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

Juan Merlo a, b, *, 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
c Demographic and Social Statistics, Institute of Statistics and Cartography of Andalusia, Seville, Spain
d Institute of Economics, Geography and Demography, Center for Humanities and Social Sciences, Spanish National Research Council, Madrid, Spain

A R T I C L E   I N F O

Article history:
Available online xxx

Keywords:
Spain
Mortality
Multilevel analysis
Ecological analysis
Small-area analysis
Social environment
Socio-economic factors

A B S T R A C T

We performed a multilevel analysis (including individuals, households, census tracts, municipalities and provinces) on a 10% sample (N = 230,978) from the Longitudinal Database of the Andalusian Population (LDAP). We aimed to investigate place effects on 8-year individual mortality risk. Moreover, besides calculating association (yielding odds ratios, ORs) between area socio-economic circumstances and individual risk, we wanted to estimate variance and clustering using the variance partition coefficient (VPC). We explicitly proclaim the relevance of considering general contextual effects (i.e. the degree to which the context, as a whole, affects individual variance in mortality risk) under at least two circumstances. The first of these concerns the interpretation of specific contextual effects (i.e. the association between a particular area characteristic and individual risk) obtained from multilevel regression analyses. The second involves the interpretation of geographical variance obtained from classic ecological spatial analyses. The so-called “ecological fallacy” apart, the lack of individual-level information renders 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 analyses. We observed an association between percentage of people with a low educational level in the census tract and individual mortality risk (OR, highest v. lowest quintile = 1.14; 95% confidence interval, CI 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 (VPCM = 0.2% and VPCCT = 0.3%). Conversely, the household (H) level appeared much more relevant (VPCH = 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.

© 2012 Elsevier Ltd. All rights reserved.

Introduction

An individual’s health is affected by the societal context in which the individual lives. Since many social processes take place over space (Cummins, Curtis, Diez-Roux, & Macintyre, 2007; Kaplan, 1999; Macintyre & Ellaway, 2000; Macintyre, Ellaway, & Cummins, 2002; Merlo, 2011), it is a fundamental issue in public health to identify the social and geographical environments that determine individual health variance. Moreover, it is necessary to identify the specific characteristics of these contexts that explain such variance and are associated with individual disease risk (Cummins et al., 2007; Merlo, Ohlsson, Lynch, Chaix, & Subramanian, 2009).

It is well established that multilevel modelling (Bingenheimer & Raudenbush, 2004; Duncan, Jones, & Moon, 1998; Goldstein, 2003; Merlo, Chaix et al., 2006; Merlo, Chaix, Yang, Lynch, & Rastam, 2005a, 2005b; Merlo, Yang, Chaix, Lynch, & Rastam, 2005; Subramanian, Jones, & Duncan, 2003) is a useful instrument for quantitative analyses of place effects on individual health. Many previous multilevel analyses have focused on the study of associations and they have considered the analysis of variance as secondary or irrelevant information (Blakely & Woodward, 2000; Diez Roux, 2008). By contrast, some scholars have explicitly concluded that the analysis of variance provides indispensable information for understanding place effects on health (Boyle &

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, 2004; Borrell, Mari-Dell’Olmo, Serral, Martínez-Beneito, & Gotsens, 2010; Gobierno de Aragon, Departamento de Salud y Consumo, 2010; MacNab & Dean, 2002; Middleton, Sterne, & Gunnell, 2008; Ocana-Riola & Mayoral-Cortes, 2010; Ocana-Riola, Mayoral-Cortes, et al., 2008; Pickle, Mungiole, Jones, & White, 1999; Shaw, 2008; Turrell & Mengersen, 2000). From an empirical perspective, the advantages of multilevel vs. ecological regression analyses were clearly identified by the seminal work performed by Aitkin and Longford (1986) as well as researchers such as Jones, Duncan, Subramanian and others (Bulien, Moon, & Jones, 1996; Duncan et al., 1993; Duncan, Jones, & Moon, 1995, 1996, 1998, 1999; Jones, Moon, & Clegg, 1991; Subramanian, Jones, Kaddour, & Krieger, 2009; Twigg, Moon, & Jones, 2000). We also contributed to this discussion in a previous work (Merlo et al., 2001).

Beyond the conventional study of specific contextual effects (i.e. a study investigating the association between particular contextual variables and individual outcomes), the use of multilevel regression and analogous techniques of analysis (Katz, Carey, Zeger, & Sommer, 1993) has allowed us to identify general contextual effects that are quantified by measures of variance and clustering rather than by measures of association (Merlo, 2003; Merlo et al., 2009, 2001; Petronis & Anthony, 2003; Subramanian, Gymlour, & Kawachi, 2007). However, while the explicit distinction between these two types of effects enriches our understanding of contextual influences on individual health, these methods are still fairly infrequent in social epidemiology (Merlo et al., 2009).

The study of general contextual effects shows the extent to which the geographical constrasts we use for defining a context (e.g. neighbourhoods, small areas, census tracts) determine individual outcomes (e.g. mortality) without specifying any contextual characteristic other 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 influences individual health, one should expect not only that there would be a statistically significant spatial variation (as frequently detected in “small-area variation studies”) (Ibanez et al., 2009), but also that this spatial variation affects a meaningful proportion of the total individual-level variation (Boyle & Willms, 1999; Merlo, 2003; Merlo et al., 2004, 2009; Subramanian et al., 2007). The work by Duncan et al. (1993) cited above was among the first studies to interpret general contextual effects on health.

In spite of the fundamental relevance of general contextual effects when investigating socio-geographical influences on individual health, only a relatively small number of the multilevel analyses published hitherto have reported measures of variance (Riva et al., 2007), and still fewer 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 about the well-known “ecological fallacy” (Morgenstern, 1998; Robinson, 2009) and the advantages of multilevel analysis for preventing it (Merlo et al., 2001; Subramanian et al., 2009), but also about the inability of ecological analyses to quantify 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 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 & Mishra, 2009; Merlo, 2010; Merlo, Bengtsson-Bostrom, Lindblad, Rastam, & Melander, 2006; Yang, Eldridge, & Merlo, 2009). The correlation between individuals within the family/household context regarding their health is much larger than among the residents of places defined by administrative boundaries (Boyle & Willms, 1999; Merlo et al., 2009; Yang et al., 2009). Therefore, the family/household level should be considered in analyses of place effects.

Against the above described background, we performed a multilevel analysis to investigate place effects on individuals’ mortality risk. We explicitly considered general contextual effects (i.e. the degree to which provinces, municipalities, census tracts and familial households affect individual variance in mortality risk) when interpreting specific contextual effects (i.e. the association between small-area socio-economic circumstances and individual mortality risk). In addition, when performing the multilevel analyses, we explicitly questioned the appropriateness of ecological spatial analyses of small-area variations in mortality. These ecological studies are frequent not only in Spain but all over the world, and may provide incorrect information on place effects on individual health.

We performed our analyses on a representative sample of the Andalusian population who participated in the Spanish Population and Housing Census, 2001 (“Instituto Nacional de Estadística, 2001”), who were included in the recently created Longitudinal Database of the Andalusian Population (LDAP) (Viciana, Montañez Cobo, Canovas Balboa, & Pozacruz, 2010). As far as we know, our study is the first multilevel analysis investigating mortality on a large population-based cohort from southern Europe.

Population and methods

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

The Institute of Statistics and Cartography of Andalusia, in collaboration with the Spanish National Research Council through a project funded by the European Social Fund, has created the LDAP research database (Viciana et al., 2010). The LDAP records information on all individuals residing in Andalusia, according to the Municipal Register elaborated by the Spanish National Institute of Statistics. The database also tracks vital statistics information of every Andalusian resident, including those who die outside Andalusia. Following the Statistical Planning, the Institute of Statistics and Cartography of Andalusia links the records of migration and mortality to other statistical databases using an internal numeric key of the LDAP. The database was constructed following the Law approved by the Andalusian Parliament (Ley 4/2007 del Plan Estadístico de Andalucía 2007–2012), that includes specific conditions of confidentiality and ethical standards (Ley 4/1989 de Estadística de la Comunidad Autónoma de Andalucía) and allows the scientific use of data (art. 25.3). All statistical analyses were performed using anonymized information after signing a formal confidentiality agreement.

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 45–79 years identified in the 10% census sample of dwellings, we analysed 94% (230,978/244,972) individuals. The individuals analysed had valid identification numbers and were Spanish-born; furthermore, information on their vital status
(i.e. dead or alive) was available to us from the follow-up, which ended on 31 December 2009. These individuals were living in 142,516 households within 5381 census tracts in 770 municipalities comprising the eight provinces of Andalusia (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.

The analysis included a gender comparison across 5-year age groups (i.e. 45–49, ... 75–79 years). We classified the occupation of the individuals into nine 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 divided the civil status of the individuals into single, married, widowed, separated, or divorced. We used the married category as reference in the comparisons.

Furthermore, we classified educational level into four categories, very low (illiterate, i.e. having <1 year of formal education), low (primary to elementary school, i.e. 5–9 years of formal education), medium (elementary baccalaureate or similar degree, i.e. 9–10 years of formal education) and high (superior baccalaureate or higher educational level, i.e. ≥12 years of formal education).

The census of 2001 contained an item indicating the municipality of residence at the moment of the census was completed. We compared this municipality with the groups (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 divided the civil status of the individuals into single, married, widowed, separated, or divorced. We used the married category as reference in the comparisons.

Concerning housing tenure, we distinguished between individuals who (i) owned or (ii) rented their homes, and who used (iii) other forms of housing tenure, such as low-rent public housing. 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 with 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 Table 2). Thereafter, we categorised this predicted probability into 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, Kaltenbach, Ding, and Ray (2008) for an extended explanation of 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 or a low educational level was a specific contextual characteristic that reflected the socio-economic and material circumstances in the census tract. We performed a direct age standardisation using the equivalent average rate (Yule, 1934) for 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 a very low or low educational level as reference.

Multilevel regression analyses

The data presented a hierarchical structure consisting of individual (level 1), individual nested within households (level 2) and nested within census tracts (level 3), census tract nested within municipalities (level 4) and municipality nested within the province (level 5). Therefore, we performed multilevel logistic regression analyses (Goldstein, 2003; Snijders & Bokser, 1999).

The multilevel regression analysis 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 in order to obtain correct statistical estimations of uncertainty (i.e. standard errors), and also, substantive information on the distribution of the individual variance across these levels (Merlo et al., 2005a).

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 or low educational levels in the census tracts. Finally, in the fourth model, we included the eight Andalusian provinces as a fixed effect, with Seville as the reference category.

We also performed age- and sex-adjusted multilevel logistic regression analyses with the individual nested within census tracts, and with mortality as well as low/very low educational level modelled separately. The purpose of this analysis was to compare geographical (e.g. census tracts) differences in socio-economic characteristics (e.g. educational level) with geographical differences in mortality.

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

Specific contextual effects

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

General contextual effects

General contextual effects give information on the degree to which the areas/households under investigation affect individual differences in mortality risk. For their assessment, we did not specify any contextual characteristic other than the boundaries that delimit the level of analysis (e.g. census tract, household) (Merlo et al., 2005b; Merlo et al., 2005a; Subramanian et al., 2007).

General contextual effects are estimates of clustering and of variance. We used two different measures to calculate these, the variance partition coefficient (VPC) and the median odds ratio (MOR).
The variance partition coefficient. We calculated the VPC at the municipality ($M$), census tract ($CT$) and household ($H$) level using the latent variable method (Browne, Subramanian, Jones, & Goldstein, 2005; Goldstein, Browne, & Rasbash, 2002; Li, Gray, & Bates, 2008; Merlo et al., 2006a; Snijders & Bokser, 1999), as follows:

\[
\text{VPC}_M = \sigma_M^2 / \left( \sigma_M^2 + \sigma_C^2 + \sigma_H^2 + \pi^2/3 \right)
\]

\[
\text{VPC}_{CT} = \left( \sigma_M^2 + \sigma_C^2 \right) / \left( \sigma_M^2 + \sigma_C^2 + \sigma_H^2 + \pi^2/3 \right)
\]

\[
\text{VPC}_H = \left( \sigma_M^2 + \sigma_C^2 + \sigma_H^2 \right) / \left( \sigma_M^2 + \sigma_C^2 + \sigma_H^2 + \pi^2/3 \right)
\]

where $\sigma^2$ 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 variance in the probability of dying that existed at a concrete 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. 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.

The median odds ratio (MOR). The MOR (Larsen & Merlo, 2005; Larsen, Petersen, Budtz-Jorgensen, & Endahl, 2000; Merlo et al., 2006a) is a measure of heterogeneity, rather than of clustering as with the VPC defined above. The MOR is an alternative way of expressing area/household variation that translates this variance to the widely used OR scale, which makes the MOR comparable with the OR of individual or area variables. In the absence of variation, the MOR = 1.

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

\[
\text{MOR}_M = \exp \left( 0.95 * \sqrt{\sigma_M^2} \right)
\]

\[
\text{MOR}_{CT} = \exp \left( 0.95 * \sqrt{\sigma_M^2 + \sigma_C^2} \right)
\]

\[
\text{MOR}_H = \exp \left( 0.95 * \sqrt{\sigma_M^2 + \sigma_C^2 + \sigma_H^2} \right)
\]

where $\sqrt{\cdot}$ is the square root of the variance ($\sigma^2$) at the specific level.
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 a very low/low educational level in the census tract. The higher the percentage of low and very low education, the higher the percentage of crude all-cause mortality, as well as the mean value of the predicted probability of death, and most variables related to socio-economic deprivation and impaired health (i.e. unemployment, disability pension, retirement, receiving a widow’s/orphan’s pension, and having only one dwelling).

Mortality was consistently higher in men than in women, but more men than women had a higher educational level and were working in 2001. As the percentage of people with a very low and low educational level increased, so the percentage of “movers” decreased, from 9% to 3%. Compared with “stayers”, the “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 a very low/low educational level in the census tract (OR = 1.99, 95% CI 1.92–22.07) (data not shown in the Table).

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 low socio-economic position and living alone, on the one hand, and increased mortality on the other. Some occupational categories directly related to impaired health, like needing help with basic activities or receiving a disability pension, were clearly associated with an increased mortality risk.

Specific contextual effects

Table 3 shows that the census tract percentage of people with a very low/low educational level variable was associated with a slightly increased mortality risk. Moreover, compared with Seville, we observed a somewhat higher average mortality in Huelva and Cadiz and a somewhat 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 4 shows the general contextual effects of the different levels of analysis. The table shows 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 municipal 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 municipal (VPC_M = 0.23) and the census tract level (VPC_T = 0.33).

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, their risk of dying would marginally increase (Model 2 MOR_M = 1.09). We observed similar results for the census tract level (Model 2 MOR_T = 1.10). These low values

Table 2

Association between individual socio-economic characteristics and all-cause mortality in the period 2001–2009 among men and women aged 45–79 years and residing in Andalusia in 2001. These individual variables were included in an equation for obtaining the predicted probability of death. Values are given as odds ratios (ORs) and 95% confidence intervals (CIs).

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Association between individual socio-economic characteristics and all-cause mortality in the period 2001–2009 among men and women aged 45–79 years and residing in Andalusia in 2001. These individual variables were included in an equation for obtaining the predicted probability of death. Values are given as odds ratios (ORs) and 95% confidence intervals (CIs).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men</td>
</tr>
<tr>
<td></td>
<td>OR 95% (CI)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>45–49</td>
<td>1.00</td>
</tr>
<tr>
<td>50–54</td>
<td>1.55 (1.40–1.71)</td>
</tr>
<tr>
<td>55–59</td>
<td>2.25 (2.04–2.47)</td>
</tr>
<tr>
<td>60–64</td>
<td>3.19 (2.90–3.51)</td>
</tr>
<tr>
<td>65–69</td>
<td>4.45 (4.03–4.92)</td>
</tr>
<tr>
<td>70–74</td>
<td>7.97 (7.20–8.83)</td>
</tr>
<tr>
<td>75–79</td>
<td>15.01 (13.52–16.65)</td>
</tr>
<tr>
<td>Civil status</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>1.00</td>
</tr>
<tr>
<td>Single</td>
<td>1.59 (1.49–1.69)</td>
</tr>
<tr>
<td>Widowed</td>
<td>1.36 (1.27–1.46)</td>
</tr>
<tr>
<td>Separated</td>
<td>1.79 (1.57–2.04)</td>
</tr>
<tr>
<td>Divorced</td>
<td>1.41 (1.14–1.73)</td>
</tr>
<tr>
<td>Educational level</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>1.00</td>
</tr>
<tr>
<td>Medium</td>
<td>1.02 (0.95–1.10)</td>
</tr>
<tr>
<td>Low</td>
<td>1.12 (1.06–1.19)</td>
</tr>
<tr>
<td>Very low</td>
<td>1.44 (1.32–1.58)</td>
</tr>
<tr>
<td>Housing tenure</td>
<td></td>
</tr>
<tr>
<td>Owned</td>
<td>1.00</td>
</tr>
<tr>
<td>Rented</td>
<td>1.39 (1.29–1.49)</td>
</tr>
<tr>
<td>Other form</td>
<td>1.16 (1.07–1.26)</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
</tr>
<tr>
<td>Working</td>
<td>1.00</td>
</tr>
<tr>
<td>Receiving formal education</td>
<td>1.44 (0.99–2.11)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>1.53 (1.41–1.67)</td>
</tr>
<tr>
<td>Receiving disability pension</td>
<td>3.13 (2.91–3.36)</td>
</tr>
<tr>
<td>Receiving pension for widows/orphans</td>
<td>2.10 (1.71–2.57)</td>
</tr>
<tr>
<td>Retired</td>
<td>1.99 (1.86–2.13)</td>
</tr>
<tr>
<td>Needing help with basic activities</td>
<td>8.77 (5.96–12.90)</td>
</tr>
<tr>
<td>Performing household tasks</td>
<td>2.04 (1.68–2.47)</td>
</tr>
<tr>
<td>Other occupations</td>
<td>1.92 (1.50–2.46)</td>
</tr>
<tr>
<td>Having only one dwelling</td>
<td>1.31 (1.23–1.39)</td>
</tr>
</tbody>
</table>

Estimations

Starting with restricted iterative generalised least squares (RIGLS) estimations (Goldstein, 2003), we applied Markov chain Monte Carlo (MCMC) methods (Browne, 2009a) with parameter expansion at the household level (Browne, 2009b), with a burn-in of 5000, a chain length of 50,000, and a thinning interval of 10. More technical and conceptual information on these methods is available in the work by Browne (2003).

We obtained the median, 2.5% and 97.5% values of the posterior distribution, to calculate the point estimates of the parameters and their 95% credible intervals (Browne, 2003). Readers unfamiliar with the 95% credible interval can interpret the credible interval in an analogous way as the confidence interval.

We compared models using the Bayesian deviance information criterion (BDIC), and considered a reduction in BDIC > 10 as an indication of a better fit (Spiegelhalter, Best, Carlin, & Van der Linde, 2002).

We performed the analyses using SPSS version 18 (SPSS Inc., Chicago, IL, USA) and MLwiN version 2.22, The Centre for Multilevel Modelling, University of Bristol.
Table 3
Multilevel analysis showing the specific associations between individual and contextual characteristics and all-cause mortality in the 8-year follow-up (2002–2009) among men and women aged 45–79 years, residing in Andalusia at the end of 2001. Values are given as odds ratios (ORs) and 95% confidence intervals (CIs).

<table>
<thead>
<tr>
<th>Specific individual effects</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men v. women</td>
<td>1.12</td>
<td>1.12</td>
<td>1.12</td>
<td>1.12</td>
</tr>
<tr>
<td>Predicted probability of death&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Decile group 1 (lowest risk)</td>
<td>1.04</td>
<td>1.42</td>
<td>4.21</td>
<td>4.21</td>
</tr>
<tr>
<td>Decile group 10 (Highest risk)</td>
<td>58.15</td>
<td>57.4</td>
<td>57.4</td>
<td>57.4</td>
</tr>
<tr>
<td>“Stayers” v. “movers”</td>
<td>1.07</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specific contextual effects</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census tract low educational level</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Quartile 1 (Lowest percentage)</td>
<td>1.12</td>
<td>1.12</td>
<td>1.12</td>
<td>1.12</td>
</tr>
<tr>
<td>Quartile 2</td>
<td>1.12</td>
<td>1.12</td>
<td>1.12</td>
<td>1.12</td>
</tr>
<tr>
<td>Quartile 3</td>
<td>1.12</td>
<td>1.12</td>
<td>1.12</td>
<td>1.12</td>
</tr>
<tr>
<td>Quartile 4 (Highest percentage)</td>
<td>1.13</td>
<td>1.14</td>
<td>1.14</td>
<td>1.14</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seville</td>
<td>1.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cadiz</td>
<td>1.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huelva</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jaen</td>
<td>0.98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Almeria</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Córdoba</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Granada</td>
<td>1.02</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> According to a logistic regression modelling all-cause mortality in 2001–2009 and including as predictor variables individual age, civil status, educational level, housing tenure, having only one dwelling, and occupation (see also Table 2).

Table 4

<table>
<thead>
<tr>
<th>Variance, median (95% credible interval)</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Municipality (σ&lt;sup&gt;2&lt;/sup&gt; M)</td>
<td>0.015 (0.009–0.023)</td>
<td>0.008 (0.004–0.014)</td>
<td>0.009 (0.005–0.015)</td>
<td>0.006 (0.002–0.011)</td>
</tr>
<tr>
<td>Census tract (σ&lt;sup&gt;2&lt;/sup&gt; CT)</td>
<td>0.076 (0.064–0.091)</td>
<td>0.010 (0.000–0.005)</td>
<td>0.003 (0.001–0.010)</td>
<td>0.007 (0.003–0.013)</td>
</tr>
<tr>
<td>Household (σ&lt;sup&gt;2&lt;/sup&gt; H)</td>
<td>0.660 (0.582–0.743)</td>
<td>0.310 (0.235–0.388)</td>
<td>0.309 (0.226–0.391)</td>
<td>0.301 (0.220–0.383)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance partition coefficient (VPC)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VPC&lt;sub&gt;M&lt;/sub&gt;</td>
<td>0.4%</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.2%</td>
</tr>
<tr>
<td>VPC&lt;sub&gt;CT&lt;/sub&gt;</td>
<td>2.3%</td>
<td>0.3%</td>
<td>0.3%</td>
<td>0.4%</td>
</tr>
<tr>
<td>VPC&lt;sub&gt;H&lt;/sub&gt;</td>
<td>18.6%</td>
<td>8.9%</td>
<td>8.9%</td>
<td>8.7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Median odds ratio (MOR)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MOR&lt;sub&gt;M&lt;/sub&gt;</td>
<td>1.12</td>
<td>1.09</td>
<td>1.09</td>
<td>1.08</td>
</tr>
<tr>
<td>MOR&lt;sub&gt;CT&lt;/sub&gt;</td>
<td>1.33</td>
<td>1.10</td>
<td>1.11</td>
<td>1.11</td>
</tr>
<tr>
<td>MOR&lt;sub&gt;H&lt;/sub&gt;</td>
<td>2.29</td>
<td>1.72</td>
<td>1.72</td>
<td>1.71</td>
</tr>
<tr>
<td>Bayesian deviance information criterion (BDIC)</td>
<td>163,450</td>
<td>135,684</td>
<td>135,657</td>
<td>135,626</td>
</tr>
<tr>
<td>BDIC change compared to the previous model</td>
<td>−27,776</td>
<td>−27</td>
<td>−27</td>
<td>−31</td>
</tr>
</tbody>
</table>

<sup>a</sup> The value of the variance of the underlying individual-level variable, according to the logistic distribution π<sup>2</sup>/3 or 3.29 in all models.
much higher for modelling individual low educational level (VPCCT = 24%) than for mortality.

Fig. 2 is similar to Fig. 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 in the provinces, with considerable overlap between the province with the highest (Huelva) and the province with the lowest mortality (Córdoba).

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 be of minor relevance for individual inequalities in mortality risk. Our conclusion critically questions previous ecological spatial studies of small-area variation published in Spain (Benach et al., 2003, 2004; Borrell et al., 2010; Gobierno de Aragon, Departamento de Salud y Consumo, 2010; Ocana-Riola & Mayoral-Cortes, 2010; Ocana-Riola, Mayoral-Cortes, et al., 2008) and, by analogy, in many other parts of the world. These ecological spatial studies of small-area variation are only justified if we implicitly assume that the geographical variation between municipalities and also 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.

General contextual effects are based on variance, and they show the possible influence of a context on individual mortality risk, without specifying any contextual characteristics other than the definition of the boundary that embraces the context (Merlo et al., 2009). Geographical variance is also analysed in customary ecological spatial studies of small-area variation. 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 multilevel analysis relates the geographical/contextual variation to the total individual-level variation, as we did in our analyses.

Applying multilevel regression analysis, we observed that the crude spatial variation at the municipal and census tract levels represented only 0.4% (i.e. VPCM = 0.4%) and 2.3% (VPCCT = 2.3%), respectively, of the total individual variance in the probability of death. Moreover, these minor percentages practically disappeared (VPCM = 0.2% and VPCCT = 0.3%) when taking into account the individual composition of the areas (including 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 fit of the model, as the BDIC decreased by 27,776 units.

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

We found that a considerable proportion of the total individual variance in the probability of dying was at the household level (VPCH = 18.6%). This intra-household correlation possibly reflected the existence of shared genetic and environmental factors (Lawlor &
Mishra, 2009), as well as positive assortative mating (Epstein & Guttman, 1984). The existence of a similar genetic background, as well as learnt behaviours like eating and drinking habits, attitudes to physical activity, and coping mechanisms, may result in 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 may explain part of the household clustering of mortality (Mineau, Smith, & Bean, 2002).

While the familial household is a naturally defined context that influences individual mortality risk (Lawlor & Mishra, 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 affects mortality (Boyle & Willms, 1999). Obviously, we cannot expect that this complex system of intricate boundaries is properly captured by simple geo-administrative criteria. Apart from the familial household, other contexts, like the workplace where people spend most of their active time, may have a greater impact on mortality than the current area of residence (Muntaner, Li, Ng, Benach, & Chung, 2011). Possibly, a multiple membership multiple classification (MMMC) analysis (Browne et al., 2001), which considers both residential areas and workplaces 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 this approach performed so far in Scandinavia have found similar results to those found in Andalusia (i.e. minor clustering of mortality within administratively defined small areas) (Naess, Claussen, Smith, & Leyland, 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 in the administrative geographical areas is not surprising. Many multilevel analyses of neighbourhood effects have considered variance a nuisance or viewed it as a measure that only provides secondary information (see, e.g., Diez Roux, 2008). The reasons for this scepticism are not clear. It is frequently argued that there are technical difficulties in measuring variance in multilevel logistic regression, which prevent the substantive interpretation of variance (Blakely & Subramanian, 2006). Nevertheless, this technical argument is no longer justified as, today, components of variance and clustering can be confidently calculated in generalised linear multilevel models (Browne et al., 2005; Goldstein et al., 2002; Larsen & Merlo, 2005; Li et al., 2008).

Another argument is 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 fairly common in multilevel analysis and, if properly interpreted, provides enhanced information on contextual effects (see Merlo et al., 2009, for a longer commentary). However, the existence of minor general contextual effects may originate publication bias (Siddiqi, 2011). In other words, authors only publish or comment on contextual effects because they are statistically significant and in harmony with preconceived beliefs on contextual effects. Publication bias can give a false impression about the influence of administrative geographical boundaries on individual health, thus misleading 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 such as living in a census tract with a high percentage of people with a low educational level.

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 “decontextualised” (Clarke & Wheaton, 2007).

We detected a small but conclusive association between a high percentage of people with a low educational level in the census tract and individual mortality risk (ORlowest v. 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 level in the most deprived census tracts would reduce mortality risk for all individuals exposed. However, the overall geographical clustering of mortality was very low (VPCCT = 0.3%), and the census tract-educational variable did not explain much of the — initially minor — geographical variation. If we compared one individual from the highest with another individual from the lowest quartile of the 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 socio-economic position (like the educational variable used in our study) raise methodological concerns regarding the ability of these variables to actually estimate contextual effects “independently” of the corresponding individual-level variable (Mujahid, Diez Roux, Morenoff, & Raghunathan, 2007). The small association observed between the aggregated educational variable and mortality may just be a reflection of the individual-level association.

More elaborated contextual variables such as a composite index of area deprivation (Carstairs, 1995; Ocana-Riola, Saurina, et al., 2008), or the application of the econometric methodology (Chaix, Lindstrom, Rosvall, & Merlo, 2008; Fone, Farewell, & Dunstan, 2006; Messer, 2007; Mujahid et al., 2007; Raudenbush & Sampson, 1999) appear conceptually to be more appropriate strategies than the simple aggregate measures of socio-economic 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 using the percentage of people with low educational level at the census tract as we have done.

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) had a slightly lower average mortality (Table 4, Model 4). Among others, a possible explanation for these mortality differences is that the provincial strategies for public health and health care have different effectiveness, since in Andalusia each province is an independent health care area. 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, compared an individual from Seville with an individual from Huelva, we would not be able to predict which of them would have the higher mortality risk, even if, on average, Huelva has a slightly higher mortality rate than Seville. Fig. 2 in part illustrates this as it shows that the mortality in the census tracts within the province with the highest mortality (Huelva) and the province with lowest mortality (Cordoba) overlap considerably with each other.

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

Conclusions

Our study demonstrates that specific and general contextual effects in multilevel analyses need to be interpreted simultaneously. Otherwise, specific contextual effects may lead to misleading conclusions. We need to consider associations between means as well as the heterogeneity around these average measures (Braumoeller, 2006; Downs & Roche, 1979; Gould, 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 & Longford, 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 levels of measurement and analysis are needed (Morgenstern, 1998, p. 480).

Our investigation, however, expands this critique beyond the "ecological fallacy" (Morgenstern, 1998; Robinson, 2009; Subramanian et al., 2009) and emphasizes the inability of ecological analyses to quantify general contextual effects because they lack information on how variance is partitioned across the different levels of analysis (e.g. the individual and area levels) (Merlo, 2003; Merlo et al., 2009; Subramanian et al., 2007).

Modern ecological analyses of small-area variation apply state-of-the-art spatial analyses and Bayesian estimation methods to model variation in mortality between different geographical areas. However, these ecological studies have several interpretative limitations. First, it is not enough to consider whether spatial variation is greater than would be expected by chance (Ibanez et al., 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 influenced by geographical/contextual factors at the small-area level. Our multilevel analysis challenges this and suggests that administrative geographical boundaries are inappropriate for embracing the relevant contexts that influence mortality. However, we insist, other geographic or cultural contexts may be important and need to be identified.

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 can even give misleading information. Our conclusion concerns all-cause mortality. Nevertheless, our study findings may be relevant for many other small-area variation studies performed around the world.

Individual health is not only an individual responsibility; it 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 to be properly interpreted. From the perspective of public health policy, we believe that, rather than blaming concrete administrative geographical areas for their mean mortality rates, improving social and economic opportunities all over Andalusia we will lower individual mortality risk and, simultaneously, the current geographical socio-economic segregation.

Acknowledgements

We aim to specially acknowledge the indispensable work performed by the people at the Instituto de Estadística y Cartografía de Andalucía, which has created and maintains the Longitudinal Database of the Andalusian Population (LDAP). Our research is supported by the Swedish Council for Working Life and Social Research (FAS) [Dnr: 2010-0402, Pl: JM] and the Swedish Research Council (VR) [Dnr K2011-69X-15377-07-6, Pl: JM].

References


