Urban Environment and Mortality Differentials in Spain

How does residential environment affect mortality in South Spain?

Mathias Voigt[1], Dariya Ordanovich[2], Fransico Viciana[3], Laura Cilek[1], Rosa Cánovas[3], and Diego Ramiro[1]

[1]Center for Humanities and Social Sciences, Spanish National Research Council

[2]ESRI Spain

[3]Institute of Statistics and Cartography of Andalusia

Acknowledgements: We thank the Institute of Statistics and Cartography of Andalusia, Spain, for providing generously their data and time.

Funding: This work was supported by the LONGPOP project which has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 676060 and the project LONGHEALTH CSO2015-69834-R funded by the Spanish Ministry of Economy and Competitiveness.

Abstract

Background: The accelerating urbanization in many developing countries like India or China revived the interest in the mechanisms behind the impact of an urban environment on health and mortality. Whereas analyses on the mechanism behind urban-rural differences are either focused on vastly industrializing countries or historic populations, it is assumed that urban dwellers in highly developed countries are exposed to a mortality advantage. Estimating the effect of small area differences in the degree of ”urbanicity” and area development on long-term mortality risk in Andalusia, can give new insights on small area effects on individual well-being.
Method: Census tract based information from a geographical data base are linked to a register-based mortality follow up study and the Spanish Population and Housing Census of 2001. After controlling for spatial autocorrelation, a mixed effects Cox Proportional Hazards model is applied which accounts for area variation and the environmental impact of neighboring census tracts.
Results: Estimates suggest that the degree of ”urbanicity” in a shared area affects individual hazards of mortality. Once population and environment based effects are incorporated, the effect disappear and estimated effects suggest a 4.5% increased mortality with every unit increase in the environmental risk factor index.
Conclusion: The application of construction and population-based measures aggregated at a census tract level reveals area-related mortality disparities which persists despite controlling for individual level difference including socioeconomic position. Estimated hazard ratios indicate the existence of indirect effects of residential environment features on mortality beyond rural/urban differences.

key words: Urban Population, Health, Public Health, Epidemiology, Small Area Analysis

1 Introduction

The world is urbanizing at increasing speed. According to United Nations projections, more than two thirds of the world population will live in areas classified as urban by the year 2050 [1]. Accelerated growth of cities in predominantly developing countries will probably trigger various changes in their social structure, occupational activities, and distribution of wealth. Such an development also entails enormous challenges regarding social equity and long-term public health development [2, 3]. Urban dwellers in vastly urbanizing countries like China seem to face a double burden on their health which is induced by a combination of increasing environmental hazards and the adaption of unhealthy behaviors like smoking and high carbohydrate diets [4, 5]. Recent public health studies suggest that various unfavorable health conditions including obesity, high cholesterol, or forms of depressions are much more prevalent among individuals living in cities compared to their rural counterparts [6, 7]. Regarding the remarkable pace at which urban areas are currently developing and in contrast to the past [8–10], health and mortality disparities between rural and urban residents are rather moderate if not beneficial for inhabitants of cities. Although urban dwellers are exposed to city-related risk factors including higher air pollution, differences in mortality or disease prevalence compared to residents in rural areas have become marginal or disappeared in modern developed and most developing societies [7]. Acknowledging the improvements in care and prevention induced by public health interventions as well as the favorable socioeconomic development of urban areas within the last century [11], it is even tempting to ask if concepts like the ”urban penalty” or the urban sprawl approach [12] are rooted in a lack of individual level data or oversimplified measurement strategies regarding the complex interaction between environmental features and individual health. While traditional comparisons of rural and urban subpopulations commonly fall back on dichotomous measures to distinguish between the two area types, the interaction between environmental features and health at the individual level is often masked by social class and the clustering of more vulnerable minorities in cities [12]. Considering possible measurement errors and additional positive aspects of urban living on health, like the generally higher proximity to hospitals and specially qualified doctors, there appear to be doubts that urban residents are exposed to substantial disadvantages.

On the other hand, it is somewhat odd to assume that the environment we live does not affect our health, as inhabitants of many modern cities are expose to specifically urban risk factors like high air pollution, lack of green space or otherwise stressful environmental features which theoretically have a negative impact on their health or even their risk of dying. Even if majority of health and survival disparities can be traced back to behavioral, socioeconomic or biological differences, the exposure to environmental hazards complements to the analysis of individual risk factors substantially [13]. There are numerous studies suggesting unfavorable health effects with increasing spatial proximity to areas with high exposure even to arguably less obvious geographical or environmental features like natural gas wells or high share of nano material [14–16]. Thus, it can be assumed that health and well-being of individuals who are clustered in neighborhoods or other forms of small areas are influenced by the shared exposure to certain stress factors or environmental features. Such effect might occur independently from their individual characteristics and be mediated by the degree of ”urbanicity”, a concept developed to distinguish between different degrees of urbanization of an area [5].
By analyzing potential effects of environmental impacts and specifically urban characteristics on individual level survival over time, we aim to contribute to the debate on urban-rural differences as well as the field of small area analysis. Following the introduction of the conceptual and methodological framework with specific focus on the measurement of latent concepts of ”urbanicity” and environmental impacts, the data infrastructure and the constructed indexes to measure the latent concept are explained in further detail. A Cox proportional hazards regressions with mixed effects is applied to model individual level mortality risks in dependence of exposure to urban environment and other area factors over time and results are compared to alternative models before we conclude and discuss the main findings, possible limitations, and prospective follow-up analysis.

2 Background

2.1 Measuring *Urbanicity* and Small Area Environment

Techniques for analyzing mortality or health differentials between rural and urban sub-populations have evolved with an increasing availability of spatial data, the possibility to manage large amounts of data within reasonable time, and advanced measurement strategies. Most classic studies apply a dichotomous indicator to distinguish between urban and rural areas and commonly rely on a core set of characteristics including population size within somehow specified areas, most commonly based on administrative boundaries [5]. Such traditional approaches for classifying areas by population size or density fail to incorporate relevant urban characteristics related to infrastructure or distribution of green space [17]. Moreover, a dichotomous classification based on administrative boundaries might be tempting for analysing rural-urban differences, but does not account for the heterogeneity within administrative boundaries. In fact, cities are often surrounded by heavily populated areas which might not be part of the same administrative area but are still exposed to similar conditions. Hence, urban sociologists and other research communities have been advocating for a more nuanced approach to the subject matter [18]. In answer to one of their calls, Vhalov et al. (2002) [5] proposed a more refined conceptualization by disentangling two related concepts, urbanization and "urbanicity", a term they coined in their paper. While urbanization is defined as a process of growth of cities in terms of area and population over time, "urbanicity" relates to a state of an area which differs by the degree of certain urban characteristics, or in other words, it the nature of urban environments [4]. As these concepts are strongly dependent on the regional context and often change over time [19], the lack of precise definitions does not only affect comparability in an exclusively negative way. It also allows one to propose new measures or adapt to regional differences which can then be compared over time.

2.2 Andalusia

Although the most recent economy crisis and related reductions of health and social service system budget were expected to slow down mortality improvements in Spain [20, 21], short-term overall survival in the post-crisis years has even improved at a faster rate than before. Whereas some studies found an elevated risks for suicide and mental health issues correlating with economic downward trends and increasing unemployment [22, 23], mortality by most causes decreased after 2008 [24].

Andalusia is the southernmost and with about 8.3 million inhabitants (2016) most populated of the 17 Spanish autonomous communities [25]. Although the predominantly rural region has experienced economic bottlenecks and strong selective outmigration of young healthy individuals throughout the twentieth century, in recent decades, not only economic and social indicators but also Andalusian mortality rates approximated the Spanish average . Analyses of small area differences have indicated that currently only a group of municipalities in the south west of the community is exposed to higher mortality rates than the national average and mostly contributed by older individuals [26, 27]. With its historically rather vulnerable economy, Andalusia was hit extraordinarily hard by the recent financial crisis which has led to extreme job loss and a continuously increasing at-risk-of-poverty rate which reached with 35.4% a high point far above the Spanish national average (22.2%) in 2016 [28].

3 Data and Method

3.1 Data

The *Base de Datos Longitudinal de Población de Andalucía* (BDLPA) is a comprehensive data infrastructure based on administratively collected information of a synthetic cohort [1] first observed in the year of the Spanish population and housing census 2001 and followed up until today. Access to the mortality follow-up of a 10 % sample which was applied for the time-to-event analysis in this contribution is provided by the Institute of

[1] The synthetic cohort is based on all individuals registered for the population and housing census of 2001 which covered about 95% of the Andalusian population at the time.

Statistics and Cartography of Andalusia (IECA) [2]. A advantageous property of the BDLPA is the possibility for joint applications with other administratively collected information which allowed us to cluster individuals by residential census tract. The geographical information is obtained from the CORINE land-cover rater data base [3] and maps for exploitative spatial analysis were provided by the cartography unit of the IECA. To achieve comparability and assure individuals and single households cannot be identified, the information was provided at census tract level according to the definitions of Spanish population and housing census of 2001. Such clustering allows additionally for links between the aforementioned longitudinal mortality register and aggregated population-based information from the census itself.

3.1.1 Indicator of Urban Environment

The degree of ”urbanicity” ,or in other words the degree of how urban an area is, can be considered as multi-dimensional and latent concept. Different degrees of urbanicity cannot be measured directly and potential effects on individual health would be indirect. To account for the difficult conceptualization, a mix between theoretical and data-driven approach was chosen to construct a multi-component index which allows to distinguish between different degrees of urban environments [4, 5, 29]. After examining graphical tests and correlation coefficients between all accessible environmental variables related to urban settings, four variables are identified as main contributers to the latent concept ”urbanicity” [6]. Population density as measure of relative crowdedness by census tract was standardized, and weighted based on the overall deciles to assure comparability with other scale components. The average coverage with medical service is incorporated by estimated service area polygons which represent the distance that can be covered between a health facility and any point on the map within 30 minutes (driving-to-facility time) [4]. The population accounted area of artificial surface area per census tract is estimated based on CORINE land cover classified raster data for the year 2006 [5]. Road density is obtained by estimation of lengths of geographical line objects representing classified roads according to the generalized transport network by unit area (km per sq.km). The weights with which the single components enter the multicomponent index variable are estimated through maximum likelihood factor analysis incorporating standardized single component values [30, 31]. Factor weights are represented in table one. The resulting index variable is further normalized to be centered around zero. A Crohnbachs α score of 91% suggests sufficient internal consistency of indicator components. First graphical quality tests of the index is depicted in figure one which contains a colored map of Seville, the population-wise biggest city in Andalusia [32]. Scores for the multi-component indicator for urban environment are represented by census tract and darker colors represent a higher degree of ”urbanicity”. Highly urban areas seem to be located in the city center and are rather small indicating a good graphical fit

[2] The mortality and emigration follow-up of a 10% sample of the census based population can be accessed through the protocol on the website: www.juntadeandalucia.es

[3] www.europedataportal.eu

[4] The estimation is conducted with the *Network Analyst extension (ArcMap 10.5)*. To limit the polygons to inhabited areas, we intersect them with a 250×250 meters population grid layer and by using the basic geometric calculations in ArcMap 10.5. yield the portion of populated service areas in square kilometers

[5] According to the classification, artificial surfaces include urban fabrics (continuous and discontinuous urban fabric surfaces), industrial and commercial areas (industrial or commercial units and transport units, road and rail networks and associated land, port areas and airports), mine and dump sites (mineral extraction sites and construction sites, dump sites, construction sites), and artificial non-agricultural areas (green urban areas vegetated areas, sport and leisure facilities). The scale is 1: 100 000

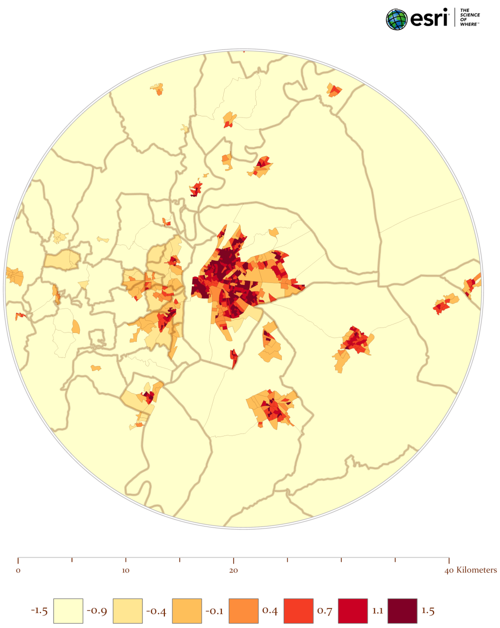

Figure 1: Municipality of Seville - Scores of the Urban Environment (left) and the Environmental Indicator (right) by census tract

3.1.2 Environmental and population-based area features

The rather constructive and physical ″urbanicity″ indicator does not allow to capture area-specific heterogeneity regarding other purely environmental hazards unrelated to ″urbanicity″ or potentially harmful population-based features. While the focus of this work is on potential effects of the residential area environment on individual level mortality, it is important to allow for heterogeneity within the urban and rural places as we observe them in reality. Population based measures and survey answers on the residential environment are aggregated by census section and are used to represent the latent socioeconomic position and area-specific environmental hazards. After a sensitivity analysis the perceived average cleanness, noise exposure, and air pollution are chosen to incorporate an environmental effect in the model, while the number of delinquencies, the proportion of employed individuals at working age, and the proportion of single households enter the model as population-based area features. All variables are obtained from the population and housing census of 2001 and standardized with reference to Andalusian mean values.

3.2 Sample Population and Individual Level Variables

There are two main sources for individual level information. The BDLPA is a quasi-continuous mortality follow-up which is semi-annually updated and corrected. By using an individual identifier, we are able to link all subjects in the follow-up study to their answers from the 2001 population and housing census. Information on the residential area and the individual socioeconomic context are therefore available for the baseline year of the study. To reduce the bias induced by potentially unobserved changes in residence and other individual time-varying information, only individuals between the ages 35 and 80 are included in the analysis. The selection of these age groups was furthermore based on informed assumptions about the household composition and other life course transitions within the age range based on common life course trajectories observed in Spain [33]. In general, individuals between age 35 and 80 have resided in their house or apartment for relatively long periods of time and are with increasing age more likely to own their dwelling. Hence, the probability of moving is rather low and somewhat assures exposure to the same environment for the time of our study and to some extent before. The choice to limit the age range of the study population is further motivated by the distribution of the event of interest as more than 90% of all deaths in Andalusia occur after age 35 but before age 92, the highest age an individual can reach at the end of the follow up [cf. 26, 27]. In consequence of the selection of age groups, the sample size decreases from 723,234 to 351,769 individuals.

To assure that neither the observed population nor the area-specific ”urbanicity” or environmental characteristics may have dramatically changed over 12 years of observation, particularly the light of substantial economic fluctuations since 2001 [20, 28], several sensitivity tests are obtained for different time periods.

One strength of this analysis, lies in the incorporation of individual level information in combination with area-specific factors in a multi-level structure. Such data structure guarantees that possible index effects are not caused by unobserved heterogeneity between sub-populations. The included individual level variables are sex, dependency status, and marital status at the baseline year. To control for socio-economic individual differences, we include several indicators for social position like highest educational degree, ownership status of the dwelling, and car ownership in our models. All socioeconomic variables are derived from the census questionnaire of 2001.

3.3 Method

The incorporation of area effects into an analysis of individual level mortality differences requires statistical testing for potential impacts of the cluster-specific effects and their spatial distribution. The graphical representation of the multicomponent ”urbanicity” index and other environmental variables suggest for example that observations are more probable to be similar if they are geographically closer to each other. Hence, these observations cannot to be assumed to be mutually independent from each other [34]. To assess the extent of spatial autocorrelation, intrinsic stationarity is assumed before calculating the row-standardized matrix of spatial weights based on the list of contiguous neighbors. At least one point of the boundary of a spatial polygon which represents a census tract has to be within snap distance of at least one point of a neighboring polygons’ border to meet the contiguity condition [35]. The product-moment correlation coefficient (Morans I) is then computed to estimate the probability of spatial correlation by central variables including the index for ”urbanicity”, smoothed standardized mortality ratios (SMRs), and the aforementioned area-specific features [36, 37].

3.3.1 Statistical Model

Mortality disparities by degree of ”urbanicity” and environmental impact are estimated applying an extension of the Cox PH model. Covariate effects on individual hazards enter the Cox model in a multiplicative fashion as expressed in the following equation [cf. 38].

$$h(t) = h_0(t) \exp(\beta_i \mathbf{X}_i) \tag{1}$$

, where h_0 is the baseline hazard and $\exp(\beta_i \mathbf{X}_i)$ the non-negative function of covariates. Hazard ratios are obtained through the maximization of the partial log likelihood with respect to $\beta_i \mathbf{X}_i$ [39, 40]. As we assume that environmental factors and the degree of "urbanicity" affect all individuals who are nested in a specific area independent from their individual characteristics, we chose a multi-level survival model which accounts for homogeneity within clusters by incorporating a stratum specific random effect [41]. The Cox PH model with mixed effects, which can be considered as shared frailty model, was recently made available for R-User [42] and proposed to extend the original Cox PH model with a normally distributed stratum specific frailty term $\mathbf{Z}_j$ as follows.

$$h(t_{ij}) = h_0(t) \exp(\mathbf{X}_i \beta + \mathbf{Z}_j) \tag{2}$$

, where $\mathbf{Z}_j$ is the design matrix for random effects which captures homogeneity within clusters. These random effect terms can be understood as relative effect of given covariate patterns on the baseline hazard which varies across census tracts [43]. Given the set-up of our analysis, the model also has to account for left truncation [44]. This adjustment affects survival estimates for everybody in the sample as their time under risk of dying before the start date of the study remains unobserved. Hence, they are selected based on their survival upon the start year of the examination.

To account for left-truncation and assure we measure age specific mortality differences, age is applied as time scale in our models. Calender time and cohort effects are accounted for by including the birth cohort as covariate.

4 Results

To account for possible mediating effects of the geographical distribution of outcome and explanatory variables, test univariate statistical tests for spatial autocorrelation are applied to the "urbanicity" index, the smoothed standardized mortality ratios and all environmental variables by census tract level. Interpretable as correlation between the variable and its spatial lag [45], univariate spatial autocorrelation is assessed through Morans product-moment correlation coefficient. Values for Morans I and the associated p-values are presented table one. Notably, the Null hypothesis that their geographical distribution is statistically random has to be rejected for all tested variables. Particularly the values close to one for the two indexes suggest that they are highly spatially autocorrelated.

Table 1: Morans I Statistics for univariate relationships of indicators with geographic dimension

	Morans I Statistic	*p-value*
Smoothed SMR	0.380	< 0.001
"Urbanicity" Index	0.784	< 0.001
Perceived Noise	0.570	< 0.001
Perceived Neighborhood Pollution	0.444	< 0.001
% Single Household	0.568	< 0.001
% Unemployed	0.500	< 0.001

4.1 Results - Mixed Effects Cox PH Model

As the analyses aims to highlight the impact of shared environmental factors on individual level survival, the population under observation is considered to be nested in geographical units which, hence, requires the application of a multi-level model structure. The estimated coefficients (fixed effects) of four separate Cox PH model with mixed effects with step-wise increased number of covariates are presented in table two and three respectively. All models are statistically different from the Null model and every subsequent model decreases the loss of information compared to the preceding model. The results for likelihood ratio test for the first model represent the comparison with a baseline model without random effects suggesting a significant improvement of model fit.

Table 2: Cox PH Model with mixed effects - Second level fixed environmental effects on individual hazards

	Dependent variable:			
	Hazard Ratios (95% CI)			
	(1)	(2)	(3)	(4)
"Urbanicity" Index	1.0061 (0.9963, 1.0158)	1.0308*** (1.0212, 1.0404)	1.0176** (1.0058, 1.0295)	1.0027 (0.9904, 1.0150)
Perceived Cleaness			1.0007** (1.0001, 1.0012)	1.0005* (0.9999, 1.0011)
Perceived Pollution			1.0020*** (1.0011, 1.0029)	1.0020*** (1.0012, 1.0029)
Perceived Noise			0.9999 (0.9991, 1.0008)	0.9995 (0.9987, 1.0004)
% Unemployed				1.0025*** (1.0013, 1.0036)
% Single HH				1.0094*** (1.0070, 1.0119)
Individual Variables		x	x	x
Observations	351,769	351,769	351,769	351,769
Integr. Log Likelihood	−542,763.5	−538,537.7	−538,510.2	−538,471.8
LR Test	128.61 (df = 2)***	8451.6 (df = 12)***	55 (df = 15)**	76.8 (df = 17)***
Note:				*p<0.1; **p<0.05; ***p<0.01

The estimated hazard ratio in model one indicates that the selected urban features of residential areas on its own do not substantially affect individual mortality differences. Socioeconomic and demographic individual level variables with well-documented indirect effects on mortality are incorporated in the second modeling step to account for within-area population heterogeneity. Although they are differentiating slightly from the Spanish national averages, the estimates for individual level variables show expected and repeatedly described patterns as shown in table two. Men between the ages of 35 to 92 are found to have a substantially higher relative risk of dying compared to their female counterparts. Individuals with functional limitations or other kind of disabilities are estimated to have a more than three times higher hazard during the observation period.

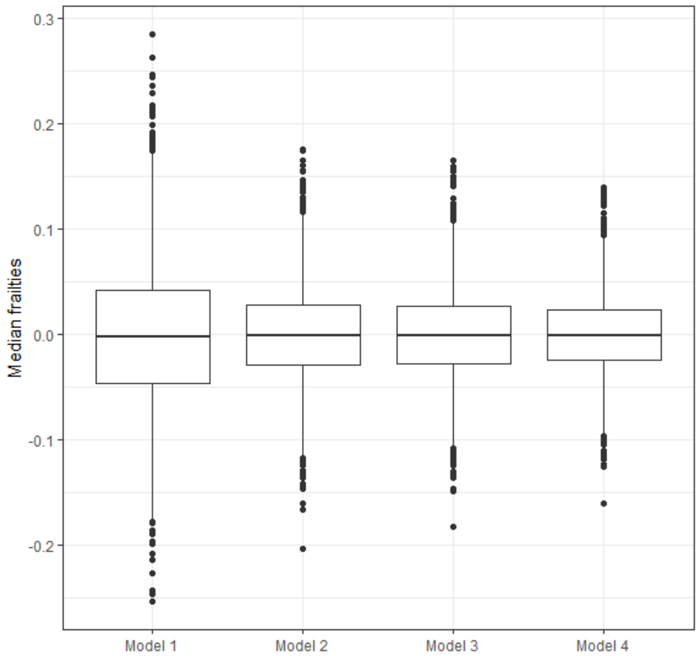

Figure 2: Random area effects (median frailty) by model

Moreover, all socioeconomic variables point towards an increased relative risk for less wealthy and educated individuals with reference to the more affluent or the ones with university education. As these estimated hazards for individual level risk factors change marginally with incorporation of more environmental and population-based area statistics, they are depicted in a separate table to avoid distraction from the effects of primary interest.

With the inclusion of individual level differences the effect of urban environment seem to become more pronounced than in the first model. In the second model every unit increase on the ”urbanicity” scale in a census tract is estimated to be associated with a significant 3 percentage point increase in the hazard of dying for its inhabitants. While the changes in the individual level impact factors are negligible between the models, the effect of the degree of ”urbanicity” on survival changes with incorporation of the additional area-level characteristics. In model three, we attempt to control for a different kind of heterogeneity between urban areas with the same degree of ”urbanicity” by incorporating the effects of perceived cleanness, air and water pollution, and noise. The estimates suggest that including these environmental area features reduces the effect of the degree of ”urbanicity” to 1.7 percentage points by unit increase. Both cleanness and pollution seem to have highly significant but seemingly small effect on survival. While perceived cleanness of the area is estimated to increase the hazard by 0.07 % by one unit increase, one percent of more perceived pollution is suggested to increase the hazard by 0.2 % per unit increase.In model four the estimated hazard for the ”urbanicity” indicator is very close to one as we additionally account for the population-based characteristics of the small areas. As the effect of physical urban environment shrinks, the percentage of unemployed individuals and the percentage of single households are estimated to increase the hazard by 0.25 and 0.94 percentage points respectively with every one percentage point increase. The impact of environmental area features on the other hand does not seem to be affected by the additional population-based measures incorporated in this model.
Going into more detail, we estimate possible effects of single indicator components on the mortality hazards in a given census tract. The results are depicted in figure two and indicate that road density seem to affect

survival negatively in a model with mixed effects without additional area variables while the percentage of artificial surface is suggested to have a slightly positive effect. Just as for the index variable, the effect of the single components vanishes when further environmental and social variables are controlled for. Such a result puts the presumable negative effects of road density, urban heat islands and other explicitly urban risk factors into perspective as the estimates indicate the existence of heterogeneity between areas with the same level of ”urbanicity”.

An advantageous property of shared frailty models over classic survival approaches is the possibility to estimate relative effects of covariate patterns on the baseline hazard across clusters. Assuming that subjects are exposed to shared environmental risk factors which affect their survival in spite of their individual level differences, we incorporated a normally distributed random effect for the residential area in the model. The estimated median frailty and its variation for all models are depicted in figure three. Naturally, the random variation is lower for models where we account for the shared additional area effects. Nevertheless, if translated into risk scores there will still be substantial differences between census tracts. For model four, a value of 0.1125, which is about one standard deviation above the mean, corresponds to a relative risk of 1.119, an almost 12 percentage point increase compared to the census tract at the mean. Further variation measures can be found in table five in the appendix. There we also provide a summary of the likelihood ratio tests between the particular model and a model with the same covariates but no random effects. The test statistics indicate that the model with additional shared frailties improves the fit significantly for all tested models. Figure four is a graphical depiction of shared frailty effects by census tracts in Andalusia. The values are exponentiated and can be understood as risk scores. The darker the color the higher is the unexplained relative mortality risk of individuals in the respective area.

Discussion

Urban life plays out in a variety of interacting settings which can affect health and mortality through multiple direct and indirect channels including the housing situation, accessibility to health centers and markets, and the cultural milieu. Hence, health of an urban population is theoretically influenced by dynamic interactions between various urban and environmental features.Hence, it is arguably more straightforward to focus on specific living conditions rather than macro level factors [12]. There are several studies on area differences between rural settings and areas with strong urbanization tendencies which findings suggest a noteworthy increase in inequality between rural and urban subpopulations. While some researchers seem to find an urban health penalty [46], others suggest health advantages for the urban population [47]. The complex relationship between health, mortality, and the residential areas features is not only heavily time varying but dependents on the overall cultural and social context as well. Moreover, these effects might work in both directions as living in cities is associated with harmful and health-promoting environments at the same time. As the adaption of unfavorable western diets in China, for instance, raises concerns about the future health of the urban population, vaccinations rates, access to hygiene, and health care, on the other hand, are significantly better in the Chinas urban centers [3].

To our knowledge, there are only few studies which address disparities between rural and urban population in larger regions of todays high income countries which might be related to the much lower rates at which

Table 3: Cox PH Model with mixed effects - Individual fixed effects corresponding to models in table 2

	Dependent variable:		
	Hazard Ratios (95% CI)		
	(2)	(3)	(4)
Male	2.0872***	2.0883***	2.0898***
	(2.0682, 2.106)	(2.0692, 2.1074)	(2.0707, 2.1089)
Reference: Female			
Physically Dependent	3.0286***	3.0180***	3.0028***
	(2.9617, 3.0954)	(2.9512, 3.0848)	(2.9360, 3.0695)
Reference: No Dependency			
Single	1.412***	1.4126***	1.4038***
	(1.3826, 1.4422)	(1.3829, 1.4424)	(1.3740, 1.4336)
Widowed	1.1836***	1.1823***	1.1808***
	(1.1586, 1.2085)	(1.1573, 1.2073)	(1.1559, 1.2058)
Divorced/Separated	1.4794***	1.4765***	1.4768***
	(1.4249, 1.5340)	(1.4219, 1.5310)	(1.4222, 1.5314)
Reference: Married			
No or Incomplete Educ.	1.3798***	1.3837***	1.3942***
	(1.3387, 1.4210)	(1.3425, 1.4249)	(1.3529, 1.4357)
Primary/Secondary Educ.	1.1595***	1.1598***	1.1652***
	(1.1155, 1.2036)	(1.1157, 1.2038)	(1.1211, 1.2093)
Reference: Tertiary Educ.			
Does not Own House	1.1625***	1.1537***	1.1399***
	(1.1362, 1.1888)	(1.1274, 1.1801)	(1.1134, 1.1664)
Reference: Does Own House/Apartment			
Does not Own a Car	1.2690***	1.2693***	1.2653***
	(1.2493, 1.2887)	(1.2497, 1.2890)	(1.2456, 1.2851)
Reference: Does Own Car(s)			
Observations	351,769	351,769	351,769
Note:			*p<0.1; **p<0.05; ***p<0.01

these cities are growing compared to China and other vastly urbanizing countries. Even if the change may be incremental, attractive job markets in or near urban centers and the growing demand for service work have let to a changing population composition between rural and urban settings with regard to the age distribution, education, and other wealth parameters. Consequential population movements and area developments will lead to environmental changes which in turn will affect the population health. Although the rather comprehensive health insurance coverage in Spain and other European countries might provide large-scale health and survival disparities, the existence of environmental risk factors, some specifically urban or rural, which affect peoples health in a accumulative and often indirect fashion can help to explain and ideally prevent an opening health and mortality gap between particular subpopulations. Therefore, it will be necessary to analyze phenomenons like urban heat islands where the asphalt and other artificial surfaces will store the summer heat and create a substantially warmer environment during the night [48], for instance, would particularly affect individuals residing or working in urban centers.

Presumably the greatest obstacles for analyzing differences in survival and disease prevalence by environmental characteristics, is the lack of a standard measurement strategies. As mentioned above, the latent concept "urbanicity" is multidimensional, dependent on the cultural and social sphere and often time varying [17]. In this contribution, we choose to disentangle the physical from the social and environmental sphere and generate a multicomponent scale indicator constructed to capture the gradient from rural to urban. The index allows for distinguishing more precisely between different degrees of "urbanicity", and can be understood as improvement over classic binary measures. Precision with regard to the area size is an additional advantage over comparable approaches, as for example the rurality index by [49], where data was aggregated at municipality level. An index based on census tract level information unarguably reduces the risk of misclassifying large areas as urban if, for instance, only a share of the overall area exhibits urban features. Moreover, the scale indicator allows for capturing multiple dimensions of urban space which will make it possible to distinguish between highly populated areas with different urban features like access to public transport or health care facilities. Results from mixed effects Cox PH models suggest that individuals residing areas with higher levels of unemployment, single households, and perceived pollution face small but highly significant survival disadvantages even after controlling for individual level risk factors. While population-based and environmental factors seem to explain geographical survival differences in modern-day Andalusia, we found no clear evidence for an effect of the physical urban environment captured through the aforementioned "urbanicity" index. The negative impact of degree of "urbanicity" on survival seem to disappear as soon as other small area variables are incorporated into the model indicating that the physical sphere of the concept "urban" seem to mask effects in the other spheres. Although there seemed to be a small effect of the more precise physical measures on individual level survival in the first place, the index does not help us to explain small area differences in mortality in Andalusia if additional information on environmental and social measures are controlled for. The effects estimated for these population-based and environmental factors on the other hand need to be investigated in greater detail to draw insightful conclusions on potential risk factors in different types of residential areas and their effect on growing inequalities in individual level survival.

The size of such effects may have been influenced by possible threats to validity related to the available data, unobserved mediators, or the general assumptions which were necessary to conduct this analysis. Whereas for example the assumption that individuals are unlikely to change their residence after age 35 can be justified with general life course trajectories in the context of Andalusia [33], values for other central explanatory variables including the ”urbanicity” indicator and some of the individual level information almost certainly have changed over time for parts of the small areas and the individuals. Since the extent of the variations of the census tract environment or individual socio-economic indicators like the ownership of cars cannot be estimated with the available single cross-sectional data point, we have to assume a relative constancy over time. As Andalusia was hit hard by the most recent financial and debt crisis, which has led to extreme job loss, a continuously increasing at-risk-of-poverty rate, and a high number of evictions [28, 50], the assumption of residential continuity might not hold for economically disadvantaged groups or individuals but cannot be tested with the available data. Moreover, there is no information about the average exposure to the environmental and social features of the residential area. The average amount of time someone spends in his or her residential area and is consequentially exposed to its environment will probably differ dependent on the persons age, employment status, and other unmeasured area features like access to ”third places” [51, 52].

Nonetheless, the results of our exploratory survival analysis not only contribute to the debate on urban health or mortality disadvantages, but also invite for implementing more refined examinations of spatial mortality and health differences in Southern Europe which further explore the mechanisms behind the spatio-temporal inequalities as prevalent in Andalusia [26]. To our knowledge, there is no other study which combines detailed small area information on census tract level with individual level variables and a survival follow up in Southern Europe. Although, we assume that we capture the most important individual and environmental measures for explaining overall mortality by allowing the effects to vary randomly, the results need to be contextualized by more refined studies which investigate the spatial composition of specific health conditions. Notwithstanding, our results suggest that environmental and arguably more important population composition features of residential small areas affect the survival probability of their residents. While a similar analysis at provincial level found an association between lower per-capita income and survival in Spain [53], our results seem to tell a different story. Although a more detailed measure of urban space might not contribute to the explanation of estimated spatial mortality differences, particular unfavorable area conditions as high levels of perceived pollution or a high percentage of unemployed co-residents are found to increase relative mortality even if other well-known protective individual characteristics might work in your favor.

References

[1] U. United Nations, “World Urbanization Prospects: The 2014 Revision,” report, Department of Economic and Social Affairs, Population Division, 2015.

[2] S. Allender, C. Foster, L. Hutchinson, and C. Arambepola, “Quantification of urbanization in relation to chronic diseases in developing countries: a systematic review,” *Journal of Urban Health*, vol. 85, no. 6, pp. 938–951, 2008.

[3] P. Gong, S. Liang, E. J. Carlton, Q. Jiang, J. Wu, L. Wang, and J. V. Remais, “Urbanisation and health in China,” *The Lancet*, vol. 379, no. 9818, pp. 843–852, 2012.

[4] D. L. Dahly and L. S. Adair, “Quantifying the urban environment: A scale measure of urbanicity outperforms the urbanrural dichotomy,” *Social Science & Medicine*, vol. 64, no. 7, pp. 1407–1419, 2007.

[5] D. Vlahov, S. Galea, and N. Freudenberg, “The urban health advantage,” *Journal of Urban Health*, vol. 82, no. 1, pp. 1–4, 2005.

[6] J. Jones-Smith and B. M. Popkin, “Understanding community context and adult health changes in China: development of an urbanicity scale,” *Social science & medicine*, vol. 71, no. 8, pp. 1436–1446, 2010.

[7] S. Eckert and S. Kohler, "Urbanization and health in developing countries: a systematic review," *World Health Popul*, vol. 15, no. 1, pp. 7–20, 2014.

[8] G. Kearns, "The urban penalty and the population history of England," *Society, health and population during the demographic transition*, pp. 213–236, 1988.

[9] D. Rosner, *Hives of sickness: public health and epidemics in New York City*. Rutgers University Press, 1995.

[10] R. Woods, "Urban-rural mortality differentials: An unresolved debate," *Population and development review*, pp. 29–46, 2003.

[11] D. Cutler and G. Miller, "The role of public health improvements in health advances: the twentieth-century United States," *Demography*, vol. 42, no. 1, pp. 1–22, 2005.

[12] N. Freudenberg, S. Galea, and D. Vlahov, "Beyond urban penalty and urban sprawl: back to living conditions as the focus of urban health," *Journal of community health*, vol. 30, no. 1, pp. 1–11, 2005.

[13] L. A. Waller and C. A. Gotway, *Applied spatial statistics for public health data*, vol. 368. John Wiley & Sons, 2004.

[14] P. M. Rabinowitz, I. B. Slizovskiy, V. Lamers, S. J. Trufan, T. R. Holford, J. D. Dziura, P. N. Peduzzi, M. J. Kane, J. S. Reif, T. R. Weiss, *et al.*, "Proximity to natural gas wells and reported health status: results of a household survey in Washington County, Pennsylvania," *Environmental Health Perspectives*, vol. 123, no. 1, p. 21, 2015.

[15] V. L. Colvin, "The potential environmental impact of engineered nanomaterials," *Nature biotechnology*, vol. 21, no. 10, p. 1166, 2003.

[16] J. E. Fergusson, *Heavy elements: chemistry, environmental impact and health effects*. Pergamon, 1990.

[17] S. Cyril, J. C. Oldroyd, and A. Renzaho, "Urbanisation, urbanicity, and health: a systematic review of the reliability and validity of urbanicity scales," *BMC Public Health*, vol. 13, no. 1, p. 513, 2013.

[18] T. W. McDade and L. S. Adair, "Defining the urban in urbanization and health: a factor analysis approach," *Social Science & Medicine*, vol. 53, no. 1, pp. 55–70, 2001.

[19] D. A. Leon, "Cities, urbanization and health," *International Journal of Epidemiology*, vol. 37, no. 1, pp. 4–8, 2008.

[20] A. Escolar-Pujolar, A. Bacigalupe, and M. San Sebastian, "European economic crisis and health inequities: research challenges in an uncertain scenario," *International Journal for Equity in Health*, vol. 13, no. 1, p. 59, 2014.

[21] H. Legido-Quigley, L. Otero, D. la Parra, C. Alvarez-Dardet, J. M. Martin-Moreno, and M. McKee, "Will austerity cuts dismantle the Spanish healthcare system?," *BMJ*, vol. 346, 2013.

[22] J. A. Lopez Bernal, A. Gasparrini, C. M. Artundo, and M. McKee, "The effect of the late 2000s financial crisis on suicides in Spain: an interrupted time-series analysis," *European Journal of Public Health*, vol. 23, no. 5, pp. 732–736, 2013.

[23] M. Gili, M. Roca, S. Basu, M. McKee, and D. Stuckler, "The mental health risks of economic crisis in Spain: evidence from primary care centres, 2006 and 2010," *Eur J Public Health*, vol. 23, 2013.

[24] E. Regidor, G. Barrio, M. J. Bravo, and L. de la Fuente, "Has health in Spain been declining since the economic crisis?," *J Epidemiol Community Health*, pp. jech–2013–202944, 2013.

[25] I. N. de Estadística, "Series detalladas desde 2002 -resultados por Comunidades Autonomas," 2017.

[26] R. Ocaña Riola and J. M. Mayoral-Cortés, "Spatio-temporal trends of mortality in small areas of Southern Spain," *BMC Public Health*, vol. 10, no. 1, p. 26, 2010.

[27] F. Viciana-Fernández, M. Ruiz-Ramos, and A. E. Pujolar, *Andalucía y España. Desigualdades en la Esperanza de Vida*, book section 3. Asociacin para la Defensa de la Sanidad Pblica de Andaluca, 2008.

[28] Eurostat, "European Union statistics on income and living conditions," report, European Union, 2017.

[29] G. Rey, E. Jougla, A. Fouillet, and D. Hmon, "Ecological association between a deprivation index and mortality in France over the period 19972001: variations with spatial scale, degree of urbanicity, age, gender and cause of death," *BMC public health*, vol. 9, no. 1, p. 33, 2009.

[30] T. A. Brown, *Confirmatory factor analysis for applied research*. Guilford Publications, 2014.

[31] K. G. Jreskog, "A general approach to confirmatory maximum likelihood factor analysis," *ETS Research Report Series*, vol. 1967, no. 2, pp. 183–202, 1967.

[32] INE, "Series detalladas desde 2002 -Resultados por Comunidades Autonomas," 2017.

[33] A. C. Pla and J. A. Cabrerizo, "Home ownership and social inequality in Spain," *Home Ownership and Social Inequality. Stanford*, pp. 233–254, 2004.

[34] R. S. Bivand, E. Pebesma, and V. Gómez-Rubio, *Applied Spatial Data Analysis with R*, vol. 10. Springer Science & Business Media, 2013.

[35] L. Anselin, *Spatial econometrics: methods and models*, vol. 4. Springer Science & Business Media, 2013.

[36] R. R. Sokal and D. E. Wartenberg, "A test of spatial autocorrelation analysis using an isolation-by-distance model," *Genetics*, vol. 105, no. 1, pp. 219–237, 1983.

[37] P. A. P. Moran, "The Interpretation of Statistical Maps," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 10, no. 2, pp. 243–251, 1948.

[38] D. G. Kleinbaum and M. Klein, *Survival analysis*, vol. 3. Springer, 2010.

[39] P. D. Allison, *Event history and survival analysis: Regression for longitudinal event data*, vol. 46. SAGE publications, 2014.

[40] T. M. Therneau and P. Grambsch, "Extending the Cox model," *Edited by P. Bickel, P. Diggle, S. Fienberg, K. Krickeberg*, p. 51, 2000.

[41] P. C. Austin, "A tutorial on multilevel survival analysis: methods, models and applications," *International Statistical Review*, vol. 85, no. 2, pp. 185–203, 2017.

[42] T. Therneau, "Mixed effects cox models," *CRAN repository*, 2015.

[43] V. S. Pankratz, M. De Andrade, and T. M. Therneau, "Random-effects cox proportional hazards model: general variance components methods for time-to-event data," *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, vol. 28, no. 2, pp. 97–109, 2005.

[44] K. C. Cain, S. D. Harlow, R. J. Little, B. Nan, M. Yosef, J. R. Taffe, and M. R. Elliott, "Bias due to left truncation and left censoring in longitudinal studies of developmental and disease processes," *American journal of epidemiology*, vol. 173, no. 9, pp. 1078–1084, 2011.

[45] L. Anselin, I. Syabri, and O. Smirnov, "Visualizing multivariate spatial correlation with dynamically linked windows," *Urbana*, vol. 51, p. 61801, 2002.

[46] E. Van de Poel, O. O'Donnell, and E. Van Doorslaer, "Is there a health penalty of China's rapid urbanization?," *Health Economics*, vol. 21, no. 4, pp. 367–385, 2012.

[47] Y. Rydin, A. Bleahu, M. Davies, J. D. Dvila, S. Friel, G. De Grandis, N. Groce, P. C. Hallal, I. Hamilton, P. Howden-Chapman, K. M. Lai, C. J. Lim, J. Martins, D. Osrin, I. Ridley, I. Scott, M. Taylor, P. Wilkinson, and J. Wilson, "Shaping cities for health: complexity and the planning of urban environments in the 21st century," *Lancet*, vol. 379, no. 9831, pp. 2079–108, 2012.

[48] T. Oke, "Urban heat island," in *The Routledge Handbook of Urban Ecology*, pp. 120–131, Routledge Abingdon, Oxon, 2011.

[49] R. Ocaña Riola and C. Sánchez-Cantalejo, "Rurality index for small areas in Spain," *Social Indicators Research*, vol. 73, no. 2, pp. 247–266, 2005.

[50] E. Romanos, "Evictions, petitions and escraches: Contentious housing in austerity Spain," *Social Movement Studies*, vol. 13, no. 2, pp. 296–302, 2014.

[51] R. Oldenburg and D. Brissett, "The third place," *Qualitative sociology*, vol. 5, no. 4, pp. 265–284, 1982.

[52] V. Mehta and J. K. Bosson, "Third places and the social life of streets," *Environment and Behavior*, vol. 42, no. 6, pp. 779–805, 2010.

[53] E. Regidor, F. Vallejo, C. Girldez-Garca, P. Ortega, J. M. Santos, P. Astasio, and L. de la Fuente, "Low mortality in the poorest areas of Spain: adults residing in provinces with lower per capita income have the lowest mortality," *European journal of epidemiology*, vol. 30, no. 8, pp. 637–648, 2015.

Appendix

Table 4: Factor loadings and uniqueness parameter from the maximum likelihood factor analysis

	Factor Loading	*Uniqueness Parameter*
Population Density	0.97	0.05
% of Artificial Surface Area	0.82	0.33
% of Health Service Area	0.90	0.19
Road Density	0.65	0.58
Note:	Estimated factor explains 71% of the variance.	

Table 5: Random effects (RE) statistics and model comparison to model without random effects

	Model 1	Model 2	Model 3	Model 4
Standard deviation RE	0.1559	0.1251	0.1202	0.1125
Variance RE	0.0243	0.0156	0.0144	0.0127
AIC	1085414	1077044	1077001	1076938
Log-Likelihood Ratio Test	127.11^{***}	56.36^{***}	48.36^{***}	37.55^{***}
Chi-Square (df)	(1)	(1)	(1)	(1)
Note:		*p<0.01; **p<0.005; ***p<0.001		

Table 6: Cox PH Model with mixed effects

	Dependent variable:		
	Hazard Ratios (95% CI)		
	(1)	(2)	(3)
Male	2.0872***	2.0883***	2.0898***
	(2.0682, 2.106)	(2.0692, 2.1074)	(2.0707, 2.1089)
Reference: Female			
Physically Dependent	3.0286***	3.0180***	3.0028***
	(2.9617, 3.0954)	(2.9512, 3.0848)	(2.9360, 3.0695)
Reference: No Dependency			
Single	1.412***	1.4126***	1.4038***
	(1.3826, 1.4422)	(1.3829, 1.4424)	(1.3740, 1.4336)
Widowed	1.1836***	1.1823***	1.1808***
	(1.1586, 1.2085)	(1.1573, 1.2073)	(1.1559, 1.2058)
Divorced/Separated	1.4794***	1.4765***	1.4768***
	(1.4249, 1.5340)	(1.4219, 1.5310)	(1.4222, 1.5314)
Reference: Married			
No or Incomplete Educ.	1.3798***	1.3837***	1.3942***
	(1.3387, 1.4210)	(1.3425, 1.4249)	(1.3529, 1.4357)
Primary/Secondary Educ.	1.1595***	1.1598***	1.1652***
	(1.1155, 1.2036)	(1.1157, 1.2038)	(1.1211, 1.2093)
Reference: Tertiary Educ.			
Does not Own House	1.1625***	1.1537***	1.1399***
	(1.1362, 1.1888)	(1.1274, 1.1801)	(1.1134, 1.1664)
Reference: Does Own House/Apartment			
Does not Own a Car	1.2690***	1.2693***	1.2653***
	(1.2493, 1.2887)	(1.2497, 1.2890)	(1.2456, 1.2851)
Reference: Does Own Car(s)			
Observations	351,769	351,769	351,769
Log Likelihood	−538,537.7	−538,510.2	−538,471.8
LR Test	8451.6 (df = 12)***	55** (df = 15)	76.8*** (df = 17)
Note:			*p<0.1; **p<0.05; ***p<0.01