

## MATERIAL DEPRIVATION AS MARKER OF HEALTH NEEDS

L. Grisotto, D. Catelan, G. Accetta, A. Biggeri

### 1. INTRODUCTION

In the epidemiological literature a relationship between socio-economic status and health was largely documented both by individual-level and ecological studies. A pubmed search on “socio-economic status and health” returns 21,319 papers (for Italy see Costa *et al.*, 2004; Dreassi *et al.*, 2005; Catelan *et al.*, 2006).

There was also a variety of health indicators (e.g. mortality, hospital admissions, out-patient visits, pharmaceutical prescriptions) and socio-economic indicators (e.g. simple vs composite; income vs social) used in studying this association. In the present paper we faced with statistical aspects that need to be addressed and that directly depend on the level of available data.

Ecological regression analysis (Morgenstern 2008) are usually performed on a response variable (e.g. Standardized Mortality Ratio) which is adjusted for some relevant confounder (such as age). This approach is also useful to reduce the dimensionality of the data since disease counts are collapsed by age classes within area. However, it is common to find that the covariates included in the ecological regression model are not adjusted for the same confounder. Rosenbaum and Rubin (1984) denoted this as the “mutual standardization” problem. They shown that in linear regression models where directly age standardized rates are used as dependent variable and covariates are not age-adjusted the regression coefficients are biased. When the independent variable is not adjusted we are implicitly assuming that it is constant throughout all the strata of the confounding variable. If this assumption is not true the result is biased (Rosenbaum and Rubin 1984, p. 440). Wakefield (2007) suggested that to correctly estimate the covariate effect it is sufficient to know the bivariate confounder by regressor joint distribution and not the full tri-variate response by confounder by regressor joint distribution.

Some aspects of poverty exert their action at an ecological and not only at an individual level. The effect of socio-economic variables on health is complex, and might be mediated by the average level of deprivation of the area in which a person lives. It means that a better off person that lives in a poor area have worst health outcomes than he would have living in a non disadvantaged area. In the sociological literature this was named “contextual effect” (Firebaugh, 1978).

Eventually, the observed mortality in a high deprived area would be bigger than that we would expect summing up the effects of all the disadvantaged conditions observed at individual level in that area.

These arguments are relevant when considering the issue of introducing socio-economic indicators in resource allocation formulas (Carr Hill *et al.*, 1994, Ministero della Salute, 1998). Generally speaking, allocation criteria are usually based on a weighted capitation formula determined by some weighting scheme (Stone and Galbraith, 2006). The weights may reflect the demographic and socio-economic structure of the population, which are strong determinants of the health needs of a given population. First attempts in this direction were done in UK with the creation in 1948 of the *National Health Service* (NHS). The initial allocation formulas considered only demographic characteristic of the population (age). The emerging evidence of a relationship between health and socio-economic status motivated the introduction in the allocation formula of some measures of social or material deprivation (Jarman, 1983, Carr Hill *et al.*, 1994). In Italy, in the 2006 State-Regions Conference (Roma, 20 aprile 2006), it was discussed the introduction of socio-economic indicators in the allocation formula.

In this work we considered a material deprivation index as predictor of health needs and we addressed the problem of mutual standardization comparing results when adjusting or not health outcome and deprivation for the same confounding variable. The availability of individual-level data allowed us to compare models with different data hierarchies and to evaluate both the individual effect of deprivation on health status and the aggregate effect, disentangling the individual from the contextual effects. In fact, we took advantage of data from the Florence 1991 census cohort (Tuscan Longitudinal Study, Biggeri *et al.*, 2001).

In section 2 we describe the data. In section 3 we present the individual and the ecological models used. Results are presented in section 4. Discussion and conclusions follow in section 5.

## 2. MATERIALS

### 2.1 Mortality data

Individual mortality data come from the Florence 1991 census cohort which is part of the Tuscan Longitudinal Study (SLT<sub>o</sub>, Biggeri *et al.*, 2001). We considered the resident population present in the city of Florence at the census day 31<sup>st</sup> October, 1991. The follow-up ended on 31<sup>st</sup> December, 1995. We analyzed mortality for all causes (ICD IX 001-999) among males aged between 18 and 75 at entry. We enrolled 144,001 people for 568,713 person years of follow up and observed 4,312 deaths (a crude mortality rate of  $7.6 \cdot 10^{-3}$ ) in 2,752 census blocks.

### 2.2 Material Deprivation data

Data on socio-economic factors derived from individual records of the 1991 census. A material deprivation index at individual level has been constructed as

the sum of four indicators of adverse events: low education (less than 6 years of formal education), unemployment, house characteristics (<25 sqm and bathroom outside the house). The index is zero if a person has none of the four events, one if a person has 1 adverse event and two if she has 2 or more adverse events.

### 3. METHODS

#### 3.1 Individual level models

Let  $Y_{jisa}$  be a Bernoulli variable which takes the value 1 if the  $j$ -th subject in the  $i$ -th level of deprivation  $s$ -th census block and  $a$ -th age class died and 0 otherwise. Indeed, the material deprivation index is discrete with value  $i=0, 1, 2$ . If we consider the frequency of people by each  $i$ -th level of deprivation,  $s$ -th census block and  $a$ -th age classes then  $Y_{isa} = \sum_j Y_{jisa}$  is the observed number of death, and it follows a Poisson distribution with mean  $(T_{isa} \times \theta_{isa})$  where  $\theta_{isa}$  is the relative risk and  $T_{isa}$  the number of person years of observation or the expected number of cases under internal indirect standardization.

In fact, to adjust by age, internal indirect standardization is used and the expected number of deaths is calculated as  $PY_{isa} \times \theta_{sa}^* = E_{isa}$ , where  $PY_{isa}$  is person year of observation,  $\theta_{sa}^*$  is the reference rate and  $E_{isa}$  the expected counts. Notice that we assume the multiplicative model  $\theta_{isa} = \theta_{is} \times \theta_{sa}$  (Breslow and Day 1978, 1986).

We specified three generalized linear models on the logarithm of the relative risk:

1) The *pooled* model:

$$\log(\theta_{isa}) = \alpha + \beta^T x_{isa} + \sum \xi_a age_a \quad (1)$$

where  $\alpha$  is the intercept,  $x_{isa}$  is the individual-level material deprivation index,  $age_a$  is an indicator variable for age class,  $\xi_a$  is the age effect ( $\xi_{a=1} = 0$  for identifiability) and  $\beta^T$  is the *pooled* individual effect of material deprivation.

2) The *contextual* model (Firebaugh, 1978):

$$\log(\theta_{isa}) = \alpha + \beta^I x_{isa} + \beta^C \bar{x}_s + \sum \xi_a age_a \quad (2)$$

where  $x_{isa}$  is the individual-level material deprivation index,  $\bar{x}_s$  is the average material deprivation of the  $s$ -th census block ( $\bar{x}_s = \sum_a^A \sum_i^{I_a} x_{isa} / \sum_a^A I_a$ ),  $\beta^I$  e  $\beta^C$  are the individual and the contextual effects, respectively.

3) The *Cronbach* model (Cronbach e Webb, 1975):

$$\log(\theta_{isa}) = \alpha + \beta^I (x_{isa} - \bar{x}_s) + \beta^A (\bar{x}_s - \bar{x}) + \sum \xi_a age_a \quad (3)$$

where  $x_{isa}$  and  $\bar{x}_s$  are, as before, the individual-level and the average census block material deprivation index,  $\bar{x}$  is the grand mean of material deprivation for the population under study,  $\beta^I$  is the individual effect and  $\beta^A$  is the aggregate (ecological) effect. In a linear effect model the aggregate effect is algebraically equivalent to the sum of the individual and the contextual effects: ( $\beta^A = \beta^I + \beta^C$ ).

The Cronbach model is computationally more efficient because the individual-level material deprivation and the census block average material deprivation are centered.

### 3.2 Ecological Models

Let assume now that data are aggregated at census block level and that  $Y_s$  is the count of observed deaths and  $\bar{x}_s$  is the mean deprivation level in the  $s$ -th census block. As before, we assume the death counts are distributed as Poisson with mean  $T_s \times \theta_s$  where  $\theta_s$  is the relative risk of the  $s$ -th census block and  $T_s$  the expected number of cases or person years of observation.

We specified a series of ecological models on the logarithm of the Relative Risk in which the response variable and the material deprivation index are alternatively age-adjusted or not ( $\bar{x}_s$  and  $\bar{x}_s^{age}$  respectively).

4) The first model is fitted on crude data, both mortality and deprivation are not age-adjusted. However, age is included in the model as the mean age by census block  $\overline{age}_s$ . In this case  $T_s$  is equal to the person years at risk in the  $s$ -th census block:

$$T_s = PY_s$$

$$\log(\theta_s^g) = \alpha + \beta^A \bar{x}_s + \xi \overline{age}_s \quad (4)$$

where  $\alpha$  is the intercept,  $\beta^A$  is the aggregate effect of material deprivation and  $\xi$  represents the linear effect of age.

5) In the second model both mortality and deprivation are adjusted for age. In this case  $T_s$  is equal to the expected number of cases given by the product of  $PY_{sa}$ , the person years of observation in  $s$ -census block and  $a$ -th age classes, and  $\xi_a^R$ , the age-specific reference rates:

$$T_s = E_s = \sum \xi_a^R \times PY_{sa}$$

$$\log(\theta_s^{age}) = \alpha + \beta^A \bar{x}_s^{age} \quad (5)$$

where  $\beta^A$  is the aggregate effect and  $\bar{x}_s^{age}$  is the age-adjusted deprivation.

6) In the third model mortality is age-adjusted while the material deprivation index is not adjusted for age:

$$\log(\theta_s^{age}) = \alpha + \beta^A \bar{x}_s \quad (6)$$

where  $\bar{x}_s$  is the crude census block average deprivation, not age-adjusted.

7) In the last model mortality is age-adjusted, the material deprivation index is not adjusted but, differently from the above model, age is now included into the regression model as a confounder:

$$\log(\theta_s^{age}) = \alpha + \beta^A \bar{x}_s + \xi \overline{age}_s \quad (7)$$

where  $\overline{age}_s$  is the age average in the  $s$ -th census block.

For all models, the material deprivation effects are reported both per unit change in deprivation and for one standard deviation change. In this way the magnitude of the regression coefficients of individual and ecological models are directly comparable.

#### 4. RESULTS

Table 1 reports the distribution of numbers of enrolled people, counts of observed and expected deaths and Standardized Mortality Ratios (SMRs) by level of material deprivation. The expected number of cases was calculated by internal indirect standardization classifying the population in 16 age classes (0-4, 5-9, ..., 75 or more) (Breslow e Day, 1975). There is a positive trend in SMR by level of material deprivation; the relative risk for a linear trend is 1.27 (0.237 on the log scale, 90% CL 0.17; 0.30) (Armitage, 1955; Cochran, 1954).

TABLE 1

*Tuscan Longitudinal Study. Number of enrolled subjects (percentage), observed and expected death counts, SMR (90% confidence limit) by category of material deprivation index. All cause mortality. Males. Florence census cohort 1991-1995*

Material deprivation index (categories)	n	%	Observed deaths	Expected deaths	SMR (CL 90%)
0	123,111	85	3,710	3,828.2	0.97 (0.94-1.00)
1	19,429	13	549	451.3	1.22 (1.13-1.31)
2	1,461	1	53	32.5	1.63 (1.28-2.05)

The logarithm of relative risks estimated by the individual models are reported in Table 2. The estimated individual effect from the pooled model (1) is 0.238 (90% CL 0.17; 0.30); it decreases to 0.166 (90% CL 0.10; 0.23) when we con-

TABLE 2

*Tuscan Longitudinal Study. Regression coefficients (standard error) of census block average material deprivation index. Results of the individual-level models (1) (2) (3) (see text). All cause mortality. Males. Florence census cohort 1991-1995*

Model	Covariate	$\beta$	$\beta / sd$
(1) <i>pooled</i>	$x_{isa}$	0.238 (0.039)	0.093 (0.015)
(2) Contextual	$x_{isa}$	0.166 (0.041)	0.065 (0.016)
	$\bar{x}_{sa}$	0.672 (0.120)	0.082 (0.015)
(3) Cronbach	$x_{isa} - \bar{x}_{sa}$	0.166 (0.041)	0.062 (0.015)
	$\bar{x}_{sa} - \bar{x}$	0.838 (0.113)	0.102 (0.014)

\* regression coefficients are log relative risk

sidered in the model the contextual effect of census block average material deprivation (model 2). The *pooled* effect  $\beta^T=0,238$  (model 1) is a biased estimate of the individual effect since it does not account for the presence of a concurrent contextual effect. In a linear effect model is easy to show that:

$$\beta^T = \beta^I + \beta^C \frac{Var(\bar{x})}{Var(x)}$$

In our case  $sd(x)=0.3897$  ;  $sd(\bar{x})=0.1219$  ;  $\frac{Var(\bar{x})}{Var(x)}=0.10$ ; and thus:

$$0.166+0.672 \times 0.10=0.233.$$

The contextual effect of material deprivation is larger than the individual effect ( $\beta^I = 0.166$ , 90% CL 0.10; 0.23, and  $\beta^C = 0.672$ , 90% CL 0.48; 0.87, respectively), even when we considered the standardized effect ( $\beta^I / sd(x) = 0.065$ , 90% CL 0.038; 0.091, and  $\beta^C / sd(\bar{x}) = 0.082$ , 90% CL 0.057; 0.107). The aggregate effect  $\beta^A$  estimated by the Cronbach model is 0.838 (90% CL 0.65; 1.02), and it is algebraically the sum of  $\beta^I$  and  $\beta^C$ . The aggregate effect  $\beta^A$  standardized by unit of standard deviations is 0.102 (90% CL 0.079; 0.125).

The results obtained fitting the ecological models are reported in Table 3. Let consider standardized coefficients by unit of standard deviations to compare the results with those obtained fitting the *pooled* and Cronbach models, respectively. Model 4 ( $b_A / sd=0,097$ , 90% CL 0.074; 0.120) and model 5 ( $b_A / sd=0.119$ , 90% CL 0.091; 0.147) reproduced the analysis conducted on individual-level data ( $\beta^A / sd=0.102$ , 90% CL 0.079; 0.125). Model 4 used crude mortality and deprivation variable and age as covariate, while model 5 used both age-adjusted mortality and deprivation. Models 6 and 7 overestimated the effect ( $\beta^A / sd = 0.143$ , 90%

CL 0.110; 0.176, and  $\beta^A / sd = 0.140$ , 90% CL 0.107; 0.173, respectively). We descriptively may compare these five estimates by a random effect meta-analysis: the heterogeneity  $I^2$  (Higgins et al., 2003) between the effect estimates was 34.4% (the variance between was  $\text{var}(\hat{\beta}) = 0.000441$  while the average within variance was  $0.000289 - \overline{\text{var}(\hat{\beta})} = 0.017$ ; then  $I^2 = (0.000441 - 0.000289) / 0.000289 = 34.4\%$ ). These heterogeneity was completely accounted by model 6 and 7, which overestimated the effect by 0.037,  $p = 0.025$  (Thompson and Sharp, 1999).

Standard errors of the estimates are rather stable through all models, but models 5, 6 and 7 show slightly more uncertainty in the estimates (table 3).

TABLE 3

*Tuscan Longitudinal Study. Regression coefficients (standard error) of census block average material deprivation index. Results of the aggregate-level models (4) (5) (6) (7) (see text). All cause mortality. Males. Florence census cohort 1991-1995*

Model	Covariate	$\beta^A$	$\beta^A / sd$
(4) crude + age as covariate	$\bar{x}_i$	0.792 (0.112)	0.097 (0.014)
(5) age-adjusted	$\bar{x}_i^{age}$	0.830 (0.116)	0.119 (0.017)
(6) age-adjusted mortality crude deprivation	$\bar{x}_i$	0.799 (0.113)	0.143 (0.020)
(7) age-adjusted mortality crude deprivation + age as covariate	$\bar{x}_i$	0.787 (0.114)	0.140 (0.020)

## 5. DISCUSSION AND CONCLUSION

We addressed the problem of mutual standardization when analyzing ecological data adjusting or not the outcome and the predictor for the same confounding variable. In particular we focused on material deprivation as predictor of all cause mortality because it is debated how to improve resource allocation formulas including the deprivation effect. Rosenbaum and Rubin (1984) discussed the problem. However they treated the linear case and a predictor defined at the subject-specific level. But, in the present context, Poisson regression models fitted to individual-level data highlighted the presence of a relevant contextual effect of material deprivation and we are interested in estimating the aggregate effect by Poisson ecological regression model.

We showed that standardized aggregate effect is unbiasedly estimated by ecological models in which either mortality and deprivation are age-adjusted or when the model on crude mortality include crude deprivation and age as predictors. These results are perfectly consistent with the findings on linear models by Rosenbaum and Rubin (1984).

Coming back to the real issue of allocation formulas, our results implies that if we want to use information on socio-economic factors and age we need either to age-standardized mortality and deprivation or regress crude mortality rates by crude deprivation and age.

Moreover, since deprivation is available at aggregate level, we are assuming the

plausibility of a contextual effect. Infact, the study of contextual effect is complicated by the identification of the appropriate level of aggregation. Census block seems to be a good compromise and the choice of a bigger geographical level (e.g. municipality) would lead to an effect dilution or an ecological bias. This suggests also to not overinterpret the association as a direct causal relation. The scope of the analysis, in the case of resources allocation, is to identify markers of health needs. A marker is associated to the outcome but not necessarily it is a determinant of it. A further complication is that deprivation indexes may assume different meanings in different geographical contexts or points in time (Dreassi *et. al.*, 2005, Catelan *et. al.*, 2006). Eventually, the association of material deprivation and mortality may be reversed and using it in a fixed allocation formula would lead to a distribution of resources not consistent with objectivity, equity and efficiency principles.

Practical limitations in the estimation of the aggregate effect associated with deprivation can be met in the availability of the data. Indeed, we do not have the joint distribution of age and deprivation at a small geographical level and therefore it is currently impossible to age-standardized the material deprivation index.

In conclusion, after adjusting for age, deprivation appeared to be a good predictor of health needs both at individual and at aggregate level (census block). The presence of a contextual effect increases the interest in using deprivation in the allocation formula, since it would permit a better distribution of resources to disadvantaged micro-areas. In the present paper, we stress the need to estimate the association between deprivation and health appropriately adjusting for age. The current common practice, in absence of individual data, to regress standardized mortality on material deprivation may be inappropriate. The bias introduced depends on the difference between adjusted and not-adjusted predictors. The alternative model that include age as predictor together with crude material deprivation may result in unbiased estimates under the restrictive assumption of absence of interaction in the scale of the linear predictor (Rosenbaum and Rubin, 1984 p. 441-442).

#### ACKNOWLEDGMENTS

The present research was funded by Ministry of Education and Scientific Research PRIN 2006131039 "Statistical Methods for Environmental Health Impact Assessment" PRIN 2008TXL73Z "Economic and statistical methods in Environmental Epidemiology". Census data were made available by the Tuscan Region to the Tuscan Longitudinal Study (SLTo).

*Department of Statistics "G. Parenti"*  
*University of Florence*  
*Biostatistics Unit*  
*ISPO Cancer Prevention and Research Institute*

LAURA GRISOTTO  
DOLORES CATELAN  
GABRIELE ACCETTA  
ANNIBALE BIGGERI

## REFERENCES

- ARMITAGE P. (1955), Test for linear trend in proportions and frequencies. *Biometrics*; **11**: 375-86.
- BIGGERI A, GORINI G, DREASSI E, KALALA N, LISI C (2001), Condizione socio-economica e mortalità in Toscana, *Studi e Ricerche*, n. 7, Edizioni Regione Toscana, Centro Stampa Giunta Regionale, Firenze.
- BRESLOW NE, DAY NE (1975), Indirect standardization and multiplicative models for rates, with reference to the age adjustment of cancer incidence and relative frequency data. *Journal of Chronic Diseases* **28**(5,6), 289-303.
- CARR-HILL RA, SHELDON TA, SMITH PC, MARTIN S, PEACOCK S, HARDMAN G (1994), Allocating resources to health authorities: development of method for small area analysis of use of inpatient services, *BMJ*, **309**: 1046-1049.
- CATELAN D, BIGGERI A, DREASSI E, LAGAZIO C (2006), Space-cohort Bayesian models in ecological studies. *Statistical Modelling*; **6**: 1-15.
- COCHRAN WG (1954), Some methods for strengthening the common chi-squared tests. *Biometrics*; **10**: 417-54.
- COSTA G, SPADEA T, CARDANO M (2004), (a cura di) Diseguaglianze di salute in Italia. *Epid Prev*, **28** (3).
- CRONBACH LJ, WEBB J (1975), Between-class and within-class effects in a reported aptitude X treatment interaction. *J. Educ. Psychology*; **67**, 6:717.
- DREASSI E, BIGGERI A, CATELAN D (2005), Space-time models with time dependent covariates for the analysis of the temporal lag between socio-economic factors and lung cancer mortality. *Statist. Med.*; **24**: 1-19.
- FIREBAUGH G (1978), A rule for inferring individual-level relationships from aggregate data. *American Sociological Review*. **43**, 557-572.
- HIGGINS JPT, THOMPSON SG, DEEKS JJ, ALTMAN DG (2003), Measuring inconsistency in meta-analyses. *British Medical Journal*, **327**: 557-560.
- JARMAN B (1983), Identification of underprivileged areas. *British Medical Journal*, **1705**-1709.
- MINISTERO DELLA SALUTE (1998), Piano Sanitario Nazionale 1998-2000. Un patto di solidarietà per la salute. Roma.
- MORGENSTERN H (2008), Ecologic Studies. In Rothman KJ, Greenland S, Lash TL *Modern Epidemiology* – third edition. Lippincott Williams & Wilkins, Philadelphia.
- ROSENBAUM P R, RUBIN D B (1984), Difficulties with regression analyses of age-adjusted rates. *Biometrics*, **40**, 2: 437-443.
- STONE M, GALBRAITH J (2006), How not fund hospital and community health services in England. *J. R. Statist. Soc. A*, **169** Part 1, 143-164.
- THOMPSON SG, SHARP SJ (1999), Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, **18**: 2693-2708.
- WAKEFIELD J (2007), Disease mapping and spatial regression with count data. *Biostatistics*, **8**, 2, 158-183.

## SUMMARY

*Material deprivation as marker of health need*

A relationship between socio-economic status and health has been widely documented both by individual-level and ecological regression studies. We addressed the problem

known in the literature as using a material deprivation index as predictor of health needs and comparing results when adjusting or not the health outcome and the deprivation index for the same confounding variables. We focus on non-linear hierarchical models. We take as example the the issue of introducing socio-economic indicators in national or regional resources allocation formulas. We fitted a series of models with different data hierarchies to evaluate both the individual effect and the aggregate (census block) effect of material deprivation on health status, disentangling the individual from the contextual effects. Individual mortality records came from the Florence census cohort 1991-1995 which is part of the Tuscan Longitudinal Study. Data on socio-economic factors derived from individual records of the 1991 census. Our results suggested that after adjusting for age, material deprivation is a good predictor of health needs both at individual and at aggregate level (census block). The presence of a contextual effect increases the interest in using deprivation in the allocation formula, since it would permit a better distribution of resources to disadvantaged micro-areas. In the present paper, we stress the need to estimate the association between deprivation and health appropriately adjusting for age. The ideal goal would be having information at small geographical level on the joint distribution of age and deprivation to age-standardize both the response and the predictor. A temporary solution should be to regress crude mortality rates on deprivation and age. The current common practice, in absence of individual data, to regress standardized mortality on material deprivation may be inappropriate.