Article

Socioeconomic status and self-reported chronic diseases among Argentina's adult population: results based on multivariate probability models

Valentina Viego, Karina Temporelli

Instituto de Investigaciones Economica y Sociales del Sur, CONICET; Department of Economics, Universidad Nacional del Sur, Argentina

Significance for public health

Latin American countries are going through an epidemiological transition where infectious illnesses are being superseded by chronic diseases which, in turn, are related to lifestyles and socioeconomic factors. Specificities in the relationship between chronic diseases and socioeconomic status have been recorded in high income countries, but has not been sufficiently studied in low and middle income countries. We believe that analysis grounded on large scale datasets, recently available in Argentina, and based on proper statistical tools can provide useful guidance for decision making in public health policies as they highlight where population needs and risks do concentrate.

Abstract

Background. Hypertension, diabetes and hypercholesterolemia are the most frequent and diagnosed chronic diseases in Argentina. They contribute largely to the burden of chronic disease and they are strongly influenced by a small number of risk factors. These risk factors are all modifiable at the population and individual level and offer major prospects for their prevention. We are interested in socioeconomic determinants of prevalence of those 3 specific diseases.

Design and methods. We estimate 3-equation probit model, combined with 3 separate probit estimations and a probit-based Heckman correction considering possible sample selection bias. Estimations were carried out using secondary self-reported data coming from the 2013 Risk Factor National Survey.

Results. We find a negative association between socioeconomic status and prevalence of hypertension, cholesterolemia and diabetes; main increases concentrate in the transition from low to high SES in hypertension and diabetes. In cholesterol, the major effect takes place when individual crosses from low to middle SES and then vanishes. Anyway, in Argentina SES exhibit and independent effect on chronic diseases apart from those based on habits and body weight.

Conclusions. Public strategies to prevent chronic diseases must be specially targeted at women, poorest households and the least educated individuals in order to achieve efficacy. Also, as the probability of having a condition related to excessive blood pressure, high levels of cholesterol or glucose in the blood do not increase proportionally with age, so public campaigns promoting healthy diets, physical activity and medical checkups should be focused on young individuals to facilitate prophylaxis.

Introduction

In the second half of the twentieth century, Latin American countries have experienced a marked change in their health conditions. At the present epidemiological transition, infectious diseases are progressively being displaced by chronic diseases related to lifestyles, while morbidity is increasingly replacing mortality rate as the leading health indicator. Also related to changes in the epidemiological profile, demographic dynamics has been determined by a decline in mortality rates, especially infant-juvenile, as well as a decrease in the fertility of populations. This process has been accompanied by improvements in life expectancy, which in recent decades have begun to include older groups, leading to the development of a process of gradual aging of the population.[1-5]

Emerging epidemiological, demographic and socioeconomic transformations impose new challenges on health systems, which are now threatened by the expectation of a geometric increase in the incidence of chronic diseases in the next years. In developing countries, the challenge is twofold. On the one hand, health systems must face the increase in the demand for health services caused by the huge increase in the prevalence of chronic diseases that require long and costly treatments.[6] On the other hand, these countries still need to solve the problem of transmittable diseases.[6-10]

This paper focuses on 3 specific chronic diseases, hypertension, diabetes and hypercholesterolemia, largely extended chronic diseases in the world. Every year about 9 million people die worldwide from high blood pressure-related causes. Moreover, hypertension is the second leading global cause of loss of life years due to disability.[11,12]

In turn, raised cholesterol increases the risks of heart disease and stroke. Globally, one third of ischemic heart disease cases are attributable to high cholesterol. Overall, raised cholesterol is estimated to cause 2.6 million deaths around the world and 29.7 million disability adjusted life years (DALYS).[11]

It is estimated that almost 382 million people suffered from diabetes in 2013 globally, which is equivalent to a prevalence of 8.3%.[11] In sum, they contribute largely to the burden of chronic disease and they are strongly influenced by a small number of risk factors. Chronic diseases and their risk factors show significant variation across population groups in terms of their incidence, prevalence, prevention, management, and associated health outcomes.[13]

The socioeconomic status (SES) is a strong predictor of health and risk of injury.[14] Usually, SES is proxied by education, occupational status and income,[15,16] although some studies include also health insurance as a social marker.[17] Nevertheless, socioeconomic

status is not considered a classical Farmington risk factor as it would not directly influence chronic diseases, but through behaviour.[18] Still, while studies linking socioeconomic conditions with chronic diseases are abundant in high-income countries,[19,20] their relation is yet not well understood in middle and low income economies.[21,22]

Argentina is not free from this problem; the selected diseases are the most frequent and diagnosed among adult population. Local epidemiology indicators show the growing importance of chronic diseases in the profile of population morbidity and mortality. Also, risk factors such as obesity, overweight and physical inactivity have been steadily rising, a trend that is compatible with the increase in the prevalence of the conditions mentioned above.

According to the *2013 Risk Factor National Survey* (hereinafter RFNS), nearly one third of the adult urban population exhibit high blood pressure and high cholesterol figures, while nearly 1 in 10 are diabetics. Previous national surveys report confidence intervals for prevalence of each of the 3 diseases depicting a stable situation; nevertheless the point estimates suggest an upward trend. Also, between 2005-2013 overweight and obesity figures increased significantly; based on *RFNS* data, figures for population ≥18 y.o. who self-reported weight and height figures leading to overweight or obesity were 50.6% in 2005; 54.9% in 2009 and 58.9% in 2013. Their corresponding confidence intervals do not overlap, confirming the increasing trend.

Although a significant proportion of the adult population (20-35%) does not receive periodic medical controls, between those adults who were checked by doctors, near 36% (7.8 million) were founded as hypertensive, 30.5% (4.9 million) recognize suffering from cholesterolemia and about 9.8% (2.5 million) were diagnosed as diabetic. Moreover, near 5% of individuals who were controlled (725,000 patients) experiment those 3 conditions simultaneously. These data show the relevance of a problem that needs to be approached by using all available information so that policies can focus on the most affected populations.

The objective of this paper is to determine the contribution of variables such as biological constitution, habits and socioeconomic status to high blood pressure, hypercholesterolemia and diabetes in the adult population in Argentina.

Materials and Methods

Model specification and variables definition

One shortcoming of the majority of public health research is that analysis are carried out on a bivariate basis: correlation between each risk factor and the probability of disease is checked separately. That approach could suffer from internal validity as it does not control other latent determinants and could be improved using partial correlation or regression-based techniques.

Consider the 3-equation probit model:

$$y^*_{im} = \beta'_m X_{im} + \epsilon_{im} \qquad i=1...N; m=1, 2, 3 \qquad [1]$$
$$y_{im} = 1 \text{ if } y^*_{im} > 0 \text{ and } 0 \text{ otherwise}$$

where y_{im}^* is a latent variable that represents the probability of the individual suffering from the *m*-th disease. This probability cannot be directly observed. Instead, the presence or absence of disease is observed through y_{im}, which takes only two values. ϵ_{im} represents random i.i.d. error terms.

Parameters of equation [1] can be estimated more efficiently through a 3-equation model, instead of estimating single equations

for each disease separately since this model takes into account that cross-equation error terms can be correlated as they originate in the same individual. A joint estimation relaxes the independence of error terms in [1] so the variance-covariance matrix *V* has now unity values on the leading diagonal and correlations $\rho_{jk} = \rho_{kj}$ as off-diagonal elements. The model is known as triprobit.

Moreover, each dependent variable is binary, so it should be estimated applying maximum likelihood estimators, which guarantee consistency of results. In the literature, a 3-equation model with 3 binary dependent variables is usually estimated by the method of simulated maximum likelihood (SML). In particular, the estimator uses the Geweke-Hajivassiliou-Keane (GHK) simulator to evaluate the 3-dimensional Normal integrals in the likelihood function. For a brief description of the GHK smooth recursive simulator, see Greene,[23] who also provides references to the literature.

Under standard conditions, the SML estimator is consistent as the number of observations and the number of draws tend to infinity, and is asymptotically equivalent to the true maximum likelihood estimator as the ratio of the square root of the sample size to the number of draws tends to zero. In this case, estimations used 5 random variates drawn when calculating the simulated likelihood. We also tried 133 draws (the root of the number of observations, n= √17520), following the recommendation of Cappellari and Jenkins.[24]

In addition, we applied the Huber-White sandwich estimator for the variance of the error term in order to take into account possible heteroskedasticity.

Variables descriptive statistics are displayed on Table 1 (operative definitions are exhibited in Table A1 in the Appendix).

Multivariate techniques would not justify the inclusion of social risk factors together with behavioural variables, as SES´s exert would express itself through habits. However, we believe that socioeconomic factors can exert an independent influence beyond registered behaviour; low SES can be associated with higher stress and pressure that in turn can precede chronic diseases.

Although the inclusion of most of the explanatory variables on selected chronic diseases requires no additional explanation, we provide some details for the variables related to SES. Besides from education level, occupational status, and income we also individual´s health insurance in order to better capture SES; market research and population surveys usually consider health insurance as a proxy for SES.[25] It may be tempting to expect bias in estimations emerging from adverse selection issues (unhealthy individuals more prone to buy health care plans). We must acknowledge that although this is a problem in some countries where health care usually takes part of individuals' choices, in our case its extent is reduced; in Argentina contribution to health insurance is mandatory for all formal actual and retired workers, also covering household members. That element explains why near 70-75% of population has health insurance, higher than US figures during the same period.[26]

It is worth mentioning that predictors related to family record, which probably add to the prospects of suffering any of the three diseases, were omitted from our estimations. The *RFNS* does not collect data about individual´s history in order to size appropriately hereditary factors conditioning disease emergence. Nonetheless, as those factors are not related to lifestyles and socioeconomic status, their omission does not affect the consistency of estimators.

The model also included the square of age in order to capture non-linear effects over the likelihood of any of the three diseases.

We compare the results from performing a joint estimation for the 3 diseases with those obtained from results obtained estimating equation by equation.

Population universe and data sources

The analysis is based on secondary data coming from the *2013 Risk Factor National Survey*, which was directed by the National Ministry of Health and conducted by provincial statistic bodies during the fourth quarter of 2013.

Health authorities constructed variables with information arising from direct interviews using a probabilistic sample of individuals 18 y.o. or older living in cities with 5000 inhabitants or more. That design confines the domain of inferential exercises to urban adult population, excluding thus children, adolescents and rural inhabitants. The survey collects data about housing conditions, demographic, educational characteristics and employment status of the household head. Also, individual variables about education, employment as well as information on risk factors for non-communicable diseases were recorded based on self-reported information.

Individual data emanates from a single adult member per household. Although 2013 NRFS actual sample size is 32,365, our estimations are based on varying samples ranging from 28,489 to 17,520 observations for various reasons; 2,075 individuals (6.4%) did not report weight or height data, additionally 34 individuals with achondroplasia (judged from height information) were excluded from the sample as they constitute a different population with specific risk factors and related diseases. Also, just 65% of the individuals have controlled their cholesterol levels. Finally, near 5% of observations were missing due to lack of response in the rest of the dataset.

Although income categories exhibit the highest figures of missing observations (Table 1), it must be recognized that this variable has in fact only 101 missing cases (<1%) for the whole sample. The additional missing observations originate from the fact that descriptive statistics exclude individuals with achondroplasia, which is in turn based on height categories, the dimension with highest missing values. Besides body weight categories constitute the variable with highest missing values and could be an indicator of poor self-knowledge that could affect individual´s behavior about health care, that element does not introduce significant bias in estimations in Argentina.[27]

Even considering missing values, sample means are still similar to population averages, with slight over-representation of occupied and insured individuals (Table 1).

We must anticipate a kind of trade-off between estimation strategies; on the one hand, the joint estimation may increase the precision of results as it exploits data coming from the same individual. On the other, the fact a lower rate of individuals have they cholesterol levels checked, truncates the sample in joint estimation. In individual estimation strategy the reduced sample size affects only the equation for hypercholesterolemia while not the other 2. Thus we assess the effect of varying size sample on results.

Sample selection issues

We recognize that the definition of prevalence in diseases reduces the universe to those who had medical controls, who in turn may differ significantly from those haven't been controlled, introducing thus bias in the estimates. In statistical literature this phenomenon is known as sample selection problem. In fact, sample selection rises a specific kind of endogeneity in regression analyses.[28]

The solution to this possible source of bias was proposed by Heckman.[29] As the dependent observed variable, y_{im}, is not always observed as some individuals did not visit doctors, it is possible to add to the expression [1] the condition which captures if the dependent variable for the *i*-th observation is observed, thus Equation [1] can be re-written as:

$$y^*_{im} = \beta'_m X_{im} + \epsilon_{1im} \qquad i=1...N;\ m=1,2,3 \qquad [2]$$

$y_{im} = 1$ if $y^*_{im} > 0$ and 0 otherwise

$\alpha'_m Z_{im} + \epsilon_{im} > 0$ where

$\epsilon_{1im} \sim N(0,\sigma)$

$\epsilon_{2im} \sim N(0,1)$

$corr\,(\epsilon_{1im}, \epsilon_{2im}) = \rho$

Where γ is the correlation between the unobserved determinants of propensity to having medical control, ϵ_{1im} and unobserved determinants of suffering the *i*-th chronic disease,. The task here is predicting the likelihood of having been controlled for each individual, $\alpha_m Z_{im}$, using also a probit model. In our case, the probability of having received medical care was explained by its age, per capita income and the indicator of health coverage. When $\rho \neq 0$ the standard regression techniques applied to the first equation (probability of disease) yield biased results. On the contrary, when $\rho=0$ bias emerging from sample selection is negligible and Equation [2] estimates are still consistent but inefficient. Heckman proposed consistent, asymptotically efficient estimates for all the parameters in

Table 1. Descriptive statistics (excludes individuals with achondroplasia).

Variables	Descriptive statistics	Non missing values, no.
Medical checks, %		
Blood pressure	89.3	30,195
Cholesterol level	65.0	29,978
Glucose in blood	79.3	29,427
Disease prevalence, %		
Hta	35.4	26,814
Col	30.2	19,412
Diab	10.4	23,341
Basic information		
Age, years	44.6 [43.3]	30,257
Woman %	55.0 [52.8]	30,257
Economic factors, %		
Occupied,	62.9 [57.0]	30,257
Unemployed,	3.6 [3.5]	30,257
Not active	33.5 [39.5]	30,257
Illiterate / incompl primary school	10.1 [10.5]	30,257
Primary education	37.5 [40.1]	30,257
Secondary education	36.1 [35.6]	30,257
Tertiary education	16.4 [13.8]	30,257
Income quintile 1	21.1	30,164
Income quintile 2	19.2	30,164
Income quintile 3	19.4	30,164
Income quintile 4	18.8	30,164
Income quintile 5	21.5	30,164
Health insurance	74.9 [69.7]	29,931
Habits, %		
Alcohol abuse	7.3	29,801
Smoker	24.9	30,257
Past smoker	17.8	30,257
Sitting, hours per day	4.49	28,818
Intense physical activity	14.3	30,257
Moderate physical activity	31.6	30,257
Weight category, %		
Overweight	37.5	30,257
Class 1 obesity	15.3	30,257
Class 2 obesity	4.4	30,257
Class 3 obesity	1.7	30,257

Source: own based on *2013 RFNS*. [In brackets]: national urban figures based on *2013 Urban Households Annual Survey*.

models under sample selection suspect.[29] It is worth to mention that although the canonical model considered a continuous outcome, subsequent progress has adapted the former proposal for binary dependent variable in the outcome equation.

Unlike Equation [1] it is not possible yet to estimate [2] jointly for the 3 diseases, so we estimated them separately by each disease. Estimations and analysis were carried out in October 2016 using Stata 13.

Results

As with any regression model with binary dependents estimated by maximum likelihood, coefficients cannot be interpreted as treatment effects. Instead, results usually focus on coefficients' statistical significance, sign and relative magnitude. Marginal contributions of explanatory variables should be then computed manually (*e.g.* assigning values for each individual predictor and then varying some of them). We choose an individual profile for computing marginal effects considering a woman with average age (*e.g.* 44 y.o.), middle SES, sedentary and normal body weight (Table A2). As our main goal is to highlight and also assess the contribution of SES level on disease risk, we focused the effect calculations to variables related to SES and choose the rest of the values at the modal level.

Table 2 exhibits marginal and discrete effects of the 3 methods applied, that is, the joint triprobit model, the probit approach and the Heckman correction. In the latter two cases, estimations were performed equation by equation and may suffer some inefficiency. Including individuals with achondroplasia does not change sign, significance and effects of any explanatory. Excluding habits from estimations (that is, alcohol abuse, sitting daily hours and exercise) decrease near 2 pp the effect of sex and income on the probability of high levels of blood pressure. That variation can be attributed to the high association between SES and habits, as mainstream literature pose. We do not include results from that checks because of length restrictions but are available on demand. In turn, the fact that SES-related coefficients keep individual significance and sign shows that collinearity between repressors is not serious and highlights the probable independent (emotional) channel through which SES exerts influence on chronic diseases.

In the case of cholesterolemia and diabetes, we found that there is no significant estimation bias based on sample selection. In uncontrolled individuals by blood pressure, differences in marginal effects are significant but do not reach the 2 digit level. In particular, Heckman correction does not change size effects for non-participation in the labour force, elementary school, and habits, but sex and unemployment display higher effects on probability of hypertension and income and body weight categories show a lower effect than the triprobit specification.

Explanatory variables for the selection equation in Heckman model were all significant at 1%. Additional results are available by request to authors. They are not included here in order to attend journal's policy about articles' lenght. Also, the suspect of adverse selection between health insurance and the probability of either of the 3 diseases should turn positive the coefficient of health insurance either in separate or joint probit estimations. None of the estimated coefficients were positive, weakening evidence in favour of adverse selection and endogeneity in probability models of selected chronic diseases.

Thus, while in hypertension sample selection could pose an issue when evaluating risk factor gradients, in cholesterol or diabetes, a simultaneous estimation strategy could lead to more effi-

cient results. In the case of triprobit estimation, the statistic that checks if equations are correlated is significant at 1% revealing that joint estimation lead to more efficient estimates. Also, individual significance, sign and relative magnitude of the coefficient do not change significantly with the number of draws, so our estimations are stable. However, separate probit estimations increase significantly the sample size for high blood pressure and diabetes as they can include individuals that did not controlled their cholesterol levels. Although results do not change significantly between separate and joint probit estimation in hypertension and hypercholesterolemia, the variation in sample size affect the results for some of determinants of diabetes; in particular, the effect of age, sex and low occupational status are higher in probit than in triprobit scheme while education and income exhibit lower effect on the probability of diabetes. Nevertheless, the relative importance of explanatories tend to be stable between models. If medical checks rates were similar among the population, joint (triprobit) estimates would be preferable to separate equations as samples would not vary substantially.

Considering their magnitude, the most important predictors of hypertension, hypercholesterolemia, and diabetes is weight category. In particular, morbid and severe obesity (class 2 and 3) exhibit the highest coefficients on the individual probability of suffering from any of the considered diseases. Their effect is more important in hypertension and diabetes than in hypercholesterolemia.

In the second place, the probability of chronic diseases among individuals who gained access to higher education is lower. Also, extreme poverty (*i.e.* the poorest 20% of households) increases the probability of suffering from hypertension and diabetes. In turn, per capita income levels affect the probability of being diagnosed with hypercholesterolemia only in the highest strata.

Habits appear in the third place in terms of coefficient magnitude. In particular, individuals that recognize regular risky consumption of alcohol exhibit higher probability of hypertension and hypercholesterolemia. On the contrary, the coefficient of alcohol abuse in the equation for diabetes prevalence is opposed to the expected, as it is individually significant but negative. Also, having smoked in the past contributes to diabetes and high blood pressure.

The magnitude of the probability of females suffering from high blood pressure is similar to the one corresponding to education or poverty; likely, gender exerts an effect over the probability of occurrence of hypercholesterolemia that is similar in degree to the one exerted by habits.

Finally, as expected, ageing increases the prospects of chronic diseases, but at a decreasing pace. This finding implies that risks of suffering from chronic diseases are age related, but are especially higher for individuals with an unfavourable past (that age enhances). Otherwise, if hypertension, hypercholesterolemia or diabetes were mere consequences of ageing, chronic diseases would increase exponentially with age (*e.g.* that is, the coefficient of the square of age would be positive). This result implies that prevention campaigns should be focused on young individuals and should be long lasting in order to favour changes towards increasingly healthier habits.

As the main focus of our analysis is directed to assess the burden of socioeconomic status on chronic diseases, we've estimated predicted individual probabilities of suffering each disease. Table 3 exhibits the estimated probability of each chronic disease for women with average age, habits and body weight profiles (see Table A2 for details) with different SES profiles.

Joint size effects of SES on the probability of suffering from chronic diseases vary from one estimation method to another one. Low SES increases between 5.3 and 7.5 percentage points (pp) the probability of suffering hypertension relative to middle SES and

between 7-9 and 16.4 pp relative to high SES levels. The impact of SES in the probability of cholesterolemia is rather lower: low SES increases between 1.3 and 5.7 pp the probability of high cholesterol levels in blood relative to middle SES and 2.8-4.9 pp relative to high SES. Also, the effects from middle to high SES are negligible. Finally, low SES increases 2.2-3.4 pp the probability of suffering diabetes in comparison with individuals with middle SES and the difference reaches 3.5-5.3 pp compared to individuals

Table 2. Marginal effects from triprobit, probit and Heckman estimates for a 44 y.o. woman, middle SES, sedentary and normal body weight.

Variables	Triprobit°			Separate probits			Heckman correction		
	phta	pcol	pdiab	phta	pcol	pdiab	phta	pcol	pdiab
Basic information									
Age	0.0434***	0.0597***	0.0312***	0.0331***	0.0592***	0.0400***	0.0362***	0.0589***	0.0360***
	(0.00389)	(0.00392)	(0.00459)	(0.00311)	(0.00383)	(0.00375)	(0.00301)	(0.00584)	(0.00445)
Agesq	-0.0002***	-0.0004***	-0.0002***	-9.52e-05***	-0.0004***	-0.0003***	-0.0001***	-0.0004***	-0.0002***
	(3.77e-05)	(3.76e-05)	(4.34e-05)	(3.12e-05)	(3.68e-05)	(3.65e-05)	(3.05e-05)	(4.22e-05)	(3.94e-05)
Woman	0.137***	0.0403*	0.0225	0.177***	0.0410*	0.101***	0.172***	0.0410*	0.0365
	(0.0224)	(0.0226)	(0.0269)	(0.0190)	(0.0222)	(0.0236)	(0.0184)	(0.0222)	(0.0244)
Socioeconomic status									
Unemployed	0.0517	0.0238	0.0800	0.0940*	0.0185	0.150**	0.0889*	0.0186	0.155**
	(0.0659)	(0.0665)	(0.0806)	(0.0503)	(0.0654)	(0.0629)	(0.0483)	(0.0655)	(0.0672)
Not active	0.0843***	0.0247	0.142***	0.0858***	0.0217	0.159***	0.0813***	0.0218	0.154***
	(0.0262)	(0.0266)	(0.0314)	(0.0221)	(0.0262)	(0.0275)	(0.0215)	(0.0262)	(0.0285)
Illiteracy	0.270***	0.0707	0.183***	0.256***	0.0644	0.0875*	0.251***	0.0644	0.176***
	(0.0432)	(0.0430)	(0.0503)	(0.0375)	(0.0424)	(0.0448)	(0.0364)	(0.0424)	(0.0463)
Primary education	0.187***	0.0585*	0.0799**	0.192***	0.0588*	0.0211	0.186***	0.0589*	0.0810**
	(0.0314)	(0.0314)	(0.0385)	(0.0278)	(0.0310)	(0.0351)	(0.0271)	(0.0311)	(0.0357)
Secondary educ	0.125***	-0.0123	0.0458	0.112***	-0.00940	0.0172	0.109***	-0.00931	0.0374
	(0.0296)	(0.0295)	(0.0369)	(0.0265)	(0.0291)	(0.0339)	(0.0258)	(0.0292)	(0.0344)
Income quintile 1	0.183***	0.0279	0.150***	0.157***	0.0278	0.0947**	0.134***	0.0296	0.143***
	(0.0369)	(0.0368)	(0.0436)	(0.0305)	(0.0361)	(0.0376)	(0.0300)	(0.0460)	(0.0417)
Income quintile 2	0.135***	-0.0288	0.118***	0.0826***	-0.0294	0.0567	0.0659**	-0.0279	0.0781**
	(0.0333)	(0.0333)	(0.0396)	(0.0285)	(0.0328)	(0.0353)	(0.0280)	(0.0405)	(0.0380)
Income quintile 3	0.0838***	0.00435	0.0366	0.0768***	-7.73e-05	0.0389	0.0634**	0.00117	0.0442
	(0.0313)	(0.0312)	(0.0379)	(0.0274)	(0.0307)	(0.0341)	(0.0268)	(0.0368)	(0.0362)
Income quintile 4	0.0578*	-0.0777**	0.0208	0.0419	-0.0769***	0.0127	0.0328	-0.0760**	0.0156
	(0.0303)	(0.0303)	(0.0371)	(0.0270)	(0.0299)	(0.0339)	(0.0264)	(0.0335)	(0.0350)
Health insurance	-0.0691**	-0.0321	-0.0481	-0.0354	-0.0228	0.0534*	0.00201	-0.0254	0.0305
	(0.0297)	(0.0300)	(0.0355)	(0.0231)	(0.0294)	(0.0291)	(0.0229)	(0.0512)	(0.0408)
Habits									
Alcohol abuse	0.106**	0.0805*	-0.119**	0.0716**	0.0971**	-0.108**	0.0693**	0.0970**	-0.0753
	(0.0431)	(0.0430)	(0.0548)	(0.0346)	(0.0418)	(0.0459)	(0.0335)	(0.0418)	(0.0484)
Smoker	-0.00540	0.00546	-0.0419	-0.00410	0.00691	-0.0334	-0.00333	0.00687	-0.0235
	(0.0266)	(0.0267)	(0.0328)	(0.0219)	(0.0261)	(0.0279)	(0.0213)	(0.0261)	(0.0292)
Past smoker	0.0981***	0.109***	0.131***	0.109***	0.111***	0.136***	0.106***	0.111***	0.128***
	(0.0261)	(0.0260)	(0.0301)	(0.0228)	(0.0257)	(0.0271)	(0.0221)	(0.0257)	(0.0280)
Sitting, daily hs	-8.34e-05	0.00774**	0.0169***	0.001000	0.00720**	0.0194***	0.00109	0.00718**	0.0168***
	(0.00363)	(0.00360)	(0.00418)	(0.00313)	(0.00356)	(0.00372)	(0.00304)	(0.00357)	(0.00384)
Intense phys act	-0.0376	-0.0157	-0.0708*	-0.0441	-0.0291	-0.104***	-0.0414	-0.0292	-0.0905**
	(0.0331)	(0.0332)	(0.0416)	(0.0271)	(0.0327)	(0.0359)	(0.0263)	(0.0327)	(0.0378)
Moderate phys act	-0.0557**	-0.0323	-0.0158	-0.0532***	-0.0346	-0.0160	-0.0514***	-0.0347	-0.0180
	(0.0229)	(0.0231)	(0.0275)	(0.0194)	(0.0227)	(0.0240)	(0.0189)	(0.0227)	(0.0249)
Weight category									
Overweight	0.290***	0.201***	0.237***	0.254***	0.202***	0.222***	0.246***	0.202***	0.220***
	(0.0244)	(0.0245)	(0.0306)	(0.0205)	(0.0241)	(0.0264)	(0.0200)	(0.0241)	(0.0277)
Class 1 Obesity	0.532***	0.297***	0.443***	0.492***	0.305***	0.427***	0.475***	0.305***	0.411***
	(0.0302)	(0.0305)	(0.0358)	(0.0258)	(0.0300)	(0.0314)	(0.0257)	(0.0300)	(0.0332)
Class 2 Obesity	0.731***	0.471***	0.621***	0.692***	0.487***	0.619***	0.668***	0.487***	0.593***
	(0.0469)	(0.0466)	(0.0518)	(0.0408)	(0.0458)	(0.0457)	(0.0401)	(0.0458)	(0.0480)
Class 3 Obesity	0.814***	0.414***	0.723***	0.796***	0.415***	0.755***	0.767***	0.415***	0.737***
	(0.0723)	(0.0722)	(0.0752)	(0.0642)	(0.0705)	(0.0674)	(0.0626)	(0.0705)	(0.0707)
Constant	-2.470***	-2.502***	-2.562***	-2.255***	-2.497***	-3.038***	-2.445***	-2.481***	-2.831***
	(0.111)	(0.112)	(0.133)	(0.0861)	(0.109)	(0.107)	(0.0829)	(0.286)	(0.165)
No. obs	17,52	17,52	17,52	25,053	18,171	28,011	28,253	28,489	27,765
ρ							0,519	-0,011	0,221
Statistic indep eq	591,043***						30,24***	0,01	1,58

°Run with 5 draws. (Robust standard errors in parenthesis). ***1%, **5%, *10%.

with high SES. Major differences in size effects concentrate in hypertension, where Heckman correction probed more appropriate than triprobit. In that case, the triprobit approach though more efficient than separate equations, tends to underestimate the effect of SES on hypertension. In hypercholesterolemia and diabetes, where sample selection bias was not significant, the gain in efficiency attained by simultaneous estimation can be offset by missing observations. A summary result, independent from the estimation method, is that negative association between SES and probability of disease is gradual (monotonous) in hypertension; while in cholesterol SES effect emerges mainly from low to middle SES categories and in diabetes the effect takes place only in radical SES shifts (from low to high SES), the transition between intermediate ones categories (low to middle, middle to high) does not exhibit substantial impact on the individual probability of diabetes.

Any of the three variants exhibit satisfactory sensitivity (recall) figures, especially in cholesterolemia and diabetes which exhibit rates higher than 70%. None of the three models exhibit absolute superiority in terms of goodness of fit; triprobit performs better in hypertension and diabetes but Heckman specification achieves higher sensitivity figures in cholesterol. In turn separate probit produces results with better specificity indicators. Also, models display rather high false positive rates, which may be interpreted as positive predictive values in individuals that do not report positive diagnosis, either because of memory failure or lack of adequate control, although their predictor variables exhibit warning values.

Discussion

This section discusses some contradictory results, uncover aspects that could affect the validity of our results and proposes ways of improve them.

First, prediction figures (Table A3 in the Appendix) show a considerable number of individuals with high probability of hypertension, hypercholesterolemia or diabetes that have not been diagnosed. Technically, they are classified as false positive observations (type I error). This could be interpreted in several ways; first, the model should be improved in order to raise precision. Second, false positives may be masking memory failures in individuals who do not remember a positive diagnosis. Third, a great proportion of population under risk has not been correctly identified by health care systems. While the first interpretation calls for improve models in order to predict more accurately the risk of each disease,

the last two alternatives highlight the importance of medical checkups. Related to that, we must acknowledge that underreporting and/or inadequate controls may be concentrated in low SES individuals, affecting possibly our results. In a sense, this is also linked to the sample selection problem emerging from individuals who have not received medical checks. In favour of our results we must emphasize that Heckman correction does not produce significant bias in estimates for cholesterolemia and diabetes, which in turn show the highest false positive rates. Finally, as self-reported data are the only source of information, we cannot measure the incidence of underreporting. But if underreporting was higher in low SES individuals, SES effects on chronic diseases would be also higher than our estimations. So marginal effects reported here could be considered as a minimum.

Second, the NRFS does not collect information about (current and past) labour tasks and workday. Both can play an important role in hypertension and hypercholesterolemia (as they influence diet and stress levels). That information would allow more effective prophylactic measures to be taken (e.g. medical control at workplaces, regulation on workday duration, promotion of restaurants offering healthy food, etc.).

Third, the negative relation between risky consumption of alcohol and the probability an individual suffers from diabetes may originate in endogeneity between alcohol consumption and diabetes diagnosis; individuals with unusual high levels of glucose in blood may be medically restricted in alcohol intake causing a coefficient negative and masking the true effect of alcohol abuse on diabetes. Future research should consider that source of bias in estimation strategy.

Fourth, the chosen approach is a kind of unmatched case-control study where cases (individuals affected by chronic diseases) are compared by regression techniques with controls (healthy individuals) with the aim of quantify the association between selected environmental variables and chronic diseases (adjusting for possible confounding factors). Matched case-controls studies are usually preferred over the unmatched, especially for diseases with low prevalence, as they better control for confounding factors and provide efficiency gains.[30] In that view, our results, particularly diabetes determinants, could be improved with a conditional analysis (i.e. matched case-control study). Anyway, the fact that our data are based on an extensive probabilistic sample of urban adult population and models included wide variety of determinants (sex, age, SES, habits, and weight categories), results are still accurate for guiding medical practices and health policies.

Finally, figures of hypertension, cholesterolemia and glycemic prevalence are based on self-reported data that were not validated by the NRFS with medical indicators (applied, for example in a subsample). This could affect reliability of our estimations. Anyway, several studies regard self-reported data as reasonably accurate for surveillance of chronic disease trends.[31-33]

Conclusions

The main contribution of this work is to highlight the role of the socioeconomic status in chronic diseases in order to guide public health policies focusing. Empirical approaches addressing this task may apply different individual probability modelling (e.g. probit or logit, joint estimation as triprobit or Heckman procedure when sample selection bias is suspected). In any case, they can provide a framework with stronger internal validity as they can consider simultaneously various types of risk factors, instead of assessing them based on bivariate statistics. In that view, our study

Table 3. Probability of chronic disease in women with average age, normal weight, and daily average sitting hours, by SES level.

Triprobit	phta	pcol	pdiab
Middle SES	0.262[t]	0.230[t]	0.089[t]
	0.304[p]	0.338[p]	0.092[p]
	0.266[h]	0.242[p]	0.091[h]
Low SES	0.321[t]	0.243[t]	0.123[t]
	0.379[p]	0.381[p]	0.091[p]
	0.319[h]	0.280[h]	0.111[h]
High SES	0.243[t]	0.271[t]	0.088[t]
	0.215[p]	0.333[p]	0.061[p]
	0.190[h]	0.237[h]	0.058[h]

t=mean probability estimated with a 3-equation probit model (triprobit); p=mean probability estimated with separate probit models for each disease; h=mean probability estimated with Heckman correction for possible sample selection bias.

illustrates pros and cons of every multivariate available method and also highlights some robust results concerning chronic diseases among adult population.

Strategies to reduce overweight and obesity as well as changing habits regarding alcohol and tobacco consumption play a key role in the prevention of chronic diseases in Argentina, also worsened by unfavourable socioeconomic conditions. So strategies must be specially targeted at women, poorest households and the least educated individuals in order to achieve efficacy.

Also, as the probability of having a condition related with excessive blood pressure, high levels of cholesterol or glucose in blood does not increase in direct proportion to age, public campaigns promoting healthy diets, physical activity and medical checkups should be focused on young individuals to facilitate prophylaxis and long lasting prevention.

References

1. Omran AR. The epidemiologic transition: a theory of the epidemiology of population change. Milbank Q 1971;49:509-83.
2. Caselli G. Health transition and cause-specific mortality. In: Schofield R, Reher D, Bideau A, eds. The decline of mortality 3. Frenk J, Frejka T, Bobadilla JL, et al. La transición epidemiológica en América Latina. Bol Oficina Sanit Panam 1991;111:485-96.
4. Robles Gonzalez E, Bernabeu J, Benavides FG. La transición sanitaria: una revisión conceptual. Bol Asoc Demografía Histor 1996;14:117-44.
5. Bernabeu Mestre J, Robles González E. Demografía y problemas de salud. Unas reflexiones críticas sobre los conceptos de transición demográfica y sanitaria. Rev Politica Sociedad 2000;35:45-54.
6. Olivares-Tirado P. Perfil epidemiológico del adulto mayor en Chile. Gobierno de Chile, Superintendencia de Salud. Departamento de Estudios y Desarrollo 2006: 1-20.
7. Morse S. Factors in the emergence of infectious diseases. Emerg Infect Dis 1995;1:7-15.
8. Satcher D. Emerging infectious: getting ahead of the curve. Emerg Infect Dis 1995;1:1-6.
9. Arriagada I, Arandana V, Miranda F. Políticas y programas de salud en América Latina. Problemas y propuestas. CEPAL 2005. Serie Políticas Sociales Nro. 114. LC/L.2450-P
10. Di Cesare M. El perfil epidemiológico de América Latina y el Caribe: desafíos, límites y acciones. CEPAL Documento de Proyecto (LC/W.395) 2005.
11. WHO. Información general sobre hipertensión en el mundo. 2013. Available from: http://apps.who.int/iris/bitstream/10665/87679/1/WHO_DCO_WHD_2013.2_spa.pdf
12. López A, Mathers C, Ezzati M, et al. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet 2006;367:1747-57.
13. AIHW. Chronic diseases and associated risk factors in Australia, 2006. Cat. No. PHE 81. Canberra: AIHW. Available from: http://www.aihw.gov.au/publication-detail/?id=6442467914. Accessed: 27 February 2016.
14. Lynch J, Kaplan, G, Pamuk E, et al. Income Inequality and Mortality in Metropolitan Areas of the United States. Am J Public Health 1998;88:1074-80.
15. Grotto I, Huerta M, Sharabi Y. Hypertension and socioeconomic status. Curr Opin Cardiol 2008;23:335-9.
16. Leng B, Jin Y, Li G, et al. Socioeconomic status and hypertension: a meta analysis. J Hypertension 2015;33:221-9.
17. Rahman S, Howard H, McNeely E, et al. Social and environmental risk factors for hypertension in African Americans. Fla Public Health Rev 2008;1:64-72.
18. Lam C. The socioeconomics of hypertension. Hypertension 2011;58:140-1.
19. Cavelaars AE, Kunst AE, Geurts JJ, et al. Differences in self-reported morbidity by educational level: a comparison of 11 western European countries. J Epidemiol Commun Health 1998;52:219-27.
20. Dalstra JAA, Kunst AE, Borrell C, et al. Socioeconomic differences in the prevalence of common chronic diseases: an overview of eight European countries. Int J Epidemiol 2005;34:316-26.
21. Koch E, Romero T, Romero CX, et al. Impact of education, income and chronic disease risk factors on mortality of adults: does a pauper-rich paradox exist in Latin American societies? Public Health 2010;124:39-48.
22. Linetzky B, De Maio F, Ferrante D, et al. Sex-stratified socioeconomic gradients in physical inactivity, obesity, and diabetes: evidence of short-term changes in Argentina. Int J Public Health 2013;58:277-84.
23. Greene WH. Econometric analysis. 5th ed. Upper Saddle River: Prentice-Hall; 2002.
24. Cappellari L, Jenkins S. Multivariate probit regression using simulated maximum likelihood. Stata J 2003;3:278-84.
25. SAIMO. El nivel socioeconómico en Argentina, 2015. Estratificación y variables. 2015. Available from: http://www.saimo.org.ar/archivos/observatorio-social/El-NSE-en-la-Argentina-2015-Estratificacion-y-Variables.pdf
26. Kaiser Family Foundation. State health facts. Health insurance coverage of the total population. 2013. Available from: http://www.kff.org/other/state-indicator/total-population/?currentTimeframe=2&selectedRows=%7B%22wrapups%22:%7B%22united-states%22:%7B%7D%7D%7D&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D. Accessed: 26 June 2017.
27. Temporelli K, Viego V. Malnutrición: impacto de los hábitos, variables constitutivas y condiciones socioeconómicas en la población adulta urbana en Argentina. Rev Ciencias Salud 2015;13:165-76.
28. Antonakis, J, Bendahan S, Jacquart P, Lalive R. On making causal claims: a review and recommendations. Leadership Q 2010;21:1086-120.
29. Heckman J. The common structure of statistical models of truncation, sample selection, and limited dependent variables

and a simple estimator for such models. Ann Econ Soc Measur 1976;5:475-92.

30. Rose S, van der Laan M. Why match? Investigating matched case-control study designs with causal effect estimation. 2008. UC Berkeley Division of Biostatistics Working Paper Series. Working Paper 240. Available from: http://biostats.bepress.com/ucbbiostat/paper240

31. Vargas C, Burt V, Gillum R, Pamuk E. Validity of self-reported hypertension in the national health and nutrition examination survey III, 1988-1991. Prev Med 1997;26:678-85.

32. Martin L, Leff M, Calonge N, et al. Validation of self-reported chronic conditions and health services in a managed care population. Am J Prev Med 2000;18:215-8.

33. Okura Y, Urban LH, Mahoney DW, et al. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. J Clin Epidemiol 2004;57: 1096-103.