Discussion Papers in Economics

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

Shivani Gupta and Sangeeta Bansal

Discussion Paper 19-04

Centre for International Trade and Development

School of International Studies

Jawaharlal Nehru University

India

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

Shivani Gupta* and Sangeeta Bansal**

June, 2019

Abstract: Diabetes prevalence has escalated in India during the recent decades. The recent nutritional transition brought about by the rapid emergence of overnutrition may have a relationship with the growing diabetes problem in India. This paper examines the causal effect of an increase in body mass index on the likelihood of suffering from diabetes using an individual level data from National Family Health Survey for the year 2015-16. The study uses two alternative indicators for measuring diabetes – self-reported diabetes status and blood glucose levels (ordinal measure). The novel contribution of the study is that it takes into account the role played by unobserved genetic and other related factors in the determination of the relationship between body mass index and diabetes status of an individual by instrumenting individual's body mass index with a non-biologically related household member's body mass index. The results show that the likelihood of being diabetic is three times among the overweight and obese individuals as compared to the non-overweight individuals.

JEL Classification: C35, C36, I12

Keywords: Body mass index, diabetes, instrumental variable, overweight and obesity, ordered probit model, probit model and IV-probit model.

* PhD Scholar, Centre for International Trade and Development, School of International Studies, Jawaharlal Nehru University, New Delhi-110067, India. Email: shivanisuccess22@yahoo.com
** Professor of Economics, Centre for International Trade and Development, School of International Studies, Jawaharlal Nehru University, New Delhi-110067, India. Email: sangeeta.bansal7@gmail.com

1 Introduction

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, India, people suffering from diabetes are likely to go up to 80 million by 2025, making India the 'Diabetes Capital' of the world. Analysing National Family Health Survey (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 United States find that the risk of diabetes is increasing in Body Mass Index (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 have a relationship with the growing diabetes problem in India. The broad objective of this study is to examine the effect of overnutrition on diabetes in India. Overnutrition is associated with the increased risk of mortality and co-morbidities (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 increased susceptibility towards non-communicable diseases (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 Report, WHO 2000).[1]

In this study, 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 at large in India while accounting for the potential endogeneity arising from the unobserved genetic and other related factors. 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 the 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. NFHS is a large-scale, multi-round survey conducted in a representative sample of households throughout India and provides data for the female population having age 15-49 years and the male population having age 15-54 years.

Studies such as Gray et al. 2011 and Sepp et al. 2014 have estimated the correlation between overnutrition and NCDs. Most studies are, however, based on the small sample size, and, these 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

[1] We examined the IDF and WHO recent estimates on the diabetes prevalence and obesity rates for adult population and found that some of the 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. 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.

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 Report, WHO 2000).[2]

Further, 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 South Delhi and show that an increase in BMI has a statistically significant impact on diabetes among married women in Delhi, India. The study by Ramachandran et al. (2001) finds a positive association between diabetes and BMI among urban population of India using the national urban diabetes survey.[3] Also, none of the studies has considered the effect of overnutrition on diabetes for India, using WHO Asian BMI classification which defines an individual having BMI $\geq$ 23 kg/m^2 as overweight or obese.

In this paper, our main dependent variable is diabetes status of an individual. We use two alternative measures for indicating diabetes status across population, 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 level gives us an advantage of estimating the effect of BMI on prediabetes as well. In addition, it also takes care of any measurement error in the self-reported diabetes status.

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, and 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

[2] 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 states that prevalence of abdominal obesity in South Asian population is around 69% which is much greater than 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).

[3] In this survey, data was collected from a representative sample in each of six major cities of India during January to August, 2000.

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, regions – rural and urban and different wealth quintiles.

The findings of this paper 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.[4] 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 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 increased health care costs (Cawley and Meyerhoefer 2012; Ramachandran et al. 2007 and Yesudian et al. 2014).[5]

The rest of the paper is organized as follows. Section 2 presents the conceptual framework and the methodology applied. Section 3 discusses the data set used along with definition of the variables considered in the analysis. This section also presents the descriptive statistics.

[4] 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 in people under the age of 60.

[5] 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 2018).

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 fugue is 119.4 US dollars per person (IDF). These low mean expenditures per person can be attributed to overall low levels of health expenditure among these countries. However, these estimates do highlight the huge economic burden associated with diabetes in the form of considerably high health care costs.

Section 4 presents the estimation results and their interpretation. Section 5 presents a discussion on the findings of the study. Finally, section 6 presents the concluding remarks.

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 caused by the changes in any of these factors 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 diabetic, that is, an individual may become prediabetic (defined in next paragraph) initially and later diabetic with 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. Also, as discussed in section 1, the risk of diabetes is expected to increase 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:[6]

[6] 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 also used to diagnose diabetes amongst individuals and it also based on the above defined categories.

(i) Less than or equal to140 mg/dl corresponds to low or moderate blood glucose – *Normal Blood Glucose Levels*

(ii) Between 141 and 200 mg/dl corresponds to high blood glucose – *Prediabetes*

(iii) Greater than 200 mg/dl corresponds to very high blood glucose – *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 health care 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 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 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. 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 classification of BMI which defines individuals having a BMI of 25 kg/m^2 or above as overweight or obese. Additionally, we also test our hypotheses using WHO Asian BMI classification, which defines individuals having a BMI of 23 kg/m^2 or above as overweight or obese.

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 is 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., increases 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 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. 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 Report, WHO 2000 and Bener et al. 2005).[7] Bener et al. (2005) find that the reported diabetes is higher among population with a family history of diabetes. Thus, it is important to account for the unobserved genetic and other related factors while establishing the link between BMI and diabetes. Therefore, we resort to an Instrument Variable Approach to address the potential endogeneity. Unobserved genetic and other related factors influence both diabetes as well as overweight or obesity status of an individual, thereby, causing endogeneity resulting from the omitted variable bias (OVB). Endogeneity issue is elaborated later in this section.

This study examines the micro level relationship between BMI and diabetes for population at large in India while accounting for the potential endogeneity arising from the unobserved genetic and other related factors thereby testing for the existence of any causal impact of overweight and obesity on diabetes. The novel contribution of the study is that it takes into account the role played by genetic factors and other related factors in determining the effect of BMI on diabetes status of an individual.

[7] 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.

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. The following model is estimated, having D_i^* as the dependent variable:

$$D_i^* = \beta' X_i + \nu_i \tag{1}$$

where,

$$D_i = \begin{cases} 0 \; if \; individual \; is \; non-diabetic, \\ 1 \; if \; individual \; is \; diabetic. \end{cases} \tag{2}$$

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 order ($k * 1$) of controls including BMI for i^{th} individual;[8]

ν_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. 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, $Cov\ (X_i, \nu_i) = 0$. We estimate the average marginal effects of BMI on self-reported diabetes status and examine their signs and magnitudes. We are interested in finding out 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:

$$\left[\frac{\partial P(D=1|\ X)}{\partial BMI};\ if\ BMI\ \geq 25\ kg/m^2\right] > \left[\frac{\partial P(D=1|\ X)}{\partial BMI};\ if\ BMI < 25\ kg/m^2\right] > 0 \tag{3}$$

[8] Coefficient of X_i, β, is a vector of order ($k * 1$).

that is, the change (or increase) in 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 OVB resulting from the unobserved genetic or 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 overcomes the endogeneity caused by both OVB 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 genetic and other related factors that explain variations in the diabetes status of an individual, i.e., the instrument must be uncorrelated with the error term:

$$Cov\ (BMI^S, \nu\ ) = 0 \tag{4}$$

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

$$Cov\ (BMI^S, BMI\ ) \neq 0 \tag{5}$$

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).

With an objective to measure the causal effect of the BMI on self-reported diabetes status of individual and to address the potential endogeneity problem, we consider equation (1) and estimate an IV-Probit model. The first stage equation for this model can be written as:

$$BMI_i = \delta_0 + \delta_1 BMI_i^s + \delta_2 x_i + \eta_i \tag{6}$$

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 this second 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^*:

$$BG_i^* = \alpha' X_i + \varepsilon_i \tag{7}$$

Blood glucose levels (dependent variable) are sorted into $j + 1$ categories, where j = 0, 1, 2:[9]

$$BG_i = \begin{cases} 0 \ \ if \ \ BG_i^* \leq \mu_1 \\ 1 \ \ if \ \mu_1 < BG_i^* \leq \mu_2 \\ 2 \ \ if \ \mu_2 < BG_i^* \end{cases} \tag{8}$$

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

$$P(\mu_j < BG_i^* \leq \mu_{j+1}) = \Phi(\mu_{j+1} - \alpha' X_i) - \Phi(\mu_j - \alpha' X_i) \tag{9}$$

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

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

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

a unit increase in BMI is higher among the overweight or obese individuals as compared to the non-overweight individuals:

$$\left[\frac{\partial P(BG=j|\,X)}{\partial BMI};\ if\ BMI\ \geq 25\ kg/m^2\right] > \left[\frac{\partial P(BG=j|\,X)}{\partial BMI};\ if\ BMI < 25\ kg/m^2\right] > 0\ ;\ j = 1, 2 \qquad (10)$$

3 Data and Descriptive Statistics

The study extracts individual level data from Demographic and Health Surveys (DHS) of India, namely, National Family Health Survey (NFHS). We consider the fourth round of NFHS for the year 2015-16. The NFHS is a large-scale, multi-round survey conducted in a representative sample of households throughout India. This survey has rich information on household characteristics, individual characteristics – 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. 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 variables included in the study along with their definitions is provided in Table 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 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.[10] This section presents the descriptive statistics for the full sample data.

[10] 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}.

Descriptive Statistics

Table 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 been 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 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 3 presents the descriptive statistics grouped by overweight and obesity status. We also report the mean difference across two groups with its statistical significance. The highlighting feature is 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 which is lower than the mean for the total sample (21.71 kg/m^2), and the mean BMI among overweight or obese individuals 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 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 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 the 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.

Now, we graphically analyse the relationship between self-reported diabetes status and BMI. Figure 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 1, the solid red line represents BMI distribution for the diabetic population while dash-dotted blue line represents the BMI distribution for 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 ≥ 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.

We also analyse the relationship between the blood glucose levels and BMI graphically. Figure 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 2) and then we plot three separate BMI distributions for each blood glucose category. In Figure 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 ≥ 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).[11] We may infer that amongst the population having BMI ≥ 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.[12]

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 level of an individual is the outcome of interest.

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 our IV-Probit model, we are using 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 6 presents the average marginal effects of BMI on the self-reported diabetes status for the sample data that is restricted to married couples. We estimate the average marginal effects for overweight or obese and non-overweight individuals. The results based on Probit and IV-Probit models' estimation are reported in Table 6. Based on estimated probit model, we compute the marginal effect of BMI on the self-reported diabetes status across overweight or

[11] We also plotted distributions for males and females separately and found similar observations for both males and females.

[12] 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%.

obese and non-overweight individuals. These marginal effects are reported for two classifications, first, for WHO International BMI classification in column (1) and, then, for WHO Asian BMI classification in column (2). Within 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.

In all the models, we include same set of controls so that the marginal effects can be compared across different BMI categories. We control for demographic and socio-economic variables for individual and household characteristics, behavioural risk factors, eating habits and state fixed effects. We report Wald chi2 test statistic for both Probit and IV-Probit models along with their P-values. 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. Here, 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 approximate guide for the quality our instrument. All the estimates are found to be robust to the inclusion or exclusion of controls. In Table 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 models are substantially higher than those for the corresponding Probit models indicating that correlation 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. Whereas 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 0.4% 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. We find that the marginal effects of BMI on the self-reported diabetes status differ significantly across non-overweight and overweight or obese individuals.

For both the IV-Probit model, the Wald test of exogeneity is rejected at 1% significance levels 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 considerably high value.

We also estimated the Probit model for the full sample data. These results are reported in Table 7. The presentation of results in done in similar fashion as explained for Table 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 within the overweight and obese individuals, which sections of the population are at a greater risk. 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 8 presents the results obtained from Probit and IV-Probit model, based on the restricted sample. The marginal effect of BMI on 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 models (2.6 times $\approx$ 3 times in IV model). The marginal effects across regions and wealth quintiles differ statistically significantly.

4.2 Effect of Body Mass Index on Prediabetes and Diabetes: Ordered Probit Model Estimates

Table 9 presents the average marginal effects of BMI on the ordinal blood glucose levels based on the Ordered Probit model estimation. For the estimated Ordered Probit model, we compute the average marginal effects of BMI on 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). 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.

We control for the demographic and socio-economic variables for individual and household characteristics, behavioural risk factors, eating habits and 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 $\mu_1 < \mu_2$. All the estimates are found to be robust to the inclusion or exclusion of controls. In Table 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 is highly statistically significant.

We now examine that within the overweight and obese individuals, which sections of the population are at a greater risk of prediabetes and diabetes. 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 10 presents these results. Males are at a slightly higher risk of being both prediabetic (0.5%) and diabetic (0.3%) compared to 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 rural. 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.

5 Discussion

The study finds that the overweight or obese individuals are more likely to be diabetic as well prediabetic as compared to the non-overweight individuals. These results are line with the studies by Sepp et a. (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 providing an evidence on 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 consistent across both indicators used for measuring for diabetes.

Also, 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 a rise in BMI or weight gain increases the risk of diabetes regardless of individual being overweight or not. However, the level of risk is expected vary with weight or BMI of 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 additionally plot the average marginal effects of BMI on diabetes for the full sample data to examine how does the marginal effects vary across different subgroups with age and BMI. Figure 3 illustrates the graphical plot of the average marginal effects of BMI on self-reported diabetes status. We plot the average marginal effect of BMI on self-reported diabetes status for 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 effect of BMI on self-reported diabetes status is 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 self-reported diabetes status increases with both age and BMI across all subgroups. Further research may examine the effects on overnutrition of other NCDs such as cardiovascular diseases, hypertension, etc. Researchers may also quantify the health care burden associated with diabetes.

6 Conclusion

Recognising the recently growing problem of overnutrition and diabetes in India, the study quantifies the causal effect of overweight and obesity on diabetes in India. The novel contribution of the study is that it addresses that potential endogeneity problem 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 and 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 correlation estimates highly understate the causal impact of the rise in BMI on diabetes. To best of our knowledge this is the first study that addresses the role played by unobserved genetic and other related factors in the relationship between BMI and diabetes using an instrumental variable approach in 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.

Present study faces some limitations. First, the causal effects can be generalised only for married couples living in the same household of whom either is the head of the family. However, our correlation estimates can be generalised for the population at large in India. Second, 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 or diabetic.

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, stokes, 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 70% of the total health expenditure is financed by households privately in the form of out of pocket health expenditures. It is also crucial to note that diabetes is not only restricted to urban areas but is also prevalent among rural areas and it is no longer a disease of 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 and there can be huge economic gains.

References

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

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

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

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

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

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.

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

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.

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.

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.

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

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.

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 Part V and Review of Course Topics. Stanford University. URL: https://web.stanford.edu/class/polisci100a/regress5.pdf

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.

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

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.

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

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.

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.

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.

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.

Nelson, M. C., Larsen, P. G., North, K. E. and Adair, L. S. (2012). 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

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.

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.

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.

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.

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.

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/

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.

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.

WHO, World Health Organization, Global Health Observatory Data Repository. URL: http://apps.who.int/gho/data/node.home

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

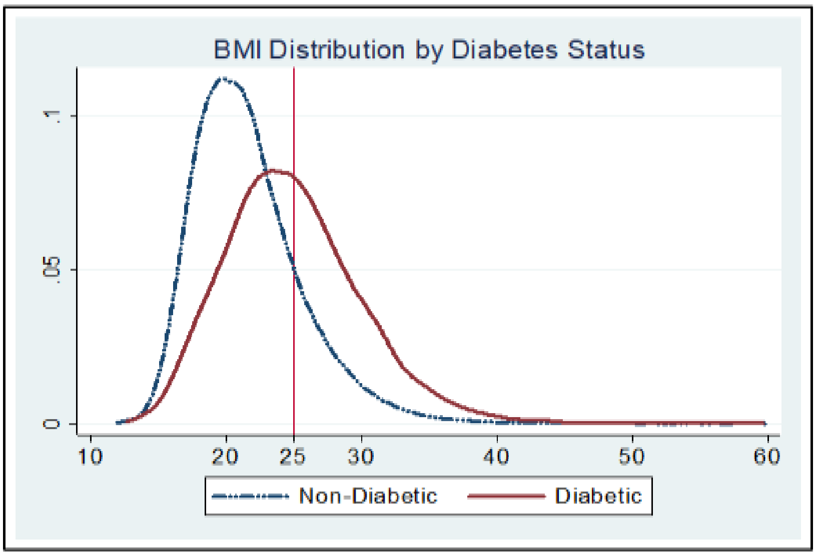
Wooldridge, J.M. (2006). Introductory Econometrics: A Modern Approach. 3rd Edition, Thomson/South-Western, Cengage Learning, Mason.

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

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.

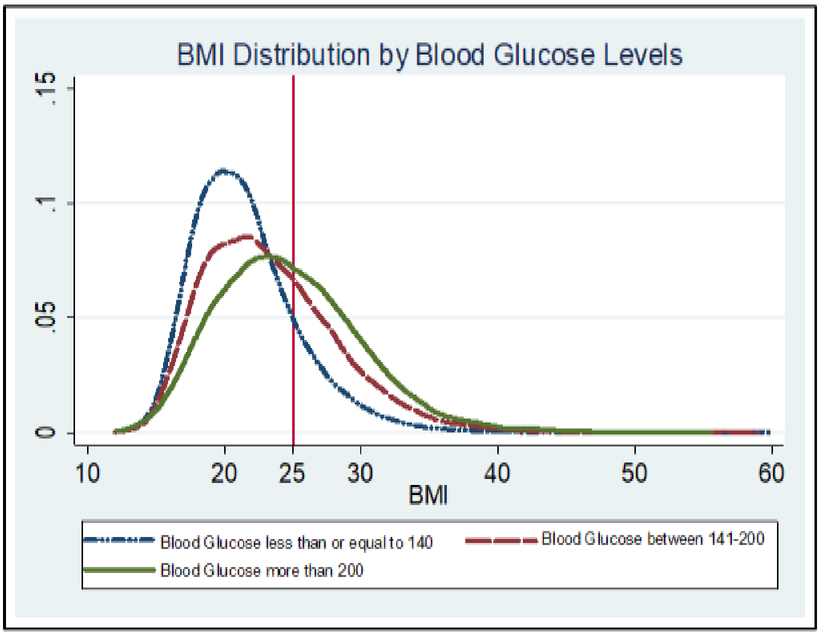
Appendix

Figure 1: BMI Distribution by Self-Reported Diabetes Status



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

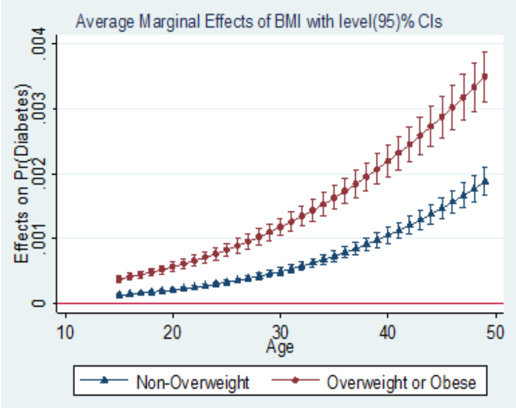
Figure 2: BMI Distribution by Blood Glucose Levels



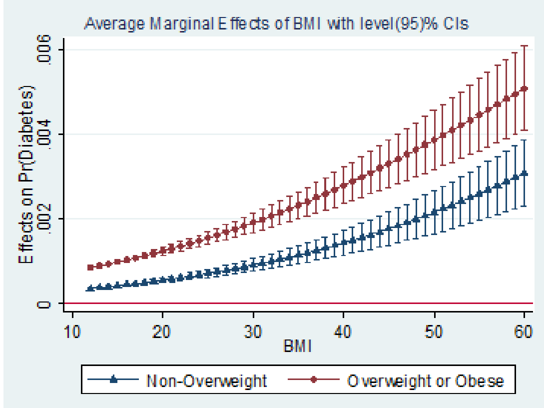
Source: Figure constructed by author based on NFHS data for year 2015-16.
Note: Blood glucose levels are measured in mg/dl.

Figure 3: Margins Plot for the Effect of BMI on the Self-Reported Diabetes Status

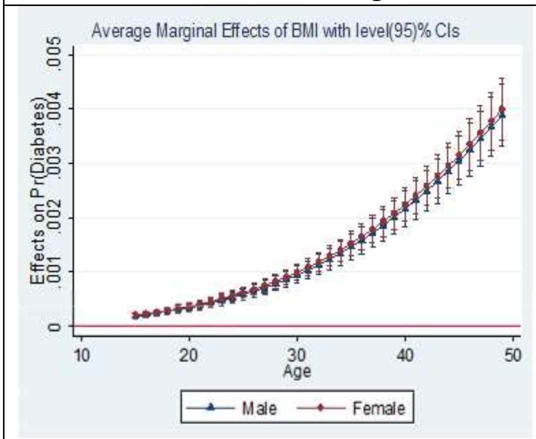
(1)
ME across Non-Overweight and Overweight or Obese Individuals w.r.t. Age



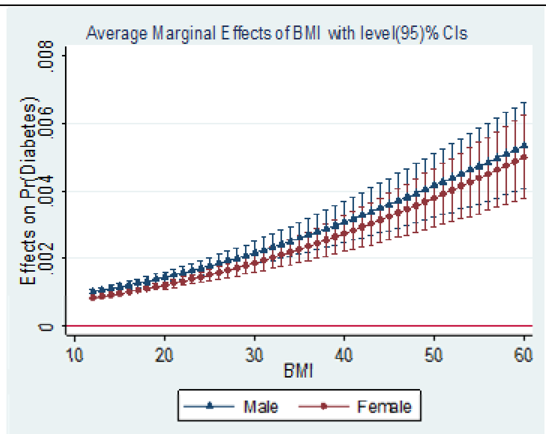
(2)
ME across Non-Overweight and Overweight or Obese Individuals w.r.t. BMI



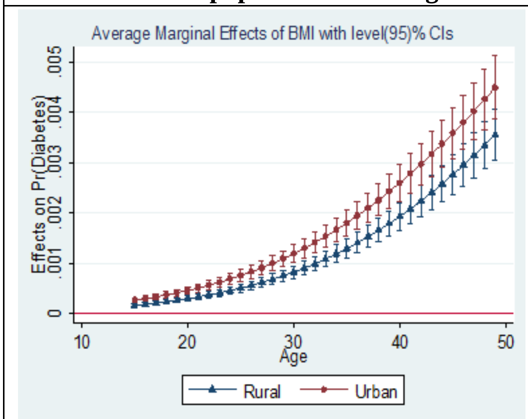
(3)
ME across Overweight or Obese Males and Females w.r.t. Age



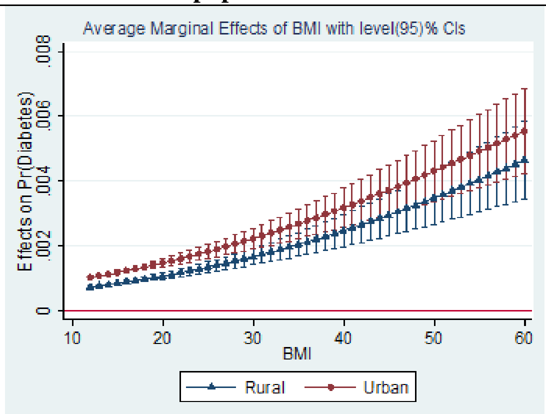
(4)
ME across Overweight or Obese Males and Females w.r.t. BMI



(5)
ME across Overweight or Obese Rural and Urban population w.r.t. Age



(6)
ME across Overweight or Obese Rural and Urban population w.r.t. BMI



Source: Figure constructed by authors.
ME = Average marginal effect of BMI on self-reported diabetes status.
Notes: In all graphs (1–6), the dark dot or triangle represents the average marginal effect of a unit rise in BMI on probability of being diabetic (measured on Y-axis). On X-axis, we have plotted either age or BMI (as labelled in each graph).

Table 1: List of Variables with Definition and Type

Variable	Definition	Type
Health Outcome Variables:		
Ordinal Blood Glucose Levels	• BG = 0 if blood glucose is less than or equal to140 mg/dl • BG = 1 if between 141 and 200 mg/dl • BG = 2 if higher than 200 mg/dl	Ordinal
Self-Reported Diabetes Status	• D = 0 if non-diabetic • D = 1 if 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 (kg/m^2).	Continuous
Age	Age 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@ [13] • = 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 drink	Time since last drink (in hours), something other than plain water. Time is recorded before blood glucose measurements are taken.	Continuous
Behavioural Risk Factors[14]	• = 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
Eating Habits[15]	• = 1 if eats fried food daily or weekly, 0 otherwise@ • = 1 if drinks aerated drink daily or weekly, 0 otherwise@	Binary

[13] Includes married but gauna not done.
[14] Contains a set of eight dummy variables.
[15] Contains a set of two dummy variables.

Table 1 (Continued)		
Variable	**Definition**	**Type**
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[16]	• = 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.

[16] Contains a set of three dummy variables.

Table 2: Descriptive Statistics

Variable	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 drink (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
SC	811465	0.181	0.385	0	1
ST	811465	0.182	0.386	0	1
OBC	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 3: Descriptive Statistics by Overweight or Obesity Status

Variable	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 drink (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 3 (Continued)

Variable	**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)
SC	0.151	0.358	0.187	0.390	-0.036*** (-32.613)
ST	0.120	0.324	0.196	0.397	-0.076*** (-68.823)
OBC	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 4: Distribution of the Individuals based on Overweight and Obesity Status and Diabetes

Variable	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: Descriptive Statistics for Sub-Sample of Married Couples

Variable	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 drink (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
SC	45205	0.187	0.390	0	1
ST	45205	0.198	0.398	0	1
OBC	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.

Table 6: Average Marginal Effects of BMI on Self-Reported Diabetes Status: Probit and IV-Probit Model Estimates for Married Couples Sub-Sample

	Probit Model		IV-Probit Model	
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	
<u>First Stage</u>				
F – statistic			153.28	
R^2			0.2038	

*** represents significance at 1% significance 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).
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 7: Average Marginal Effects of BMI on Self-Reported Diabetes Status: Probit Model Estimates based on Full Sample Data

	Probit Model	
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	

*** represents significance at 1% significance level.
Delta-Method standard errors are reported in parentheses.
[#] Difference is ME(Overweight and Obese) – ME(Non-Overweight). And dof = degrees of freedom.
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 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 8: Average Marginal Effects of BMI on Self-Reported Diabetes Status amongst Overweight or Obese Individuals (BMI ≥ 25 kg/m²): Probit and IV-Probit Model Estimates for 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
Marginal Effects	0.0026***	0.0031***	0.0035***	0.0175*	0.0202**	0.0221**
	(0.0006)	(0.0007)	(0.0008)	(0.0093)	(0.0103)	(0.0108)
	Female	Rural	Poorest	Female	Rural	Poorest
Marginal Effects	0.0019**	0.0022***	0.0011***	0.0133	0.0152*	0.0086
	(0.0009)	(0.0005)	(0.0003)	(0.0098)	(0.0085)	(0.0058)
	Difference#	Difference#	Difference#	Difference#	Difference#	Difference#
	0.0007	0.0009***	0.0024***	0.0041	0.0049**	0.0135**
	(0.0009)	(0.0003)	(0.0006)	(0.0053)	(0.0020)	(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 * represents significance at 1%, 5% and 10% significance 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.

Table 9: Average Marginal Effects of BMI on Ordinal Blood Glucose Levels: Ordered Probit Model Estimates based on Full Sample Data

	Ordered Probit Model					
	WHO International BMI Classification			WHO Asian BMI Classification		
Marginal Effects	Blood Glucose ≤ 140	141 ≤ Blood Glucose ≤ 200	Blood Glucose > 200	Blood Glucose ≤ 140	141 ≤ Blood Glucose ≤ 200	Blood Glucose > 200
	Normal Blood Glucose	Prediabetes	Diabetes	Normal Blood Glucose	Prediabetes	Diabetes
	(1)	(2)	(3)	(4)	(5)	(6)
Overweight or Obese	-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					

*** represents significance at 1% significance 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 10: Average Marginal Effects of BMI on Ordinal Blood Glucose Levels amongst Overweight or Obese Individuals (BMI ≥ 25 kg/m²): Ordered Probit Model Estimates based on Full Sample Data

	Ordered Probit Model		
Marginal Effects	**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		

*** represents significance at 1% significance 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.