** (1.6e+02)
Bank Account	0.944	0.230	0.907	0.290	0.037*** (45.656)
Time since last ate (in hours)	3.104	3.620	3.138	3.526	-0.034*** (-3.335)
Time since last drank (in hours)	4.031	10.141	5.685	14.761	-1.654*** (-40.679)
Behavioural Risk Factors					
Smokes cigarette	0.025	0.156	0.024	0.153	0.001** (2.536)
Smokes pipe	0.0005	0.022	0.001	0.025	-0.0002** (-2.392)
Chews Tobacco	0.010	0.099	0.012	0.110	-0.002*** (-7.352)
Snuffs	0.001	0.033	0.001	0.034	-0.000 (-0.345)
Smokes cigar	0.001	0.036	0.001	0.037	-0.000 (-0.427)
Chews paan or gutkha	0.039	0.192	0.051	0.221	-0.013*** (-20.544)
Chews paan with Tobacco	0.045	0.207	0.043	0.203	0.002*** (3.325)
Alcohol	0.062	0.241	0.065	0.247	-0.003*** (-4.370)
Eating Habits					
Fried Food	0.472	0.499	0.451	0.498	0.021*** (14.535)
Aerated Drinks	0.280	0.449	0.234	0.423	0.046*** (37.081)

Table 5.3 (Continued)

Variables	Overweight or Obese		Non-Overweight		Difference#
	Mean	Standard Deviation	Mean	Standard Deviation	(t-statistic)
Household Characteristics					
Wealth Quintile	2.745	1.207	1.814	1.363	0.930*** (2.4e+02)
Religion	0.590	1.271	0.505	1.258	0.086*** (23.631)
Scheduled Caste	0.151	0.358	0.187	0.390	-0.036*** (-32.613)
Scheduled Tribe	0.120	0.324	0.196	0.397	-0.076*** (-68.823)
Other Backward Classes	0.390	0.488	0.387	0.487	0.003** (2.353)
Insurance	0.278	0.448	0.258	0.438	0.020*** (15.830)
Below Poverty Line	0.286	0.452	0.408	0.491	-0.122*** (-87.177)
Family Structure	0.499	0.500	0.504	0.500	-0.006*** (-4.089)
Number of Household Members	5.531	2.696	5.825	2.638	-0.294*** (-38.524)
Region	0.452	0.498	0.257	0.437	0.195*** (1.5e+02)

*** and ** indicates significance at 1% and 5% significance level.

Difference = mean(Overweight or Obese) - mean(Non-Overweight). A positive value indicates that the mean is higher for overweight or obese population while a negative value indicates that the mean is higher for non-overweight population. The t-statistic is obtained from two-sample mean-comparison test with equal variances.

Table 5.4: Proportion of Individuals across Different Categories for Selected Binary and Ordinal Variables based on Overweight and Obesity Status

Variables	Proportion of Individuals (in %)		
	Full Sample	Non-Overweight Individuals Sub-Sample #	Overweight or Obese Individuals Sub-Sample#
Ordinal Blood Glucose:			
Normal Blood Glucose	94.11	95.47	87.93
Prediabetic	4.82	3.95	8.77
Diabetic	1.07	0.58	3.30
Self-Reported Diabetes Status:			
Non-Diabetic	98.54	99.04	96.30
Diabetic	1.46	0.96	3.70
Gender:			
Male	13.74	13.83	13.34
Female	86.26	86.17	86.66
Region:			
Rural	70.77	74.32	54.80
Urban	29.23	25.68	45.20
Wealth Quintile:			
Poorest	18.82	21.77	5.51
Poorer	21.40	23.47	12.02
Middle	21.17	21.47	19.80
Richer	19.88	18.11	27.84
Richest	18.74	15.17	34.83

As per WHO International BMI classification using a BMI cut-off of 25 kg/m^2.

Table 5.5: Descriptive Statistics for Sub-Sample of Married Couples

Variables	Observations	Mean	Standard Deviation	Minimum Value	Maximum Value
Individual Characteristics					
Self-Reported Diabetes Status	43664	0.029	0.167	0	1
Ordinal Blood Glucose Levels	44986	0.144	0.422	0	2
Blood Glucose Levels - Actual Values (in mg/dl)	44986	112.519	40.807	20	499
Body Mass Index (in kg/m^2)	45205	22.470	3.826	12.32	59.8
Age (in years)	45205	39.676	8.013	15	54
Gender	45205	0.010	0.101	0	1
Education	45039	1.497	0.939	0	3
Married	45184	0.998	0.043	0	1
Bank Account	45156	0.905	0.293	0	1
Time since last ate (in hours)	44926	3.164	3.526	0	48
Time since last drank (in hours)	44743	4.817	12.866	0	95
Behavioural Risk Factors					
Smokes Cigarette	44255	0.162	0.369	0	1
Smokes Pipe	44255	0.004	0.063	0	1
Chews Tobacco	44255	0.038	0.191	0	1
Snuffs	44255	0.001	0.037	0	1
Smokes Cigar	44255	0.006	0.076	0	1
Chews Paan or Gutkha	44255	0.171	0.376	0	1
Chews Paan with Tobacco	44255	0.090	0.286	0	1
Alcohol	44255	0.392	0.488	0	1
Eating Habits					
Fried Food	44255	0.440	0.496	0	1
Aerated Drinks	44255	0.267	0.443	0	1
Household Characteristics					
Wealth Quintile	45205	1.877	1.369	0	4
Religion	45205	0.518	1.275	0	9
Scheduled Caste	45205	0.187	0.390	0	1
Scheduled Tribe	45205	0.198	0.398	0	1
Other Backward Classes	45205	0.383	0.486	0	1
Insurance	44982	0.274	0.446	0	1
Below Poverty Line	45123	0.375	0.484	0	1
Family Structure	45205	0.303	0.460	0	1
Number of Household Members	45205	4.898	1.802	2	24
Region	45205	0.297	0.457	0	1

Note: Values are based on restricted sample comprising of married couples living together in the same household of whom either is head of the household.

5.4 Estimation Results and Interpretation

In this section, we first present the estimation results for the Probit and IV-Probit models using self-reported diabetes status as the outcome variable. We then present the results obtained from the estimation of an Ordered Probit model in which the ordinally defined blood glucose levels of the individuals is the outcome of interest.

5.4.1 Effect of Body Mass Index on the Self-Reported Diabetes Status: Probit and IV-Probit Model Estimates

This sub-section contains results pertaining to the outcome variable self-reported diabetes status. For IV-Probit model, we use BMI of the spouse as an instrument, therefore, our sample gets restricted to only married couples living in the same households either of whom is head of the family. For the sake of comparison, we also report the results of the Probit model using the restricted sample data.

Table 5.6 presents the average marginal effects of BMI on the self-reported diabetes status for the sample data which is restricted to married couples. We estimate the average marginal effects for overweight or obese and non-overweight individuals. The coefficient estimates obtained from the estimation of the Probit model are provided in column (1) of Table A.5.2 of Appendix and for the IV-Probit model these results are provided in columns (1) and (2) of Table A.5.3 of Appendix. Based on the estimated Probit model, we compute the average marginal effects of BMI on the self-reported diabetes status across overweight or obese, and non-overweight individuals, as reported in Table 5.6. These marginal effects are reported for two classifications: for WHO International BMI classification in column (1) and, for WHO Asian BMI classification in column (2) of Table 5.6. Within the each column the average marginal effects of BMI, i.e., the change in probability of being diabetic due to a unit rise in BMI $\left(\frac{\partial P(D=1|\,X)}{\partial BMI}\right)$ is reported for overweight or obese individuals and non-overweight individuals along with the difference between the marginal effects across these two categories. Similarly, columns (3) – (4) report the results obtained from the IV-Probit model.

Both the models include same set of controls so that the marginal effects can be compared across them. We control for the demographic and socio-economic variables for individual and household characteristics, behavioural risk factors, eating habits and the state fixed effects. Wald chi2 test statistic along with the p-value is reported for both Probit and IV-Probit models. For IV-Probit model, we use Wald test of exogeneity to check endogeneity of BMI. The null hypothesis of this test states that there is no endogeneity. A rejection of null hypothesis indicates that BMI is endogenous. A non-rejection indicates that corresponding Probit model is appropriate. We also report R^2 and F statistic for the first stage regression of the IV-Probit model as an approximate guide for the quality of our instrument. All the estimates are found to be robust to the inclusion or exclusion of controls. In Table 5.6, we report the results from regressions that include all the control variables.

Comparing Probit and IV-Probit models in each column, we find that marginal effects of BMI on self-reported diabetes status for IV-Probit model are substantially higher than those for the corresponding Probit model indicating that Probit model estimates highly underestimate the casual effect of BMI on diabetes.

Comparing the marginal effects across overweight or obese individuals and non-overweight individuals in columns (1) and (2), based on Probit model, we find that the increase in the probability of being diabetic due to a unit rise in BMI is twice among overweight or obese individuals as compared to the non-overweight individuals. However, comparing the marginal effects across overweight or obese individuals and non-overweight individuals in columns (3) and (4), based on IV-Probit model, we find that the increase in the probability of being diabetic due to a unit rise in BMI is thrice among overweight or obese individuals as compared to the non-overweight individuals. The marginal effect of BMI on the self-reported diabetes status for non-overweight individuals is 0.5% and for the overweight or obese individuals it is 1.5% for the IV-Probit model, while the same figures for Probit model are 0.16% and 0.3%, respectively, based on WHO International BMI classification. We find that the marginal effects of BMI on the self-reported diabetes status differ statistically significantly across non-overweight, and overweight or obese individuals.

For IV-Probit model, the Wald test of exogeneity is rejected at 1% significance level indicating that BMI is endogenous. Also, F-statistic for the corresponding first-stage regression is found to be much higher than the conventional minimum value of 10 and R^2 also takes a reasonably high value.

We also estimated the Probit model for the full sample data (provided in column (3) of Table A.5.2 of Appendix). Based on the estimated Probit model, we compute the average marginal effects of BMI on the self-reported diabetes status. These results are reported in Table 5.7. The presentation of results in done in similar fashion as explained for Table 5.6. Comparing the marginal effects across overweight or obese individuals and non-overweight individuals in columns (1) and (2), we find that the increase in the probability of being diabetic due to a unit rise in BMI is almost thrice among overweight or obese individuals as compared to the non-overweight individuals. In column (1), the marginal effect of BMI on the self-reported diabetes status for non-overweight individuals is 0.08% and for the overweight or obese individuals it is 0.23%. Similar results are obtained by applying WHO Asian BMI classification, in column (2).

Having shown that the overweight and obese individuals are at a higher risk of diabetes, we next examine that within the overweight or obese individuals, which sections of the population are at a greater risk of diabetes. For the purpose, we consider a sub-sample comprising of overweight or obese individuals (having BMI $\geq$ 25 kg/m^2), which is about 18% of the total sample, and examine if the marginal effects of an increase in BMI on the likelihood of being diabetic differ across genders – male and female, regions – urban and rural, and wealth quintiles – poorest and richest. Table 5.8 presents the average marginal effects obtained from the estimation of the Probit model (provided in column (2) of Table A.5.2 of Appendix) and IV-Probit model (provided in columns (3) and (4) of Table A.5.3 of Appendix), based on the restricted sample. The marginal effect of BMI on the self-reported diabetes status is higher among males as compared to that for females in both specifications. However, these results do not differ statistically significantly. In both the models, the urban population is about 1.3 times more likely to be diabetic than the rural population. Also, the individuals from richest wealth quintile are 3 times more likely to be diabetic as compared to the poorest wealth quintile in both the models (2.6 times $\approx$ 3 times in IV-Probit model). The marginal effects across regions and wealth quintiles differ statistically significantly.

Table 5.6: Average Marginal Effects of BMI on the Self-Reported Diabetes Status: Probit and IV-Probit Model Estimates based on the Restricted (or Married Couples) Sub-Sample

	Probit Model		IV-Probit Model	
Average Marginal Effects	**WHO International BMI Classification**	**WHO Asian BMI Classification**	**WHO International BMI Classification**	**WHO Asian BMI Classification**
	(1)	**(2)**	**(3)**	**(4)**
Overweight or Obese Individuals	0.0032*** (0.0004)	0.0028*** (0.0003)	0.0148*** (0.0038)	0.0115*** (0.0028)
Non-Overweight Individuals	0.0016*** (0.0001)	0.0014*** (0.0001)	0.0046*** (0.0008)	0.0036*** (0.0005)
Difference#	0.0016*** (0.0002)	0.0014*** (0.0002)	0.0101*** (0.0030)	0.0079*** (0.0023)
Controls	Yes		Yes	
State Fixed Effects	Yes		Yes	
Observations	43202		43202	
Wald chi2	1010.93		234600.29	
P-Value	0.0000		0.0000	
Pseudo R^2	0.1072			
Wald test of exogeneity, chi2			18.73	
P-Value			0.0000	
First Stage				
F – statistic			153.28	
R^2			0.2038	

*** denotes significance at 1% level.
Delta-Method standard errors are reported in parentheses. "The delta method is used to estimate the standard errors of a non-linear function of model parameters (such as ordered probit, probit or IV-probit models). The delta method finds a linear approximation of the non-linear function to calculate the variance" (Corneliben, 2005).
Difference is ME(Overweight and Obese) – ME(Non-Overweight) where ME denotes average marginal effects.
Note: Probit and IV-Probit models do not include marital status as a control. Marital status is omitted in the restricted sample as the sample comprises of only married individuals.
Controls include individual and household characteristics, behavioural risk factors and eating habits.
Individual and household characteristics include age, gender, education, bank account, household characteristics such as wealth quintile, religion, caste, insurance, below poverty line, family structure, number of household members and region.
Behavioural risk factors include smoking cigarette, smoking pipe, chewing tobacco, snuffing, smoking cigar, chewing paan or gutkha, chewing paan with tobacco and drinking alcohol.
Eating habits include daily or weekly consumption of fried foods and aerated drinks.

Table 5.7: Average Marginal Effects of BMI on the Self-Reported Diabetes Status: Probit Model Estimates based on Full Sample Data

	Probit Model	
Average Marginal Effects	**WHO International BMI Classification**	**WHO Asian BMI Classification**
	(1)	**(2)**
Overweight or Obese Individuals	0.0023*** (0.00008)	0.0019*** (0.00006)
Non-Overweight Individuals	0.0008*** (0.00002)	0.0007*** (0.00002)
Difference[#]	0.0015*** (0.00006)	0.0013*** (0.00005)
Controls	Yes	
State Fixed Effects	Yes	
Observations	776394	
Wald chi2	10987.47	
P-Value	0.0000	
Pseudo R^2	0.1256	

*** denotes significance at 1% level.
Delta-Method standard errors are reported in parentheses.
[#] Difference is ME(Overweight and Obese) – ME(Non-Overweight).
Controls include individual and household characteristics, behavioural risk factors and eating habits.
Individual and household characteristics include age, gender, education, bank account, household characteristics such as wealth quintile, religion, caste, insurance, below poverty line, family structure, number of household members and region.
Behavioural risk factors include smoking cigarette, smoking pipe, chewing tobacco, snuffing, smoking cigar, chewing paan or gutkha, chewing paan with tobacco and drinking alcohol.
Eating habits include daily or weekly consumption of fried foods and aerated drinks.

Table 5.8: Average Marginal Effects of BMI on the Self-Reported Diabetes Status amongst Overweight or Obese Individuals (BMI ≥ 25 kg/m^2): Probit and IV-Probit Model Estimates based on the Restricted (or Married Couples) Sub-Sample

	Probit Model			IV-Probit Model		
	Gender	Region	Wealth Quintile	Gender	Region	Wealth Quintile
	(1)	(2)	(3)	(4)	(5)	(6)
	Male	**Urban**	**Richest**	**Male**	**Urban**	**Richest**
Average Marginal Effects	0.0026*** (0.0006)	0.0031*** (0.0007)	0.0035*** (0.0008)	0.0175* (0.0093)	0.0202** (0.0103)	0.0221** (0.0108)
	Female	**Rural**	**Poorest**	**Female**	**Rural**	**Poorest**
Average Marginal Effects	0.0019** (0.0009)	0.0022*** (0.0005)	0.0011*** (0.0003)	0.0133 (0.0098)	0.0152* (0.0085)	0.0086 (0.0058)
	Difference#	**Difference#**	**Difference#**	**Difference#**	**Difference#**	**Difference#**
	0.0007 (0.0009)	0.0009*** (0.0003)	0.0024*** (0.0006)	0.0041 (0.0053)	0.0049** (0.0020)	0.0135** (0.0055)
Controls		Yes			Yes	
State Fixed Effects		Yes			Yes	
Observations		9622			9711	
Wald chi2		394.56			106298.91	
P-Value		0.0000			0.0000	
Pseudo R^2		0.1039				
Wald test of exogeneity, chi2					3.69	
P-Value					0.0547	

***, ** and * denote significance at 1%, 5% and 10% level.
Delta-Method standard errors are reported in parentheses.
(1) Difference is ME(Male) – ME(Female); (2) Difference is ME(Urban) – ME(Rural) and (3) Difference is ME(Richest) – ME(Poorest).
Note: Probit and IV-Probit models do not include marital status as a control. Marital status is omitted in the restricted sample as the sample comprises of only married individuals.
Controls include individual and household characteristics, behavioural risk factors and eating habits.
Individual and household characteristics include age, gender, education, bank account, household characteristics such as wealth quintile, religion, caste, insurance, below poverty line, family structure, number of household members and region.
Behavioural risk factors include smoking cigarette, smoking pipe, chewing tobacco, snuffing, smoking cigar, chewing paan or gutkha, chewing paan with tobacco and drinking alcohol.
Eating habits include daily or weekly consumption of fried foods and aerated drinks.

5.4.2 Effect of Body Mass Index on the Ordinal Blood Glucose Levels: Ordered Probit Model Estimates

Table 5.9 presents the average marginal effects of BMI on the ordinal blood glucose levels based on the Ordered Probit model estimates (provided in column (1) of Table A.5.4 of Appendix). For the estimated Ordered Probit model, we compute the average marginal effects of BMI on the ordinal blood glucose levels across overweight or obese and non-overweight individuals. The marginal effects based on WHO International BMI classification are reported in columns (1) – (3) of Table 5.9. We report the marginal effect, i.e., the change in probability of belonging to a specific blood glucose category due to a unit rise in BMI $\left(\frac{\partial P(BG=j|\ X)}{\partial BMI};\ \ j = 0, 1, 2\right)$ for the three blood glucose categories. Each column, first, reports these marginal effects for the overweight or obese individuals and then for the non-overweight individuals along with the difference in the marginal effects across the two categories. Similarly, columns (4) – (6) report the results using WHO Asian BMI classification.

In addition to controlling for the demographic and socio-economic variables for individual and household characteristics, behavioural risk factors, eating habits and the state fixed effects, we also control for the time since the individual last ate and drank (in hours) since these variables are expected to influence individual's blood glucose levels (Moebus et al., 2011). In the estimated model, the threshold coefficients, μ_1 and μ_2 , are found to be positive, and $\mu_1 < \mu_2$ (the values of threshold coefficients are provided in the Appendix with the respective table). All the estimates are found to be robust to the inclusion or exclusion of controls. In Table 5.9, we report the results from regression that includes all the control variables.

Comparing marginal effects across overweight or obese and non-overweight individuals reported in column (2), we find that the increase in the probability of being prediabetic due to a unit rise in BMI is almost twice among overweight or obese individuals as compared to the non-overweight individuals. The marginal effect of BMI on prediabetes for non-overweight individuals is 0.27% and for the overweight or obese individuals it is 0.48%. In column (3), the marginal effect of BMI on diabetes is 0.07% among non-overweight individuals and 0.2% among overweight or obese individuals. Here, it can be inferred that the increase in probability of being diabetic due to a unit rise in BMI is about three times among overweight or obese individuals as compared to the non-overweight individuals. Similar results are obtained by applying WHO Asian BMI classification, in columns (5) and (6). Also, the differences in the marginal effects are highly statistically significant.

We now examine that within the overweight or obese individuals, which sections of the population are at a greater risk of prediabetes and diabetes. For this purpose, we consider a sub-sample comprising of overweight or obese individuals (having BMI ≥ 25 kg/m^2), and examine if the marginal effects of an increase in BMI on the likelihood of being prediabetic or diabetic differ across genders – male and female, regions – urban and rural, and wealth quintiles – poorest and richest. Table 5.10 presents these results obtained from the Ordered Probit model estimates (provided in column (2) of Table A.5.4 of Appendix).[52] The results show that the males are at a slightly higher risk of being both prediabetic (0.5%) and diabetic (0.3%) as compared to the females (0.4% and 0.2% respectively). Also, the marginal effects for prediabetes and diabetes are slightly higher in the urban regions as compared to the rural regions. For the wealth quintiles, the individuals from the richest wealth quintile are 1.5 times more likely to be diabetic, and 1.2 times more likely to be prediabetic as compared to the poorest wealth quintile. The marginal effects across genders, regions and wealth quintiles differ statistically significantly.

[52] The results in Table 5.10 have been presented differently. The average marginal effects of BMI for a specific blood glucose category are presented across rows.

Table 5.9: Average Marginal Effects of BMI on the Ordinal Blood Glucose Levels: Ordered Probit Model Estimates based on Full Sample Data

	Ordered Probit Model					
	WHO International BMI Classification			WHO Asian BMI Classification		
Average Marginal Effects	**Blood Glucose ≤ 140 (BG = 0)**	**141 ≤ Blood Glucose ≤ 200 (BG = 1)**	**Blood Glucose > 200 (BG = 2)**	**Blood Glucose ≤ 140 (BG = 0)**	**141 ≤ Blood Glucose ≤ 200 (BG = 1)**	**Blood Glucose > 200 (BG = 2)**
	Normal Blood Glucose	**Prediabetes**	**Diabetes**	**Normal Blood Glucose**	**Prediabetes**	**Diabetes**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Overweight or Obese Individuals	-0.0068*** (0.0001)	0.0048*** (0.00009)	0.0020*** (0.00005)	-0.0061*** (0.0001)	0.0044*** (0.00008)	0.0017*** (0.00004)
Non-Overweight Individuals	-0.0035*** (0.00005)	0.0027*** (0.00004)	0.0007*** (0.00001)	-0.0031*** (0.00004)	0.0025*** (0.00004)	0.0006*** (0.00001)
Difference#	-0.0034*** (0.00008)	0.0021*** (0.00005)	0.0013*** (0.00003)	-0.0029*** (0.00007)	0.0019*** (0.00004)	0.0011*** (0.00003)
Controls	Yes					
State Fixed Effects	Yes					
Observations	748,995					
Wald chi2	26968.90					
P-Value	0.0000					
Pseudo R^2	0.0901					

*** denotes significance at 1% level.
Delta-Method standard errors are reported in parentheses.
Difference is ME(Overweight and Obese) – ME(Non-Overweight).
Controls include individual and household characteristics, behavioural risk factors and eating habits.
Individual and household characteristics include age, gender, education, marital status, bank account, household characteristics such as wealth quintile, religion, caste, insurance, below poverty line, family structure, number of household members, region and time since last ate and drank.
Behavioural risk factors include smoking cigarette, smoking pipe, chewing tobacco, snuffing, smoking cigar, chewing paan or gutkha, chewing paan with tobacco and drinking alcohol.
Eating habits include daily or weekly consumption of fried foods and aerated drinks.

Table 5.10: Average Marginal Effects of BMI on the Ordinal Blood Glucose Levels amongst Overweight or Obese Individuals (BMI ≥ 25 kg/m^2): Ordered Probit Model Estimates based on Full Sample Data

	Ordered Probit Model		
	Gender		
	Male	**Female**	**Difference#**
Normal Blood Glucose (Blood Glucose ≤ 140)	-0.0078*** (0.0003)	-0.0064*** (0.0002)	-0.0014*** (0.0001)
Prediabetes (141 ≤ Blood Glucose ≤ 200)	0.0046*** (0.0002)	0.0041*** (0.0001)	0.0005*** (0.00004)
Diabetes (Blood Glucose > 200)	0.0032*** (0.0001)	0.0023*** (0.00008)	0.0009*** (0.00007)
	Region		
	Urban	**Rural**	**Difference#**
Normal Blood Glucose (Blood Glucose ≤ 140)	-0.0070*** (0.0002)	-0.0062*** (0.0002)	-0.0008*** (0.00007)
Prediabetes (141 ≤ Blood Glucose ≤ 200)	0.0043*** (0.0001)	0.0040*** (0.0001)	0.0003*** (0.00003)
Diabetes (Blood Glucose > 200)	0.0026*** (0.0001)	0.0022*** (0.00008)	0.0005*** (0.00004)
	Wealth Quintile		
	Richest	**Poorest**	**Difference#**
Normal Blood Glucose (Blood Glucose ≤ 140)	-0.0070*** (0.0002)	-0.0054*** (0.0002)	-0.0016*** (0.0002)
Prediabetes (141 ≤ Blood Glucose ≤ 200)	0.0043*** (0.0001)	0.0036*** (0.0001)	0.0007*** (0.00008)
Diabetes (Blood Glucose > 200)	0.0026*** (0.0001)	0.0017*** (0.00009)	0.0009*** (0.00008)
Controls	Yes		
State Fixed Effects	Yes		
Observations	135,630		
Wald chi2	7482.14		
P-Value	0.0000		
Pseudo R^2	0.0704		

*** denotes significance at 1% level.
Delta-Method standard errors are reported in parentheses.
(1) Difference is ME(Male) – ME(Female); (2) Difference is ME(Urban) – ME(Rural) and (3) Difference is ME(Richest) – ME(Poorest).
Controls include individual and household characteristics, behavioural risk factors and eating habits.
Individual and household characteristics include age, gender, education, marital status, bank account, household characteristics such as wealth quintile, religion, caste, insurance, below poverty line, family structure, number of household members, region and time since last ate and drank.
Behavioural risk factors include smoking cigarette, smoking pipe, chewing tobacco, snuffing, smoking cigar, chewing paan or gutkha, chewing paan with tobacco and drinking alcohol.
Eating habits include daily or weekly consumption of fried foods and aerated drinks.

5.5 Discussion

In this chapter, we find that the overweight or obese individuals are more likely to be diabetic as well as prediabetic as compared to the non-overweight individuals. These results are line with the studies by Sepp et al. (2014) and Huffman et al. (2011) which show that a rise in BMI is positively associated with the blood glucose levels and diabetes. This study contributes to the existing literature by quantifying the impact of overnutrition on both prediabetes and diabetes in India. The results obtained from the study are consistent across both WHO International and Asian BMI classifications for defining overweight and obesity status of the population.

The marginal effects obtained from the estimation of a Probit model for the full sample data with self-reported diabetes status as the outcome variable and the marginal effects obtained from the estimation of an Ordered Probit model defining diabetes status based on the blood glucose levels (above 200 mg/dl) are qualitatively similar indicating that our results are consistent across both indicators used for measuring diabetes status of the individuals.

The change in probability of being diabetic or prediabetic with an additional unit gain in BMI is positive for non-overweight individuals as well. This suggests that weight gain increases the risk of diabetes for non-overweight individuals as well. However, the level of risk is expected to vary with weight or BMI of an individual. This finding is line with the study by Colditz et al. (1995) which states that the risk of diabetes is faced by population at all levels of BMI.

Based on Probit model estimates, we plot the average marginal effects of BMI on the self-reported diabetes status for the full sample data to examine how does the marginal effects vary across different subgroups with age and BMI. Figure A.5.1 of Appendix, illustrates the graphical plot of the average marginal effects of BMI on the self-reported diabetes status. We plot the average marginal effects of BMI at different values of age and BMI, and compare it across different subgroups – overweight or obese and non-overweight; male and female; and rural and urban. We find that the average marginal effects of BMI on the self-reported diabetes status are considerably higher among overweight or obese individuals as compared to the non-overweight individuals. We do not witness any considerable difference in these marginal effects across genders. Also, the average marginal effects are higher for the urban population as compared to the rural population. One important result is that the average marginal effect of BMI on the self-reported diabetes status increases with both age and BMI across all subgroups.

5.6 Conclusion

Recognising the recently growing problem of overnutrition and diabetes in India, the present chapter quantifies the causal effect of overweight and obesity on diabetes in India. The novel contribution of the study is that it addresses the potential endogeneity problem resulting from the omitted variable bias while estimating the effect of BMI on diabetes. We examine the change in the likelihood of being diabetic and prediabetic with a rise in BMI across different subgroups of the population. Considering two different health outcome variables – self-reported diabetes status and ordinal blood glucose levels, we find that the marginal effect of BMI on diabetes is positive and statistically significant. Also, these effects are found to be much higher for the overweight or obese individuals as compared to the non-overweight individuals. However, the magnitude of the marginal effect of BMI on diabetes differ across different model specifications – Ordered Probit model, Probit model and IV-Probit model. It is found that Ordered Probit and Probit model estimates highly understate the causal impact of the rise in BMI on diabetes. To the best of our knowledge, this is the first study that addresses unobserved heterogeneity in the relationship between BMI and diabetes using an instrumental variable approach in the Indian context.

Heterogeneity analysis across different subgroups of the population suggests that among the overweight and obese individuals, males, population living in the urban areas and population belonging to the richest wealth quintile face a higher risk of being diabetic and prediabetic as compared to females, population living in the rural areas and population belonging to the poorest wealth quintile respectively.

Our findings have significant implications for the policy formulation as diabetes has a substantial health and economic burden associated with it. Diabetes elevates the risk of having other NCDs such as cardiovascular diseases, strokes, etc., thereby further aggravating the health burden. The economic burden associated with diabetes is large given its substantially high health care costs. The cost burden associated with diabetes may have severe adverse impact on the households as more than 60% of the total healthcare costs or expenditures are directly financed by these households in the form of out of pocket health expenditures. It is also important to note that diabetes is not only restricted to urban areas but is also prevalent among rural areas and is no longer a disease of the rich. Diabetes among poor households may have catastrophic implications and lead to extreme impoverishment. Therefore, policies that target overweight and obesity prevalence may also reduce diabetes prevalence, which could result in huge economic gains.

Chapter 6

Determinants of Overweight and Obesity in India

6.1 Introduction

In the previous chapters, we have shown the health effects of a rise in overnutrition both in terms of a decline in life expectancy at birth and a rise in diabetes, therefore, it is important to determine the factors that may contribute to the rise in overweight and obesity in India. Overweight and obesity prevalence in India has almost doubled during 2005-06 and 2015-16 (NFHS). The rise in the overweight and obesity prevalence during the recent decade is attracting a lot of attention from both the researchers and the policy makers. Some of the policy interventions have already been made, fat tax imposition on the junk foods in Kerala in year 2016 being one such example, and many policy suggestions have been laid out by several institutions and organisations. However, identification of the factors contributing to the rise in overweight and obesity can help in formulation of more targeted and effective policies to control the overweight and obesity problem in India. In this chapter, we aim at identifying the factors that contribute to the increase overweight and obesity in India.

Overnutrition is an outcome of positive energy balance (or net calorie intake), that is, an individual is likely to gain weight if the calories consumed exceed the calories expended (Rashad, 2006 and Chou et al., 2004). Available literature has identified several potential factors that may lead to a rise in the overnutrition. These factors include transformation of the consumption basket towards food products that are high on fat and sugar, increase in restaurant food consumption, increase in sedentary behaviour, increased use of private transportation, etc. (Cutler et al., 2003; Alston et al., 2006; Fletcher et al., 2011; Young and Nestle, 2002 and Spanier et al., 2006).

Amongst all the factors that may lead to overweight and obesity, urbanisation is a major factor associated with the rise in overnutrition problem (WHO; FAO; Popkin, 1993; Banwell et al., 2009 and Popkin et al., 2012). WHO states that rapid urbanisation leads to major changes in the lifestyle of the population which could affect their overweight and obesity status. Popkin et al. (2012) reviewing the available literature states that urbanisation could be driving the obesity prevalence in low- and middle-income countries. The study by Garden and Jalaludin (2009) finds that individuals living in the urban areas with low population density face an increased risk of overweight and obesity.

Urban lifestyle is more conducive to higher net calorie intake due to the increased availability of calorie-intense or unhealthy food choices and decline in the physical effort required to carry out both household and work-related activities, that is, urban environment tends to discourage the physical activity and promotes calorie-intense consumption. Also, improper urban planning limits the opportunities for the physical activities in the form of insufficient space for the recreational or sports facilities and public parks, heavy reliance on transportation that limits energy expenditure, etc. One may expect that moving to the urban areas or overtime urbanisation of the existing areas would transform the lifestyle of the individuals in terms of their occupational structure, availability of the consumption choices (especially in the form of ready to eat or processed food products), exposure to the sedentary technology (television, computer and other related gadgets), etc. All these changes may influence the BMI levels and thereby overweight and obesity status of the population.

The studies that have examined the cause of overnutrition in India have focused upon estimating the effect of demographic and socioeconomic indicators on the overweight and obesity status of individuals (Sengupta et al., 2015 and Ackerson et al., 2008). The study by Griffiths and Bentley (2001), based on women in Andhra Pradesh, includes the diet pattern as measured in the terms of consumption of fruits and vegetables along with demographic and socioeconomic indicators, however, the study does not take into account the effect of consumption of calorie-intense foods. Griffiths and Bentley (2001) also analyse the effect of living in urban areas and finds that overweight and obesity is higher among the urban areas. These studies are based on women samples or on small sample data and do not represent the total population in India. Also, these studies have not accounted for the effect of unobserved factors which may affect the relationship between place of residence and BMI levels of the individuals.

In the present chapter, we estimate the causal impact of urbanisation on BMI levels for the population in India. For this purpose, we define urbanisation in terms of the place of residence of the individuals – urban and rural. We expect that living in urban areas is likely to be correlated with the omitted or unobserved lifestyle related factors that determine BMI levels of the individuals. This may lead to the potential endogeneity problem in the relationship between the BMI levels of the individuals and their place of residence in the form of omitted variable bias. To examine the effect of living in urban areas on the BMI levels of individuals, we estimate a 2SLS model with instrument variable. We use the same sample data as used in the previous chapter. Further, to understand other potential factors that may govern the overweight and obesity status of individuals, we also examine the effects of sedentary lifestyle and consumption pattern on the overweight and obesity status of the individuals by estimating a logistic regression model. In both the econometric specifications, we include several covariates on individual and household characteristics, and behavioural risk factors.

The results obtained from this study are expected to have implications in the policy framing especially in the context of urban planning. To put a restrain on the increasing overnutrition prevalence, it is important to have sufficient evidence on what factors affect overweight and obesity in India, and which section of the population is at a higher risk from overweight and obesity. Present study provides an evidence for such effects. Our study may also have important implications for the health and nutrition policies.

The rest of the chapter is organised as follows. Section 6.2 presents the conceptual framework and the empirical methodology adopted. Section 6.3 discusses the data and descriptive statistics along with the definition of the variables used. Section 6.4 reports the estimation results. Section 6.5 presents a discussion on the results. Finally, Section 6.6 concludes this chapter.

6.2 Conceptual Framework and Methodology

The net calorie intake or energy balance at time t, NC_t, is defined as the total calorie intake, C_t, minus the total calories expended or activity level, E_t:

$$NC_t = C_t - E_t \tag{6.1}$$

BMI of an individual is not just affected by energy balance at time *t* alone instead it is affected by the cumulative energy balance at time *t*, denoted by, $\sum_t NC_t$, that is, the BMI function of an individual can be written as (Chou et al., 2004):

$$BMI = f(\sum_t NC_t, \varepsilon) \tag{6.2}$$

that is, an individual will gain weight if the calories consumed exceed the calories expended and vice-versa. Here, ε is the vector of individual specific variables and relates to individual's predisposition towards overweight and obesity, and includes individual's demographic and socioeconomic indicators, consumption preferences, occupational characteristics, place of residence, built environment and other related factors, etc.

The study by Banwell et al. (2009), estimating a logistic regression model, reports a strong association between the urban residence and the obesity levels. Another study by Assah et al. (2015), analysing the physical activity patterns, finds that urban population is more likely to be obese as compared to the rural population. These evidences indicate that lifestyle in the urban areas could be promoting an increase in BMI through reduced physical activity levels and changes in consumption pattern. Built environment of the urban areas in the form of proximity to parks, neighbourhood layout, street connectivity, transportation, etc. is likely to affect the BMI of the individuals through its effect on the physical activity levels. In addition, availability of ready to eat food products, etc., may also affect BMI through changes in the consumption pattern. In this chapter, we, first, empirically estimate the effect of living in urban areas on the BMI levels of individuals. Following the above discussion, we expect that living in urban areas is associated with higher levels of BMI as compared to the living in rural areas. Secondly, we attempt to quantify the effect of sedentary lifestyle and consumption pattern on the overweight and obesity status of the individuals.

We test the following hypotheses:

Hypothesis 1: Living in urban areas leads to a higher body mass index of individuals.

Hypothesis 2: Sedentary lifestyle leads to overweight and obesity among individuals.

Hypothesis 3: Consumption of calorie-intense foods leads to overweight and obesity among individuals.

To test our first hypothesis, we use the BMI values of the individuals as the outcome variable and for testing our second and third hypotheses, we define these BMI values in a binary variable form which takes value 1 if an individual is overweight or obese and 0 otherwise. We apply WHO International BMI classification to categorise individuals as overweight or obese and non-overweight. For the purpose of analysis, we define urbanisation in terms of the place of residence of the individuals - urban or rural.

In first hypothesis, the explanatory variable of interest is place of residence of the individuals. We also include variables on the demographic and socioeconomic indicators, occupation type, behavioural risk factors, consumption patterns and the proxy measures for the physical activity levels or sedentary behaviour. The same set of variables is used to test the second and third hypotheses. We consider several individual and household level characteristics such as age, gender, educational attainment, household income and other socioeconomic indicators. One may expect the occupational characteristics to affect the overweight and obesity status through work related physical activity levels. Griffiths and Bentley (2001) have shown that agricultural and manual workers are less prone to overweight and obesity. Therefore, we control for the broad type of occupation of the individuals in our model. We also control for tobacco and alcohol consumption as they are expected to affect BMI, and overweight and obesity status of an individual. Tobacco consumption/smoking negatively affects the BMI of an individual by suppressing the appetite whereas alcohol consumption is expected to increase BMI (Chou et al., 2004; Fagard and Nilsson, 2009 and Howard et al., 2004).

To estimate the effect of sedentary lifestyle on overweight and obesity status of the individuals, we use proxy measures for the sedentary lifestyle in terms of frequency of television watching for an individual, and household's possession of assets such as mobile, computer and washing machine. Increased usage of these assets may reduce the physical activity levels of the individuals. There exists evidence stating that higher screen time is associated with obesity (Spanier et al., 2006 and Lopes et al., 2014). Available literature suggest that mode of transportation used also affects obesity. Higher usage of private transport as compared to the public transit is linked to obesity (Teimann et al., 2008 and She et al., 2017). Therefore, we include variables on the household's possession of car/truck and scooter/motorcycle as additional proxy measures for the physical activity levels or sedentary lifestyle.

To estimate the effect of consumption pattern on overweight and obesity status of the individuals, we include variables on daily or weekly consumption of milk and curd, dark green leafy vegetables, fruits, fried foods and aerated drinks. Here, consumption of fried foods and aerated drinks are used to proxy for the consumption of calorie-intense foods.

Empirical Methodology

I. Body Mass Index and Urbanisation: IV-2SLS Model

Urbanisation as measured by the place of residence of individuals could be correlated with the unobserved lifestyle related factors that affects their BMI. This may lead to endogeneity problem in the form of omitted variable bias. We apply an IV-2SLS model using total road length per Km of state area as an instrument for the place of residence. We consider total road length of the state (in Km) and divide it by the area of that state (in Km^2).

Urban areas are commonly associated with higher road connectivity or network. The study by Imam and Banerjee (2016) states that urbanisation is linked to the higher density of road network. Having better road connectivity may increase access to the improved and more convenient modes of transportation. Also, in the absence of proper urban planning, higher road density may also restrict the space available for parks, etc., thereby putting a restrain on the physical activity opportunities. Better road connectivity is likely to be uncorrelated with the unobserved factors that determine BMI of the individual. However, it is possible that higher road connectivity may affect BMI of the individuals through transportation facilities (private or public). Therefore, we include variables on the household's ownership of the transports or vehicles in our model to control for such effects.

We estimate the following model:

$$BMI_{is} = \alpha_0 + \alpha_1 \, URB_{is} + \alpha_2 \, X_{is} + \varepsilon_{is} \tag{6.3}$$

where, i = 1, 2, …., n and s = 1, 2, …., N, represents i^{th} individual in state s;

BMI_{is} represents body mass index of i^{th} individual in state s;

URB_{is} is a dummy variable for the place of residence which takes value 1 if the i^{th} individual in state s lives in an urban area and 0 for rural area;

X_{is} represents the vector of controls for i^{th} individual in state s;

ε_{is} represents the error term.

The corresponding first stage regression can be written as:

$$URB_{is} = \gamma_0 + \gamma_1 \, RD_s + \gamma_2 \, X_{is} + \mu_{is} \tag{6.4}$$

where, RD_s is the total road length per Km of state area and is defined at the state level.

II. Sedentary Lifestyle, Consumption Pattern, and Overweight and Obesity Status: Logistic Regression Model

To examine the effect of sedentary lifestyle and consumption pattern on the overweight and obesity status of the individuals, we estimate a logistic regression model having binary dependent variable, OW_i, defined as,

$$OW_i = \begin{cases} 0 \; if \; the \; individual \; is \; non-overweight \\ 1 \; if \; the \; individual \; is \; overweight \; or \; obese \end{cases} \tag{6.5}$$

We estimate the following model:

$$OW_i^* = \beta' X_i + \nu_i \tag{6.6}$$

where, i = 1, 2, …., n, represents i^{th} individual;

OW_i^* represents latent selection variable for overweight and obesity status of i^{th} individual and is unobserved;

X_i represents the vector of individual and household specific characteristics for i^{th} individual;

ν_i represents the error term.

In case of logit models, the non-linear function, $\mathrm{G}(.)$, which is applied to the response function is the standard logistic cumulative distribution function. The maximum likelihood estimation is used to estimate the model. We estimate the odds ratio, that is, odds in favour of a response which is simply the exponent of the parameter estimate of a variable (Greene and Hensher, 2010).

6.3 Data and Descriptive Statistics

For the empirical analysis in the present chapter, we use the same data set as used in the Chapter 5, which is the individual level data from the fourth round of NFHS for the year 2015-16. However, for the purpose of the analysis in the present chapter, we include some additional variables as well. The list of variables included in the chapter along with their definitions is provided in the Table 6.1. In Table 6.1, we report only additional variables that are included and the variables which have been redefined (in binary or ordinal form) for the purpose of analysis in this chapter. The definition for the remaining variables has been provided in Table 5.1 of Chapter 5. The variables to be referred from Table 5.1 include individual characteristics such as gender, education, marital status and household characteristics such as wealth quintile, religion, caste or tribe, below poverty line or not, family structure (nuclear or joint), number of household members and region (urban or rural). In our sample data, we include only those individuals for whom BMI values are reported. The total sample size is about 0.8 million observations. As discussed in the Section 6.2, we also consider a variable on the total road length per Km of state area for which we have taken the data on the state-wise total road length (in Km) from the Handbook of Statistics on Indian States published by Reserve Bank of India and data on the area of the state (in Km^2) is taken from the Ministry of Statistics and Programme Implementation, Government of India.

Descriptive Statistics

Table 6.2 presents the descriptive statistics. The mean BMI is 21.71 kg/m^2 and 18.2% of the individuals in our sample are either overweight or obese. Examining the eating habits of the individuals, we find that more two-third of the individuals consume milk or curd and dark green leafy vegetables on daily or weekly basis, more than two-fifth of the individuals consume fruits and fried food on daily or weekly basis and about one-fourth of the individuals consume aerated drinks on daily or weekly basis.

We also present the descriptive statistics for some selected variables based on the place of residence in Table 6.3. For these selected variables, we report the proportion of individuals belonging to each sub-category across both rural and urban areas. This allows for a better comparison of the population across the two areas. We find that the proportion of overweight or obese individuals is twice in the urban areas as compared to the rural areas. Educational attainment and wealth distribution vary considerably across the two areas with urban areas performing better on the education indicators and being more likely to belong to the higher wealth quintiles. The occupational structure also varies across the two areas. Rural areas have agriculture as the major occupation while the urban areas are better characterised by the occupations such as professional/technical/managerial, clerical, sales, services and manual (skilled and unskilled). Also, the proportion of individuals who watch television almost every day is considerably higher among the urban areas as compared to the rural areas. The possession of household assets such as car/truck, motorcycle/scooter, mobile, computer and washing machine is higher among the urban areas.

In Figure 6.1, we graphically illustrate the relationship between the mean BMI Values (in kg/m^2) at state level and the percentage of the individuals living in the urban areas within a specific state. We have computed these values at the state level based on our sample data. The figure suggests a positive relationship between the mean BMI levels and urban residence.

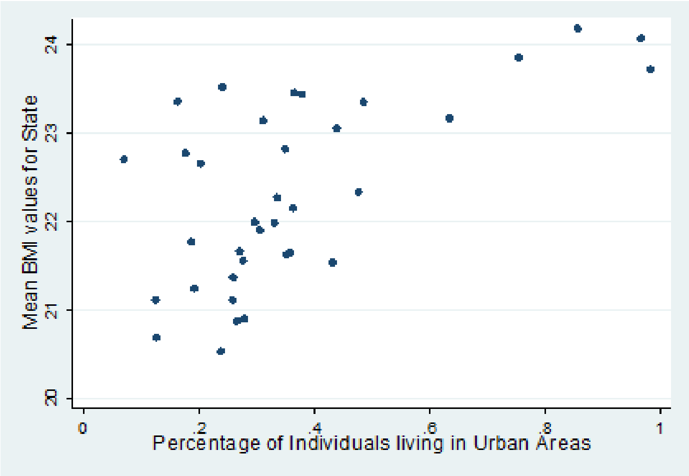


Figure 6.1: Mean BMI Values for the State (in kg/m^2) and Percentage of the Individuals living in Urban Areas within a State

Source: Figure constructed by authors using NFHS data.

Table 6.1: List of Variables with Definition and Type

Variable	Definition	Type
Dependent Variables:		
Body Mass Index	Same as defined earlier (in kg/m^2).	Continuous
Overweight and Obesity Status	• OW = 0 if individual is non-overweight • OW = 1 if individual is overweight or obese	Binary
Independent Variables:		
Individual Characteristics:		
Age	• = 0 if age is between 15-24 years$^{@}$ • = 1 if age is between 25-34 years • = 2 if age is between 35-44 years • = 3 if age is between 45 years and above	Ordinal
Behavioural Risk Factors[53]	• = 1 if consumes tobacco products in any form,[54] 0 otherwise$^{@}$ • = 1 if drinks alcohol, 0 otherwise$^{@}$	Binary
Eating Habits[55]	• = 1 if consumes milk or curd daily or weekly, 0 otherwise$^{@}$ • = 1 if eats dark green leafy vegetables daily or weekly, 0 otherwise$^{@}$ • = 1 if eat fruits daily or weekly, 0 otherwise$^{@}$ • = 1 if eats fried food daily or weekly, 0 otherwise$^{@}$ • = 1 if drinks aerated drink daily or weekly, 0 otherwise$^{@}$	Binary
Occupation	• = 0 if not working/no occupation$^{@}$ • = 1 if professional/technical/managerial • = 2 if clerical • = 3 if sales • = 4 if agricultural • = 5 if services • = 6 if manual (skilled and unskilled)	Ordinal
Television Watching	• = 0 if not at all$^{@}$ • = 1 if less than once a week • = 2 if at least once a week • = 3 if almost every day	Ordinal
Household Characteristics:		
Household Assets[56]	• = 1 if owns car/truck, 0 otherwise$^{@}$ • = 1 if owns motorcycle or scooter, 0 otherwise$^{@}$ • = 1 if owns mobile phone, 0 otherwise$^{@}$ • = 1 if owns computer, 0 otherwise$^{@}$ • = 1 if owns washing machine, 0 otherwise$^{@}$	Binary

[53] Contains a set of two dummy variables.
[54] Includes smoking cigarette, smoking pipe, chewing tobacco, snuffing, smoking cigar, chewing paan or gutkh and chewing paan with tobacco.
[55] Contains a set of five dummy variables.
[56] Contains a set of five dummy variables.

Table 6.1 (Continued)		
Variable	**Definition**	**Type**
Instrument Variable		
State wise Total Road Length	State wise total road length (in Km) for year 2015.	Continuous
Area of State	Area of state (in Km^2).	Continuous

[@] Indicates the base category.

Table 6.2: Descriptive Statistics

Variables	Observations	Mean	Standard Deviation	Minimum Value	Maximum Value
Individual Characteristics					
Body Mass Index (in kg/m^2)	811465	21.714	4.094	12.01	59.96
Overweight and Obesity Status	811465	0.182	0.385	0	1
Age	811465	1.113	1.014	0	3
Occupation	226929	2.268	2.385	0	6
Television Watching	795856	2.021	1.253	0	3
Behavioural Risk Factors					
Consumes Tobacco	795856	0.114	0.318	0	1
Consumes Alcohol	795856	0.065	0.246	0	1
Eating Habits					
Milk or curd	795856	0.645	0.479	0	1
Dark Green Leafy Vegetables	795856	0.856	0.351	0	1
Fruits	795856	0.434	0.496	0	1
Fried Food	795856	0.455	0.498	0	1
Aerated Drinks	795856	0.242	0.429	0	1
Household Characteristics					
Household Asset – Car	811465	0.070	0.256	0	1
Household Asset – Motorcycle or Scooter	811465	0.388	0.487	0	1
Household Asset – Mobile	811465	0.937	0.243	0	1
Household Asset – Computer	811465	0.091	0.287	0	1
Household Asset – Washing Machine	811465	0.146	0.353	0	1
Total Road Length Per Km of State Area	811465	1.280	1.874	0.0007	21.623

Notes: Variables such as body mass index, alcohol consumption and consumption of fried food and aerated drinks are reported here again.

Table 6.3: Proportion of Individuals across Different Categories for Selected Binary and Ordinal Variables based on the Place of Residence

Variables	Proportion of Individuals (in %)		
	Total	**Rural**	**Urban**
Overweight and Obesity Status:			
Non-Overweight	81.85	85.94	71.92
Overweight and Obesity	18.15	14.06	28.08
Education:			
No education or preschool	25.07	29.48	14.40
Primary	13.46	14.68	10.50
Secondary	49.59	47.99	53.46
Higher	11.88	7.85	21.63
Occupation:			
Not working/no occupation	47.24	45.54	51.07
Professional/technical/managerial	4.20	2.89	7.14
Clerical	1.08	0.76	1.79
Sales	4.98	3.32	8.73
Agricultural	22.25	29.69	5.42
Services	5.08	3.98	7.55
Manual (Skilled and Unskilled)	15.18	13.81	18.30
Television Watching:			
Not at all	23.16	29.52	7.80
Less than once a week	7.98	9.42	4.49
At least once a week	12.42	13.49	9.81
Almost every day	56.45	47.56	77.90
Wealth Index:			
Poorest	18.82	25.30	3.12
Poorer	21.40	27.04	7.73
Middle	21.17	23.08	16.55
Richer	19.88	15.52	30.41
Richest	18.74	9.05	42.19
Household Assets:			
Car/truck	7.03	4.88	12.21
Motorcycle/scooter	38.82	33.12	52.65
Mobile Phone	93.69	92.04	97.67
Computer	9.08	4.74	19.60
Washing Machine	14.58	7.99	30.54

6.4 Estimation Results and Interpretation

6.4.1 Body Mass Index and Urbanisation: IV-2SLS Estimates

The estimation results obtained from the IV-2SLS model are presented in Table 6.4. In all the models, standard errors are clustered at the state level. Table 6.4, we report the coefficient estimates for the urban residence only, which is our main explanatory variable of interest. We also report the Ordinary Least Squares (OLS) estimates for comparison sake. The complete result table containing coefficient estimates of all the variables included in the model is provided in the Table A.6.1 of Appendix.

In all the models, the coefficient of urban residence is positive and statistically significant. The results obtained from both OLS and IV-2SLS models, suggest that living in the urban areas is associated with higher levels of BMI. The coefficient estimate of the urban residence for IV-models is considerably higher as compared to the corresponding OLS estimate.

In column (2), we controlled for demographic and socioeconomic indicators. The results show that the average BMI is 1.7 kg/m^2 units higher for the individuals living in the urban areas as compared to those living in the rural areas. In column (4), we include additional controls on behavioural risk factors, daily and weekly consumption of certain food items and proxy measures for physical activity levels. We find that the coefficient of urban residence is robust to the inclusion of these variables. However, controlling for the type of occupation of the individuals (in column (6)), the coefficient estimate of the urban residence decreases by about 0.3 kg/m^2 units indicating that the occupational differences across rural and urban areas may also influence the BMI differences.

The Durbin-Wu-Hausman test for endogeneity, having null hypothesis that the regressor is exogenous, is rejected at 1% significance level for all the IV-models indicating that urban residence is indeed endogenous. Durbin-Wu-Hausman test statistic has a p-value of 0.0000 for all the IV-models.[57] Our instrument meets the F statistic condition. The Kleibergen-Paap rk Wald F statistic in all the IV-models has a higher value than the conventionally acceptable value of 10 or above indicating that our

[57] Durbin-Wu-Hausman test statistic has a value of 198.10, 152.2 and 35.60 for the models given in columns 2, 4 and 6 respectively.

instrument is not weak (Staiger and Stock, 1997). Since, we have considered one instrument variable for our endogenous variable, therefore, Hansen J statistic reports that the equation is exactly identified. The state fixed effects are omitted in the IV-models as our instrument is defined at state level. Therefore, we do not include the state fixed effects in the OLS models as well, however, including the state fixed effects do not alter our results.

To check for robustness of our results, we define our dependent variable in the binary variable form taking value 1 if the individual is overweight or obese and 0 otherwise.[58] We estimate a linear probability model with instrument variable. We use same instrument from the previous model. These results are reported in the Table A.6.2 of Appendix. The coefficient of urban residence is found to be positive and statistically significant. We find that the urban areas on average have 16% higher overweight and obesity prevalence as compared to the rural areas.

6.4.2 Effect of Sedentary Lifestyle and Consumption Pattern on Overweight and Obesity Status: Odds Ratio based on the Logistic Regression Model

Table 6.5 presents the results obtained from estimation of the logistic regression model. We report the odds ratio along with 95% confidence interval for each variable. Here, the dependent variable is overweight and obesity status of the individuals, which is defined based on WHO International BMI classification.[59] In both the models, we include variables on individuals' consumption pattern and proxy measures for sedentary lifestyle. We also include control variables on the individuals' demographic and socioeconomic characteristics as well as household characteristics. In model (2), we include additional controls on behavioural risk factors and type of occupation of the individuals.

We focus on model specification (2). The results suggest that sedentary lifestyle statistically significantly increases the likelihood of being overweight or obese. Our first measure for sedentary lifestyle or behaviour, frequency of television watching,

[58] Based on WHO International BMI classification.
[59] The binary dependent variable takes value 1 if the individual's BMI ≥ 25 kg/m^2 and value 0 if BMI < 25 kg/m^2.

indicates that more frequent television watching is associated with higher odds for being overweight or obese. Individuals who watch television almost every day are 21% more likely to be overweight or obese as compared to the individuals who do not watch television at all. Our second set of measures for sedentary lifestyle which pertain to the household's ownership of assets that may influence physical activity or promote sedentary lifestyle such as private transportation (car/truck and motorcycle/scooter) and electronic items (mobile phone and washing machine) show that individuals living in the households which own these assets are more likely to be overweight or obese.

The consumption of calorie-intense foods, daily or weekly consumption of fried food and aerated drinks, do not have any statistically significant effect on the overweight and obesity status of the individuals. Amongst the variables on the consumption pattern, only daily or weekly consumption of fruits has a statistically significant impact, however, the odds ratio is quite close to 1 indicating that there exists no substantial difference in the overweight and obesity status based on the frequency of fruit consumption.

We also find that with an increase in age, individuals are more likely to be overweight or obese. Females are more likely to be overweight or obese as compared to the males. We find that individuals having completed higher education are considerably more likely to be overweight or obese (1.3 times or 30% more likely) as compared to the individuals having no education or having completed only preschool. Married individuals are about twice more likely to be overweight or obese as compared to the unmarried individuals.

The likelihood of being overweight or obese is higher among the individuals belonging to the higher wealth quintile households. Individuals from the richest households are four times more likely to be overweight or obese as compared to the individuals from the poorest wealth quintile households. Examining the caste factors, we observe that the odds ratios are less than one indicating that the individuals belonging to the schedule caste, schedule tribe and other backward classes are less likely to be overweight or obese compared to the individuals who do not belong to these castes, respectively. Similar observations can be made for the individuals belonging to the below poverty line households. Also, the likelihood of being overweight or obese differs across individuals belonging to different religions.

In line with our earlier findings, we find that the individuals living in the urban areas are more likely to be overweight or obese as compared to the individuals living in rural areas. We find that tobacco consumption/smoking by individuals reduces their likelihood of being overweight or obese whereas for alcohol consumption we observe a positive effect. Occupation type has a statistically significant impact on the overweight and obesity status (except for services). For all the occupations (except sales), we find an odds ratio of less than one indicating that working population is less likely to be overweight or obese as compared to the non-working population. The population belonging to the agricultural occupations is least likely to be overweight or obese.

Table 6.4: Effect of Living in the Urban Areas on Body Mass Index: IV-2SLS Estimates

Variables	OLS	IV-2SLS	OLS	IV-2SLS	OLS	IV-2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Urban Residence	0.446*** (0.011)	1.773** (0.803)	0.440*** (0.012)	1.712** (0.850)	0.295*** (0.021)	1.397* (0.789)
Demographic and Socio-economic Factors	Yes	Yes	Yes	Yes	Yes	Yes
Behavioural Risk Factors	No	No	Yes	Yes	Yes	Yes
Consumption Pattern and Physical Activity Measures	No	No	Yes	Yes	Yes	Yes
Occupation Type	No	No	No	No	Yes	Yes
Observations	779519	779519	764454	764454	219781	219781
F Statistic	7311.21	1565.07	4772.10	12774.89	1240.29	3926.42
Prob > F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R^2	0.2067	0.1908	0.2106	0.1964	0.2110	0.1998
Kleibergen-Paap rk Wald F statistic		25.46		22.723		17.913

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses.
Notes: State fixed effects are omitted as the instrument variable used is a state level variable.
Demographic and Socio-economic factors include age, gender, education, household characteristics such as wealth quintile, religion, caste or tribe, below poverty line household, family structure and number of household members.
Behavioural risk factors include tobacco and alcohol consumption.
Consumption pattern includes daily or weekly consumption of milk or curd, dark green leafy vegetables, fruits, fried food and aerated drinks.
Physical activity measures include frequency of watching television and household's ownership of assets such as car/truck, motorcycle/scooter, mobile, computer and washing machine.
Occupation type includes occupations such as not working, professional/technical/managerial, clerical, sales, agricultural, services and manual (skilled and unskilled).

Table 6.5: Effect of Sedentary Lifestyle and Consumption Pattern on Overweight and Obesity Status: Odds Ratio with 95% Confidence Interval

Variables	(1)	(2)
Watch Television (Base = Not at all)		
Less than once a week	1.060*** (1.026 - 1.094)	1.078** (1.017 - 1.143)
At least once a week	1.101*** (1.072 - 1.131)	1.108*** (1.053 - 1.165)
Almost every day	1.231*** (1.203 - 1.260)	1.210*** (1.158 - 1.265)
Household Assets (Base = Do not own)		
Car/truck	1.067*** (1.042 - 1.092)	1.105*** (1.059 - 1.153)
Motorcycle/scooter	1.062*** (1.045 - 1.079)	1.167*** (1.133 - 1.202)
Mobile Phone	1.125*** (1.082 - 1.170)	1.121*** (1.040 - 1.208)
Computer	0.999 (0.977 - 1.021)	1.001 (0.962 - 1.042)
Washing Machine	1.143*** (1.119 - 1.168)	1.173*** (1.129 - 1.219)
Consumption Pattern[@] (Base = Occasionally or never consume)		
Milk or curd	0.952*** (0.937 - 0.967)	0.993 (0.963 - 1.024)
Dark green leafy vegetables	1.056*** (1.035 - 1.078)	1.024 (0.985 - 1.064)
Fruits	1.102*** (1.086 - 1.119)	1.096*** (1.067 - 1.126)
Fried food	0.998 (0.984 - 1.012)	0.999 (0.973 - 1.026)
Aerated drinks	0.992 (0.976 - 1.007)	0.985 (0.957 - 1.013)
Age (Base = 15-24 years)		
25-34 years	2.484*** (2.428 - 2.541)	2.514*** (2.407 - 2.625)
35-44 years	4.104*** (4.005 - 4.204)	3.851*** (3.672 - 4.039)
45 years and above	4.703*** (4.575 - 4.835)	4.151*** (3.939 - 4.375)
Gender (Base = Male)		
Female	1.189*** (1.166 - 1.212)	1.115*** (1.079 - 1.152)
Education (Base = No Education or Preschool)		
Primary	1.161*** (1.135 - 1.188)	1.157*** (1.107 - 1.210)
Secondary	1.222*** (1.198 - 1.246)	1.167*** (1.124 - 1.213)
Higher	1.231*** (1.198 - 1.264)	1.255*** (1.193 - 1.321)
Marital Status (Base = Unmarried)		
Married	2.027*** (1.979 - 2.077)	1.829*** (1.753 - 1.909)
Wealth Index (Base = Poorest)		
Poorer	1.670*** (1.619 - 1.723)	1.654*** (1.556 - 1.757)
Middle	2.562*** (2.479 - 2.647)	2.435*** (2.286 - 2.594)
Richer	3.674*** (3.547 - 3.806)	3.454*** (3.229 - 3.695)
Richest	4.614*** (4.427 - 4.808)	4.197*** (3.884 - 4.535)

Table 6.5 (Continued)		
Variables	**(1)**	**(2)**
Caste or Tribe (Base = Not SC, ST or OBC)		
Schedule Caste	0.875*** (0.857 - 0.894)	0.886*** (0.851 - 0.921)
Schedule Tribe	0.705*** (0.686 - 0.725)	0.750*** (0.714 - 0.788)
Other Backward Classes	0.894*** (0.879 - 0.909)	0.942*** (0.913 - 0.972)
Religion (Base = Hindu)		
Muslim	1.341*** (1.314 - 1.369)	1.213*** (1.168 - 1.261)
Christian	1.036* (0.997 - 1.077)	1.011 (0.944 - 1.082)
Sikh	1.247*** (1.184 - 1.314)	1.326*** (1.203 - 1.462)
Buddhist/Neo-Buddhist	1.271*** (1.199 - 1.346)	1.279*** (1.155 - 1.416)
Jain	1.096 (0.959 - 1.252)	0.938 (0.722 - 1.219)
Jewish	1.088 (0.199 - 5.950)	0.808 (0.0804 - 8.118)
Parsi/Zoroastrian	1.221 (0.737 - 2.021)	2.984** (1.256 - 7.088)
No religion	1.352** (1.049 - 1.741)	1.539* (0.934 - 2.536)
Other	1.105*** (1.030 - 1.185)	1.149** (1.005 - 1.314)
Below Poverty Line (Base = Not BPL)		
Yes	0.903*** (0.889 - 0.917)	0.906*** (0.880 - 0.933)
Residence (Base = Rural)		
Urban	1.293*** (1.273 - 1.314)	1.213*** (1.178 - 1.249)
Behavioural Risk Factors (Base = No)		
Consumes Tobacco		0.890*** (0.860 - 0.921)
Consumes Alcohol		1.033* (0.995 - 1.072)
Occupation (Base = Not working/no occupation)		
Professional/ Technical/Managerial		0.941** (0.889 - 0.996)
Clerical		0.891** (0.805 - 0.987)
Sales		1.100*** (1.043 - 1.161)
Agricultural		0.774*** (0.744 - 0.806)
Services		0.996 (0.944 - 1.051)
Manual (Skilled and Unskilled)		0.883*** (0.846 - 0.920)
Constant	0.015*** (0.014 - 0.017)	0.029*** (0.024 - 0.035)
State Fixed Effects	Yes	Yes
Observations	764,454	219,781
Wald chi2	89215.59	25259.92
Prob > chi2	0.0000	0.0000
Pseudo R^2	0.1643	0.1580

***, ** and * denote significance at 1%, 5% and 10% level.
95% confidence interval reported in parentheses.
Notes: Both the models also include variables on family structure and number of household members.
[@]For consumption pattern, we report odds ratio for daily or weekly consumption of different food items.

6.5 Discussion

In this chapter, we find that urbanisation leads to a rise in average BMI levels of the individuals, that is, individuals living in the urban areas have a higher BMI as compared to the individuals living the rural areas. This result is line with the studies by Banwell et al. (2009) and Garden and Jalaludin (2009) which find a positive association between living in urban areas and obesity. The results drawn from the present study on the causal effect of urbanisation on BMI levels can be generalised for the Indian population.

The coefficient estimates of urban residence obtained from IV-2SLS model and logistic regression model provide qualitatively similar results for the effect of urbanisation on BMI levels, and overweight and obesity status respectively. This suggests that our result on the effect of urbanisation on overnutrition is consistent across different model specifications.

Our result on the relationship between more frequent television watching, and overweight and obesity is line with studies by Spanier et al. (2006) and Lopes et al. (2014) which suggest that increased screen time causes obesity. Another important result is that individuals belonging to the households having ownership of private transportation and electronic items are more likely to be overweight or obese. This result is line with studies by Teimann et al. (2008) and She et al. (2017) which state that usage of private transportation is associated with obesity. However, due to data limitations, we could not assess the effect of the frequency of usage of these assets or time spent using these assets, which would be a more accurate measure for the physical activity levels.

The estimates of calorie-intense food consumption do not provide statistically significant results highlighting the need to measure consumption pattern using more appropriate and narrowly defined variables.

The results on the effects of tobacco consumption or smoking and alcohol consumption, based on the logistic regression model, are consistent with the available literature which states that smoking reduces obesity and alcohol consumption increases obesity (Fagard and Nilsson, 2009). In addition, the result that population working in agriculture and related occupation is less likely to be overweight or obese is quite expected and consistent with the results of the study by Griffiths and Bentley (2001) which suggests that individuals working in jobs that require more physical activity are less prone to becoming overweight or obese.

6.6 Conclusion

The present chapter provides an evidence for the effect of urbanisation on overnutrition in India. Additionally, the chapter also quantifies the effect of sedentary lifestyle and consumption of calorie-intense food on overweight and obesity for the population in India. The chapter, first, examines the effect of living the urban areas on BMI levels of the individuals using an IV-2SLS model. The results showed that individuals living in urban areas have higher average BMI levels as compared to the individuals living the rural areas. This effect is robust to the inclusion of controls on the behavioural risk factors, consumption pattern, physical activity measures and occupation type. The results suggest that factors related to the urban lifestyle such as built environment, consumption patterns, modes of transportation, access to sedentary technology, etc. contribute to the rise in the overweight and obesity. We expect these results to be stronger across towns and cities within the urban areas.

We also analysed the effect of sedentary lifestyle and consumption or dietary patterns on overweight and obesity status of the individuals using a logistic regression model. The measures of sedentary lifestyle such as household’s ownership of private transportation and electronic items, and higher frequency of watching television (almost every day) are found to be associated with higher overweight and obesity. The variables used to capture the effect of consumption pattern, do not explain the variations in the overweight and obesity status, however, to better understand the effect of eating habits, one could consider variables on frequency of restaurant food consumption, consumption of ready to eat packaged food products, etc.

Chapter 7

Conclusion

Examining the nutritional pattern in India during the period of 2005-06 and 2015-16, we found that India is experiencing a nutritional transition in the form of a shift away from undernutrition towards overnutrition. The available data showed that the gap between prevalence of undernutrition and overnutrition at all India level has drastically reduced over the past ten years, from 23.6% to about only 2%. If similar trends continue, the prevalence of overweight and obesity would dominate the prevalence of underweight very soon. The nutritional transition was also evident at the state level.

This rapid emergence of overnutrition is expected to pose a major health challenge in India. Overnutrition imposes both health and economic burden. Overnutrition increases the risk of non-communicable diseases and has an adverse impact on the health outcomes such as longevity, mortality, quality of life, etc. In addition, it also increases healthcare spending and contributes to the economic burden.

The goal of the thesis was to examine the effects of overnutrition on health outcomes. It also examined the potential factors that contribute to the rise in overnutrition. For the purpose, we examined three research objectives in this thesis. In first research objective, we investigated the adverse impacts of overnutrition on health outcomes at macro level by considering life expectancy at birth as a measure of health status. For this, we considered two countries, the United States and India. In second research objective, the effect of overnutrition on diabetes is examined at micro level in the Indian context. Finally, in third research objective, we investigated the factors that contribute to the rise in overnutrition in India at micro level.

To assess the impacts of overnutrition on life expectancy at birth, we conducted an empirical analysis for two countries, the United States and India, using state level aggregate data. For the United States, a Fixed Effects model and a System GMM model were estimated. We considered state level data for 50 states over a fifteen-year period during 2000-14. The results showed that life expectancy at birth has a concave relationship with the obesity prevalence, that is, life expectancy at birth is increasing in obesity prevalence at a decreasing rate. The threshold obesity prevalence beyond which life expectancy at birth begins to fall is 26%, based on GMM estimates. It is found that obesity prevalence across many states of the United States has passed the threshold level and most of the states are placed on the downward sloping segment of the concave curve indicating that longevity is declining in obesity prevalence.

A similar analysis was conducted for India, using data across 21 states for the period of 2005-06 and 2015-16. We, first estimated a Fixed Effects model and to address the potential endogeneity, we estimated an IV-2SLS model. We constructed a weighted average index of household's possession of assets such as television and mobile, and used it to instrument the overweight and obesity prevalence. The results suggested for existence of a concave relationship between life expectancy at birth, and overweight and obesity prevalence. The threshold overweight and obesity prevalence beyond which life expectancy at birth declines is 26%, based on IV-2SLS estimates. For India, the overweight and obesity prevalence is below the threshold level in many states and most of the states are placed on the upward sloping segment of the concave curve suggesting that longevity is increasing in overweight and obesity prevalence. These results suggest that the extent of overnutrition governs the nature of the effect of overnutrition on longevity.

Another important result is that per capita health expenditure counters the adverse effects of overnutrition on longevity. The gains in longevity associated with a rise in per capita health expenditure are much higher among low- and middle-income countries like India as compared to the high-income countries. We find that a 1000 US dollar increase in per capita health expenditure may improve life expectancy at birth by about 0.24 months in the United States whereas the same amount may increase the life expectancy at birth in India by about 580 months. This suggests that the marginal health benefits associated with an increase in per capita health expenditure are substantially higher for India.

Recognising the recently growing problem of diabetes in India and the potential adverse effects of overnutrition on diabetes, we quantified the effect of an additional unit gain in body mass index on the likelihood of being prediabetic and diabetic for the population in India. For this purpose, we considered individual level data from the fourth round of National Family Health Survey for the year 2015-16. We measured the diabetes status of the population using two alternative indicators – self-reported diabetes status and ordinal blood glucose levels. To address the endogeneity problem in the form of omitted variable bias, we estimated an IV-Probit model using self-reported diabetes status as the outcome variable. We instrumented the body mass index of an individual with the body mass index of a non-biologically related household member (individual's spouse). The results showed that overweight or obese individuals are thrice more likely to be diabetic as compared to the non-overweight individuals. The change in probability of being diabetic due to a unit rise in body mass index is found to be 1.5% amongst the overweight or obese individuals.

Using the ordinally defined blood glucose levels, we estimated an Ordered Probit model and examined the marginal effects of a rise in body mass index on the probability of being prediabetic and diabetic. The results showed that the increase in the probability of being diabetic (prediabetic) due to a unit rise in BMI is thrice (twice) among overweight or obese individuals as compared to the non-overweight individuals. Our results on the effect of overweight and obesity on diabetes are found to be qualitatively similar across different model specifications and both the measures of diabetes. Also, application of both WHO International and Asian BMI classifications for defining overweight and obesity status of the population generates similar results. Additionally, we also found that amongst overweight or obese individuals, men, population living in the urban areas and population belonging to the richest wealth quintile face a higher risk of being prediabetic and diabetic as compared to women, population living in the rural areas and population belonging to the poorest wealth quintile respectively.

Finally, we examined the factors that affect overnutrition for the population in India. For this purpose, we used the same sample data that was used in the analysis of the second research objective. We, first, examined the effect of living in the urban areas on body mass index levels of the individuals. For this, an IV-2SLS model was estimated using the total road length per Km of state area as an instrument for the place of residence. The results showed that individuals living in the urban areas on average have at least 1.4 kg/m^2 units higher body mass index as compared to the individuals living the rural areas. We also estimated the effect of sedentary lifestyle and consumption of calorie-intense foods on overweight and obesity status of the individuals using a logistic regression model. Using proxy measures for sedentary lifestyle and consumption pattern, we found that the physical activity measures such as household's ownership of assets (private transportation and certain electronic items) and higher frequency of watching television (almost every day) are associated with higher overweight and obesity. This provides an evidence for the effects of an increased access to sedentary technology on the overweight and obesity in India. We could not find any significant effect of the consumption patterns on overweight and obesity.

This study contributes to the literature, first, by providing an evidence for the adverse effects of overnutrition on the health outcomes as measured by life expectancy at birth. Ours is the first study that examines the concave relationship between life expectancy at birth and overweight and/or obesity prevalence by considering state level aggregate data in the United States and India. We provide an evidence on how the health effects of overnutrition varies across countries based on their respective nutritional status. We provide an estimate for the causal effect of overnutrition on longevity. Second, we provide evidence on the health benefits associated with an increase in per capita health expenditure. Third, we provide an estimate for the causal effect of overweight and obesity on diabetes for the population in India. This result is of huge significance given the recent growth witnessed in both overweight and obesity, and diabetes prevalence in India. Fourth, we show that the urban population in India is at a higher risk of overnutrition and associated health effects. Lastly, we provide an evidence for the effect of sedentary lifestyle on the overweight and obesity in India.

The policy implication of this study is that the policy makers can reduce the burden of non-communicable diseases and improve health outcomes by addressing the rise in overweight and obesity prevalence. Effects of overnutrition on the health status are heterogenous across states and a uniform health policy may not be appropriate. Therefore, government can design different policies based on the nutritional status of the state to improve the health conditions. For the states having overnutrition problem, various policy options include taxing calorie intense foods such as foods/drinks having high sugar, salt and saturated fat, subsidising foods having high nutritive value, improvements in product labelling, educating people about the ill-health effects associated with overnutrition so that they can be self-incentivised to adopt a healthy lifestyle, etc., and more research can examine which policy suits best.

The study faces some limitations. First, the estimates for the effect of overnutrition on longevity in India are based on only two time periods data set due to data limitations. A larger panel data including more time periods may help in improving over the existing estimates. Second, the causal effect of overweight and obesity on diabetes is representative for the married couples subsample, living in the same household of whom either is the head of the family. A richer data set could generalise these results. The estimates based on the Probit and Ordered Probit models, however, can be generalised for the population at large in India. Also, the estimates are based on one time period analysis. A panel data set will facilitate better understanding of how the past values and overtime changes in BMI of an individual affects the likelihood of being prediabetic and diabetic. Third, our estimates on the effects of urbanisation and sedentary lifestyle on overnutrition are based on one time period analysis. A panel data set may help in better understanding of how the overtime changes in urbanisation and access to sedentary technology has contributed to the overweight and obesity in India.

Future research can investigate the impact of overnutrition on health outcomes such as longevity and death rate at a more disaggregated level data such as district level. This will help in better understanding of the relationship between overnutrition and longevity or death rate by capturing the within country heterogeneity more accurately. Researchers can also examine the effects of overnutrition on other NCDs such as cardiovascular diseases, hypertension, etc. Researchers may also quantify the healthcare burden associated with diabetes. To better understand the effects of sedentary lifestyle and consumption pattern on the overweight and obesity status of the population, more narrowly defined measures can be considered.

References

Ackerson, L. K., Kawachi, I., Barbeau, E. M. and Subramanian, S. V. (2008), "Geography of Underweight and Overweight among Women in India: A Multilevel Analysis of 3204 Neighborhoods in 26 States", *Economics and Human Biology*, 6(2): 264–280.

Allison, D.B., Fontaine, K.R., Manson, J. E., Stevens, J. and VanItallie, T. B. (1999), "Annual Deaths Attributable to Obesity in the United States", *Journal of the American Medical Association*", 282(16): 1530–38.

Alston, J. M., Sumner, D. A. and Vosti, S. A. (2006), "Are Agricultural Policies Making Us Fat? Likely Links between Agricultural Policies and Human Nutrition and Obesity, and Their Policy Implications", *Applied Economic Perspectives and Policy*, 28(3): 313–322.

Anderson, M. L. and Matsa, D. A. (2011), "Are Restaurants Really Supersizing America?", *American Economic Journal: Applied Economics*, 3(1): 152-88.

Andreyeva, T., Michaud, P. C. and Soestc, A. V. (2007), "Obesity and Health in Europeans Aged 50 years and Older", *Public Health*, 121: 497–509.

Arellano, M. and Bond, S. (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, 58(2): 277–297.

Arellano, M. and Bover, O. (1995), "Another Look at the Instrumental Variable Estimation of Error-Components Models", *Journal of Econometrics*, 68(1): 29–51.

Assah, F., Mbanya, J. C., Ekelund, U., Wareham, N. and Brage, S. (2015), "Patterns and Correlates of Objectively Measured Free-Living Physical Activity in Adults in Rural and Urban Cameroon", *Journal of Epidemiology and Community Health*, 69(7) :700-707.

Astrup, A. (2005), "Super-Sized and Diabetic by Frequent Fast-Food Consumption?", *The Lancet*, 365(9453): 4-5.

Bansal, S. and Zilberman, D. (2016), "Does Health Care Expenditure Counter Adverse Effects of Obesity on Health: Evidence from Global Data", Presented on January 6-8, 2017 at Allied Social Sciences Association (ASSA) Annual Meeting Chicago, Illinois.

Banwell, C., Lim, L., Seubsman, S. A., Bain, C., Dixon, J. and Sleigh, A. (2009), "Body Mass Index and Health-Related Behaviours in a National Cohort of 87 134 Thai Open University Students", *Journal of Epidemiology and Community Health*, 63(5): 366-372.

Becker, W. E. and Kennedy, P. E. (1992), "A Graphical Exposition of the Ordered Probit", *Econometric Theory*, 8(1): 127-131.

Behavioral Risk Factor Surveillance System (2018), *Prevalence and Trends Data*, Centers for Disease Control and Prevention.

Bener, A., Zirie1, M. and Al-Rikabi, A. (2005), "Genetics, Obesity, and Environmental Risk Factors Associated", *Croatian Medical Journal*, 46(2): 302-7.

Berndt, E. (2001), "The U.S. Pharmaceutical Industry: Why Major Growth in Times of Cost Containment?", *Health Affairs*, 20: 100-14.

Bhattacharya, J. and Sood, N. (2006), "Health Insurance and the Obesity Externality", *Advances in Health Economics and Health Services Research*, 17: 281-321.

Bhattacharya, J. and Sood, N. (2011), "Who Pays for Obesity?", *The Journal of Economic Perspectives*, 25(1): 139-157.

Blundell, R. and Bond, S. (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics*, 87(1): 115–143.

Boes, S. and Winkelmann, R. (2006), "Ordered Response Models", *Allgemeines Statistisches Archiv*, 90(1): 167-181.

Brunello, G., Michaud, P. C. and Galdeano, A. S. (2009), "The Rise of Obesity in Europe: An Economic Perspective", *Economic Policy*, 24(59): 551-596.

Bureau of Economic Analysis (2018), *Regional Data*, U. S. Department of Commerce. URL:https://apps.bea.gov/iTable/iTable.cfm?reqid=70&step=1&isuri=1&acrdn=2#reqid=70&step=1&isuri=1&acrdn=2

Carlsson, S., Hammar, N., Grill, V. and Kaprio, J. (2003), “Alcohol Consumption and the Incidence of Type 2 Diabetes”, *Diabetes Care*, 26(10): 2785-2790.

Cash, S. B., Sunding, D. L. and Zilberman, D. (2007), “Fat Taxes and Thin Subsidies: Prices, Diet, and Health Outcomes”, *Acta Agriculturae Scandinavica, Section C*, 2(3-4): 167-174.

Cawley, J. (2000), “Body Weight and Women's Labor Market Outcomes”, *National Bureau of Economic Research Working Paper* 7841.

Cawley, J. and Meyerhoefer, C. (2012), “The Medical Care Costs of Obesity: An Instrumental Variables Approach”, *Journal of Health Economics*, 31: 219–230.

Centers for Medicare and Medicaid Services (2018), *National Health Expenditure Accounts*, Department of Health and Human Services

Chang, S. A. (2012), “Smoking and Type 2 Diabetes Mellitus”, *Diabetes and Metabolism Journal*, 36(6): 399-403.

Chiburis, R. and Lokshin, M. (2007), “Maximum Likelihood and Two-Step Estimation of an Ordered-Probit Selection Model”, *The Stata Journal*, 7(2): 167–182.

Chou, S. Y., Grossman, M. and Saffer, H. (2004), “An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System”, *Journal of Health Economics*, 23(3): 565-87.

Cobb, L. K., McAdams-DeMarco, M. A., Gudzune, K. A., Anderson, C. A. M., Demerath, E., Woodward, M., Selvin, E. and Coresh, J. (2015), “Changes in Body Mass Index and Obesity Risk in Married Couples Over 25 Years - The ARIC Cohort Study”, *American Journal of Epidemiology*, 183(5): 435–443.

Colditz, G.A., Willett, W. C., Rotnitzky, A. and Manson, J. E. (1995), “Weight Gain as a Risk Factor for Clinical Diabetes Mellitus in Women”, *Annals of Internal* Medicine, 122(7): 481-6.

Corneliben, T. (2005), “Standard Errors of Marginal Effects in the Heteroskedastic Probit Model”, *Institute of Quantitative Economic Research*, University of Hannover, Germany, Discussion Paper No. 320.

Cutler, D. M., Glaeser, E. L. and Shapiro, J. M. (2003), "Why Have Americans Become More Obese?", *The Journal of Economic Perspectives*, 17(3): 93-118.

Dall, T. M., Yang, W., Halder, P., Pang, B., Massoudi, M., Wintfeld, M., Semilla, A. P., Franz, J. and Hogan, P. F. (2014), "The Economic Burden of Elevated Blood Glucose Levels in 2012: Diagnosed and Undiagnosed Diabetes, Gestational Diabetes Mellitus, and Prediabetes", *Diabetes Care,* 37(12): 3172-3179.

Demographic and Health Surveys (2018), *Data Collection, Survey for India, NFHS*. URL: https://dhsprogram.com/data/

Dhana, K., Nano, J., Ligthart, S., Peeters, A., Hofman, A., Nusselder, W., Dehghan, A. and Franco, O. H. (2016), "Obesity and Life Expectancy with and without Diabetes in Adults Aged 55 Years and Older in the Netherlands: A Prospective Cohort Study", *PLOS Medicine*, 13(7): 1-13.

Diabetes Foundation, India (2018), Assessed December 2018. URL: http://www.diabetesfoundationindia.org/about.htm

Ehrlich, I. and Becker, G. S. (1972), "Market Insurance, Self-Insurance and Self Protection", *Journal of Political Economy*, 80(4): 623-48.

Fagard, R. H. and Nilsson, P. M. (2009), "Smoking and Diabetes-The Double Health Hazard!", *Primary Care Diabetes*, 3(4): 205-209.

Falba, T. A. and Sindelar, J. L. (2008), "Spousal Concordance in Health Behavior Change", *Health Services Research*, 43(1): 96-116.

Fearon, J. D. (2001), "Regression and Review of Course Topics", Lecture Notes, Stanford University. URL: https://web.stanford.edu/class/polisci100a/regress5.pdf

Flegal, K. M., Williamson, D. F., Pamuk, E. R. and Rosenberg, H. M. (2004), "Estimating Deaths Attributable to Obesity in the United States", *American Journal of Public Health*, 94(9): 1486–89.

Fletcher, J. M., Frisvold, D. E. and Tefft, N. (2011), "Are Soft Drink Taxes an Effective Mechanism for Reducing Obesity?", *Journal of Policy Analysis and Management*, 30(3): 655-662.

Fontaine, K. R., Redden, D. T., Wang, C., Westfall, A. O. and Allison, D. B. (2003), "Years of Life Lost Due to Obesity", *Journal of the American Medical Association*, 289(2): 187-93.

Food and Agricultural Organization of the United Nations (2018), URL: http://www.fao.org/about/en/

Garden, F. L. and Jalaludin, B. B. (2009), "Impact of Urban Sprawl on Overweight, Obesity, and Physical Activity in Sydney, Australia", *Journal of Urban Health*, 86(1): 19–30.

Geiss, L. S., Kirtland, K., Lin, J., Shrestha, S., Thompson, T., Albright, A. and Gregg, E. W. (2017), "Changes in Diagnosed Diabetes, Obesity, and Physical Inactivity Prevalence in US Counties, 2004-2012", *PLOS One*, 12(3): e0173428.

Global Health Data Exchange (2018), *Institute for Health Metrics and Evaluation*, University of Washington. URL: http://ghdx.healthdata.org/us-data

Government of India (2016), *Area and Population by States*, Ministry of Statistics and Programme Implementation.

Government of India (2013), *Cause of Death Statistics*, Office of the Registrar General & Census Commissioner, Ministry of Home Affairs.

Government of India (2005), *National Family Health Survey 2005-06*, Ministry of Health and Family Welfare.

Government of India (2015), *National Family Health Survey 2015-16*, Ministry of Health and Family Welfare.

Government of India (2004), *National Health Accounts 2004-05*, Ministry of Health and Family Welfare.

Government of India (2015), *National Health Accounts 2014-15*, Ministry of Health and Family Welfare.

Government of India (2015), *National Sample Survey Office*, Ministry of Statistics and Programme Implementation.

Government of India (2015), *Sample Registration System*, Ministry of Home Affairs.

Government of India (2010), *State-wise Indicators of Poverty*, Planning Commission.

Gray, L. J., Yates, T., Davies, M. J., Brady, E., Webb, D. R., Sattar, N. and Khunti, K. (2011), "Defining Obesity Cut-Off Points for Migrant South Asians", *PLOS One*, 6(10): e26464.

Greene, W. H. and Hensher, D. A. (2010), *Modeling Ordered Choices: A Primer*, New York: Cambridge University Press.

Gregg, E. W. and Guralnik, J. M. (2007), "Is Disability Obesity's Price of Longevity?", *Journal of the American Medical Association*, 298(17): 2066–67.

Griffiths, P. L. and Bentley, M. E. (2001), "The Nutrition Transition Is Underway in India", *The Journal of Nutrition*, 131(10): 2692-700.

Gulati, S. and Misra, A. (2014), "Sugar Intake, Obesity, and Diabetes in India", *Nutrients*, 6(12): 5955-74.

Hewitt, J. K. (1997), "The Genetics of Obesity: What Have Genetic Studies Told Us About the Environment", *Behavior Genetics*, 27(4): 353-358.

Howard, A. A., Arnsten, J. H. and Gourevitch, M. N. (2004), "Effect of Alcohol Consumption on Diabetes Mellitus: A Systematic Review", *Annals of Internal Medicine*, 140(3): 211–219.

Huffman, M. D., Prabhakaran, D., Osmond, C., Caroline, H. D. F., Fall, C. H. D., Tandon, N., Lakshmy, R., Ramji, S., Khalil, A., Gera, T., Prabhakaran, P., Biswas, S. K. D., Reddy, K. S., Bhargava, S. K. and Sachdev, H. S. (2011), "Incidence of Cardiovascular Risk Factors in an Indian Urban Cohort: Results From the New Delhi Birth Cohort", *Journal of American College of Cardiology*, 57(17): 1765–1774.

IDF (2017), International Diabetes Federation, IDF SEA members, Accessed December 2018, URL: https://www.idf.org/our-network/regions-members/south-east-asia/members/94-india

IDF (2017), International Diabetes Federation, IDF Diabetes Atlas - 8th Edition, Accessed March 2019, URL: https://diabetesatlas.org/across-the-globe.html

Imam, A. U. K. and Banerjee, U. K. (2016), “Urbanisation and Greening of Indian Cities: Problems, Practices, and Policies”, *Ambio*, 45(4): 442–457.

Kahn, B. B. and Flier, J. S. (2000), “Obesity and Insulin Resistance”, *The Journal of Clinical Investigation*, 106(4): 473-481.

Katzmarzyk, P. T., Hebebrand, J. and Bouchard, C. (2002), “Spousal Resemblance in the Canadian Population: Implications for the Obesity Epidemic”, *International Journal of Obesity and Related Metabolic Disorders*, 26(2): 241-6.

Kaufman, D. W., Kelly, J. P., Rosenberg, L., Anderson, T. E. and Mitchell, A. A. (2002), “Recent Patterns of Medication Use in the Ambulatory Adult Population of the United States”, *Journal of the American Medical Association,* 282: 337-44.

Kerr, D., Penfold, S., Zouwail, S., Thomas, P. and Begley, J. (2009), “The Influence of Liberal Alcohol Consumption on Glucose Metabolism in Patients with Type 1 Diabetes: A Pilot Study”, *QJM: An International Journal of Medicine*, 102(3): 169-174.

Kostova, D. (2011), “Can the Built Environment Reduce Obesity? The Impact of Residential Sprawl and Neighborhood Parks on Obesity and Physical Activity”, *Eastern Economic Journal*, 37(3): 390-402.

Lakdawalla, D., Philipson, T. and Bhattacharya, J. (2005), “Welfare-Enhancing Technological Change and the Growth of Obesity”, *American Economic Review*, 95 (2): 253-257.

Lichtenberg, F. R. (2002), “Sources of U.S. Longevity Increase, 1960-1997”, *National Bureau of Economic Research (NBER)* Working Paper 8755.

Lichtenberg, F. R. (2011), “The Quality of Medical Care, Behavioral Risk Factors and Longevity Growth”, *International Journal of Health Care Finance Economics*, 11:1-34.

Lindeboom, M., Lundborg, P. and Klaauw, B. V. D. (2009), “Obesity and Labor Market Outcomes: Evidence from the British NCDS”, *Institute for the Study of Labor* Discussion Paper 4099.

Lopes, A. S., Silva, K. S., Filho, V. C. B., Bezerra, J., Oliveira, E. S. A. and Nahas, M. V. (2014), "Trends in Screen Time on Week and Weekend Days in a Representative Sample of Southern Brazil Students", *Journal of Public Health*, 36(4): 608–614.

Lundin, D. (2000), "Moral Hazards in Physician Prescription Behavior", *Journal of Health Economics*, 19: 639-62.

Malley, O. G., Santoro, N., Northrup, V., D'Adamo, E., Shaw, M., Eldrich, S. and Caprio. S. (2010), "High Normal Fasting Glucose Level in Obese Youth: A Marker for Insulin Resistance and Beta Cell Dysregulation", *Diabetologia*, 53(6): 1199-209.

Mapchart.net (2018), *Online Map Tool*, URL: https://mapchart.net/india.html

Mayhew, L., Richardson, J. and Rickayzen, B. (2009), "A Study into the Detrimental Effects of Obesity on Life Expectancy in the UK", Cass Business School, London, United Kingdom. URL: https://www.actuaries.org.uk/system/files/documents/pdf/Mayhew_obesity.pdf

McCrory, M. A., Suen, V. M. M. and Roberts, S. B. (2002), "Biobehavioral Influences on Energy Intake and Adult Weight Gain", *The Journal of Nutrition*, 132(12): 3830S–3834S. https://doi.org/10.1093/jn/132.12.3830S

Meenakshi, J. V. (2016), "Trends and Patterns in the Triple Burden of Malnutrition in India", *Agricultural Economics*, 47: 115–134.

Mehta, N. K. and Chang, V. W. (2009), "Mortality Attributable to Obesity among Middle-Aged Adults in the United States", *Demography*, 46(4): 851–872.

Mello, M. M., Studdert D. M. and Brennan, T. A. (2006), "Obesity - The New Frontier of Public Health Law", *New England Journal of Medicine*, 354(24): 2601-10.

Moebus, S., Gores, L., Losch, C. and Jockel, K. H. (2011), "Impact of Time Since Last Caloric Intake on Blood Glucose Levels", *European Journal of Epidemiology*, 26(9): 719–728.

National Vital Statistics Reports (2016), *Deaths: Leading Causes for 2016*. U.S. Department of Health and Human Services, Centers For Disease Control and Prevention (CDC), National Center For Health Statistics.

Nelson, M. C., Larsen, P. G., North, K. E. and Adair, L. S. (2006), "Body Mass Index Gain, Fast Food, and Physical Activity: Effects of Shared Environments over Time", *Obesity*, 14(4): 701-9.

Olinto, M. T. A., Theodoro, H. and Canuto. R. (2017), *Epidemiology of Abdominal Obesity. Adiposity - Epidemiology and Treatment Modalities*, IntechOpen Publishers. URL: file:///C:/Users/shivani/Downloads/52576%20(1).pdf

Organisation for Economic Co-operation and Development (2015), *Health Spending.* URL: https://data.oecd.org/healthres/health-spending.htm

Organisation for Economic Co-operation and Development (2017), *Obesity Update,* URL: http://www.oecd.org/

Owens, C.S., Crone, D., De Ste Croix, M. B. A., Gidlow, C. J. and James, D. V. B. (2013), "Physical Activity and Screen Time in Adolescents Transitioning Out of Compulsory Education: A Prospective Longitudinal Study", *Journal of Public Health,* 36(4): 599-607.

Patel, S., Bhopal, R., Unwin, N., White, M., Alberti, K. G. M. M. and Yallop, J. (2001), "Mismatch between Perceived and Actual Overweight in Diabetic and Non-Diabetic Populations: A Comparative Study of South Asian and European Women", *Journal of Epidemiology and Community Health*, 55(5): 332-333.

Pereira, M. A., Kartashov, A., Ebbeling, C. B., Horn, L. V., Slattery, M. L., Jacobs, D. R. and Ludwig, D. S. (2005), "Fast-Food Habits, Weight Gain, and Insulin Resistance (The CARDIA Study): 15-Year Prospective Analysis", *The Lancet,* 365(9453): 36-42.

Philipson, T. and Posner, R. (2008), "Is the Obesity Epidemic a Public Health Problem? A Review of Zoltan J. Acs and Alan Lyles's: Obesity, Business and Public Policy", *Journal of Economic Literature, 46*(4): 974-982.

Popkin, B. M. (1993), "Nutritional Patterns and Transitions", *Population and Development Review,* 19(1): 138-157.

Popkin, B.M., Adair, L. S. and Ng, S. W. (2012), "Now and Then: The Global Nutrition Transition: The Pandemic of Obesity in Developing Countries", *Nutrition Review*, 70(1): 3-21.

Preston, S. H. and Stokes, A. (2011), “Contribution of Obesity to International Differences in Life Expectancy”, *American Journal of Public Health*, 101(11): 2137-2143.

Preston, S. H., Stokes, A., Mehta, N. K. and Cao, B. (2014), “Projecting the Effect of Changes in Smoking and Obesity on Future Life Expectancy in the United States”, *Demography*, 51(1): 27–49.

Ramachandran, A., Snehalatha, C., Kapur, A., Vijay, V., Mohan, V., Das, A. K., Rao, P. V., Yajnik, C. S., Kumar, P. K. M. and Nair, J. D. (2001), “High Prevalence of Diabetes and Impaired Glucose Tolerance in India: National Urban Diabetes Survey”, *Diabetologia*, 44(9): 1094–1101.

Ramachandran, A., Snehalatha, C., Yamuna, A., Mary, S. and Ping, Z. (2007), “Cost-Effectiveness of the Interventions in the Primary Prevention of Diabetes among Asian Indians: Within-Trial Results of the Indian Diabetes Prevention Programme (IDPP)”, *Diabetes Care*, 30(10): 2548-52.

Rashad, I. (2006), “Structural Estimation of Calorie Intake, Exercise, Smoking and Obesity”, *National Bureau of Economic Research* Working Paper 11957.

Razak, F., Anand, S. S., Shannon, H., Vuksan, V., Davis, B., Jacobs, R., Teo, K. K., McQueen, M. and Yusuf, S. (2007), “Defining Obesity Cut Points in a Multiethnic Population”, *Circulation*, 115(16): 2111-8.

Reinhardt, U. E. (2001), “Perspectives on the Pharmaceutical Industry”, *Health Affairs*, 20: 136-49.

Reserve Bank of India (2015), *Handbook of Statistics on Indian Economy*. URL: https://www.rbi.org.in/Scripts/AnnualPublications.aspx?head=Handbook%20of%20Statistics%20on%20Indian%20Economy

Reserve Bank of India (2015), *Handbook of Statistics on Indian States 2015*. URL: https://www.rbi.org.in/Scripts/AnnualPublications.aspx?head=Handbook%20of%20Statistics%20on%20Indian%20States

Roodman, D. (2009), “How to do xtabond2: An Introduction to Difference and System GMM in Stata”, *The Stata Journal*, 9(1): 86-136.

Rowley, W. R., Bezold, C., Arikan, Y., Byrne, E. and Krohe, S. (2017), “Diabetes 2030: Insights from Yesterday, Today, and Future Trends”, *Population Health Management*, 20(1): 6-12.

Schmidhuber, J. and Shetty, P. (2005), “The Nutrition Transition to 2030. Why Developing Countries are Likely to Bear the Major Burden”, *Acta Agriculturae Scandinavica Section C*, 2(3-4): 150-166.

Sengupta, A., Angeli, F., Syamala, T. S., Dagnelie, P. C. and Schayck, C. P. V. (2015), “Overweight and Obesity Prevalence among Indian Women by Place of Residence and Socio-Economic Status: Contrasting Patterns from 'Underweight States' and 'Overweight States' of India”, *Social Science and Medicine*, 38:161-69.

Sepp, E., Kolk, H., Loivukene, K. and Mikelsaar, M. (2014), “Higher Blood Glucose Level Associated with Body Mass Index and Gut Microbiota in Elderly People”, *Microbial Ecology in Health and Disease*, 25:10.3402/mehd.v25.22857. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4048595/

She, Z., King, D. M. and Jacobson, S. H. (2017), “Analyzing the Impact of Public Transit Usage on Obesity”, *Preventive Medicine*, 99: 264-268.

Sikdar, K. C., Wang, P. P., MacDonald, D. and Gadag, V. G. (2010), “Diabetes and its Impact on Health-Related Quality of Life: A Life Table Analysis”, *Quality of Life Research*, 19(6): 781-787.

Spanier, P. A., Marshall, S. J. and Faulkner, G. E. (2006), “Tackling the Obesity Pandemic: A Call for Sedentary Behaviour Research”, *Canadian Journal of Public Health*, 97(3): 255-257.

Staiger, D. and Stock, J. H. (1997), “Instrumental Variables Regression with Weak Instruments”, *Econometrica*, 65(3): 557-586.

Stewart, S. T., Cutler, D. M. and Rosen, A. B. (2009), “Forecasting the Effects of Obesity and Smoking on U.S. Life Expectancy”, *The New England Journal of Medicine*, 361: 2252-60.

Teufel-Shone, N. I., Jiang, L., Beals, J., Henderson, W. G., Zhang, L., Acton, K. J., Roubideaux, Y. and Manson, S. M. (2014), “Demographic Characteristics and Food

Choices of Participants in the Special Diabetes Program for American Indians Diabetes Prevention Demonstration Project", *Ethnicity and Health*, 20(4): 327-340.

Thornton, J. A. and Rice, J. L. (2008), "Determinants of Healthcare Spending: A State Level Analysis", *Applied Economics*, 40: 2873-89.

Thorpe K. E., Florence, C. S., Howard D. H. and Joski, P. (2004), "The Impact of Obesity on Rising Medical Spending", *Health Affairs*, w4: 480-486.

Tiemann, T. K., Miller, P. and Lamanna, E. (2008), "Active Transport, Public Transportation, and Obesity in Metropolitan Areas of the United States", *Elon University Working Paper*-07.

Traill, W. B. and Mazzocchi, M. (2005), "The Economics and Policy of Diet and Health", *Acta Agriculturae Scandinavica Section C*, 2(3-4): 113-116.

Vandegrift, D. and Datta, A. (2006), "Prescription Drug Expenditures in the United States: The Effects of Obesity, Demographics and New Pharmaceutical Products", *Southern Economic Journal*, 73(2): 515-529.

Variyam, J. N., Blaylock, J. and Smallwood, D. (1999), "Information, Endogeneity, and Consumer Health Behaviour: Application to Dietary Intakes", *Applied Economics*, 31: 217-226.

Withrow, D. and Alter, D. A. (2011), "The Economic Burden of Obesity Worldwide: A Systematic Review of the Direct Costs of Obesity", *Obesity Reviews*, 12(2): 131-41.

Wooldridge, J. M. (2001), "Applications of Generalized Method of Moments Estimation", *The Journal of Economic Perspectives*, American Economic Association, 15(4): 87-100.

Wooldridge, J.M. (2006). Introductory Econometrics: A Modern Approach. 3rd Edition, Mason: Thomson/South-Western.

World Bank (2016). *Cause of Death, By Non-Communicable Diseases*. URL: https://data.worldbank.org/indicator/SH.DTH.NCOM.ZS?end=2016&name_desc=true&start=2000

World Bank (2017). *World Bank Country and Lending Groups Classification.* URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

World Health Organization (2000). The Asia-Pacific Perspective: Redefining Obesity and its Treatment. *International Association for the Study of Obesity.* URL: http://www.wpro.who.int/nutrition/documents/docs/Redefiningobesity.pdf

World Health Organization (2015). *Global Health Observatory Data Repository*. URL: http://apps.who.int/gho/data/node.home

World Health Organization (2016). *Global Health Observatory Data Repository*. URL: http://apps.who.int/gho/data/node.home

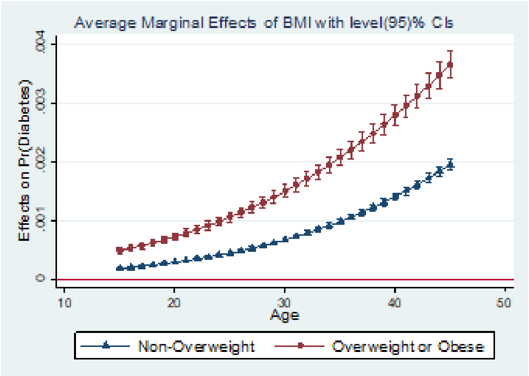
World Obesity Federation (2018), *Global Obesity Observatory*. URL: https://www.worldobesitydata.org/

Yesudian, C. A. K., Grepstad, M., Visintin, E. and Ferrario, A. (2014), "The Economic Burden of Diabetes in India: A Review of the Literature", *Global Health*, 10: 80.

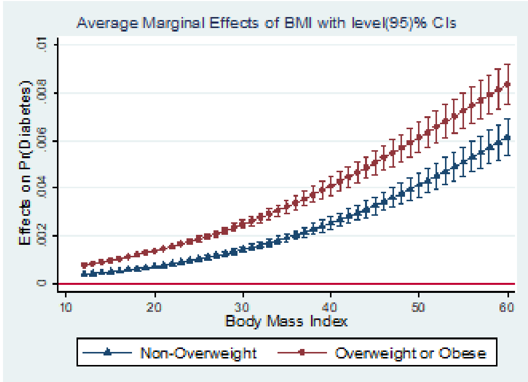
Young, L. R. and Nestle, M. (2002), "The Contribution of Expanding Portion Sizes to the US Obesity Epidemic", *American Journal of Public Health*, 92(2): 246-49.

Appendix

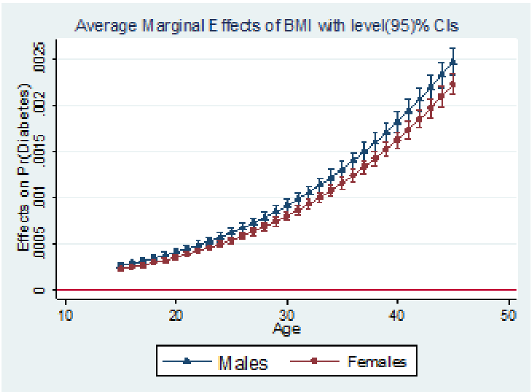
(1)
ME across Non-Overweight and Overweight or Obese Population w.r.t. Age



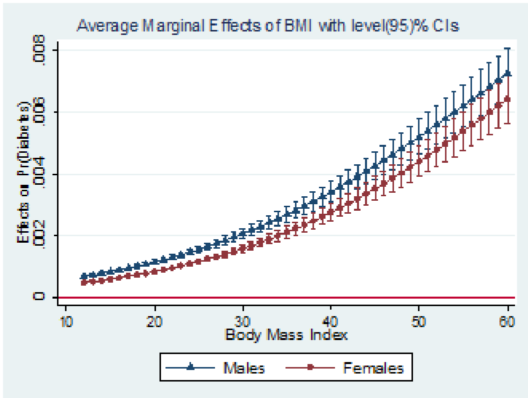
(2)
ME across Non-Overweight and Overweight or Obese Population w.r.t. BMI



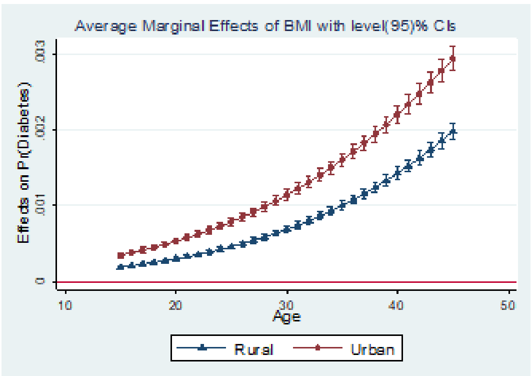
(3)
ME across Males and Females w.r.t. Age



(4)
ME across Males and Females w.r.t. BMI



(5)
ME across Rural and Urban Population w.r.t. Age



(6)
ME across Rural and Urban Population w.r.t. BMI

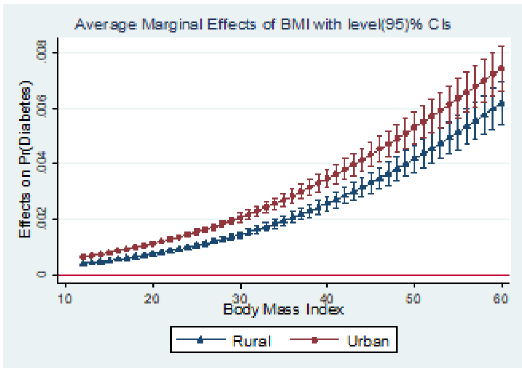


Figure A.5.1: Margins Plot for the Effect of BMI on the Self-Reported Diabetes Status based on Probit Model Estimates for the Full Sample

Source: Author's calculations based on NFHS data.
Note: ME denotes average marginal effect of BMI on the self-reported diabetes status and CIs indicates 95% confidence intervals. These graphs are based on the Probi model estimates for the full sample data which includes individuals belonging to all BMI categories.

Table A.4.1: List of States included in the Analysis, The United States

Alabama	Indiana	Nebraska	South Carolina
Alaska	Iowa	Nevada	South Dakota
Arizona	Kansas	New Hampshire	Tennessee
Arkansas	Kentucky	New Jersey	Texas
California	Louisiana	New Mexico	Utah
Colorado	Maine	New York	Vermont
Connecticut	Maryland	North Carolina	Virginia
Delaware	Massachusetts	North Dakota	Washington
Florida	Michigan	Ohio	West Virginia
Georgia	Minnesota	Oklahoma	Wisconsin
Hawaii	Mississippi	Oregon	Wyoming
Idaho	Missouri	Pennsylvania	
Illinois	Montana	Rhode Island	

Table A.4.2: Data Sources, The United States

Variables	Source
Life Expectancy at Birth	Global Health Data Exchange (GHDx), Institute for Health Metrics and Evaluation (IHME), University of Washington.
Obesity Prevalence, Educational Attainment – High School, Tobacco Consumption or Smoking, Alcohol Consumption, Income $50,000 or Above, Activity Status or Exercise, Age 65 years or Above and Cholesterol Checked[60]	Behavioral Risk Factor Surveillance System (BRFSS).
Per Capita Total Health Expenditure	Centers for Medicare and Medicaid Services (CMS).
Per Capita Real GDP	Bureau of Economic Analysis.

[60] For the variable on cholesterol checked, data is available only for years 2001, 2003, 2005, 2007, 2009, 2011 and 2013.

Table A.4.3: List of States included in the Analysis, India

Andhra Pradesh[61]	Haryana	Madhya Pradesh	Uttar Pradesh
Assam	Himachal Pradesh,	Maharashtra	Uttarakhand
Bihar	Jammu and Kashmir	Odisha	West Bengal
Chhattisgarh	Jharkhand	Punjab	
Delhi	Karnataka	Rajasthan	
Gujarat	Kerala	Tamil Nadu	

[61] Data for undivided Andhra Pradesh has been used in this study. In year 2014, the State of Andhra Pradesh was divided into two states, namely, Andhra Pradesh and Telangana. The data was calculated using weighted average wherever the it was available for new partitioned Andhra Pradesh and Telangana.

Table A.4.4: Data Sources, India

Variables	Source
Life Expectancy at Birth[62]	Life tables of Sample Registration System published by Ministry of Home Affairs, Government of India.
Overweight and Obesity Prevalence, Literacy Rate, Tobacco Consumption or Smoking, Alcohol Consumption, Television, Mobile, Computer and Car.	State reports of National Family Health Survey round - 3 and 4 for years 2005-06 and 2015-16 published by Ministry of Health and Family Welfare, Government of India.
Per Capita Total Health Expenditure	National Health Accounts Estimates for India published by National Health Accounts Cell and National Health Accounts, Technical Secretariat, National Health Systems Resource Centre, Ministry of Health and Family Welfare, Government of India.
Per Capita Net State Domestic Product and Total Food Grains Production	Handbook of Statistics on Indian States published by Reserve Bank of India, Government of India.
Head Count Ratio	Handbook of Statistics on Indian Economy published by Reserve Bank of India, Government of India.
Gini Coefficient[63]	State-wise Indicators of Poverty, Planning Commission, Government of India.
Monthly Per Capita Consumer Expenditure[64]	Key Indicators of Household Consumer Expenditure in India, NSS 68th Round (July 2011 – June 2012) and Household Consumer Expenditure in India, NSS 60th Round (January – June 2004) published by National Sample Survey Office, Ministry of Statistics and Programme Implementation, Government of India.

[62] Life expectancy at birth is available for a period of 5 years interval rather than annually, therefore, we use life expectancy at birth for year 2011-15 as an estimate for the year 2015. Similarly, data for the year 2001-05 is used as an estimate for year the 2005 and is extracted from 'Compendium of India's Fertility and Mortality Indicators,1971 – 2013' published by Ministry of Home Affairs, Government of India. Life expectancy for undivided Andhra Pradesh is considered. Life expectancy estimates for year 2001-05 of Bihar and Madhya Pradesh includes Jharkhand and Chhattisgarh respectively.

[63] Estimates for year 2004-05 and 2009-10 are used.

[64] "Uniform Reference Period MPCE (or MPCE URP) is the measure of MPCE obtained by the NSS consumer expenditure survey (CES) when household consumer expenditure on each item is recorded for a reference period of 'last 30 days' (preceding the date of survey). MPCE is based on total expenditure on food and non-food items" (NSSO).

Table A.4.5: Effect of Overweight and Obesity Prevalence on Life Expectancy at Birth, Robustness Check - IV-2SLS Estimates

Variables	(1)	(2)	(3)	(4)
Overweight and Obesity Prevalence (in %)	0.262# (0.164)	0.310*** (0.102)	0.326*** (0.111)	0.328*** (0.111)
Overweight and Obesity Prevalence Square (in %)	-0.005** (0.002)	-0.006*** (0.001)	-0.006*** (0.002)	-0.006*** (0.002)
Per Capita Health Expenditure (in Rs.)	0.0003 (0.001)	0.001 (0.001)	0.001 (0.001)	
Literacy Rate (in %)	0.182*** (0.038)	0.174*** (0.034)	0.194*** (0.033)	0.195*** (0.035)
Smoking (in %)	0.066* (0.038)	0.060 (0.044)	0.051 (0.036)	0.059 (0.042)
Alcohol Consumption (in %)	-0.109** (0.045)	-0.114*** (0.044)	-0.104** (0.041)	-0.134* (0.072)
Per capita NSDP (in Rs)	5.74e-06 (0.000)	7.91e-06 (0.000)	7.73e-06 (0.000)	0.00001 (0.000)
Gini Coefficient	4.340 (4.425)	4.419 (4.720)	2.673 (3.838)	2.802 (3.465)
Head Count Ratio (in %)	-0.050 (0.052)			
Total Food Grains Production (in Thousand Tonnes)		0.00007 (0.000)		
Per Capita Health Expenditure Trimmed at 5% (in Rs.)				0.0004 (0.001)
Observations	128	128	128	112
R^2	0.888	0.883	0.881	0.874
F Statistic	52.18	47.26	45.56	50.55
Prob > F	0.0000	0.0000	0.0000	0.0000
Kleibergen-Paap rk LM Statistic	9.508	10.237	6.194	7.593
Chi2 p-value	0.0020	0.0014	0.0128	0.0059
Kleibergen-Paap rk Wald F Statistic	10.556	15.083	5.547	15.643

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses. Standard errors are clustered at the state level.
[#] Weakly significant at 10.9%.

Table A.4.6: Effect of Overweight and Obesity Prevalence on Life Expectancy at Birth, Heterogeneity Analysis - First Stage Regressions' Estimates (for Model 1)

Variables	**Overweight and Obesity Prevalence**	**Overweight and Obesity Prevalence Square**	**Overweight and Obesity Prevalence * Female Dummy**	**Overweight and Obesity Prevalence Square * Female Dummy**
	(1)	**(2)**	**(3)**	**(4)**
Per Capita Health Expenditure (in Rs.)	0.002 (0.005)	0.146 (0.233)	0.0007 (0.003)	0.036 (0.151)
Literacy Rate (in %)	0.031 (0.251)	-1.560 (13.077)	0.012 (0.158)	1.019 (8.982)
Smoking (in %)	-0.012 (0.087)	-3.739 (4.862)	0.037 (0.039)	1.994 (2.448)
Alcohol Consumption (in %)	0.112 (0.134)	3.386 (7.137)	-0.041 (0.035)	-1.977 (2.159)
Per capita NSDP (in Rs.)	-0.0001 (0.0002)	-0.006 (0.008)	-0.00006 (0.00008)	-0.002 (0.005)
Gini Coefficient	19.258 (18.929)	1537.037 (1115.479)	7.085 (8.170)	733.053 (577.432)
Instruments				
Index	-0.026 (0.053)	-10.606*** (2.622)	-0.003 (0.017)	-0.216 (0.949)
Index Square	0.002*** (0.0005)	0.153*** (0.024)	0.0001 (0.0003)	0.006 (0.017)
Index * Female Dummy	0.118** (0.052)	4.568 (2.942)	0.106* (0.063)	-5.558 (3.666)
Index Square * Female Dummy	-0.001** (0.0006)	-0.041 (0.030)	0.0005 (0.0008)	0.103*** (0.038)
F Statistic	105.68	157.05	172.96	106.60
Prob > F	0.0000	0.0000	0.0000	0.0000

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses. Standard errors are clustered at the state level.
Note: Sanderson-Windmeijer multivariate F test of excluded instruments is reported.
Female Dummy takes value 1 for female gender and 0 for male gender.

Table A.4.7: Effect of Overweight and Obesity Prevalence on Life Expectancy at Birth, Heterogeneity Analysis - First Stage Regressions' Estimates (for Model 2)

Variables	**Overweight and Obesity Prevalence**	**Overweight and Obesity Prevalence Square**	**Overweight and Obesity Prevalence * Female Dummy**	**Overweight and Obesity Prevalence Square * Female Dummy**
	(1)	**(2)**	**(3)**	**(4)**
Per Capita Health Expenditure (in Rs.)	0.001 (0.003)	0.097 (0.160)	0.0008 (0.002)	0.041 (0.111)
Literacy Rate (in %)	0.287 (0.172)	9.651 (8.253)	0.170 (0.108)	7.952 (6.003)
Smoking (in %)	0.001 (0.111)	-3.396 (5.941)	0.085 (0.058)	4.607 (3.384)
Alcohol Consumption (in %)	-0.042 (0.159)	-3.387 (7.587)	-0.163** (0.064)	-7.837** (3.515)
Monthly Per Capita Consumer Expenditure (in Rs.)	-0.006** (0.002)	-0.249** (0.119)	-0.004*** (0.001)	-0.205*** (0.064)
Gini Coefficient	28.546 (25.097)	1889.719 (1367.883)	14.214 (10.738)	0.064 (696.802)
Instruments				
Index	-0.139** (0.068)	-15.530*** (3.364)	-0.088** (0.034)	-4.131** (1.705)
Index Square	0.004*** (0.0006)	0.213*** (0.029)	0.034*** (0.0004)	0.061*** (0.019)
Index * Female Dummy	0.090 (0.060)	3.590 (3.340)	0.088 (0.066)	-6.204 (3.867)
Index Square * Female Dummy	-0.001** (0.0005)	-0.042 (0.029)	0.0005 (0.0007)	0.101*** (0.037)
F Statistic	137.30	134.53	151.34	118.86
Prob > F	0.0000	0.0000	0.0000	0.0000

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses. Standard errors are clustered at the state level.
Note: Sanderson-Windmeijer multivariate F test of excluded instruments is reported.
Female Dummy takes value 1 for female gender and 0 for male gender.

Table A.5.1: List of States and Union Territories included in the Analysis

Andaman and Nicobar Islands	Gujarat	Manipur	Tamil Nadu
Andhra Pradesh	Haryana	Meghalaya	Tripura
Arunachal Pradesh	Himachal Pradesh	Mizoram	Uttar Pradesh
Assam	Jammu and Kashmir	Nagaland	Uttarakhand
Bihar	Jharkhand	Delhi	West Bengal
Chandigarh	Karnataka	Odisha	Telangana
Chhattisgarh	Kerala	Puducherry	
Dadra and Nagar Haveli	Lakshadweep	Punjab	
Daman and Diu	Madhya Pradesh	Rajasthan	
Goa	Maharashtra	Sikkim	

Table A.5.2: Results for the Probit Model having Self-Reported Diabetes Status as the Dependent Variable

Variables	**Restricted Sample**	**Restricted Sample (for Individuals having BMI ≥ 25 kg/m^2)**	**Full Sample**
	(1)	**(2)**	**(3)**
Body Mass Index	0.033***	0.027***	0.032***
	(0.003)	(0.006)	(0.001)
Age	0.033***	0.046***	0.034***
	(0.002)	(0.004)	(0.001)
Gender (Base = Male)			
Female	-0.201	-0.310	-0.015
	(0.178)	(0.302)	(0.013)
Education (Base = No Education or Preschool)			
Primary	0.060	0.029	0.082***
	(0.047)	(0.093)	(0.014)
Secondary	0.144***	0.080	0.108***
	(0.041)	(0.080)	(0.012)
Higher	0.203***	0.067	0.038**
	(0.054)	(0.096)	(0.017)
Bank Account	0.101*	0.306**	-0.032*
	(0.059)	(0.132)	(0.017)
Wealth Quintile (Base = Poorest)			
Poorer	0.018	0.222	0.022
	(0.053)	(0.145)	(0.017)
Middle	0.073	0.194	0.067***
	(0.054)	(0.141)	(0.017)
Richer	0.190***	0.359**	0.174***
	(0.057)	(0.141)	(0.018)
Richest	0.322***	0.422***	0.260***
	(0.063)	(0.146)	(0.020)
Religion (Base = Hindu)			
Muslim	0.074	-0.015	0.120***
	(0.046)	(0.077)	(0.013)
Christian	-0.085	-0.090	-0.012
	(0.070)	(0.116)	(0.022)
Sikh	-0.043	-0.180	-0.003
	(0.119)	(0.158)	(0.036)
Buddhist/Neo-Buddhist	0.073	-0.010	-0.065
	(0.122)	(0.215)	(0.042)
Jain	0.151	-0.042	0.037
	(0.259)	(0.373)	(0.085)
Jewish			
Parsi/Zoroastrian			
No religion	0.422		0.248*
	(0.538)		(0.139)

Table A.5.2 (Continued)			
Variables	**(1)**	**(2)**	**(3)**
Other	0.164	0.298	0.028
	(0.132)	(0.258)	(0.045)
Scheduled Caste	0.016	-0.023	0.020
	(0.044)	(0.072)	(0.013)
Scheduled Tribe	-0.149***	-0.237**	-0.051***
	(0.056)	(0.110)	(0.018)
Other Backward Classes	-0.106***	-0.156***	-0.055***
	(0.037)	(0.058)	(0.011)
Insurance	0.001	0.013	0.021**
	(0.032)	(0.053)	(0.010)
Below Poverty Line	-0.021	-0.024	0.004
	(0.033)	(0.061)	(0.010)
Family Structure (Base = Nuclear)	-0.032	-0.017	0.010
	(0.033)	(0.055)	(0.009)
Number of Household Members	-0.006	0.026*	-0.006***
	(0.009)	(0.015)	(0.002)
Region (Base = Rural)	-0.002	0.074	0.049***
	(0.032)	(0.052)	(0.010)
Smokes Cigarette	0.057	0.081	0.082***
	(0.037)	(0.065)	(0.025)
Smoke Pipes	-0.118	0.107	0.172
	(0.247)	(0.468)	(0.117)
Chews Tobacco	-0.034	-0.028	0.004
	(0.078)	(0.141)	(0.037)
Snuffs	0.651**	0.288	0.298***
	(0.255)	(0.552)	(0.090)
Smoke Cigar	0.222	-0.137	0.277***
	(0.135)	(0.276)	(0.078)
Chews Paan or Gutkha	-0.091**	-0.124	0.012
	(0.042)	(0.081)	(0.020)
Chews Paan with Tobacco	0.116**	0.163*	0.052***
	(0.048)	(0.085)	(0.020)
Drinks Alcohol	0.047	-0.010	0.022
	(0.031)	(0.053)	(0.018)
Eats fried food daily or weekly	0.040	0.011	0.011
	(0.029)	(0.050)	(0.009)
Drinks aerated drink daily or weekly	-0.103***	-0.119**	-0.025**
	(0.033)	(0.054)	(0.010)
Constant	-4.078***	-4.970***	-3.900***
	(0.202)	(0.377)	(0.059)
State Fixed Effects	Yes	Yes	Yes
Observations	43,202	9,622	776,394

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses.
Notes: Based on these Probit model estimates, we compute average marginal effects of BMI on self-reported diabetes status.
Definition of all the variables included in the above models is given in Table 5.1.

Table A.5.3: Results for the IV-Probit Model having Self-Reported Diabetes Status as the Dependent Variable

Variables	Restricted Sample		Restricted Sample (for Individuals having BMI $\geq$ 25 kg/m^2)	
	Self-Reported Diabetes	First Stage Regression for BMI	Self-Reported Diabetes	First Stage Regression for BMI
	(1)	(2)	(3)	(4)
Body Mass Index	0.097***		0.150***	
	(0.016)		(0.058)	
Age	0.031***	0.006**	0.043***	-0.004
	(0.003)	(0.003)	(0.005)	(0.003)
Gender (Base = Male)				
Female	-0.196	-0.057	-0.373	0.699**
	(0.269)	(0.290)	(0.342)	(0.275)
Education (Base = No Education or Preschool)				
Primary	0.038	0.261***	0.016	0.093
	(0.042)	(0.059)	(0.065)	(0.105)
Secondary	0.103***	0.470***	0.060	0.096
	(0.036)	(0.055)	(0.065)	(0.084)
Higher	0.137**	0.812***	0.057	0.011
	(0.067)	(0.092)	(0.078)	(0.086)
Bank Account	0.092	0.129**	0.281**	0.044
	(0.069)	(0.050)	(0.124)	(0.153)
Wealth Quintile (Base = Poorest)				
Poorer	-0.017	0.441***	0.210	-0.066
	(0.058)	(0.049)	(0.142)	(0.169)
Middle	-0.016	1.110***	0.162	0.055
	(0.064)	(0.070)	(0.115)	(0.148)
Richer	0.040	1.827***	0.292**	0.174
	(0.079)	(0.093)	(0.137)	(0.125)
Richest	0.104	2.640***	0.316**	0.405***
	(0.101)	(0.096)	(0.161)	(0.130)
Religion (Base = Hindu)				
Muslim	0.055*	0.170*	-0.043	0.195*
	(0.029)	(0.098)	(0.082)	(0.110)
Christian	-0.097	0.211*	-0.128	0.363**
	(0.083)	(0.112)	(0.107)	(0.166)
Sikh	-0.116	0.945***	-0.230	0.446
	(0.150)	(0.291)	(0.166)	(0.339)
Buddhist/Neo-Buddhist	0.032	0.513***	-0.065	0.403***
	(0.072)	(0.179)	(0.105)	(0.154)
Jain	0.105	0.627*	-0.000	-0.236
	(0.162)	(0.327)	(0.386)	(0.320)
Jewish				
Parsi/Zoroastrian				
No religion	0.280	1.860**	-4.509***	3.291**
	(0.612)	(0.883)	(0.248)	(1.635)
Other	0.153*	0.056	0.294	-0.137
	(0.085)	(0.179)	(0.287)	(0.264)

Table A.5.3 (Continued)				
Variables	**(1)**	**(2)**	**(3)**	**(4)**
Scheduled Caste	0.032	-0.208***	0.024	-0.332***
	(0.048)	(0.078)	(0.070)	(0.116)
Scheduled Tribe	-0.134**	-0.091	-0.189*	-0.206*
	(0.060)	(0.143)	(0.099)	(0.106)
Other Backward Classes	-0.088***	-0.215***	-0.107*	-0.289***
	(0.029)	(0.064)	(0.060)	(0.085)
Insurance	0.004	-0.033	0.003	0.067
	(0.034)	(0.064)	(0.050)	(0.104)
Below Poverty Line	-0.005	-0.177***	-0.020	-0.002
	(0.043)	(0.034)	(0.087)	(0.080)
Family Structure (Base = Nuclear)	-0.035	0.065	-0.003	-0.100
	(0.040)	(0.044)	(0.069)	(0.091)
Number of Household Members	-0.005	0.008	0.020	0.034
	(0.010)	(0.012)	(0.016)	(0.024)
Region (Base = Rural)	-0.018	0.152***	0.058	0.044
	(0.037)	(0.054)	(0.050)	(0.056)
Smokes Cigarette	0.062	-0.119*	0.076	-0.019
	(0.045)	(0.067)	(0.057)	(0.067)
Smoke Pipes	-0.082	-0.511*	0.057	0.219
	(0.314)	(0.288)	(0.514)	(0.877)
Chews Tobacco	-0.014	-0.305***	0.005	-0.246
	(0.061)	(0.080)	(0.146)	(0.153)
Snuffs	0.627	0.178	0.100	1.322
	(0.408)	(0.546)	(0.583)	(1.084)
Smoke Cigar	0.219*	-0.014	-0.070	-0.522*
	(0.123)	(0.142)	(0.247)	(0.267)
Chews Paan or Gutkha	-0.082	-0.073	-0.130*	0.121
	(0.051)	(0.071)	(0.071)	(0.092)
Chews Paan with Tobacco	0.106*	0.121	0.135*	0.151
	(0.060)	(0.088)	(0.077)	(0.118)
Drinks Alcohol	0.053*	-0.076*	-0.008	0.004
	(0.032)	(0.045)	(0.034)	(0.070)
Eats fried food daily or weekly	0.032	0.094**	-0.004	0.111
	(0.033)	(0.040)	(0.041)	(0.090)
Drinks aerated drink daily or weekly	-0.108***	0.109**	-0.117*	0.054
	(0.041)	(0.052)	(0.066)	(0.061)
Body Mass Index of Spouse		0.150***		0.053***
		(0.007)		(0.006)
Constant	-5.395***	19.084***	-8.095***	26.850***
	(0.327)	(0.241)	(1.314)	(0.242)
State Fixed Effects	Yes	Yes	Yes	Yes
Observations	43,202	43,202	9,711	9,711

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses.
Notes: Based on these IV-Probit model estimates, we compute average marginal effects of BMI on self-reported diabetes status.
Definition of all the variables included in the above models is given in Table 5.1.

Table A.5.4: Results for the Ordered Probit Model having Ordinal Blood Glucose levels as the Dependent Variable

Variables	Full Sample	Full Sample (for Individuals having BMI ≥ 25 kg/m^2)
	(1)	(2)
Body Mass Index	0.038***	0.036***
	(0.001)	(0.001)
Age	0.031***	0.037***
	(0.000)	(0.001)
Gender (Base = Male)		
Female	-0.162***	-0.164***
	(0.008)	(0.015)
Education (Base = No Education or Preschool)		
Primary	0.030***	0.002
	(0.008)	(0.016)
Secondary	0.006	-0.023*
	(0.007)	(0.014)
Higher	-0.069***	-0.091***
	(0.011)	(0.018)
Bank Account	-0.001	-0.007
	(0.010)	(0.021)
Wealth Quintile (Base = Poorest)		
Poorer	0.017*	0.066**
	(0.009)	(0.026)
Middle	0.045***	0.114***
	(0.009)	(0.025)
Richer	0.092***	0.184***
	(0.010)	(0.025)
Richest	0.101***	0.196***
	(0.012)	(0.027)
Religion (Base = Hindu)		
Muslim	0.064***	0.069***
	(0.008)	(0.014)
Christian	-0.036**	0.003
	(0.015)	(0.026)
Sikh	0.004	-0.045
	(0.023)	(0.034)
Buddhist/Neo-Buddhist	-0.128***	-0.178***
	(0.025)	(0.045)
Jain	0.070	-0.029
	(0.057)	(0.086)
Jewish	-3.253***	-3.493***
	(0.117)	(0.205)
Parsi/Zoroastrian	0.003	0.021
	(0.176)	(0.393)
No religion	0.015	0.084
	(0.101)	(0.182)
Other	-0.005	0.048
	(0.025)	(0.050)
Scheduled Caste	0.008	0.052***
	(0.008)	(0.015)
Scheduled Tribe	-0.003	-0.041*
	(0.010)	(0.022)
Other Backward Classes	-0.004	-0.007
	(0.007)	(0.012)

Table A.5.4 (Continued)		
Variables	**(1)**	**(2)**
Insurance	0.002	0.014
	(0.006)	(0.011)
Below Poverty Line	0.000	-0.013
	(0.006)	(0.012)
Family Structure (Base = Nuclear)	0.014**	-0.005
	(0.006)	(0.011)
Marital Status (Base = Unmarried)	-0.105***	-0.060***
	(0.008)	(0.020)
Number of Household Members	-0.002	0.003
	(0.001)	(0.002)
Region (Base = Rural)	0.001	0.028***
	(0.006)	(0.011)
Smokes Cigarette	-0.009	0.017
	(0.015)	(0.030)
Smoke Pipes	-0.205**	-0.322*
	(0.094)	(0.193)
Chews Tobacco	0.037*	0.074*
	(0.020)	(0.042)
Snuffs	0.009	0.062
	(0.066)	(0.119)
Smoke Cigar	0.060	-0.023
	(0.058)	(0.120)
Chews Paan or Gutkha	0.032***	0.053**
	(0.011)	(0.023)
Chews Paan with Tobacco	0.017	0.032
	(0.012)	(0.023)
Drinks Alcohol	-0.009	-0.021
	(0.011)	(0.021)
Eats fried food daily or weekly	-0.012**	-0.020*
	(0.005)	(0.010)
Drinks aerated drink daily or weekly	-0.027***	-0.026**
	(0.006)	(0.011)
Time since last ate	-0.065***	-0.060***
	(0.001)	(0.002)
Time since last drank	0.001***	0.001*
	(0.000)	(0.000)
$\boldsymbol{\mu_1}$	3.053***	3.243***
	(0.042)	(0.078)
$\boldsymbol{\mu_2}$	3.874***	3.967***
	(0.042)	(0.078)
State Fixed Effects	Yes	Yes
Observations	748,995	135,630

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses.
Notes: Based on these Ordered Probit model estimates, we compute average marginal effects of BMI on ordinal blood glucose levels.
Definition of all the variables included in the above models is given in Table 5.1.

Table A.6.1: Effect of Living in the Urban Areas on Body Mass Index – OLS and IV-2SLS Estimates

Variables	OLS (1)	IV-2SLS (2)	OLS (3)	IV-2SLS (4)	OLS (5)	IV-2SLS (6)
Urban Residence	0.446***	1.773**	0.440***	1.712**	0.295***	1.397*
	(0.011)	(0.803)	(0.012)	(0.850)	(0.021)	(0.789)
Age (Base = 15-24 years)						
25-34 years	1.342***	1.317***	1.345***	1.326***	1.429***	1.421***
	(0.012)	(0.063)	(0.012)	(0.059)	(0.023)	(0.063)
35-44 years	2.325***	2.301***	2.337***	2.321***	2.224***	2.214***
	(0.014)	(0.088)	(0.014)	(0.083)	(0.028)	(0.081)
45 years and above	2.547***	2.543***	2.563***	2.564***	2.277***	2.276***
	(0.018)	(0.098)	(0.018)	(0.095)	(0.033)	(0.086)
Gender (Base = Male)						
Female	0.109***	0.116	0.117***	0.112	0.181***	0.175**
	(0.012)	(0.078)	(0.013)	(0.075)	(0.020)	(0.086)
Education (Base = No Education or Preschool)						
Primary	0.413***	0.405***	0.376***	0.379***	0.323***	0.326***
	(0.014)	(0.047)	(0.015)	(0.044)	(0.028)	(0.051)
Secondary	0.664***	0.670***	0.577***	0.595***	0.477***	0.518***
	(0.012)	(0.061)	(0.012)	(0.060)	(0.024)	(0.071)
Higher	0.863***	0.818***	0.712***	0.677***	0.766***	0.785***
	(0.018)	(0.069)	(0.019)	(0.069)	(0.034)	(0.079)
Marital Status (Base = Unmarried)	1.139***	1.179***	1.140***	1.166***	0.999***	1.022***
	(0.012)	(0.074)	(0.012)	(0.067)	(0.023)	(0.057)
Wealth Quintile (Base = Poorest)						
Poorer	0.592***	0.519***	0.428***	0.358***	0.396***	0.336***
	(0.012)	(0.093)	(0.013)	(0.095)	(0.024)	(0.102)
Middle	1.279***	1.042***	0.989***	0.758***	0.909***	0.720***
	(0.013)	(0.207)	(0.015)	(0.218)	(0.028)	(0.192)
Richer	1.996***	1.472***	1.584***	1.064***	1.466***	1.051***
	(0.015)	(0.381)	(0.018)	(0.411)	(0.033)	(0.348)
Richest	2.725***	1.900***	2.006***	1.171*	1.852***	1.168**
	(0.018)	(0.560)	(0.025)	(0.627)	(0.045)	(0.554)
Religion (Base = Hindu)						
Muslim	0.573***	0.400	0.605***	0.450*	0.537***	0.434**
	(0.014)	(0.267)	(0.014)	(0.247)	(0.025)	(0.204)
Christian	0.649***	0.617***	0.668***	0.667***	0.648***	0.637***
	(0.017)	(0.174)	(0.018)	(0.144)	(0.033)	(0.139)
Sikh	0.633***	0.971***	0.544***	0.843***	0.586***	0.808***
	(0.033)	(0.244)	(0.033)	(0.237)	(0.061)	(0.197)
Buddhist/Neo-Buddhist	0.996***	1.051*	0.947***	1.021*	1.123***	1.179***
	(0.037)	(0.561)	(0.037)	(0.550)	(0.066)	(0.454)

Table A.6.1 (Continued)						
Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Jain	-0.133	-0.425	-0.158	-0.449	-0.417*	-0.655**
	(0.129)	(0.262)	(0.132)	(0.286)	(0.232)	(0.301)
Jewish	0.296	0.363	0.166	0.265	1.807***	2.067***
	(0.806)	(0.782)	(0.790)	(0.775)	(0.547)	(0.529)
Parsi/Zoroastrian	-0.079	-0.095	-0.104	-0.109	0.938	0.832
	(0.281)	(0.245)	(0.281)	(0.240)	(0.891)	(0.746)
No religion	1.222***	1.264***	1.236***	1.264***	1.513***	1.534***
	(0.166)	(0.318)	(0.168)	(0.245)	(0.439)	(0.595)
Other	0.728***	0.668**	0.726***	0.668**	0.753***	0.714**
	(0.035)	(0.333)	(0.035)	(0.309)	(0.066)	(0.349)
Scheduled Caste	-0.308***	-0.396***	-0.289***	-0.367***	-0.301***	-0.353***
	(0.014)	(0.099)	(0.014)	(0.094)	(0.026)	(0.079)
Scheduled Tribe	-0.425***	-0.443***	-0.407***	-0.423***	-0.354***	-0.378**
	(0.015)	(0.165)	(0.015)	(0.158)	(0.027)	(0.150)
Other Backward Classes	-0.306***	-0.365***	-0.284***	-0.348***	-0.230***	-0.292***
	(0.012)	(0.106)	(0.012)	(0.098)	(0.021)	(0.090)
Below Poverty Line	-0.144***	-0.118**	-0.148***	-0.115**	-0.125***	-0.097*
	(0.009)	(0.057)	(0.009)	(0.054)	(0.017)	(0.059)
Family Structure (Base = Nuclear)	0.015	0.086*	0.004	0.064	0.048***	0.095***
	(0.010)	(0.045)	(0.010)	(0.042)	(0.018)	(0.036)
Number of Household Members	-0.047***	-0.038***	-0.047***	-0.042***	-0.062***	-0.056***
	(0.002)	(0.012)	(0.002)	(0.010)	(0.004)	(0.009)
Smokes			-0.271***	-0.324***	-0.165***	-0.199***
			(0.014)	(0.102)	(0.021)	(0.060)
Drinks Alcohol			0.207***	0.200***	0.219***	0.211***
			(0.018)	(0.074)	(0.023)	(0.052)
Consumes Milk/curd Daily or Weekly			0.033***	0.085	0.112***	0.151**
			(0.009)	(0.064)	(0.018)	(0.060)
Consumes Dark Green Leafy Vegetables Daily or Weekly			0.096***	0.092	0.069***	0.077
			(0.012)	(0.063)	(0.022)	(0.055)
Consumes Fruits Daily or Weekly			0.206***	0.168***	0.235***	0.211***
			(0.010)	(0.045)	(0.018)	(0.043)
Consumes Fried Food Daily or Weekly			0.002	-0.014	-0.021	-0.037
			(0.009)	(0.047)	(0.016)	(0.040)
Consumes Aerated Drinks Daily or Weekly			0.008	-0.010	0.003	-0.007
			(0.011)	(0.037)	(0.019)	(0.038)
Household Assets (Base = Do not own)						
Mobile			0.265***	0.280***	0.315***	0.331***
			(0.017)	(0.066)	(0.032)	(0.075)

Table A.6.1 (Continued)

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Computer			0.037*	-0.038	0.049	-0.023
			(0.019)	(0.059)	(0.034)	(0.058)
Car/Truck			0.227***	0.368***	0.279***	0.403***
			(0.020)	(0.099)	(0.036)	(0.096)
Motorcycle/Scooter			0.056***	0.186	0.206***	0.302***
			(0.011)	(0.115)	(0.020)	(0.104)
Washing Machine			0.387***	0.369***	0.372***	0.362***
			(0.018)	(0.082)	(0.032)	(0.073)
Watch Television (Base = Not at all)						
Less than once a week			0.151***	0.131***	0.131***	0.123***
			(0.016)	(0.036)	(0.029)	(0.028)
At least once a week			0.199***	0.173***	0.222***	0.214***
			(0.015)	(0.041)	(0.027)	(0.030)
Almost every day			0.373***	0.291***	0.369***	0.317***
			(0.013)	(0.070)	(0.024)	(0.050)
Occupation (Base = Not working/no occupation)						
Professional/ Technical/Managerial					0.173***	0.131***
					(0.043)	(0.050)
Clerical					0.217***	0.154
					(0.078)	(0.108)
Sales					0.414***	0.283***
					(0.040)	(0.102)
Agricultural					-0.233***	-0.097
					(0.022)	(0.113)
Services					0.247***	0.147*
					(0.039)	(0.085)
Manual (Skilled and Unskilled)					-0.023	-0.106
					(0.026)	(0.069)
Constant	18.036***	17.916***	17.618***	17.534***	17.794***	17.650***
	(0.023)	(0.155)	(0.029)	(0.168)	(0.052)	(0.219)
Observations	779,519	779,519	764,454	764,454	219,781	219,781
R-squared	0.207	0.191	0.211	0.196	0.211	0.200
First Stage Regression						
F Statistic		25.46		22.72		17.91
Prob > F		0.0000		0.0000		0.0000

***, ** and * denote significance at 1%, 5% and 10% level.
Robust standard errors are reported in parentheses.
Definition of all the variables included in the above models is given in Tables 6.1 and 5.1.

Table A.6.2: Effect of Living in the Urban Areas on Overweight and Obesity Status – Linear Probability Model Estimates

Variables	(1)	(2)	(3)
Urban Residence	0.168**	0.164**	0.151*
	(0.074)	(0.083)	(0.081)
Age (Base = 15-24 years)			
25-34 years	0.092***	0.092***	0.093***
	(0.005)	(0.005)	(0.006)
35-44 years	0.180***	0.182***	0.169***
	(0.008)	(0.008)	(0.008)
45 years and above	0.209***	0.211***	0.185***
	(0.009)	(0.009)	(0.009)
Gender (Base = Male)			
Female	0.025***	0.020***	0.021***
	(0.005)	(0.005)	(0.006)
Education (Base = No Education or Preschool)			
Primary	0.036***	0.035***	0.032***
	(0.004)	(0.004)	(0.005)
Secondary	0.059***	0.054***	0.046***
	(0.005)	(0.005)	(0.006)
Higher	0.053***	0.043***	0.050***
	(0.006)	(0.006)	(0.006)
Marital Status (Base = Unmarried)	0.075***	0.074***	0.070***
	(0.005)	(0.004)	(0.004)
Wealth Quintile (Base = Poorest)			
Poorer	0.026***	0.016**	0.014*
	(0.007)	(0.007)	(0.008)
Middle	0.063***	0.043**	0.038**
	(0.018)	(0.020)	(0.019)
Richer	0.098***	0.068*	0.066*
	(0.035)	(0.039)	(0.036)
Richest	0.128**	0.074	0.069
	(0.054)	(0.061)	(0.058)
Religion (Base = Hindu)			
Muslim	0.026	0.029*	0.021
	(0.019)	(0.018)	(0.014)
Christian	0.002	0.006	0.008
	(0.013)	(0.011)	(0.012)
Sikh	0.070***	0.060***	0.057***
	(0.023)	(0.023)	(0.020)
Buddhist/Neo-Buddhist	0.054	0.054	0.073**
	(0.035)	(0.035)	(0.033)
Jain	-0.033	-0.035	-0.063**
	(0.022)	(0.025)	(0.030)
Jewish	0.026	0.020	0.028
	(0.071)	(0.072)	(0.049)
Parsi/Zoroastrian	0.020	0.020	0.137*
	(0.031)	(0.030)	(0.082)
No religion	0.072**	0.069***	0.103**
	(0.029)	(0.023)	(0.047)
Other	0.028	0.028	0.036
	(0.018)	(0.017)	(0.024)
Scheduled Caste	-0.029***	-0.027***	-0.025***
	(0.009)	(0.009)	(0.007)
Scheduled Tribe	-0.042***	-0.041***	-0.036***
	(0.010)	(0.010)	(0.009)
Other Backward Classes	-0.025***	-0.023***	-0.017**
	(0.009)	(0.008)	(0.008)

Table A.6.2 (Continued)

Variables	(1)	(2)	(3)
Below Poverty Line	-0.006	-0.006	-0.005
	(0.006)	(0.005)	(0.006)
Family Structure (Base = Nuclear)	0.006	0.004	0.006
	(0.004)	(0.004)	(0.004)
Number of Household Members	-0.002**	-0.002***	-0.003***
	(0.001)	(0.001)	(0.001)
Smokes		-0.027***	-0.022***
		(0.008)	(0.006)
Drinks Alcohol		0.003	0.007
		(0.006)	(0.006)
Consumes Milk/curd Daily or Weekly		0.001	0.007
		(0.006)	(0.005)
Consumes Dark Green Leafy Vegetables Daily or Weekly		0.007	0.007
		(0.005)	(0.005)
Consumes Fruits Daily or Weekly		0.012***	0.015***
		(0.004)	(0.004)
Consumes Fried Food Daily or Weekly		-0.003	-0.006
		(0.004)	(0.004)
Consumes Aerated Drinks Daily or Weekly		-0.003	-0.003
		(0.004)	(0.004)
Household Assets (Base = Do not own)			
Mobile		0.012***	0.013***
		(0.004)	(0.004)
Computer		-0.006	-0.008
		(0.005)	(0.007)
Car/Truck		0.034***	0.042***
		(0.010)	(0.011)
Motorcycle/Scooter		0.017*	0.028***
		(0.010)	(0.010)
Washing Machine		0.027***	0.030***
		(0.008)	(0.009)
Watch Television (Base = Not at all)			
Less than once a week		0.006***	0.009***
		(0.002)	(0.003)
At least once a week		0.011***	0.013***
		(0.003)	(0.003)
Almost every day		0.022***	0.024***
		(0.007)	(0.005)
Occupation (Base = Not working/no occupation)			
Professional/ Technical/Managerial			0.006
			(0.006)
Clerical			-0.005
			(0.011)
Sales			0.014
			(0.011)

Table A.6.2 (Continued)

Variables	(1)	(2)	(3)
Agricultural			-0.017
			(0.011)
Services			0.0002
			(0.009)
Manual (Skilled and Unskilled)			-0.024***
			(0.007)
Constant	-0.112***	-0.124***	-0.110***
	(0.010)	(0.011)	(0.016)
Observations	779,519	764,454	219,781
R-squared	0.118	0.122	0.118
First Stage Regression			
F Statistic	25.46	22.72	17.91
Prob > F	0.0000	0.0000	0.0002

***, ** and * denote significance at 1%, 5% and 10% level.

Robust standard errors are reported in parentheses.

In all models, we have used the total road length per Km of state area as an instrument for place of residence.

Definition of all the variables included in the above models is given in Tables 6.1 and 5.1.