DISCUSSION PAPER SERIES

IZA DP No. 3719

The Construction of Neighbourhoods and its Relevance for the Measurement of Social and Ethnic Segregation: Evidence from Denmark

Anna Piil Damm
Marie Louise Schultz-Nielsen

September 2008

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Construction of Neighbourhoods and its Relevance for the Measurement of Social and Ethnic Segregation: Evidence from Denmark

Anna Piil Damm
Aarhus School of Business

Marie Louise Schultz-Nielsen
Rockwool Foundation Research Unit
and IZA

Discussion Paper No. 3719
September 2008

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3719
September 2008

ABSTRACT

The Construction of Neighbourhoods and its Relevance for the Measurement of Social and Ethnic Segregation: Evidence from Denmark*

In this paper we propose a model for constructing neighbourhoods based on geo-referenced data and administrative data. The 431,233 inhabited hectare cells in Denmark are clustered into 9,404 small and 2,296 large neighbourhoods, inhabited on average in 2004 by 572 and 2,343 persons respectively. The priorities in the clustering process are to obtain neighbourhoods that are unaltered over time, delineated by physical barriers, compact, homogeneous in terms of type of housing and ownership, relatively small, homogeneous in terms of number of inhabitants, and comprised of a contiguous cluster of cells. To illustrate the importance of detailed neighbourhood information we compare social and ethnic segregation measured by Isolation and Dissimilation indices on the levels of municipalities and of small neighbourhoods. Our findings demonstrate substantial variation in the residential mix in neighbourhoods within a given municipality, and thus show the importance of having information on a more detailed geographical level than that of the municipality.

JEL Classification: I3, J61, R2

Keywords: geo-referenced data, neighbourhoods, segregation

Corresponding author:

Marie Louise Schultz-Nielsen
The Rockwool Foundation Research Unit
Sejrøgade 11
2100 Copenhagen Ø
Denmark
Email: mls@rff.dk

* We are grateful for financial support from the Rockwool Foundation Research Unit and the Danish Social Science Research Agency (grant 24-03-0288). We thank Torben Tranæs for his valuable comments.

I Introduction

Until now, Danish research into the effects of residential segregation has been hindered by the lack of a catalogue of neighbourhoods defined in a consistent manner for the entire country. Some Danish studies of neighbourhood effects have been carried out at the municipality level. But the results of these studies cannot be interpreted as genuine neighbourhood effects, since municipalities tend to be too large, both in terms of number of inhabitants and physical area, to be good specifications of neighbourhoods (Heinesen 2002). Other Danish studies have identified residential areas which received governmental financial support for social housing initiatives or the counselling of residents in the mid-1990s, and which have been referred to as “vulnerable” areas (see e.g. Hummelgaard et al. 1997). However, even for these socially vulnerable areas it is still not possible to study neighbourhood effects, since the rest of country is not divided into neighbourhoods with which they can be compared. Moreover, the vulnerable areas are administratively defined. Ideally, residential areas should be divided into neighbourhoods based on objective statistical criteria. A final approach taken in the Danish neighbourhood studies literature has been to use questionnaires to collect information about the residential area of respondents (see Constant and Schultz-Nielsen 2004). This approach does not allow for an unequivocal definition of neighbourhoods and would be too costly if all areas in Denmark were to be covered.

Our aim is to construct neighbourhoods for all residents in Denmark which satisfy the following criteria. First, a neighbourhood should correspond to the geographical area within which an individual has social contact with other residents. Second, the neighbourhood delineation should be unaltered over a specified period of time, in order to allow comparisons over time. This criterion rules out use of administrative divisions such as parishes and school districts. Finally, we should be able to combine the neighbourhoods thus defined with administrative register information.

For practical reasons and to comply with the rules of confidentiality of Statistics Denmark we construct two types of neighbourhoods: small neighbourhoods which have a minimum of 150 households, to be used in the analyses of residential segregation, and large neighbourhoods which have a minimum of 600 households, to be used for descriptive purposes.

II Construction of Neighbourhoods

II.A The Data Foundation: Geo-Referenced Data

We use geo-referenced data provided by The National Square Grid – Denmark to construct residential neighbourhoods (see http://www.kms.dk/NR/rdonlyres/4B374089-A734-4C5C-807D-F2D55B9615B1/0/systemspecifikation_danskekvadretnet_gb.pdf). The National Square Grid – Denmark is a national system of vector grids constructed by the National Survey and Cadastre (Kort & Matrikelstyrelsen) and Statistics Denmark. The National Square Grid divides Denmark into cells (squares).

Each cell has a unique name which corresponds to the cell size plus the co-ordinates of the lower-left corner of the cell, i.e. cell size_north_east, where the cell size refers to the length and height of the square, which can be from 100 metres to 100 kilometres. Thus, the National Square Grid constitutes an unambiguous geographical division of Denmark into spatial units which is stable and yet flexible (in that one can choose the size of cells to work with), and also independent of administrative boundaries.

It is possible to link these cells with administrative register information related to the residents of each cell. However, access to administrative register information requires that the division into cells meets certain confidentiality requirements of Statistics Denmark. The principle is that the more information one wants to have about the residents in each cell, the larger the number of households in each cell has to be. For analyses of residential segregation, the confidentiality requirement is a minimum of 150 households per neighbourhood. For descriptive purposes the confidentiality requirement is a minimum of 600 households per neighbourhood. We therefore have the choice between either using a grid with large cells or clustering small cells into groups of cells.

For the analysis of residential segregation, the latter choice is greatly preferable to the former, as it allows for the more precise delineation of neighbourhoods. Therefore, we construct neighbourhoods on the basis of the grid using the smallest cell size, 10,000 square metres, henceforth referred to as hectare cells. We cluster hectare cells until they meet the confidentiality requirements of a minimum of 150 and 600 households for segregation analysis and descriptive purposes respectively.

The starting point for the clustering of cells into neighbourhoods is the 431,233 hectare cells (100 m x 100 m) which were inhabited on Jan. 1st 1985 or Jan 1st 2004. Table 1 shows the distribution of hectare cells with respect to the number of households and individuals at these two dates. Note that 10.1% of the hectare cells were not inhabited in 1985 but were inhabited in 2004. Similarly, 3.9% of the cells were not inhabited in 2004 but were inhabited in 1985.

As shown in Table 1, only 0.2% of the hectare cells contained enough households in both 1985 and 2004 to meet the confidentiality requirement of a minimum of 150 households. Around 65% of the cells were inhabited by fewer than five households in both years. This calls for an extensive clustering of cells into neighbourhoods.

Table 1
Distribution of hectare cells in terms of numbers of households and residents on Jan. 1st 1985 and Jan. 1st 2004.

Number of households /residents in the cell	Percentage distribution, number of households		Percentage distribution, number of residents	
	1985	2004	1985	2004
		%		
0	10.1	3.9	10.1	3.9
1-4	63.9	65.7	44.1	46.8
5-9	14.2	15.5	16.0	16.4
10-19	8.0	9.9	11.5	13.2
20-49	2.5	3.4	15.0	16.0
50-99	1.0	1.1	2.1	2.3
100-149	0.3	0.3	0.7	0.7
150-	0.2	0.2	0.6	0.6
All	100.2	100.0	100.1	99.9

Source: Own calculations based on data from Statistics Denmark and Geomatic.
Note: Number of inhabited cells in either 1985 or 2004: 431,233.

II.B Clustering Criteria

Our aim is to cluster hectare cells into groups to obtain neighbourhoods which

- are inhabited by at least 150/600 households
- are unaltered over time (in this instance, across the two dates for which we have data)
- are delineated by physical barriers
- comprise a contiguous cluster of cells
- are compact
- are homogeneous in terms of type of housing and ownership of the housing unit
- are relatively small
- are homogeneous in terms of number of inhabitants

The criteria are listed in order of priority and have been used in the construction of both small and large groups of hectare cells. Large hectare cell groups have been constructed according to an additional criterion, namely that their boundaries should respect the boundaries of the small hectare cell group boundaries. Therefore, any given small hectare cell group lies entirely within a given large hectare cell group. In the following, we explain the choice of clustering criteria.

The neighbourhood should reflect the geographical area within which the individual has most opportunities for contact with other residents. We assume that the likelihood of social interaction with other residents in the area decreases with the spatial distance between the housing units, so that individuals who live on the same stairway are more likely to interact than individuals who live on different stairways of the same block of flats and in different blocks. This assumption is supported by Butt (2002), Wellman (1996) and Latané et al. (1995). Since the small hectare cell groups are meant to represent social networks of the individuals in the neighbourhood of residence, residential neighbourhoods should be small. For this reason the small hectare cell groups should be just large enough to meet the confidentiality rules of Statistics Denmark, that is, they should contain 150-200 households which as far as possible are closer to one another than to other households. For the same reason the small hectare cell groups should be compact and consist of contiguous hectare cells.

Furthermore, we assume that some visible features in residential areas such as water areas (oceans, inlets and big lakes), large forests and major roads reduce social interaction between individuals living on different sides of the visible feature. We therefore use such physical barriers as boundaries of neighbourhoods. Moreover, since physical boundaries are stable over time, their use as neighbourhood boundaries ensures unaltered neighbourhoods over time. Unaltered neighbourhoods over time are important for reasons of clarity of the analyses and to guarantee that neighbourhoods are uniquely defined on the basis of the hectare cells. For this reason the neighbourhoods must meet the minimum requirements with respect to number of households in both 1985 (start year of analysis) and 2004 (end year of analysis). Note, however, that the average number of households increases a little over time due to demographic changes.

The criterion of contiguous cells ensures that the clustering of cells into groups respects physical barriers in the form of water areas and forests. In practice, a few cells on different sides of water areas do have to be clustered together. This happens, for example, in the case of islands with so few households that the island does not meet the minimum requirements in terms of number of households. In the case of there being a ferry or bridge connection between the island and the mainland, the island is added to the group of hectare cells on the mainland to which there is a connection. In the absence of a connection, the island is clustered with the nearest inhabited hectare cells. All larger roads (road class 1-3) are regarded as physical barriers; in principle, there is no clustering of cells on opposite sides of these roads. There are however a few cases where the number of households between two large roads does not meet the rules of confidentiality. If the cells are connected with a group of inhabited cells by a road (possibly smaller than road class 1-3), they are clustered with these. In the absence of a road connection to other inhabited cells, the cells are clustered with inhabited cells on the opposite side of the smallest of the surrounding roads. Although we would have liked to do so, we have not been able to take account of railways in the clustering process.

We aim at constructing neighbourhoods which are homogeneous with respect to demographic and socio-economic characteristics of the residents because we assume that social interaction is greatest among individuals with similar demographic and socio-economic characteristics, henceforth referred to as equals. This assumption is supported by Becker (1957) and Bailey (1959), for example, who were the first to put forward the hypothesis that an individual's preference for place of residence includes preferences for the race of their neighbours. White Americans may, for instance, dislike having Black neighbours. However, clustering of individuals into residential neighbourhoods based on demographic and socio-economic similarity would cause the segregation indices to be overestimated. Moreover, a goal of homogeneity in terms of personal attributes of the residents would contradict the goal of neighbourhoods that were unaltered over time. Urban renewal, repairs and the like could, for instance, lead to a more heterogeneous demographic and socioeconomic composition in the neighbourhoods. Under such circumstances, satisfaction of a homogeneity criterion would require a new delineation of neighbourhoods.

However, equals are likely to sort into the same type of housing and house ownership, e.g. two individuals who live in two flats on opposite sides of a street are more likely to be equals than an individual who lives in a flat and an individual who lives in a house across the street. Similarly, house owners are more likely to be equals than a house owner and a tenant. The distribution of different types of housing across the local housing market is likely to be an important determinant of the choice of residential neighbourhood for different demographic and socio-economic groups and exogenous in the location decision of the household. Moreover, the housing stock is relatively static. Therefore, we use housing type and ownership information instead of the demographic and socio-economic characteristics of individuals in the construction of residential neighbourhoods.

We seek to construct neighbourhoods which are homogeneous in terms of number of inhabitants for the following reasons. First, if residential segregation is measured at the municipality level, we cannot tell whether the potential effects of residential segregation capture social interaction with the nearest 200 or the nearest 2,000 neighbours. If the neighbourhoods have roughly the same number of inhabitants, the effects of residential segregation can be estimated with more precision. Second, differences in segregation indices between two neighbourhoods with varying numbers of inhabitants may result from the difference in the number of inhabitants, because the variation in residential mix is likely to be greater in the larger neighbourhood. Thirdly, according to a recent British study, Bolster et al. (2004), neighbourhood effects decrease with the number of inhabitants in the neighbourhood. Finally, we expect neighbourhood effects to increase with the average duration of residence in the neighbourhood; and the larger the neighbourhood, the longer the average duration of residence. With larger neighbourhoods, people may move to a new house but within the same neighbourhood.

Note also that the use of physical barriers and the criterion of homogeneity in terms of housing stock to delineate neighbourhoods imply that the size of the neighbourhoods is exogenous in the analyses of the determinants of socioeconomic and ethnic segregation. In contrast, the neighbourhood size would be endogenous in such analyses if instead we had used a criterion of homogeneity in terms of socioeconomic and demographic characteristics.

II.C Implementation of the Clustering Algorithm

An algorithm for the clustering of hectare cells into neighbourhoods according to the criteria outlined above was written and implemented in collaboration with the firm Geomatic. A short description of the clustering process follows. For the sake of convenience, we will henceforth refer to the small hectare cell groups as small neighbourhoods and the large hectare cell groups as large neighbourhoods.

II.C.1 Construction of Neighbourhoods

The first factor to be taken into consideration in the clustering process of cells is physical barriers in the form of larger roads (road classes 1-3). The road net divides Denmark into approximately 500 areas within which first large neighbourhoods and subsequently small neighbourhoods are constructed. The construction of both types of neighbourhood follows the same principles.

In each area, the hectare cell with the largest number of households is selected. The housing and ownership characteristics of this cell are compared with the eight neighbouring cells (within 100 metres). The cell is then clustered with the neighbouring cell with the closest resemblance to it, cf. Section II.C.3. Next, the mean characteristics of this provisional neighbourhood are compared with other hectare cells within 100 metres. In the absence of inhabited neighbouring cells within 100 metres, the search area is expanded to 200 metres, etc. The clustering process continues until the provisional neighbourhood has at least 600 or 150 households. At this point, a check is made to see if the number of households in the provisional neighbourhood actually has twice the number of households necessary. If so and if possible, it is split into two, cf. Section II.C.3.

The process is repeated for the next provisional neighbourhood, which to begin with is the hectare cell with the largest number of households among the remaining cells. The characteristics of this cell are compared to the characteristics of the neighbouring cells and it is clustered with the most similar neighbouring cell. However, there is one exception to this rule. If the cell is neighbour to another provisional neighbourhood and is more similar to this neighbourhood than to other neighbouring cells, the two provisional neighbourhoods are clustered into one; upon which a check is made to see whether it is large enough to be split into two.

II.C.2 Neighbourhoods which are too large are split into two if possible

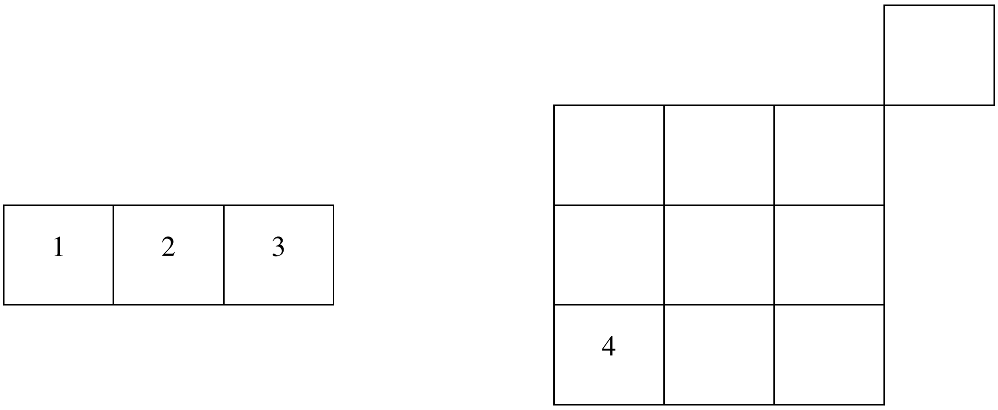
Each time a provisional neighbourhood is expanded with new hectare cells, a check is made as to whether the newly-constructed neighbourhood has more than twice as many households as the minimum required. If this is the case, the neighbourhood is split into two new provisional neighbourhoods. Since the neighbourhood is homogenous in terms of housing and ownership type it is split solely on the basis of a proximity criterion.

The split is made as follows. The distances between the hectare cells in the neighbourhood are calculated. The hectare cell that has the largest average distance to the other cells is selected, and is clustered with the nearest cell to it. If there is a tie for which is the nearest cell, the most distant cell is clustered with the nearest cell to it which leads to the lowest average distance between the remaining cells. This maintains the compact form of the provisional neighbourhood which it is desired to split into two. For this reason the clustering can only take place provided that the incorporation of new cells does not lead to an increase in the average distance between the remaining cells. If this is the case, the division of the original neighbourhood is given up.

Figure 1 shows an example of a division of a provisional neighbourhood into two. The most distant hectare cell is cell 1. It is clustered with the nearest cells, i.e. cells 2 and

3. The clustering process is continued from cell 3 to 4, because in this way the distance between the remaining cells is reduced.

Figure 1
Example of division of over-large provisional neighbourhoods into two based on a proximity criterion



II.C.3 Which hectare cells are similar?

As mentioned before, hectare cells are clustered so as to make neighbourhoods homogeneous in terms of housing characteristics. The basis for comparison of the housing characteristics of cells is administrative register information about each housing unit with respect to housing type and ownership. Each variable is divided into four categories:

Housing type:
1) Farmhouse or detached house
2) Townhouse or small block of flats
3) Large block of flats
4) Second home or other house

Ownership:
1) Private ownership
2) Privately owned rental
3) Publicly owned rental
4) Private cooperative housing

The distribution of housing units with respect to housing type and ownership is found for each hectare cell. Next, the neighbouring cell is found which has the most similar distribution.

In the calculation of which hectare cell is the most similar, housing type is given a weight of 0.3 and ownership a weight of 0.7. One reason that these weights were chosen is that we wish to use the neighbourhoods to estimate the effects of living in a "vulnerable" area, that is an area with an overrepresentation of marginalised groups on the labour market. For this reason the clustering process takes into consideration the subsidiary goal that constructed neighbourhoods should be in accordance with currently defined vulnerable neighbourhoods (see Section II.C.5). The ownership type has been found to be the main determinant of the delineation of vulnerable areas. The variation in housing type mix is substantial in the vulnerable areas (townhouses and small and large blocks of flats), whereas the ownership is almost exclusively public (notably non-profit, state subsidised cooperative housing associations).

II.C.4 Neighbourhoods

The principles for clustering of hectare cells into small and large neighbourhoods were described earlier. Table 2 shows the results of the clustering process that we carried out using the method described above. The 431,233 inhabited hectare cells were clustered into 9,086 small and 2,295 large neighbourhoods.

The small neighbourhoods were inhabited by on average 234 households (556 residents) in 1985 and 273 households (592 residents) in 2004. The median in all four cases was slightly lower than the mean, indicating that for more than half of the small neighbourhoods the numbers of both residents and households were below the mean. A small number of neighbourhoods with large populations contributed to a substantial increase in the mean value. The size of small neighbourhoods was on average 47.5 hectares, 22 hectares being the median and 190 hectares the 95^{th} percentile. That is to say, for 5% of the neighbourhoods the area exceeded 190 hectares; these were predominantly situated in the countryside.

In spite of our goal of having homogeneous neighbourhoods in terms of number of households, some small and large neighbourhoods had twice as many households as the minimum numbers required of 150 and 600 respectively. This was primarily caused by the additional criterion of compact neighbourhoods.

Heterogeneity of large neighbourhoods with respect to number of households was of little concern, since they were to be used primarily for descriptive purposes. In contrast, homogeneity of small neighbourhoods was considered important, because these neighbourhoods were to be used in analyses of determinants of neighbourhood mobility and evaluations of neighbourhood effects.

Table 2
Characteristics of small and large neighbourhoods

	Percentiles					All		
	5 %	25 %	50 %	75 %	95 %	Mean	Std.	Number of obs.
Small neighbourhoods								
Households 1985	158	180	217	272	366	234.9	73.9	9086
Households 2004	163	196	245	313	474	272.7	115.0	9086
Persons 1985	289	409	513	653	949	555.6	220.1	9086
Persons 2004	289	411	526	700	1114	592.2	285.5	9086
Size (in hectares)	2	6	22	58	190	47.5	64.6	9086
Large neighbourhoods								
Households 1985	631	712	859	1072	1460	929.9	293.4	2295
Households 2004	653	798	985	1237	1831	1079.7	396.2	2295
Persons 1985	1160	1583	1994	2608	3878	2200.0	904.2	2295
Persons 2004	1180	1618	2090	2807	4310	2344.5	1039.0	2295
Size (in hectares)	7	27	88	268	668	187.9	236.3	2295
No. of small neighbourhoods	3	3	4	5	6	4.0	1.3	2295

Source: Own calculations based on data from Statistics Denmark and Geomatic.

For this reason we made another attempt to split small neighbourhoods that were too large, but now using a different method. We chose the most southerly cells in the neighbourhood and clustered them with the nearest hectare cells until the provisional neighbourhood had at least 150 households. The subsequent neighbourhoods were constructed in an analogous way, but starting from the most easterly, northerly and westerly cells respectively. When the number of households in the remaining cells was insufficient to constitute a separate neighbourhood, each of the remaining hectare cells was clustered with the nearest provisional neighbourhood. Next, we checked that the provisional neighbourhoods were compact. If so, they were made permanent.

The 552 small neighbourhoods with the largest number of households were involved in this second attempt at division. As a result, 333 additional provisional neighbourhoods were constructed. Of these, 23 were not compact and were for that reason not made permanent. Using this division process, the number of small neighbourhoods came out at 9,396.

It turned out that no neighbourhoods had been constructed for the residential area of Vollsmose in Odense because street names were changed in the area between 1985 and 2004. We corrected the addresses for one year and divided Vollsmose into neighbourhoods on the basis of the same criteria as outlined in Section II.B, which resulted in the formation of one large and eight small neighbourhoods. The total number of small and large neighbourhoods was thus increased to 9,404 and 2,296 neighbourhoods.

Table 3
Characteristics of small and large neighbourhoods – after corrections

	Percentiles					All		
	5 %	25 %	50 %	75 %	95 %	Mean	Std.	Number of obs.
Small neighbourhoods								
Households 1985	157	178	214	265	340	227.0	60.2	9391
Households 2004	161	192	240	305	438	263.5	103.8	9404
Persons 1985	289	408	506	637	888	536.9	181.3	9391
Persons 2004	290	409	517	679	1010	572.2	251.4	9404
Size (in hectares)	2	7	23	62	172	45.9	56.9	9404
Large neighbourhoods								
Households 1985	631	712	859	1071	1459	929.1	291.1	2294
Households 2004	653	798	985	1237	1831	1079.2	394.4	2296
Persons 1985	1160	1583	1994	2607	3874	2197.6	899.0	2294
Persons 2004	1180	1618	2090	2805	4310	2343.5	1034.1	2296
Size (in hectares)	7	27	88	268	668	187.8	236.3	2296
No. of small neighbourhoods	3	3	4	5	6	4.1	1.4	2296

Source: Own calculations based on data from Statistics Denmark and Geomatic.

Table 3 shows that there still remained some neighbourhoods with relatively large populations after this round of division. One explanation is that the number of residents was rather large in some hectare cells in the large cities and we could not split up hectare cells. Another reason is that some neighbourhoods saw a dramatic increase in the number of households between 1985 and 2004.

II.C.5 Focus areas

In order to check whether the neighbourhoods constructed were appropriate for our purposes, we compared them with well-known residential areas (focus areas) during the clustering process. Among other things, we intend to use the constructed neighbourhoods in the analyses of immigrant settlement patterns. We therefore compared the neighbourhoods we had constructed with twenty socially vulnerable areas with a high proportion of immigrants identified by the Government's Programme Committee for Avoidance of Ghetto areas (see the Danish newspaper *Jyllands-Posten*, December 12, 2004).[4]

In one case, Mjølnerparken in Copenhagen, the overlap was unsatisfactory. The reason for the incomplete overlap is that Mjølnerparken, a large public-housing estate, was built after 1985. Due to the requirement that neighbourhoods should remain unaltered over time, the hectare cells of Mjølnerparken were clustered with the nearby hectare cells of Midgårdsgade and Mimersgade to ensure a sufficiently large neighbourhood in terms of number of households in 1985. However, the level of residential mixing increased.

In general, the overlap improved with the size of the housing units, the housing type homogeneity and the population density in the focus area.

[4] In the first round of checks, the overlap turned out to be satisfactory for most focus areas, but less so for areas with mixed housing types. At this point the ownership type was given less weight in the clustering process. Because the vulnerable areas are largely characterised by publicly-owned rental property, in particular cooperative housing societies, we decided to change the weights, so as to increase the weight for ownership type at the expense of the weight for housing type. In general, this led to few changes in the delineation of neighbourhoods, but the overlap improved for eight out of the twenty focus areas. There was only one case of a deterioration in the overlap. Therefore, we decided to give the ownership type more weight in the clustering process than the housing type.

Figures 2 and 3 illustrate a successful and less successful result of the clustering process respectively, if evaluated in terms of overlap between the constructed neighbourhood and the focus area.

Figure 2 shows the constructed neighbourhoods in the area of Taastrupgård, Høje Tåstrup. Addresses which are part of the focus area are marked by dots. Taastrupgård is seen to be divided into three small neighbourhoods with 313, 274 and 289 households in 2004, each of which consists of between 5 and 7 hectare cells. All addresses in Taastrupgård are included in one of these three neighbourhoods and all other addresses are excluded from the neighbourhoods. In other words, there is close correspondence between the total area of the three neighbourhoods and Taastrupgård.

Figure 2
Small neighbourhoods in the area of Taastrupgård, Høje Tåstrup

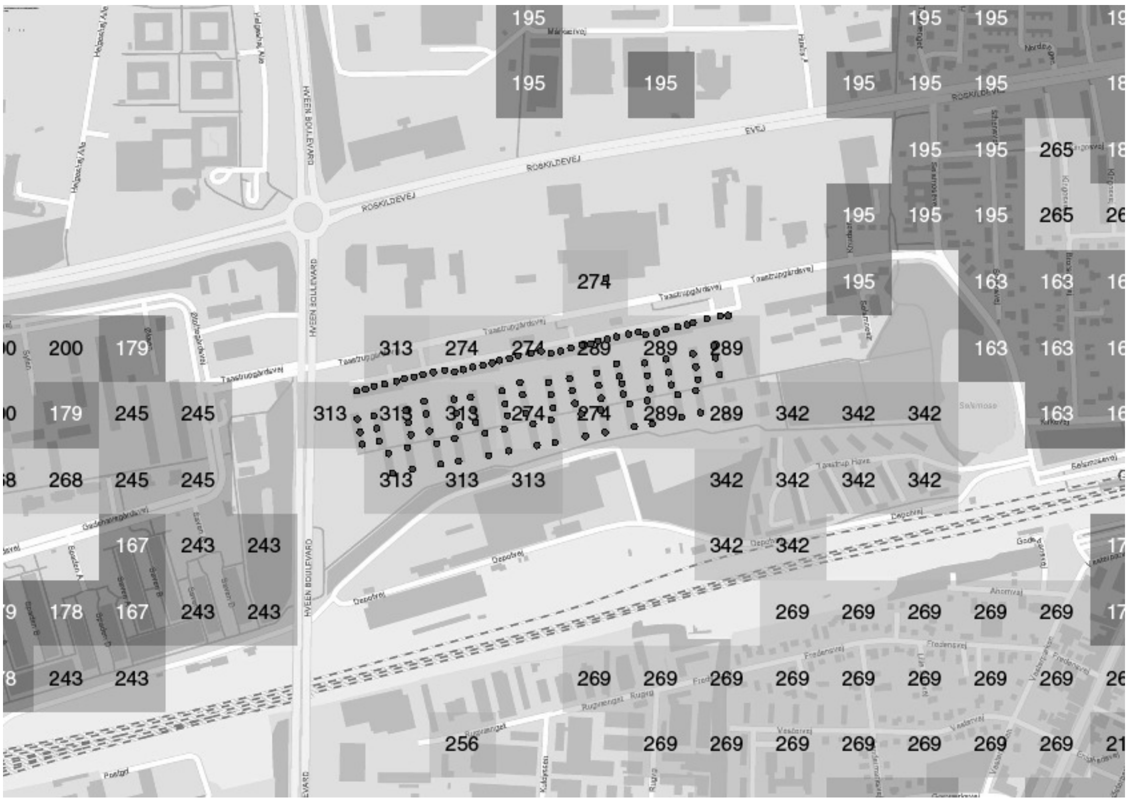
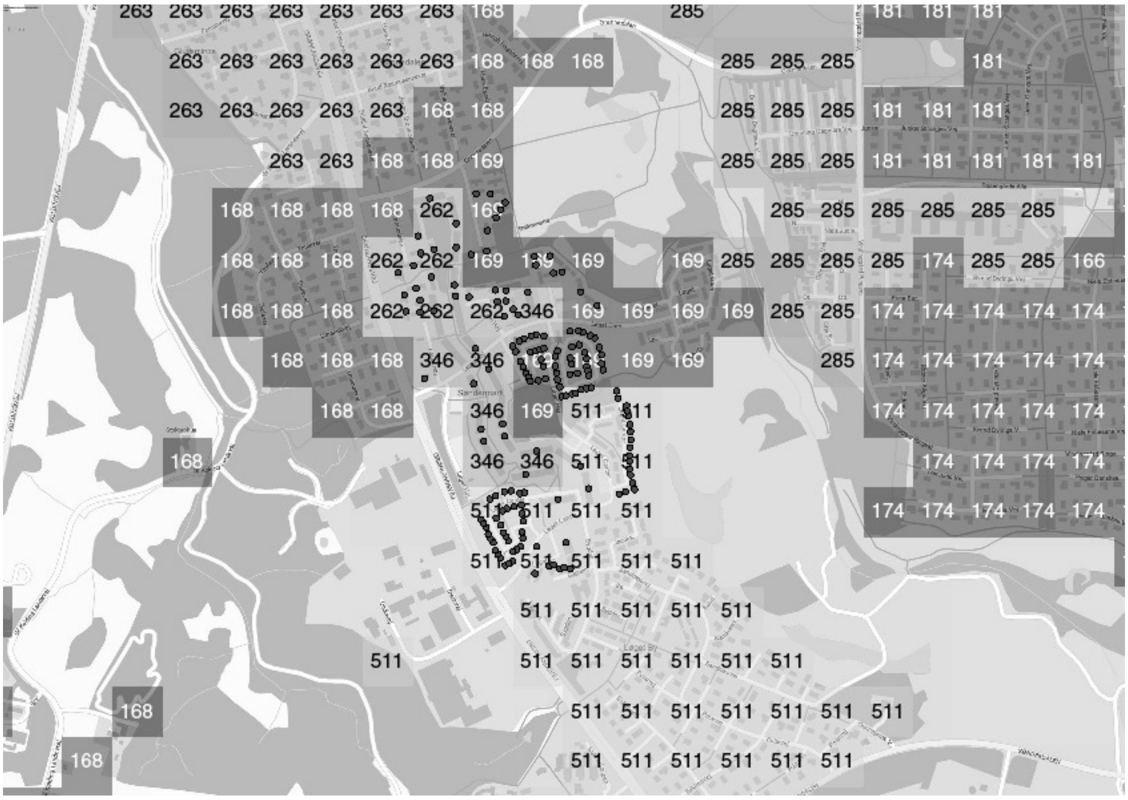


Figure 3, which shows neighbourhoods in the area of Løget By, Vejle, looks slightly different. The focus area is spread across five small neighbourhoods which contain many other hectare cells than those in the focus area. Note, however, that Figure 2 is a more representative example of the overlap than Figure 3.

Figure 3
Small neighbourhoods in the area of Løget By, Vejle



Another reason for which we are satisfied with the clustering process in spite of examples such as the one shown in Figure 3 is that the difference in the result of the clustering process reflects differences in the two types of focus area. Focus areas in the provinces (e.g. Løget By) tend to have fewer inhabitants and a lower population density than focus areas in cities (e.g. Taastrupgård). It is therefore not surprising that it is more difficult to delineate focus areas in the provinces.

III Relevance of Neighbourhoods for Measurement of Segregation

The aim of this section is to discuss the central measures of segregation and see how these measures are affected by the choice of area unit.

III.A Measurement of Segregation: Dissimilarity and Isolation Indices

According to Massey and Denton (1988), segregation in relation to settlement can be divided into five dimensions: Evenness, Exposure, Concentration, Centralization and Clustering. Our calculations will concentrate on the two most central dimensions, Exposure and Evenness, which in accordance with Massey and Denton (1988) will be measured by the Isolation index and the Dissimilarity index.

Exposure refers to the degree of potential contact or social interaction between different population groups, e.g., two ethnic groups (the majority and the minority) in neighbourhoods of a town. Exposure indices measure the degree to which minority and majority group members can potentially meet physically – in this case, seen in relation to area of residence. This factor can be measured by the Isolation index which, according to Massey and Denton (1988), can be defined as:

$${}_xP^*_x = \sum_i^I \frac{x_i}{X}\frac{x_i}{t_i} \quad \text{, where}$$

I: Number of area units

x_i: Number of individuals belonging to the minority group in area *i*

t_i: Number of residents in area *i*

X: Total number of individuals belonging to the minority

$\frac{x_i}{X}$: Share of the minority group living in area *i*

The interpretation of the formula is that for each minority group member, the chance of randomly meeting another minority group member in the area is x_i/t_i. These probabilities are weighted with the share of the minority group members living in area *i* (x_i/X), and finally the Isolation index is found by summing over all areas (I).

Note that this calculation is based on those areas where members of the minority group live. There may be many or few areas without any minority group members; what matters is the composition in those areas where the minority group members live. Note also that the Isolation index is sensitive to the size of the minority group, as the potential contact between minority group members increases with the share of the population that are minority group members.

Evenness refers to the spread of different groups across areas. If all areas have the same relative share of the majority and minority groups, Evenness is maximized and segregation minimized. According to Duncan and Duncan (1955), the Dissimilarity index can be defined as:

$$D = \frac{1}{2}\sum_{i}^{I}\left|\frac{x_i}{X} - \frac{t_i - x_i}{T - X}\right|$$, where

T: Total number of residents

$\frac{t_i - x_i}{T - X}$: Share of the majority group living in area *i*

The Dissimilarity index calculates the difference between the relative share of the majority and minority groups in each area, and afterwards these numerical differences are summed and divided by two. This measure can be interpreted as the number of members of the minority group that would have to move to another area in order for the minority group to have the same settlement pattern as the majority group. In contrast to the Isolation index, the Dissimilarity index is unaffected by the *size* of the minority group.

In a comment on our related work presented in Damm et al. (2006), Kærgård (2007) notes that the Dissimilarity index does not give as much weight to the outliers as a calculation based on the sum of the squares would have done, and he asks why we do not instead calculate the simple correlation coefficient between the share of the majority and the minority groups. The reason for not doing this is that the correlation coefficient measures the *linear* relationship between the two variables and in our data there is no evidence of a linear relationship between the two variables, or any other (single-valued) functional relationship for that matter. As expected, the data suggest that the relationship between the relative shares of the majority and the minority groups has to be represented by a two-dimensional graph. Therefore, it would be potentially misleading to use this statistic and complicate the interpretation of the result.

III.B Measurement of Segregation: Neighbourhood versus Municipality Level

Having described the central segregation measures we now turn to the question: what is the difference in the index if small neighbourhoods are used as the geographical area instead of municipalities? Or to put it differently, what extra insight is obtained from more detailed information about geographic areas? To find out, we measure the geographic segregation of socio-economic and ethnic groups in Denmark at two different geographical levels, the municipality level and the small neighbourhood level, and compare the results.

According to Statistics Denmark the population of Denmark consisted of 5,397,640 persons at the end of 2003. For 97.6 % of the population we can link the address information in the Central Personal Register (CPR) to a small neighbourhood.

In order to investigate the social dimension of geographical settlement we divide the population aged between 18 and 60 years into two groups, welfare recipients and non-welfare recipients, depending on whether they were registered as recipients of some sort of social benefit in 2003. Welfare recipients are defined as people who are not employed or studying, but who do receive early retirement pension, social security assistance or unemployment benefit.[5] These are all benefits that indicate a weak attachment to the labour market. All others are considered to be non-welfare recipients.

In 2003, welfare recipients constituted 14.4% of the population aged between 18 and 60 years. The isolation index would be equal to the share of welfare recipients in the population, i.e. 14.4%, if there was an even distribution of welfare recipients across locations. As shown in Table 3, in 2003 the isolation index calculated at the municipality level was 15.2%, suggesting an almost even distribution of welfare recipients across municipalities. However, if instead we calculate the isolation index at the small neighbourhood level, the value increases to 20.5%, demonstrating that welfare recipients are in fact overrepresented in certain geographical areas. The value of the Dissimilarity index measured at the municipality level indicates that 10.3 % of the welfare recipients would have to move to a municipality with a below average share of welfare recipients in order to obtain an even distribution of welfare recipients. This proportion increases to 27.3% if we measure segregation using small neighbourhoods instead of municipalities.

[5] See Damm et. al. (2006) for a detailed description of the definition.

Table 4
Measures of geographic segregation in 2003. By area unit.

Dimension	Segregation index	Municipalities	Small neighbourhoods
Social			
(welfare recipient share)	Isolation index	15.2	20.5
	Dissimilarity index	10.3	27.3
Ethnic			
(non-Western immigrant share)	Isolation index	8.7	23.4
	Dissimilarity index	29.8	54.1
Average no. of inhabitants		19,844.3	572.2
(std. dev.)		(38,840.8)	(251.4)

Source: Own calculations based on data from Statistics Denmark and Geomatic.

Turning to the ethnic dimension, we define the ethnic minority as non-Western immigrants and their descendants according to the definitions used by Statistics Denmark; the majority group consists of all other individuals. Using this definition, we find that the ethnic minority group constituted 6.0% of the total population at the end of 2003. The Dissimilarity index for ethnic residential segregation is 29.8 when measured at the municipality level, but it increases to no less than 54.1 when measured at the small neighbourhood level. The high level of segregation of non-Western immigrants within municipalities means that the Isolation index increases from 8.7 to 23.4 when measured at the small neighbourhood level instead of the municipality level.

We can thus see that within a municipality, a welfare recipient has a greater chance of meeting another welfare recipient than an immigrant has of meeting another immigrant. But if we calculate at the small neighbourhood level it becomes clear that within this smaller area the immigrant actually has a greater probability of meeting another minority group member than is the case for welfare recipients meeting one another.

Overall, our findings demonstrate substantial variation in the residential mix in neighbourhoods within a given municipality, and emphasise the importance of analysing residential segregation on a more detailed geographical level than that of the municipality. Residential segregation measured at the municipality level underestimates the real extent of residential segregation and could influence the results in, for example, empirical analyses of neighbourhood effects, or the push or pull factors in location decisions.

IV Conclusions

Until now, empirical research on determinants, effects and measurement of residential segregation in Denmark has been hindered complicated by the lack of a division of inhabited areas into neighbourhoods. In this paper we present a method for dividing inhabited areas into neighbourhoods and illustrate how use of neighbourhood information can improve the measurement of residential segregation in Denmark.

Small and large neighbourhoods are constructed by clustering hectare cells until they meet the requirements of a minimum of 150 and 600 households respectively. These hectare cells are clustered to obtain neighbourhoods that are unaltered over time, are delineated by physical barriers, comprise a contiguous cluster of cells, are compact, are homogeneous in terms of type of housing and ownership, are relatively small, and are homogeneous in terms of number of inhabitants.

The 431,233 inhabited hectare cells are clustered into 9,404 small and 2,296 large neighbourhoods, inhabited in 2004 by on average 572 and 2,343 persons respectively.

To illustrate the importance of detailed neighbourhood information we compare segregation as measured by Isolation and Dissimilation indices on the levels of municipalities and small neighbourhoods. As expected, we find that segregation in both the social and ethnic dimensions is higher when it is measured using the smaller area unit (neighbourhoods). But our example also illustrates that, depending on the settlement structure, some index values are more sensitive to the choice of area unit than others. When we use small neighbourhoods as the area unit, we conclude from the value of the Isolation Index that residential segregation of immigrants exceeds residential segregation of welfare recipients. We reach the opposite conclusion when the segregation analysis is based on municipalities.

This is just one example of the importance of the use of neighbourhood information. One could think of numerous of studies, including studies about determinants and effects of residential segregation, where neighbourhood information would be vital. Furthermore, as the construction is based on a combination of information from The National Square Grid and register data, this method could also be used in other

Scandinavian countries and the Netherlands, where administrative register and geo-referenced data are also available. This would enable us to compare settlement, segregation and the effects of the segregation on a much more accurate basis than today, when neighbourhood information is lacking or comes from very different sources.

References

Becker, G. 1957. *The Economics of Discrimination.* Chicago: University of Chicago.

Bailey, M. 1959. Note on the economics of residential zoning and urban renewal. *Land Economy* 35: 288-290.

Bolster, A., S. Burgess, R. Johnston, K. Jones, C. Propper and R. Saker. 2004. Neighbourhoods, Households and Income Dynamics: A Semi-Parametric Investigation of Neighbourhood Effects. CEPR Discussion Paper Series no. 4611.

Butts, C. 2002. *Spatial Models of Large-Scale Interpersonal Networks*. Doctoral Dissertation, Department of Social and Decision Sciences, Carnegie Mellon University.

Constant, A. and M.L. Schultz-Nielsen. 2004. Chapter 5 ´Labor Force Participation and Unemployment: Incentives and Preferences' in *Migrants, Work and the Welfare State* edited by T. Tranæs and K.F. Zimmermann. Syddansk universitetsforlag.

Damm, A.P., M.L. Schultz-Nielsen, T. Tranæs. 2006. En befolkning deler sig op? Gyldendal.

Duncan, O. B. & Duncan, B. 1955. A methodological analysis of segregation indexes. *American Sociological Review* 20:210-217.

Heinesen, E. 2002. Effekter af boligmæssig segregering. AMID Working paper 15/2002.

Hummelgaard, H., B.K. Graversen, D. Lemmich and J.B. Nielsen. 1997. *Udsatte boligområder i Danmark.* Akf forlaget.

Jyllands-Posten. 09.12.04. ´Fiasko for støtte til private udlejere"

Kærgård, N. 2007. Book review "En befolkning deler sig op?" Nationaløkonomisk Tidsskrift, no. 1, June 2007.

Latané B, J. Liu, A. Nowak and L. Zhenget. 1995. Distance matters: Physical space and social impact. *Personality and Social Psychology Bulletin* 21 (8): 795-805.

Massey, D.S. and N.A. Denton. 1988. The Dimensions of Residential Segregation. *Social Forces* 67 (2): 281-315.

Wellmann, B. 1996. Determinants of Recent Immigrants' Locational Choices. Federal Reserve Bank of Atlanta, Working Paper 98-3.