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 Kivimäki1,5

INTRODUCTION

Evidence that the adverse socioeconomic characteristics of residential areas are risk factors for all-cause mortality is mixed, comprising both positive1–21 and null findings.6 14 22 In these studies the spatial unit to which area data has been aggregated has varied considerably 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 towns14 22 zip-code areas11 21 26 27 census tracts1–3 5 6 14 28 blocks and wards9 29 and other statistical or geographical

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 area 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.
units. 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 or that the smaller ones provide stronger effect estimates. 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. 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 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.

The index of socioeconomic deprivation was then calculated by taking the mean value across all z-scores 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. 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. 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. This classification is determined by the activities performed in each job and education, and we have previously used this in our studies. 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 Finland 39 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 (VARIODRAG 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 figure 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 table 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 etable 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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Spatial unit*</th>
<th>250×250 m Mean (SD)</th>
<th>1×1 km</th>
<th>10×10 km</th>
<th>Zip-code area</th>
<th>Town</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of units (n)</td>
<td>†</td>
<td>27 495</td>
<td>10 272</td>
<td>1296</td>
<td>1858</td>
<td>422</td>
</tr>
<tr>
<td>Unit size (km²)</td>
<td>†</td>
<td>0.0625 (0)</td>
<td>1 (0)</td>
<td>100 (0)</td>
<td>34.3 (96.5)</td>
<td>377</td>
</tr>
<tr>
<td>Participants per unit* (n)</td>
<td>146600</td>
<td>29 (32)</td>
<td>227 (234)</td>
<td>5777 (4260)</td>
<td>762 (642)</td>
<td>13073 (8883)</td>
</tr>
<tr>
<td>Socioeconomic deprivation (z-score)</td>
<td>139600</td>
<td>−0.28 (0.68)</td>
<td>−0.33 (0.65)</td>
<td>−0.43 (0.64)</td>
<td>−0.44 (0.66)</td>
<td>−0.51 (0.60)</td>
</tr>
<tr>
<td>Annual income (k€)</td>
<td>130278</td>
<td>18.5 (4.8)</td>
<td>17.9 (3.7)</td>
<td>17.7 (2.7)</td>
<td>17.7 (3.1)</td>
<td>17.4 (2.5)</td>
</tr>
<tr>
<td>Low education (%)</td>
<td>130376</td>
<td>30.7 (10.9)</td>
<td>32.0 (8.5)</td>
<td>32.8 (5.6)</td>
<td>31.8 (6.8)</td>
<td>32.2 (5.0)</td>
</tr>
<tr>
<td>Unemployment rate (%)</td>
<td>139595</td>
<td>8.5 (6.7)</td>
<td>8.4 (4.5)</td>
<td>10.8 (3.5)</td>
<td>7.8 (3.4)</td>
<td>7.9 (2.5)</td>
</tr>
<tr>
<td>Household crowding (m² per person)</td>
<td>134075</td>
<td>37.7 (7.8)</td>
<td>37.7 (5.0)</td>
<td>37.6 (2.4)</td>
<td>35.1 (2.8)</td>
<td>35.3 (1.7)</td>
</tr>
</tbody>
</table>

*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.
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.
Spatial unit for socioeconomic characteristics and risk of death

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

<table>
<thead>
<tr>
<th>Socioeconomic deprivation*</th>
<th>Household crowding†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
</tr>
<tr>
<td>250×250 m</td>
<td></td>
</tr>
<tr>
<td>Q1 (lowest)</td>
<td>1</td>
</tr>
<tr>
<td>Q2</td>
<td>1.10</td>
</tr>
<tr>
<td>Q3</td>
<td>1.15</td>
</tr>
<tr>
<td>Q4</td>
<td>1.31</td>
</tr>
<tr>
<td>Q5 (highest)</td>
<td>1.48</td>
</tr>
<tr>
<td>1x1 km</td>
<td></td>
</tr>
<tr>
<td>Q1 (lowest)</td>
<td>1</td>
</tr>
<tr>
<td>Q2</td>
<td>0.96</td>
</tr>
<tr>
<td>Q3</td>
<td>0.92</td>
</tr>
<tr>
<td>Q4</td>
<td>0.96</td>
</tr>
<tr>
<td>Q5 (highest)</td>
<td>1.03</td>
</tr>
<tr>
<td>10x10 km</td>
<td></td>
</tr>
<tr>
<td>Q1 (lowest)</td>
<td>1</td>
</tr>
<tr>
<td>Q2</td>
<td>1.05</td>
</tr>
<tr>
<td>Q3</td>
<td>1.08</td>
</tr>
<tr>
<td>Q4</td>
<td>1.18</td>
</tr>
<tr>
<td>Q5 (highest)</td>
<td>1.00</td>
</tr>
<tr>
<td>Zip-code area</td>
<td></td>
</tr>
<tr>
<td>Q1 (lowest)</td>
<td>1</td>
</tr>
<tr>
<td>Q2</td>
<td>0.91</td>
</tr>
<tr>
<td>Q3</td>
<td>0.86</td>
</tr>
<tr>
<td>Q4</td>
<td>0.82</td>
</tr>
<tr>
<td>Q5 (highest)</td>
<td>0.85</td>
</tr>
<tr>
<td>Town</td>
<td></td>
</tr>
<tr>
<td>Q1 (lowest)</td>
<td>1</td>
</tr>
<tr>
<td>Q2</td>
<td>0.97</td>
</tr>
<tr>
<td>Q3</td>
<td>0.80</td>
</tr>
<tr>
<td>Q4</td>
<td>0.83</td>
</tr>
<tr>
<td>Q5 (highest)</td>
<td>0.87</td>
</tr>
</tbody>
</table>

All models include age, sex, occupational status, level of education, housing tenure, county and simultaneous adjustments for household income and 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 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 area were 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 area were weaker for the larger areas.33

Why might spatial unit matter

There are several reasons for the different area effects at different spatial levels. First, the scattered spatial pattern 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.45 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 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
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.

REFERENCES


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, Jussi Vahtera, Tuula Oksanen, Jaana Pentti, Marianna Virtanen, Markus Jokela, Ana V Diez-Roux and Mika Kivimäki

*BMJ Open* 2013 3:
doi: 10.1136/bmjopen-2012-002474

Updated information and services can be found at:
http://bmjopen.bmj.com/content/3/4/e002474

These include:

**Supplementary Material**
Supplementary material can be found at:
http://bmjopen.bmj.com/content/suppl/2013/04/03/bmjopen-2012-002474.DC1

**References**
This article cites 40 articles, 17 of which you can access for free at:
http://bmjopen.bmj.com/content/3/4/e002474#BIBL

**Open Access**
This is an open-access article distributed under the terms of the Creative Commons Attribution Non-commercial License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the license. See: http://creativecommons.org/licenses/by-nc/3.0/ and http://creativecommons.org/licenses/by-nc/3.0/legalcode

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Topic Collections**
Articles on similar topics can be found in the following collections

Epidemiology (2040)
Public health (2136)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/