Southern Africa Labour and Development Research Unit

Socio-economic correlates with the prevalence and onset of diabetes in South Africa: Evidence from the first four waves of the National Income Dynamics Study

by

Velenkosini Matsebula and Vimal Ranchhod

N.i.D.S.
NATIONAL INCOME DYNAMICS STUDY

SALDRU Working Paper
Number 181

NIDS Discussion Paper
2016/8

About the Author(s) and Acknowledgments

Velenkosini Matsebula: Researcher, SALDRU, UCT. Email: velimkholo@gmail.com, corresponding author.

Vimal Ranchhod: Chief Research Officer, SALDRU, UCT. Email:vimal.ranchhod@uct.ac.za

Acknowledgements:
Funding for this research from the Department of Planning, Monitoring and Evaluation is gratefully acknowledged.

Recommended citation

Matsebula, V., Ranchhod, V. (2016). Socio-economic correlates with the prevalence and onset of diabetes in South Africa: Evidence from the first four waves of the National Income Dynamics Study. Cape Town: SALDRU, University of Cape Town. SALDRU Working Paper Number 181/ NIDS Discussion Paper 2016/8.

ISBN: 978-1-928281-42-9

Working Papers can be downloaded in Adobe Acrobat format from www.saldru.uct.ac.za.
Printed copies of Working Papers are available for R25.00 each plus vat and postage charges.

Orders may be directed to:
The Senior Administrative Officer, SALDRU, University of Cape Town, Private Bag, Rondebosch, 7701,
Tel: (021) 650 5698, Fax: (021) 650 5697, Email: tania.hendricks@uct.ac.za

Socio-economic correlates with the prevalence and onset of diabetes in South Africa: Evidence from the first four waves of the National Income Dynamics Study

Velenkosini Matsebula and Vimal Ranchhod

Saldru Working Paper 181
NIDS Discussion Paper 2016/8
University of Cape Town, August 2016

Abstract

We make use of multiple waves of National Income Dynamics Study data, from 2008 to 2015, to investigate the socio-economic factors that correlate with the prevalence and onset of diabetes. Our analysis follows a cohort of 3470 older adults aged forty and above, who are interviewed four times over a six year period. We use linear probability models and estimate the likelihood of diabetes as a function of age, race, gender, education, income, exercise, and obesity. Our primary findings are that age and obesity correlate strongly with diabetes, while income does not have a statistically significant effect, conditional on the other covariates. Our regression estimates indicate that, of individuals who reported not being diabetic in Wave 1, those who were obese and morbidly obese were 12.9 and 16.7 percentage points more likely to have experienced the onset of diabetes respectively, relative to those with a BMI in the healthy range. In addition, frequent exercise does appear to have a slight protective effect against the onset of diabetes, and there is some evidence that better educated people have a lower risk of onset of the disease.

1. Introduction

What are the socio-economic and demographic correlates with the onset of diabetes in the first four waves (2008-2015) of the National Income Dynamics Study (NIDS)? We investigate this question empirically in this paper. Diabetes has relatively low levels of prevalence, but there is increasing evidence from developed countries that diabetes represents a major and growing health issue, and that the likelihood of onset is substantially increased by a mixture of diet and lifestyle choices. Thus, as low income countries become middle income countries, as is the case of South Africa, we are likely to observe an increased prevalence rate of the disease.

Diabetes is a general term used to describe a group of diseases characterized by hyperglycemia (which is chronic high blood glucose levels), that comes about as a result of the body's failure to produce any or enough insulin to control high glucose levels. The exact causes of diabetes are unknown, but the factors that correlate with the risk of developing different types of diabetes mellitus include race, age pregnancy, genetics or family history, stress, high cholesterol, and being overweight or obese.[1]

According to the International Diabetes Federation (IDF), more than four million individuals who suffer from diabetes die each year because of the disease on a global scale.[2] Moreover, tens of millions more are threatened by complications that come with diabetes, such as heart attacks, amputation, kidney failure, and strokes. In 2011, the IDF estimated that the number of people who were suffering from diabetes globally had already reached an estimated 366 million in total, and 280 million more were identified as at high risk of developing the disease (Whiting et al., 2011). The authors report a prevalence rate of 6.5% for South Africa in 2011, among adults aged 20-79 years old. They expect this to rise to 7.2% by 2030.

The relationship between income and health is well researched internationally, but less so in developing countries in general. This is particularly true for diabetes, especially in the South African context.[3] There are some medical and epidemiological studies on diabetes in South Africa, but these do not interrogate the socio-economic correlates of the disease. For example, Bertram, Jaswal, Van Wyk, Levitt and Hofman (2013) document a study looking at the non-fatal disease burden caused by type 2 diabetes in South Africa in 2009. They perform a systematic review of the literature to identify published prevalence and mortality diabetes studies in South Africa, from 1990 to 2011. According to their study the prevalence of type 2 diabetes in people aged 30 years and above was estimated at 9% in South Africa, representing almost two million diabetes cases at the time of the study. The authors note that the risk factors that accompany urbanisation and rising unhealthy lifestyles are contributing to the rising diabetes epidemic in the country. The study modelled 8000 new cases of blindness and 2000 new amputations due to diabetes annually. It was recorded that there were 78900 years lost due to disability attributed to diabetes. 64% of disabilities were caused by diabetes alone, 9% were related to strokes, 7% were attributed to ischemic heart disease, and 6% were from amputations. The authors also claim that about half of diabetes cases remain undiagnosed and untreated, and forecast an increase in the prevalence of diabetes in South Africa.

A more localized study by Motala, Esterhuizen, Gouws, Pirie, and Omar (2008) undertook a cross-sectional survey of adult individuals aged 15 years and older, conducted by random cluster sampling, in a rural community in the Ubombo district in KwaZulu-Natal, South Africa. Out of the 1300 people who were selected for the study, 1025 participants responded. It was discovered that there was a 3.9% overall prevalence of diabetes, which was similar in both males and females. Impaired glucose

---

[1] The Obesity Society, URL: http://www.obesity.org/content/weight-diabetes. Accessed on May 25th 2016.
[2] Available at http://www.diabetesatlas.org/. Accessed on May 25th, 2016.
[3] Many studies have been done in countries such as the United States, Canada, the United Kingdom and Germany.

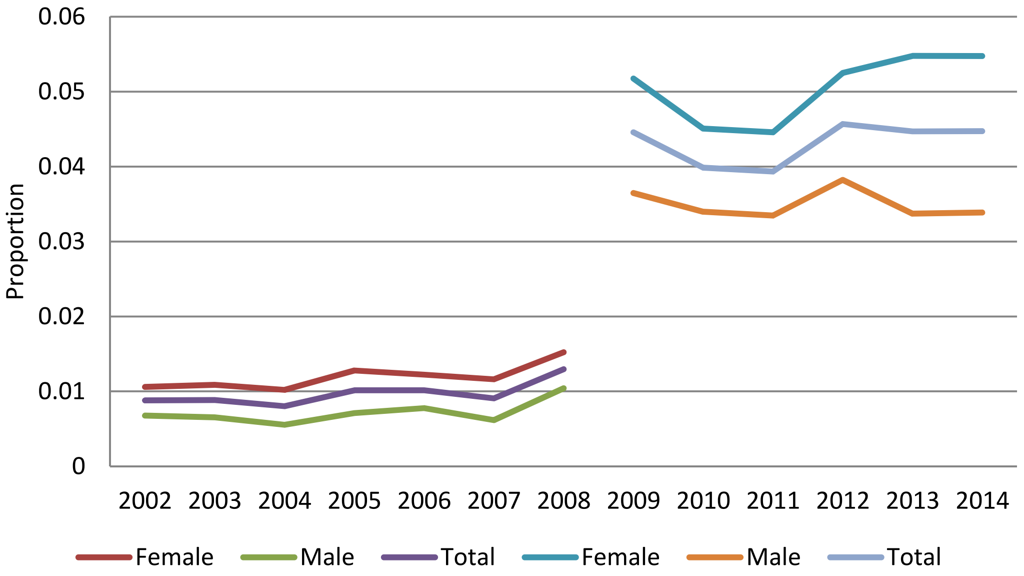
tolerance was recorded to be 4.8%, similar for both gender groups, and impaired fasting glycaemia was at 1.5%. For both genders, the prevalence of diabetes and impaired glucose increased with age, with a peak in diabetes prevalence in the age range of 55 to 64 years, and a peak for impaired glucose tolerance in the age group 65 years and older. The study revealed a higher prevalence of impaired glucose tolerance or disorders of glycaemia than of diabetes in this community.

In this paper, we explore the socio-economic factors that correlate with the onset of diabetes in South Africa. In Section 2 we provide a brief overview of prevalence rates and trends, using multiple waves of nationally representative General Household Survey (GHS) data. In Section 3, we discuss the methods that we use. Section 4 presents an overview of the NIDS data that we use, as well as a discussion on sample sizes, attrition, and some descriptive statistics on our estimation sample. In Section 5 we present results on the prevalence of the disease conditional on various covariates. Section 6 contains our regression results with a corresponding discussion. Section 7 concludes.

2. Trends in prevalence rates using General Household Survey data

While the core results in our paper are obtained using the NIDS dataset, the dataset is restricted by having a relatively small sample size, and only begins in 2008. The GHS data, which is a survey conducted by the national statistical agency Statistics South Africa (StatsSA), has been run annually since 2002. It contains a large, nationally representative cross-section of individuals, and has about 55 000 adult respondents in each wave.

Figure 1: Time trend in self-reported diabetes: GHS data
(2002-2008 & 2009-2014)



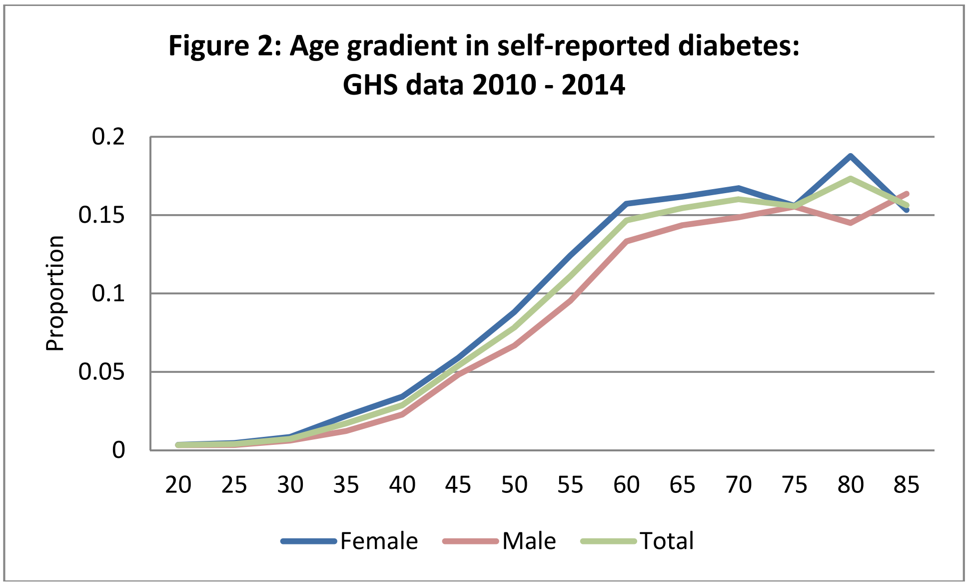
In Figure 1, we present the time trend obtained by calculating the mean proportion of adults aged between 20 to 89 years who self-report as being diabetic.[4] There is a trend break between 2008 and 2009, as the questions asked in the survey changed. Prior to 2008, respondents were asked a single

[4] All of the GHS results presented in this section were calculated using the sampling weights provided.

question about whether they were ill with diabetes, whereas from 2009 onwards there are three questions relating to diabetes, including additional ones on chronic illnesses and medication.[5]

The general point is that in either trend, there is an upward gradient over time. While there are times when the series does decrease, the longer term trend is positive, although at a fairly low rate. In the GHS data, there is also a systematic gender difference, with females being slightly more likely to report having diabetes. While the prevalence estimates are fairly low, never going above 6 percentage points, it must be borne in mind that many people could have the disease and not be aware of it.

In Figure 2, we provide the age gradient in prevalence rates for a pooled dataset spanning the most recent five years of GHS data. What is clear is that diabetes is a disease that mostly manifests among middle aged people and older. At age 45, only about 5% of the sample reports having diabetes, but this increases steadily until about age 60, where the prevalence plateaus at approximately 15%.[6]



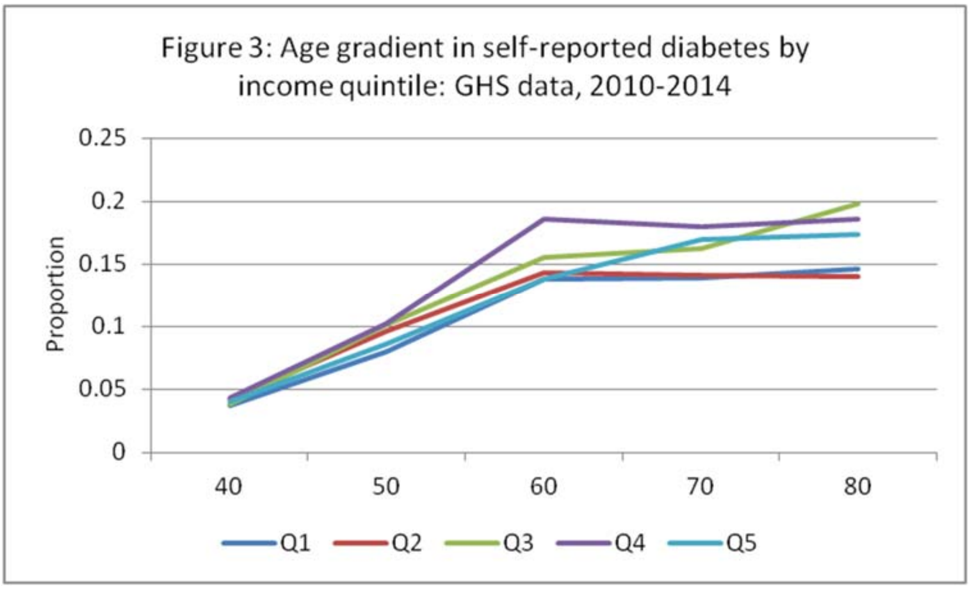
In Figure 3, we consider how the age gradient in prevalence varies with income. We use the derived total household income variable provided, and convert it to a per capita value by dividing the income by the number of members in the household. We then divide our data into income quintiles using this per capita value for each 10 year age group, and compute the mean prevalence rates within these age groups.

What we observe in the graph is that, while all income groups have an increase in the prevalence rate with age, the rate of increase differs across income groups. The poorest group (Q1) has the smallest gradient and the lowest prevalence rates, reaching a stable rate of about 14% in the age group from 60 to 69. This gradient increases with income until the fourth quintile, which would be mostly a middle class group. This group has a distinctly steeper gradient, and reaches a peak of about 18.5% in the age group from 60 to 69. The highest quintile, which would include the upper middle class and the wealthy, has a similar gradient in prevalence rates as the poorest group, but continues its upward trajectory for longer. It plateaus in the 70 to 79 age group at approximately 17%, although it is possible that this reflects greater survival probabilities in this group, relative to all of the others.

[5] In the 2014 GHS, only two such questions were asked, although they appear to maintain the trend reasonably well.

[6] We use this information in our main analysis, where we restrict our estimation sample to individuals aged 40 and above.

With this background information for context, we next turn to the primary contributions of this paper, namely, the analysis of the cohort data obtained from NIDS.



3. Methods

The methods that we use to analyse the NIDS data are fairly conventional. First, we calculate the prevalence rates of our cohort across different waves. We then perform the same calculations conditional on a host of Wave 1 characteristics in a bivariate setting. The characteristics that we use are gender, race, educational attainment, income quintile, how often respondents exercise, and BMI. For the BMI categories, we classify people into groups as underweight, healthy weight, overweight, obese, and morbidly obese. All of the prevalence rates are calculated using only members of the balanced panel, i.e. respondents who are observed in all four waves of NIDS. To adjust for attrition, as well as for sampling design effects, we calculate all our prevalence rates using the panel weights provided in the Wave 4 public release data.

The final set of analyses that we undertake is to fit a series of multivariate regression models to the data. These take the form of linear probability models, and we estimate these using OLS. We weight our observations using the panel weights provided, and estimate robust standard errors that are clustered at the sample cluster level, which is obtained from a variable provided in Wave 1. We estimate four regression models with prevalence as the dependent variable, one for each wave, and estimate a further three regression models to investigate the onset of diabetes in our cohort by Wave 2, Wave 3, and Wave 4. The regressions with 'onset of diabetes' as the dependent variable are estimated by excluding people who were already diabetic in Wave 1. The covariates that we use in our regressions are the same as the characteristics used in our bivariate analyses discussed above, and they are included as indicator variables for different groups within each variable. In addition, we include a quadratic function of age in our analyses.

4. Data

The data that we use is the cohort of people who are observed in all four waves of NIDS, and are aged forty or above in wave 1. We imposed this age restriction based on the preliminary analysis discussed above, using the GHS data. Of the original 16871 adults observed in Wave 1, only 9371 are observed in each of the subsequent waves, and of these only 3862 were aged forty or above in Wave 1. We then also exclude any individual who did not have valid educational information or a valid BMI measure in Wave 1, thus yielding our final sample of 3470 observations.

Table 1: Description of sample in NIDS balanced panel

	N	summary statistic
Overall sample	3470	
Age (mean)	3470	54.4
Prevalence		%
No diabetes (Wave 1)	3207	92.42
Diabetes (Wave 1)	263	7.58
Gender		
Female	2408	69.39
Male	1062	30.61
Race		
African	2804	80.81
Coloured	491	14.15
Asian/Indian	44	1.27
White	131	3.78
Education category		
<matric	3143	90.58
matric	157	4.52
matric plus	170	4.9
BMI category		
Underweight (BMI<18)	190	5.48
Healthy weight (18<=BMI<25)	1070	30.84
Overweight (25<=BMI<30)	854	24.61
Obese (30<=BMI<35)	674	19.42
Morbidly obese (35<=BMI)	682	19.65
HH income per capita		
Quintile 1	630	18.16
Quintile 2	811	23.37
Quintile 3	846	24.38
Quintile 4	802	23.11
Quintile 5	381	10.98
Frequency of exercise		
Never	2883	83.08
Less than once per week	139	4.01
Once/week	111	3.2
Twice per week	109	3.14
At least thrice per week	213	6.14
Exercise data missing	15	0.43

In Table 1, we present the number of observations and the mean or mean proportions of the groups of covariates that we use in our analyses. These are sample statistics and are unweighted. All of the covariates are obtained from Wave 1 data. Of our 3470 observations, the mean age is 54.4 years, and the diabetes prevalence rate is 7.58%. There are many more females in our sample, at 69.39%. Africans make up 80.8% of our sample, followed by Coloureds, who represent 14.15%. Indians and Whites are under-represented, and with Indians there are only 44 observations. Over 90% of the sample did not finish secondary school, while 4.9% have some tertiary educational experience.

In the BMI categories, the modal group is in the healthy weight range, but there is significant mass in the distribution in the overweight, obese, and morbidly obese groups as well, at approximately 24.6%, 19.4% and 19.7% respectively. The income groups are also fairly spread out, although there is some concern about the highest income quintile, which only accounts for about 11% of the sample. This probably reflects non-random attrition in the balanced panel. In terms of the frequency of exercise, we see that the vast majority of respondents report no exercise in a typical week, although 213 individuals, or 6.14% of the sample, report exercising at least three times per week.

One note of explanation is required with regard to how we generated the 'diabetes' variable. The relevant question on whether a person has diabetes is asked in each wave.[7] However, a number of people who responded 'yes' in a prior wave, later responded 'no' to the same question in a subsequent wave. This could reflect measurement error, problematic recall, or people who are feeling better either due to improved diet, exercise, or medication. For our purposes, we treated the prevalence of diabetes as sequentially cumulative; i.e. if person X reported having diabetes in a particular wave, then we impose that they have diabetes in every subsequent wave as well. Thus, by construction, our measures of prevalence and onset must increase with time.

We next consider the distribution of diabetes prevalence across the multiple waves, conditional on the different subsets of the groups described above.

5. Trends in prevalence by groups

In Table 2, we show the prevalence rates by demographic group, as well as how this changes with time. Overall, the prevalence rate increases quite dramatically, from 7.5% in Wave 1 to almost 20% in Wave 4, a growth of 166%. The levels and growth rates are fairly similar by gender, although by Wave 4 females are 1.9 percentage points more likely to report having diabetes.

[7] There is also a change in the structure of the questionnaire in Wave 4, which required that the 'diabetes' variable in that wave be generated using responses from two different questions. It is not clear whether this will affect the proportion classified as having diabetes. Logically, there should not be a difference, but survey responses can be sensitive to changes in the questionnaire. To the extent that this is a relevant issue, we note that most of our results will still hold even if we only use Wave 1, Wave 2, and Wave 3.

Table 2: Diabetes prevalence by demographic characteristics

	Ever reported diabetes			
	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
Overall	0.075	0.113	0.143	0.199
Gender				
Female	0.076	0.112	0.148	0.206
Male	0.073	0.115	0.136	0.187
Race				
African	0.067	0.104	0.131	0.184
Coloured	0.099	0.158	0.209	0.258
Asian/Indian	0.282	0.326	0.375	0.524
White	0.057	0.095	0.123	0.176
Education category				
<matric	0.072	0.115	0.144	0.201
matric	0.117	0.147	0.182	0.241
Some tertiary	0.064	0.071	0.106	0.138

Note:

1. Proportions are weighted using the balanced panel weights for wave 1 to wave 4.

We see substantial divergences by race. Indians have much higher rates of diabetes than any other group, reaching over 50% by Wave 4, although this is based on a very small subsample of 44 individuals. The Coloured subsample also has much higher rates of diabetes, about 50% greater than the African subsample in each wave, and this is calculated using a non-trivial subsample of 491 individuals.

When we consider groups defined by different levels of educational attainment, we see that the lowest prevalence rates are observed in the most highly educated group. Amongst those with some tertiary schooling, the prevalence rate increases from 6.4% in Wave 1 to 13.8% in Wave 4. The highest prevalence rates are observed in the group with intermediate levels of education. Those who have completed secondary school have a prevalence rate of 11.7% in Wave 1, and this increases to 24.1% by Wave 4. The least educated group, those who never completed secondary school, have prevalence rates that fall in-between the two more educated groups. This rises from 7.2% in Wave 1, to a high 20.1% by Wave 4. Qualitatively, this is consistent to what we observed with the GHS data and the age gradient for different income groups.

In Table 3, we perform the same calculations for groups defined by household level per capita income quintile in Wave 1. In each group we see substantial increases in the prevalence rates across waves. The prevalence rates for the lowest two quintiles are lower than they are for the higher quintiles, but they increase more than threefold between Wave 1 and Wave 4. The levels are fairly similar for the middle and upper quintiles. Each of these starts off below 10% in Wave 1, and each increases by well over 100% during the observation period, ending up at above 20% for each group. It seems to be a reasonable conjecture that there is some mechanism by which income affects the prevalence of diabetes.

Table 3: Distribution of diabetes prevalence by household per capita income quintile

	Ever reported diabetes			
	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
Quintile 1	0.046	0.088	0.113	0.147
Quintile 2	0.059	0.101	0.130	0.178
Quintile 3	0.078	0.125	0.151	0.228
Quintile 4	0.098	0.118	0.149	0.225
Quintile 5	0.083	0.129	0.167	0.203

Notes:

1. Proportions are weighted using the balanced panel weights for wave 1 to wave 4.

2. The quintiles correspond to the income variable from wave 1.

In Table 4, we consider the prevalence rates by the frequency with which the respondents exercise, as reported in Wave 1. For the most part, it is difficult to make sense of the table. Those who exercise once or twice per week tend to have higher prevalence rates than those who never exercise, but the group who exercise at least three times per week has the lowest prevalence rates. In all likelihood, the exercise rates reflect an implicit correlation with something else, possibly income, which then confounds the interpretation. Nonetheless, all groups show substantial growth in prevalence rates, and even the lowest group has a prevalence rate of 18.5% by Wave 4. An additional factor that may be relevant is that the group that exercises the most frequently is comprised of people who exercise 'at least' three times per week, and will also include anyone who exercises strenuously on a daily basis. Since the variable is thus open ended, it will include a broader dispersion within it, relative to the other groups, and this might help explain the inverted 'U-shaped' pattern that we observe in the table.

Table 4: Distribution of diabetes prevalence by frequency of exercise

	Ever reported diabetes			
	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
Never	0.072	0.113	0.141	0.195
Less than once per week	0.069	0.111	0.124	0.209
Once per week	0.145	0.160	0.213	0.223
Twice per week	0.092	0.140	0.191	0.242
At least thrice per week	0.054	0.074	0.102	0.185
No information on exercise	0.000	0.084	0.084	0.084

Note:

1. Proportions are weighted using the balanced panel weights for wave 1 to wave 4.

In Table 5 we present the prevalence rates for groups with different BMI levels in Wave 1. The pattern here is quite remarkable. The underweight and healthy weight groups start off with low levels of diabetes, which then approximately triples by Wave 4, to reach about 10%. The overweight group starts out with a prevalence rate that is close to double that of the healthy weight group, and the prevalence rate also triples, to reach 19.2% by Wave 4. The highest prevalence rates to begin with are in the obese and morbidly obese groups, which are almost double that of the overweight group. Their growth rates in prevalence are also substantial, such that their prevalence rates by Wave 4 are 28.7% and 29.5% respectively, almost triple that of the respondents who were in the healthy weight range in Wave 1.

Table 5: Distribution of diabetes prevalence by Wave 1 BMI category

	Ever reported diabetes			
	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**
Underweight (BMI<18)	0.036	0.046	0.068	0.101
Healthy weight (18<=BMI<25)	0.033	0.054	0.066	0.098
Overweight (25<=BMI<30)	0.062	0.120	0.138	0.192
Obese (30<=BMI<35)	0.126	0.172	0.232	0.287
Morbidly obese (35<=BMI)	0.113	0.151	0.196	0.295

Note:

1. Proportions are weighted using the balanced panel weights for wave 1 to wave 4.

Overall, we see some clear bivariate patterns emerging. The overall prevalence of diabetes has increased substantially in our sample over the approximately six year period. Coloureds, and Indians in particular, seem to be more predisposed to this disease. High levels of education are associated with lower levels of diabetes. Frequent exercise seems to have some protective effect against the onset of diabetes. Overweight people are about twice as likely to have diabetes as people in the healthy weight group, and people who are obese or morbidly obese are about three times as likely to be diabetic. Indeed, by Wave 4, almost 3 out of every 10 people who were morbidly obese in Wave 1 are diabetic, an increase of 18.2 percentage points. The poorest respondents are the least likely to have diabetes, which most likely reflects a difference in diet and possibly lifestyle.

In the next section, we present our multivariate regression estimates that will allow us to identify the factors that correlate with diabetes once the other covariates are explicitly controlled for.

6. Regression results

Table 6: Multivariate regression results for prevalence and onset of diabetes

	Prevalence				Onset		
VARIABLES	**Wave 1**	**Wave 2**	**Wave 3**	**Wave 4**	**Wave 2**	**Wave 3**	**Wave 4**
(age-40)	0.00881***	0.0106***	0.0105***	0.00908**	0.00233*	0.00269	0.00188
	(0.00163)	(0.00186)	(0.00205)	(0.00363)	(0.00121)	(0.00173)	(0.00317)
(age-40)^2	-0.00016***	-0.00018***	-0.00017***	-9.29e-05	-2.56e-05	-2.60e-05	4.78e-05
	(4.24e-05)	(4.84e-05)	(5.24e-05)	(0.000107)	(3.68e-05)	(4.68e-05)	(9.49e-05)
male	0.0234	0.0404*	0.0327	0.0477**	0.0206	0.0149	0.0347*
	(0.0192)	(0.0214)	(0.0225)	(0.0233)	(0.0129)	(0.0159)	(0.0190)
Coloured	0.0271	0.0560**	0.0796**	0.0688*	0.0349	0.0627	0.0552
	(0.0223)	(0.0220)	(0.0329)	(0.0412)	(0.0231)	(0.0388)	(0.0482)
Indian	0.194*	0.209	0.219**	0.332***	0.0317	0.0597	0.240***
	(0.114)	(0.132)	(0.111)	(0.0620)	(0.0654)	(0.0557)	(0.0611)
White	-0.0534	-0.0504	-0.0663	-0.0412	-0.000596	-0.0225	0.000731
	(0.0453)	(0.0548)	(0.0534)	(0.0604)	(0.0283)	(0.0323)	(0.0444)
Matric	0.0572	0.0357	0.0427	0.0415	-0.0204	-0.00888	0.00102
	(0.0369)	(0.0408)	(0.0446)	(0.0512)	(0.0249)	(0.0375)	(0.0495)
Some tertiary	0.00327	-0.0458	-0.0459	-0.0682	-0.0508**	-0.0521*	-0.0705*
	(0.0288)	(0.0373)	(0.0356)	(0.0422)	(0.0244)	(0.0294)	(0.0369)
Underweight	-0.00131	-0.0195	-0.00811	-0.0106	-0.0200**	-0.00852	-0.0128
	(0.0172)	(0.0196)	(0.0235)	(0.0281)	(0.00954)	(0.0180)	(0.0239)
Overweight	0.0256	0.0639***	0.0688***	0.0907***	0.0415***	0.0472***	0.0713***
	(0.0180)	(0.0222)	(0.0235)	(0.0250)	(0.0147)	(0.0162)	(0.0203)
Obese	0.0909***	0.122***	0.167***	0.192***	0.0387***	0.0936***	0.129***
	(0.0238)	(0.0259)	(0.0281)	(0.0308)	(0.0145)	(0.0210)	(0.0274)
Morbidly obese	0.0877***	0.116***	0.147***	0.224***	0.0358***	0.0733***	0.167***
	(0.0260)	(0.0277)	(0.0316)	(0.0342)	(0.0131)	(0.0201)	(0.0305)
quintile 2	0.00109	-0.00379	-0.00417	-0.000663	-0.00448	-0.00591	-0.00305
	(0.0158)	(0.0223)	(0.0244)	(0.0281)	(0.0128)	(0.0175)	(0.0221)
quintile 3	0.0128	0.0123	0.00609	0.0371	0.000429	-0.00643	0.0264
	(0.0161)	(0.0188)	(0.0206)	(0.0247)	(0.0106)	(0.0146)	(0.0225)
quintile 4	0.0178	-0.00952	-0.0112	0.0131	-0.0284**	-0.0300*	-0.00240
	(0.0198)	(0.0242)	(0.0248)	(0.0307)	(0.0142)	(0.0169)	(0.0255)
quintile 5	0.0131	0.0310	0.0398	0.0130	0.0209	0.0309	-0.00540
	(0.0286)	(0.0372)	(0.0419)	(0.0422)	(0.0291)	(0.0374)	(0.0396)
Exercise <1x/week	-0.00769	-0.00694	-0.0241	0.00836	0.000701	-0.0156	0.0253
	(0.0297)	(0.0364)	(0.0376)	(0.0487)	(0.0225)	(0.0271)	(0.0461)
Exercise 1x/week	0.0680	0.0406	0.0626	0.0234	-0.0243	0.00644	-0.0396
	(0.0585)	(0.0566)	(0.0629)	(0.0659)	(0.0152)	(0.0541)	(0.0520)
Exercise 2x/week	0.00573	0.0124	0.0356	0.0419	0.00600	0.0309	0.0393
	(0.0426)	(0.0486)	(0.0542)	(0.0562)	(0.0322)	(0.0453)	(0.0498)
Exercise >=3x/week	-0.0254	-0.0539*	-0.0592*	-0.0154	-0.0319*	-0.0382*	0.00648
	(0.0269)	(0.0316)	(0.0332)	(0.0402)	(0.0176)	(0.0225)	(0.0376)
Exercise data missing	-0.0419	0.0131	-0.00812	-0.0438	0.0540	0.0318	-0.00482
	(0.0295)	(0.0593)	(0.0591)	(0.0640)	(0.0614)	(0.0596)	(0.0605)
Constant	-0.0661***	-0.0641**	-0.0510*	-0.0434	-0.00461	0.00342	0.000471
	(0.0219)	(0.0255)	(0.0288)	(0.0323)	(0.0137)	(0.0213)	(0.0260)
Observations	3,470	3,470	3,470	3,470	3,207	3,207	3,207
R-squared	0.070	0.072	0.081	0.097	0.029	0.035	0.058

Notes: 1. Robust standard errors in parentheses.

2. *** p<0.01, ** p<0.05, * p<0.1.

3. Standard errors are clustered at the sampling cluster level.

4. Regression results incorporate panel weights from wave 4 for the balanced panel.

5. The sample for the Onset regressions excludes people who reported being diabetic in wave 1.

In the first four columns in Table 6, we present our estimates for prevalence in Wave 1 to Wave 4 respectively. In columns 5 to 7, we present corresponding estimates for the onset of diabetes in our sample. Note that the two may be quite different, as the prevalence is a `stock' variable while onset is a `flow' variable. For example, we can see that the age variable is clearly significant for prevalence, but generally not significant for onset. The male coefficients are always positive and in some cases are marginally significant. The race coefficients are as expected, based on our findings in the previous section: Coloured and Indians are much more likely to have diabetes than Africans. The education coefficients are generally not significant, except for the onset of diabetes, where individuals have some tertiary education. Having some tertiary qualification is associated with a 5 to 7 percentage point decrease in the likelihood of getting diabetes, conditional on not being diabetic in Wave 1, and this difference is statistically significant at either the 5% or 10% level of significance, depending on which regression we are inspecting. There is some evidence, although not very strong, that exercising at least three times per week is correlated with lower prevalence and onset of the disease.

The most striking finding is the robustness of the BMI coefficients, in conjunction with the general lack of significance of the income quintile coefficients. The coefficients on being overweight, obese, or morbidly obese are large, almost always statistically significant, and get larger as time passes. By Wave 4, the likelihood of experiencing the onset of diabetes amongst people who were morbidly obese in Wave 1 was 16.7 percentage points higher than someone similar who was in a healthy weight range in Wave 1. This supports the conjecture that the effects of income on diabetes are mediated via some mixture of diet and lifestyle, which are then being manifest in people's BMIs.

7. Discussion

We set out to use a large, nationally representative dataset with a rich set of covariates, in order to explore the socio-economic variables that correlate with the prevalence and onset of diabetes in South Africa. Our findings confirm some of the well-established arguments from the medical literature: that genetics and age play an important role, that losing weight matters a lot, and exercising regularly also matters, but less than maintaining a healthy BMI. We also find that higher education has some protective effects, even after controlling for income, BMI, and exercise. This suggests that there may be some gains from well targeted education campaigns. Also of interest is that income has almost no significant explanatory power in the model, which indicates that all of the mechanisms by which income has an effect on diabetes are fully accounted for by the other covariates in the model.

The key policy implication is that the main pathways to target that could decrease the prevalence of diabetes are policies that work by getting people to maintain a healthy weight.

References

Bertram, M. Y., Jaswal, A. V., Van Wyk, V. P., Levitt, N. S., & Hofman, K. J. (2013). The non-fatal disease burden caused by type 2 diabetes in South Africa, 2009. *Glob Health Action*, *6* (Suppl 1), 206-212.

Motala, A. A., Esterhuizen, T., Gouws, E., Pirie, F. J., & Omar, M. A. (2008). Diabetes and other disorders of glycemia in a rural South African community: Prevalence and associated risk factors. *Diabetes care*, *31*(9), 1783-1788.

Whiting, D. R., Guariguata, L., Weil, C., & Shaw, J. (2011). IDF diabetes atlas: Global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes research and clinical practice*, *94*(3), 311-321.

southern africa labour and development research unit

The Southern Africa Labour and Development Research Unit (SALDRU) conducts research directed at improving the well-being of South Africa's poor. It was established in 1975. Over the next two decades the unit's research played a central role in documenting the human costs of apartheid. Key projects from this period included the Farm Labour Conference (1976), the Economics of Health Care Conference (1978), and the Second Carnegie Enquiry into Poverty and Development in South Africa (1983-86). At the urging of the African National Congress, from 1992-1994 SALDRU and the World Bank coordinated the Project for Statistics on Living Standards and Development (PSLSD). This project provide baseline data for the implementation of post-apartheid socio-economic policies through South Africa's first non-racial national sample survey.

In the post-apartheid period, SALDRU has continued to gather data and conduct research directed at informing and assessing anti-poverty policy. In line with its historical contribution, SALDRU's researchers continue to conduct research detailing changing patterns of well-being in South Africa and assessing the impact of government policy on the poor. Current research work falls into the following research themes: post-apartheid poverty; employment and migration dynamics; family support structures in an era of rapid social change; public works and public infrastructure programmes, financial strategies of the poor; common property resources and the poor. Key survey projects include the Langeberg Integrated Family Survey (1999), the Khayelitsha/Mitchell's Plain Survey (2000), the ongoing Cape Area Panel Study (2001-) and the Financial Diaries Project.

www.saldru.uct.ac.za
Level 3, School of Economics Building, Middle Campus, University of Cape Town
Private Bag, Rondebosch 7701, Cape Town, South Africa
Tel: +27 (0)21 650 5696
Fax: +27 (0) 21 650 5797
Web: www.saldru.uct.ac.za