

# **Migration from Rural to Urban Areas in Peru: Impact on Health Outcomes<sup>1</sup>**

**Juan Jose Miranda  
Georgia State University**

**November 1<sup>st</sup>, 2009**

## **Abstract**

This paper evaluates the impact of rural to urban migration on health outcomes in Peru contrasting three different groups: rural, migrant and urban people. This paper finds that migration has heterogeneous effects on health outcomes. Migration increases level of body-mass index BMI (obesity) comparing rural group and migrant group. Rural group has lower BMI level than migrant group, while urban group has the highest BMI level. Non significant results were found in blood pressure (hypertension), whereas in glucose (diabetes) result varies according to the method used.

---

<sup>1</sup> I am grateful to Shiferaw Gurmu, Jaime Miranda, Ragan Petrie, and Inash Rashad for helpful comments and suggestions. Usual disclaimer applies.

## 1. Introduction

Peru suffered political violence since the late 1970's that continued for 20 years. This violence was mainly widespread —but not exclusive— in the Andean region which caused a mass-migration from rural to urban areas. Ayacucho, an Andean and mostly rural department, was one of the most severely affected areas during this period of violence. According to the Peruvian Truth and Reconciliation Commission, more than 50% of all deaths attributed to the violence between Shining Path guerrilla and Peruvian militia occurred in Ayacucho. Given these circumstances, the Peruvian context proposes a natural experiment to evaluate the impact of migration from rural areas to urban in health outcomes for adult people such as body-mass index BMI (obesity), blood pressure (hypertension) and glucose (diabetes).

The study on these health outcomes is relevant because of their worldwide magnitude. The World Health Organization (WHO) estimated that non communicable diseases, such as cardiovascular diseases, hypertension and diabetes, accounted for 60% of total death worldwide estimated for 2005. This number of deaths double the number of deaths from all infectious diseases (including HIV/AIDS, tuberculosis and malaria), maternal and perinatal conditions, and nutritional deficiencies all combined (WHO, 2005). 80% of chronic disease-related deaths occur in low and middle income countries (WHO, 2005). Furthermore, WHO estimated that by 2020 coronary heart disease and stroke will be the leading causes of death and disability adjusted life years, and that two thirds of the global health burden will be due to chronic diseases (WHO, 2002).

Peru, a developing economy, does not escape to this phenomenon. For example, with respect to obesity, a leading cause of cardiovascular diseases and coronary heart disease, Peru present high average level of BMI index (on average 25.8 kg/m<sup>2</sup>) which lies on overweight<sup>2</sup> range. In comparison to other countries, this average value is similar to adult mean across North America and Europe (25–27 kg/m<sup>2</sup>), but higher than Africa and Asia (20–23 kg/m<sup>2</sup>).

---

<sup>2</sup> Overweight is measured as BMI greater or equal than 25 kg/m<sup>2</sup>, and lower than 30 kg/m<sup>2</sup>.

According to the most recent Peru's National Survey of Nutritional, Biochemical, Socioeconomic and Cultural indicators related to Chronic Degenerative Diseases, urban areas have higher prevalence of overweight and obesity, hypertension, and diabetes.

In Lima 40% of inhabitants have overweight which is higher than any other region: Jungle (34.9%), rest of Coast (31.7%), Urban Andes (32.9%), and Rural Andes (22.1%). Furthermore, Coast areas —excluding Lima— have higher level of obesity<sup>3</sup> (20.2%) in comparison to Lima (18.8%), Jungle (15.2%), Urban Andes (10.8%), and Rural Andes (9.2%) (INS, 2006).

With respect to hypertension, Lima has a prevalence of 11.6% and the rest of Coast of 11.2%, higher than any other region: Urban Andes (5.2%), Rural Andes (7.2%), and Jungle (9.1%). National prevalence of hypertension is 9.5% (INS, 2006).

Finally, with respect to diabetes Lima has the highest prevalence, analogous to the previous two variables. Prevalence of diabetes in Lima is 4.6%, superior to rest of Coast (2.5%), Jungle (2.5%), Urban Andes (0.7%) and Rural Andes (0.3%), being the national prevalence of 2.8% (INS, 2006).

In this perspective, the objective of this study is to evaluate the impact of migration in Peru (which involves changes in life styles, dietary intake, habits and customs)<sup>4</sup> from rural areas to urban areas on health outcomes (BMI, blood pressure and glucose levels).<sup>5</sup> These outcomes were selected as they are the basis to identify obesity, hypertension and diabetes, respectively.

Migration studies suggest that healthier and more educated people are more willing to migrate (Sasin and McKenzie, 2007). I hypothesize that migration worsen health indicators for rural people after controlling self-selection phenomenon. In other words, migrants vary to less healthy.

---

<sup>3</sup> Obesity is measured as BMI greater or equal than 30 kg/m<sup>2</sup>.

<sup>4</sup> For instance, Andean and rural people in Peru work more in agricultural activities requiring more individual effort, use less transportation services, are accustomed to walk more, and eat healthier food. In contrast, urban people live in highly concentrated areas, like Lima (Peruvian capital city), are accustomed to use more public transportation due to long distances, and also they are adapted to spend more time on non high-effort activities outside their jobs.

<sup>5</sup> Notice that migration is not restricted only to the process of moving from rural areas to urban areas. It involves the process of moving, life styles changes, new economic opportunities, and adaptation to a new environment.

Hence, it should be found a pattern that rural people have better health indicators followed by migrants, and the latter followed by urban people. This hypothesis, in the case of BMI, is the opposite of developed countries, such as the U.S., where has been found that obesity is higher in rural than in urban settings (Jackson et. al., 2005; Boehm et. al., 2005).

To test this hypothesis I apply Instrumental Variable and Propensity Score Matching methods. If I control for the right set of observables variables (in the case of PSM) and find a good instrument (in the case of IV), I should be able to find similar results in both methods for different health outcomes.

I found robust evidence that migration from rural to urban areas increases level of BMI (obesity). No significant results were found in blood pressure (hypertension), whereas in glucose (diabetes) result varies according to the method used. These results suggest that migration impact on health outcomes is not homogeneous.

The rest of the paper proceeds as follows. Section 2 provides a literature review. Section 3 presents the data. Section 4 provides the empirical strategy and results. The last section concludes.

## **2. Literature Review**

The seminal theoretical approach related to migration from rural to urban areas is based on Harris and Todaro (1970) approach who stated that the migration decision depends on expected income differential between urban and rural areas. This income differential in favor of urban areas can be explained because of governmental policies such as union workers, minimum wage laws, unemployment benefits, so on. This income differential produces urban unemployment for which Harris and Todaro (1970) stated that wage-subsidy or migration-restriction policies lead to welfare improvements.

Several empirical studies followed the previous approach and focused on rural to urban migration and wage differentials. For the Peruvian reality, there are few studies. Sabates (2000)

evaluates rural-urban migration and rural-rural migration, concluding that wage-rate differentials are important determinants for the migrant group, but more importantly for secondary jobs between rural areas suggesting complementarities between secondary activities in rural areas and the main agricultural activity.

To the best of my knowledge, there are few studies that evaluate impacts of migration on health outcomes. Jokela et al. (2009) provides interesting insights with a 21-year panel data study in Finland. Authors found that people living in Finland rural areas has higher BMI. Furthermore, they found that overweight and obese adolescents are less likely to migrate to urban areas (independently of other socioeconomic variables such as education, occupational class, marital status, and parenthood status) supporting the hypothesis that healthier people are more willing to migrate.

Szklarska et al. (2008) evaluated health differences between inhabitant of Wroclaw – Poland and migrants (those who moved at age of 16 years) to that city finding that migrants men attain more favorable health conditions, e.g. lower blood pressure, cholesterol and glucose; while females migrants have lower blood pressure and heart rate. Also, they found that migrants were slightly slimmer than their Wroclaw-born peers.

Finally, Torun et al. (2002) compared migrants to and non-migrants of Guatemala City, a developing country. Authors found that migrants increase in sedentarism and undesirable eating habits becoming men fatter, supporting our hypothesis that migration to urban areas makes worse health outcomes. On the other hand, rural and urban women have similar prevalence of overweight.

Besides their interesting conclusions, selection bias problem is still present. As Sasin and McKenzie (2007) stated, usually healthier, more educated and wealthier households are more likely to migrate. Migration decision is not a random process. Hence, in this paper I aim to correct for this problem between migrants and non-migrants (rural) people.

### 3. Data

For the purpose of the study I will use a cross section survey done in 2007 which is based on the WHO STEPwise approach to Surveillance (STEPS) survey<sup>6</sup>. In this survey three different groups of comparison were selected: (a) people who always have lived in a rural environment, always lived in Ayacucho; (b) people who migrated from rural to urban areas, migrated from Ayacucho to Lima; (c) people who have always lived in an urban environment, always lived in Lima.

Total sample is 1077 individuals and it has no national representation neither by geographical location. Individuals were contacted in person. Stratification selection was done according to age and sex to ensure sufficient number of people in each stratum. Response rate at enrolment was 73% and overall response rate at completion of the study was 61.6%. Response rates were higher in the rural group and lower in the urban group. For survey's details refer to Miranda et. al. (2009).

All groups were selected randomly within Santillana district, Huanta province in Ayacucho (for the first group) and within the district of San Juan de Miraflores, a periurban shantytown in the south of Lima (for the second and third group). For all study groups, individuals from both sexes aged 30 years-old and over<sup>7</sup>, permanently living in their residence were considered eligible to take part on this study, however, pregnant women and those with mental illnesses were excluded from the study.

The district of Santillana and San Juan de Miraflores face dissimilar socio-economic characteristics. In 2001, according to the Ministry of Economics' Poverty Map (MEF, 2001) poverty level in Santillana was 99% (in the department of Ayacucho was 52%), while in San

---

<sup>6</sup> The STEPS method focuses on obtaining core data on the established risk factors that determine the major disease burden. It covers three different levels of "steps" of risk factor assessment: (i) Questionnaire, (ii) Physical measurements, and (iii) Biochemical measurements. For detailed explanation, please refer to: <http://www.who.int/chp/steps/en/>

<sup>7</sup> This survey focused on people aged 30 years-old or more in order to capture the migrant group during the 1980 – 1990 period in Peru.

Juan de Miraflores was 41% (in the department of Lima was 21%).<sup>8</sup> Table A1 (appendix) provides detailed differences between districts, their departments and the national level according to 2007 Peruvian Census, the latest census. On the whole, Santillana is a rural and poor district with high level of illiteracy, dedicated to agricultural activities. San Juan de Miraflores is an urban, but also poor district compared to its surroundings within Lima.

Table 1 shows some descriptive statistics for each group and the total sample.<sup>9</sup> Column (1) displays statistics for the entire sample. Columns (2), (3), and (4) display statistics for rural, migrant and urban groups, respectively. Columns (5), (6), and (7) display cross differences between groups and statistical significance. Average age is 48 years old in all groups and less than 50% of interviewee are males, following stratification selection process. People from rural areas have lower level of education (on average primary school<sup>10</sup>), while migrants and urban people have higher level of education (on average incomplete high school<sup>11</sup>).

With respect to their main economic activity, in rural areas residents are more employed on agricultural activities (farmers) while in urban areas migrants and urban residents are employed more on commercial activities (merchant) or non-skilled activities (primarily “*obreros*”). Similar conclusions arise in Table A1.

With respect to household assets, there is also an enormous difference between rural group and the other two groups confirming previous socioeconomic differences. Mostly all urban households have water pipe and wastepipe connected to the public service, they use gas as the main fuel for cooking and their house is built with concrete or cement. On the contrary, roughly all rural people use firewood as the main fuel for cooking and their house are built with adobe and they do not have concrete floor. Once again, similar conclusions are found in Table A1.

---

<sup>8</sup> A household is considered below poverty level if total expenses are lower than a basic consumption basket which includes a food basket and a non food basket. Food basket is based on calories intake and it is different across regions. In 1997 dollars, poverty line varied between US\$ 53.6 dollars/monthly per-capita (for rural highlands) to US\$ 80.2 dollars/monthly per-capita (for Metropolitan Lima).

<sup>9</sup> Table A2 (appendix) provides complementary information for Table 1: minimum, maximum and standard deviation.

<sup>10</sup> Primary school is considered from 0 to 6 years of education.

<sup>11</sup> High school is considered from 7 to 11 years of education.

**Table 1**  
**Descriptive Statistics: Socioeconomic Variables**

Variable	(1) Total Sample	(2) Rural	(3) Migrant	(4) Urban	(5) Diff (2)-(3)	(6) Diff (2)-(4)	(7) Diff (3)-(4)
Age (years)	47.96	48.32	47.78	48.14	0.53	0.18	-0.36
Male	0.47	0.47	0.48	0.46	0.00	0.01	0.01
Completed years of education	7.14	4.11	7.15	10.11	-3.04 ***	-6.00 ***	-2.96 ***
Total members per household	5.76	5.31	5.89	5.83	-0.58 ***	-0.52 ***	0.06
Total members > 18 years per household	3.59	2.59	3.89	3.69	-1.30 ***	-1.10 ***	0.20
<b>Main Economic Activity</b>							
Farmer	0.09	0.45	0.00	0.00	0.45 ***	0.45 ***	0.00
Merchant	0.15	0.05	0.20	0.12	-0.15 ***	-0.07 ***	0.09 ***
Unskilled Worker	0.20	0.02	0.24	0.29	-0.21 ***	-0.26 ***	-0.05
<b>Assets</b>							
Waterpipe and wastepipe	0.82	0.37	0.94	0.93	-0.57 ***	-0.56 ***	0.01
Gas use for cooking	0.72	0.01	0.90	0.92	-0.88 ***	-0.91 ***	-0.03
Firewood use for cooking	0.21	0.99	0.02	0.00	0.97 ***	0.99 ***	0.02
Concrete floor	0.72	0.03	0.89	0.88	-0.86 ***	-0.85 ***	0.01
Concrete external walls	0.68	0.00	0.86	0.85	-0.86 ***	-0.85 ***	0.01
Concrete ceiling	0.49	0.00	0.63	0.60	-0.63 ***	-0.60 ***	0.02
Sample Size	1077	289	589	199			

Note: \*\*\* Difference in means at 99% level (ttest).

Source: Survey database

Table 2 displays descriptive statistics for health outcomes.<sup>12</sup> Once more, column (1) displays statistics for the entire sample. Columns (2), (3), and (4) display statistics for rural, migrant and urban groups, respectively. Columns (5), (6), and (7) display cross differences between groups and statistical significance. On average, rural people (1.52 mts.) heights lower than migrants (1.54 mts.) and urban people (1.56 mts.). Furthermore, rural people have lower weight (54 kg.) in contrast to migrants (64 kg.) and urban inhabitants (69 kg.). For BMI, rural group have also the lowest level (23.2 Kg/m<sup>2</sup>) in comparison to migrants (27.0 Kg/m<sup>2</sup>) and urban group (28.3 Kg/m<sup>2</sup>). Also, there is a higher proportion of obese and overweight in migrants and the urban group.

<sup>12</sup> Table A3 (appendix) provides complementary information for Table 2: minimum, maximum and standard deviation.

In the case of hypertension (measure as Systolic Blood Pressure  $\geq 140$  mm Hg or Diastolic Blood Pressure  $\geq 90$  mm Hg<sup>13</sup>), on average this illness does not appear given that mean values are lower than defined thresholds, though I can state that migrant and rural have similar SBP (Systolic Blood Pressure) but lower than the urban group; and, migrant group have the lowest level of DBP (Diastolic Blood Pressure) comparing with the rural and urban group (see Table 2).

Finally, average values for glucose are lower that they cannot be considered as diabetes (defined as glucose  $\geq 126$  mg/dL<sup>14</sup>). Measuring Glucose, rural group has the lowest level (80 mg/dL) followed by migrant group (89 mg/dL) and urban group (96 mg/dL).

**Table 2**  
**Descriptive Statistics: Health Outcomes**

Variable	(1) Total Sample	(2) Rural	(3) Migrant	(4) Urban	(5) Diff (2)-(3)	(6) Diff (2)-(4)	(7) Diff (3)-(4)
Height (cm)	154.22	152.11	154.12	156.61	-2.01 ***	-4.49 ***	-2.48 ***
Weight (Kg)	63.11	53.88	64.13	69.43	-10.24 ***	-15.55 ***	-5.31 ***
BMI (Kg/m2)	26.50	23.23	27.02	28.28	-3.79 ***	-5.05 ***	-1.26 ***
Obese (%) <sup>1/</sup>	0.20	0.03	0.21	0.34	-0.18 ***	-0.31 ***	-0.13 ***
Overweight (%) <sup>2/</sup>	0.38	0.16	0.46	0.37	-0.30 ***	-0.20 ***	0.09
Systolic Blood Pressure (mm Hg) <sup>3/</sup>	121.75	120.92	119.86	128.16	1.06	-7.24 ***	-8.31 ***
Diastolic Blood Pressure (mm Hg) <sup>3/</sup>	72.89	74.21	71.34	76.16	2.87 ***	-1.95	-4.82 ***
Glucose (mg/dL) <sup>4/</sup>	88.83	80.74	89.08	96.24	-8.35 ***	-15.51 ***	-7.16 ***
Sample Size	1077	289	589	199			

1/ BMI  $\geq 30$  Kg/m2

2/  $25 \text{ Kg/m}^2 \leq \text{BMI} < 30 \text{ Kg/m}^2$

3/ Hypertension is defined as: SBP - Systolic Blood Pressure  $\geq 140$  mm Hg or DBP - Diastolic Blood Pressure  $\geq 90$  mm Hg

4/ Diabetes is defined as glucose  $\geq 126$  mg/dL

Note: \*\*\* Difference in means at 99% level (ttest).

Source: Survey database

<sup>13</sup> There are three criteria to determine hypertension: (i) SBP  $\geq 140$  mm Hg, or (ii) DBP  $\geq 90$  mm Hg, or (iii) self report of physician diagnosis and currently receiving antihypertensive medication. In this study I just focus on the first two criteria.

<sup>14</sup> Similarly, there are two criteria to determine diabetes: (i) Fasting glucose  $\geq 126$  mg/dL, or (ii) self report of physician diagnosis and currently receiving antidiabetic medication. In this study I only focus on the first criterion.

However, average values are likely to hide information, especially for health variables of interest. In this sense, Figure 1 provides the z-score distribution, while Graph 2 provides the Lorenz curve for each group (migrant, urban and rural).

Graph 1 does not depict much differences among the distribution of each group. Since, the distributions are very close, therefore, we cannot infer conclusively. However, in the case of BMI and Glucose<sup>15</sup>, we can state that migrant group's distribution is more similar to urban group rather than rural group. It is important to mention that in the case of log of Glucose, rural group's distribution is very different (flatter with higher variance) than the distribution of the other two groups (migrant and urban). In the case of hypertension, distributions for DBP are very similar in the three groups, while for SBP the migrant group has a flatter distribution.

In Figure 2, for BMI, we observe that urban group dominates stochastically to rural group, and the latter dominates to the migrant group. This suggests that there is less dispersion and higher mean in urban group, followed by the rural and the migrant group, respectively.

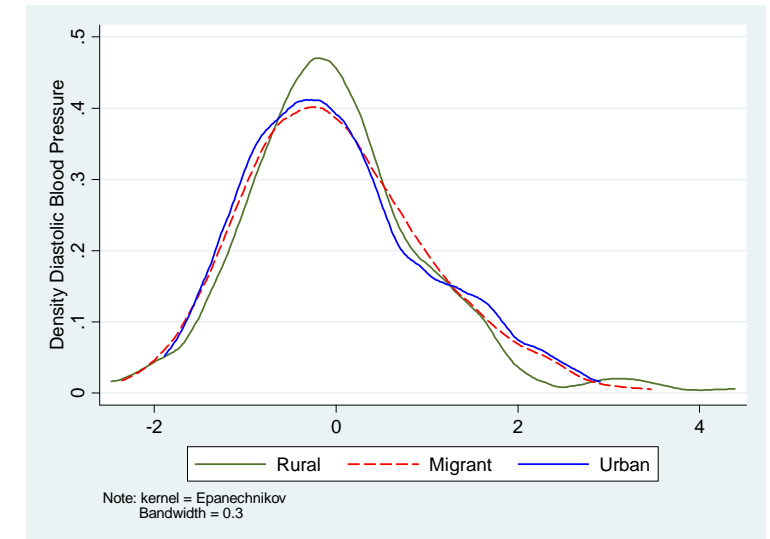
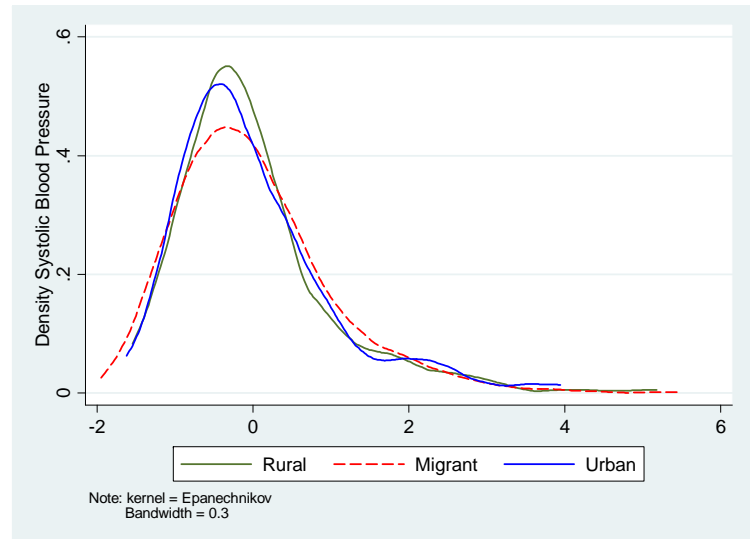
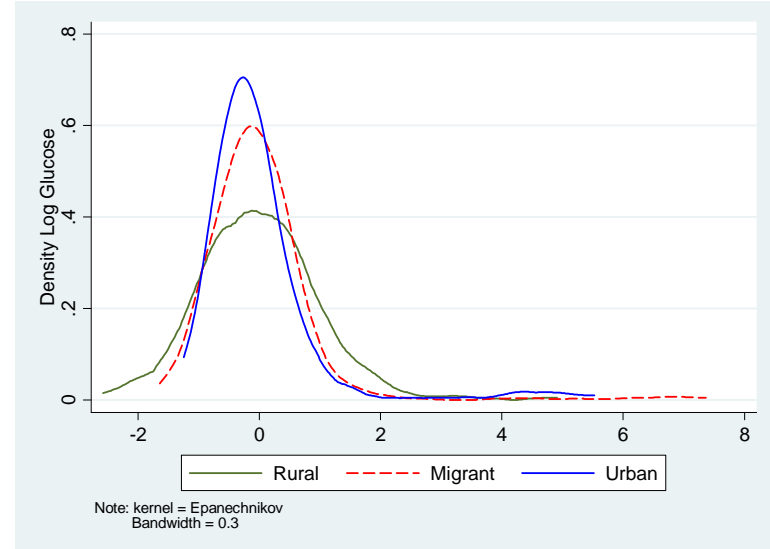
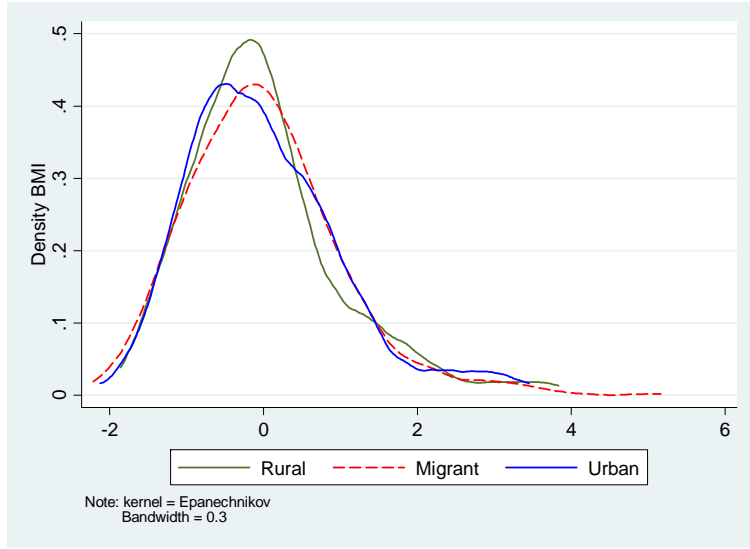
For the other three variables in analysis (Glucose, BDP, SBP) this effect is less clearly, nevertheless Glucose follows same patron as BMI. SBP's curve shows that urban group dominates stochastically to rural and migrant group, while DBP's curve shows that migrant group is dominated stochastically by migrant and urban group suggesting heterogeneous patterns and migration effects.

In this sense, with previous information provided, if migration was a random process, we can conjecture that rural people have lower level of BMI than migrants, and the latter has lower BMI than urban people, respectively, with statistical difference among the three groups. However, the same conclusion cannot be drawn for hypertension and blood pressure. Although, since randomness assumption is not accurate, I must correct for the selection bias problem in order to drawn conclusive thoughts.

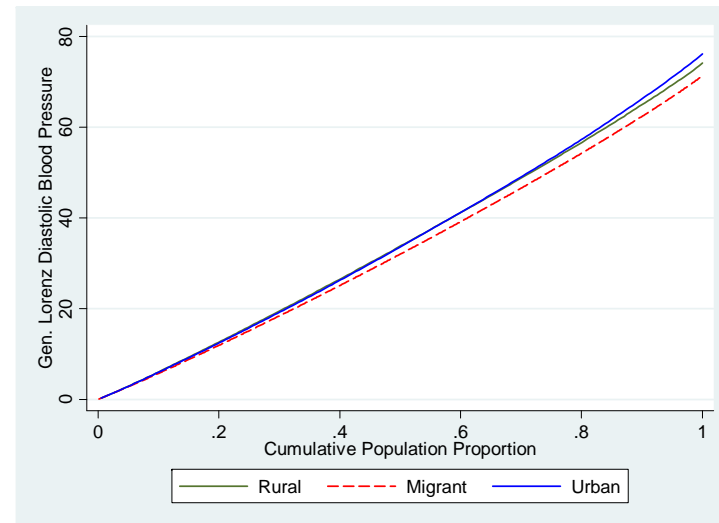
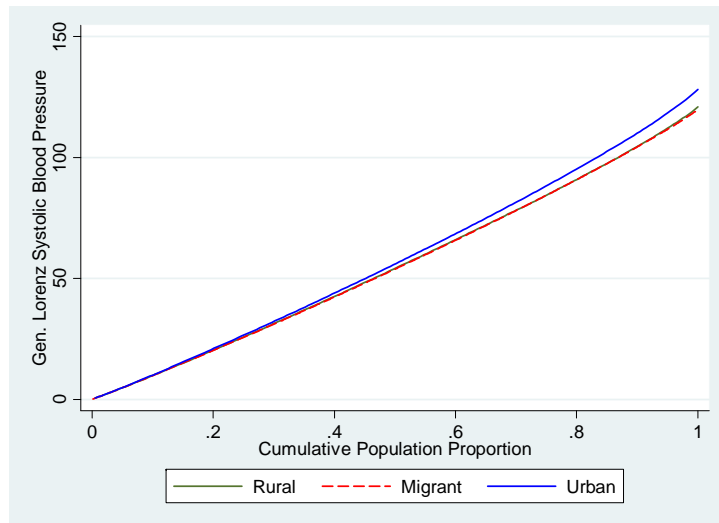
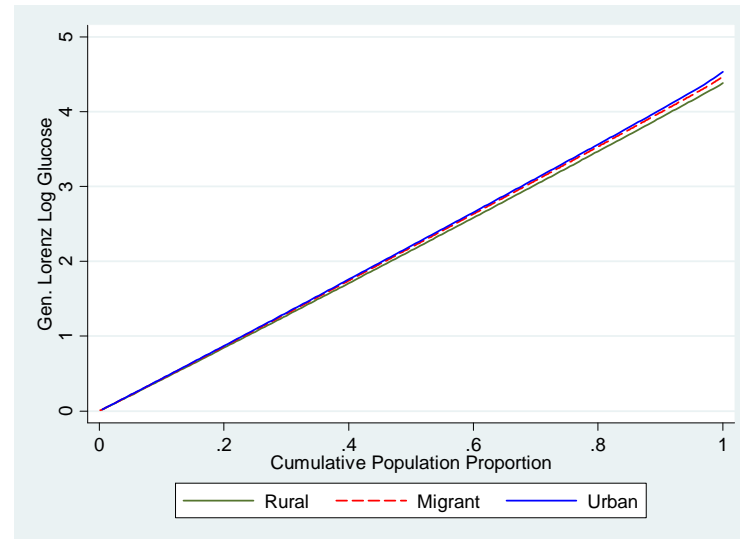
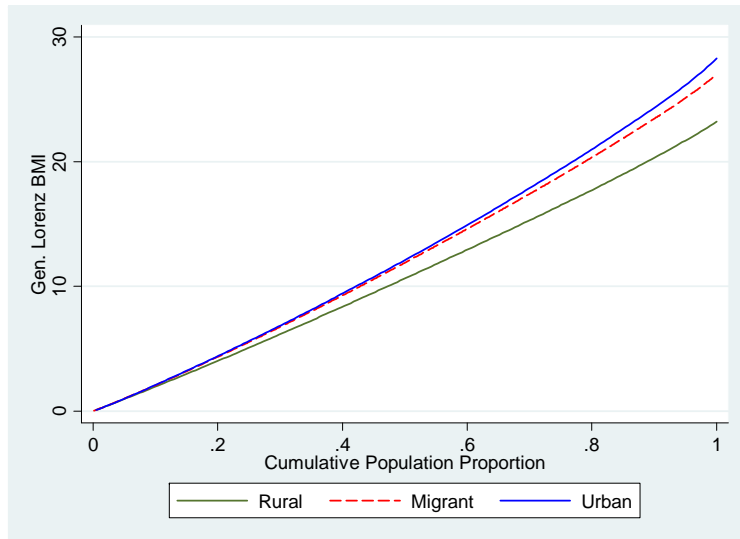
---

<sup>15</sup> In case of Glucose, I show logarithm of Glucose given that it is not uniform distributed.

**Figure 1**  
**Z-Score: BMI, SBP, DBP, Glucose**



**Figure 2**  
**Stochastic Dominance: BMI, SBP, DBP, Glucose**



#### 4. Model Specification and Results

To address the impact of migration from rural areas two approaches can be handled: Instrumental Variables (IV) and Propensity Score Matching (PSM). Both approaches help out to correct the problem of selection bias which arises in most of the migration studies. If I control for the right set of observables variables (in the case of PSM) and find a good instrument (in the case of IV), we should find similar results in both methods for different health outcomes. Recent applications of both approaches in a migration context can be found in Hagen-Zanker and Azzarri (2008).

As Sasin and McKenzie (2007) state, problem of selection bias arises when migrant “self-select” from non-migrants. Usually, healthier, more educated and wealthier households might be more likely to migrate. Sasin and McKenzie (2007) called this process “positive selection”. Therefore, comparing simple means of non-migrants with migrants does not capture real migration effects.

Since selection bias problem only arises between rural group and migrants from rural areas to urban areas, hence urban group is not considered in the analysis.

##### 4.1. Instrumental Variable Estimation

Problem of selection bias generates inconsistent estimators since the variable migration is not exogenous to the error term in the equation of interest. In this context, IV estimation provides a solution handling selection problem through an exogenous instrument which is correlated to migration decision (first stage), but is not correlated to error term of the equation of interest (second stage). This instrument must be exogenous and only affects the outcome variable of interest through its migration’s effect. This method provides consistent estimators (Angrist et. al., 1996, Cameron and Trivedi, 2005).<sup>16</sup>

---

<sup>16</sup> However, the main lack of IV method is that it lies under the assumption that the treatment effect is homogeneous within the population. If the effect is heterogeneous, under certain assumptions (such as exclusion restriction, nonzero causal effect of the IV on the treatment status, monotonicity) IV method estimates the Local Average Treatment Effect, not the Global effect (DiPrete and Gangl, 2004).

In particular, the instrumental variable must be correlated with the decision of migrate, but it cannot be correlated with health outcomes. Usually, good instruments come from natural experiments or exogenous factors.

In this context, I use number of deaths by districts due to political violence which started in the late 1970's as an instrument. This instrument account for death from 1980-1998. Remember that Ayacucho was one of the most severely affected areas during this period of violence. More than 50% of all deaths attributed to the violence between Shining Path guerrilla and Peruvian militia occurred in Ayacucho. This information was collected by the Peruvian Truth and Reconciliation Commission through family members, but it does not represent the total number of death, therefore, it is an approximation. Other consequences of political violence, aside death, were kidnapping, forced disappearance, tortures, unfair detentions, serious crimes and violations to human rights.

Number of deaths within Peruvian political violence process is an appropriate instrument because it changes social circumstances within each district without influencing individual health outcome, particularly variables such as BMI, blood pressure and hypertension. I hypothesize number of death, the instrument, can affect the decision of migration on both sides: positively or negatively. It pushes inhabitants to migrate due to poor safety conditions: people will try to run away from that insecure environment. However, this premise assumes that people can migrate. As mentioned earlier, poverty rate<sup>17</sup> in the district of Santillana was 99%, and the extreme poverty rate<sup>18</sup> was 67% in 2001 (MEF, 2001) while the comparable rural districts at the same time share similar conditions: high level of insecurity and very high levels of poverty. Hence the other possible outcome, given poorly economic conditions, is that political violence process reduces or eliminates their possibilities to migrate.

In this context, the following expressions were estimated:

---

<sup>17</sup> Poverty rate: Households are considered poor if their total per capita expenditure is lower than a basic consumption basket which includes expenses on food intake and non-food items (MEF, 2001).

<sup>18</sup> Extreme poverty rate: Households are considered extreme poor if their total per capita expenditure is lower than a basic nutritional basket which includes the minimum level of food intake (MEF, 2001).

$$Health_i = \beta + \gamma Migration + X_i\varphi + \epsilon_i \quad (1)$$

$$Migration_i = \beta + \gamma Death + X_i\varphi + \mu_i \quad (2)$$

$$Health_i = \alpha + \vartheta \widehat{Migration} + X_i\tau + v_i \quad (3)$$

Were equation (1) shows a simple OLS estimation, and equation (2) and (3) correct problem selection bias introducing as instrument the number of death from 1980-1988 through 2SLS estimation. Notice that equation (2) is estimated through a linear regression rather than a Probit estimation, following Angrist and Krueger (2001).<sup>19</sup> Covariate vector X includes sex, age, monthly household income, economic activity, education level at first migration, and house characteristics. Table 3 provides these estimations for each health outcome variable (BMI, SBP, DBP, and Glucose).<sup>20</sup> Regressions were estimated with standard robust errors and cluster them by district.

Columns (2), (4), (6) and (8) display OLS estimation for equation (1) for BMI, SBP, DBP, Glucose, respectively. These results suggest that migration has only a significant and positive impact for BMI and Glucose. In the case of BMI, migrants have 2.3 Kg/m<sup>2</sup> more than non-migrants. This result, without correcting for selection bias, can be considered as a conservative estimate given that migrants are usually healthier, more educated, and wealthier (biased towards zero). In the case of Glucose, results should be considered cautiously given that the R-squared is very low (0.061) suggesting that covariates incorporated in the model do not explain correctly that outcome variable.

Column (1) display estimation for equation (2). Columns (3), (5), (7) and (9) display estimation for equation (3) for BMI, SBP, DBP, Glucose, respectively. According to IV estimations, only BMI has a significant and positive effect: migrants have 3.6 Kg/m<sup>2</sup> more than non-migrants. The F-statistic for the first stage is 95.57, which according to Stock, Wright and Yogo (2002), the

<sup>19</sup> Angrist and Krueger (2001) state that “ In two-stage least squares, consistency of the second-stage estimates does not turn on getting the first-stage functional form right [...] So using a linear regression for the first-stage estimates generates consistent second-stage estimates even with a dummy endogenous variable” (p. 80).

<sup>20</sup> Table A4 (appendix) shows the correlation matrix of variables used for OLS and IV estimations including the instrument.

instrument used is not weak because the F-statistic exceeds 10. In the case of SBP, DBP and Logarithm of Glucose, we did not find statistically significant effect of migration.

In addition, it is worth to mention that, according to these regressions, males have lower BMI level, but higher SBP and DBP, coefficients statistically significant at 1%. With respect to BMI, this result is consistent with those provided by the National Survey of Nutritional, Biochemical, Socioeconomic and Cultural indicators related to Chronic Degenerative Diseases where women has higher BMI than men.

Interestingly, higher years of education decreases BMI and Glucose level which is coherent because higher level of education provides more knowledge about food quality (a direct effect), but also can provide higher income level, hence access to food with higher quality (an indirect effect).<sup>21</sup> Those variables are only significant for BMI and Glucose, but not for blood pressure suggesting that socioeconomic conditions are not important for explaining blood pressure.

In the case of BMI, another interesting result contradicting the previous conclusion is that years of education level at first migration is positive and significant. This means if an individual migrate with higher level of education, his/her probability to have higher BMI increases. In other words, a person who migrates at older age (given that the year of education level at first migration is a good proxy) will have higher level of BMI than a person who migrates at younger age.

In relation to Merchant as a main economic activity, I do not have a clear answer why people who work as a Merchant have higher BMI (and only significant for this variable). A plausible hypothesis is that mostly of merchants in Peru are sedentary: they stay sitting on their small store waiting for clients. Also, usually they work in small markets and they buy food within the market which has the feature of lower price, high quantity, but also low quality.

---

<sup>21</sup> However, notice that the income variable for the whole household is not statistically significant for BMI and Glucose.

Notice that most of socioeconomic variables (such as house conditions and economic activities) do not explain SBP, BDP, and Glucose. This effect, again, suggest that these variables are not determinants to explain blood pressure neither hypertension. Biological variables are more important in these cases (such as age or sex) than socioeconomic variables.

Finally, as robustness check I consider a different instrument: number of death within period 1980-2000, including 12 more years of data. Results with previous estimations should not vary given that early death (1980-1988) must influence migration decision, but not late deaths (1988-2000). If we find otherwise, it is a symptom of weak instrument. Results are show in Table A5 (appendix). In general, results are similar to previous estimations and now the impact of migration on BMI increases to  $4.0 \text{ Kg/m}^2$  which is not extremely different than previous estimation ( $3.6 \text{ Kg/m}^2$ ) supporting the instrument chosen.

**Table 3**  
**Effect of Migration on Health Outcomes: OLS and IV**  
**(Instrument: Death due to political violence 1980-1988)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Migration	BMI		SBP		DBP		Ln (Glucose)	
	First Stage	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Number of Death by district (1980 - 1988)	-0.00106*** (0.000387)								
Migration (1=migrant / 0=rural) 1/		2.310** (1.054)	3.612** (1.527)	2.268 (3.498)	7.144 (5.254)	-0.832 (2.174)	-2.438 (2.730)	0.0943*** (0.0282)	0.0819 (0.0649)
Male (1=Male) 1/	-0.0555** (0.0211)	-1.823*** (0.405)	-1.726*** (0.403)	6.733*** (1.635)	7.095*** (1.680)	4.879*** (1.040)	4.759*** (1.070)	-0.0206 (0.0200)	-0.0215 (0.0208)
Age	0.00007 (0.000845)	0.000375 (0.0182)	-0.000750 (0.0185)	0.597*** (0.0766)	0.593*** (0.0757)	0.0729 (0.0498)	0.0743 (0.0505)	0.00187*** (0.000677)	0.00188*** (0.000696)
Monthly Household Income	0.0442*** (0.0107)	0.0856 (0.198)	0.0132 (0.227)	-2.003*** (0.641)	-2.274*** (0.646)	-0.366 (0.386)	-0.276 (0.375)	0.00863 (0.00831)	0.00931 (0.00905)
Years of Education	-0.00168 (0.00353)	-0.128*** (0.0465)	-0.129*** (0.0474)	0.0257 (0.207)	0.0198 (0.210)	-0.0175 (0.122)	-0.0155 (0.121)	-0.00419** (0.00198)	-0.00418** (0.00198)
Merchant (=1) 1/	0.0209 (0.0180)	1.099** (0.501)	1.074** (0.498)	-2.133 (1.578)	-2.229 (1.624)	1.081 (1.280)	1.112 (1.274)	-0.00344 (0.0246)	-0.00320 (0.0246)
Unskilled worker (=1) 1/	0.0520*** (0.0162)	-0.166 (0.405)	-0.287 (0.395)	-0.608 (1.432)	-1.060 (1.573)	0.408 (1.009)	0.557 (1.062)	-0.00222 (0.0221)	-0.00105 (0.0223)
Education level at first migration	-0.0159*** (0.00600)	0.375** (0.160)	0.392** (0.156)	-0.0993 (0.628)	-0.0348 (0.629)	0.263 (0.377)	0.242 (0.378)	0.0118** (0.00531)	0.0117** (0.00516)
Water inside the house 1/	0.00003 (0.0350)	0.323 (0.734)	0.331 (0.712)	0.118 (2.667)	0.148 (2.718)	1.397 (2.272)	1.387 (2.271)	0.0171 (0.0218)	0.0169 (0.0220)
Floor (1 = Cement or Stone tiles) 1/	0.212*** (0.0629)	0.974 (1.087)	0.604 (1.130)	-1.454 (2.781)	-2.843 (3.571)	-2.623 (1.903)	-2.166 (2.142)	0.0400 (0.0317)	0.0435 (0.0391)
Wall (1= Cement) 1/	0.146** (0.0553)	-0.653 (0.782)	-0.938 (0.795)	-1.738 (2.372)	-2.806 (2.694)	-1.174 (1.456)	-0.822 (1.609)	-0.0538 (0.0440)	-0.0511 (0.0413)
Constant	0.664*** (0.133)	24.27*** (1.077)	23.81*** (1.024)	94.90*** (4.819)	93.17*** (5.563)	68.23*** (2.641)	68.79*** (2.872)	4.266*** (0.0343)	4.270*** (0.0367)
First Stage F-Stat	95.57								
Observations	461	461	461	461	461	461	461	460	460
R-squared	0.701	0.141	0.137	0.225	0.221	0.108	0.107	0.061	0.061

1/ Dummy Variables

Robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Estimation clustered by districts.

Source: Survey database.

## 4.2. Matching Estimation

An alternative way of estimating the effect of migration on health outcomes is the matching method. This method involves the direct pairing of treated and non-treated groups and compares the mean difference in their health outcomes.

One matching method is the propensity score. Rosenbaum and Rubin (1983) demonstrate that when outcomes are independent of the treatment or program participation, the probability of receiving a treatment can be estimated through a univariate parametric probability, known as the propensity score.

In general, the goal of this method is to find a statistically identical household for the migrant household given some observable characteristics (Dehejia and Wahba 1999, 2002). This methodology has some benefits since it does not require any distributional form assumptions, but also has some weaknesses due to the fact that the comparison group is selected only on observable characteristics. This method provides reliable results under large samples and high number of explanatory variables.

In relation to the weaknesses, Smith and Todd (2005) state that matching estimators to achieve lower bias must assure three conditions: (i) data collected must include a rich set of variables related to program participation, (ii) non-experimental comparison group must be drawn from the same sample as participants, and (iii) dependent variable must be measured in the same way for participants and nonparticipants. Those three conditions hold for the database used in this paper.

In this setting, following Dehejia and Wahba (2002), the optimal strategy is to find the control group before the Peruvian migration wave in order to reduce problems of unobservables. In other words, if we can find two statistically similar households before the migration wave, therefore we would be able to attribute that difference on health outcomes to the migration effect rather than socioeconomic and household conditions.

The database used for this analysis provides a group of question related to household circumstances during childhood (when interviewee was 10 to 12 years old) to handle this problem. The trade-off is losing observations since there can be some immigrants that moved to Lima city when they were younger than 10 or 12 years, hence, those household condition's questions are referring to their household in Lima, rather than in Ayacucho.<sup>22</sup>

Again, since self-selection arises between rural group and migrants from rural to urban areas, urban group is not considered in the analysis of migration decision. The following equation was estimated using a Probit model specification, where  $\Phi$  is the cumulative distribution function of the standard normal distribution given  $Z$ , a covariate vector of explanatory variables.

$$\Pr (Migration = 1|Z)_i = \Phi(\omega + Z_i\sigma) \quad (4)$$

In equation (4), covariate vector  $Z$  includes sex and age of migrant, parent's education and economic activity, number of people per room, and childhood household condition. Table 4 display results for equation (4). Column (1) provides the coefficients estimated, and column (2) provides marginal effects.

---

<sup>22</sup> 192 migrants moved when they had less than 12 years old. Those observations must be dropped for PSM analysis.

**Table 4**  
**Probit Estimation for Migration Decision**

Dependent Variable: Migration (1=migrant / 0=rural) 1/		
	(1)	(2)
	Coefficient	Marginal Effect
Male (1=Male) 1/	0.0400 (0.131)	0.0140 (0.0459)
Age	0.00946 (0.00639)	0.00332 (0.00223)
Father's Education Level: None	-0.863*** (0.278)	-0.300*** (0.0926)
Father's Education Level: Primary Incomplete	-0.624** (0.289)	-0.227** (0.107)
Father's Education Level: Primary complete	-0.639** (0.311)	-0.242** (0.121)
Father's Economic Activity: Farmer	-1.382*** (0.437)	-0.323*** (0.0545)
Father's Economic Activity: Merchant	-1.184* (0.620)	-0.445** (0.201)
Father's Economic Activity: Unskilled	-0.947* (0.516)	-0.363* (0.190)
Mother's Economic Activity: Farmer	0.806*** (0.159)	0.249*** (0.0416)
Mother's Economic Activity: Merchant	0.0557 (0.517)	0.0192 (0.176)
Number of people per room	0.0888*** (0.0239)	0.0311*** (0.00832)
Childhood Household Condition: Water & Sanitation inside the House	-0.525 (0.544)	-0.200 (0.217)
Childhood Household Condition: Firewood as Energy for cooking	-0.696* (0.376)	-0.199** (0.0816)
Childhood Household Condition: Cement floor	0.163 (0.551)	0.0548 (0.176)
Constant	2.016*** (0.640)	
Observations	450	
Pseudo R-squared	0.121	

1/ Dummy variables

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 4 illustrates that the lower father's education attained when interviewee were 10 to 12 years old, the lower probability to migrate. Also, if father's economic activity was farmer, merchant and unskilled ("*obrero*"), the interviewee had lower probability to migrate. It is quite surprising that in the case of mother's economic activity as farmer, affects positively in migration which is different as the sign found for the father, being both statistically significantly at 1%. I do not have a clear supporting reason for this finding however maybe land property rights may explain this difference and/or who is household head. Unfortunately, both hypotheses cannot be tested with this database. Finally, the number of people living per room (a measure of overcrowding) affects migration positively.

Controlling for those variables mentioned above, I estimate the impact of migration from rural to urban areas on health outcomes with several matching methods. These methods differ on how to match a treated observation (migrants) with a non-treated observation (rural) having a trade-off of bias and precision. I include Mahalanobis matching, Kernel matching, Radius Caliper matching and Local Linear Regression matching in order to check if results differ significantly or not across methods.<sup>23</sup> Results are shown in Table 5.

---

<sup>23</sup> Matching methods have a trade-off between bias and precision of estimates. Hence, it is better to show different methods in order to recognize how much they differ, if they do.

**Table 5**  
**Average Treatment Effect on the Treated (ATT) for Health Outcomes**<sup>1/</sup>

Matching Method	(1)	(2)	(3)	(4)
	BMI	Glucose	SBP	DBP
Mahalanobis Matching	3.866 *** (0.507)	10.048 *** (3.002)	0.212 (3.567)	-3.299 (2.035)
Kernel Matching				
- Bandwidth = 0.1	3.602 *** (0.533)	9.127 *** (2.645)	0.460 (3.134)	-2.869 (1.945)
- Bandwidth = 0.2	3.678 *** (0.467)	9.526 *** (3.459)	0.488 (3.205)	-2.843 (1.967)
Radius Matching				
- Caliper = 0.001	4.112 *** (0.446)	11.287 *** (3.326)	0.713 (3.099)	-2.819 (1.985)
- Caliper = 0.0001	4.047 *** (0.495)	1.179 *** (3.122)	-0.026 (3.559)	-4.539 (1.981)
Local Linear Regression Matching 2/				
- Bandwidth = 0.1	3.597 *** (0.510)	8.670 *** (2.998)	0.720 (3.192)	-2.680 (1.976)
- Bandwidth = 0.2	3.679 *** (0.511)	9.079 *** (3.151)	0.830 (3.208)	-2.695 (1.870)

1/ Standard Errors in parenthesis are estimated with Bootstrap (200 repetitions).

2/ Local Linear Regression estimated with Normal Kernel type.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Note: All Matching method restricted to common support.

Statistically significant differences are found for BMI and Glucose among all methods, but not in the case of blood pressure (SBP and BDP). On average, results show that migrants have higher BMI level by around 3.8 units of Kg/m<sup>2</sup> more than rural people (maximum level is 4.1 Kg/m<sup>2</sup> and the minimum level is 3.5 Kg/m<sup>2</sup>). In the case of Glucose, rural group has lower glucose level on average by 8.4 mg/dL (maximum difference is 11.3 mg/dL, and minimum difference is 8.7 mg/dL). This result is close to the mean average without controlling for observable covariates suggesting that the problem of self-selection bias and unobservables is not considerable.<sup>24</sup>

<sup>24</sup> Further estimation results can be found in Table A6.1 – A.6.4 (appendix).

Notice that in the case of Glucose different conclusion with IV estimation arises. In IV difference on Glucose between rural people and migrants is not significant, while in PSM this difference is significant. However, given that biological variables are more important for hypertension rather than socioeconomic variables, I interpret PSM's result as not conclusive because I only controlled for socioeconomic observable variables and not for biological observable variables.

After estimating the average treatment effect on the treated, I tested whether the two PSM main assumptions are satisfied or not. For this purpose, I evaluate whether the common support is similar for the control and treated group. Results are provided in Figure A1 (appendix) showing that histograms by treatment status are very similar.

Also, I calculate the Rosenbaum bounds to evaluate the presence of unobserved heterogeneity (meaning hidden bias) between treatment and control group. This test is based on the Wilcoxon Sign Rank test.<sup>25</sup> Results are provided in Table A7 (appendix). Results are robust up to a gamma factor of 4.0 which provides support that our ATT estimates remains significantly different from zero at 99%, hence are robust to the presence of unobserved bias.

## 5. Conclusions

This paper evaluates the impact of rural migration on health outcomes in Peru. This paper finds that migration has heterogeneous effect on health outcomes. Results suggest that the non communicable diseases of people who migrate from rural areas to an urban setting have a distinct pattern. These patterns differs compared to both people remaining in the rural area they migrated from, but also differs from people living in the urban area into which migration occurred. The findings challenge the view suspicions that following migration all risk factors amongst migrants will come to mirror the urban population, highlighting the complexity of migration and urbanization.

---

<sup>25</sup> As Rosenbourn (2005) mentions, in a randomized experiment  $\Gamma$  (gamma) is equal to 1 where everyone has the same chance of receiving the treatment. "A study is highly sensitive to hidden bias if the conclusions change for  $\Gamma$  just barely larger than 1, and it is insensitive if the conclusions change only for quite large values of  $\Gamma$ " (pp. 1810).

In particular, migration increases level of BMI comparing rural group and migrant group. Rural group has lower BMI level than the Migrant group, after correcting for selection bias. For the correction of selection bias I used two approaches: Instrumental Variable Method and Propensity Score Matching. Both methods provide similar conclusions, hence I can state clearly that migration from rural to urban areas increases level of fatness such as BMI.

On the other hand, migration has no effect on blood pressure (measured as SBP – Systolic Blood Pressure or DBP – Diastolic Blood Pressure). Furthermore, migration has impact on hypertension, but only with PSM and not with Instrumental Variable approach. Observing covariates that might have effects on hypertension we found that socioeconomic variables do not have high explanatory power. Hence, I conclude that socioeconomic conditions does not influence highly on hypertension, therefore results from PSM are not robust given that I am controlling only for socioeconomic observable characteristics, and not for biological variables.

It is important to mention that this study only reflects effect of migration from rural areas in the Andes to urban areas in the Coast (from Ayacucho to Lima). Hence, we cannot draw general conclusions about the impact of rural to urban migration in Peru. However, those results provide a big picture about its possible consequences in health outcomes.

To conclude, this study provides evidence on rural migration and its impact on health outcomes. Further studies are required in order to establish clear migration's impacts on health status and mechanisms of urbanization.

## 6. References

Angrist, Joshua; Guido Imbens; and Donald Rubin (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, June 1996, Vol. 91, 434, pp. 444-455.

Angrist, Joshua; and Alan Krueger (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives*, Vol. 15 (4), Fall 2001, pp.69–85.

Boehm, Bernhard; Simone Claudi-Boehm; Suzan Yildirim, et. al. (2005). Prevalence of the metabolic syndrome in southwest Germany. *Scandinavian Journal of Clinical & Laboratory Investigation*, Vol. 65, Suppl. 240, pp. 122–128.

Cameron, Colin and Pravin Trivedi (2005). Microeconometrics: Methods and Applications. Cambridge University Press.

Dehejia, Rajeev and Sadek Wahba (1999). Causal Effects in Nonexperimental studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*, Vol.94, 448. (Dec.1999), pp. 1053-1062.

Dehejia, Rajeev and Sadek Wahba (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *The Review of Economics and Statistics*, Vol. 84, 1 (February, 2002), pp. 151 – 161.

DiPrete, Thomas and Markus Gangl (2004). Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments. *Sociological Methodology*, Vol. 34, pp. 271-310.

Hagen-Zanker, Jessica and Carlo Azzarri (2008). Are internal migrants in Albania leaving for the better? Maastricht Graduate School of Governance Working paper MGSOG/2008/WP008.

Harris, John and Michael Todaro (1970). Migration, Unemployment and Development: A Two-Sector Analysis. *The American Economic Review*, Vol. 60, 1, pp. 126-142.

Instituto Nacional de Salud (2006). *Encuesta Nacional de Indicadores Nutricionales, Bioquímicos, Socioeconómicos y Culturales relacionados con las Enfermedades Crónicas Degenerativas – Informe Ejecutivo de la Encuesta*. Lima: Centro Nacional de Alimentación y Nutrición – Instituto Nacional de Salud. Lima, 2006.

INEI (2009). *Resultados Censo Nacional 2007 XI de Población y VI de Vivienda a nivel distrital*. Instituto Nacional de Estadística e Informática – INEI. Available at: <http://www.inei.gob.pe/>

Jackson, Elizabeth; Mark Doescher; Anthony Jerant; Gary Hart (2005). A national study of obesity prevalence and trends by type of rural county. *The Journal of Rural Health*, Vol. 21 (2), pp. 140–148.

Jokela, Markus; Mika Kivimaki; Marko Elovainio; Jorma Viikari; Olli Raitakari; Liisa Keltikangas-Jarvinen (2009). Urban/rural differences in body weight: Evidence for social selection and causation hypotheses in Finland. *Social Science & Medicine* Vol. 68, pp. 867-875.

Miranda, Jaime; Robert H Gilman; Héctor H García; and Liam Smeeth (2009). The effect on cardiovascular risk factors of migration from rural to urban areas in Peru: Peru Migrant Study. *BMC Cardiovascular Disorders* Vol. Vol. 9 (23).

MEF (2001). *Hacia la búsqueda de un nuevo instrumento de focalización para la asignación de recursos destinados a la inversión social adicional en el marco de la lucha contra la pobreza*. Working Paper – Ministry of Economic and Finance of Peru. November, 2001.

Rosenbaum, Paul and Donald Rubin (1983). The Central Role of the Propensity Score in Observational Studies for Casual Effects. *Biometrika*, Vol. 70 (1), pp. 41–55.

Rosenbaum, Paul (2005). Sensitivity Analysis in Observational Studies. Vol. 4, pp. 1809-1814. In: Brian Everitt and David Howell (Eds.). Encyclopedia of Statistics in Behavioral Science. John Wiley & Sons, Ltd, Chichester, 2005.

Sabates, Ricardo (2000). Job Search and Migration in Peru. *The Journal of Regional Analysis & Policy*, Vol. 30 (2), pp. 55-79.

Smith, Jeffrey and Petra Todd (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, Vol. 125, pp. 305-353.

Stock, James; Jonathan Wright; and Motohiro Yogo (1992). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics*, October 2002, Vol. 20, No. 4, pp. 518-529.

Szklarska, Alicja; Anna Lipowicz; Monika Lopuszanska; Tadeusz Bielicki; Slawomir Koziel (2008). Biological Condition of Adult Migrants and Nonmigrants in Wroclaw, Poland. *American Journal of Human Biology*, Vol. 20, pp. 139-145.

Sasin, Marcin and David McKenzie (2007). Migration, Poverty and Human Capital. Migration Operation Vehicle – Operational Note # 1. World Bank, Washington D.C., January 2007.

Torun, Benjamin; Aryeh Stein; Dirk Schroeder; Ruben Grajeda; Andrea Conlisk; Monica Rodriguez; Humberto Mendez; Reynaldo Martorell (2002). Rural-to-urban migration and cardiovascular disease risk factors in young Guatemalan adults. *International Journal of Epidemiology*, Vol. 31, pp. 218-226

WHO (2005). Preventing chronic diseases: A vital investment. WHO Global Report. WHO – World Health Organization. Geneva, 182 pp.

WHO (2002). The World Health Report 2002 - Reducing risks, promoting healthy life. WHO – World Health Organization. Geneva, 248 pp.

## APPENDIX

**Table A1**  
**Descriptive Statistics: San Juan de Miraflores and Santillana Districts**

Variable	Peru		Lima Department		San Juan de Miraflores District		Ayacucho Department		Santillana District	
	Total	%	Total	%	Total	%	Total	%	Total	%
<b>POPULATION</b>										
<b>Urban/Rural Residence</b>	<b>27,412,157</b>	<b>100</b>	<b>8,445,211</b>	<b>100</b>	<b>362,643</b>	<b>100</b>	<b>612,489</b>	<b>100</b>	<b>7,215</b>	<b>100</b>
Urban	20,810,288	75.9	8,275,823	98.0	362,643	100.0	355,384	58.0	625	8.7
Rural	6,601,869	24.1	169,388	2.0			257,105	42.0	6,590	91.3
<b>MIGRATION</b>										
Migrant Population (by place of birth)	5,460,296	19.9	2,781,145	32.9	180,859	49.9	39,805	6.5	245	3.4
Migrant Population (by place of residence 5 years ago)	1,537,099	6.2	614,648	8.0	46,030	13.9	25,697	4.7	145	2.4
<b>EDUCATION</b>										
Population with college degree (15 years and older)	5,922,674	31.1	2,610,357	41.4	115,206	42.7	75,742	19.4	88	2.3
Population without education (15 years old and older)	1,359,558	7.1	132,148	2.1	6,425	2.4	69,922	17.9	1,262	33.2
<b>HEALTH</b>										
Population with health insurance	11,598,698	42.3	3,525,794	41.7	135,595	37.4	335,322	54.7	5,082	70.4
Population with public health insurance ( <i>Seguro Integral de Salud</i> )	5,075,779	18.5	622,009	7.4	24,122	6.7	261,470	42.7	4,935	68.4
Population with private health insurance ( <i>ESSALUD</i> )	4,920,046	17.9	2,039,531	24.2	81,920	22.6	59,235	9.7	91	1.3
<b>MOTHER TONGUE (Children ≥ 5 years)</b>										
Spanish	20,718,227	83.9	7,202,159	93.2	293,867	88.4	194,216	35.7	175	2.9
Native Language 1/	3,919,314	15.9	510,385	6.6	38,072	11.5	348,720	64.1	5,898	97.0
<b>MAIN ECONOMIC SECTORS (≥ 14 years)</b>										
<b>Labor Force Population</b>	<b>10,163,614</b>	<b>100</b>	<b>3,611,300</b>	<b>100</b>	<b>157,868</b>	<b>100</b>	<b>191,173</b>	<b>100</b>	<b>1,512</b>	<b>100</b>
Agriculture and livestock	2,342,493	23.0	137,829	3.8	933	0.6	87,986	46.0	1,199	79.3
Construction	559,306	5.5	213,580	5.9	13,374	8.5	9,078	4.7	128	8.5
Commerce	1,689,396	16.6	746,364	20.7	34,599	21.9	22,840	11.9	36	2.4
Manufacture	943,954	9.3	477,799	13.2	19,620	12.4	7,342	3.8	7	0.5
Transport and Communication	848,916	8.4	362,467	10.0	16,425	10.4	7,774	4.1		
Education	610,159	6.0	205,789	5.7	8,280	5.2	15,067	7.9	24	1.6
Others	3,169,390	31.1	1,467,472	40.6	64,637	41.0	41,086	21.4	118	7.8
<b>DWELLING CHARACTERISTICS</b>										
<b>Wall materials</b>										
Brick or concrete	2,991,627	46.7	1,505,535	78.3	58,804	84.1	24,675	15.6	12	0.7
Adobe or mud	2,229,715	34.8	171,766	8.9	352	0.5	108,648	68.7	1,610	90.5
Wood	617,742	9.7	153,375	8.0	8,856	12.7	8,216	5.2		
Stone and mud	106,823	1.7	3,393	0.2	21	0.0	11,697	7.4	140	7.9
Others	454,224	7.1	87,880	4.6	1,909	2.7	5,025	3.2	17	1.0

<b>Floor materials</b>										
Dirt	2,779,676	43.4	320,107	16.7	9,218	13.2	123,349	77.9	1,728	97.1
Cement	2,441,884	38.2	977,720	50.9	43,012	61.5	29,329	18.5	44	2.5
Stone tiles	597,734	9.3	359,788	18.7	13,136	18.8	3,571	2.3		
Others	580,837	9.1	264,334	13.7	4,576	6.6	2,012	1.2	7	0.4
<b>Water Supply</b>										
Inside the House	3,504,658	54.8	1,412,156	73.5	58,739	84.0	63,842	40.3	328	18.4
Outside the House, but inside the compound	568,800	8.9	142,583	7.4	2,657	3.8	17,222	10.9	685	38.5
Pump	243,241	3.8	74,108	3.9	1,850	2.6	6,527	4.1	110	6.2
<b>Sanitation Facilities</b>										
Public network connection inside the House	3,073,327	48.0	1,393,858	72.5	58,416	83.5	39,967	25.3	209	11.7
Public network connection outside the House	393,506	6.1	142,661	7.4	3,126	4.5	7,794	4.9	177	9.9
Pit Latrine	1,396,402	21.8	169,856	8.8	4,526	6.5	47,291	29.9	997	56.0
<b>HOUSEHOLD ASSETS</b>										
<b>Total Households</b>	<b>6,754,074</b>	<b>100</b>	<b>2,075,091</b>	<b>100</b>	<b>82,591</b>	<b>100</b>	<b>163,147</b>	<b>100</b>	<b>1,784</b>	<b>100</b>
<b>Equipements</b>										
Radio	4,869,621	72.1	1,596,320	76.9	63,834	77.3	106,773	65.4	1,108	62.1
TV	4,116,857	61.0	1,775,672	85.6	71,598	86.7	51,153	31.4	110	6.2
Computer	998,222	14.8	540,414	26.0	17,438	21.1	10,210	6.3	2	0.1
<b>Communication</b>										
Home Phone	1,868,953	27.7	1,029,973	49.6	41,810	50.6	11,843	7.3	3	0.2
Mobile Phone	2,898,406	42.9	1,294,720	62.4	50,679	61.4	29,896	18.3		
Internet	458,158	6.8	301,028	14.5	7,929	9.6	2,398	1.5		
Cable TV	1,045,708	15.5	655,191	31.6	17,112	20.7	4,160	2.5	1	0.1
<b>Source of Energy for cooking</b>										
Gas	3,751,930	55.6	1,747,753	84.2	72,922	88.3	35,956	22.0	25	1.4
Firewood	2,036,901	30.2	75,060	3.6	772	0.9	115,341	70.7	1,624	91.0
Others	756,506	11.2	184,005	8.9	7,117	8.6	8,287	5.1	93	5.2

1/ Include Quechua, Aymara, Ashaninka and others.  
Source: INEI (2009). Peruvian Census 2007.

**Table A2**  
**Descriptive Statistics: Socioeconomic Variables**  
**(Including Minimum, Maximum and Standard Deviation)**

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Mean	Total Sample			Mean	Rural			Mean	Migrant			Mean	Urban		
		Min	Max	SD		Min	Max	SD		Min	Max	SD		Min	Max	SD
Age (years)	47.96	29	92	12.01	48.32	29	92	13.12	47.78	30	88	11.66	48.14	30	87	11.90
Male	0.47	0	1	0.50	0.47	0	1	0.50	0.48	0	1	0.50	0.46	0	1	0.50
Completed years of education	7.14	0	19	4.59	4.11	0	18	4.37	7.15	0	18	4.32	10.11	0	19	3.51
Total members per household	5.76	1	19	2.53	5.31	1	18	2.49	5.89	1	19	2.45	5.83	1	17	2.76
Total members > 18 years per household	3.59	0	14	1.79	2.59	1	6	1.10	3.89	0	14	1.89	3.69	0	10	1.71
Main Economic Activity																
Farmer	0.09	0	1	0.29	0.45	0	1	0.50	0.00	0	0	0.00	0.00	0	0	0.00
Merchant	0.15	0	1	0.36	0.05	0	1	0.22	0.20	0	1	0.40	0.12	0	1	0.32
Unskilled Worker	0.20	0	1	0.40	0.02	0	1	0.16	0.24	0	1	0.43	0.29	0	1	0.45
Assets																
Waterpipe and wastepipe	0.82	0	1	0.38	0.37	0	1	0.48	0.94	0	1	0.24	0.93	0	1	0.26
Gas use for cooking	0.72	0	1	0.45	0.01	0	1	0.12	0.90	0	1	0.30	0.92	0	1	0.26
Firewood use for cooking	0.21	0	1	0.41	0.99	0	1	0.12	0.02	0	1	0.13	0.00	0	0	0.00
Concrete floor	0.72	0	1	0.45	0.03	0	1	0.18	0.89	0	1	0.31	0.88	0	1	0.32
Concrete external walls	0.68	0	1	0.47	0.00	0	0	0.00	0.86	0	1	0.35	0.85	0	1	0.36
Concrete ceiling	0.49	0	1	0.50	0.00	0	0	0.00	0.63	0	1	0.48	0.60	0	1	0.49

Source: Survey database

**Table A3**  
**Descriptive Statistics: Health Outcomes**  
**(Including Minimum, Maximum and Standard Deviation)**

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Mean	Total Sample			Mean	Rural			Mean	Migrant			Mean	Urban		
		Min	Max	SD		Min	Max	SD		Min	Max	SD		Min	Max	SD
Height (cm)	154.22	128.3	186.8	8.05	152.11	128.3	173	7.66	154.12	134	175	7.65	156.61	132	186.8	8.93
Weight (Kg)	63.11	32.66	126.61	12.21	53.88	35.41	77.14	8.16	64.13	37.63	108.75	10.63	69.43	32.66	126.61	14.54
BMI (Kg/m <sup>2</sup> )	26.50	16.90	49.22	4.61	23.23	18.14	33.74	2.74	27.02	17.51	49.22	4.30	28.28	16.90	46.78	5.36
Obese (%) <sup>1/</sup>	0.20	0	1	0.40	0.03	0	1	0.17	0.21	0	1	0.41	0.34	0	1	0.48
Overweight (%) <sup>2/</sup>	0.38	0	1	0.49	0.16	0	1	0.37	0.46	0	1	0.50	0.37	0	1	0.48
Systolic Blood Pressure (mm Hg) <sup>3/</sup>	121.75	88	218.5	18.60	120.92	92	218	18.73	119.86	88	209	16.35	128.16	91	218.5	22.91
Diastolic Blood Pressure (mm Hg) <sup>3/</sup>	72.89	49.5	114.5	9.94	74.21	51.5	114.5	9.19	71.34	49.5	103.5	9.28	76.16	54.5	109.5	11.48
Glucose (mg/dL) <sup>4/</sup>	88.83	58	328	24.83	80.74	58	148	10.79	89.08	66	308	23.55	96.24	70	328	34.45

1/ BMI  $\geq$  30 Kg/m<sup>2</sup>

2/ 25 Kg/m<sup>2</sup>  $\leq$  BMI < 30 Kg/m<sup>2</sup>

3/ Hypertension is defined as: SBP - Systolic Blood Pressure  $\geq$ 140 mm Hg or DBP - Diastolic Blood Pressure  $\geq$  90 mm Hg

4/ Diabetes is defined as glucose  $\geq$  126 mg/dL

Source: Survey database

**Table A4**  
**Correlation Matrix for variables used in IV Estimation**

	BMI	SBP	DBP	Glucose	Male	Age	Income	Years Education	Merchant	Unskilled worker	Education first migration	Water	Floor	Wall	Migration	Death (80-88)	Death (80-00)
BMI	1																
SBP	0.1448	1															
DBP	0.2144	0.7203	1														
Glucose	0.1502	0.1427	0.0598	1													
Male	-0.2737	0.2161	0.2752	-0.087	1												
Age	0.0206	0.397	0.0732	0.1011	0.0536	1											
Income	0.0508	-0.0874	-0.0355	0.0424	0.1481	0.0085	1										
Years Education	-0.1131	-0.0854	0.053	-0.0777	0.2882	-0.3	0.2547	1									
Merchant	0.1535	-0.1057	-0.0394	-0.0015	-0.2205	-0.0251	-0.0155	-0.004	1								
Unskilled worker	-0.0954	0.0564	0.0896	0.002	0.4208	-0.0001	0.1961	0.1128	-0.2711	1							
Education at first migration	-0.003	-0.0417	0.0873	0.0081	0.2535	-0.2013	0.1751	0.5038	-0.0118	0.0936	1						
Water	0.1031	0.0328	-0.0152	0.0669	-0.0908	0.1992	0.2018	0.0169	0.0778	0.1006	-0.0251	1					
Floor	0.1699	-0.0849	-0.1585	0.1001	-0.0743	0.0066	0.3335	0.1284	0.0456	0.1379	0.0395	0.308	1				
Wall	0.1514	-0.0351	-0.1251	0.0432	-0.0663	0.1701	0.3419	0.0658	0.1182	0.1069	0.0104	0.4519	0.6588	1			
Migration	0.2124	-0.0438	-0.1392	0.1242	-0.1134	0.0736	0.4027	0.0551	0.0702	0.1759	-0.0284	0.3174	0.6276	0.6155	1		
Death (1980 - 1988)	-0.1538	-0.0256	0.0922	-0.0644	0.0143	-0.0691	-0.2598	-0.1154	-0.02	-0.1603	-0.06	-0.2026	-0.3891	-0.3984	-0.6921	1	
Death (1980 - 2000)	-0.1548	-0.0196	0.0828	-0.063	0.0146	-0.0646	-0.2547	-0.0845	-0.0009	-0.1491	-0.0355	-0.1849	-0.3749	-0.3821	-0.6635	0.9368	1

Source: Survey database

**Table A5**  
**Effect of Migration on Health Outcomes: OLS and IV**  
**(Instrument: Death due to political violence 1980-2000)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Migration	BMI		SBP		DBP		Ln (Glucose)	
	First Stage	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Number of Death by district (1980 - 2000)	-0.000889** (0.000373)								
Migration (1=migrant / 0=rural) 1/		2.310** (1.054)	4.021** (1.550)	2.268 (3.498)	7.570 (6.542)	-0.832 (2.174)	-1.307 (2.818)	0.0943*** (0.0282)	0.0943 (0.0629)
Male (1=Male) 1/	-0.0576*** (0.0212)	-1.823*** (0.405)	-1.696*** (0.414)	6.733*** (1.635)	7.127*** (1.716)	4.879*** (1.040)	4.843*** (1.085)	-0.0206 (0.0200)	-0.0206 (0.0208)
Age	0.000211 (0.000932)	0.000375 (0.0182)	-0.00110 (0.0186)	0.597*** (0.0766)	0.593*** (0.0757)	0.0729 (0.0498)	0.0733 (0.0501)	0.00187*** (0.000677)	0.00187*** (0.000695)
Monthly Household Income	0.0431*** (0.0103)	0.0856 (0.198)	-0.00951 (0.221)	-2.003*** (0.641)	-2.297*** (0.687)	-0.366 (0.386)	-0.339 (0.376)	0.00863 (0.00831)	0.00863 (0.00881)
Years of Education	-0.000716 (0.00348)	-0.128*** (0.0465)	-0.130*** (0.0479)	0.0257 (0.207)	0.0193 (0.210)	-0.0175 (0.122)	-0.0169 (0.121)	-0.00419** (0.00198)	-0.00419** (0.00199)
Merchant (=1) 1/	0.0283 (0.0175)	1.099** (0.501)	1.066** (0.497)	-2.133 (1.578)	-2.237 (1.626)	1.081 (1.280)	1.090 (1.274)	-0.00344 (0.0246)	-0.00344 (0.0247)
Unskilled worker (=1) 1/	0.0586*** (0.0173)	-0.166 (0.405)	-0.325 (0.408)	-0.608 (1.432)	-1.099 (1.613)	0.408 (1.009)	0.452 (1.071)	-0.00222 (0.0221)	-0.00222 (0.0223)
Education level at first migration	-0.0151** (0.00590)	0.375** (0.160)	0.397** (0.157)	-0.0993 (0.628)	-0.0292 (0.626)	0.263 (0.377)	0.257 (0.378)	0.0118** (0.00531)	0.0118** (0.00513)
Water inside the house 1/	-0.00218 (0.0348)	0.323 (0.734)	0.333 (0.708)	0.118 (2.667)	0.150 (2.726)	1.397 (2.272)	1.394 (2.276)	0.0171 (0.0218)	0.0171 (0.0218)
Floor (1 = Cement or Stone tiles) 1/	0.214*** (0.0665)	0.974 (1.087)	0.487 (1.137)	-1.454 (2.781)	-2.964 (3.872)	-2.623 (1.903)	-2.488 (2.176)	0.0400 (0.0317)	0.0399 (0.0378)
Wall (1= Cement) 1/	0.151** (0.0586)	-0.653 (0.782)	-1.027 (0.834)	-1.738 (2.372)	-2.899 (2.755)	-1.174 (1.456)	-1.070 (1.644)	-0.0538 (0.0440)	-0.0538 (0.0416)
Constant	0.633*** (0.150)	24.27*** (1.077)	23.67*** (1.075)	94.90*** (4.819)	93.02*** (5.851)	68.23*** (2.641)	68.39*** (2.828)	4.266*** (0.0343)	4.266*** (0.0367)
First Stage F-Stat	89.4								
Observations	461	461	461	461	461	461	461	460	460
R-squared	0.687	0.141	0.134	0.227	0.220	0.108	0.108	0.053	0.061

1/ Dummy Variables

Robust standard errors in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Estimation clustered by districts.

Source: Survey database.

**PSM & Average Treatment Effect on the Treated (ATT)**

**Table A6.1  
ATT for BMI**

Matching Method	Mean		ATT	Bootstrap S.E.1/	
	Treated	Untreated			
Mahalanobis Matching	27.189	23.323	3.866	0.507	***
Kernel Matching					
- Bandwidth = 0.1	27.189	23.588	3.602	0.533	***
- Bandwidth = 0.2	27.189	23.511	3.678	0.467	***
Radius Matching					
- Caliper = 0.001	27.370	23.257	4.112	0.446	***
- Caliper = 0.0001	26.883	22.836	4.047	0.495	***
Local Linear Regression Matching 2/					
- Bandwidth = 0.1	27.189	23.593	3.597	0.510	***
- Bandwidth = 0.2	27.189	23.510	3.679	0.511	***

1/ 200 repetitions

2/ Regress with Normal Kernel type

Note: All restricted to common support

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table A6.2  
ATT for Glucose**

Matching Method	Mean		ATT	Bootstrap S.E.1/	
	Treated	Untreated			
Mahalanobis Matching	90.904	80.857	10.048	3.002	***
Kernel Matching					
- Bandwidth = 0.1	90.904	81.778	9.127	2.645	***
- Bandwidth = 0.2	90.904	81.378	9.526	3.459	**
Radius Matching					
- Caliper = 0.001	91.813	80.526	11.287	3.326	***
- Caliper = 0.0001	84.357	83.179	1.179	3.122	***
Local Linear Regression Matching 2/					
- Bandwidth = 0.1	90.904	82.234	8.670	2.998	***
- Bandwidth = 0.2	90.904	81.825	9.079	3.151	***

1/ 200 repetitions

2/ Regress with Normal Kernel type

Note: All restricted to common support

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table A6.3**  
**ATT for SBP**

Matching Method	Mean		ATT	Bootstrap S.E.1/
	Treated	Untreated		
Mahalanobis Matching	120.105	119.893	0.212	3.567
Kernel Matching				
- Bandwidth = 0.1	120.105	119.645	0.460	3.134
- Bandwidth = 0.2	120.105	119.617	0.488	3.205
Radius Matching				
- Caliper = 0.001	119.550	118.838	0.713	3.099
- Caliper = 0.0001	117.868	117.895	-0.026	3.559
Local Linear Regression Matching 2/				
- Bandwidth = 0.1	120.105	119.386	0.720	3.192
- Bandwidth = 0.2	120.105	119.275	0.830	3.208

1/ 200 repetitions

2/ Regress with Normal Kernel type

Note: All restricted to common support

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table A6.4**  
**ATT for DBP**

Matching Method	Mean		ATT	Bootstrap S.E.1/
	Treated	Untreated		
Mahalanobis Matching	71.779	75.077	-3.299	2.035
Kernel Matching				
- Bandwidth = 0.1	71.779	74.648	-2.869	1.945
- Bandwidth = 0.2	71.779	74.621	-2.843	1.967
Radius Matching				
- Caliper = 0.001	70.950	73.768	-2.819	1.985
- Caliper = 0.0001	70.289	74.829	-4.539	1.981
Local Linear Regression Matching 2/				
- Bandwidth = 0.1	71.779	74.458	-2.680	1.976
- Bandwidth = 0.2	71.779	74.473	-2.695	1.870

1/ 200 repetitions

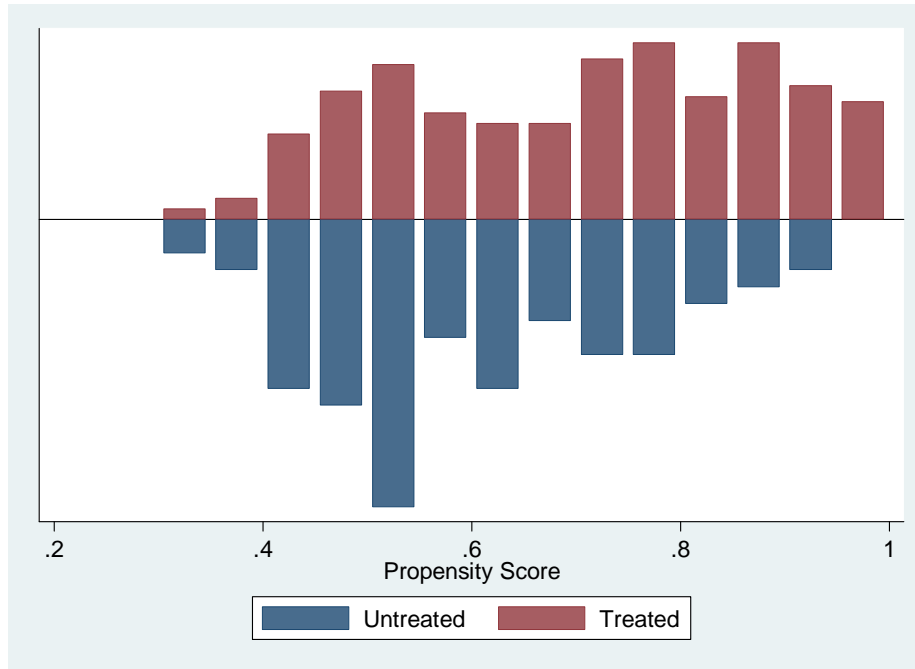
2/ Regress with Normal Kernel type

Note: All restricted to common support

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## PSM Post Estimation Analysis

**Figure A1**  
**Propensity Score Common Support**



**Table A7**  
**Rosenbaum sensitivity analysis**

Gamma	sig+	sig-	CI+	CI-
1.0	0	0	3.03308	4.14901
1.5	1.10E-16	0	2.19321	5.04809
2.0	5.50E-11	0	1.60537	5.70564
2.5	1.20E-07	0	1.15754	6.2375
3.0	0.000017	0	0.798594	6.66894
3.5	0.000444	0	0.494709	7.0299
4.0	0.004334	0	0.244499	7.35002
4.5	0.021789	0	0.025386	7.66221
5.0	0.069261	0	-0.181722	7.90917
5.5	0.158824	0	-0.366985	8.14507
6.0	0.287308	0	-0.528176	8.34762
6.5	0.436727	0	-0.670984	8.55862
7.0	0.58421	0	-0.816864	8.75688

sig+: upper bound significance level.

sig-: lower bound significance level

CI+: upper bound confidence interval.

CI- - lower bound confidence interval