WPS5575

Policy Research Working Paper 5575

Chronic Diseases and Labor Market Outcomes in Egypt

Lorenzo Rocco
Kimie Tanabe
Marc Suhrcke
Elena Fumagalli

The World Bank
Human Development Network
Health, Nutrition and Population Unit
February 2011

Policy Research Working Paper 5575

Abstract

By causing a sizeable reduction in employment 6 percent and labor supply 19 percent, chronic diseases are responsible for a major efficiency loss in the Egyptian economy. Furthermore the impact of chronic diseases on the labor market is not uniformly distributed. The older and the less educated suffer a larger drop in the probability of being employed and in their supply of working hours. The authors estimate the reduced form equations of individual employment status, labor supply and the usual wage equation. They control for unobserved ability and individual preferences by means of a within-siblings estimator. Measurement errors in our self-reported health variable have been accounted for.

This paper is a product of the Health, Nutrition and Population Unit, Human Development Network. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at ktanabe@worldbank.org.

Chronic Diseases and Labor Market Outcomes in Egypt.

Lorenzo Rocco [a] , Kimie Tanabe [b], Marc Suhrcke [c], and Elena Fumagalli [d]

Keywords: chronic diseases, Egypt, labor supply, wage

[a] University of Padova, Assistant Professor of Economics

[b] The World Bank, Economist (MNSHD)

c University of East Anglia, Associate Professor of Health Economics

[d] University of East Anglia, Research Associate

Correspondence Details:
Kimie Tanabe, Human Development Department (MNSHD), The World Bank. 1818 H Street NW, Washington DC, 20433, USA. Telephone: (202) 458-5333; fax: (202) 614-4161;
e-mail: ktanabe@worldbank.org ; website: http://www.worldbank.org .

Introduction.

According to WHO, in Egypt chronic diseases accounted for 78% of all deaths in 2002, well above communicable diseases (18%) and injuries (4%) (WHO, 2005). As a result of the epidemiological transition, the burden communicable diseases has declined, at the expense of an increasing burden of chronic. The transition was helped by massive vaccination campaigns implemented in the past several years: child immunization coverage increased from 86% in 1990 to 92% in 2007, and the under-five-mortality rate was reduced dramatically from 104 to 36.2 in the last two decades[1]. While life expectancy in Egypt has reached 70 years in 2007[2], the incidence of chronic diseases hinders large proportion of Egyptians to enjoy a long, healthy and prosperous life. Nonetheless, although age and genetic endowment contribute directly to the emergence of chronic diseases, risk factors such as smoking, mal-nutrition, lack of physical exercise play a major role.

In addition to their burden of pain and suffering, chronic diseases bring about important economic consequences. On one hand, they increase the demand of health care and, on the other hand, they may hamper people's ability to generate income, especially by increasing absenteeism at work or by ultimately impeding people to work.

In this paper we focus on the latter potential consequence. If chronic diseases reduce employment and labor supply, this would cause an efficiency loss to the economy as a whole, as the endowment of labor is not fully used and the number of "efficiency units" per worker remains below its potential. Therefore chronic conditions could contribute to keeping the country's economy below its production frontier. Whether or not this is the case should be of interest to economic policy-makers, whose objective it might be to maximizing the economy's production.

The primary purpose of this paper is to investigate whether such efficiency losses do in fact occur in Egypt. More specifically, we measure the causal impact of chronic diseases on the probability of being employed, on the number of working hours supplied per week and on the hourly wage earned in Egypt, among people in their working age 16-64. Although many papers have looked at the impacts of health on the labor market (for recent surveys, see Currie and Madrian, 1999; Rocco and Suhrcke, 2008), few have specifically focused on chronic diseases (Gammon, 2005; Wilson, 2001; Loprest et al. 1995; Suhrcke et al., 2007) and to the best of our knowledge no study has looked at middle income countries in MENA. The reason for focusing on Egypt is that it is one of the largest middle income countries, with a population of more than 70 million and, in many respects, it is representative of the Arab world, a particularly distinguished region around the

[1] The World Bank, WDI 2010 data: % of children under 1 year immunized against measles; and under 5 mortality rate per 1,000.
[2] The World Bank, WDI 2010 data: Life expectancy increased from 55 years in 1978 to 70 in 2008.

world where religious restriction of alcohol and pork meat consumption could influence the prevalence of chronic diseases.

The results of our analysis indicate that the presence of chronic conditions reduces the probability of being employed by about 25 percentage points (from 50 percent on average). Among the employed, the amount of working time supplied is reduced by about 22 hours per week. No impact results on wage rates. In a country where about one quarter of working-age individuals report to suffer from chronic diseases, the burden on employment and on aggregate supply of working hours is substantial. Extrapolating the results at the country level would suggest that the current employment rate of about 50 percent would be below its potential by about 6 percentage points (in the absence of chronic diseases the employment rate would be about 56 percent[3]). The aggregate labor supply, which combines the loss of employment and the lower number of hours worked by people reporting chronic conditions and still employed, is about 19 percent below its potential. Assuming that the national production function is of the usual Cobb-Douglas form in labor and capital, defined as $Y = AL^{0.6}K^{0.4}$, the implied production loss would be about 12% of Egyptian GDP.

It is not hard to imagine that the above effects are not uniformly distributed across society. Hence, the second purpose of the paper is to analyze how the negative impact of chronic diseases varies along dimensions such as age and education. Especially regarding the probability of being employed and hours of work supplied, the negative effects of chronic diseases are larger among the older and the less educated and those belonging to the informal economy. Therefore the impact of chronic diseases is not uniformly distributed across socio-economic groups, *on top of the fact* that the prevalence of chronic diseases is higher among the less educated, the older and to some extent the vulnerable groups. This implies that chronic diseases increase economic inequality in the population, because they hit the lowest socio-economic groups more frequently and harder.

The identification of the causal impact of chronic diseases is challenging, especially because observable and, mainly, unobservable individual characteristics might influence simultaneously both chronic conditions and work-related decisions, producing a spurious correlation between chronic conditions and labor market outcomes. For instance, smoking and diet can be a consequence of individual discount rate (which may stand for individuals' "patience") and risk aversion, and more generally of individual preferences which simultaneously determine the choice of participating into the labor market. Likewise, genetic endowment might determine both individual propensity to develop a chronic disease and cognitive ability (Butcher et al. 2006). Perhaps more importantly, family background and the experiences of childhood might influence both individual decisions about health and education, and preferences about work-related issues. For

[3] A 25 percent reduction in the probability of being employed among the 25 percent share of working-age people reporting chronic diseases is equivalent to about 6 percent points in terms of employment rate (0.25*0.25).

instance a child grown up in a wealthy household, with parents employed in highly ranked jobs, might have been pushed to pursue a high profile career and, at same time, might have been used and educated to a healthy life-style with a vegetable-rich diet and regular sport practice, and thus in a position to reduce the probability of subsequent chronic conditions.

To address these empirical problems, we look at the siblings living in those households surveyed by the Egypt Household Health Utilization and Expenditure Survey (EHHUES) 2002, a cross section composed of about 33,000 observations. Siblings are more similar in terms of genetic endowment, preferences and family background than any two randomly matched individuals. Therefore any pair of siblings shares a common family fixed effect that we can control for by means of within-siblings estimates. Of course, especially among siblings of rather different ages, it is likely that an important idiosyncratic component remains uncontrolled for (Griliches, 1979), as family background itself might change overtime (Ermish and Francesconi, 2001). To avoid this problem, within-siblings estimators have been often applied to subsamples of twins (Ashenfelter and Krugman, 1994; Ashenfelter and Rouse, 1998), as twins share exactly the same family background on top of an almost identical genetic endowment. Unfortunately the number of twins is too small in our data to be of any use. Thus, to check the robustness of our identification strategy, we approximated the ideal case of twins by carrying out our analysis on the restricted set of siblings whose age differs by at most two or three years, obtaining results closely in line with those obtained from the unrestricted sample.

A second empirical issue that we have taken into account is the fact that the indicator of chronic diseases is self-reported and therefore subject to possibly relevant measurement errors (responsible for an attenuation bias in the estimates).

The remainder of the paper is organized as follows. Section 2 presents the theoretical framework from which the empirical models are derived. Section 3 describes which variables are included in the model. The identification strategy and a description of the data and the samples are discussed in Section 4. The main results are discussed in Section 5. Section 6 contains several robustness checks to inspect the validity of our identification strategy. Section 7 describes how the effects of chronic diseases vary depending on education, age, household income and gender. Finally, Section 8 concludes. Two technical appendixes contain further technical details.

The model

An individual i residing in household j is endowed with the following utility function which summarizes his preferences over the set of work alternatives:

$$U_{ij} = w_{ij}(X_{ij}^{w})L_{ij} - \psi(L_{ij}, X_{ij}^{\psi}) + \varepsilon_{ij}$$

where $w_{ij}(X_{ij}^{w})$ is the wage rate the individual expects to earn, conditional on the characteristics X_{ij}^{w} relevant to his or her productivity, L_{ij} is the amount of time spent at work, $\psi(L_{ij}, X_{ij}^{\psi})$ is the sum of direct and opportunity costs of working. Direct costs include physical, psychological and monetary costs. They are increasing and convex in L_{ij} and depend on the individual characteristics X_{ij}^{ψ} (in general not coincident with X_{ij}^{w}). Of course individual health conditions are supposed to influence both wage rates and the disutility of labor. Finally, some zero-mean random shocks ε_{ij} might affect individual utility. Let $F(.)$ be the cumulate distribution of ε_{ij}, supposed symmetric.

In spite of its simplicity, this model is general enough to encompass three important relationships between health and labor market outcomes, i.e. how health conditions influence the decision of being employed, and, conditional on this, the amount of labor to supply and whether wage rates depend on individual health.

Employment

Denote by $X_{ij} = \left[X_{ij}^{w}, X_{ij}^{\psi}\right]$ the full set of individual characteristics. Given the concavity of U_{ij} in L_{ij}, there exists a unique utility maximizer $L_{ij}^{*} = L(X_{ij})$ which is a function of the full set X_{ij}. An agent in his working age chooses to be employed if $U_{ij}(L_{ij}^{*}) > 0$. This implies that the probability of being employed, conditional on characteristic X_{ij}, is $\Pr\left(E_{ij} \middle| X_{ij}\right) = F\left(w_{ij}(X_{ij}^{w})L_{ij}^{*} - \psi(L_{ij}^{*}, X_{ij}^{\psi})\right)$ which can be linearized as $\Pr\left(E_{ij} \middle| X_{ij}\right) = E\left(E_{ij} \middle| X_{ij}\right) = X_{ij}\beta$. This is the first equation of interest, which is estimable as a the linear probability model

$$E_{ij} = X_{ij}\beta + \mu_{ij} \tag{1}$$

where μ_{ij} is a zero mean error term independent of all X_{ij}.

Labor supply

A rational agent will supply the amount of work which maximizes his utility function. Such quantity will solve the FOC

$$w_{ij}\left(X_{ij}^{w}\right) = \psi'_{L}(L_{ij}, X_{ij}^{\psi})$$

so that

$$L_{ij} = \psi_{L}^{\prime -1}(w_{ij}\left(X_{ij}^{w}\right), X_{ij}^{\psi}) \tag{2}$$

which depends on the full set of individual characteristics X_{ij}. Upon appropriate linearization (see Appendix A) we get the second equation of interest,

$$L_{ij} = \lambda_0 + \lambda_1 X_{ij}^{w} + \lambda_2 X_{ij}^{\psi} + \upsilon_{ij} \tag{2'}$$

which relates the number of hours optimally supplied to X_{ij}.

Wage equation

Finally we state a usual Mincer equation to model the relationship between log wages and human capital inputs.

$$\ln(w_{ij}) = X_{ij}^{w}\alpha + \tau_{ij} \tag{3}$$

This specification is usually adopted to represent the fact that a worker's productivity (and hence his wage) depends on his own human capital and the fact that wages are dispersed also among people holding the same occupation.

Empirical specification

So far we have defined the model in very general terms. Nonetheless important hints about which variables to include have already emerged. For instance, the fact that wages are a function of individual characteristics, wages can be substituted out in both the employment and the labor supply equations. Moreover, there is a clear indication that the (only) set of variables to be included in our reduced form equations are those influencing individual wages (productivity) and those influencing the disutility from labor (X_{ij}^{w} and X_{ij}^{ψ}). We stress here that both observable and unobservable/non measurable variables do enter these sets.

As mentioned earlier, it is reasonable to assume that poor health reduces worker productivity and increases the disutility of labor. Specifically, we are interested in the role of chronic diseases and disabilities and the presence of such conditions is our measure of health conditions. In the dataset we use, Egypt Household Health Utilization and Expenditure Survey 2002 (EHHUES2002), a representative sample of people is asked "During the last 12 months have you been complaining from any persistent health problem for at least 3 months (including disability, disease, injury) or any other chronic disease?". About 25% of the subsample of people aged 16-64 (working age) answer affirmatively (27% of females and 22% of males). Chronic conditions are also the most important single factor influencing people's negative health conditions, according to EHHUES2002. For instance the probability of reporting bad health increases by 25% in case of chronic diseases, after controlling for age, gender and governatorate of residence.

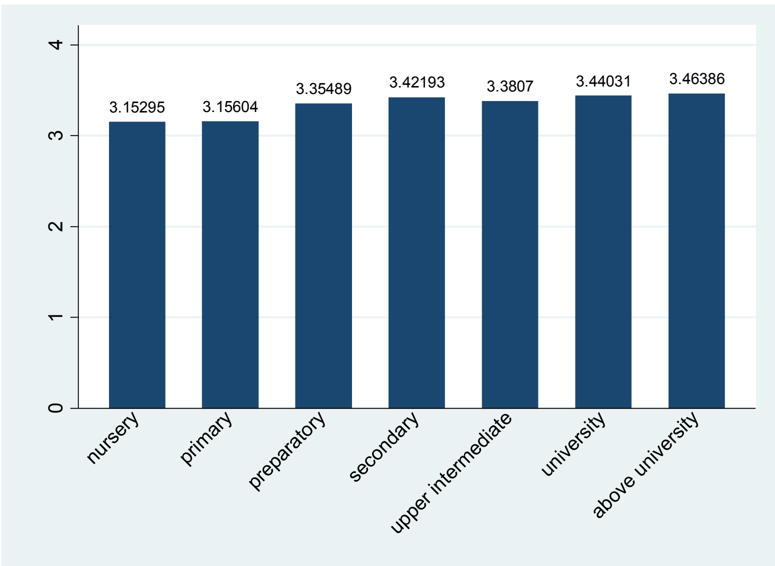
Following the literature of labor economics, the set X_{ij}^{w} includes human capital dimensions such as gender, education, experience accrued on the labor market (proxied by age), current health conditions (D_{ij}^{*}) and endowed ability (A_{ij}). Hence

$$\ln(w_{ij}) = \alpha_0 + \alpha_1 male + \alpha_2 edu + \alpha_3 age + \alpha_4 age^2 + \alpha_5 D^*_{ij} + A_{ij} + \tau_{ij}$$

There are two major empirical problems. First, endowed ability is largely unobservable. Unfortunately its omission from the wage equation will cause severe bias in the estimated parameters as ability and education are likely to be correlated. Although we speak about ability, A_{ij} has to be interpreted in broader terms: besides pure cognitive ability, it includes also individual preferences and specificities, such as risk aversion, patience, genetic endowment. Therefore A_{ij} is likely to be correlated with health conditions, too. For instance a low degree of risk aversion might induce the individual to take risks which negatively impact upon health. A high discount rate (indicating impatience) is likely to favor behaviors which have delayed bad health consequences, such as smoking and drinking. Finally the genetic endowment has been proven to be responsible of many chronic diseases. Controlling for A_{ij} is therefore crucial to obtain reliable estimates.

We decompose ability into its idiosyncratic and family component, by stating that $A_{ij} = I_{ij} + F_j$. The idiosyncratic component I_{ij} is, by definition specific to each individual, while F_j is the part common to household members. Needless to say, the more "similar" the household members, the greater the common component of their ability. For instance, genes would be much alike between siblings (and even more

Figure 1 - Self-reported health by education level



between twins) than between spouses.

The second empirical problem, especially relevant for health conditions, is that of misreporting. The correctly measured variable to be included in the model is D^*_{ij} but we observe only self-reported health conditions which might be distorted. Typically more educated people are likely to know their health status better than the poorly educated, because they are more informed, take frequent check-ups and are, in general, more attentive to their health (Bago d'Uva et al, 2008). In other words less educated people are likely to over-report their health conditions with respect to their true conditions. This is consistent with the surprisingly small positive gradient of self-reported health conditions that can be observed in Figure 1. On top of this, there are also random measurement errors. Let D_{ij} be self-reported health. We assume the following relationship between self-reported and true health conditions

$$D_{ij} - D_{ij}^* = \delta_0 + \delta_1 edu + \pi_{ij}$$

where δ_1 should be negative according to our hypothesis. The resulting equation to be estimated is then

$$\ln(w_{ij}) = (\alpha_0 - \delta_0) + \alpha_1 male + (\alpha_2 - \alpha_5\delta_1)edu + \alpha_3 age + \alpha_4 age^2 + \alpha_5 D_{ij} + \alpha_6 occupation +$$

$$+ I_{ij} + F_j + \tau_{ij} + \pi_{ij} \tag{3'}$$

By construction D_{ij} results to be correlated with the error term $\tau_{ij} + \pi_{ij}$. Typically, measurement errors result into an attenuation bias of the impact of health.

The variables included in the set X_{ij}^{ψ} are those which determine the disutility of labor. Only variables varying at individual level have to be included, while all those referring to household characteristics and, more generally, all those common to all siblings[4] are already captured by family fixed effects F_j. Marital status of women is likely to influence opportunity costs of working, because of the mutual support established within the couple. Among the variables affecting the direct costs, the type of occupation influences the costs of working to a large extent. Next, certainly bad health will increase the disutility of working and will also require spending additional resources to be able to keep on working. Health conditions are captured by the self-reported presence of chronic diseases of disability, as discussed above. Summing up, labor supply equation (2'), which includes the full set of variables $X_{ij} = \left[X_{ij}^{w}, X_{ij}^{\psi}\right]$, is specified as follows:

$$L_{ij} = \beta_0 + \beta_1 male + \beta_2 edu + \beta_3 age + \beta_4 age^2 +$$

$$+ \beta_5 D_{ij} + \beta_6 married + \beta_7 occupation + I_{ij} + F_j + \pi_{ij} + \mu_{ij} \tag{2''}$$

which differs from the wage equation only for the inclusion of the variable *married*. Of course the parameters β do not coincide with α. Except for *occupation*, the same specification applies to the employment equation (1). In this case, education and gender are assumed to capture the expected occupation of the employees.

[4] Relevant variables likely to modify the opportunity costs of working and common to all siblings are labor market conditions and commuting costs as well as indicators of family endowment and gender composition. For instance availability of land and/or livestock create an opportunity for housework, especially for women.

Identification strategy, sample selection and data

There are few surveys in Egypt including information on both health and labor market variables. Several rounds of the Demographic and Health Survey (DHS) focus mainly on health and only little attention is paid to labor. EHHUES 2002 is relatively rich in both dimensions, as it includes information on occupation, hours worked, cash and in kind payments, frequency of payments, although its primary focus is on health conditions, health care utilization and expenditure. Unfortunately this survey is cross-sectional only, further complicating the challenge of identifying the causal impact of chronic conditions. It has long been recognized that especially education and health are likely to be correlated with unobservable individual characteristics. These same individual characteristics are also determinants of individual productivity and individual propensity to supply work. Therefore, the omission of unobservables would make both education and health endogenous. A strategy to deal with unobservables is that of removing them by exploiting the within-individual variation, i.e. by exploiting the variations over-time in labor market outcomes and health of each individual. As unobservables can often be assumed constant, within-individual estimators, such as first-difference, are able to remove them from the model and still to identify the structural parameters. In our context, within-individual estimators would net out both idiosyncratic and common components of ability. This option is ruled out by the cross-sectional nature of our data. Another possible identification strategy is that of Instrumental Variables (IV). Here we make use of them only to solve additional problem of measurement error – as we shall discuss below –, but IV methods are often used to disentangle the impact of health and education from that of unobservables. The EHHUES2002 does not include variables suitable to be used as instruments, and external data, for instance, on reforms are difficult to find. Although the levels of enforcement of any reform in different governorates are likely to differ, increasing the attractiveness of IV, reliable and complete data on the effective enforcement at different times and places are difficult to obtain, if they exist at all.

The strategy we have followed in this paper dates back at least to the 1970ies (see Griliches, 1979 for an early survey) and exploits the fact that siblings and especially twins have similar characteristics. For siblings, part of the genetic endowment is common and, especially if they have lived together within the same family for long time, the family background they have experienced is similar (see Ermish and Francesconi 2001 for a discussion on the extent to which family background can be assumed to be really similar). Of course this argument applies far better to twins, where both genetic endowment and family background essentially coincide. In this case, rather than exploiting within-individual variation, it is possible to exploit within-family or within-siblings variation, removing the common component F_j of unobservables among siblings. Such a strategy, applied to samples of twins, has been followed by Ashenfelter and Krugman (1994), Ashenfelter and Rouse (1998) among many others. Recently Oreopoulos el. al. (2008) adopted it to identify the long term effects of child health on future occupation. However the list of papers resting on within-siblings variation is very long and covers many fields in economics (see e.g. Krashinsky, 2008) measures family effects on voting preferences). However, when applied to siblings rather than twins, one should recognize that the idiosyncratic component which cannot be removed may be relevant. Griliches (1979) and Neumark (1999) caution that neglecting the individual component might induce a distortion in within-siblings estimates even larger than in pooled OLS estimates. To limit this problem we have included among the regressors the order of births of the siblings, which in combination with age should capture differential allocations of resources within the family as well as differential expositions to the family background. In Section 6 we specifically address this concern more in detail, by approximating the ideal case of twins by restricting the sample to the siblings whose age difference is at most two or three years.

We have extracted from the EHHUES2002, which counts about 27,000 observations of people aged 16-64 (working age according to Egyptian regulations), the subsample of all heads' sons and daughters residing in each household from which we have removed the only sons. The resulting sample, labeled "sons", is then composed of at least two siblings per household. Furthermore, we have extracted from the full sample the sample composed of household heads and of their own brothers, exploiting the detailed information on household members' relationships. In Egypt, as in many middle-income countries, it is rather common to find extended families, where several generations live together. We label the resulting sample "head/bros". The union of "sons" and "head/bros" produces the sample "siblings" that we have used throughout the paper. In Table 1 we have reported some key summary statistics referred to "siblings", to its component subsets, to the full sample of the aged 16-64 and to an additional sample composed of the "only sons". The latter is the sample of heads' sons who live with the household heads and have no other brother or sister living with them. This does not mean that they have no brothers at all, but only that their brothers possibly moved out from the original household.

We are interested in investigating whether the sample of siblings can be representative of the general population or whether the choice of siblings introduces serious self-selection bias. While selection on observables is not problematic, as it will be accounted for by the included controls, our concern is that the sample selection depends on unobservable characteristics. For instance, it could be the case that less endowed sons prefer to remain with their parents and brothers, in order to receive support and benefit from their family. In this case unobservable individual endowment will determine both the selection and the performance on the labor market. Fortunately, the within-siblings estimator, by removing a large share of individual unobservables, is quite robust to sample selection, as we shall discuss in detail in appendix B. In particular it is fully robust, if all siblings share the same endowment, i.e. if $I_{ij} = I_j$ for all i in household j. Otherwise, the interpretation of the within-siblings estimates is that of marginal effects in a Heckman-type model.

Looking at Table 1, the differences across samples in most of the variables can be explained in terms of the age-gender composition of the samples, while there are no hints of a heavy selection on unobservables. For instance, it is natural to expect that “sons” and “head/bros” show different characteristics because they differ quite a lot by age, given that individuals in “head/bros” are typically one generation older than in “sons”. The same is true when we compare “siblings” with the “full sample”, because by construction we have removed most heads, all heads’ spouses and other relatives. Thus average age in “siblings” is much lower than in the “full sample”. Compared with the “full sample”, we observe that the percentage of males is much higher in “siblings”, indicating that females are more likely to exit from their original family, to enter in the husband’s family or to establish a new family. Alternatively, males are more likely to marry and stay in their original family or to exit at later ages.

Table 1 Sample description

	Siblings			Full sample	Only sons
	Sons	**head/brothers**	**total**		
observations	5900	1200	7100	27186	1045
age	23.74	32.44	25.20	35.30	25.70
age [min-max]	16-57	16-64	16-64	16-64	16-58
% male	68.3	74.9	69.4	49.5	64.3
years of education	8.31	6.77	8.05	6.58	8.34
% reporting chronic diseases and disabilities	11.1	19.3	12.5	24.8	15.2
% reporting bad general health conditions	6.3	11.9	7.2	12.9	8.8
% married	15.62	45.0	20.5	67.0	22.39
% employed	55.2	70.3	57.7	49.5	56.26
% informal sector	53.0	46.3	51.9	47.2	56.2
hours of work supplied (among employed)	43.42	45.91	43.91	40.95	42.96
hourly wage [pounds] (among employed)	1.45	1.28	1.41	1.79	1.35
obs. per household	2.98	2.85	2.96	3.62	
obs. per household [min-max]	2-8	2-8	2-8	1-19	

A useful benchmark against which to compare “sons” is the sample “only sons”. Their composition in terms of age and gender is similar, so that by comparing them, it is possible to guess if members of “sons” are systematically different in terms of education, health conditions, employment, wage and labor supply. By looking at the summary statistics we observe that the two samples are very similar in terms of education and labor market outcomes, although they differ somewhat regarding health, “only sons” performing rather worse than “sons”, possibly because mean age is about two years higher in “only sons”[5].

The last empirical issue to discuss relates to measurement errors. Already Griliches (1979) warned that measurement errors might be emphasized in within-siblings estimation due to the correlation between siblings’ health conditions. Our indicator of chronic disease and disabilities is self-reported and there might be misreporting by recall errors, little attention in answering the questionnaire or, more importantly, chronic conditions may be systematically under-reported when in fact they are present, especially among less educated people, less aware of their true health and less keen to submit themselves to regular check-ups. The additional “justification bias” (Bound, 1991) which induces people to over-report bad health conditions to justify their absence from the labor market or their lower labor supply has been assumed away in this analysis, as the most recent literature shows mixed results about its existence and relevance (Kapteyn et al., 2009; Jones, 2007; Kreider and Pepper, 2007). The measurement error bias (likely an attenuation bias) can be removed by an IV technique, which assumes that the random component of the measurement error is independent across two alternative self-reported measures (Wooldridge, 2002). We have then instrumented the self-reported indicator of chronic diseases and disabilities with an indicator of poor general health conditions which takes 1 when people report that their general health is worse or much worse than the average health of people of their same age[6].

Results

Estimates are reported in Tables 3-5. Columns 1-3 are discussed in this section and columns 4-7 report robustness checks that we shall address in the following one. For each equation we have reported OLS estimates based on the pooled sample of siblings (augmented by full set of dummies indicating the governorate of residence and an indicator of urban/rural residence), within-siblings (fixed effect FE) estimates and within-siblings-IV (FE-IV), where we take into account the measurement error.

Employment

Chronic diseases and disabilities reduce the probability of being employed by about 7% according to both OLS and FE estimates. It is worth noting that such estimates do not differ much, indicating that individual ability (family effect), which is fully accounted in the FE column, does not influence much the impact of chronic conditions, once the type and the place of residence have been controlled for. However, the estimated impact of education is not significant with OLS whereas, the other variable correlated with

[5] We have also tested whether the sample “siblings” was systematically different from the “full sample”, conditional on age and gender: we have stacked the sample “siblings” with the “full sample” and have estimated a probit model, where the outcome took 1 for individuals in “siblings” and 0 otherwise. A fourth-degree polynomial of age and gender interacted with age was used to properly control for age and gender. The variables education, employment status and our two measures of health conditions resulted significant, but their marginal effect was on the order of 1%.

[6] The measurement error embodied in the indicator of poor health can depend on education, a problem that can be overcome by including education in the main equation.

individual ability, becomes positive and significant according to the FE estimate. Looking at column 3, we observe that attenuation bias induced by measurement errors is particularly strong. Once measurement error is corrected by means of IV, it results that an individual reporting chronic diseases is 25 percentage points less likely to be employed (the average probability is about 50 percent). Applying our estimates to the entire Egyptian population, the employment rate in Egypt is 6 percentage points below its potential (only 50 percent, rather than a potential of 56 percent).

Labor Supply

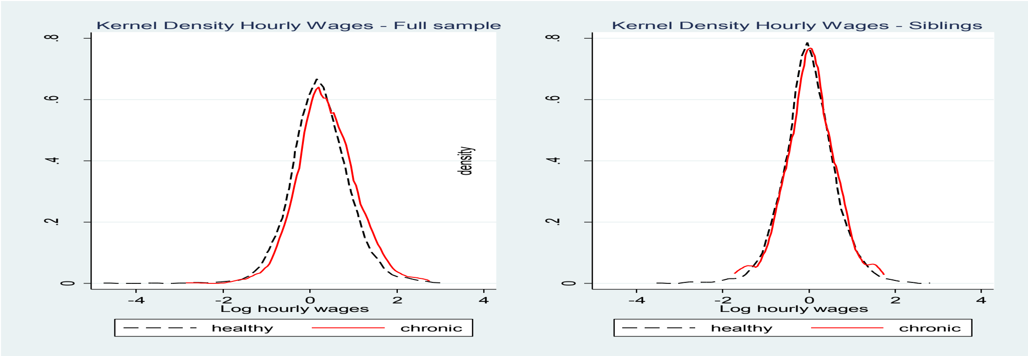
As can be expected, chronic conditions substantially reduce the number of hours supplied weekly by the employees. OLS and FE estimates indicate that the number of hours supplied is reduced by about seven hours because of chronic diseases. In this case OLS and FE estimates differ by about 10% from each other (-7.55 vs. -6.80). Once measurement error is taken into account, chronic conditions cause a much larger contraction in labor supply, up to about 22 hours per week. This reduction accounts for about 50 percent of the average working hours in the "full sample". The aggregate labor supply is about 19 percent below its potential. This gap includes both the loss in employment and in individual labor supply. The implied efficiency loss at the macro-economic level is quite large. To suggest an order of magnitude, suppose that the production function which summarizes Egyptian economy is the widely adopted Cobb-Douglas $Y = AL^{0.6}K^{0.4}$. In this case the resulting output gap is of about 12%, i.e. actual Egyptian GDP is 12% short of its potential level, due to the burden of chronic diseases.

Wage rates

We do not observe an impact of chronic conditions on hourly wage rates. This might be due to rigidities in the labor market, where wages are determined according to occupation and age rather than according to actual individual productivity. Indeed, occupation dummies (non-reported) were significant here, while they were not in the labor supply equation. We also checked that this is not due to sample selection. Figure 2 plots the log hourly wage in the full sample of working age workers and in the sample of siblings, distinguishing between workers with chronic diseases and healthy workers. There is no evidence of sample selection and plots confirm that chronic diseases do not influence wage rates. Note that once we control for individual ability, education does not play any role in explaining individual wages on average. Conditional on occupation and age, a higher level of education does not imply higher wages. Even when occupation dummies are omitted, education remains non significant[7]. One possible explanation is the fact the Egyptian education system is rather ineffective, since "half of those who leave school by the end of the mandatory stage remain illiterate" according to Assaad and Barsoum (2007).

[7] This result is valid a fortiori when we recall that the estimated coefficient of education includes a positive component coming from the mis-reporting in health (see equation 3').

Figure 2 - Log hourly wage distributions



Robustness checks

Columns 4-7 of Tables 3-5 report some of the robustness checks we have performed. All equations have been estimated on three alternative subsets, the subsample of "siblings" aged 25 and more, the sample "sons" and the sample "head/bros". When we focus on those aged 25+ we are looking at about the older 50% of the sample "siblings" (median age is 23 and mean age 25), that part which has necessarily completed his education. In this case we have no concerns about possible problems of over-selection into the sample of people who ended their education just after the primary or secondary attainment level. The direction and the significant level of the results are consistent across the three subsets, although their magnitude changes somewhat, especially regarding the impact of chronic diseases on employment status. The broad picture that emerges is that the effect of chronic diseases is certainly negative but does get larger as people grow older.

The main empirical issue is to what extent our within-siblings estimators control for the idiosyncratic component of individual endowment. We have addressed this point by following two strategies.

First, we have approximated the ideal situation of twins by restricting the sample to siblings of similar age. For each eligible household we have selected all siblings whose age difference was less maximum(????) 3 years. The resulting sample is composed of 3713 individuals, and its mean age is 23.8. Using this sample we have estimated the employment equation and the resulting coefficient for *chronic* is -0.281 [95% C.I -0.508, -0.055]. By restricting the age difference to just 2 years at most, the resulting sample further reduces to 2770 observations (mean age 23.5) and the coefficient of *chronic* is -0.206 [95% C.I. -0.459, 0.048]. In both cases the point estimate is not significantly different from our estimate reported in Table 3 (-0.2568).

As a second strategy we have added proxies to further capture individual specific unobservables. Specifically we included indicators of whether the respondent smokes and of whether he stopped smoking in the past. Smoking is sometimes used as a proxy for (im)patience (Fersterer and Winter-Ebmer, 2003) and the decision to quit is seen as as a sign of willpower (Kan, 2007), two characteristics that are correlated with both health conditions and labor market outcomes. Their inclusion has not altered significantly neither the impact of *chronic* nor that of education, supporting our assumption that only a small share of individual specific ability remains outside the family component that is captured by the within-siblings estimators and the proxy *order* (order of birth).

Following the same idea of focusing on siblings of similar age, we have estimated the employment equation by age intervals. This procedure implies a drastic reduction in the number of observations, so that the choice of age intervals had to accommodate the need of a reasonable sample size. Indeed, the model has been estimated in each of the following age intervals: 16-20, 21-25, 26-30, 31-40, 41-64. Results are reported in Table 2 (line *). Unsurprisingly, the precision of the estimates is much smaller and their magnitude is increasing with age, a pattern already observed above. The benchmark against which to compare these estimates are those of equation (1), augmented by the interaction *chronic X age* (which provides age specific effects of chronic diseases)[8], obtained by using the whole “siblings” sample. In the lower part of Table 2 we have reported such age specific effect (line **), calculated at the mean age of the corresponding interval, and its standard error. Moreover we add the p-value of a test where we check if age-specific estimates (**) differ from estimates (*). The two are not significantly different at 95% level of confidence. Actually, except for the age interval 21-25 the null of equality cannot be rejected at much lower levels of confidence.

Finally we have checked whether there are long-term effects of the business cycle that an individual finds at 16, the minimum legal age for being employed, and at the time when he actually enters the labor market, after the end of his education period. For this purpose we have included the per capita GDP growth rate in these years (sources: WDI 2007 and Maddison). They always resulted in non-significant and comparatively uncorrelated with all the other variables included into the model.

Table 2 – Employment equation. Estimates by age intervals

		(1)	(2)	(3)	(4)	(5)
	VARIABLES	16-20	21-25	26-30	31-40	41-64
*	Chronic	-0.0376	-0.0577	-0.2121	-0.4761	-0.4968
	s.d.	(0.1729)	(0.1916)	(0.3022)	(0.3174)	(0.3291)
	Observations	1314	1204	506	407	200
	Number of generations	609	569	243	190	96

[8] This model extension is described in detail in the next section.

	mean age	18	23	28	35	48
**	chronic + chronic X mean age	-0.0741	-0.1877	-0.3012	-0.4629	-0.7556
	s.d.	(0.0924)	(0.0704)	(0.0711)	(0.1067)	(0.2057)
	p-value	0.6927	0.0648	0.2097	0.8822	0.2083

Extensions

Having observed in the previous section that the impact of chronic conditions might not be constant across age, we explore in this section whether and if so how it varies also across levels of education, family income, gender and sector of the economy (formal or informal).

In Tables 6 to 8 we have reported the estimates of each equation augmented by an interaction term between chronic conditions and respectively age (column 1), education (column 2), household income (column 3), gender (column 4) and an indicator of informal sector (column 5).

By looking at columns (1), confirming the findings discussed above, chronic diseases have a stronger effect among older people. Their negative impact on the probability of being employed is about 17 percent at age 22 (25th percentile of age distribution in the "full sample") and about 73.6 percent at age 47 (75th percentile of age distribution). The impact on labor supply is about 21 hours lost at age 22 and about 24 hours at age 47 (both statistically different from zero at 95%).

As estimates of Table 6 column (2) indicate, the negative effect of *chronic* on the probability of being employed is lower among the more educated. Particularly the loss in the probability of being employed is only 10 percent for individuals with a university degree (i.e. at least 16 years of schooling) while it is 25 percent in the total population and 43 percent among people with no schooling. By looking at Table 7 the drop in working hours amounts to 30 hours per week among people without education (significantly different from zero at 95%) and only 15 hours (significant at 90%) among those with a university degree..

Regarding the variation of the effect of *chronic* across levels of household (monthly) income – columns (3) – we observe that the probability of being employed is reduced by 28 percent among households earning 300 pounds (25th percentile of income distribution) and by 25 percent among household earning 700 pounds (75th percentile of income distribution) (Table 6). Labor supply falls more among poorer households (30 hours/week with a monthly income of 300 pounds and 21 hours/week with an income of 700 pounds, both significant at 99%).(Table 7)

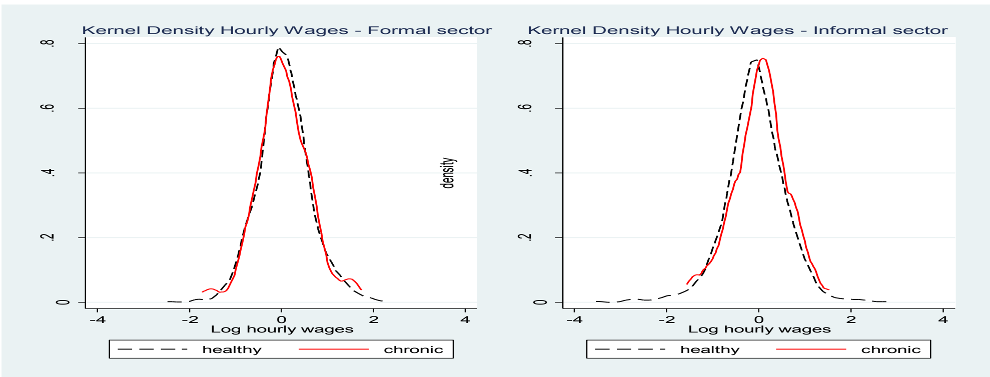
Next, the probability of being employed decreases more for males than for females and the difference across genders is significant (Table 6 - column (4)). Surprisingly, the impact of *chronic* on the probability of being employed is not significantly different from zero for females. Also, the impact on labor supply does not significantly differ across genders (Table7).

Finally we analyzed how the effect of chronic diseases changes among people employed in the informal sector, where workers do not benefit from any employment protection, disability benefits or health insurance. In our data there is no explicit information on which sector a worker belongs to, though extensive information on health insurance coverage is available. In Egypt only employees of the formal economy are entitled to a health insurance which typically does not extend to other members of the family. Indeed our indicator of informal sector is based on whether people are covered by a health insurance. We attribute individuals to the informal sector when they report that in their household no one is covered by a health insurance and those who report of having a health insurance provided by the school, which covers only the children[9]. According to this classification about 51 percent of the sample belongs to the informal sector, a figure not dissimilar from the 49 percent in 1998 which is estimated using the Egypt Labor Market Survey 1998.

In column (5) of Tables 6-8 we have reported estimates of the specification including the interaction between chronic diseases and the informal sector indicator. Not surprisingly, chronic conditions reduce employment probability much more among workers employed in the informal sector (37 percent compared to a statistically insignificant 13 percent of those belonging to the formal sector) (Table 6). Similarly, labor supply falls by 36 hours per week in the informal sector compared to less than 13 in the formal (Table 7).

From Table 8 we notice that the absence of relation between chronic diseases and wages persists also within each sector. This result is further confirmed by Figure 3 which plots log hourly wage distribution by sector for the sample of siblings. There is no evidence that chronic conditions influence wages differently in the informal sector.

Figure 3 Wages in formal and informal sectors



[9] Information about health insurance is collected at the household level and it does not vary across household members.

Conclusions and policy implications.

Egypt experienced a dramatic increase in life expectancy and a more than threefold increase in GDP per capita in the last fifty years. The usual transition of morbidity from communicable to non-communicable, chronic diseases occurred during the past decades. Chronic diseases have by now clearly become the largest threat to people's health. Besides the obvious negative effect on the quality of people's life, chronic diseases have negative and sizeable economic consequences, both in terms of efficiency and equity.

In this paper we have assessed the impact of chronic diseases on labor-market-related decisions, i.e. employment status and labor supply, and their impact on wage rates. We have found that the probability of being employed is 25 percentage points lower among people reporting chronic disease conditions (the average probability is about 50 percent) and the amount of working time supplied is reduced by 22 hours per week (out of about 40). The impact of chronic diseases is larger among the more elderly, the less educated and the workers of the informal sector. Instead, among people with a university degree, the probability of being employed is reduced in case of chronic conditions only by 10 percent.

Hence, chronic conditions on the one hand, potentially, cause a large efficiency loss at the macroeconomic level and on the other hand could worsen economic inequality, as they disproportionally hit vulnerable people with low socio-economic status.

It is unclear at this stage whether the countervailing role played by education can justify the enlargement and improvement of education as a health policy. Of course education is important *per se* and in Egypt there is much room for improving it (Assaad and Barsoum, 2007). What is less clear is if such policy would be effective in attenuating the negative effects of chronic diseases. The reason is that we are not able to disentangle between two alternative explanations for the attenuation role played by education that we have found. On the one hand more educated people are more likely to be employed in non-manual and less physically demanding occupations, where chronic conditions are less likely to dictate their exit from the labor market. Therefore in this case education would be associated with a small impact of chronic diseases, simply because education gives access to white-collar positions: if so, it is the kind of occupation and not education which is the cause of the reduced impact of chronic diseases among the more educated. On the other hand, education might enable people to better manage their chronic conditions and delay significantly the moment when they must leave their job (Cutler and Lleras Muney, 2006). Only if the latter explanation were dominant, would education be an effective instrument to improve health outcomes.

The strategy followed to identify the causal impact of chronic diseases has taken advantage of the similarity in genetic endowment, preferences and family background among siblings. Several robustness checks have been performed to test its appropriateness, which allow us to consider our results reliable. Incidentally, it is surprising to notice that there are relatively few papers in health economics that have followed this approach, although it looks especially promising when important unobservables are genetic or related to the family background.

Given the nature of chronic diseases, their naturally increasing trend due to the success in extending life expectancy, and in light of the economic burden associated with those diseases, a sound health and social policy to limit the main risk factors of chronic diseases is critical in Egypt. While an important driver of chronic disease prevalence is population aging, evidence also indicates that a large share of the chronic disease burden are preventable or can at least be postponed by changes in health behavior (Ezzati et al 2003). Effecting such changes in chronic disease-relevant areas like diet, physical activity, alcohol and smoking is no small challenge, but there is a significant evidence base on effective and cost-effective interventions that could and should be considered in Egypt (Institute of Medicine 2010). Such action appears to be particularly urgent in the case of Egypt (and possibly a number of other MENA countries) which is facing, for instance, an enormous (and widely under-recognized) challenge of obesity, well beyond what other most countries have encountered so far, both in the developing and developed countries (Fumagalli, Suhrcke & Rocco 2010).

It would be beyond the scope of this article to discuss the interventions to manage and prevent or postpone chronic diseases in Egypt. Suffice to say that an adequate policy response will be located within *and* outside the health care system proper. Extending health insurance to the vulnerable population at large, including self employed, unpaid family workers, and other informal sector workers, is only the first step towards at least improved access to relevant health care services, especially chronic disease management. Measures outside the health care system include tobacco control policies (especially tobacco taxation), control of environmental pollution via regulation and legislation. Education policies are a further potentially important avenue for reducing the burden of chronic disease in Egypt.

Appendix A - Linearization

In this appendix we provide a description of the linearization involved in the labor supply equation, which is useful to understand the meaning of the coefficients that we have estimate.

Denote with $\Psi(.)$ the function $\psi_L'^{-1}(.)$. By first order Taylor expansion around $(w(\bar{X}_{ij}^{w}), \bar{X}^{\psi})$, equation (2) becomes

$$L_{ij} = \left[\Psi\left(\bar{w}, \bar{X}^{\psi}\right) - \Psi_{\mathrm{w}}'\left(\bar{w}, \bar{X}^{\psi}\right)\bar{w} - \Psi_{\mathrm{X}}'\left(\bar{w}, \bar{X}^{\psi}\right)\bar{X}^{\psi}\right] + \Psi_{\mathrm{w}}'\left(\bar{w}, \bar{X}^{\psi}\right)w_{ij}\left(X_{ij}^{w}\right) + \Psi_{\mathrm{X}}'\left(\bar{w}, \bar{X}^{\psi}\right)X_{ij}^{\psi} + \upsilon_{ij}$$

where $\bar{w} = w(\bar{X}_{ij}^{w})$, which amount to

$$L_{ij} = \lambda_0 + \lambda_1 w_{ij}\left(X_{ij}^{w}\right) + \lambda_2 X_{ij}^{\psi} + \upsilon_{ij}$$

Now by a further linearization, we have $w_{ij}\left(X_{ij}^{w}\right) = \exp\left(X_{ij}^{w}\alpha + \tau_{ij}\right) \approx X_{ij}^{w}\theta + \vartheta_{ij}$, that can be substituted above to obtain equation (2'). Note that the coefficients have a "technological" nature, as they are the first derivatives of the function $\psi_L'^{-1}(w_{ij}\left(X_{ij}^{w}\right), X_{ij}^{\psi})$, i.e. a transformation of the labour disutility. Precisely λ_1 reflects the reciprocal of ψ_L''. The more convex ψ, the smaller will be our coefficients. The economic meaning is that a larger wage will induce a small increase in labor supply if the marginal disutility of labor increases much in response to a marginal increase in labor supply.

Appendix B - Sample selection.

Observation1.

The sample "siblings" has been drawn in a non random way from the "full sample". It is possibly self-selected on unobservables, as the decision to remain into the household depends on individual and household preferences. The same preferences might enter in the labor equations inducing a sample selection bias in the estimates. Formally, let

$$Y_{ij}^{0} = X_{ij}^{s}\gamma_1 + W_j^{s}\gamma_2 + \eta_{ij}^{s}$$

determine the selection equation, where an individual selects into the sample if $Y_{ij}^{0} \geq 0$, and

$$Y_{ij} = X_{ij}\beta + \eta_{ij}$$

is the equation of interest. The fact that unobserved preferences $I_{ij} + F_j$ enter in both η_{ij}^{s} and η_{ij} makes them correlated. In the selected sample the conditional mean would be

$$E(Y_{ij}|Y_{ij}^{0} \geq 0) = X_{ij}\beta + \sigma_{12}\lambda(X_{ij}^{s}\gamma_1 + W_j^{s}\gamma_2)$$

where $\lambda(X_{ij}^{s}\gamma_1 + W_j^{s}\gamma_2)$ is the inverse mills ratio. Now the inverse mills ratio is very much linear (Cameron and Trivedi, 2005, p.540), and little harm is produced by linearizing it as $\lambda\left(X_{ij}^{s}\gamma_1 + W_j^{s}\gamma_2\right) = X_{ij}^{s}\gamma_1\Lambda + W_j^{s}\gamma_2\Lambda$.

Suppose for simplicity that X_{ij}^{s} is a subset of X_{ij}, so that the additional instrument to be included in the selection equation for identification purposes varies at household level. We can then rewrite

$$E(Y_{ij}\middle|Y_{ij}^{0} \geq 0) = X_{ij}^{s}(\beta + \sigma_{12}\gamma_1\Lambda) + X_{ij}^{-s}\beta + W_j^{s}\gamma_2\Lambda\sigma_{12} + I_{ij} + F_j$$

where X_{ij}^{-s} is the complement of X_{ij}^{s} in X_{ij}. Note that by means of the within-siblings transformation, the component $W_j^{s}\gamma_2\Lambda + F_j$ is netted out and we estimate $(\beta + \sigma_{12}\gamma_1\Lambda)$. This quantity is the marginal effect of

X_{ij}^{s} on the selected sample, i.e. the sum of the impact of X_{ij}^{s} on the probability of being selected and the impact of X_{ij}^{s} on Y_{ij} among the selected (see Greene, 2003, p.783).

Observation 2.

Consider again the model described above and write it as

$$Y_{ij}^{0} = X_{ij}^{s}\gamma_1 + W_j^{s}\gamma_2 + A_{ij} + \theta_{ij}^{s}$$

and

$$Y_{ij} = X_{ij}\beta + A_{ij} + \theta_{ij}$$

recalling that we have defined $A_{ij} = I_{ij} + F_j$, $\eta_{ij}^{s} = A_{ij} + \theta_{ij}^{s}$ and $\eta_{ij} = A_{ij} + \theta_{ij}$. Suppose that $cov\left(\theta_{ij}^{s}, \theta_{ij}\right) = 0$. By applying the within-siblings transformation we have

$$Y_{ij} - \bar{Y}_{ij} = (X_{ij} - \bar{X}_{ij})\beta + (A_{ij} - \bar{A}_{ij}) + (\theta_{ij} - \bar{\theta}_{ij})$$

Now suppose that A_{ij} is fully common across siblings, so that $(A_{ij} - \bar{A}_{ij}) = 0$. Then it follows that $cov\left(A_{ij} + \theta_{ij}^{s},\ \theta_{ij} - \bar{\theta}_{ij}\right) = 0$ and there is no sample selection bias.

References

Ashenfelter, O. and Krueger, A. (1994) "Estimates of the Economic Return to Schooling from a New Sample of Twins", The American Economic Review, 84(5) :1157-1173

Ashenfelter, O. and Rouse, C. (1998) "Income, Schooling, and Ability: Evidence from A New Sample of Identical Twins", Quarterly Journal of Economics, 113(1): 253-284

Assaad, R. and G. Barsoum (2007) "Youth Exclusion in Egypt: In Search of Second Chances", The Middle East Youth Initiative Working Paper, n.2/2007

Bago d'Uva, T., van Doorslaer, E., Lindeboom, M., O'Donnell, O. and Chatterji, S. (2008) "Does reporting heterogeneity bias the measurement of socio-economic inequalities in health?", Health Economics, 17(3) :351 - 375

Bound, J. (1991), "Self-reported versus objective measures of health in retirement models," *Journal of Human Resources,* 26(1), 106-138.

Butcher, L.M., Kennedy, J.K., Plomin, R. (2006) "Generalist genes and cognitive neuroscience", Current Opinion in Neurobiology, 16:145–151

Cameron, A.C., and Trivedi, P.K. (2005) “Microeconometrics Methods and Applications” Cambridge University Press

Currie, J. and Madrian B.C. (1999) “Health, health insurance and the labor market”, in Handbook of Labor Economics, Volume 3, Edited by O. Ashenfelter and D. Card, cap.50, pp.3309-3415

Cutler, D.M. and Lleras-Muney, A. (2006) “Education and health: evaluating theories and evidence”, NBER working paper n.12352

Ermish, J.F. and Francesconi, M. (2001) “Family structure and children’s achievements”, Journal of Population Economics, 14:249-270

Ezzati, M. et al (2003). Estimates of global and regional potential health gains from reducing multiple major risk factors. The Lancet. 26;362(9380):271-80.

Fersterer, J. and Winter-Ebmer, R. (2003) “Smoking, discount rates, and returns to education” Economics of Education Review, 22:561–566

Fumagalli, E; Suhrcke, M. and Rocco, L. (2010). Obesity in Egypt: trends, socioeconomic inequalities and determinants. Mimeo, University of East Anglia: Norwich.

Gannon, B. (2005) “A dynamic analysis of disability and labor force participation in Ireland.” Health Economics. 14: 925–938.

Greene, W.H. (2002) “Econometric Analysis - 5th edition”, Prentice Hall

Griliches, Z. (1979) “Sibling Models and Data in Economics: Beginnings of a Survey”, Journal of Political Economy, 87(5):S37-S64

Institute of Medicine (2010). Promoting Cardiovascular Health in the Developing World: A Critical Challenge to Achieve Global Health. Committee on Preventing the Global Epidemic of Cardiovascular Disease: Meeting the Challenges in Developing Countries; Institute of Medicine.

Jones, M.K. (2007), “How accurate are estimates of the impact of disability on labor market participation? Evidence from the Health Survey for England,” Swansea University, School of Business, Working Paper.

Kan, K. (2007) “Cigarette smoking and self-control”, Journal of Health Economics 26:61–81

Kapteyn A., Smith J.P., van Soest A. (2009), Work Disability, Work, and Justification Bias in Europe and the U.S., NBER Working paper 15245

Krashinsky , H. (2007) “The impact of social mobility and within-family learning on voter preferences: Evidence from a sample of twins”, Journal of Public Economics 91:97–116

Kreider, B. and J.V. Pepper (2007), “Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors,” *Journal of the American Statistical Association*, 102(478), 432-441.

Loprest P, Rupp K, Sandell SH (1995) “Gender, Disabilities, and Employment in the Health and Retirement Study”. Journal of Human Resources, 30, supplement: S293-S318.

Neumark, D. (1999) “Biases in twin estimates of the return to schooling”, Economics of Education Review, 18:143–148

Oreopoulos, P., Stabile M., Walld, R. and Roos, LL, (2008) “Short-, Medium-, and Long-Term Consequences of Poor Infant Health An Analysis Using Siblings and Twins”, The Journal of Human Resources, XLIII(1):88-138

Rocco, L. and Suhrcke, M. (2008) “Microeconomic consequences of non-communicable diseases at the individual or household level, with a focus on labor market outcomes”. Presented at the WHO Consultation “Methods for identifying the economic consequences of diseases and injuries”, 12 - 13 June, Venice

Suhrcke, M., Rocco, L., McKee, M. (2007) “Health: a vital investment for economic development in eastern Europe and central Asia”. European Observatory on Health Care Systems, (London) 2007

WHO (2005) “Preventing CHRONIC DISEASES a vital investment” WHO Press, Geneva

Wilson, S.E. (2001) “Work and the accommodation of chronic illness: A re-examination of the health-labor supply relationship”, Applied Economics, 33: 1139-1156

Wooldridge, J.M. (2002) “Econometric analysis of cross section and panel data”, The MIT Press

Table 3 Employment

VARIABLES	(1) OLS	(2) FE	(3) FE-IV	(4) FE-IV	(5) FE-IV	(6) FE-IV	(7) FE-IV
				age>=25	sons	head/broth	add. contr.
chronic	-0.0668***	-0.0698***	-0.2599***	-0.3759***	-0.2030***	-0.5108***	-0.2616***
	(0.015)	(0.020)	(0.066)	(0.099)	(0.073)	(0.167)	(0.066)
male	0.4703***	0.4646***	0.4598***	0.5413***	0.4439***	0.5795***	0.4493***
	(0.013)	(0.015)	(0.015)	(0.031)	(0.016)	(0.037)	(0.016)
age	0.0432***	0.0446***	0.0449***	0.0433***	0.0483***	0.0445***	0.0442***
	(0.004)	(0.006)	(0.006)	(0.013)	(0.009)	(0.011)	(0.006)
age2	-0.0576***	-0.0580***	-0.0566***	-0.0506***	-0.0615***	-0.0544***	-0.0558***
	(0.006)	(0.008)	(0.009)	(0.016)	(0.013)	(0.014)	(0.009)
grade	-0.0010	0.0050***	0.0045***	0.0063***	0.0035*	0.0078**	0.0046***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
married	0.1691***	0.1346***	0.1335***	0.1072***	0.1305***	0.0701*	0.1307***
	(0.013)	(0.018)	(0.018)	(0.028)	(0.021)	(0.037)	(0.018)
order	0.0110**	0.0156	0.0164*	-0.0000	0.0134	0.0246	0.0156
	(0.005)	(0.010)	(0.010)	(0.018)	(0.011)	(0.022)	(0.010)
smoker							0.0239
							(0.016)
stopsmoke							0.0460
							(0.036)
Constant	-0.4875***	-0.5563***					
	(0.059)	(0.088)					
Observations	7100	7100	7100	2307	5905	1195	7100
R-squared	0.335	0.343	0.328	0.363	0.308	0.406	0.328
Number of generations		2680	2680	987	2216	464	2680
F test			206.8	65.36	172.3	34.41	208.3

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4 Labor Supply

VARIABLES	(1) OLS	(2) FE	(3) FE-IV	(4) FE-IV	(5) FE-IV	(6) FE-IV	(7) FE-IV
				age>=25	sons	head/broth	add. contr.
chronic	-7.5536***	-6.8425***	-23.5298***	-24.8231***	-21.6075***	-28.9538**	-23.5076***
	(1.194)	(1.569)	(6.784)	(8.058)	(7.886)	(13.332)	(6.794)
male	3.7486***	4.3957**	4.1258**	4.2628	4.4442*	4.0570	5.1094**
	(1.371)	(1.987)	(2.095)	(3.165)	(2.346)	(4.631)	(2.098)
age	0.7993**	0.9878*	1.1125*	1.9133*	0.5670	1.9284*	1.1449*
	(0.337)	(0.578)	(0.594)	(1.146)	(0.786)	(0.991)	(0.595)
age2	-1.1273**	-1.1625	-1.0701	-1.8484	-0.2051	-2.0789*	-1.0882
	(0.511)	(0.750)	(0.785)	(1.392)	(1.093)	(1.245)	(0.784)
grade	0.0283	-0.0732	-0.1004	-0.3700*	-0.0875	0.0643	-0.1166
	(0.094)	(0.140)	(0.143)	(0.192)	(0.163)	(0.328)	(0.142)
married	4.4346***	3.7367**	3.7626**	4.0741**	4.8599***	-0.5130	3.8009**
	(1.048)	(1.505)	(1.515)	(2.060)	(1.732)	(3.126)	(1.526)
order	-1.0360***	-1.8184**	-2.0507**	-3.3387**	-1.9351*	-1.9967	-1.9547**
	(0.387)	(0.903)	(0.927)	(1.367)	(1.085)	(2.060)	(0.923)
smoker							-2.7863**
							(1.268)
stopsmoke							4.1514
							(2.726)
occupation dummies	YES	YES	YES	YES	YES	YES	YES
Constant	43.7528***	34.5412**					
	(6.464)	(13.555)					
Observations	4528	4528	3754	1549	3007	747	3754
R-squared	0.096	0.058	0.007	0.041	0.013	0.032	0.012
Number of generations		2329	1555	685	1244	311	1555
F test			88.94	42.26	64.77	23.20	89.77

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5 – Wage Equation

VARIABLES	(1) OLS	(2) FE	(3) FE-IV	(4) FE-IV	(5) FE-IV	(6) FE-IV	(7) FE-IV
				age>=25	sons	head/broth	add. contr.
chronic	0.0182	0.0591	-0.1501	-0.4794	-0.1411	-0.1181	-0.1607
	(0.033)	(0.042)	(0.173)	(0.322)	(0.197)	(0.353)	(0.175)
male	0.2774***	0.2646***	0.2550***	0.2976***	0.2506***	0.2444*	0.2411***
	(0.042)	(0.056)	(0.057)	(0.102)	(0.064)	(0.136)	(0.058)
age	0.0525***	0.0504**	0.0566***	0.1107***	0.0584**	0.0656*	0.0561***
	(0.011)	(0.020)	(0.020)	(0.038)	(0.025)	(0.035)	(0.020)
age2	-0.0581***	-0.0735**	-0.0775***	-0.1373***	-0.0805**	-0.0878*	-0.0766***
	(0.018)	(0.029)	(0.028)	(0.046)	(0.036)	(0.045)	(0.028)
grade	0.0177***	0.0053	0.0055	0.0026	0.0042	0.0074	0.0055
	(0.003)	(0.004)	(0.004)	(0.007)	(0.005)	(0.010)	(0.004)
order	0.0015	0.0533*	0.0444	-0.0242	0.0459	0.0221	0.0426
	(0.012)	(0.030)	(0.030)	(0.049)	(0.034)	(0.069)	(0.030)
smoker							0.0273
							(0.038)
stopsmoke							0.0620
							(0.089)
Occupation dummies	YES	YES	YES	YES	YES	YES	YES
Constant	-0.6095***	-0.6351					
	(0.209)	(0.394)					
Observations	2603	2603	1738	725	1391	347	1738
R-squared	0.198	0.118	0.100	-0.046	0.106	0.122	0.099
Number of generations		1629	764	330	614	150	764
F test			40.54	17.58	35.24	6.107	41.13

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in parentheses

Table 6 Employment (extensions)

VARIABLES	(1) int. var. = age	(2) int. var. = grade	(3) int. var. = income	(4) int. var. = gender	(5) int. var. = informal
chronic	-0.4312***	0.3133	-0.3074***	-0.1087	-0.1351
	(0.104)	(0.222)	(0.118)	(0.076)	(0.092)
chronic X grade	0.0233**				
	(0.010)				
chronic X age		-0.0220***			
		(0.008)			
chronic X income			0.0001		
			(0.000)		
chronic X male				-0.2459**	
				(0.102)	
chronic X informal					-0.2402*
					(0.129)
chronic + chronic X int. var.				-0.355***	-0.375***
				0.0861	0.0931
male	0.4598***	0.4573***	0.4597***	0.4899***	0.4582***
	(0.015)	(0.015)	(0.015)	(0.020)	(0.015)
age	0.0453***	0.0440***	0.0449***	0.0450***	0.0438***
	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)
age2	-0.0567***	-0.0449***	-0.0568***	-0.0565***	-0.0560***
	(0.009)	(0.012)	(0.009)	(0.009)	(0.009)
grade	0.0010	0.0044***	0.0045***	0.0047***	0.0046***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
married	0.1240***	0.1230***	0.1326***	0.1412***	0.1333***
	(0.019)	(0.019)	(0.018)	(0.018)	(0.018)
order	0.0179*	0.0097	0.0169*	0.0144	0.0188*
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Observations	7100	7100	7100	7100	7100
R-squared	0.327	0.317	0.328	0.325	0.324
Number of generation	2680	2680	2680	2680	2680
F bad health	110.4	92.76	104.0	104.7	103.5
F bad health X int. var.	88.43	104.0	87.26	81.95	55.63

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors in parentheses

Table 7 Labour Supply (extensions)

VARIABLES	(1) int. var. = age	(2) int. var. = grade	(3) int. var. = income	(4) int. var. = gender	(5) int. var. = informal
chronic	-34.9014**	-20.9279	-39.3331***	-26.4338*	-12.8083
	(16.137)	(20.859)	(12.857)	(13.706)	(8.332)
chronic X grade	1.2950				
	(1.401)				
chronic X age		-0.0894			
		(0.611)			
chronic X income			0.0237		
			(0.016)		
chronic X male				3.4387	
				(14.172)	
chronic X informal					-23.4749*
					(14.059)
chronic + chronic X int. var.				-23.00***	-36.28***
				7.182	11.60
male	4.1402**	4.1383**	4.1318*	3.5931	4.1087*
	(2.090)	(2.096)	(2.122)	(3.060)	(2.138)
age	1.1890*	1.1024*	1.2323**	1.1053*	0.9572
	(0.609)	(0.599)	(0.613)	(0.596)	(0.613)
age2	-1.1209	-1.0163	-1.2785	-1.0630	-0.9124
	(0.800)	(0.850)	(0.820)	(0.787)	(0.809)
grade	-0.2730	-0.1005	-0.0861	-0.0995	-0.1070
	(0.226)	(0.143)	(0.145)	(0.144)	(0.148)
married	3.3051**	3.7512**	3.2868**	3.7466**	3.4194**
	(1.586)	(1.522)	(1.555)	(1.516)	(1.549)
order	-1.9776**	-2.0801**	-1.9579**	-2.0426**	-1.6670*
	(0.922)	(0.943)	(0.939)	(0.930)	(0.963)
Occupation dummies	YES	YES	YES	YES	YES
Observations	3754	3754	3754	3754	3754
R-squared	-0.004	0.009	-0.014	0.005	-0.021
Number of generations	1555	1555	1555	1555	1555
F bad health	52.20	45.50	36.90	44.57	44.59
F bad health X int. var.	50.03	43.56	44.56	43.06	22.07

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8 Wage equation (extensions)

VARIABLES	(1) int. var. = age	(2) int. var. = grade	(3) int. var. = income	(4) int. var. = gender	(5) int. var. = informal
chronic	-0.3048	-0.4900	-0.1093	-0.0579	-0.2240
	(0.408)	(0.520)	(0.288)	(0.248)	(0.244)
chronic X grade	0.0168				
	(0.033)				
chronic X age		0.0124			
		(0.018)			
chronic X income			-0.0001		
			(0.000)		
chronic X male				-0.1289	
				(0.322)	
chronic X informal					0.1793
					(0.362)
chronic + chronic X int. var.				-0.187	-0.0447
				0.216	0.256
male	0.2514***	0.2511***	0.2550***	0.2727***	0.2568***
	(0.058)	(0.058)	(0.057)	(0.072)	(0.057)
age	0.0591***	0.0568***	0.0565***	0.0578***	0.0570***
	(0.021)	(0.020)	(0.020)	(0.020)	(0.020)
age2	-0.0819***	-0.0831***	-0.0773***	-0.0785***	-0.0774***
	(0.029)	(0.031)	(0.028)	(0.028)	(0.028)
grade	0.0034	0.0053	0.0054	0.0057	0.0051
	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
order	0.0467	0.0492	0.0440	0.0430	0.0418
	(0.030)	(0.031)	(0.030)	(0.030)	(0.031)
Occupation dummies	YES	YES	YES	YES	YES
Observations	1738	1738	1738	1738	1738
R-squared	0.095	0.096	0.101	0.096	0.100
Number of generation	764	764	764	764	764
F bad health	22.61	18.28	18.47	12.54	20.42
F bad health X int. var.	24.63	20.22	20.68	26.28	9.302

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1