HEDG
Health, Econometrics and Data Group
The University of York

WP 11/33

Health in the cities: when the neighborhood matters more than income

Marcel Bilger

Vincenzo Carrieri

December 2011

york.ac.uk/res/herc/hedgwp

Health in the cities: when the neighborhood matters more than income*

Marcel Bilger

The Harris School of Public Policy Studies, University of Chicago

Vincenzo Carrieri†

Department of Economics and Statistics and CELPE, University of Salerno

ABSTRACT

Using a rich Italian cross-sectional data set, we estimate the effect of a neighborhood quality index based on pollution, crime, and noise on self-assessed health, presence of chronic conditions and limitations in daily activities. We address the self-selection of the residents in their neighborhoods, as well as the possible endogeneity of income with respect to health, through instrumental variable methods and several endogeneity tests. The main novelty is the sound estimation of the neighborhood effect on health using observational data, which has the advantage of providing general results that are not dependent on any experimental design. This allows us to fully compare the neighborhood effect with the traditional socioeconomic determinants of health. Our main findings are that low quality neighborhoods are strongly health damaging. This effect is comparable to the primary/upper secondary education health differential and is even higher than the impact that poor economic circumstances have on health.

JEL classification: I10; R23; C31.
Key-words: Neighborhood quality; Health; Endogeneity; Instrumental Variables

*We are very grateful to Rachel Griffith and participants at the Royal Economic Society Conference (London 2011) for useful comments and suggestions. We also thank participants at the Italian Health Economists Annual Meeting (Bergamo 2009), at the Italian Economists Annual Conference (Rome 2009) and Elena Granaglia, Giuseppe Costa and Luca Salvatici for their comments and suggestions on an earlier version of this paper. The usual disclaimer applies.

† Corresponding author: Vincenzo Carrieri, Department of Economics and Statistics and CELPE, Via Ponte Don Melillo, 84084, University of Salerno, Italy. E-mail: vcarrieri@unisa.it

1. INTRODUCTION

An important part of the research on urban-related issues is dedicated to the effect that the characteristics of the neighborhoods in which people live have on a wide range of social and economic outcomes, such as wages, educational attainments, and criminal behavior. This research is carried out with many difficulties and even some skepticism. Indeed, especially in the field of economics, there is a strong debate about the possibility of properly identifying neighborhood effects given the self-selection of the residents in their neighborhood. This point is crucial, as this self-selection can bias the estimation of any model where neighborhood characteristics are used as determinants of any given outcome (Manski, 1993).

The difficulty to identify a clear causal effect has deterred economists from exploring the consequences of the neighborhood, especially when the outcome of interest is thought to be influenced by unobservable factors that are also correlated with neighborhood choice. One of these outcomes is, undoubtedly, health. Interestingly, although social scientists in other disciplines mostly make use of observational data and do so without dealing with the identification problem, the economic literature on the neighborhood effect on health heavily relies on one single experiment: the Moving To Opportunity (MTO) program. The objective of this experiment was to determine whether families who moved from inner-city, high-poverty areas to low-poverty areas could attain better outcomes, including health-related ones. All of the studies based on this experiment have found significant health benefits for moving to better areas. However, the evidence provided is far from being general, as this experiment only took place in five US cities (Baltimore, Boston, Chicago, Los Angeles, and New York) and, more importantly, was limited to low income families.

The lack of evidence on the effect of the place of residence on health is especially penalizing in urban economics because geographical health inequalities are a dramatic urban issue. Geographical variations in health are most often present at aggregate levels (e.g., at the county level in the US, Murray et al., 2006, and at the regional level in Italy, Costa et al., 2003), but health differentials become truly outstanding between bad- and good-quality neighborhoods within cities. A thought-provoking example is given by Hanlon et al. (2006), who calculated that the discrepancy in life expectancy at birth between the richest (Calton) and poorest (Lenzie) districts of Glasgow amounts to as much as 30 years. Remarkably, this health differential is comparable in magnitude with the average increase in life expectancy that took place in the most developed countries throughout the entire 20th century (Cutler, Deaton and Lleras-Muney, 2006).

In this paper, we aim at filling this literature gap by making use of observational data from a national representative survey while simultaneously dealing with the identification issue. The purpose is to produce more general results that relate to the whole population and not only to few specific segments of it. Our study also compares the neighborhood effect with the traditional determinants of health used in economic models, namely income and education, which have been the subject of an extensive multidisciplinary literature. Unlike previous evidence on the health effects of the neighborhood coming from the US, we conduct our analysis in a European country. Italy is a natural candidate for this analysis, as it shows the highest geographical variation in health among OECD countries. To conduct this research, we used data from the income and living conditions survey carried out in 2004 by the Italian Statistical Office. This survey presents the considerable advantage of directly recording neighborhood characteristics along with health and individual characteristics, a feature that is not common in surveys on income and living conditions.

The empirical strategy consists of analyzing the effect of an index of neighborhood quality with respect to pollution, crime, and noise on three different health outcomes: Self-Assessed Health (SAH), chronic conditions, and limitations in Activities of Daily Living (ADLs). The identification issue is addressed by performing a whole series of endogeneity tests based on instrumental variable estimations. Another important characteristic of this study is that it also deals with the possible endogeneity of income with respect to health by means of an instrumental variable method. Note that not only retrieving the causal effect of income on health is important *per se*, but so is removing any possible second-order bias that this important determinant of neighborhood choice can transmit to the neighborhood effect on health. Finally, this additional sophistication allows us to properly compare the effect of the neighborhood to that of income and other individual determinants of health. Such a comparison is very important, as it enables the assessment of the potential of neighborhood-level policies relative to more traditional measures that aim at enhancing public health in urban communities.

The rest of the paper is organized as follows. The second part gives an overview of the literature on the neighborhood effect on health. The third describes the econometric model and empirical strategy adopted. The fourth presents the data and main variables used in the analysis. The fifth displays and interprets the results. Finally, the last part discusses and concludes the paper.

2. RELATED LITERATURE

A growing literature has documented numerous associations between various characteristics of the environment in which people live and different types of health outcomes. There is strong evidence of a positive association between the socio-economic characteristics of the place of residence and several health outcomes, such as cardiovascular risk factors (Davey Smith et al., 1998), diseases (Diez-Roux et al., 1997), mortality (Anderson et al., 1997), SAH and chronic diseases (Costa et al., 2003), as well as several health-affecting behaviors, such as smoking (Karvonen and Rimpela, 1996; Reijneveld, 1998), alcohol use and a lack of physical activity (Karvonen and Rimpela, 1996). Such evidence has also been found in Italy with respect to SAH, chronic diseases and ADLs (Carrieri and Bilger, 2010) and to mortality in the Italian city of Turin (Marinacci *et al.*, 2004).

Neighborhood characteristics can potentially affect health through several channels. Some of these involve aggregates of individual characteristics such as average income and educational levels. For instance, a high concentration of poor, less-educated individuals might adversely affect health because of the faster spreading of unhealthy lifestyles that are more concentrated among people in depressed socio-economic circumstances (Christakis and Fowler, 2007; Christakis and Fowler, 2008; Crane, 1991; Evans et al. 1992; Trogdon et al., 2008). In addition, deprived neighborhoods might have reduced levels of social capital, trust and social cohesion, all of which are positive determinants of health (Kakwachi et al., 1997, Kakwachi and Kennedy, 1997, Wilkinson, 1996).

Health could also be affected by other neighborhood characteristics that are purely contextual in the sense that they do not directly depend on individual characteristics and, in particular, not on the case mix of income and education. One of these purely contextual factors is the environmental quality of the neighborhood, which notably includes pollution, filth, noise and the presence of toxic substances. Empirical evidence indicates that these neighborhood-level variables are very strongly associated with health (Cadum *et al.*, 2002; Chappie, and Lester, 1982; Joyce et al. 1989; Seskin, 1979). Another important factor is the level of crime and vandalism in the neighborhood, which might adversely affect individual health both directly, by harming physical integrity, and indirectly, by leading to social isolation, stress and lack of physical activity (Macyntre *et al.*, 1993; Piro et al., 2006; Sooman and Macyntre, 1995). Crime is even thought to be the most important determinant of bad health in highly deprived neighborhoods in the United States (Minkler, 1992).

The main methodological limitation of all the above studies is the difficulty of identifying the neighborhood effect. Indeed, if health is a determinant of neighborhood choice, then a reverse-causality bias would be present. Such a perspective is however not held in the urban economics literature, as the studies that have investigated the determinants of residential choice do not consider health as being a potential determinant (Bayoh et al., 2006; Ioannides and Topa, 2010). A second type of endogeneity bias might arise because individuals might select their neighborhood according to unobservable factors that are correlated with their health. This is particularly relevant for the spreading of unhealthy lifestyles through peer effects. In this case, the adoption of a given lifestyle could be the determinant of the neighborhood choice (i.e., the individuals select peers with similar lifestyles) but, at same time, be the consequence of living in a place with peers having that lifestyle. This is the well-known *reflection problem*; that is, the difficulty of inferring whether the average behavior in some group influences the behavior of the individuals who compose the group (Manski, 1993).

Very few papers have dealt with these methodological challenges in the literature. Those that have successfully overcome the self-selection problem, to our knowledge, are all based on the MTO experiment. All of the MTO studies confirm the presence of health benefits resulting from living in a high-quality place. Kling et al. (2007) found significant mental health benefits for adults and for female youths. Katz et al. (2001), Ludwig et al. (2001), Rosenbaum and Harris (2001) found benefits for overall health (physical and mental, safety, boys' problem behavior, and well-being). In the next section, we present the identification strategy which makes it possible to complement this evidence with more general national representative data, and to compare the neighborhood effect with individual income and education.

3. IDENTIFICATION STRATEGY

3.1. Empirical model

Our model explaining individual health can be outlined as follows:

(1) $$H^* = \alpha * NI + \beta * Inc + \gamma * Edu + \delta * X + \varepsilon$$

where *H** represents latent health, *NI* an index of neighborhood quality, *Inc* a measure of individual income, *Edu* education, *X* a vector of control variables, α, β, γ, δ parameters to be estimated, and ε an error term. This is a latent variable model, as our measures of health are all qualitative, either ordinal or binary. *NI* measures neighborhood quality with respect to crime, pollution, and noise. The construction of this index is described in Section 4.3. As for the vector of control variables, *X*, this includes demographic variables (gender, age, household composition, and nationality), tenure status to account for wealth, housing problems (humidity, lightness, and heating problems), and regional fixed effects. Housing problems variables prevent from biasing the neighborhood effect by the effect that some housing characteristics have on health (see Graham, 2000, for a review). Finally, regional fixed effects are particularly relevant control variables given the high regional heterogeneity in health status existing in Italy.

3.2. Possible sources of bias

One may argue that the strong positive association found between neighborhood quality and health arises because of the self-selection of the residents in their neighborhood based on their initial health status. It is indeed possible that the individuals choose their place of residence according to the presence of healthcare facilities and other health services. In addition, disabled individuals might also choose their neighborhood because of its accessibility. However, we think that these problems should remain quantitatively limited as we are considering general health and not specific acute conditions. In addition, Italy distinguishes itself by having an extremely well-widespread healthcare supply and by providing good access to healthcare facilities. It is even possible that unhealthy individuals try to avoid bad neighborhoods, which would partly cancel out or even reverse the causality. As for the bias that might arise from missing variables in the health model, this should also be limited, as we notably control for the most important behavioral determinants of both neighborhood choice and health, i.e. income and education (Bayoh et al., 2006; Ioannides and Topa, 2010).

Another potential source of bias might come from the income-health relationship. Income is a positive determinant of health, as it increases access to healthcare, is necessary to buy health-promoting goods, and contributes to maintaining a high social status. However, reverse causality is also possible as good health increases earnings opportunities. It should be stressed that this identification problem of the income effect might also transmit a second-order bias to the estimate of the neighborhood effect. To deal with this issue, we use a classical instrumental variable strategy. To instrument income, we use two variables of non-earnings income, namely interests and dividends received, and income from the rental of a property or land. Non-earnings income does not come from activities that require being in good health, as is the case for work activities (Ettner, 1996; Lindhal, 2005). Although our instruments might depend on health history, we think that they are reasonable in the context of general health, which tends to vary more rapidly than acute health conditions, which are more permanent. It is finally worth noting that the endogeneity of income in our model explaining individual health should remain fairly limited, as our income variable is measured at household level.

3.3. Testing the endogeneity of the neighborhood effect

We use an instrumental variable approach for testing for the potential endogeneity of the neighborhood effect. The main difficulty lies in finding an exogenous instrument given the lack of theory on the precise mechanisms linking neighborhood and health. Our approach consists of applying the IV method despite this difficulty and complementing it with theoretical reasoning, additional estimations, and an over-identification test. Our instrument is the urbanization degree of the place of residence, measured by a categorical variable indicating whether the area is thinly populated (less than 100 inhabitants per Km^2), intermediate (100-500), or densely populated (more than 500). This choice of instrument is motivated by two factors. The first is that urbanization is greatly correlated with neighborhood choice, which provides power to the IV method. The second is that urbanization is easy to interpret and can handily fit in a theoretical reasoning. For instance, we could assume that urban areas are not harmful per se but give rise to various problems (e.g., pollution, crime, and noise), which in turn adversely affect individual health.

The challenge is that, like for the neighborhood, living in the city or the country results from a choice and thus could be endogenous in the health model. If this were the case, urbanization would not be an adequate instrument to correct the endogeneity bias. Nevertheless, we argue that it can still be used to both detect many forms of endogeneity and exclude many possible explanations supporting the endogeneity of neighborhood quality. To show that, let us consider the three cases: i) urbanization is exogenous; ii) only urbanization is endogenous; iii) both urbanization and neighborhood quality are endogenous. In the first case, the instrument is valid, as is the endogeneity test. In the second case, the association between health and urbanization is easily detected by including the latter into the model explaining the former, and by testing its significance. It is here not possible to determine whether neighborhood quality is endogenous or not, but it is at least clearly established that the IV method fails. The third case is less clear cut. However, we argue that if health (or any of its unobservable factors) influences both urbanization and neighborhood quality, it should do so in approximately the same magnitude to remain undetected by both the endogeneity and significance of the urbanization variable tests. This is especially true in our study, as the large sample size makes these tests fairly sharp. Moreover, from a theoretical perspective, we do not a priori expect health to influence both urbanization and neighborhood quality in the exact same way, as the underlying choices are of different nature and likely to involve different factors.

Consequently, when the endogeneity test is negative and urbanization is not statistically significant in the health model, the number of possible explanations supporting the endogeneity of neighborhood quality is very limited. We are left with two types of explanations, namely that sick individuals, either directly or because of unobservable health factors, tend to: a) simultaneously, and in the same order of magnitude, move to poor-quality neighborhoods and to the city (or, equivalently, stay in good neighborhoods and in the country), or b) simultaneously move to poor-quality neighborhoods and to the country. In the first case, we expect the association between health and neighborhood quality to be higher in the city due to the arrival of sick individuals in deprived neighborhoods, and lower in the country due to the departure of sick individuals to the city. The opposite reasoning holds for case b). To further exclude potential explanations supporting endogeneity, we include interaction terms between urbanization and neighborhood quality and test their significance.

Finally, after the above tests, endogeneity could still be supported—at most—by one explanation of either type a) or b). To further restrict the number of possibilities, we take advantage of the fact that, being categorical, our urbanization variable has to be introduced into the model by means of dummy variables. Therefore, it is technically possible to perform a test of over-identifying restrictions. We argue that the potential endogeneity should be of about the same magnitude for each urbanization degree to remain undetected. If this were not the case, one category would constitute an instrument that would be good enough to detect the endogeneity of the neighborhood effect. To conclude, when the endogeneity test is negative, urbanization is not statistically significant in the health model, and the over-identification test does not reject the exogeneity of urbanization, we are left with a very unlikely explanation (of type a or b) supporting the endogeneity of neighborhood quality. It would also be necessary to establish reasonable theoretical grounds for such explanation.

4. DATA

4.1. Sample

We use data from the income living conditions survey carried out by the Italian Statistical Office (Istat) in 2004. This cross-sectional survey collects information on 61,429 individuals living in 24,204 households. Data are provided at both the individual and the household level and relates to the following types of living conditions: income, education, health, work conditions, social exclusion, and housing. We exploit an appealing feature of this dataset, which is the recording of characteristics of the places in which people live. Such information are absent in other Italian survey on living conditions (i.e., The Bank of Italy survey on income and wealth). It may also be noted that these data were collected independently of the Italian ECHP longitudinal sample, which does not contain neighborhood variables; consequently, no panel dimension is available for our study. The lack of neighborhood variables in longitudinal data is common among all European surveys on living conditions.

4.2. Measures of health

To check the robustness of our results, we use three different variables to measure health status. The first is SAH, which is measured by the five conventional answers: very good, good, fair, bad, and very bad. SAH is widely used in the literature, as it has been shown to be closely correlated with objective health measures such as mortality (Idler and Benyamini, 1995). In addition, it is a convenient aggregate that covers all aspects of health. However, it has also been shown that its use can be problematic for geographical analysis, as it is vulnerable to adaptive preferences according to the context in which individuals live (Sen, 2002). For this reason, we also carry out our analysis with two other health variables, which are characterized by a greater level of objectivity. These are the presence of chronic illness as diagnosed by a physician (yes or no), and the presence of limitations in ADLs, with three possible answers: strongly limited, limited, and not limited.

4.3. Neighborhood quality index

The characteristics of the place of residence are self-assessed by the individuals who are asked to report the presence or absence of the three following problems: *i)* pollution, filth, and other environmental problems, *ii)* noise caused by neighbors, traffic, and industry, among other sources, and *iii)* crime, violence and vandalism in the area. As for the place of residence itself, it is defined as the area where the individuals usually shop and walk back home. This is consistent with the definition of neighborhood used in the economic literature (Manski, 1993). A self-assessed—and thus subjective—evaluation of the neighborhood characteristics should provide reasonable estimates, as it has been found that such measures are strongly correlated with the objective characteristics of the neighborhood (Elo et al., 2009). We have aggregated these three measures of neighborhood problems into a single index indicating neighborhood quality. A higher index indicates more neighborhood problems and lower neighborhood quality. We have exploited the fact that there are only eight different possible combinations between the three binary neighborhood variables. We first introduced these combinations into the health model by means of dummy variables, and, by testing different groupings, found that we can simply count the number of neighborhood problems. We then introduced one dummy variable per number of neighborhood problems into the health model. We finally tested various simple functional forms and established that the square root of the number of neighborhood problems is an adequate specification for our neighborhood quality index.

4.4. Control variables and descriptive statistics

After neighborhood quality, our most important covariate is individual income, which we measure as the equivalent household income[1]. Income is usually linearly introduced into health models (e.g., Ettner, 1996), but we find this assumption to be overly restrictive. That is why we have compared various specifications of income using the Akaike goodness-of-fit criterion and found that that the square root is the best specification, which notably allows the income effect to decrease with income. Another important covariate is education, which is measured as a four-category variable indicating whether the individual education level is: primary or less, secondary I, secondary II, or

[1] Income is measured as total household disposable income divided by the equivalized household size (EHS), where EHS: 1+ 0.5 * ($HM14_+$ -1) + 0.3 * HM13, with $HM14_+$ and $HM13_-$ as the number of household members aged 14 or more and 13 and less, respectively.

tertiary. Note also that, by using Wald tests, we have established that the optimal specification of age is a degree 4 polynomial, with interaction terms to account for the different patterns between men and women.

Finally, Table 1 provides detailed descriptive statistics on our key variables. The sample mean of the remaining covariates can be found in Table A.1. Table 1 clearly shows that neighborhood problems are highly correlated. For instance, in 51.8 percent of the neighborhoods where crime is reported, pollution is also deemed to be a problem. This has to be compared with the unconditional mean, which indicates that only 21.1 percent of the neighborhoods are deemed polluted in the whole population. Table 1 also shows that health problems are more prevalent when neighborhoods problems are also reported. For instance, 25.0 percent of the individuals living in polluted neighborhoods have a chronic condition, in contrast to 21.6 percent in the general population. Interestingly, health is better in densely populated areas. This is likely to result from an income effect, as income is significantly higher in the cities, as shown in the last column of the table. However, it is important to stress that health remains an important urban problem. This is revealed by the fact that crime, pollution, and noise are dramatically more concentrated in densely populated areas, which in turn affects the health of those who live in these deprived neighborhoods. For instance, pollution is a problem for 58.4 percent of those living in the cities as compared with only 10.8 percent in the country.

Table 1: Descriptive statistics of the main variables

	Urbanization			**Bad health**			**Neighborhood**			**Income**
	thin	*interm.*	*dense*	*SAH*	*ADLs*	*Chronic*	*crime*	*pollution*	*noise*	*mean*
Urbanization										
Thin	100	0	0	11.3	16.8	22.2	3.2	9.4	15.8	14214
interm.	0	100	0	10.2	15.0	21.5	8.1	16.6	22.8	15449
Dense	0	0	100	9.9	14.1	21.3	24.5	33.8	34.4	16308
Bad health										
SAH	26.6	38.7	34.7	100	79.2	85.2	14.8	24.5	29.2	12985
ADLs	27.0	39.0	34.0	54.4	100	82.6	14.3	24.0	28.4	13641
Chronic	25.0	39.0	36.0	40.9	57.8	100	14.3	24.4	29.0	14496
Neighborhood										
Crime	6.1	24.7	69.3	11.9	16.7	23.9	100	51.8	52.1	15147
pollution	10.8	30.8	58.4	12.0	17.2	25.0	31.7	100	63.8	15826
Noise	15.2	35.3	49.5	11.9	16.9	24.7	26.5	53.1	100	15456
Unconditional means										
	24.3	39.2	36.5	10.4	15.1	21.6	12.9	21.1	25.4	15462

Table presents the mean of the column variables conditionally to the row variables. Percentages are displayed except for income, whose unit is equivalent Euros. The following categories have been grouped: “bad” and “very bad” for SAH, and “limitations” and “severe limitations” for chronic diseases.

5. RESULTS

In this section, we present our estimations of the empirical model described in Section 3.1. We started by estimating exogenous ordered probit (for SAH and ADLs) and probit (for chronic conditions) models by maximum likelihood. Then, to account for the potential endogeneity of income, we applied a two-stage residual inclusion IV method (Newey, 1987) with the instruments discussed in Section 3.2. Finally, to perform the endogeneity tests described in Section 3.3, we applied the same two-stage method using instruments for both income and neighborhood quality. For all estimates, we have computed the standard errors by bootstrapping with 1,000 repetitions.

The results of the endogeneity tests discussed in Section 3.3 are summarized in Table 2. The first part shows that the instruments are very powerful, as all F-test statistics are very large. The tests displayed in the second part do not reveal any endogeneity for neighborhood quality for either of the health variables. As for income, we find it to be endogenous for SAH and ADLs but not for chronic conditions. Our next tests show that urbanization is not correlated with health when neighborhood quality is included into the models for SAH and chronic conditions. However, urbanization is directly correlated with ADLs to some extent; therefore, the exogeneity of this instrument is questionable. In the fourth part, we show that the interaction terms between neighborhood quality and urbanization are statistically significant for all three health models. In addition, the coefficients are all negative, which means that the neighborhood effect is greater (in absolute terms) in thinly populated areas. Therefore, any explanation involving that the positive relationship between good health and neighborhood quality is due to the departure of sick individuals from the country to live in good neighborhoods in the cities would not be consistent with our findings. Finally, the tests of over-identifying restrictions do not detect any sign of endogeneity of urbanization for any of the three health variables. As for the instruments for income, they are found to be endogenous only for ADLs.

The above tests indicate that neighborhood quality is very unlikely to be endogenous for the SAH and chronic conditions models. Note that it was not possible to properly test for the presence of endogeneity in the case of ADLs, as our instrument is endogenous for this health variable. This is probably due to the fact that individuals with physical limitations tend to live in densely populated areas for their better accessibility. Interestingly, as we will see below, the neighborhood effect for ADLs is very comparable to those of the two other health conditions, which suggests no or little endogeneity bias for ADLs as well. That is why we present the results for ADLs along with those of SAH and chronic conditions. The absence of endogeneity might seem surprising at first, but it should be borne in mind that the study takes place in a country where healthcare is publicly provided and its distribution particularly well widespread, and that general health is analyzed. In addition, the urban economics literature shows that the main determinants of neighborhood choice are income and education (Bayoh et al., 2006; Ioannides and Topa, 2010) which are accounted for in our model. All these factors eliminate or strongly reduce the importance of health in the neighborhood choice, which in turn limits the bias that might affect estimates of the neighborhood effect.

Table 2: Endogeneity tests

Statistical Test	**Instrumented Variable**	**Health Model** SAH	Chronic	ADLs
Weak instruments test	neighborhood	1787	1787	1787
(partial F-test statistic)	income	1943	1943	1943
	both	7711	7711	7711
Endogeneity test[a]	neighborhood	0.856	0.795	0.160
(H_0: exogenous variable, p-value)	income	0.002	0.544	0.016
Urbanization degree, Wald test[b] (p-value)	none	0.501	0.841	0.025
intermediate area, coefficient		-0.009	-0.003	-0.017
		(0.012)	(0.017)	(0.018)
thinly populated area, coefficient		-0.017	-0.013	-0.058
		(0.015)	(0.021)	(0.022)
Interaction terms, Wald test[b] (p-value)	none	0.040	0.000	0.004
neighb. quality × intermediate area, coef.		-0.033*	0.004	-0.013
		(0.018)	(0.023)	(0.028)
neighb. quality × thinly populated area, coef.		-0.053**	-0.121***	-0.109***
		(0.024)	(0.031)	(0.034)
Lee (1992) over-identifying restrictions test	neighborhood	0.297	0.744	0.158
(H0: exogenous instruments, p-value)	income	0.133	0.196	0.014
	both	0.246	0.407	0.017

[a]Test of significance of the residuals included into the second stage of the IV estimation.
[b]The p-values for categorical variables correspond to the joint test that all categories have the same effect.
Standard errors displayed in parentheses, and significance levels are: *** = 1%, ** = 5%, and * = 10%.

According to the above endogeneity tests, we selected the most efficient unbiased models. These are here the income-instrumented model for SAH, and the exogenous model for chronic conditions and ADLs. The estimated coefficients of these models are displayed in Table A.1. We then calculated the corresponding incremental effects of neighborhood quality, income, and education, which are displayed in Tables 3, 4, and 5, respectively.

Table 3: Neighborhood effect on health

	Probability	Incremental effect		
	(no problems)	1 problem	2 problems	3 problems
SAH (ref: very good)				
very bad	0.004	0.001	0.002	0.002
	(0.000)	(1.4e-04)	(2.2e-04)	(2.9e-04)
Bad	0.042	0.008	0.012	0.015
	(0.002)	(8.1e-04)	(0.001)	(0.001)
Chronic disease (ref: no)				
Yes	0.167	0.035	0.050	0.062
	(0.009)	(0.003)	(0.005)	(0.006)
ADLs (ref: no limitations)				
Severe	0.024	0.006	0.009	0.011
	(0.002)	(8.9e-04)	(0.001)	(0.002)
Moderate	0.072	0.012	0.017	0.021
	(0.004)	(0.002)	(0.002)	(0.003)

Standard errors displayed in parenthesis.

First of all, it can be observed that the effects of all these three determinants are unambiguously positively associated with good health. Indeed, better neighborhood quality, higher income, and higher education respectively increase the likelihood of reporting better health, not suffering from chronic conditions, and not having severe or moderate limitations in ADLs. These results are in line with the vast literature on the socioeconomic determinants of health, where it is established that both income and education are positive determinants of health (see reviews by Adams et al., 2003, and Wilkinson and Pickett, 2006). Regarding neighborhood quality, its positive effect on health has been found in many other papers, but except for the MTO studies, none of them have dealt with the identification of neighborhood quality (see Section 2). As for the effects of the other covariates, we find similar estimates across all three health models, suggesting that the subjective evaluation of health is not dissimilar from more objective measures, which was also found by other authors (Idler and Benyamini, 1995). Finally, we also observe a strong influence of housing conditions on health, which is in line with many other empirical studies (see Macyntre *et al.*, 2000, for a review), and a strong regional variability in health conditions, which is consistent with previous empirical research in Italy (Costa *et al.*, 2003; Carrieri, 2008).

Table 4: Income effect on health

	Probability	**Incremental effect**	
	(1st quartile)	Median	3rd quartile
SAH (ref: very good)			
very bad	0.006	-0.011	-0.001
	(5.3e-04)	(0.000)	(0.000)
Bad	0.055	-0.007	-0.013
	(0.003)	(0.001)	(0.003)
Chronic disease (ref: no)			
Yes	0.189	-0.005	-0.010
	(0.009)	(0.001)	(0.002)
ADLs (ref: no limitations)			
Severe	0.029	-0.002	-0.004
	(0.003)	(0.000)	(0.000)
Moderate	0.082	-0.003	-0.007
	(0.005)	(0.000)	(0.001)

Standard errors displayed in parenthesis.

Table 5: Educational effect on health

	Probability	**Incremental effect**		
	(primary or less)	secondary I	secondary II	tertiary
SAH (ref: very good)				
Very Bad	0.008	-0.004	-0.005	-0.007
	(7.0e-04)	(4.4e-04)	(5.2e-04)	(6.1e-04)
Bad	0.069	-0.024	-0.045	-0.044
	(0.003)	(0.002)	(0.002)	(0.003)
Chronic disease (ref: no)				
Yes	0.225	-0.004	-0.059	-0.101
	(0.011)	(0.006)	(0.006)	(0.008)
ADLs (ref: no limitations)				
Severe	0.044	-0.016	-0.025	-0.033
	(0.004)	(0.002)	(0.002)	(0.003)
Moderate	0.106	-0.027	-0.045	-0.063
	(0.006)	(0.003)	(0.003)	(0.004)

Standard errors displayed in parentheses.

We find that the neighborhood and educational effects on health are of comparable magnitude. For instance, the difference in the likelihood of suffering from chronic conditions between good-quality and severely deprived (i.e. where three problems are reported) neighborhoods is comparable to the health difference between individuals with no more than a primary education and those with an upper-secondary education (with an incremental effect of around 6%). Similarly, the increase in likelihood of having severe limitations in ADLs in a poor-quality neighborhood is comparable to the health differential between individuals with no more than a primary education and those with an upper-secondary education (with an incremental effect of around 1%). Only the primary-tertiary gap has a stronger effect on health than neighborhood quality, with incremental effects being close to 10% on average. With respect to income and neighborhood comparison, we find that the effect of the latter is substantially higher than the effect of the former for all three health variables.

Finally, given that the Italian population is of around 56,000 inhabitants, the incremental effect of 1% we find for ADLs means that 560,000 individuals suffer from limitations in ADLs because of poor neighborhood quality. Similarly, the 6% incremental effect we find for chronic diseases represents more than 3 million individuals having a chronic disease because of their neighborhood.

6. CONCLUSION

In this paper, we estimate the effect of a neighborhood quality index based on pollution, crime, and noise on three different measures of individual health, namely Self-Assessed Health (SAH), presence of chronic conditions, and limitations in Activities of Daily Living (ADLs). An original feature of this study is the production of general results by means of non-experimental data from the 2004 Italian income and living conditions survey. This requires accounting for the self-selection of the residents in their neighborhood. To address this issue, we made extensive use of econometric tests to detect the possible endogeneity of the neighborhood quality index in the health model. In addition, we also dealt with the possible endogeneity of the income variable.

We find that poor-quality neighborhoods are strongly health damaging, even more so than economic deprivation and about as much so as having the lowest educational levels, which represent the two traditional socioeconomic determinants of bad health. For instance, we find that the difference in the likelihood of suffering from chronic conditions between good and bad neighborhoods is comparable to the health difference between those who have no more than a primary education and those who have an upper-secondary education (an incremental effect of around 6%). In terms of public health, we find for instance, that out of 10 million individuals who suffer from chronic diseases, more than 3 million cases are caused by poor-quality neighborhoods in Italy.

The results of this paper suggest that neighborhood-level policies represent an important tool to improve public health. In Italy, such policies are already being carried out. In 2002, the new profession of neighborhood police officers was created, the task of which is to prevent and punish episodes of crime and vandalism occurring at the neighborhood level. However, in 2006, only 10% of the Italian territory was covered by this new kind of police officers. With respect to pollution and noise problems, there are some zones where cars are not allowed; these zones were gradually created by communes over time. Note that these areas are not present everywhere and only exist in city centers. Considering the substantial negative impact that we find for poor-quality neighborhoods, it seems advisable to strengthen this kind of policies to enhance public health in urban communities.

A limitation of our study is that we use a neighborhood quality index based on the problems perceived by the residents rather than on objective indicators. However, we are confident about the robustness of our results for two reasons. First, it has been shown that objective and subjective measures of neighborhood quality are strongly correlated (Elo et al., 2009). Second, we observed a strong impact of neighborhood characteristics on both subjective and objective health measures.

As for future directions of research, it would be interesting to analyze the effect of life-time exposures to bad neighborhood quality as this effect is likely to be cumulative. To do this, though, information on past places of residency would be necessary, which would require undertaking a survey dedicated to this topic. It would also be fruitful for economists to systematically consider neighborhood effects in health models. More generally, we think that more attention should be paid to the role played by the neighborhood in all areas of economics and not only when explaining health outcomes. Indeed, even when one restricts the scope to the neighborhood characteristics considered here (i.e., crime, pollution, and noise), one can anticipate that these characteristics will have a considerable impact on a wide range of social and economic outcomes. Strikingly, economic research on these topics is thin compared to that in other social sciences. This is probably due to the fact that economists have been a priori deterred by the challenge raised by the self-selection problem. In this paper, we show that this does not necessarily represent an insuperable problem when assessing the neighborhood effect on health. In particular, we show that the sorting mechanisms according to neighborhood quality strongly depends on observable covariates such as income and education which can easily be included into econometric models. In addition, in countries like Italy with free universal coverage, these sorting mechanisms are unlikely to depend on health status. This paper shows that in such cases, it is possible to obtain an accurate and general measure of the effect that neighborhood quality has on health, and this using readily available survey data

REFERENCES

Adams, P., Hurd, M.D., McFadden, D., Merrill, L., Ribeiro, T., 2003. Healthy, Wealthy, and Wise? Tests for Direct Causal Paths between Health and Socioeconomic Status. *Journal of Econometrics*, 112, 1, 3-56.

Anderson, R., Sorlie, P., Backlund, E., Johnson, N., Kaplan, G.A., 1997. Mortality effects of community economic status, Epidemiology, 8,42–7.

Bayoh, I., Irwin, E.G.G., Haab, T.C., 2006. Determinants of Residential Location Choice: How Important are Local Public Goods in Attracting Homeowners to Central City Locations? Journal of Regional Science, 46, 1, 97-120.

Cadum E., Demaria M., Martuzzi M., Costa G., 2002. Problemi ambientali e salute nella percezione delle persone (Analisi dei dati dell'Indagine Multiscopo Istat 1998) In Sabbadini, L.L., Costa, G., (Eds.) *Informazione statistica e politiche per la promozione della salute. Atti del convegno 10-12 settembre 2002*. Istat, Roma.

Carrieri, V., 2008. Disuguaglianze di salute e condizioni socioeconomiche in Italia: esiste una questione meridionale?, Politiche sanitarie, 9,1,15-24.

Carrieri, V., Bilger, M., 2010. Dimmi dove vivi e ti dirò come stai: Un'indagine empirica sugli effetti del quartiere di residenza sui risultati di salute, Department of Economics and Statistics, University of Calabria, Working papers series, 3/2010.

Chappie, M., Lester, L., 1982. The Health Effects of Air Pollution: A Reanalysis, Journal of Urban Economics, 12, 346-376.

Christakis, N.A, Fowler, J.H, 2008. The collective dynamics of smoking in a large social network, New England Journal of Medicine, 358, 2249–2258.

Christakis, N.A., Fowler, J.H, 2007. The spread of obesity in a large social network over 32 years, New England Journal of Medicine 357, 370–379.

Costa, G., Marinacci, C., Caiazzo, A., Spadea, T., 2003. Individual and contextual determinants of inequalities in health. The Italian Case, International Journal of Health Services, 33, 4, 635-667.

Crane, J., 1991. The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Childbearing, The American Journal of Sociology, 96, 5, 1226-1259.

Cutler, D., Deaton, A., Lleras-Muney, A., 2006. The Determinants of Mortality, Journal of Economic Perspectives, 20, 3, 97–120.

Davey Smith, G., Hart, C., Watt, G., Hole, D., Hawthorne, V., 1998. Individual social class, area-based deprivation, cardiovascular disease risk factors, and mortality: the Renfrew and Paisley study. Journal of Epidemiology and Community Health, 52, 399–405.

Diez-Roux, A., Nieto, F., Muntaner, C., Tyroler, H.A., Comstock, G.W., Shahar, E., Cooper, L.S., Watson, R.L., Szklo, M., 1997. Neighborhood environments and coronary heart disease: a multilevel analysis, American Journal of Epidemiology, 146, 48–63.

Elo, I., Mikyta, L., Margolis, R., Culhane, J., 2009. Perceptions of Neighborhood Disorder: The Role of Individual and Neighborhood Characteristics, Social Science Quarterly, 90, 5, 1298-1320.

Ettner, S.L., 1996. New evidence on the relationship between income and health, Journal of Health Economics, 15, 67-85.

Evans, W.N., Oates, W.E., Schwab, R.M., 1992. Measuring peer group effects: a study of teenage behavior, Journal of Political Economy 100, 5, 966–991.

Hanlon, P., Walsh, D., Whyte, B., 2006. Let Glasgow Flourish, Glasgow Centre for Population Health, Glasgow.

Idler, E.L. Benyamini, Y., 1997. Self rated health and mortality: a review of 27 community studies, Journal of Health and social behaviour, 38,1, 21-37.

Ioannides, M.Y., Topa, G., 2010. Neighborhood Effects: Accomplishments And Looking Beyond Them, Journal of Regional Science, 50, 1, 343-362.

Joyce, T.J., Grossman, M., Goldman, F., 1989. An assessment of the benefits of air pollution control: The case of infant health, Journal of Urban Economics 25, 32-51.

Karvonen, S., Rimpela, A., 1996. Socio-regional context as a determinant of adolescents' health in Finland. Social Science and Medicine, 43,1467–74.

Katz, L.F., Kling, J., Liebman, J.B., 2001. Moving to Opportunity in Boston: early results of a randomized mobility experiment, Quarterly Journal of Economics, 116, 2, 607–54.

Kawachi, B., Kennedy,P., Lochner, L., Prothrow-Stith, D., 1997. Social capital, income inequality, and mortality, American Journal of Public Health, 87, 9, 1491-1498.

Kawachi, B., Kennedy, P., 1997. Socioeconomic determinants of health: Health and social cohesion: why care about income inequality?, British Medical Journal, 314:1037.

Kling, J.R., Liebman, J.B., Katz, L.F., 2007. Experimental Analysis of Neighborhood Effects, Econometrica, Econometric Society, 75, 1, 83-119.

Lindahl, M., 2005. Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income, Journal of Human Resources, 40, 144-168.

Ludwig, J., Hirschfield, P., Duncan, G.J., 2001. Urban poverty and juvenile crime: evidence from a randomized housing-mobility experiment, Quarterly Journal of Economics 116 , 2, 665–79.

Macyntre, S., Hiscock, R., Kearns, A., Ellaway, A., 2000. Housing tenure and health inequalities: a three-dimensional perspective on people, homes and neighborhoods, in Graham, H., (Eds.), Understanding health inequalities, Open University Press, 129-142.

Macintyre, S., Maciver, S., Sooman, A., 1993. Area, class and health: should we be focusing on places or people? Journal of Social Policy, 22, 213–34.

Manski, C.F., 1993. Identification of Endogenous Social Effects: The Reflection Problem, The Review of Economic Studies, 60, 3, 531-542.

Marinacci, C., Spadea, T., Buggeri, A., Demaria, M., Caiazzo, A., Costa, G., 2004. The role of individual and contextual socio-economic circumstances on mortality: analysis of time variations in a city of Northwest Italy, Journal of Epidemiology and Community Health, 58, 3,199-207.

Minkler, M., 1992. Community organizing among the elderly poor in the United States: a case study, International Journal of Health services, 22, 2, 303-16.

Murray, C.J.L., Kulkarni, S.C., Michaud, C., Tomijima, N., Bulzacchelli, M.T., Iandiorio, T.J., Ezzati, M., 2006. Eight Americas: Investigating Mortality Disparities across Races, Counties, and Race-Counties in the United States, PLoS Medicine, 3, 9: e260.

Newey, W.K., 1987. Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables, Journal of Econometrics, 36, 231–250.

Piro, F.N., Noess, Ø.,Claussen, B., 2006. Physical activity among elderly people in a city population: the influence of neighbourhood level violence and self perceived safety, Journal of Epidemiology and Community Health, 60, 626-632.

Reijneveld, S., 1998. The impact of individual and area characteristics on urban socioeconomic differences in health and smoking. International Journal of Epidemiology, 27, 33–40.

Rosenbaum, J.E., Harris, L.E., 2001. Low-income families in their new neighborhoods, Journal of Family Issues 22, 2, 183–21.

Sen, A., 2002. Health: perception Vs Observation, British Medical Journal, 324, 860-861.

Seskin, E., 1979. Pollution and health in Washington DC, Journal of Urban Economics, 6, 275-291.

Sooman, A., Macyntre, S., 1995. Health and perception of the local environment in socially contrasting neighborhoods in Glasgow, Health and place, 1, 1, 15-26.

Trogdon, J.G., Nonnemaker, J., Pais, J., 2008. Peer effects in adolescent overweight, Journal of Health Economics 27, 5, 1388-1399.

Wilkinson, R.G., 1996. Unhealthy Societies: The Afflictions of inequality. New York: Routledge.

Wilkinson, R.G., Pickett, K, E, 2006. Income inequality and population health: a review and explanation of the evidence, Social Science and Medicine, 62, 7, 1768-84.

APPENDIX

Table A.1: Sample means and estimated coefficients of the health models

	Sample means	**SAH** (IV income)	**Chronic** (exogeneous)	**ADLs** (exogeneous)
Neighborhood index		-0.091***	-0.130***	-0.098***
		(0.008)	(0.011)	(0.012)
Square root of equivalized income		9.7e-06***	2.8e-06***	4.9e-06***
		(2.0e-06)	(7.9e-07)	(8.6e-07)
education (ref: primary or less)				
secondary 1	0.284	0.237***	0.176***	0.209***
		(0.017)	(0.020)	(0.022)
secondary 2	0.328	0.360***	0.216***	0.365***
		(0.020)	(0.021)	(0.024)
Tertiary	0.086	0.533***	0.403***	0.565***
		(0.031)	(0.031)	(0.039)
Male	0.477	-0.007	-0.569	0.041
		(0.385)	(0.677)	(0.723)
Age	47.277	-0.056*	-0.018	0.057
		(0.029)	(0.048)	(0.051)
age^2		1.5e-04	-7.2e-04	-0.004**
		(0.001)	(0.002)	(0.002)
age^3		5.1e-06	1.7e-05	7.1e-05***
		(1.5e-05)	(2.3e-05)	(2.5e-05)
age^4		-5.6e-08	-1.2e-07	-4.5e-07***
		(7.8e-08)	(1.2e-07)	(1.2e-07)
male × age		7.5e-04	0.047	-0.023
		(0.040)	(0.067)	(0.071)
male × age^2		2.8e-04	-0.001	0.001
		(0.001)	(0.002)	(0.002)
male × age^3		-7.9e-06	1.1e-05	-2.4e-05
		(2.1e-05)	(3.3e-05)	(3.4e-05)
male × age^4		5.8e-08	-2.6e-08	1.5e-07
		(1.1e-07)	(1.7e-07)	(1.7e-07)
Foreigner	0.026	0.132***	0.217***	0.083
		(0.033)	(0.051)	(0.055)
houshold type (ref: 1 adult)				
2 adults, no children	0.242	-0.010	0.025	0.004
		(0.017)	(0.023)	(0.023)
other, without children	0.230	0.023	0.086***	0.027
		(0.020)	(0.024)	(0.025)
1 adult, 1 child	0.019	0.035	0.149**	0.156**

		(0.042)	(0.060)	(0.074)
2 adults, 1 child	0.110	0.049**	0.158***	0.200***
		(0.022)	(0.031)	(0.035)
2 adults, 2 children	0.122	0.109***	0.225***	0.260***
		(0.021)	(0.032)	(0.035)
2 adults, 3 children	0.030	0.184***	0.271***	0.311***
		(0.034)	(0.052)	(0.060)
other, with children	0.129	0.065***	0.158***	0.136***
		(0.021)	(0.028)	(0.030)
home owner	0.758	0.010	-0.013	0.020
		(0.012)	(0.017)	(0.017)
humid dwelling	0.232	-0.178***	-0.199***	-0.177***
		(0.013)	(0.017)	(0.017)
cold dwelling	0.096	-0.209***	-0.183***	-0.242***
		(0.019)	(0.023)	(0.023)
dark dwelling	0.087	-0.053***	-0.095***	0.081***
		(0.019)	(0.024)	(0.025)
residuals, income equation		-5.5e-06** (2.2e-06)	-	-

Regional fixed effects and latent variable thresholds have been estimated but are not displayed. Standard errors displayed in parentheses, significance levels: *** = 1%, ** = 5%, and * = 10%.