Bringing the individual back to small-area variation studies: A multilevel analysis of all-cause mortality in Andalusia, Spain.

Merlo, Juan; Viciana-Fernández, Francisco J; Ramiro-Fariñas, Diego

Published in:
Social Science and Medicine

DOI:
10.1016/j.socscimed.2012.06.004

2012

Bringing the individual back to small-area variation studies: a multilevel analysis of all-cause mortality in Andalusia, Spain

Juan Merlo[a,b] (Corresponding author)

Francisco J. Viciana-Fernández[c]

Diego Ramiro-Fariñas[d]

For the Research Group of the Longitudinal Database of the Andalusian Population (LDAP)

[a] Unit for Social Epidemiology, CRC, Faculty of Medicine, Lund University, Malmö, Sweden
[b] Centre for Economic Demography at Lund University, Malmö, Sweden
juan.merlo@med.lu.se

[c] Demographic and Social Statistics, Institute of Statistics and Cartography of Andalusia, Seville, Spain
franciscoj.viciana@juntadeandalucia.es

[d] Institute of Economics, Geography and Demography, Center for Humanities and Social Sciences, Spanish National Research Council. Madrid, Spain
diego.ramiro@cchs.csic.es

Corresponding author
Juan Merlo
MD, PhD, Tenured Professor
Unit for Social Epidemiology, CRC, Faculty of Medicine,
Skåne University Hospital
Jan Waldenströms street 35. SE-20502 Malmö, Sweden
Tel. +46 40391329/+46 70-5092265
Juan.merlo@med.lu.se

LUNDS UNIVERSITET

ABSTRACT

We perform a multilevel analysis (individuals, households, census tracts, municipalities, and provinces) on a 10% sample (N = 230,978) from the Longitudinal Database of the Andalusian Population. We aimed to investigate place effects on 8-year individual mortality risk. Moreover measures of association (i.e., odds ratios, OR) between area socioeconomic circumstances and individual risk, we estimated measures of variance and clustering like the variance partition coefficient (VPC).

We explicitly proclaim the relevance of considering *general contextual effects* (that is, in which degree the context, as whole, conditions individual variance in mortality risk) in at least two circumstances. The first one concerns the interpretation of *specific contextual effects* (that is, the association between a particular area characteristic and individual risk) obtained from multilevel regression analyses. The second situation involves the interpretation of geographical variance obtained from classical ecological/spatial analyses. Besides the recognized ecological fallacy, the lack of individual level information leaves geographical variance unrelated to the total individual variation and, therefore, difficult to interpret. Finally we stress the importance of considering the familial household in multilevel analysis.

We observed an association between percentage of people with low educational achievement in the census tract and individual mortality risk (OR highest vs. lowest quintile = 1.14, 95% confidence interval 1.08−1.20). However, only a minor proportion of the total individual variance in the probability of dying was at the municipality (M) and census tract (CT) levels (VPC_M = 0.2%, and VPC_{CT} = 0.3%, respectively). Conversely, the household (H) level appeared much more relevant (i.e. VPC_H = 18.6%) than the administrative geographical areas.

Without considering *general contextual effects,* both multilevel analyses of *specific contextual effects* and ecological studies of small-area variation may provide a misleading picture that overstates the role of administrative areas as contextual determinants of individual differences in mortality.

Keywords: Mortality, Multilevel analysis, Ecological analysis, Small-area analysis, Social environment, Socioeconomic factors

Individual health is not only an individual responsibility; it also depends on the societal context in which the individual resides. Since many social processes take place over space (Cummins et al., 2007; Kaplan, 1999; Macintyre S & Elleway A, 2000; Macintyre et al., 2002; Merlo J, 2011), it is a fundamental issue in public health to identify the social and geographical environments that condition individual health variance. Moreover, is necessary to ascertain the specific characteristics of these contexts that explain such variance and are associated individual disease risk (Cummins et al., 2007; Merlo et al., 2009).

It is today well established that multilevel modelling (Bingenheimer & Raudenbush, 2004; Duncan et al., 1998; Goldstein, 2003; Merlo et al., 2006c; Merlo et al., 2005a, c; Merlo et al., 2005d; Subramanian et al., 2003) is a worthy instrument for the quantitative analyses of place effects on individual health. However, many multilevel analyses performed so far have been mainly focused on the study of associations between contextual variables and individual health, considering the analysis of variance as secondary information (Blakely & Woodward, 2000; Diez Roux, 2008). In contrast, others scholars have explicitly concluded that the analysis of variance provides indispensable information for understanding place effects on health (Boyle MH & Willms JD, 1999; Clarke P & Wheaton B, 2007; Duncan et al., 1993; Merlo, 2003; Merlo et al., 2004; Merlo et al., 2009; Riva et al., 2007)

Moreover, all around the world, a persistent amount of observational information on place effects is still being obtained from ecological/spatial studies of “small-area variations” , frequently in the form of coloured atlases and disease maps (Benach et al., 2003; Benach et al., 2004; Borrell et al., 2010; Collaboration, 2010; MacNab & Dean, 2002; Middleton et al., 2008; Ocana-Riola & Mayoral-Cortes, 2010; Ocana-Riola et al., 2008a; Pickle et al., 1999; Shaw, 2008; Turrell & Mengersen, 2000). From an empirical perspective, the advantages of multilevel versus ecological regression analyses were clearly identified by the seminal work performed by Aitkins and Longford (Aitkin M & Longford N, 1986) as well as by Jones, Duncan, Moon, Subramanian and colleagues (Bullen et al., 1996; Duncan et al., 1993, 1995, 1996, 1998, 1999; Jones et al., 1991; Subramanian et al., 2009; Twigg et al., 2000). We also contributed to this discussion in a previous publication (Merlo et al., 2001).

Beyond the conventional study of *specific contextual effects* (i.e., measures of association between particular contextual variables and individual outcomes), the use of multilevel regression and analogous techniques of analyses (Katz et al., 1993) has allowed us to identify *general contextual effects* that are based on measures of variance and clustering rather than on measures of association (Merlo, 2003; Merlo et al., 2009; Merlo et al., 2001; Petronis & Anthony, 2003; Subramanian SV et al., 2007). However, while the explicit distinction between these two types of effects enriches our understanding on contextual influences on individual health, this approach is still rather infrequent in social epidemiology (Merlo et al., 2009).

General contextual effects informs on which extend the geographical constructs we use for defining a context (e.g., neighborhoods, small areas, census tracts) condition individual outcomes (e.g., mortality) without specifying any other contextual characteristic than the very boundaries we used for defining the context (Merlo et al., 2005a). That is, if the geographical administrative boundaries actually capture a relevant context that conditions individual health, one should expect not only a statistically significant spatial variation (as frequently detected in "small- area variation studies") (Ibanez et al., 2009), but also that this spatial variation conditions a meaningful proportion of the total individual-level variation (Boyle MH & Willms JD, 1999; Merlo, 2003; Merlo et al., 2004; Merlo et al., 2009; Subramanian SV et al., 2007). Pioneer in interpreting general contextual effects was also the work by Duncan, Jones, and Moon referred above (Duncan et al., 1993).

In spite of the fundamental relevance of general contextual effects when investigating socio-geographical influences on individual health, only a relatively small part of the multilevel analyses published until today have reported measures of variance (Riva et al., 2007), and still fewer of them explicitly discuss general and specific contextual effects within the same discourse (Merlo et al., 2009).

Regarding ecological studies of small-area variations the concerns are not only against the legendary "ecological fallacy"(Morgenstern, 1998; Robinson, 2009) and the advantages of the multilevel analysis for preventing it (Merlo et al., 2001; Subramanian et al., 2009), but rather on the incapacity of ecological analyses for quantifying general contextual effects because they lack information on how variance is partitioned across the different levels of analysis (e.g., individual and areas) (Merlo, 2003; Merlo et al., 2009; Subramanian SV et al., 2007).

Besides, while the analysis of place effects on health is normally based on geographical areas defined by administrative boundaries (e.g. census tracts, municipalities), other—maybe more relevant—contexts have received less consideration. One of these contexts is the familial household (Lawlor DA et al., 2009; Merlo, 2010; Merlo et al., 2006a; Yang et al., 2009). The correlation of individual health within family/household context is much larger than among the residents of places defined by administrative boundaries (Boyle MH & Willms JD, 1999; Merlo et al., 2009; Yang et al., 2009). Therefore, the family/household level should, be considered in the analyses.

Against the above-described background, we perform a multilevel analysis to investigate place effects on individual mortality risk. We explicitly considered general contextual effects (i.e., in which degree the provinces, municipalities, census tracts, and familial households condition individual variance in mortality risk) when interpreting specific contextual effects (i.e., the association between small area socioeconomic circumstances and individual mortality risk). In addition, when performing the multilevel analyses, we explicitly question the appropriateness ecological/spatial analyses of small area variations in mortality. These ecological studies are still very frequent not only in Spain, but also all over the world and might provide unsuitable information on place effects on individual health.

We perform our analyses on a representative sample of the Andalusian population who participated in the Spanish Population and Housing Census 2001 ("Censos de Población y Viviendas 2001,"), and that is included on the recently created Longitudinal Database of the Andalusian Population (LDAP) (Viciana F et al., 2010). As far we known, our study is the first multilevel analysis investigating mortality on a large population based cohort from south Europe.

POPULATION AND METHODS

Andalusia is located in the southernmost part of the European Union, and it is the most populous (8,370,975 inhabitants in 2010) of the 17 autonomous communities of Spain. Andalusia has one of the lowest per capita incomes in the nation, and it also accounts for the highest mortality rate.

The Andalusian Institute of Statistics, in collaboration with the Spanish National Institute of Statistics and the Spanish National Research Council, has created the LDAP research database (Viciana F et al., 2010). The LDAP records information on all the individuals residing in Andalusia according to the National Population Register. The database also tracks the vital status of all the Andalusian residents, including those who die outside the Andalusian territory.

According to the law governing statistical planning approved by the Andalusia Parliament, the Andalusian Institute of Statistics linked the records of migration and mortality to the

Spanish Population and Housing Census 2001, using an internal numeric key of the LDAP. For the purpose of our investigation, the Andalusian Institute of Statistics drew a 10% random sample of all the Andalusian dwellings that were recorded in the 2001 census. From the original 244,972 individuals aged 45 to 79 years identified in the 10% census sample of dwellings, we analysed 94% (230,978/244,972) unique individuals with correct identification numbers, who were born in Spain and had information on vital status (i.e. dead or alive) at the end of the follow-up by 31 December 2009. These individuals were living in 142,516 households, within 5,381 census tracts in 770 municipalities comprising the eight provinces of Andalusia (i.e. Almeria, Cadiz, Cordoba, Granada, Huelva, Jaen, Malaga, and Seville).

Assessment of individual variables

The outcome of this study is individual, all-cause mortality, assessed from the time of the 2001 census until 31 December 2009.

In the analysis we compared men with women, and included age in 5-year groups (i.e. 45–49 … 75–79).

We classified the *occupation* of the individuals in 11 categories, as follows: (i) working, (ii) unemployed, (iii) receiving some kind of formal education, (iv) being on permanent disability pension, (v) collecting widow's or orphan's pension, (vi) collecting pension for retirement or early retirement, (vii) needing help with basic activities, (viii) performing household tasks, and (ix) other occupations. We used the working category (i) as a reference group in the comparisons.

We considered the *civil status* of the individuals as single, married, widowed, separated, or divorced. We considered the married category as reference in the comparisons.

We classified *educational achievement* into four categories, very low (analphabetic, i.e., less than one year of formal education), low (assistance to elementary school, i.e., five to 9 years of formal education), medium (elementary baccalaureate or similar degree, i.e., nine to 10 years of formal education) and high (superior baccalaureate or higher educational achievement, i.e., 12 years or more of formal education).

The census 2001 contained an item indicating the municipality where the individual resided 10 years before the time when the census was completed. We contrasted this municipality with the municipality of residence at the moment of the census, and compared those who moved (i.e. "*movers*") with those who stayed in the same municipality (i.e. "*stayers*").

Assessment of household variables

Concerning *housing tenure*, we distinguished whether the individuals (i) owned or (ii) rented their homes, and (iii) other forms of housing tenure, such as by donation at a low price. We compared those who owned their homes with the other categories.

We identified the individuals who owned a *second dwelling* beyond the principal one, and compared this group of people to those who had only one dwelling.

Predicted probability of death

We performed individual-level logistic regression analyses separately in women and in men to obtain the predicted probability (i.e., risk score) of death as a function of the individual and household characteristics described above (see also the Table 2). Thereafter, we categorised this predicted probability in ten groups by deciles, and used the group with the lowest probability as reference in the comparisons. We used this variable to adjust for individual and household characteristics in the multilevel regression analyses. We refer to the work by Arbogast et al (Arbogast et al., 2008) for an extended explanation on the risk score (i.e., predicted probability) methodology for confounding adjustment.

Assessment of contextual characteristics

We assumed that the *percentage of people with a very low/low educational achievement* was a specific contextual characteristic that reflected the socioeconomic and material circumstances in the census tract. We performed a direct age standardisation using the method of the equivalent average rate (Yule GU, 1934) on 5-year age groups. Thereafter, we categorised the census tracts according to the age-standardised educational variable in four groups by quartiles, and used the group with the lowest percentage of people with very low/low educational achievement as reference in the comparisons.

Multilevel regression analyses

The data presented a hierarchical structure consisting of individual (level 1), nested within households (level 2), nested within census tracts (level 3), and census tracts, in turn, nested within municipalities (level 4), which were embraced by the eight provinces of Andalusia (level 5). Therefore, we performed multilevel logistic regression analyses (Goldstein, 2003; Snijders & Bokser, 1999).

The multilevel regression analysis offered two main advantages. On the one hand, it accounted for the possible correlation of the individual-level information within households, the correlation of households within census tracts, and the spatial correlation of census tracts within municipalities, and municipalities with provinces. Accounting for this correlation is necessary for obtaining correct statistical estimations of uncertainty (i.e., standard errors), and also substantive information on the distribution of the individual variance across those levels (Merlo et al., 2005a).

a) Models

The first model was a random intercept model that only included random terms for the household, census tract, and municipality levels. This model provided information only on the way individual variance in the probability of dying was distributed across the different levels of analysis.

The second model added the gender-stratified, individual, predicted probability of death; the gender residual effect; and the residential mobility variable, to adjust for differences in the individual composition of the households and areas.

The third model added the percentage of people with very low/low educational achievement in the census tracts.

Finally, in the fourth model, we included the eight Andalusian provinces (i.e. Almeria, Cadiz, Cordoba, Granada, Huelva, Jaen, Malaga, Seville) as a fixed effect, and considered Seville as the reference category.

We also did age- and sex-adjusted multilevel logistic regression analyses with the individual nested within census tracts and modelled separately mortality, as well as low/very low educational achievement. The purpose of these analyses was to compare geographical (i.e. census tracts) differences in socioeconomic characteristics (i.e. educational achievement) with geographical differences in mortality.

When interpreting the multilevel regression analyses, we distinguished between specific and general contextual effects.

c) Specific contextual effects

Specific contextual effects provide information about the existence of an association between concrete area characteristics (i.e. area low educational achievement) and individual outcome (i.e. mortality). We estimated these effects using regression coefficients expressed as odds ratios (OR) 95% confidence intervals (CI), as we also did for estimating individual-level effects.

d) General contextual effects

General contextual effects give information on which degree the areas/households under investigation condition individual differences in mortality risk. For their assessment, we do not specify any contextual characteristic other than the very boundaries that delimit the level of analysis (e.g. census tract, household) (Merlo et al., 2005b; Merlo et al., 2009; Subramanian SV et al., 2007). General contextual effects are estimates by measures of variance and clustering. For this purpose we calculated two different measures.

The variance partition coefficient (VPC)

We calculated the VPC at the municipality ($_M$), at the census tract ($_C$), and at the household ($_H$) level, according to the latent variable method (W. J. Browne et al., 2005; Goldstein et al., 2002; Li J et al., 2008; Merlo et al., 2006b; Snijders & Bokser, 1999), as follows:

$$VPC_M = \sigma^2_M / (\sigma^2_M + \sigma^2_C + \sigma^2_H + \pi^2/3)$$

$$VPC_C = (\sigma^2_M + \sigma^2_C) / (\sigma^2_M + \sigma^2_C + \sigma^2_H + \pi^2/3)$$

$$VPC_H = (\sigma^2_M + \sigma^2_C + \sigma^2_H) / (\sigma^2_M + \sigma^2_C + \sigma^2_H + \pi^2/3)$$

where σ^2 represents the variance at the specific level. The value of the variance of the underlying individual-level variable according to the logistic distribution is $\pi^2/3$ or 3.29. In our study the VPC provides information on the percentage of the total individual variance in the probability of dying that existed at a concrete level. In our case, the VPC can also be interpreted as the correlation in the probability of death between two individuals randomly selected from the same municipality (VPC_M), from the same census tract and municipality (VPC_C), or from the same household from the same census tract and municipality (VPC_H). If the level did not condition the individual probability of death, the VPC would be close to zero, and there would not be general contextual effects.

The median odds ratio (MOR)

The MOR (Larsen & Merlo, 2005; Larsen et al., 2000; Merlo et al., 2006b) is a measure of heterogeneity, rather than of clustering, as the VPC is. The MOR is an alternative way of expressing the area/household variation that translates this variance in the widely used odds ratio scale, which makes the MOR comparable with the ORs of individual or area variables. If there is no variation, the MOR = 1.

The MOR is defined as the median value of the distribution of odds ratios obtained when randomly picking out two individuals from different areas and comparing the individual from the highest risk area with the individual from the lowest risk area. In simple terms, the MOR could be interpreted as the increased (median) odds of dying if the individual were living in another area with higher risk. We compute the MOR as follows:

$$MOR_M \approx \exp\left(0.95^* \sqrt{\sigma^2_M}\right)$$

$$MOR_C \approx \exp\left(0.95^* \sqrt{(\sigma^2_M + \sigma^2_C)}\right)$$

$$MOR_H \approx \exp\left(0.95^* \sqrt{(\sigma^2_M + \sigma^2_C + \sigma^2_H)}\right)$$

where $\sqrt{}$ is the square root of the variance (σ^2) at the specific level.

e) Estimations

Starting with restricted iterative generalised least squares (RIGLS) estimations (Goldstein, 2003), we applied Markov chain Monte Carlo (MCMC) methods (W.J. Browne, 2009a) with parameter expansion at the household level (W.J. Browne, 2009b), a burning in length of 5,000, a monitoring chain length of 50,000, and a thinning of 10. More technical and conceptual information on these methods is available in the work by Browne (Browne WJ, 2003).

We obtained the median, 2.5%, and 97.5% values of the posterior distribution to calculate the point estimations of the parameters and their 95% credible intervals (Browne WJ, 2003). The reader unfamiliar with the 95% credible interval can interpret the credible interval in the same way as in the case of usual confidence intervals. .

We compared models using the Bayesian deviance information criterion (BDIC), and considered a reduction of the BDIC greater than 10 as an indication of a better fit (Spiegelhalter et al., 2002).

We performed the analyses using SPSS version 18 and MLwiN version 2.22

RESULTS

Characteristics of the population

Table 1 shows the characteristics of the population of men and women residing in Andalusia at the time of the 2001 census by quartiles of percentage of people with very low/low educational achievement at the census tract. As this percentage increased, so did the crude all-cause mortality, the mean value of the predicted probability of death, and most variables related to socioeconomic deprivation and impaired health (i.e. unemployment, disability pension, retirement, receiving widow's/orphan's pension, and having only one dwelling).

Mortality was always higher in men than in women, but more men than women had a higher educational achievement and were working in 2001. The percentage of "movers" decreased from 9% to 3%, as the percentage of people with very low/low educational achievement at the census tract increased. Compared with "stayers", "movers" had a lower crude mortality (OR =

0.73, 95% CI: 0.69–0.78) and an increased probability of residing in the quartile with the lowest percentage of people with very low/low educational achievement at the census tract (OR = 1.99, 95% CI: 1.92–22.07) (Not shown in tables).

Association between the individual- and household-level variables and mortality

Table 2 shows the association between the individual- and household-level variables used for the calculation of the predicted probability of death. These analyses replicated the well-known associations between, on the one hand, low socioeconomic position and living alone, and on the other, increased mortality. Some occupational categories directly related to impaired health, like needing help with basic activities or being on disability pension, were clearly associated with an increased mortality risk.

Specific contextual effects

Table 4 shows that the census tract *percentage of people with a very low/low educational achievement* variable was associated with a slightly increased mortality risk. Moreover, comparing with Seville, we observed a somewhat higher average mortality in Huelva and Cadiz, and to some extent, lower mortality in Jaen and Cordoba. We also observed that the association between individual and household characteristics (summarised by categories of predicted probability of death) and mortality was clearly stronger than the corresponding association concerning the contextual variable.

General contextual effects

Table 3 shows the general contextual effects of the different levels of analysis. We can see that a considerable proportion of the total individual-level variance in the probability of dying was at the household level. However, this proportion was very small at the municipality and census tract area levels. In the adjusted analyses, when randomly picking up two individuals from the same geographical unit, the correlation in their probability of dying was very low at both the municipality (i.e. VPC_M = 0.2%) and the census tract level (i.e. PCV_C = 0.3%).

Using a measure of heterogeneity like the MOR, rather than a measure of clustering like the VPC, we found that if an individual moved to another municipality, his/her risk of dying would marginally increase (Model 2 MOR_M = 1.09). We observed similar results for the census tract level (Model 2 MOR_C = 1.11). These low values contrasted with the high MOR values obtained at the household level (Model 2 MOR_H = 1.72).

Adjustment in Model 2 for the categories of the predicted probability of death had an impressive influence on model fit, reducing the BDIC by 25,746 units. This procedure also explained 53% [(0.660–0.310)/0.660] of the variation at the household level, as well as 87% [(0.076-0.010)/0.076] of the census tract and 47% [(0.015-0.008)/0.015] of the municipality variances. However, these last two components of variance were originally very small. After adjustments, the VPCs of the municipality and census tract levels became negligible.

The inclusion of the census tract educational variable in Model 3 and of the province variable in Model 4 improved the model fit, and it reallocated the geographical variance between municipalities and census tracts. However, these results needs be interpreted within the very small size of the geographical variation.

Figure 1 shows the geographical differences in mortality risk and in individual low educational achievement obtained from multilevel logistic regression models with only two levels (i.e., individuals within census tracts). These differences are expressed as odds ratios (ORs) comparing each census tract with the overall mean in the population. The ORs are in the logarithmic (log.) scale so an OR=1 (i.e., no difference) correspond with the log OR = 0. In this analysis the $VPC_{census\ tract}$ for mortality is equal to 1%, which means that the correlation in the probability of death between two individuals randomly selected from the same census tract is only 1%. However, the geographical clustering was much higher when modelling individual low educational achievement ($VPC_{census\ tract}$ = 24%) than for mortality.

Figure 2 is similar to the figure 1 but it illustrates the crude (Model 1) and adjusted (Model 2) geographical differences in mortality risk between census tracts obtained from the full multilevel logistic regression analysis with individuals, census tracts, municipalities and provinces. It shows the geographical differences in mortality between the census tracts embraced by the provinces with the highest (i.e. Huelva) and the lowest mortality (Cordoba) overlaps considerably each other.

DISCUSSION

General contextual effects: on geographical variation

According to the analysis of general contextual effects, in Andalusia, the municipalities and the census tracts appeared to have a minor relevance for understanding individual inequalities in mortality risk. Our conclusion critically questions previous ecological/spatial studies of small-area variation published hitherto in Spain (Benach et al., 2003; Benach et al., 2004; Borrell et al., 2010; Collaboration, 2010; Ocana-Riola & Mayoral-Cortes, 2010; Ocana-Riola et al., 2008a). These ecological/spatial studies of small-area variation are only justified if we implicitly assume that the geographical variation between municipalities and between census tracts represents an important factor for understanding individual inequalities in mortality risk. In other words, if there is a considerable general contextual effect. Our results, however, do not support this assumption. By analogy, our results could be generalized to many other parts of the world where ecological/spatial studies of small-area variation are used to evaluate geographical effects.

General contextual effects are based on measures of variance, and they show the possible influence of a context on individual mortality risk, without specifying any contextual characteristics other than the very definition of boundary that embraces the context (Merlo et al., 2009). Geographical variance is also analysed in customary ecological/spatial studies of "small-area variation". However, ecological/spatial studies basically aim to determine whether variation between the areas is higher than would be expected by chance (Ibanez et al., 2009), while the multilevel analysis relates the geographical/contextual variation to the total individual-level variation, as we do in our analyses.

Applying multilevel regression analysis, we observed that the crude spatial variation at the municipalities and census tracts levels represented only 0.4% (i.e. VPC_M = 0.4%) and 2.3% (VPC_M = 2.3%), respectively, of the total individual variance in the probability of death. Moreover, these minor percentages practically disappeared (i.e. VPC_M = 0.2% and VPC_M = 0.3%), when taking into account the individual composition of the areas (i.e. gender, residential mobility, and the categories of predicted probability of death). In fact, the inclusion of individual and household level information had a strong influence on the goodness of the fit of the model, as the BDIC decreased by 27,765 units.

Interestingly, using the same analytical approach as in the analysis of mortality, we detected (Figure 1) a very high geographical clustering of individuals with low educational achievement (Model 1 VPC_C = 24%) within census tracts, which contrasts with the minor geographical clustering of mortality.

On the other hand, we found that a considerable proportion of the total individual variance in the probability of dying was at the household level (i.e. VPC_H = 18.6%). This intra-household correlation possibly reflected the existence of shared genetic and environmental factors (Lawlor DA et al., 2009), as well as positive assortative mating (Epstein & Guttman, 1984). The existence of a similar genetic background, as well as learned behaviours like eating and drinking habits, attitudes to psychical activity, and coping mechanisms, may condition a comparable probability of dying between members of the same household. In fact, the household variance was considerably reduced after adjustment for the categories of predicted probability of death that included both individual- and household-level characteristics. Furthermore, it is known that the death of one partner increases the risk of death of the other, which might explain part of the household clustering of mortality (Mineau et al., 2002).

While the familial household is a naturally defined context that conditions individual mortality risk (Lawlor DA et al., 2009), the geographical boundaries of the municipalities and census tracts are defined by administrative criteria. Therefore, their relevance as determinant of individual health is not as obvious as in the case of the familial household. Even so, municipalities and census tracts are often taken for granted, without questioning their appropriateness for delimiting the real context that conditions mortality (Boyle MH & Willms JD, 1999). It is just this aspect that we have tried to assess in the present study, and we found that the municipalities and the census tracts appear to play a minor role for understanding inequalities in mortality risk. There are several reasons supporting our findings.

Mortality is influenced by an array of circumstances across the life course, but in the present study, we investigate a population of adult and elderly people and the municipalities and the census tracts where they resided in 2001. Possibly, geographical/contextual influences in childhood are more relevant than in adulthood (Lawlor DA et al., 2009; Lynch & Smith, 2005). Besides, human beings do not belong to one simple context, but to a changing spatiotemporal, relational (Cummins et al., 2007), multiple membership and cross-classified mixture of contexts (Browne WJ et al., 2001). Obviously, we cannot expect that this complex system of intricate boundaries is properly captured by simple geo-administrative criteria. Moreover the familial household, other contexts, like the work place where people spend most of their active time, may have a higher impact on mortality that the current area of residence

(Muntaner et al., 2011). Possibly, a multiple membership multiple classification (MMMC) analysis (Browne WJ et al., 2001) that considers both residential areas and work places across the life course would be a more appropriate analytical design. However, the information required to perform MMMC analyses is difficult to obtain, and the few studies with such approach performed so far in Scandinavia have found similar results as in Andalusia (i.e. minor clustering of mortality within administratively defined small areas) (Naess et al., 2008; Ohlsson & Merlo, 2011).

In the early 20th century, it was common for a person to be born, reside, work, marry, and die within the same neighbourhood. In the 21st century, however, we sleep in one place and commute to work in another. Today, our habits and lifestyles may depend much more on values transmitted by global communication networks than by the influence of our neighbourhoods. Therefore, the low clustering of mortality risk at the administrative geographical areas is not surprising.

Many multilevel analyses of neighborhood effects performed so far have considered variance as a nuisance or as a measure that only provides secondary information (Diez Roux, 2008). The reasons for this skepticism are not clear. It is frequently argued that there are technical difficulties in the measurement of the variance in multilevel logistic regression which prevent the substantive interpretation of measures of variance (Blakely TA & Subramanian, 2006). Nevertheless, this technical argument is no longer justified as, today, components of variance and clustering can be confidently calculated in generalized linear multilevel models (W. J. Browne et al., 2005; Goldstein et al., 2002; Larsen & Merlo, 2005; Li J et al., 2008).

Another argument could be the fact that it is possible to find statistically significant specific contextual effects (i.e., associations between contextual variables and the individual outcome) with very small general contextual effects (e.g., intra-area correlation of around 1%). This apparently counterintuitive situation is rather common in multilevel analyses and, if properly interpreted, provides enhance information on contextual effects (see elsewhere for a longer commentary on this aspect) (Merlo et al., 2009). However, the existence of minor *general* contextual effects might originate publication bias (Siddiqi, 2011). In other words, authors only publish or comment *specific* contextual effects because they are statistically significant and in harmony with preconceived believes on contextual effects. Publication bias can give a false impression about the influence of administrative geographical boundaries on individual health and mislead decision makers in public health.

Specific contextual effects: on the association between contextual characteristics and mortality

In this section we discuss the associations found between individual mortality risk and specific contextual characteristics like living in a census tract with a high percentage of people with a low educational achievement.

We emphasise that specific contextual effects need to be interpreted side by side with general contextual effects (Merlo, 2003; Merlo et al., 2009). In fact, without knowledge of the general contextual effects, the interpretation of area-level specific contextual effects becomes "decontextualized'' (Clarke P & Wheaton B, 2007).

We detected a small but conclusive association between a high percentage of people with a low educational achievement at the census tract and individual mortality risk ($OR_{\text{lowest vs. highest quartile}}$ = 1.14, 95% CI: 1.08–1.20). A customary interpretation of this result would be that, over and above the individual characteristics studied, improving the educational achievement in the most deprived census tracts would reduce mortality risk for all individuals exposed. However, the overall geographical clustering of mortality was very low (VPC_C = 0.3%), and the census tract educational variable did not explain so much of the—initially minor—geographical variation. If we compared one individual from the highest with another individual from the lowest quartile of census tract educational variable, we would not be able to predict which of them would have the higher mortality risk, even if, on average, the individuals in the highest quartile had a slightly higher mortality than those in the lowest.

Additionally, we need to bear in mind that aggregate measures of socioeconomic position (like the educational variable used in our study) raise methodological worries regarding the ability of these variables to actually estimate contextual effects “independently” of the corresponding individual-level variable (Mujahid et al., 2007). The small association observed between the aggregated educational variable and mortality might just be reflecting the individual-level association.

More elaborated contextual variables like composite index of area deprivation (Carstairs, 1995; Ocana-Riola et al., 2008b), or the application of the *ecometric* methodology (Chaix et al., 2008; Fone et al., 2006; Messer, 2007; Mujahid et al., 2007; Raudenbush SW & Sampson RJ, 1999) appear conceptually to be more appropriate strategies than the simple aggregate measures of socioeconomic position that we used. However, simple measures of deprivation are highly correlated with more complicated ones, and therefore, yield similar information (Bingenheimer & Raudenbush, 2004; Folwell, 1995). We do not think the results and conclusion of our study would be substantially changed by using more elaborate contextual variables, instead of the percentage of people with low educational achievement at the census tract.

Independently of the individual, household, and census tract variables considered in the multilevel regression analysis, we found that compared with Seville, the provinces of Huelva and Cadiz had a somewhat higher (OR = 1.10), and the provinces of Jaen (OR = 0.93) and Cordoba (OR = 0.86), a slightly lower average mortality (Table 4, Model 4). Among others, a possible explanation for these mortality differences could be that the provincial strategies for public health and healthcare have different effectiveness, since in Andalusia the provinces are also the independent healthcare areas. However, when interpreting specific contextual effects (i.e. the OR of mortality in Huelva compared with Seville) it is also necessary to consider the size of the general contextual effects. Since the number of provinces was low, we included the provincial level as a fixed, rather than as a random, effect in the multilevel regression analysis. Inclusion of the province variable decreased the BDIC by 31 units (Table 3), but did not have a major effect on the initially small geographical variance. Therefore, since the underlying geographical clustering within municipalities and census tracts was very low, if we, for example, compare an individual from Seville with an individual from Huelva, we will not be able to predict which of them will have the higher mortality risk, even if, on average, Huelva has a slightly higher mortality than Seville. Figure 2 illustrates, in part, this idea, as it shows that the mortality of the census tracts within the province with the highest (i.e. Huelva) and the province with lowest mortality (Cordoba) overlap considerably with each other.

A last caveat concerns the investigation of causality using quantitative observational analyses. While this subject is a challenge in all observational epidemiology (Hernan & Robins, 2006) the analysis of contextual causal effects on health presents specific difficulties (Merlo & Chaix, 2006; Oakes JM, 2003)

CONCLUSIONS

Our analyses demonstrate that the interpretation of specific and general contextual effects in multilevel analyses need to be be performed simultaneously. Otherwise specific contextual effects may lead to misleading conclusions. We need to consider both associations between means and the heterogeneity around these average measures (Braumoeller B, 2006; Downs GW & Roche DM, 1979; Gould SJ, 1996; Merlo, 2003; Merlo et al., 2009)

Our study also gives further evidence concerning the unsuitability of ecological/spatial analyses for investigating contextual effects (Aitkin M & Longford N, 1986; Bullen et al., 1996; Duncan et al., 1993, 1995, 1996, 1998, 1999; Jones et al., 1991; Merlo et al., 2001; Subramanian et al., 2009; Twigg et al., 2000). As expressed by Morgenstern, "Several epidemiologists have recently called for a greater emphasis on understanding differences in health status between populations—a return to a public health orientation in contrast to the individual (reductionist) orientation of modern epidemiology…. This recommendation … cannot be met by conducting ecologic studies; multiple level of measurement and analysis are needed" (Morgenstern, 1998). Our investigation, however, expands this critique beyond the "ecological fallacy"(Morgenstern, 1998; Robinson, 2009; Subramanian et al., 2009), and emphasizes the incapacity of ecological analyses of quantifying general contextual effects because they lack information on how variance is partitioned across the different levels of analysis (e.g., individual and areas) (Merlo, 2003; Merlo et al., 2009; Subramanian SV et al., 2007).

Modern ecological analyses of small-area variation applied state-of-the-art spatial analyses and Bayesian estimation methods to model variation in mortality between small geographical areas. However, these ecological studies have several interpretative limitations. First, it is not enough to consider whether spatial variation is higher than would be expected by chance (Ibanez et al., 2009) . It is necessary to quantify how much of the total individual variation is at the area level (W. J. Browne et al., 2005; Clarke P & Wheaton B, 2007; Larsen & Merlo, 2005; Merlo, 2003; Merlo et al., 2009), as we have done in the present study by calculating the VPCs. Furthermore, because of technical difficulties, ecological/spatial analyses only consider a few individual characteristics (e.g. age and sex), and therefore, the observed geographical variation may reflect only residual differences in the individual composition of the areas, as we have demonstrated in our analysis.

Ecological/spatial analyses give the impression that inequalities in health are conditioned by geographical/contextual factors at the small-area level. Our multilevel analysis challenges this belief and suggests that administrative geographical boundaries seem inappropriate for embracing the relevant contexts that influence mortality. However, other geographic or cultural contexts may be important.

Against this background, performing spatial analyses to model geographical variation in mortality without knowledge of the size of the general contextual effects seems meaningless, and it can even give misleading information. Our conclusion concerns all-cause mortality in the municipalities and census tracts of Andalusia. Nevertheless, our reasoning may be pertinent for many other small-area variation studies performed around the world.

Individual health is not only an individual responsibility, but also depends on the social contexts that affect the individual (Cummins et al., 2007; Kaplan, 1999; Krieger, 2001; Macintyre et al., 2002). Therefore, our results need be properly interpreted. From the perspective of public health policy, we think that, instead of blaming concrete administrative geographical areas for their mean mortality, a collective system that increases social and economic opportunities all over Andalusia will decrease individual mortality risk, and simultaneously, the current socioeconomic geographical segregation.

REFERENCES

Aitkin M, & Longford N (1986). Statistical modelling issues in school effectiveness. *J R Statist Soc A,* 149, 1-43.

Arbogast, P.G., Kaltenbach, L., Ding, H., & Ray, W.A. (2008). Adjustment for multiple cardiovascular risk factors using a summary risk score. *Epidemiology,* 19, 30-37.

Benach, J., Yasui, Y., Borrell, C., Rosa, E., Pasarin, M.I., Benach, N., et al. (2003). Examining geographic patterns of mortality: the atlas of mortality in small areas in Spain (1987-1995). *Eur J Public Health,* 13, 115-123.

Benach, J., Yasui, Y., Martinez, J.M., Borrell, C., Pasarin, M.I., & Daponte, A. (2004). The geography of the highest mortality areas in Spain: a striking cluster in the southwestern region of the country. *Occup Environ Med,* 61, 280-281.

Bingenheimer, J.B., & Raudenbush, S.W. (2004). Statistical and substantive inferences in public health: issues in the application of multilevel models. *Annu Rev Public Health,* 25, 53-77.

Blakely TA, & Subramanian, S.V. (2006). Multilevel studies In J.M. Oakes, & J.S. Kaufman (Eds.), *Methods in social epidemiology*. San Francisco, CA: Jossey-Bass.

Blakely, T.A., & Woodward, A.J. (2000). Ecological effects in multi-level studies. *J.Epidemiol.Community Health,* 54, 367-374.

Borrell, C., Mari-Dell'olmo, M., Serral, G., Martinez-Beneito, M., & Gotsens, M. (2010). Inequalities in mortality in small areas of eleven Spanish cities (the multicenter MEDEA project). *Health Place,* 16, 703-711.

Boyle MH, & Willms JD (1999). Place effects for areas defined by administrative boundaries. *Am.J.Epidemiol.,* 149, 577-585.

Braumoeller B (2006). Explaining Variance; Or, Stuck in a Moment We Can't Get Out Of. . *Political Analysis,* 14, 268-290.

Browne, W.J. (2009a). MCMC Estimation in MLwiN (Version 2.13) Centre for Multilevel Modelling, University of Bristol *http://www.bristol.ac.uk/cmm/software/mlwin/download/mcmc-print.pdf*.

Browne, W.J. (2009b). Parameter Expansion. . *MCMC Estimation in MLwiN (Version 2.13)* Bristol: Centre for Multilevel Modelling, University of Bristol

Browne WJ (2003). *MCMC estimation in MLwiN. Version 2.0*. London: Centre for Multilevel Modelling. Institute of Education. University of London.

Browne WJ, Goldstein H, & Rasbash J (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling,* 1, 103-124.

Browne, W.J., Subramanian, S.V., Jones, K., & Goldstein, H. (2005). Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 168, 599-613.

Bullen, N., Moon, G., & Jones, K. (1996). Defining localities for health planning: a GIS approach. *Soc Sci Med,* 42, 801-816.

Carstairs, V. (1995). Deprivation indices: their interpretation and use in relation to health. *J Epidemiol Community Health,* 49 Suppl 2, S3-S8.

Censos de Población y Viviendas 2001. Instituto Nacional de Estadistica de España.

Chaix, B., Lindstrom, M., Rosvall, M., & Merlo, J. (2008). Neighbourhood social interactions and risk of acute myocardial infarction. *J Epidemiol Community Health,* 62, 62-68.

Clarke P, & Wheaton B (2007). Addressing Data Sparseness in Contextual Population Research: Using Cluster Analysis to Create Synthetic Neighborhoods. *Sociological Methods Research,* 35, 311 - 351.

Collaboration (2010). Atlas VPM. Atlas de Variaciones en la Práctica Médica en el Sistema Nacional De Salud *http://www.atlasvpm.org/avpm/inicio.inicio.do*.

Cummins, S., Curtis, S., Diez-Roux, A.V., & Macintyre, S. (2007). Understanding and representing 'place' in health research: A relational approach. *Soc Sci Med,* 65, 1825-1838.

Diez Roux, A.V. (2008). Next steps in understanding the multilevel determinants of health. *J Epidemiol Community Health 1470-2738 (Electronic),* 62, 957-959.

Downs GW, & Roche DM (1979). Interpreting Heteroscedasticity. *American Journal of Political Science,* 23, 816-828.

Duncan, C., Jones, K., & Moon, G. (1993). Do places matter? A multi-level analysis of regional variations in health-related behaviour in Britain. *Soc Sci Med,* 37, 725-733.

Duncan, C., Jones, K., & Moon, G. (1995). Psychiatric morbidity: a multilevel approach to regional variations in the UK. *J Epidemiol Community Health,* 49, 290-295.

Duncan, C., Jones, K., & Moon, G. (1996). Health-related behaviour in context: a multilevel modelling approach. *Soc Sci Med,* 42, 817-830.

Duncan, C., Jones, K., & Moon, G. (1998). Context, composition and heterogeneity: using multilevel models in health research. *Soc Sci Med.,* 46, 97-117.

Duncan, C., Jones, K., & Moon, G. (1999). Smoking and deprivation: are there neighbourhood effects? *Soc Sci Med,* 48, 497-505.

Epstein, E., & Guttman, R. (1984). Mate selection in man: evidence, theory, and outcome. *Soc Biol,* 31, 243-278.

Folwell, K. (1995). Single measures of deprivation. *J Epidemiol Community Health,* 49 Suppl 2, S51-S56.

Fone, D.L., Farewell, D.M., & Dunstan, F.D. (2006). An ecometric analysis of neighbourhood cohesion. *Popul Health Metr,* 4, 17.

Goldstein, H. (2003). *Multilevel Statistical Models*: Hodder Arnold, London.

Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning variation in generalised linear multilevel models. *Understanding Statistics,* 1, 223-232.

Gould SJ (1996). *Full House: The Spread of Excellence from Plato to Darwin* New York: Three Rivers Press.

Hernan, M.A., & Robins, J.M. (2006). Estimating causal effects from epidemiological data. *J Epidemiol Community Health,* 60, 578-586.

Ibanez, B., Librero, J., Bernal-Delgado, E., Peiro, S., Lopez-Valcarcel, B.G., Martinez, N., et al. (2009). Is there much variation in variation? Revisiting statistics of small area variation in health services research. *BMC Health Serv Res,* 9, 60.

Jones, K., Moon, G., & Clegg, A. (1991). Ecological and individual effects in childhood immunisation uptake: a multi-level approach. *Soc Sci Med,* 33, 501-508.

Kaplan, G.A. (1999). What is the role of the social environment in understanding inequalities in health? *Ann. N.Y .Acad .Sci.,* 896, 116-119.

Katz, J., Carey, V.J., Zeger, S.L., & Sommer, A. (1993). Estimation of design effects and diarrhea clustering within households and villages. *Am J Epidemiol,* 138, 994-1006.

Krieger, N. (2001). Theories for social epidemiology in the 21st century: an ecosocial perspective. *International Journal of Epidemiology,* 30, 668-677.

Larsen, K., & Merlo, J. (2005). Appropriate assessment of neighborhood effects on individual health: integrating random and fixed effects in multilevel logistic regression. *Am J Epidemiol,* 161, 81-88.

Larsen, K., Petersen, J.H., Budtz-Jorgensen, E., & Endahl, L. (2000). Interpreting parameters in the logistic regression model with random effects. *Biometrics,* 56, 909-914.

Lawlor DA, Mishra GD, & Eds (2009). *Family Matters. Designing, analysing and understanding family-based studies in life-course epidemiology*. New York: Oxford University Press.

Li J, Gray BR, & Bates DM (2008). An Empirical Study of Statistical Properties of Variance Partition Coefficients for Multi-Level Logistic Regression Models. *Communications in Statistics - Simulation and Computation,* 37, 2010–2026.

Lynch, J., & Smith, G.D. (2005). A life course approach to chronic disease epidemiology. *Annu Rev Public Health,* 26, 1-35.

Macintyre S, & Elleway A. (2000). Ecological approaches: rediscovering the role of the physical and social environment. In Berkman LF, & Kawachi I (Eds.), *Social epidemiology* pp. 332-348). New York: Oxford University Press.

Macintyre, S., Ellaway, A., & Cummins, S. (2002). Place effects on health: how can we conceptualise, operationalise and measure them? *Soc Sci Med,* 55, 125-139.

MacNab, Y.C., & Dean, C.B. (2002). Spatio-temporal modelling of rates for the construction of disease maps. *Stat Med*, 21, 347-358.

Merlo, J. (2003). Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. *J Epidemiol Community Health,* 57, 550-552.

Merlo, J. (2010). (Book review) Family Matters: Designing, Analysing and Understanding Family-based Studies in Life-course Epidemiology. Debbie A Lawlor, Gita D Mishra (eds) *Int. J. Epidemiol. 39(3): 936-937 first published online January 16, 2010 doi:10.1093/ije/dyp387*

Merlo J (2011). Contextual Influences on the Individual Life Course: Building a Research Framework for Social Epidemiology. http://bit.ly/i5s64N. *Psychosocial Intervention,* 20, 109-118.

Merlo, J., Asplund, K., Lynch, J., Rastam, L., & Dobson, A. (2004). Population effects on individual systolic blood pressure: a multilevel analysis of the World Health Organization MONICA Project. *Am J Epidemiol,* 159, 1168-1179.

Merlo, J., Bengtsson-Bostrom, K., Lindblad, U., Rastam, L., & Melander, O. (2006a). Multilevel analysis of systolic blood pressure and ACE gene I/D polymorphism in 438 Swedish families--a public health perspective. *BMC Med Genet,* 7, 14.

Merlo, J., & Chaix, B. (2006). Neighbourhood effects and the real world beyond randomized community trials: a reply to Michael J Oakes. *International journal of epidemiology,* 35, 1361-1363.

Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., et al. (2006b). A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *J Epidemiol Community Health,* 60, 290-297.

Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., et al. (2006c). A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *J Epidemiol Community Health 1470-2738 (Electronic),* 60, 290-297.

Merlo, J., Chaix, B., Yang, M., Lynch, J., & Rastam, L. (2005a). A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *J Epidemiol Community Health 1470-2738 (Electronic),* 59, 443-449.

Merlo, J., Chaix, B., Yang, M., Lynch, J., & Rastam, L. (2005b). A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *J Epidemiol Community Health,* 59, 1022-1028.

Merlo, J., Chaix, B., Yang, M., Lynch, J., & Rastam, L. (2005c). A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *J Epidemiol Community Health 1470-2738 (Electronic),* 59, 1022-1028.

Merlo, J., Ohlsson, H., Lynch, K.F., Chaix, B., & Subramanian, S.V. (2009). Individual and collective bodies: using measures of variance and association in contextual epidemiology. *J Epidemiol Community Health 1470-2738 (Electronic),* 63, 1043-1048.

Merlo, J., Ostergren, P.O., Hagberg, O., Lindstrom, M., Lindgren, A., Melander, A., et al. (2001). Diastolic blood pressure and area of residence: multilevel versus ecological analysis of social inequity. *J Epidemiol Community Health,* 55, 791-798.

Merlo, J., Yang, M., Chaix, B., Lynch, J., & Rastam, L. (2005d). A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. *J Epidemiol Community Health 1470-2738 (Electronic),* 59, 729-736.

Messer, L.C. (2007). Invited commentary: Beyond the metrics for measuring neighborhood effects. *Am J Epidemiol,* 165, 868-871; discussion 872-863.

Middleton, N., Sterne, J.A., & Gunnell, D.J. (2008). An atlas of suicide mortality: England and Wales, 1988-1994. *Health Place,* 14, 492-506.

Mineau, G.P., Smith, K.R., & Bean, L.L. (2002). Historical trends of survival among widows and widowers. *Soc Sci Med,* 54, 245-254.

Morgenstern, H. (1998). Ecological studies. In Rothman KJ, & Greenland S (Eds.), *Modern epidemiology* pp. 459-480). Philadelphia: Lippincott-Raven.

Mujahid, M.S., Diez Roux, A.V., Morenoff, J.D., & Raghunathan, T. (2007). Assessing the measurement properties of neighborhood scales: from psychometrics to ecometrics. *Am J Epidemiol,* 165, 858-867.

Muntaner, C., Li, Y., Ng, E., Benach, J., & Chung, H. (2011). Work or place? Assessing the concurrent effects of workplace exploitation and area-of-residence economic inequality on individual health. *Int J Health Serv,* 41, 27-50.

Naess, O., Claussen, B., Smith, G.D., & Leyland, A.H. (2008). Life course influence of residential area on cause-specific mortality. *J Epidemiol Community Health,* 62, 29-34.

Oakes JM (2003). The (Mis)Estimation of Neighborhood Effects: Causal Inference for a Practicable Social Epidemiology. *Soc Sci Med.*

Ocana-Riola, R., & Mayoral-Cortes, J.M. (2010). Spatio-temporal trends of mortality in small areas of Southern Spain. *BMC Public Health,* 10, 26.

Ocana-Riola, R., Mayoral-Cortes, J.M., Sanchez-Cantalejo, C., Toro-Cardenas, S., Fernandez-Ajuria, A., & Mendez-Martinez, C. (2008a). [Interactive mortality atlas in Andalusia, Spain (AIMA)]. *Rev Esp Salud Publica,* 82, 379-394.

Ocana-Riola, R., Saurina, C., Fernandez-Ajuria, A., Lertxundi, A., Sanchez-Cantalejo, C., Saez, M., et al. (2008b). Area deprivation and mortality in the provincial capital cities of Andalusia and Catalonia (Spain). *J Epidemiol Community Health,* 62, 147-152.

Ohlsson, H., & Merlo, J. (2011). Place effects for areas defined by administrative boundaries: a life course analysis of mortality and cause specific morbidity in Scania, Sweden. *Soc Sci Med,* 73, 1145-1151.

Petronis, K.R., & Anthony, J.C. (2003). A different kind of contextual effect: geographical clustering of cocaine incidence in the USA. *J Epidemiol Community Health,* 57, 893-900.

Pickle, L.W., Mungiole, M., Jones, G.K., & White, A.A. (1999). Exploring spatial patterns of mortality: the new atlas of United States mortality. *Stat Med,* 18, 3211-3220.

Raudenbush SW, & Sampson RJ (1999). Ecometrics: toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology,* 29 1-41.

Riva, M., Gauvin, L., & Barnett, T.A. (2007). Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations published since July 1998. *J Epidemiol Community Health,* 61, 853-861.

Robinson, W.S. (2009). Ecological correlations and the behavior of individuals. *Int J Epidemiol,* 38, 337-341.

Shaw, M. (2008). *The Grim Reaper's road map : an atlas of mortality in Britain*. Bristol: Policy Press.

Siddiqi, N. (2011). Publication bias in epidemiological studies. *Cent Eur J Public Health,* 19, 118-120.

Snijders, T., & Bokser, R. (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. Thousand Oaks, California: Sage Publications.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J R Stat Soc B,* 64, 583-639.

Subramanian SV, Glymour MM, & Kawachi I. (2007). Identifying causal ecologic effects on health: a methodological assessment. . *In: Galea S, ed. Macrosocial determinants of population health. New York: Springer Media,*.

Subramanian, S.V., Jones, K., & Duncan, C. (2003). Multilevel methods for public health research. In I. Kawachi, & L.F. Berkman (Eds.), *Neighborhoods and Health* pp. 65-111). New York, NY: Oxford University Press.

Subramanian, S.V., Jones, K., Kaddour, A., & Krieger, N. (2009). Revisiting Robinson: the perils of individualistic and ecologic fallacy. *Int J Epidemiol,* 38, 342-360; author reply 370-343.

Turrell, G., & Mengersen, K. (2000). Socioeconomic status and infant mortality in Australia: a national study of small urban areas, 1985-89. *Soc Sci Med,* 50, 1209-1225.

Twigg, L., Moon, G., & Jones, K. (2000). Predicting small-area health-related behaviour: a comparison of smoking and drinking indicators. *Soc Sci Med*, 50, 1109-1120.

Viciana F, Montañez Cobo V, Canovas Balboa MR, & PozaCruz E (2010). Base de Datos Longitudinal de Población de Andalucía (BDLPA): Modelo de datos y sistema de gestión. . *Jornadas de Estadistica de la s Comunidades Autonomas 2010;http://www.jecas.org/ponencias/jueves/tarde/desarrollosII/BDlogitudinalIAE.pdf (by 6, June, 2011).*

Yang, M., Eldridge, S., & Merlo, J. (2009). Multilevel survival analysis of health inequalities in life expectancy. *Int J Equity Health*, 8, 31.

Yule GU (1934). On some points relating to vital statistics, more especially statistics of occupational mortality. *J Roy Stat Soc 97, part 1*, 1-84.

FIGURES LEGENDS

Figure 1: Age- and sex-adjusted geographical differences in all-cause mortality (A) and in individual low educational achievement (B) obtained from a multilevel logistic regression analysis with only two levels (i.e., individuals within census tracts). These differences are expressed as ORs and 95% confidence interval. The values are in the logarithmic (log.) scale so an OR=1 (i.e., no difference) correspond with the log OR = 0. In this analysis the $VPC_{census\ tract}$ for mortality is equal to 1%, which means that the correlation in the probability of death between two individuals randomly selected from the same census tract is only 1%. However, the geographical clustering was much higher when modelling individual low educational achievement ($VPC_{census\ tract}$ = 24%) than for mortality

Figure 2: Census tract differences in all-cause mortality expressed as residuals and 95% confidence intervals obtained from multilevel regression analyses before (A) and after (B) adjustment for sex, residential mobility, and individual predicted probability of death (i.e. predicted probability for all-cause mortality as a function of individual age, civil status, educational achievement, housing tenure, having only one dwelling, and occupation). The values are in the logarithmic (log.) scale so an OR=1 (i.e., no difference) correspond with the log OR = 0. In this analysis the $VPC_{census\ tract}$ for mortality is equal to 2.3% in the unadjusted analysis, which means that the correlation in the probability of death between two individuals randomly selected from the same census tract is only 2.3%. However, the geographical clustering in mortality was considerably reduced in the adjusted analysis ($VPC_{census\ tract}$ = 0.3%). The figure shows only the census tracts in the provinces that have the highest (Huelva, red) and lowest (Cordoba, blue) mortality in Andalusia. The rest of the census tracts are in the analysis, but they are not shown in the figure.

Table 1. Characteristics of the population of men and women ages 45–79 years residing in Andalusia in 2001, by quartiles of percentage of people with very low/low educational achievement at the census tract. Values are percentages, unless otherwise indicated.

	Percentage of people low educational achievement at the census tract							
	1st quartile		2nd quartile		3rd quartile		4th quartile	
	Women	Men	Women	Men	Women	Men	Women	Men
All-cause mortality 2001–2009	7	12	9	14	9	16	10	17
Predicted probability of death*	7	12	8	14	9	16	10	17
Age (mean years)	59	58	60	59	61	60	62	61
Civil status								
Married	11	6	8	7	6	9	6	11
Single	69	87	70	86	72	84	72	83
Widowed	15	3	18	4	19	4	20	5
Separated	3	2	2	2	2	2	1	2
Divorced	2.5	1.5	1.3	1.0	0.9	0.6	0.5	0.5
Educational achievement								
Very low	3	1.1	7	3	11	5	19	8
Low	43	31	62	54	71	69	71	77
Medium	29	25	23	26	14	18	7	11
High	26	42	8	16	4	9	2	4
Housing tenure								
Owned	89	89	89	89	89	88	88	87
Rented	8	8	6	6	6	6	5	6
Other form	3	3	4	4	6	6	7	8
Having only a dwelling	75	75	87	87	91	90	93	93
Occupation								
Working	25	55	17	48	14	42	12	37
Receiving formal education	1.3	0.4	1.1	0.2	1.3	0.2	1.3	0.2
Unemployed	4	6	5	7	6	9	9	11
Disability pension	3	6	4	8	4	8	5	9
Pension for widow/orphan	11	0.4	13	0.5	14	0.5	15	0.7
Retirement	10	30	11	35	14	39	17	42
Needing help, basic activities	0.5	0.1	0.4	0.2	0.4	0.1	0.4	0.1
Household work	43	0.8	48	0.7	46	0.7	40	0.6
Other occupations	0.7	0.7	0.5	0.5	0.5	0.5	0.4	0.4
“Movers”**	7.7	9.1	5.4	6.2	4.1	4.7	2.9	3.3

*According to a logistic regression modelling all-cause mortality 2001–2009 and including as predictor variables all the individual characteristics indicated in the table. **Individuals residing in a different municipality 10 years before the 2001 census.

Table 2. Association between individual socioeconomic characteristics and all-cause mortality in the next eight years (2001–2009) among men and women ages 45–79 years residing in Andalusia in 2001. These individual variables were included in an equation for obtaining the predicted probability of death. Values are odds ratios (OR) and 95% confidence intervals (CI).

	Men			Women		
	OR	95% CI		OR	95% CI	
Age (years)						
45–49	1.00			1.00		
50–54	1.55	1.40	1.71	1.31	1.13	1.52
55–59	2.25	2.04	2.47	2.04	1.78	2.35
60–64	3.19	2.90	3.51	3.03	2.64	3.47
65–69	4.45	4.03	4.92	5.08	4.46	5.79
70–74	7.97	7.20	8.83	9.86	8.67	11.22
75–79	15.01	13.52	16.65	20.67	18.16	23.52
Civil status						
Married	1.00			1.00		
Single	1.59	1.49	1.69	1.32	1.22	1.43
Widowed	1.36	1.27	1.46	1.15	1.06	1.24
Separated	1.79	1.57	2.04	1.23	1.02	1.49
Divorced	1.41	1.14	1.73	0.97	0.73	1.29
Educational achievement						
High	1.00			1.00		
Medium	1.02	0.95	1.10	0.94	0.83	1.06
Low	1.12	1.06	1.19	1.13	1.02	1.25
Very low	1.44	1.32	1.58	1.59	1.42	1.78
Housing tenure						
Owned	1.00			1.00		
Rented	1.39	1.29	1.49	1.34	1.24	1.46
Other form	1.16	1.07	1.26	1.14	1.04	1.26
Occupation						
Working	1.00					
Receiving formal education	1.44	0.99	2.11	0.93	0.72	1.19
Unemployed	1.53	1.41	1.67	1.03	0.86	1.23
Disability pension	3.13	2.91	3.36	3.05	2.68	3.48
Pension for widow/orphan	2.10	1.71	2.57	1.55	1.36	1.76
Retirement	1.99	1.86	2.13	1.53	1.36	1.72
Needing help with basic activities	8.77	5.96	12.90	9.48	7.61	11.80
Performing household tasks	2.04	1.68	2.47	1.29	1.16	1.44
Other occupations	1.92	1.50	2.46	1.36	1.00	1.84
Having only one dwelling (yes/no)	1.31	1.23	1.39	1.23	1.14	1.33

Table 3. Multilevel analysis of variance showing general contextual effects on all-cause mortality in the next eight years (2002–2009) among men and women ages 45–79 years residing in Andalusia at the end of 2001. Values are median (95% confidence interval), unless otherwise indicated.

General contextual effects	Model 1		Model 2		Model 3		Model 4	
Variance								
Municipality (σ^2_{M})	0.015	(0.009 - 0.023)	0.008	(0.004 - 0.014)	0.009	(0.005 - 0.015)	0.006	(0.002 - 0.011)
Census tract (σ^2_{C})	0.076	(0.064 - 0.091)	0.010	(0.000 - 0.005)	0.003	(0.001 - 0.010)	0.007	(0.003 - 0.013)
Household (σ^2_{H})	0.660	(0.582 - 0.743)	0.310	(0.235 - 0.388)	0.309	(0.226 - 0.391)	0.301	(0.220 - 0.383)
Individual[a]								
Variance partition coefficient (VPC)								
VPC_M	0.4%		0.2%		0.2%		0.2%	
VPC_C	2.3%		0.3%		0.3%		0.4%	
VPC_H	18.6%		8.9%		8.9%		8.7%	
Median odds ratio (MOR)								
MOR_M	1.12		1.09		1.09		1.08	
MOR_C	1.33		1.10		1.11		1.11	
MOR_H	2.29		1.72		1.72		1.71	
Bayesian deviance information Criterion (BDIC)	163450		135684		135657		135626	
BDIC change compared to the previous model			-27765		-27		-31	

[a]The value of the variance of the underlying individual-level variable according to the logistic distribution $\pi^2/3$ or 3.29 in all models.

Table 4. Multilevel analysis showing the specific individual and contextual characteristics of all-cause mortality in the next eight years (2002–2009) among men and women ages 45–79 years residing in Andalusia at the end of 2001. Values are odds ratios (OR) and 95% confidence intervals (CI).

	Model 1	Model 2		Model 3		Model 4	
		OR	95% CI	OR	95% CI	OR	95% CI
Specific individual effects							
Men vs. Women		1.12	(1.09 - 1.16)	1.12	(1.09 - 1.16)	1.12	(1.09 - 1.16)
Predicted probability of death[a]							
Decile group 1 (Lowest risk)		Reference		Reference			
Decile group 5		1.94	(1.68 - 2.24)	4.21	(3.69 - 4.81)	4.21	(3.71 - 4.85)
Decile group 10 (Highest risk)		58.15	(51.62 - 66.2)	57.4	(50.3 - 65.3)	57.45	(51.06 - 65.69)
"Stayers" vs. "movers"		1.07	(0.99 - 1.15)	1.05	(0.98 - 1.13)	1.05	(0.98 - 1.13)
Specific contextual effects							
Census tract low educational achievement							
Quartile 1 (Lowest percentage)				Reference			
Quartile 2				1.12	(1.07 - 1.18)	1.12	(1.07 - 1.18)
Quartile 3				1.12	(1.06 - 1.17)	1.12	(1.06 - 1.17)
Quartile 4 (Highest percentage)				1.13	(1.07 - 1.19)	1.14	(1.08 - 1.20)
Region							
Seville						Reference	
Cadiz						1.10	(1.03 - 1.17)
Huelva						1.10	(1.02 - 1.19)
Jaen						0.93	(0.87 - 0.99)
Almeria						0.98	(0.90 - 1.06)
Córdoba						0.86	(0.80 - 0.92)
Granada						0.99	(0.92 - 1.05)
Malaga						1.02	(0.96 - 1.09)

[a] According a logistic regression modelling all-cause mortality 2001–2009 and including as predictor variables individual age, civil status, educational achievement, housing tenure, having only one dwelling, and occupation (see also Table 2).

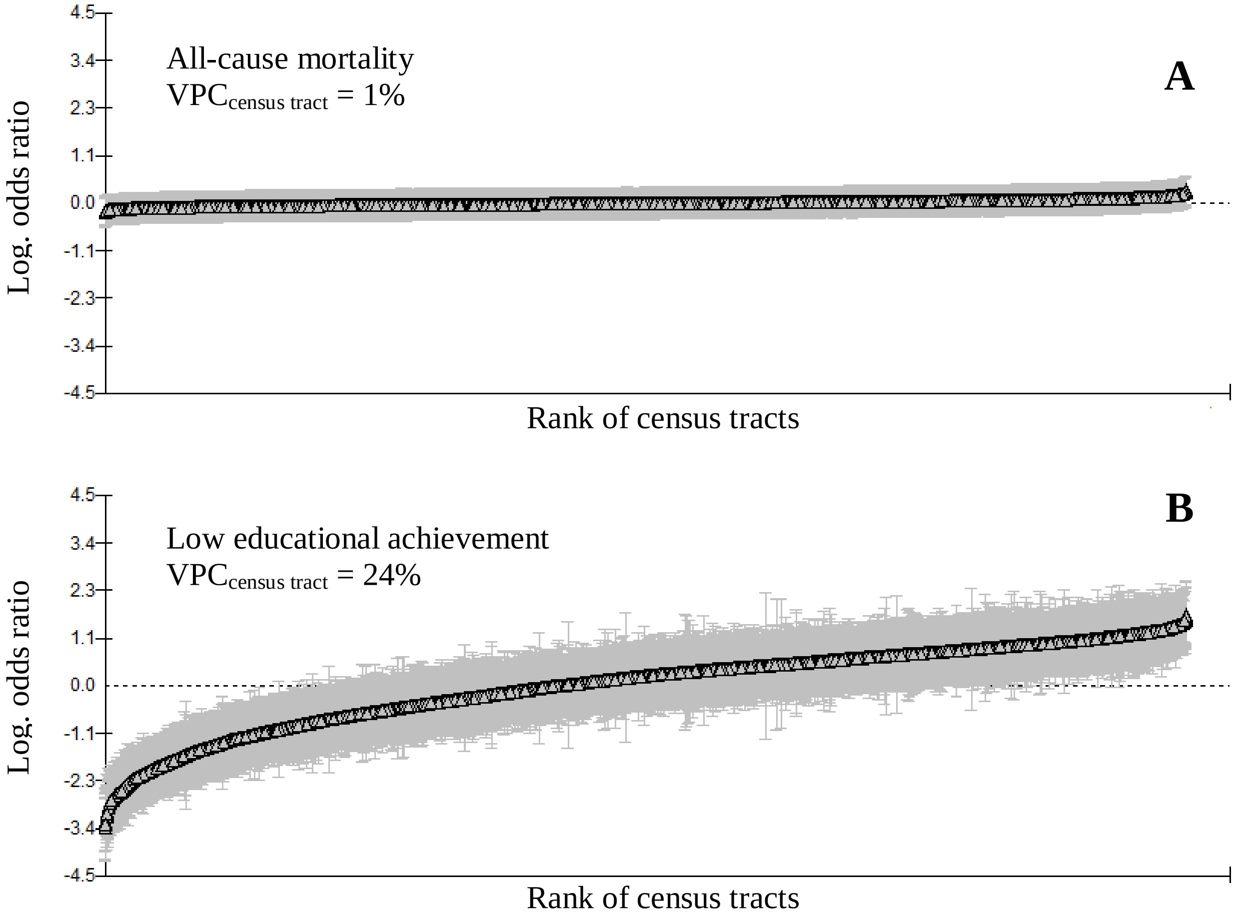
4.5
3.4
2.3
1.1
0.0
-1.1
-2.3
-3.4
-4.5
Log. odds ratio
All-cause mortality
VPCcensus tract = 1%
A
Rank of census tracts
4.5
3.4
2.3
1.1
0.0
-1.1
-2.3
-3.4
-4.5
Log. odds ratio
Low educational achievement
VPCcensus tract = 24%
B
Rank of census tracts

Figure 1

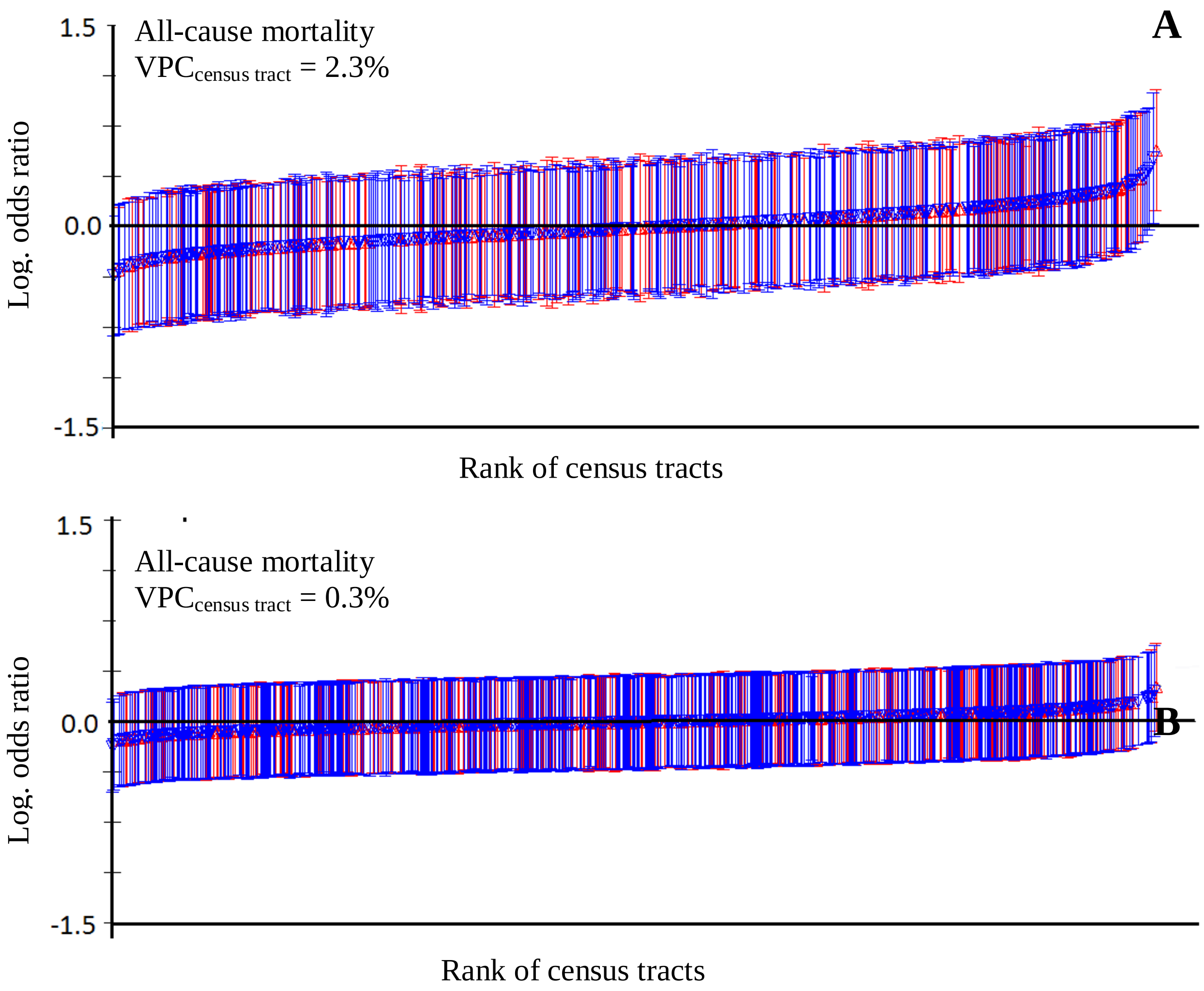
1.5
0.0
-1.5
Log. odds ratio
All-cause mortality
VPC census tract = 2.3%
A
Rank of census tracts
1.5
0.0
-1.5
Log. odds ratio
All-cause mortality
VPC census tract = 0.3%
B
Rank of census tracts

Figure 2