PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Socioeconomic characteristics of residential areas and risk of death: Is variation in spatial units for analysis a source of heterogeneity in observed associations?
AUTHORS	Halonen, Jaana; Vahtera, Jussi; Oksanen, Tuula; Pentti, Jaana; Virtanen, Marianna; Jokela, Markus; Diez Roux, Ana; Kivimäki, Mika

VERSION 1 - REVIEW

REVIEWER	Gina S. Lovasi, PhD, MPH Assistant Professor of Epidemiology Columbia University Mailman School of Public Health New York City, NY, USA I have no competing interests to report.
REVIEW RETURNED	30-Dec-2012

THE STUDY	I was not convinced that the literature review was well tailored to the focus on spatial scale, or had properly considered the range of explanations for the observed associations (e.g., air quality, stress).
GENERAL COMMENTS	Major comments: Some key potential explanations for the observed crowding association with mortality are oddly not explored. For example, are there institutional settings (e.g. Nursing homes or assisted living environments) that are driving some of these associations because of ill or elderly individuals being relocated into dense settings? Perhaps some exploration of the data is feasible either by identifying housing type through the population data, or by excluding mortality that occurred in the first two years of follow-up. I was happy to see the use of quantiles across spatial units, which like standardization is something I think makes the analyses across spatial units easier to compare. In contrast, a median income difference of $10,000 dollars as discussed for ref 11 may represent a more extreme contrast for larger spatial units. I think this would be a good distinction to draw out in discussing the literature on whether associations have been previously observed to be smaller at smaller spatial aggregations. Looking beyond mortality to other health analyses may also be helpful. I found it odd that the grid and administrative areas were not considered on the same scale (either by including smaller administrative areas or larger grid cells comparable to the scale of zip codes and towns. I think this is a missed opportunity to disentangle the scale vs zoning issues of MAUP. Minor comments: It seemed to me that some copyediting could be useful (as an example, the term “gradually” seems odd for a dose-response relationship, vs something more precise like monotonically or linearly), and additional details (e.g. on geocoding, on the

	construction of deprivation z-scores) could make the methods clearer. In Table e3, I don't understand the utility of juxtaposing these correlations, particularly when the variables take different forms (continuous age, dichotomous sex, ordered categorical SES, nominal categorical County). I was not convinced that the literature review was well tailored to the focus on spatial scale, or had properly considered the range of explanations for the observed associations (e.g., air quality, stress).

REVIEWER	Prof Ian Gregory Professor of Digital Humanities Lancaster University UK I have no competing interests
REVIEW RETURNED	22-Jan-2013

GENERAL COMMENTS	This is a competent and thorough investigation of an aspect of statistical research – namely the shape and size of the administrative units used – that is frequently taken for granted but may, as the research reveals, have a significant impact on the results. The research is interesting in that it starts with individual level data and aggregates these in four different ways – two of which are different sizes of grid squares and two of which are different sizes of irregular administrative unit. It has long been known that using data aggregated in different ways results in different results from analyses. This is generally called the Modifiable Areal Unit Problem and was first described by Stan Openshaw in 1977 – I am slightly surprised that neither this term or a reference to Openshaw's work appear in the paper and would recommend that perhaps they should be. The results of the study are perhaps not all that surprising – in general hazard ratios seem to fall with increasingly aggregate data and this might be expected as the scores for each administrative unit will become increasingly averaged as the size of the unit increases. Nevertheless, the fact that the researchers have taken this on and shown that the effect happens is interesting and commendable. These effects will occur in a lot of studies and this paper allows other papers to provide better evidence when they speculate on the likely effects of aggregation on their results. For this reason I would recommend that it is published although I would suggest including a reference to Openshaw's work as described above.

VERSION 1 – AUTHOR RESPONSE

Reviewer: Gina S. Lovasi, PhD, MPH
Assistant Professor of Epidemiology
Columbia University Mailman School of Public Health New York City, NY, USA

I have no competing interests to report.

Thank you for the opportunity to review this paper examining the modifiable areal unit problem (MAUP) for the case of socioeconomic status or crowding predicting mortality.

Major comments:
1. Some key potential explanations for the observed crowding association with mortality are oddly not explored. For example, are there institutional settings (e.g. Nursing homes or assisted living environments) that are driving some of these associations because of ill or elderly individuals being relocated into dense settings?

Response 1: This is an important point. In this study, we used data from the Finnish Public Sector Study registers including those aged 18-65 years in the beginning of the follow-up for whom data on the location of residence were available. Of the included participants, 95.5% were employed at the beginning of the follow-up, as now clarified on page 5: "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 organizations between January 1st 2000 and January 1st 2005 (for those contracted before 2000, the start date was January 1st 2000). Thus, of the included participants 95.5% were employed at the beginning of the follow-up." This means that those institutionalized must have been a fraction among the 4.5% remaining participants (n=6670). Therefore we believe that people in nursing homes and hospitals could not have driven these findings.

2. Perhaps some exploration of the data is feasible either by identifying housing type through the population data, or by excluding mortality that occurred in the first two years of follow-up.

Response 2: Thank you for this suggestion. We have now added sensitivity analyses excluding the first two years of follow-up. Please see Methods on page 8: "As another sensitivity analysis, we excluded mortality that occurred within the first two years of follow-up from the data."

and Results on page 10: "When the first two years of mortality follow-up were excluded the effect estimates for the top vs. bottom quintile attenuated slightly in the 250×250m unit (1.31, 95% CI 1.16-1.47 for deprivation, and 1.18, 95% CI 1.08-1.28 for crowding) and in the 1×1 km unit (1.06, 95% CI 0.97-1.15 for deprivation and 1.18, 95% CI 1.05-1.33 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)."

3. I was happy to see the use of quantiles across spatial units, which like standardization is something I think makes the analyses across spatial units easier to compare. In contrast, a median income difference of $10,000 dollars as discussed for ref 11 may represent a more extreme contrast for larger spatial units. I think this would be a good distinction to draw out in discussing the literature on whether associations have been previously observed to be smaller at smaller spatial aggregations. Looking beyond mortality to other health analyses may also be helpful.

Response 3: Thank you for the suggestion. We have now made this distinction in the discussion, please see page 11: "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 (hazard ratio for mortality per $10,000 lower median tract-based household income 1.15, 95% CI: 1.13-1.16, and zip code-based income 1.16, 95% CI: 1.14-1.17) [11]. Obviously, this comparison is cruder than our which was based on income quintiles."

4. I found it odd that the grid and administrative areas were not considered on the same scale (either by including smaller administrative areas or larger grid cells comparable to the scale of zip codes and towns. I think this is a missed opportunity to disentangle the scale vs zoning issues of MAUP.

Response 4: Thank you for this insightful comment. We have now run new analyses for 10 km grids, the size of which is closer to, although not exactly the same as that of zip code areas and towns. Please see abstract: "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 on administrative borders (zip-code area and town)." "…the hazard ratios ranging from 1.14 (95% CI 1.03-1.25) for zip-code and 1.21 (95% CI 1.11-1.31) for 250×250 m area to 1.28 (95% CI 1.10-1.50) for 10×10km area."

Methods on page 5: "Spatial units used were map-grid-based squares of 250×250 m, 1×1 km and 10×10 km [33], and administrative measures of town and zip-code areas."
and: "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 the larger administrative units."

and Results on page 9: "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 zip code, and <0.001 town). In the 10×10 km grid, the association was the strongest (1.28, 95% CI 1.10-1.50, for the top vs. bottom quintile) and the magnitudes of hazard ratios for zip-code areas and towns were only slightly lower than those for the smaller grid based units (Figure 1)."

All tables and Figure 1 have been up-dated accordingly.

Minor comments:
5. It seemed to me that some copyediting could be useful (as an example, the term "gradually" seems odd for a dose-response relationship, vs something more precise like monotonically or linearly), and additional details (e.g. on geocoding, on the construction of deprivation z-scores) could make the methods clearer.

Response 5: Thank you for the suggestions. The text has been reviewed for language by a native speaker. We have now replaced word "gradual" with "linear".

In the Methods on page 5 we now provide more information about the residential coordinate data: "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 Center's data on nearly three million residences is maintained and checked in close cooperation with municipal building supervision authorities and local register offices [31]."

In the Discussion on page 14 we also state the following: "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]. "

On page 6 we now give more information about the deprivation index as follows: ” 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 characterize the disadvantage and deprivation [33]. For each of the three indicators, we derived a standardized z-score (mean = 0, standard deviation = 1). The index of socioeconomic deprivation was then calculated by taking the mean value across all z-scores [34] when the z-score for at least one of the indicators was available.”

6. In Table e3, I don’t understand the utility of juxtaposing these correlations, particularly when the variables take different forms (continuous age, dichotomous sex, ordered categorical SES, nominal categorical County).

Response 6: Thank you for the comment. This correlation table from the supplement has been deleted.

7. I was not convinced that the literature review was well tailored to the focus on spatial scale, or had properly considered the range of explanations for the observed associations (e.g., air quality, stress).

Response 7: We have edited the Introduction and Discussion sections, and the main changes are the following:
We have now mentioned the Modifiable Area Unit Problem and referred to this on page 4 as follows: “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].”

In the introduction we now mention larger literature about the spatial scale studies, also some that were not looking at mortality, please see page 4: “Some investigations have aggregated area characteristics to the level of states [25], 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].

and: “Prior research comparing health effects by spatial units 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].”

In the discussion we have now also mentioned studies looking other outcomes than mortality, please see page 11: “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].”
and page 12: “At least two further studies, examining other health outcomes, have reported similar neighbourhood effects regardless of the spatial unit used [27 29].”

We have also mentioned stress as a possible pathway on the deprivation-mortality association on page 12: “For deprivation, local characteristics may be particularly important because they either exert a causal effect (e.g. via psychosocial pathways and stress), or because they are proxies for the individual socioeconomic status, which is what matters causally.”

On page 13 we have added discussion about social ties as another possible explanation: “Further plausible explanations involve neighborhood 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."

And we mention the lack of air pollution data as a limitation of this study on page 14: "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 by the socioeconomic status of the individual or the area, respectively."

Reviewer: Prof Ian Gregory
Professor of Digital Humanities
Lancaster University
UK

I have no competing interests

This is a competent and thorough investigation of an aspect of statistical research – namely the shape and size of the administrative units used – that is frequently taken for granted but may, as the research reveals, have a significant impact on the results. The research is interesting in that it starts with individual level data and aggregates these in four different ways – two of which are different sizes of grid squares and two of which are different sizes of irregular administrative unit. It has long been known that using data aggregated in different ways results in different results from analyses. This is generally called the Modifiable Areal Unit Problem and was first described by Stan Openshaw in 1977 – I am slightly surprised that neither this term or a reference to Openshaw's work appear in the paper and would recommend that perhaps they should be.
The results of the study are perhaps not all that surprising – in general hazard ratios seem to fall with increasingly aggregate data and this might be expected as the scores for each administrative unit will become increasingly averaged as the size of the unit increases. Nevertheless, the fact that the researchers have taken this on and shown that the effect happens is interesting and commendable. These effects will occur in a lot of studies and this paper allows other papers to provide better evidence when they speculate on the likely effects of aggregation on their results. For this reason I would recommend that it is published although I would suggest including a reference to Openshaw's work as described above.

Response: Thank you for this comment. We have now described the modifiable area unit problem and referred to Mr. Openshaw's work in the Introduction on page 4 as follows: "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]."

and in the Discussion on page 11: "Discordant findings across different socioeconomic exposures empirically illustrate the Modifiable Area Unit Problem and suggest that differences in spatial units used in the analyses are a source of heterogeneity in observed associations between residential area characteristics and the risk of death."

23. Openshaw S. The modifiable are unit problem. Norwich, UK: Geo Books, 1983.
24. Openshaw S. Optimal zoning systems for spatial interaction models Environment and Planning A 1977;9:169-84

VERSION 2 – REVIEW

REVIEWER	Gina S. Lovasi Assistant Professor Department of Epidemiology Columbia University Mailman School of Public Health New York, NY, USA
REVIEW RETURNED	01-Mar-2013

- The reviewer completed the checklist but made no further comments.