

Socioeconomic characteristics of residential areas and risk of death: is variation in spatial units for analysis a source of heterogeneity in observed associations?

Jaana I Halonen,¹ Jussi Vahtera,^{1,2} Tuula Oksanen,¹ Jaana Pentti,¹ Marianna Virtanen,¹ Markus Jokela,³ Ana V Diez-Roux,⁴ Mika Kivimäki^{1,5}

To cite: Halonen JI, Vahtera J, Oksanen T, *et al*. Socioeconomic characteristics of residential areas and risk of death: is variation in spatial units for analysis a source of heterogeneity in observed associations?. *BMJ Open* 2013;**3**:e002474. doi:10.1136/bmjopen-2012-002474

► Prepublication history and addition material for this paper are available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2012-002474>).

JV and MK contributed equally.

Received 10 December 2012
Revised 5 February 2013
Accepted 4 March 2013

This final article is available for use under the terms of the Creative Commons Attribution Non-Commercial 2.0 Licence; see <http://bmjopen.bmj.com>

For numbered affiliations see end of article.

Correspondence to
Dr Jaana I Halonen;
jaana.halonen@ttl.fi

ABSTRACT

Objectives: Evidence on the association between the adverse socioeconomic characteristics of residential area and mortality is mixed. We examined whether the choice of spatial unit is critical in detecting this association.

Design: Register-linkage study.

Setting: Data were from the Finnish Public Sector study's register cohort.

Participants: The place of residence of 146 600 cohort participants was linked to map grids and administrative areas, and they were followed up for mortality from 2000 to 2011. Residential area socioeconomic deprivation and household crowding were aggregated into five alternative areas based on map grids (250×250 m, 1×1 km and 10×10 km squares), and administrative borders (zip-code area and town).

Primary and secondary outcome measures: All-cause mortality.

Results: For the 250×250 m area, mortality risk increased with increasing socioeconomic deprivation (HR for top vs bottom quintile 1.36, 95% CI 1.21 to 1.52). This association was either weaker or missing when broader spatial units were used. For household crowding, excess mortality was observed across all spatial units, the HRs ranging from 1.14 (95% CI 1.03 to 1.25) for zip code, and 1.21 (95% CI 1.11 to 1.31) for 250×250 m areas to 1.28 (95% CI 1.10 to 1.50) for 10×10 km areas.

Conclusions: Variation in spatial units for analysis is a source of heterogeneity in observed associations between residential area characteristics and risk of death.

INTRODUCTION

Evidence that the adverse socioeconomic characteristics of residential areas are risk factors for all-cause mortality is mixed, comprising both positive^{1–21} and null findings.^{6 14 22} In these studies the spatial unit to which area data has been aggregated has varied considerably

ARTICLE SUMMARY

Article focus

- There is no strong consensus on which spatial units are best for determining the health effects of residential areas.
- Few studies have been able to compare area-level socioeconomic effects using several alternative spatial units.

Key messages

- Data on residential area socioeconomic deprivation and household crowding were aggregated into five alternative areas based on map grids (250×250 m, 1×1 km and 10×10 km squares), and administrative borders (zip-code area and town/city).
- High areal socioeconomic deprivation and household crowding, as aggregated into the smallest of the five spatial units, 250×250 m square, were associated with increased mortality. For household crowding, excess mortality risk was also observed using the other spatial units.
- These data show that aggregating data in different ways leads to different results in the analyses of the associations between residential area characteristics and risk of death.

Strengths and limitations of this study

- Individual socioeconomic variables were adequately controlled for.
- As the study population consisted of Finnish public sector employees, the generalisability of the results needs to be confirmed in other studies.

and is a possible source of inconsistencies, a feature known as the Modifiable Area Unit Problem (MAUP).^{23 24} Some investigations have aggregated area characteristics to the level of states²⁵ towns^{14 22} zip-code areas^{11 21 26 27} census tracts^{1–3 5 6 14 28} blocks and wards^{9 29} and other statistical or geographical

units.^{3 7 8 12 16 17 29 30} Towns and other large administrative units can capture differences in the provision of community health and welfare services, but smaller spatial units, such as zip codes, may cover local variability in people's social environments as well as 'local health-related cultures' that may also contribute to mortality differences between areas.

Prior research comparing health effects by spatial units has suggested that no differences exist between spatial measures^{11 16 27 29} or that the smaller ones provide stronger effect estimates.^{4 13 18 28} However, few studies have systematically examined this issue across different area characteristics and various spatial units within a single analytic setting and adequately adjusting for individual socioeconomic variables. We sought to undertake such a study by comparing five different spatial units (towns, zip-code areas and map-grid squares of 250×250 m, 1×1 km and 10×10 km) in relation to two widely used socioeconomic area characteristics, deprivation and household crowding.

METHODS

Study design and population

The Finnish Public Sector study cohort consists of employees working for ten municipalities and six hospital districts in Finland. All men and women employed in these organisations for more than 6 months in any year between 1991 and 2005, and from the full spectrum of socioeconomic groups were eligible (n=151 901). Owing to the nature of public sector jobs in Finland (nurses, teachers, etc) most of the study participants were women. For this study, we selected those cohort members who were alive and aged 18–65 years at the beginning of the follow-up, which was the date on which the participant began his/her first employment contract in the target organisations between 1 January 2000 and 1 January 2005 (for those contracted before 2000, the start date was 1 January 2000). Thus, of the included participants 95.5% were employed at the beginning of the follow-up. Participants were followed until the end of December 2011, a move abroad, or death, whichever came first. The global positioning system (GPS)-coordinates of the residential buildings of 146 831 participants were obtained from the population information system of the Population Register Center using personal identification codes. The centre's data on nearly three million residences is maintained and checked in close cooperation with municipal building supervision authorities and local register offices.³¹ The outcome was all-cause mortality, including deaths from diseases and external causes. The dates of death were obtained from Statistics Finland on the basis of personal identification codes. The Ethics Committee of the Hospital District of Helsinki and Uusimaa approved the study.

Area characteristics and spatial units

The spatial units used were map-grid-based squares of 250×250 m, 1×1 km and 10×10 km and administrative measures of town and zip-code areas.³² The smallest units

represent areas in which people communicate with their neighbours, and conduct their businesses by foot. The zip-code area is a unit used in prior literature^{4 11 21} and is based on a defined postal area, usually larger than 1 km square, but smaller than a town. Towns are thus assumed to form boundaries within which people conduct most of their daily activities, and larger grids, in the 10×10 km scale, represent units possibly comparable to administrative units. All these units cover the whole Finland.

The participants were linked to the map squares by using the GPS-coordinates of their residential buildings and to their administrative areas by their postal addresses. For each spatial unit, information on deprivation and household crowding was calculated by Statistics Finland on the basis of data from the population register, registers of the Finnish Tax Administration, and Statistic Finland's employment register. In all these registers the total population residing in Finland at the end of data collection year served as the universe. For each spatial unit, we defined an index of socioeconomic deprivation using information on median income (median household income in the area logarithmically transformed and then coded as additive inverse in order to obtain higher values for greater deprivation), education attainment (proportion of those aged >18 whose highest education level was elementary school), and unemployment rate (unemployed people belonging to the labour force/total labour force). These are standard variables, used either separately or jointly, to characterise the disadvantage and deprivation.³³ For each of the three indicators, we derived a standardised z-score (mean=0, SD=1). The index of socioeconomic deprivation was then calculated by taking the mean value across all z-scores³⁴ when the z-score for at least one of the indicators was available. Because we had no information on, for example, car ownership or crime rates at each area level, we could not build an index identical to the Townsend deprivation index or the Scottish Index of Multiple Deprivation used in prior studies.³⁵ However, we defined household crowding (residential area (m²) per person) for each spatial unit, as household overcrowding is a variable also used in the Townsend deprivation index.³⁵ A small portion of grid database information was missing as Statistic Finland does not release information on areas with <10 residents, those with missing data were excluded from the analyses.

Covariates

The covariates obtained from the employers' administrative records were age, sex and occupational title. Occupational titles were used as one indicator of individual level socioeconomic status (SES). We classified individuals into three groups: high=upper grade non-manual workers (eg, physicians and teachers), intermediate=lower grade non-manual workers (eg, registered nurses, technicians) and low=manual workers (eg, cleaners, maintenance workers) based on the classification of occupations.³⁶ This classification is determined by the activities performed in each job and education, and we have previously used this in our studies.^{37 38} However, it may not

correspond to classifications used in other countries, because, for example, teachers in Finland are required to have a university degree. Other indicators for individual SES were level of education (high=university degree, intermediate=high school or vocational school, low=comprehensive school) obtained from Statistics Finland, and housing tenure (owner vs other) from the Population Register Center. Because the study participants were spread all over the country (though mainly in the Southern and Western parts, see online supplementary figure 1), and because of regional variation in mortality rates in Finland³⁹ we included a four-category area-level variable for county in the analyses (South, West, East and North obtained from Statistics Finland).

Statistical analyses

The associations between the two area characteristics and total mortality for each spatial unit were assessed using Cox proportional hazards regression (PHREG procedure of SAS V.9.2) with a robust variance estimator that accounts for the correlation of individuals residing within the same spatial units.⁴⁰ We also examined the spatial variance using the survival package in R software (using frailty with gamma distribution), which showed that the spatial correlation in mortality in these data is very small (p for frailty=0.41, variance of random effect 0.0017). Non-significant interaction terms between logarithmically transformed follow-up time and the area characteristics suggested that the proportionality assumption was not violated (all $p>0.05$). The results are presented as HR with 95% CI by quintiles of area characteristics, with the most favourable quintile as the reference group. Analyses were adjusted for age, sex, individual-level SES variables and county.

In further analyses we stratified the models by individual SES, and ran them including all spatial units of a given area characteristic simultaneously. As another sensitivity analysis, we excluded mortality that occurred within the first 2 years of follow-up from the data. To study the nature and strength of the spatial patterning of the area

characteristics, we calculated a pooled spatial autocorrelation index over five major town areas as suggested by Moran (VARIOGRAM procedure). Pooled Moran's indexes were used because cohort participants were scattered over the whole of Finland, but the majority (87%) resided within these five town areas (see online supplementary efigure 1). The distance between the areas within each town was calculated as the distance between the bottom-left coordinates of the 250×250 m map squares. To visualise the spatial distribution of the area characteristics we generated maps using the 250×250 m map grid over a sample town.

RESULTS

During 1.45 million person-years of follow-up 3832 participants died. The median follow-up period was 12 years (interquartile range 10.0–12.0), and median time of residence 6.4 years (see online supplementary efigure 1). The means and SD of the area characteristics, population size and numbers of participants by spatial units are shown in table 1. Correlations between the spatial units for each area characteristic are provided in the web appendix (see online supplementary efigure 2).

For the 250×250 m area, mortality increased linearly (p values for trend <0.001) with increasing quintiles of both area characteristics (figure 1). The hazards ratio for mortality in the top versus bottom quintile of socioeconomic deprivation was 1.36 (95% CI 1.21 to 1.52), and that of crowding 1.21 (95% CI 1.11 to 1.38) in the adjusted models.

In the analyses for the 1×1 km map grid squares, crowding was linearly (p value for trend <0.001) associated with increases in mortality, the HR for the top versus bottom quintile being 1.20 (95% CI 1.05 to 1.37). For deprivation the association was weaker (HR 1.10, 95% CI 1.01 to 1.20, for the top vs bottom quintile).

When using the 10×10 km grids, zip-code areas and towns, household crowding was again linearly associated with mortality (p values for trend <0.001 10×10 km, 0.02

Table 1 Means and SD of area characteristics by spatial units and numbers of participants included in the analyses

Variable	Spatial unit*					
	N	250×250 m Mean (SD)	1×1 km	10×10 km	Zip-code area	Town
Number of units (n)	†	27 495	10 272	1296	1858	422
Unit size (km ²)	†	0.0625 (0)	1 (0)	100 (0)	34.3 (96.5)	377 (370)
Participants per unit* (n)	146600	29 (32)	227 (234)	5777 (4260)	762 (642)	13073 (8883)
Socioeconomic deprivation (z-score)	139600	-0.28 (0.68)	-0.33 (0.65)	-0.43 (0.64)	-0.44 (0.66)	-0.51 (0.60)
Annual income‡ (k€)	130278	18.5 (4.8)	17.9 (3.7)	17.7 (2.7)	17.7 (3.1)	17.4 (2.5)
Low education (%)	130376	30.7 (10.9)	32.0 (8.5)	32.8 (5.6)	31.8 (6.8)	32.2 (5.0)
Unemployment rate (%)	139595	8.5 (6.7)	8.4 (4.5)	10.8 (3.5)	7.8 (3.4)	7.9 (2.5)
Household crowding (m ² per person)	134075	37.7 (7.8)	37.7 (5.0)	37.6 (2.4)	35.1 (2.8)	35.3 (1.7)

*In statistic Finland's linkage dataset, the SD of the total number of residents in the spatial units with study participants is 285 (294) per 250×250 m unit, 2365 (2351) per 1×1 km unit, 68 251 (53 153) per 10×10 km unit, 8096 (5133) per zip-code area and 152 355 (122 064) per town.

†The number of participants per unit ranges between 1 and 21 921, depending on the spatial unit.

‡Median income per resident aged >18 with taxable income in the spatial unit.

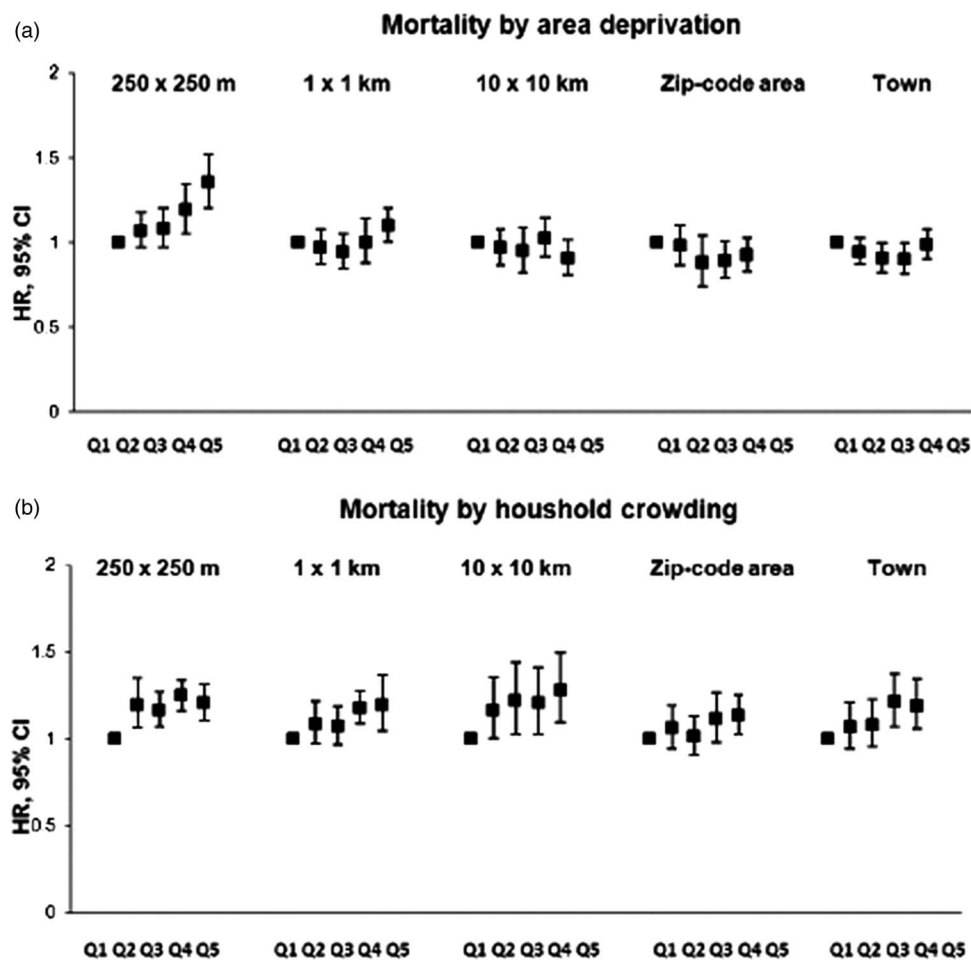


Figure 1 Mortality by (A) area deprivation and (B) household crowding. HRs (95% CIs) using five alternative spatial units. Models are adjusted for age, sex, occupational status, level of education, housing tenure and county.

zip code and <0.001 town). The association was the strongest in the 10×10 km grid (1.28, 95% CI 1.10 to 1.50, for the top vs bottom quintile), and the magnitudes of HRs for zip-code areas and towns were only slightly lower than those for the smaller grid based units (figure 1). Deprivation, however, was not associated with mortality in the larger units. To investigate potential confounding by individual-level SES, we ran the analyses for area deprivation stratified by individual-level SES variables in the smallest spatial unit. Mortality increased linearly by area socioeconomic deprivation in the 250×250 m square within each occupational group, each level of education and by housing tenure (figure 2). We then examined whether any given spatial unit drove the associations. In the analyses including all spatial units simultaneously we found the strongest associations for deprivation in the 250×250 m area (table 2). Associations for crowding became non-significant, possibly because of the high correlation between crowding in the 10 km grids and crowding in towns (Pearson $r=0.79$). Owing to this multicollinearity, we analysed crowding in the 250 m, 1 km and 10 km grids and in zip-code areas simultaneously, which resulted in significant associations for all units except zip-code area (HR in the top vs bottom quintile 1.12, 95%

CI 1.00 to 1.26 for 250 m grid, and 1.20, 95% CI 1.02 to 1.42 for 10 km grid). Analyses of the 250 m and 1 km grids with zip-code areas and towns also produced significant associations (HR 1.12, 95% CI 0.99 to 1.26 for 250 m grid, and 1.12, 95% CI 0.97 to 1.29 for town).

When the first 2 years of mortality follow-up were excluded, the effect estimates for the top versus bottom quintile attenuated slightly in the 250×250 m unit (1.31, 95% CI 1.16 to 1.47 for deprivation, and 1.18, 95% CI 1.08 to 1.28 for crowding), and in the 1×1 km unit (1.06, 95% CI 0.97 to 1.15 for deprivation and 1.18, 95% CI 1.05 to 1.33 for crowding). In the larger units, associations between crowding and mortality remained similar to those in the whole data, and deprivation was not associated with mortality (data not shown).

The spatial patterning of the area characteristics, based on pooled Moran's Indexes for the 250×250 m spatial units in the five major town areas were modest—0.043 (95% CI 0.013 to 0.074) for socioeconomic deprivation and 0.030 (95% CI 0.019 to 0.042) for household crowding. Figure 3 illustrates how broadening spatial units affect the identification of local areas with disadvantages in an example town. A substantial proportion of the 250×250 m squares that belong to the least

favourable quintile in terms of socioeconomic deprivation and household crowding may remain unidentified if disadvantage is defined on the basis of the surrounding areas instead of the local area itself.

DISCUSSION

This study has two key findings. First, for socioeconomic deprivation, the smallest spatial unit, the 250×250 m square, captured the mortality associations best. The associations were substantially weaker when deprivation was defined for the 1×1 km squares, and no association with mortality was found for wider spatial units. Second, we found a graded association for household crowding and increased mortality risk across all five alternative spatial

units. There was evidence that both small and broad spatial units were representative area definitions, suggesting that high household crowding in proximal areas and as a town average captured a partially non-overlapping set of mortality risk factors. Discordant findings across different socioeconomic exposures empirically illustrate the Modifiable Area Unit Problem and suggest that differences in the spatial units used in the analyses are a source of heterogeneity in observed associations between residential area characteristics and the risk of death.

Comparison with other research

Our findings are in agreement with several previous studies. An ecological study in France, for example, observed linear mortality effects in the two smallest area

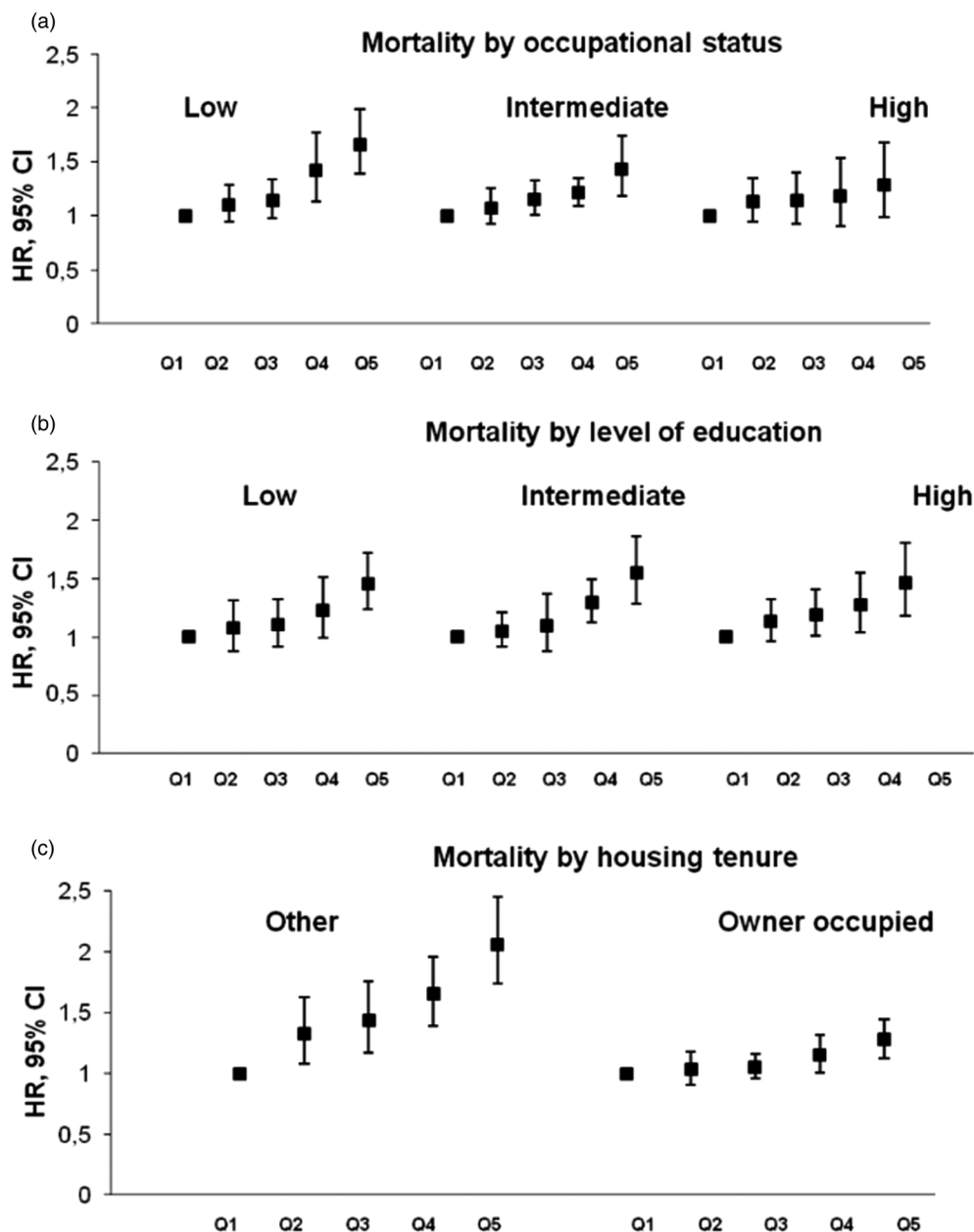


Figure 2 HRs (95% CIs) for mortality by quintiles of socioeconomic deprivation in 250×250 m grids by (A) occupational status, (B) level of education and (C) housing tenure. Models are adjusted for age, sex and county.

Table 2 HRs (95% CIs) for mortality within each area measure in association with socioeconomic deprivation and household crowding when simultaneously adjusting for all area definitions

	Socioeconomic deprivation*			Household crowding†		
	HR	95% CI		HR	95% CI	
250×250 m						
Q1 (lowest)	1			1		
Q2	1.10	0.99	1.22	1.18	1.04	1.33
Q3	1.15	1.04	1.28	1.13	1.03	1.23
Q4	1.31	1.13	1.51	1.18	1.08	1.30
Q5 (highest)	1.48	1.30	1.69	1.12	0.99	1.26
1×1 km						
Q1 (lowest)	1			1		
Q2	0.96	0.86	1.06	1.01	0.89	1.14
Q3	0.92	0.79	1.07	0.99	0.88	1.10
Q4	0.96	0.84	1.11	1.08	0.98	1.19
Q5 (highest)	1.03	0.90	1.18	1.07	0.91	1.26
10×10 km						
Q1 (lowest)	1			1		
Q2	1.05	0.90	1.22	1.18	1.02	1.37
Q3	1.08	0.85	1.36	1.19	0.99	1.43
Q4	1.18	0.97	1.43	1.17	0.93	1.46
Q5 (highest)	1.00	0.81	1.24	1.20	0.94	1.53
Zip-code area						
Q1 (lowest)	1			1		
Q2	0.91	0.83	1.00	1.01	0.86	1.18
Q3	0.86	0.76	0.96	0.91	0.81	1.02
Q4	0.82	0.73	0.93	0.97	0.88	1.08
Q5 (highest)	0.85	0.76	0.95	0.96	0.84	1.10
Town						
Q1 (lowest)	1			1		
Q2	0.97	0.79	1.19	0.95	0.84	1.07
Q3	0.80	0.62	1.02	0.93	0.78	1.11
Q4	0.83	0.65	1.07	1.06	0.90	1.24
Q5 (highest)	0.87	0.70	1.07	1.00	0.82	1.23

All models include age, sex, occupational status, level of education, housing tenure, county and simultaneous adjustments for
 *Socioeconomic deprivation for 250×250 m, 1×1 km, 10×10 km, zip-code area and town.
 †Household crowding for 250×250 m, 1×1 km, 10×10 km, zip-code area and town.

scales ('commun' and 'canton') but not for the three larger area units.¹³ Similarly, a recent meta-analysis found that area SES-mortality associations were slightly stronger in the small (relative risk 1.10, 95% CI 1.06 to 1.15) than the large (1.05, 95% CI 1.04 to 1.06) area units.¹⁸ However, only age and sex were controlled for in these models. A study from Massachusetts, USA, reported similar effects for large and small area units—the mortality incidence rate ratio in the area of lowest versus highest socioeconomic position was 1.39 (95% CI 1.33 to 1.46) using the zip-code area, and 1.31 (95% CI 1.25 to 1.38) using the block group.⁴ For crowding a significant association was seen at the block level (incidence rate ratio 1.43, 95% CI 1.23 to 1.67) but not at the level of zip-code areas (1.18, 95% CI 0.69 to 2.00).⁴ These effect estimates are in agreement with ours (HR 1.20, 95% CI 1.11 to 1.31 in the

250×250 m grids, and 1.14, 95% CI 1.03 to 1.25 in the zip-code area), although in our analysis the association at zip-code level also reached statistical significance. Stronger associations for the smaller of two spatial units have also been reported in relation to poor self-rated health, a predictor of overall mortality.²⁸

Our findings are not in agreement with data on household income from 14 US states in which similar effect estimates for two spatial units were observed (HR for mortality per US\$10 000 lower median tract-based household income 1.15, 95% CI 1.13 to 1.16, and zip-code-based income 1.16, 95% CI 1.14 to 1.17).¹¹ Obviously, this comparison is cruder than ours which was based on income quintiles. Another recent study reported linear mortality associations at tract-level with age-adjusted HRs in the most versus least deprived area—1.53 for women and 1.66 for men, but when adjusted for individual-level risk factors these HRs attenuated considerably—1.13 and 1.17, respectively.¹⁵ In a Finnish study, mortality rate ratio for the highest versus the lowest proportion of manual workers were similar at subdistrict level (1.13, 95% CI 1.01 to 1.25), and whole district level (1.10, 95% CI 0.98 to 1.23).¹⁶ At least two further studies, examining other health outcomes, have reported similar neighbourhood effects regardless of the spatial unit used.^{27 29}

Why might spatial unit matter

There are several reasons for the different area effects at different spatial levels. First, the scattered spatial patterning of deprivation in the 250×250 m units in our study may have resulted in stronger mortality associations in the small units than the large spatial units. Averaging over larger areas that are homogenous may result in a measure that does not capture the local conditions relevant to health—the association is therefore expected to be much weaker for the larger areas.³³ For deprivation, local characteristics may be particularly important because they either exert a causal effect (eg, via psychosocial pathways and stress), or because they are proxies for the individual SES, which is what matters causally. Nonetheless, we found no support for the proxy explanation because the associations for socioeconomic deprivation in the 250×250 m square remained after adjustment for individual socioeconomic position variables. These associations were also observed within each level of individual SES.

Second, the independent mortality effects of household crowding in local and broad spatial units suggest that there are likely to be multiple health-related mechanisms operating at different spatial levels. Household crowding may increase mortality risk because of its relation to increased transmission of diseases, such as respiratory and infectious diseases.^{41 42} Further plausible explanations involve neighbourhood social ties. The habit of smoking, for example, was found to spread through social ties assessed using social network analysis.⁴³ This effect may be better captured within small than large spatial units. Household crowding at the town level, in turn, was only weakly related to crowding in

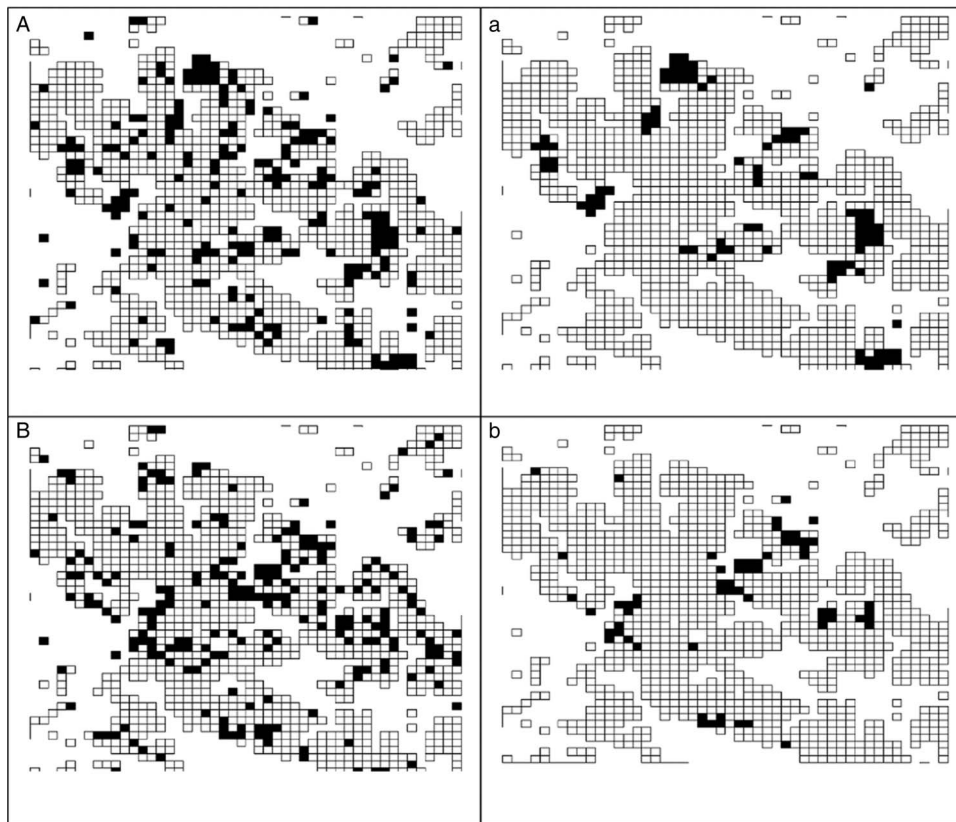


Figure 3 Illustration of the influence of area choice on detecting the local variance in area characteristics. In the maps of a sample town on the left, the 250×250 m squares that belong to the least favourable quintile of (A) area deprivation, and (B) household crowding are black, and on the right, only those squares whose neighbouring squares also belong to the worst quintile are black (a and b).

smaller areas and may therefore mark increased mortality risk through other mechanisms such as the quality of health-related resources, because the recruitment of motivated and well-trained personnel in health services may be easier in towns in which socioeconomic disadvantage is low.³⁰

Limitations

We used GPS-coordinates to link mortality and grid-based area data. The advantage of using map grids is that they can be used for the creation and maintenance of population-level spatial databases such as Statistics Finland's Grid Database. However, the lack of coordinate-based sociodemographic data in many countries is one reason why grid-based data have rarely been used for defining areas in socioepidemiological studies. In the near future, wider use of grid databases may become possible, as the European Forum for Geostatistics is currently developing guidelines for datasets and methods to link Population and Housing Census results from 2010 to 2011 to a common harmonised grid.⁴⁴ It has already combined European population grid datasets for the reference year 2006.⁴⁵ In this study, the used grid squares were fixed, whereas the administrative areas varied more in geographic and population sizes. However, as the socioeconomic exposure variables varied less within the large

(eg, mean annual income in towns 17 400€, SD=2500) than the small units (in 250 m grids 18 500€, SD=4800), and because categorical exposure variables were used, we believe the fixed versus non-fixed area choice did not have a substantial influence on the results.

At least three further methodological issues are noteworthy. First, although coordinate-based data provide many advantages for small-area research, errors in converting addresses into coordinates may occur.⁴⁶ In this study, all address-to-coordinate conversions were made by the Population Register Center. It has reported that 90% of the residential building locations in Finland are correct to within 20 m accuracy, and that the coverage is the best in the city plan areas (where most participants resided).⁴⁷

Second, it has been suggested that to obtain a comprehensive definition for a neighbourhood, both objective and subjective components of area characteristics should be incorporated.⁴⁸ However, we did not have subjective assessments of the areas, such as social ties between neighbours. Furthermore, we neither had data on the known mortality risk factors such as smoking (individual-level confounder), or air pollution (area-level confounder), which may vary according to the SES of the individual or the area, respectively.

Third, the study population was female dominated and consisted of Finnish (ie, mainly Caucasian ethnicity)

employees from the public sector. Thus, further research is needed to examine whether the specific spatial patterns observed in this study can be generalised to populations of other countries, or people with different ethnic backgrounds or of different age and gender distribution. Our results might not be generalisable to the unemployed or to populations in which socioeconomic inequalities are more pronounced than those in Finland.

CONCLUSION

Our study has demonstrated that area effects on mortality may vary in different spatial units depending on the exposure in question. This evidence suggests that the choice of spatial unit for analysis is a source of heterogeneity in observed associations and therefore an important factor in understanding area effects, interpreting previous findings and conducting future studies.

Author affiliations

¹Finnish Institute of Occupational Health, Kuopio, Finland

²Department of Public Health, University of Turku, and Turku University Hospital, Turku, Finland

³Department of Psychology, Institute of Behavioural Sciences, University of Helsinki, Helsinki, Finland

⁴Center for Integrative Approaches to Health Disparities, University of Michigan School of Public Health, Ann Arbor, Michigan, USA

⁵Department of Epidemiology and Public Health, University College of London, London, UK

Contributors JIH conceptualised the study, analysed the data, carried out the literature review and drafted the article. JV and MK conceptualised the study, acquired and interpreted the data and helped in drafting the article. JP, MJ and ADR helped to conceptualise the study, assisted with the data analysis and helped interpret the data. TO and MV helped interpret the data and draft the article. All authors have critically reviewed the drafts of the article and approved the final version.

Funding This work was supported by the Academy of Finland (projects 12 4271, 124 322, 129 262 and 126 602), the participating organisations, by the Bupa Foundation, the UK, and the National Institute on Aging (R01AG034454-01). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests None.

Ethics approval The Ethics Committee of the Hospital District of Helsinki and Uusimaa has approved the study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

- Anderson RT, Sorlie P, Backlund E, *et al.* Mortality effects of community socioeconomic status. *Epidemiology* 1997;8:42–7.
- Yen IH, Kaplan GA. Neighborhood social environment and risk of death: multilevel evidence from the Alameda County Study. *Am J Epidemiol* 1999;149:898–907.
- Bosma H, Van de Mheen HD, Borsboom GJ, *et al.* Neighborhood socioeconomic status and all-cause mortality. *Am J Epidemiol* 2001;153:363–71.
- Krieger N, Chen JT, Waterman PD, *et al.* Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter? The Public Health Disparities Geocoding Project. *Am J Epidemiol* 2002;156:471–82.
- Winkleby M, Cubbin C, Ahn D. Effect of cross-level interaction between individual and neighborhood socioeconomic status on adult mortality rates. *Am J Public Health* 2006;96:2145–53.
- Winkleby MA, Cubbin C. Influence of individual and neighbourhood socioeconomic status on mortality among black, Mexican-American, and white women and men in the United States. *J Epidemiol Community Health* 2003;57:444–52.
- Martikainen P, Kauppinen TM, Valkonen T. Effects of the characteristics of neighbourhoods and the characteristics of people on cause specific mortality: a register based follow up study of 252,000 men. *J Epidemiol Community Health* 2003;57:210–17.
- Osler M, Prescott E. Educational level as a contextual and proximate determinant of all cause mortality in Danish adults. *J Epidemiol Community Health* 2003;57:266–9.
- Borrell LN, Roux AV, Diez, Rose K, *et al.* Neighbourhood characteristics and mortality in the atherosclerosis risk in communities study. *Int J Epidemiol* 2004;33:398–407.
- Van Lenthe FJ, Borrell LN, Costa G, *et al.* Neighbourhood unemployment and all cause mortality: a comparison of six countries. *J Epidemiol Community Health* 2005;59:231–7.
- Thomas AJ, Eberly LE, Smith G, Davey, *et al.* ZIP-code-based versus tract-based income measures as long-term risk-adjusted mortality predictors. *Am J Epidemiol* 2006;164:586–90.
- Turrell G, Kavanagh A, Draper G, *et al.* Do places affect the probability of death in Australia? A multilevel study of area-level disadvantage, individual-level socioeconomic position and all-cause mortality, 1998–2000. *J Epidemiol Community Health* 2007;61:13–19.
- Rey G, Jouglé E, Fouillet A, *et al.* Ecological association between a deprivation index and mortality in France over the period 1997–2001: variations with spatial scale, degree of urbanicity, age, gender and cause of death. *BMC Public Health* 2009;9:33.
- Borrell C, Mari-Dell'olmo M, Serral G, *et al.* Inequalities in mortality in small areas of eleven Spanish cities (the multicenter MEDEA project). *Health Place* 2010;16:703–11.
- Major JM, Doubeni CA, Freedman ND, *et al.* Neighborhood socioeconomic deprivation and mortality: NIH-AARP diet and health study. *PLoS ONE* 2010;5:e15538.
- Tarkiainen L, Martikainen P, Laaksonen M, *et al.* Comparing the effects of neighbourhood characteristics on all-cause mortality using two hierarchical areal units in the capital region of Helsinki. *Health Place* 2010;16:409–12.
- Panczak R, Galobardes B, Voorpostel M, *et al.* A Swiss neighbourhood index of socioeconomic position: development and association with mortality. *J Epidemiol Community Health* 2012;66:1129–36.
- Meijer M, Rohl J, Bloomfield K, *et al.* Do neighborhoods affect individual mortality? A systematic review and meta-analysis of multilevel studies. *Soc Sci Med* 2012;74:1204–12.
- Yen IH, Syme SL. The social environment and health: a discussion of the epidemiologic literature. *Annu Rev Public Health* 1999;20:287–308.
- Robert SA. Socioeconomic position and health: the independent contribution of community socioeconomic context. *Annu Rev Sociol* 1999;25:489–516.
- McLoone P, Boddy FA. Deprivation and mortality in Scotland, 1981 and 1991. *BMJ* 1994;309:1465–70.
- Ocana-Riola R, Saurina C, Fernandez-Ajuria A, *et al.* Area deprivation and mortality in the provincial capital cities of Andalusia and Catalonia (Spain). *J Epidemiol Community Health* 2008;62:147–52.
- Openshaw S. *The modifiable areal unit problem*. Norwich, UK: Geo Books, 1983.
- Openshaw S. Optimal zoning systems for spatial interaction models. *Environ Plann A* 1977;9:169–84.
- Kaplan GA, Pamuk ER, Lynch JW, *et al.* Inequality in income and mortality in the United States: analysis of mortality and potential pathways. *BMJ* 1996;312:999–1003.
- Lakshman R, McConville A, How S, *et al.* Association between area-level socioeconomic deprivation and a cluster of behavioural risk factors: cross-sectional, population-based study. *J Public Health (Oxf)* 2010;33:234–45.
- Reijneveld SA, Verheij RA, De Bakker DH. The impact of area deprivation on differences in health: does the choice of the geographical classification matter? *J Epidemiol Community Health* 2000;54:306–13.
- Oliver LN, Hayes MV. Does choice of spatial unit matter for estimating small-area disparities in health and place effects in the Vancouver Census Metropolitan Area? *Can J Public Health* 2007;98(Suppl 1):S27–34.
- Stafford M, Duke-Williams O, Shelton N. Small area inequalities in health: are we underestimating them? *Soc Sci Med* 2008;67:891–9.

30. Kravdal O. Mortality effects of average education: a multilevel study of small neighbourhoods in rural and urban areas in Norway. *Int J Equity Health* 2009;8:41.
31. Väestörekisterikeskus, [Population Register Centre]. 2012. <http://vrk.fi/default.aspx?docid=48&site=4&id=40> (accessed 26 Jun 2012).
32. Statistics Finland. Grid Database 2007. http://www.tilastokeskus.fi/tup/ruututietokanta/index_en.html (accessed 11 Jan 2012).
33. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *J Epidemiol Community Health* 2001;55:111–22.
34. Kim D, Masyn KE, Kawachi I, et al. Neighborhood socioeconomic status and behavioral pathways to risks of colon and rectal cancer in women. *Cancer* 2010;116:4187–96.
35. Morris R, Carstairs V. Which deprivation? A comparison of selected deprivation indexes. *J Public Health Med* 1991;13:318–26.
36. Statistics Finland. *Classification of Occupations*. Helsinki: Statistics Finland, 1987.
37. Kivimäki M, Lawlor DA, Smith G Davey, et al. Socioeconomic position, co-occurrence of behavior-related risk factors, and coronary heart disease: the Finnish Public Sector study. *Am J Public Health* 2007;97:874–9.
38. Halonen JI, Kivimäki M, Pentti J, et al. Quantifying neighbourhood socioeconomic effects in clustering of behaviour-related risk factors: a multilevel analysis. *PLoS ONE* 2012;7:e32937.
39. Pitkänen K, Koskinen S, Martelin T. Geographical differences in mortality and their history [in Finnish]. *Duodecim* 2000;116:1697–710.
40. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. New York: Springer-Verlag, 2000.
41. Lietzau S, Raum E, Von Baum H, et al. Household contacts were key factor for children's colonization with resistant *Escherichia coli* in community setting. *J Clin Epidemiol* 2007;60:1149–55.
42. Taylor JC, Law GR, Boyle PJ, et al. Does population mixing measure infectious exposure in children at the community level? *Eur J Epidemiol* 2008;23:593–600.
43. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. *N Engl J Med* 2008;358:2249–58.
44. European Forum for Geostatistics. Geostat 1 A project — Representing Census data in a European population grid. Geostat 1 A project—Representing Census data in a European population grid. 2010. http://www.efgs.info/geostat-project/EFGS_GEOSTAT_flyer.pdf (accessed 8 Nov 2011).
45. European Forum for Geostatistics. GEOSTAT 2006 population grid dataset. 2012. (http://www.efgs.info/data/european-datasets/GEOSTAT_Grid_POP_2006_1K) (accessed 22 Mar 2013).
46. Krieger N, Waterman P, Lemieux K, et al. On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. *Am J Public Health* 2001;91:1114–16.
47. Väestörekisterikeskus [Population Register Centre]. Rakennusten osoite- ja koordinaattitietojen kattavuus väestötietojärjestelmässä [Report on the coverage of address and coordinate data of buildings in the population information system], 2004.
48. Weiss L, Ompad D, Galea S, et al. Defining neighborhood boundaries for urban health research. *Am J Prev Med* 2007;32(6 Suppl):S154–9.

Supplemental material

Halonen et al.

“Studying socioeconomic characteristics of residential areas and risk of death: Is choice of spatial unit an issue?”



eFigure 1. Inhabited areas in Finland (grey) and areas in which study participants live (black). Spatial unit 1×1 km square.

eTable 1. Descriptive statistics of individual-level variables.

Characteristic	Statistic	N missing
Sex, n (%)		-
Men	36 144 (24.6)	
Women	110 687 (75.4)	
Individual occupational status, n (%)		120
High	42 789 (29.2)	
Intermediate	69 688 (47.5)	
Low	34 234 (23.4)	
Individual level of education, n (%)		-
High	73 922 (50.3)	
Intermediate	52 708 (35.9)	
Low	20 201 (13.8)	
Housing tenure, n (%)		2920
Owner	82 647 (57.4)	
Other	61 262 (42.6)	
Median (IQR) age, y	41.0 (32.0-51.0)	-
Median (IQR) time of residence, y	6.0 (1.8-12.0)	-
Median (IQR) follow-up time, y	12.0 (11.0-12.0)	

IQR= interquartile (25% - 75%) range

eTable 2. Correlations between spatial units for each area characteristic.

Area characteristics	Pearson r			
	250 x 250 m vs. 1 x 1 km	250 x 250m vs. 10 x 10 km	250 x 250 m vs. Zip-code area	250 x 250 m vs. Town
Socioeconomic deprivation, <i>z-score</i>	0.63 *	0.36 *	0.47 *	0.32 *
Median household income, <i>k€</i>	0.71 *	0.47 *	0.59 *	0.45 *
Low education, %	0.66 *	0.37 *	0.53 *	0.32 *
Unemployment rate, %	0.62 *	0.37*	0.45 *	0.36 *
Household crowding, <i>m² per person</i>	0.60 *	0.30 *	0.37 *	0.29 *

* P-value <0.01