Online-only material for "Urban-rural differences in schizophrenia risk: Multilevel survival analyses of individual- and neighborhood-level indicators, urbanicity and population density in a Danish national cohort study" - Pedersen, CB et al.

Contents:

eMethods 1: Small-area zonal design for Denmark

In 1971 Denmark was divided into 2039 parishes, increasing to between 2116 and 2124 parishes from 1998 to 2010, between 2169 to 2194 parishes from 2011 to 2016, and 2159 parishes in 2018[1]. The historic changes in parish boundaries challenges longitudinal geospatial analyses using parish as the neighborhood of interest. In addition, parishes also vary considerably in population size; for example, in 2005, the 2159 parishes had a mean of 2505 residents but a large variance (IQR: 484 to 3158; Range: 33 to 37,872 residents)

From 1st January 1971 to 28th June 2003 Denmark was also divided into 275 municipalities, a higher geographic aggregation. A municipality reform in Bornholm on 29th June 2003 collapsed the five municipalities in Bornholm into one municipality, reducing the number of municipalities to 271. Then a major national municipality reform on 1 January 2007 redistributed Denmark into 98 municipalities, although the island of Ærø implemented the national municipality reform on 1st January 2006. Due to the municipality reform 32 municipality borders were unchanged; 225 municipalities were collapsed into larger municipalities without changing borders; and 14 municipality borders were changed[2]. In addition, Christiansø, a small island, located north-east of Bornholm with 92 residents is not included in any municipality boundaries.

Disregarding changes in borders, these geographic areas vary considerably in population size, which also leads to challenges in geospatial analyses[3]. For example, the 98-level municipality division has a mean population size of 55,179 residents (IQR: 28,834 to 59,848 residents; Range: 2090 to 495,507 residents), the 270-level municipality division had a mean of 20,028 residents (IQR: 7162 to 19,099 residents; Range 2090 to 495,506 residents) [all values in 2005].

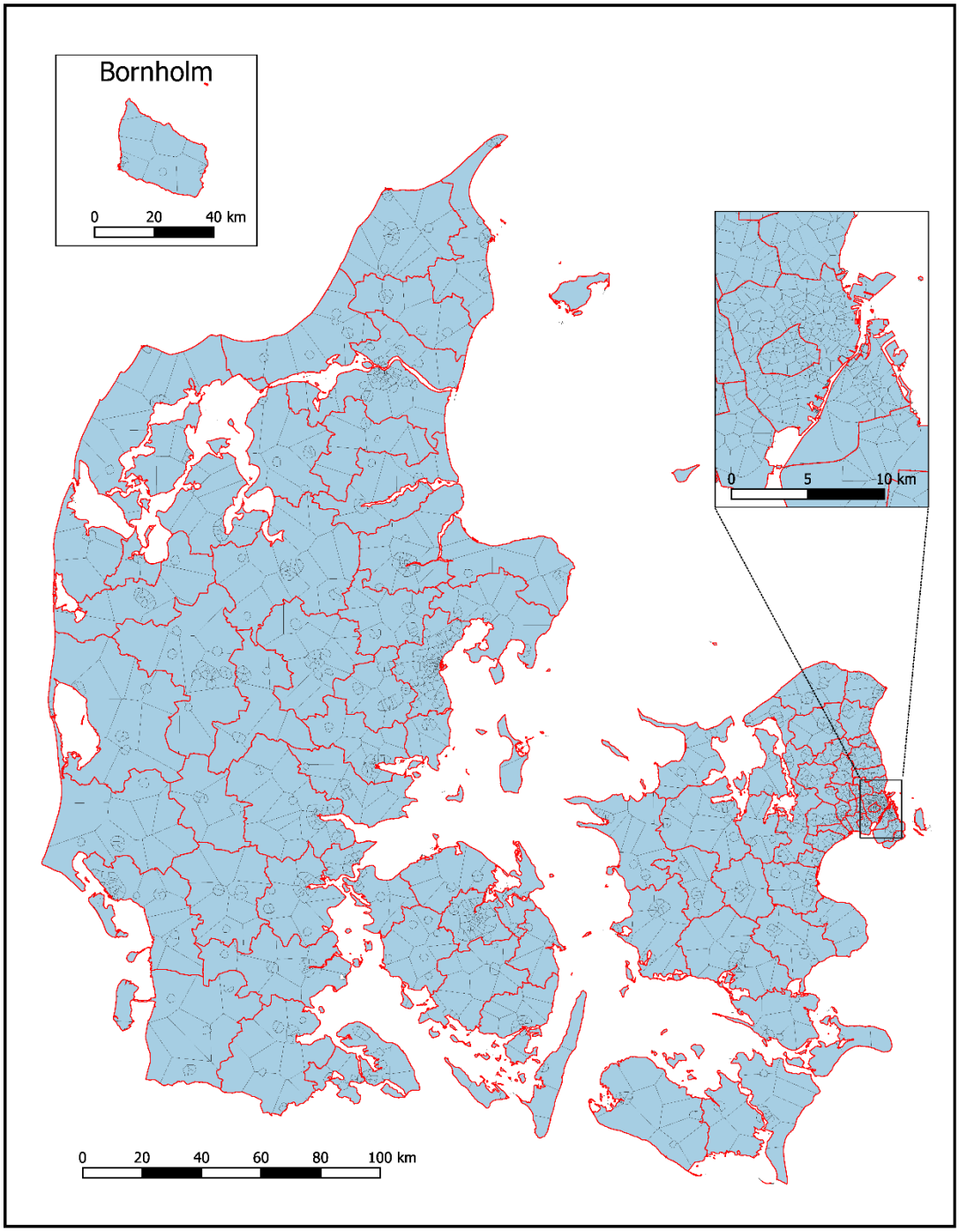
Faced with these challenges, we applied zone design algorithms on nationwide longitudinal residential geographic coordinates to divide Denmark into a total of 1885 homogeneously sized geographic areas, and to create a new level of geography, hereafter referred to as 'data zones'. Unlike other zoning algorithms which emphasized finding demographically/socioeconomically homogenized regions[4,5], our data zone algorithm aims to define geographic areas that are homogenized in population density and size. This way, we can avoid the selection bias by creating zones not by natural aggregations of residences but the variables we intend to estimate, such as socio-economic status. The data zones were defined by a data driven / automatic tessellation algorithm that finds a fine balance between respecting the clustering of community and the agnostic assignment of the area without prior knowledge. The zoning algorithm has three stages: Stage I - clustering, Stage II - tessellations, and Stage III - zone refinement. Here we describe the details of each stage. First, given a point pattern as all residents living in a given municipality, a density-based clustering algorithm is performed to identify the natural clustering of residential areas. The constraint for defining a cluster is set as 2000 to 3000 people in each data zone. The map is divided into sub-areas, as captured by the clustering results, each of them would then feed into the tessellation step. In the second stage, adaptive binning of Voronoi tessellation was performed on each sub-area. The objective function is to break down the sub-areas into smaller zones so that the number of people in each zone is close to the target number (2500 individuals in this study). The third stage is then to refine every zone within the municipality, merging neighboring and splitting zones that failed to reach the target. The Zoning algorithm was implemented in R version 3.5.0.

Data zones were constrained to nest within the 98 municipalities in Denmark. We aimed for a mean data zone size of 2500 individuals. A minimum number of residents of 1200 was chosen to ensure confidentiality for all neighborhood-level counts; any numerator in any year in any data zone must exceed five[6,7]. The procedure was applied to the annual mean number of residents at each address from 1980 to 2016. When we applied this homogenously population-sized geographic division to Denmark, the mean of annual number of residents per data zone was 2820 people (sd=809, range from 768 to 6495). Interest is also on the annual numbers, for example in 2005, the mean residents per data zone was 2869 people (sd=840, range from 793 to 7564).

eFigure 1 shows the geographic division of Denmark into data zones. The red lines are the 98-level municipality borders.

The shape file defining the data zones may be used for research purposes by contacting the first author (cbp@econ.au.dk).

eFigure 1: Geographic delineation of Denmark into data zones



Red lines show municipality borders (98 level), black lines show the 1885 data zones of homogeneous population size, with the upper left inset depicting the island of Bornholm. The upper right inset zooms in on Copenhagen and Frederiksberg municipalities.

eMethods 2: Delineation of neighborhood-level indicators

To extract information on neighborhood-level indicators, we assessed information on all people residing in Denmark on December 31 annually from 1981 to 2016. The population was divided into data zones that were nested within the 98 Danish municipalities. For each year, we first extracted information on individual-level covariates for all Danish residents on December 31. Thereafter, individual-level/household-level covariates were aggregated by data zone and year to calculate data zone-level constructs. Depending on the setting, we calculated the proportion of individuals (PI) or the proportion of households (PH). Unless noted otherwise we used the full population for calculation the data zone-level constructs. In detail:

Residential data zone
The delineation of residential data zone was based on each resident's geocoded residential information on December 31 annually from 1981 to 2016. Longitudinal geocoded information on people who have resided in Denmark is complete on a municipality and country-level from 1971 onwards and on an address-level from 1978 onwards[8]. Therefore, the earliest period from which we access complete information in a data zone is from 1978 onward.

Age distribution
We calculated the proportion of individuals in the following age groups: 0-14, 15-29, 30-49, 50-64 and 65 years and above. These values sum to unity. This information originates from the Danish Civil Registration System and is complete from 1978 onwards. From an epidemiological perspective we were interested in the potential causal effect of the neighborhood-level constructs examined. Therefore, in Table 3 in the paper, the effect of the neighborhood-level constructs were adjusted for neighborhood-level age structure.

Population Density
We calculated the population density as the number of inhabitants per square kilometer. Population density was modelled as a log-transformed continuous variable. This information was available from 1978 onwards (Table 4: Population density).

Urbanicity
Urbanicity was categorized according to each cohort member's residential address: capital, capital suburb, provincial city with more than 100,000 inhabitants, provincial town, and rural area. This classification was based on the 270-level classification of municipalities in 2006, and do not completely nest within the 98-level classification of municipalities. To ensure that urbanicity was nested within the 98-level municipality structure, urbanicity was redefined as the most frequent original urbanicity measure among people residing within each data zone at December 31, 2005. Around 5.5% of people had their urbanicity redefined. Sensitivity analyses show identical effects of original and modified urbanicity classification on schizophrenia risk. This information was available from 1978 onwards (Table 4: Urbanicity)

Material deprivation: Low income
We calculated the proportion of individuals aged 25 years and above with low gross income. It was delineated as individuals in lowest gross income quartile based on the distribution of gross income by year, gender and year of birth. This information stem from Integrated Database for Labour Market Research (IDA)[9] and is complete from 1981 onwards (Table 3: PI Low income).

In sensitivity analyses, we calculated low disposable income leading to similar results. Disposable income is the gross income minus tax and interest plus estimated rental value of residence (for homeowners).

Material deprivation: Short education
We calculated the proportion of individuals aged 25 years and above with primary school only, corresponding to a maximum length of education of seven years for people born before January 1, 1958, and nine years for people born on or after that date. Also, this was based on people born from 1921 or later due to missing information before that date. This information is based on IDA and complete from 1981 onwards (Table 3: Short education).

Material deprivation: Not employed
We calculated the proportion of individuals aged 25 years and above who were not employed. This information stem from IDA and was complete from 1981 onwards (Table 3: PI not employed). We chose the category "not employed" as opposed to "unemployed" to minimize the influence of political and organizational changes in

employment status during the study period (employed, unemployed, students, disability pension, early retirement, retired, or not working for another reason).

Material deprivation: Manual workers
We calculated the proportion of individuals aged 25 years and above who were manual workers. This information stem from IDA and was complete from 1981 onwards (Table 3: PI manual work)

Material deprivation: Overcrowded household
Number of people per room was calculated from the number of people in the family divided by the number of rooms shared by the family. Crowding was defined as the proportion of households with more than one person per room. This information is available from 1981 onwards (Table 3: PH overcrowded).

Material deprivation: No car owned in household
Households were classified as owning a car if a family member was registered as the owner of a car primarily used for private transportation. We calculated the proportion of households without a car. Information on car ownership was available from 2003 onwards (Table 3: PH no car owned).

Social fragmentation: Lone adult household
We calculated the proportion of households where an adult lived alone with or without children. This was based on Statistics Denmark's classification of family type (D and E family type). This information is complete from 1981 onwards (Table 3: PH lone adult).

Social fragmentation: Rents home
We calculated the proportion of households who lived in a rented dwelling. Information was available from 2000 onwards (Table 3: PH rents home).

Social fragmentation: Residential transience
We calculated the proportion of individuals aged 5 years or older with 2 or more residential changes during the past five years[10]. Residential transience was measured as a movement of address across municipality boundaries according to the 98-level municipality structure. This information originated from in the Danish Civil Registration System and is complete from 1978 onwards (Table 3: PI residential transience).

Social marginalization: Violent offending
We calculated the proportion of individuals aged 15 years and above who during the past 10 years had a conviction of a violent crime. Violent Crimes includes homicide, assault, robbery, aggravated burglary or arson, possessing a weapon in a public place, violent threats, extortion, human trafficking, abduction and kidnapping, rioting and other public order offenses, terrorism, and sexual offenses[11]. Information of violent crimes stem from the National Crime Register, which was available from 1980 onwards. Information on violent criminality among people aged 15 years and above was complete from 1991 onwards (Table 3: PI Violent offender).

Social marginalization: Any criminality
We calculated the proportion of individuals aged 15 years and above who during the past 10 years had a conviction of any crime (including violent crimes). Information of crimes stem from the National Crime Register, which was available from 1980 onwards. Information on any crime among people aged 15 years and above was complete from 1991 onwards (Table 3: PI Crime).

Social marginalization: Born abroad
We calculated the proportion of individuals who were foreign born in each data zone[12]. This information originates from in the Danish Civil Registration System and is complete from 1978 onwards (Table 3: PI born abroad).

Physical illness (Charlson comorbidity index)
Information on comorbid conditions was obtained from the National Patient Register with an admission date during the past 10 years. We calculated proportion of individuals aged 25 years and above with at least one disease included in the Charlson comorbidity index[13]. The following diseases are included in the Charlson comorbidity index:

Disease	ICD-8	ICD-10
Myocardial infarction	410	I21, I22, I23
Congestive heart failure	427.09, 421.10, 427.11, 427.19, 428.99, 782.49	I50, I11.0, I13.0, I13.2
Peripheral vascular disease	440, 441, 442, 443, 444, 445	I70, I71, I72, I73, I74, I77
Cerebrovascular disease	430-438	I60-I69, G45, G46

Dementia	290.09-290.19, 293.09	F00-F03, F05.1, G30
Chronic pulmonary disease	490-493, 515-518	J40-J47, J60-J67, J68.4, J70.1, J70.3, J84.1, J92.0, J96.1, J98.2, J98.3
Connective tissue disease	712, 716, 734, 446, 135.99	M05, M06, M08, M09, M30, M31, M32, M33, M34, M35, M36, D86
Ulcer disease	530.91, 530.98, 531-534	K22.1, K25-K28
Mild liver disease	571, 573.01, 573.04	B18, K70.0-K70.3, K70.9, K71, K73, K74, K76.0
Diabetes I and II	249.00, 249.06, 249.07, 249.09, 250.00, 250.06, 250.07, 250.09	E10.0, E10.1, E10.9, E11.0, E11.1, E11.9
Hemiplegia	344	G81, G82
Moderate to severe renal disease	403, 404, 580-584, 590.09, 593.19, 753.10-753.19, 792	I12, I13, N00-N05, N07, N11, N14, N17-N19, Q61
Diabetes with end-organ damage	249.01-249.05, 249.08, 250.01-250.05, 250.08	E10.2-E10.8, E11.2-E11.8
Any tumor	140-194	C00-C75
Leukemia	204-207	C91-C95
Lymphoma	200-203, 275.59	C81-C85, C88, C90, C96
Moderate to severe liver disease	070.00, 070.02, 070.04, 070.06, 070.08, 573.00, 456.00-456.09	B15.0, B16.0, B16.2, B19.0, K70.4, K72, K76.6, I85
Metastatic solid tumor	195-198, 199	C76-C80
AIDS	079.83	B21-B24

This information is complete from 1987 onwards (Table 3: PI Physical illness)

Aggregation from individual-level to neighborhood-level constructs

For the aggregation of short education, gross income, not employed, household overcrowding and one adult household, we excluded inhabitants/households with missing data from the aggregation (less than 5%). For the aggregation of rented accommodation, households with missing data were included as a separate category (since the proportion with missing data was larger than 5%).

The final dataset contains one record for each data zone (1885 possibilities) and year (36 possibilities) yielding a total of 67,860 records. eTable 1 show the distribution of the neighborhood-level constructs examined for a single year.

To reduce uncertainty of neighborhood level covariates and allow for principal component analyses of neighborhood level indicators, for each data zone we used the mean of each neighborhood level construct from 1981 to 1984 [sensitivity analyses show minor increase in effect of neighborhood level constructs as compared to using year at initiation of follow-up]. A few constructs were based on other periods due to lack of historical data: no car owned (2003-2009), rented accommodation (2000-2009), Charlson Comorbidity Index (1987-1989), violent offending (1991-1999), any criminal offending (1991-1994). We acknowledge that these five constructs were measured after initiation of follow-up, although in most instances individual people have no impact on the neighborhood covariations. Whenever possible temporality[14], should be satisfied for both individual-level and neighborhood-level variables.

Geographic maps for all neighborhood-level constructs are shown in eFigures 2A - 2S, however, distribution of violent criminality do not obey operational Danish confidentiality measures and is therefore not shown.

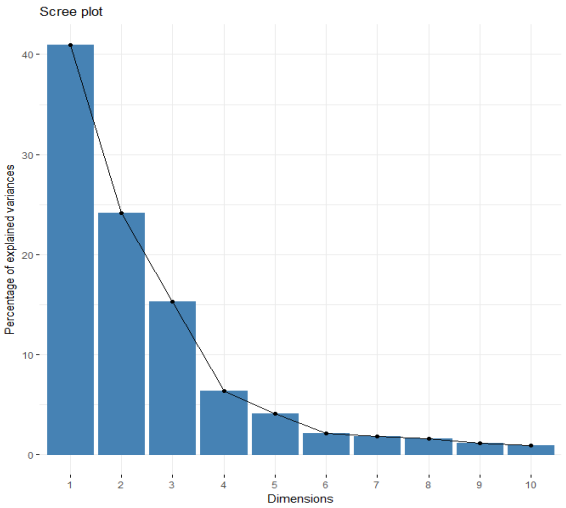
Principal Component Analyses of neighborhood level covariates

Principal component analyses of all neighborhood-level constructs. All neighborhood-level constructs were first weighted averaged 1981-1984, standardized to zero mean and unit standard deviation. Principal components with eigenvalue > 1 were included for further analyses (Table 3). Resulting principal componentss are not standardized. The Principal component analyses was fit using the inbuilt *prcomp* function in R version 3.5.0.

The eigenvalues and variances explained are:

```
       eigenvalue variance.percent cumulative.variance.percent
Dim.1  5.32629358       40.9714891                    40.97149
Dim.2  3.13808836       24.1391413                    65.11063
Dim.3  1.99178298       15.3214075                    80.43204
Dim.4  0.83203365        6.4002588                    86.83230
Dim.5  0.53202786        4.0925220                    90.92482
Dim.6  0.27342908        2.1033006                    93.02812
Dim.7  0.23915092        1.8396225                    94.86774
Dim.8  0.20409671        1.5699747                    96.43772
Dim.9  0.15166479        1.1666523                    97.60437
Dim.10 0.11790609        0.9069699                    98.51134
Dim.11 0.09068031        0.6975408                    99.20888
```

```
Dim.12 0.06618973        0.5091517                   99.71803
Dim.13 0.03665595        0.2819688                  100.00000
```



Rotation for the principal components:

```
Rotation (n x k) = (13 x 13):
                              PC1         PC2         PC3         PC4
Material deprivation
pi_lowincome            0.1959763 -0.38585831  0.03117006 -0.46982691
pi_shorteducation       0.0666679 -0.52589927  0.05246428 -0.15813243
pi_notemployed          0.3460190 -0.13332742  0.34175726 -0.05754973
pi_manualwork           0.1083769 -0.50525746 -0.08673057  0.13871054
ph_overcrowded          0.0205746 -0.25965427 -0.43024404  0.57011985
ph_nocarowned           0.3891485  0.17589751  0.08378031  0.16813436


Social fragmentation
ph_loneadult            0.3908363  0.12621700  0.17353039 -0.10868726
ph_rentshome            0.3710701  0.15736477  0.04811219  0.32074116
pi_residentialtransience 0.2538291  0.22950161 -0.34826109 -0.37278455


Social marginalization
pi_violentoffender      0.3407765 -0.11000331 -0.23216052  0.18819847
pi_crime                0.2413956 -0.13139713 -0.42580476 -0.19804749
pi_bornabroad           0.2698703  0.27176173 -0.29249222 -0.08608289

Physical health
pi_physicalillness      0.2677843 -0.08084546  0.45315933  0.20661691
```

eTable 1: National distribution of neighborhood-level indicators by urbanicity.

2005	Urbanicity			
	Capital	Provincial	Rural areas	**Overall**
Number of residents	1,397,182	2,182,144	1,828,235	5,407,561
Number of households	774,815	1,140,208	860,862	2,775,885
Number of data zones	478	744	663	1,885
Material deprivation				
PI low income	24.0 (8.6)	24.9 (6.8)	25.6 (5.1)	24.9 (6.8)
PI short education	18.2 (8.0)	24.0 (7.7)	30.4 (6.8)	24.8 (8.8)
PI not employed	36.4 (8.2)	38.7 (9.1)	36.3 (8.0)	37.3 (8.6)
PI manual work	4.6 (2.3)	4.7 (1.4)	5.2 (1.4)	4.9 (1.7)
PH overcrowded	5.0 (2.3)	4.0 (2.1)	3.8 (0.9)	4.2 (1.9)
PH no car owned	51.9 (17.8)	38.5 (14.7)	28.1 (7.5)	38.2 (16.4)
Social fragmentation				
PH lone adult	54.2 (14.9)	47.0 (13.0)	40.4 (7.7)	46.5 (13.1)
PH rents home	54.5 (22.3)	42.2 (18.3)	28.6 (10.2)	40.5 (19.9)
PI residential transience	11.0 (4.8)	7.2 (3.8)	6.0 (1.5)	7.8 (4.0)
Social marginalization				
PI violent offending	1.2 (0.7)	1.2 (0.7)	0.9 (0.4)	1.1 (0.6)
PI any criminality	16.2 (3.0)	15.8 (2.9)	17.2 (3.0)	16.4 (3.0)
PI born abroad	12.0 (6.0)	7.3 (5.4)	4.4 (1.8)	7.5 (5.5)
Physical health				
PI Physical illness	11.5 (3.0)	12.2 (2.6)	12.4 (2.5)	12.1 (2.7)
Population density	5,793 (6,585)	1,448 (1,739)	271 (313)	2,136 (4,124)
Age distribution				
PI 0-14 years	17.7 (4.3)	18.4 (4.7)	20.1 (3.0)	18.8 (4.2)
PI 15-29 years	18.9 (7.7)	18.4 (8.6)	13.9 (2.0)	16.9 (7.1)
PI 30-49 years	30.5 (3.8)	28.1 (2.9)	28.5 (2.9)	28.9 (3.3)
PI 50-64 years	18.5 (4.3)	20.0 (4.1)	21.5 (2.8)	20.2 (3.9)
PI 65+ years	14.3 (5.6)	15.1 (4.8)	16.0 (4.9)	15.2 (5.1)

PI: Proportion of Individuals, values are shown for 2005

PH: Proportion of Households, values are shown for 2005

Except for the first three rows, numbers are averages and standard deviations (in parenthesis) across data zones.

For the purpose of this table only, urbanicity was categorized as capital (city and its suburb), provincial (city or town), and rural areas.

eMethods 3: Sensitivity analyses using parish boundaries as small area delineators

To explore the potential impact on the observed results of using data zones versus other neighborhoods, we repeated all analyses using parish of the neighborhood of interest. To ensure confidentiality we first clustered small population-sized parishes with neighboring parishes to obtain a minimum of 750 residents in each parish 1980-2016. We obtained 1045 clustered parishes (mean: 5175 residents, IQR: 2924 to 6563; Range: 2005 to 37,872 residents)[value in 2005]. Neighborhood-level indicators were re-calculated based on these clustered parishes and new multilevel models were estimated. We obtained nearly identical effect estimates for all parameters, except that a) the general contextual effect for clustered parishes were slightly smaller than that for data zones, b) the specific contextual effects for clustered parishes were attenuated towards nil compared to identical specific contextual effects for data zones, and c) while urbanicity was more important than population density when neighboring boundaries were defined by data zones, population density was more important than urbanicity when neighboring boundaries were defined by clustered parishes.

Estimates presented in Tables 3 and 4 but using clustered parish as the neighborhood of interest are repeated in eTable 2 and eTable 3.

eTable 2: Neighborhood-level contextual effects - parish level

	Basic adjustment[a;b]		Full adjustment[a;c]	
	Specific contextual effect	General contextual effect	Specific contextual effect	General contextual effect
Neighborhood-level socio economic indicator[d]	Incidence Rate Ratio of neighborhood per 1 sd increase	Median Incidence Rate ratio (95% Probability Interval)	Incidence Rate Ratio of neighborhood per 1 sd increase	Median Incidence Rate Ratio (95% Probability Interval)
None	-	1.33 (1.27-1.38)		1.27 (1.22-1.33)
Material deprivation				
PI low Income	0.95 (0.90-0.99)	1.33 (1.27-1.38)	0.92 (0.88-0.96)	1.27 (1.21-1.32)
PI short education	0.90 (0.85-0.94)	1.32 (1.27-1.38)	0.88 (0.83-0.92)	1.26 (1.20-1.32)
PI not employed	1.24 (1.14-1.35)	1.31 (1.25-1.36)	1.13 (1.04-1.22)	1.27 (1.21-1.33)
PI manual work	0.99 (0.95-1.04)	1.33 (1.27-1.39)	0.96 (0.92-1.00)	1.27 (1.22-1.33)
PH overcrowded	1.12 (1.08-1.16)	1.29 (1.24-1.35)	1.08 (1.05-1.12)	1.25 (1.20-1.31)
PH no car owned	1.35 (1.28-1.42)	1.23 (1.17-1.29)	1.26 (1.20-1.33)	1.21 (1.15-1.27)
Social fragmentation				
PH lone adult	1.38 (1.29-1.49)	1.26 (1.21-1.32)	1.27 (1.19-1.36)	1.24 (1.18-1.30)
PH rents home	1.24 (1.19-1.30)	1.25 (1.19-1.31)	1.18 (1.14-1.23)	1.23 (1.17-1.29)
PI residential transience	1.14 (1.10-1.19)	1.29 (1.23-1.35)	1.10 (1.06-1.14)	1.25 (1.20-1.31)
Social marginalization				
PI violent offending	1.15 (1.11-1.19)	1.27 (1.21-1.33)	1.10 (1.06-1.13)	1.25 (1.19-1.31)
PI any criminality	1.11 (1.07-1.14)	1.30 (1.24-1.35)	1.06 (1.02-1.09)	1.26 (1.21-1.32)
PI born abroad	1.19 (1.15-1.24)	1.27 (1.21-1.33)	1.14 (1.10-1.19)	1.24 (1.18-1.30)
Physical illness				
PI Physical illness	1.13 (1.06-1.20)	1.31 (1.26-1.37)	1.11 (1.04-1.17)	1.26 (1.21-1.32)
All area-level indices[e]	-	1.22 (1.16-1.28)		1.22 (1.15-1.28)

PI: Proportion of individuals

PH: Proportion of households

[a] Estimates were based on multilevel survival analyses, neighborhoods are 1045 clustered parishes. 579,039 people born in Denmark 1972-1981 were followed for development of schizophrenia. During the follow-up period from 1982 to 2016, a total of 5103 developed schizophrenia during the 17,191,889 person-years at risk

[b] Estimates were adjusted for individual-level age and its interaction with sex and neighborhood-level age-distribution (Basic adjustment).

[c] Estimates were adjusted for Basic adjustment and individual-level residential instability, parental Charlson, parental death, parental imprisonment, parental age, parental history of mental disorders, parental income, parental employment status and parental education.

[d] Neighborhood-level covariates measure the effect of a one standard deviation increase in the proportion of each covariate. Neighborhood-level covariates were data parish-level averages 1981-1984 (details in Emethods 2). The effects of each neighborhood-level covariate was modelled separately.

[e] Neighborhood-level covariates were summarized using the first 3 Principal Components of all neighborhood-level covariates (excluding urbanicity and population density). The fixed effect (specific contextual effect) estimate is not comparable to the other neighborhood-level fixed effects and is therefore not shown.

eTable 3: Urbanicity, population density and data zone level heterogeneity by individual- and neighborhood-level adjustment scenarios – parish level

	Basic adjustment[a;b]		Full adjustment[a;c]	
	Specific contextual effect	General contextual effect	Specific contextual effect	General contextual effect
Measure of urbanization[d]	Incidence Rate Ratio of neighborhood per 1 sd increase	Median Incidence Rate ratio (95% CI)	Incidence Rate Ratio of neighborhood per 1 sd increase	Median Incidence Rate Ratio (95% CI)
No adjustment for neighborhood-level covariates				
Population density	1.23 (1.17-1.29)	1.27 (1.21-1.33)	1.20 (1.14-1.25)	1.22 (1.16-1.28)
Urbanicity	1.74 (1.52-1.98)	1.27 (1.21-1.33)	1.62 (1.43-1.84)	1.22 (1.16-1.29)
Adjusted for neighborhood-level covariates[e]				
Population density	1.10 (1.03-1.18)	1.21 (1.15-1.27)	1.11 (1.04-1.19)	1.20 (1.13-1.26)
Urbanicity	1.18 (1.00-1.39)	1.21 (1.15-1.27)	1.20 (1.01-1.42)	1.21 (1.15-1.27)
Adjusted for neighborhood-level covariates[e] and other urbanization proxy				
Population density	1.09 (1.01-1.16)	1.21 (1.14-1.27)	1.09 (1.02-1.17)	1.20 (1.14-1.26)
Urbanicity	1.10 (0.92-1.31)	1.21 (1.14-1.27)	1.11 (0.93-1.34)	1.20 (1.14-1.26)

[a] Estimates were based on multilevel survival analyses, neighborhoods are 1045 clustered parishes. 579,039 people born in Denmark 1972-1981 were followed for development of schizophrenia. During the follow-up period from 1982 to 2016, a total of 5103 developed schizophrenia during the 17,191,889 person-years at risk

[b] Estimates were adjusted for individual-level age and its interaction with sex and neighborhood-level age-distribution (Basic adjustment).

[c] Estimates were adjusted for Basic adjustment and individual-level residential instability, parental Charlson, parental death, parental imprisonment, parental age, parental history of mental disorders, parental income, parental employment status and parental education.

[d] Proxy measure for urbanization. For each data zone, population density was calculated as the number of inhabitants divided by the area's population size, log-transformed and thereafter standardized to unit standard deviation. The estimate for urbanicity measures the effect of residence in the most urban environment compared to residence in the most rural environment. It was included as a trend variable scored as: capital 1, capital suburb 0.75, provincial city 0.5, provincial towns 0.25, and rural area 0. Both variables were delineated at initiation of follow-up. Due to the different scales, effect sizes are not directly comparable in size.

[e] Neighborhood-level indicators were summarized using the first 3 Principal Components of all neighborhood-level covariates (excluding urbanicity and population density).

eTable 4: Neighborhood-level contextual effects – schizophrenia and related disorders

	Basic adjustment[a;b]	
	Specific contextual effect	General contextual effect
Neighborhood-level socio economic indicator[d]	Incidence Rate Ratio of neighborhood per 1 sd increase	Median Incidence Rate ratio (95% Probability Interval)
None	-	1.37 (1.33-1.41)
Material deprivation		
PI low Income	0.94 (0.91-0.97)	1.36 (1.32-1.40)
PI short education	0.89 (0.86-0.92)	1.35 (1.31-1.39)
PI not employed	1.32 (1.24-1.40)	1.34 (1.30-1.38)
PI manual work	1.02 (0.98-1.05)	1.37 (1.33-1.41)
PH overcrowded	1.17 (1.14-1.21)	1.33 (1.29-1.37)
PH no car owned	1.44 (1.39-1.49)	1.25 (1.21-1.29)
Social fragmentation		
PH lone adult	1.51 (1.44-1.59)	1.28 (1.24-1.32)
PH rents home	1.29 (1.25-1.33)	1.27 (1.23-1.31)
PI residential transience	1.23 (1.19-1.27)	1.32 (1.28-1.36)
Social marginalization		
PI violent offending	1.21 (1.18-1.24)	1.31 (1.27-1.35)
PI any criminality	1.15 (1.12-1.18)	1.33 (1.30-1.37)
PI born abroad	1.23 (1.20-1.26)	1.30 (1.26-1.34)
Physical illness		
PI Physical illness	1.13 (1.08-1.18)	1.36 (1.32-1.40)
All area-level indices[e]		1.24 (1.20-1.28)

PI: Proportion of individuals

PH: Proportion of households

[a] Estimates were based on multilevel survival analyses, neighborhoods are 1885 novel data zones nested in Denmark's 98 municipalities. 579,039 people born in Denmark 1972-1981 were followed for development of schizophrenia or related disorders. During the follow-up period from 1982 to 2016, a total of 9289 developed schizophrenia or related disorders during the 17,140,418 person-years at risk

[b] Estimates were adjusted for individual-level age and its interaction with sex and neighborhood-level age-distribution (Basic adjustment).

[c] Estimates were adjusted for Basic adjustment and individual-level residential instability, parental Charlson, parental death, parental imprisonment, parental age, parental history of mental disorders, parental income, parental employment status and parental education.

[d] Neighborhood-level covariates measure the effect of a one standard deviation increase in the proportion of each covariate. Neighborhood-level covariates were data zone-level averages 1981-1984 (details in eMethods 2). The effects of each neighborhood-level covariate were modelled separately.

[e] Neighborhood-level covariates were summarized using the first 3 Principal Components of all neighborhood-level covariates (excl. urbanicity and population density). The fixed effect (specific contextual effect) estimate is not comparable to the other neighborhood-level fixed effects and is therefore not shown.

eTable 5: Urbanicity, population density and data zone level heterogeneity by individual- and neighborhood-level adjustment scenarios – schizophrenia and related disorders

	Basic adjustment[a;b]	
	Specific contextual effect	General contextual effect
Measure of urbanization[d]	Incidence Rate Ratio of neighborhood per 1 sd increase	Median Incidence Rate ratio (95% CI)
Population density	1.23 (1.19-1.27)	1.31 (1.27-1.35)
Urbanicity	1.96 (1.79-2.13)	1.30 (1.26-1.34)
Population density	1.07 (1.03-1.12)	1.24 (1.19-1.28)
Urbanicity	1.23 (1.09-1.38)	1.24 (1.19-1.28)
Population density	1.06 (1.01-1.10)	1.23 (1.19-1.27)
Urbanicity	1.18 (1.04-1.34)	1.23 (1.19-1.27)

[a] Estimates were based on multilevel survival analyses, neighborhoods are 1885 novel data zones nested in Denmark's 98 municipalities. 579,039 people born in Denmark 1972-1981 were followed for development of schizophrenia. During the follow-up period from 1982 to 2016, a total of 9289 developed schizophrenia during the 17,140,418 person-years at risk

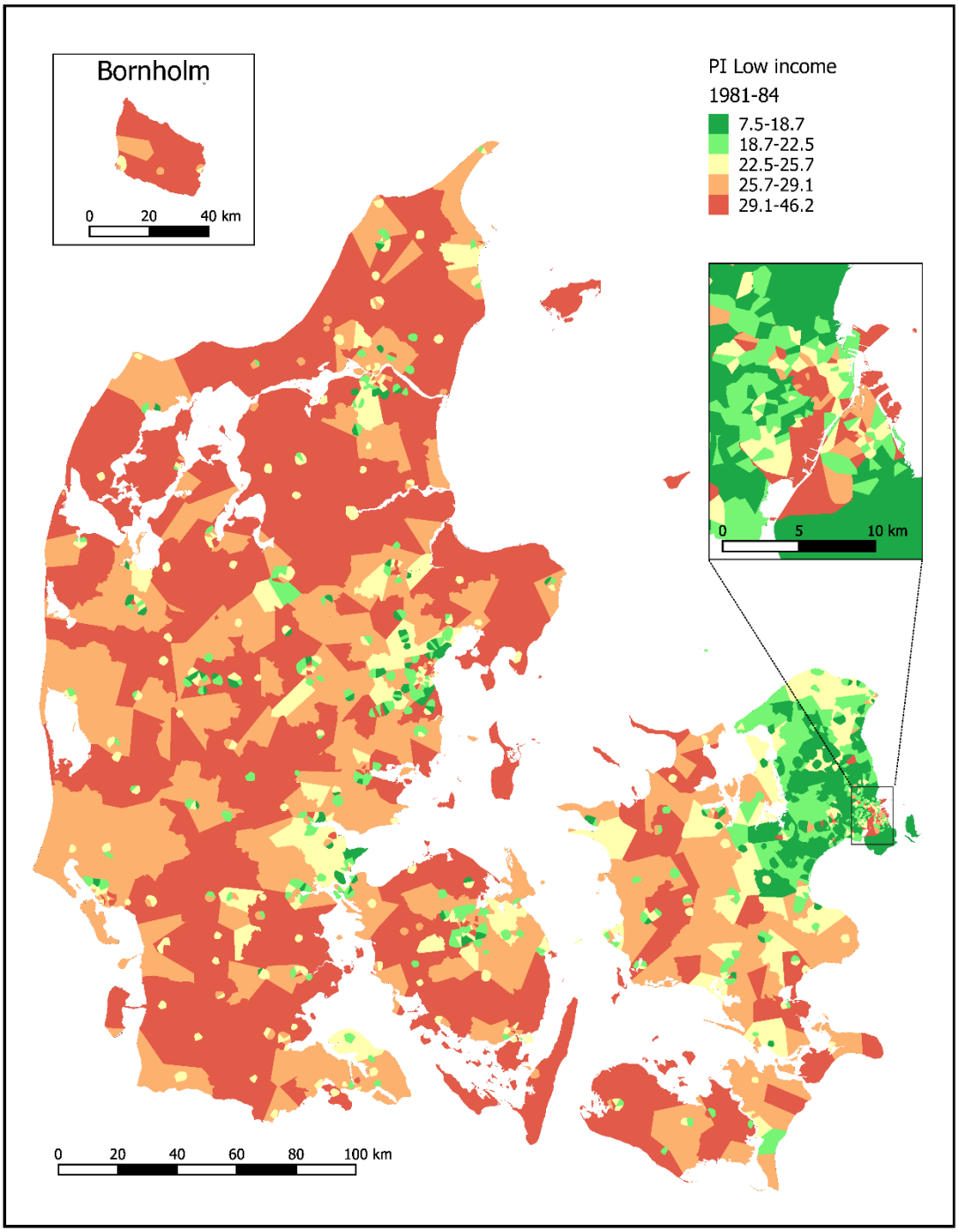
[b] Estimates were adjusted for individual-level age and its interaction with sex and neighborhood-level age-distribution (Basic adjustment).

[c] Estimates were adjusted for Basic adjustment and individual-level residential instability, parental Charlson, parental death, parental imprisonment, parental age, parental history of mental disorders, parental income, parental employment status and parental education.

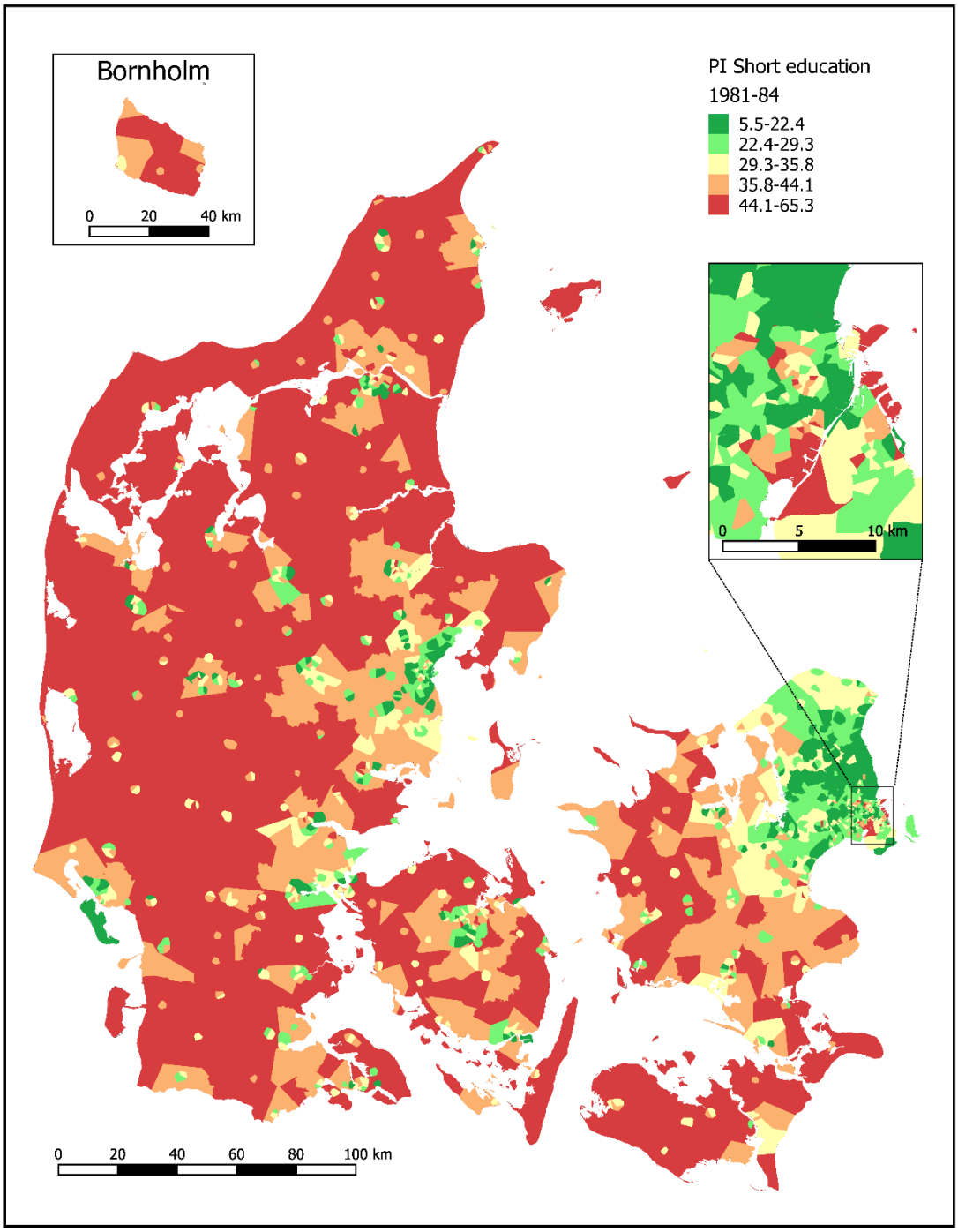
[d] Proxy measure for urbanization. For each data zone, population density was calculated as the number of inhabitants divided by the area, log transformed and thereafter standardized to unit standard deviation. The estimate for urbanicity measure the effect of residence in the most urban environment compared to residence in the most rural environment. It was included as a trend variable scored as Capital 1, capital suburb 0.75, provincial city 0.5, provincial towns 0.25, and rural area 0. Both variables were delineated at initiation of follow-up. Due to the different scales, effect sizes are not directly comparable in size.

[e] Neighborhood-level indicators were summarized using the first 3 Principal Components of all neighborhood-level covariates (excl. urbanicity and population density).

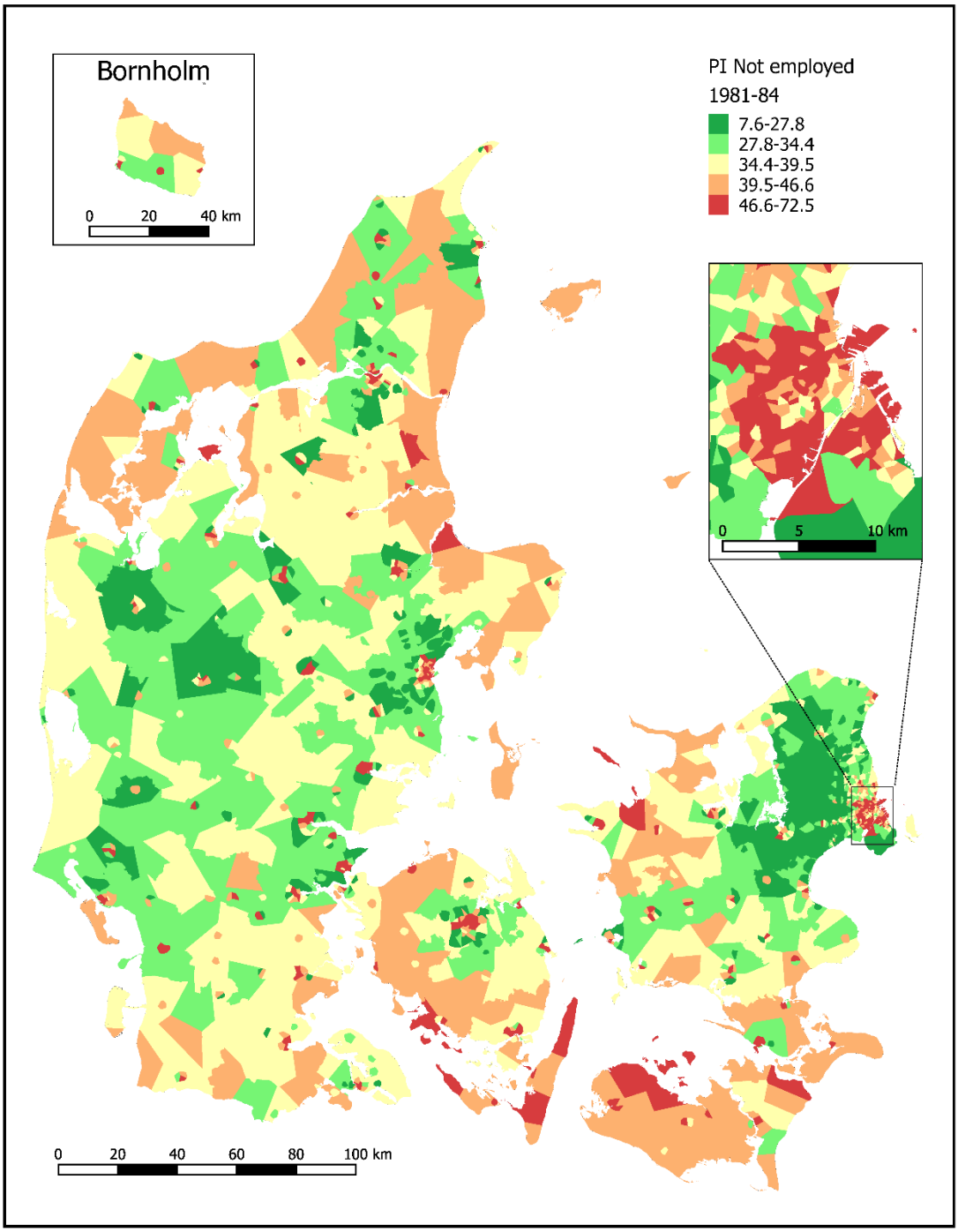
eFigure 2A: Material deprivation: Low income



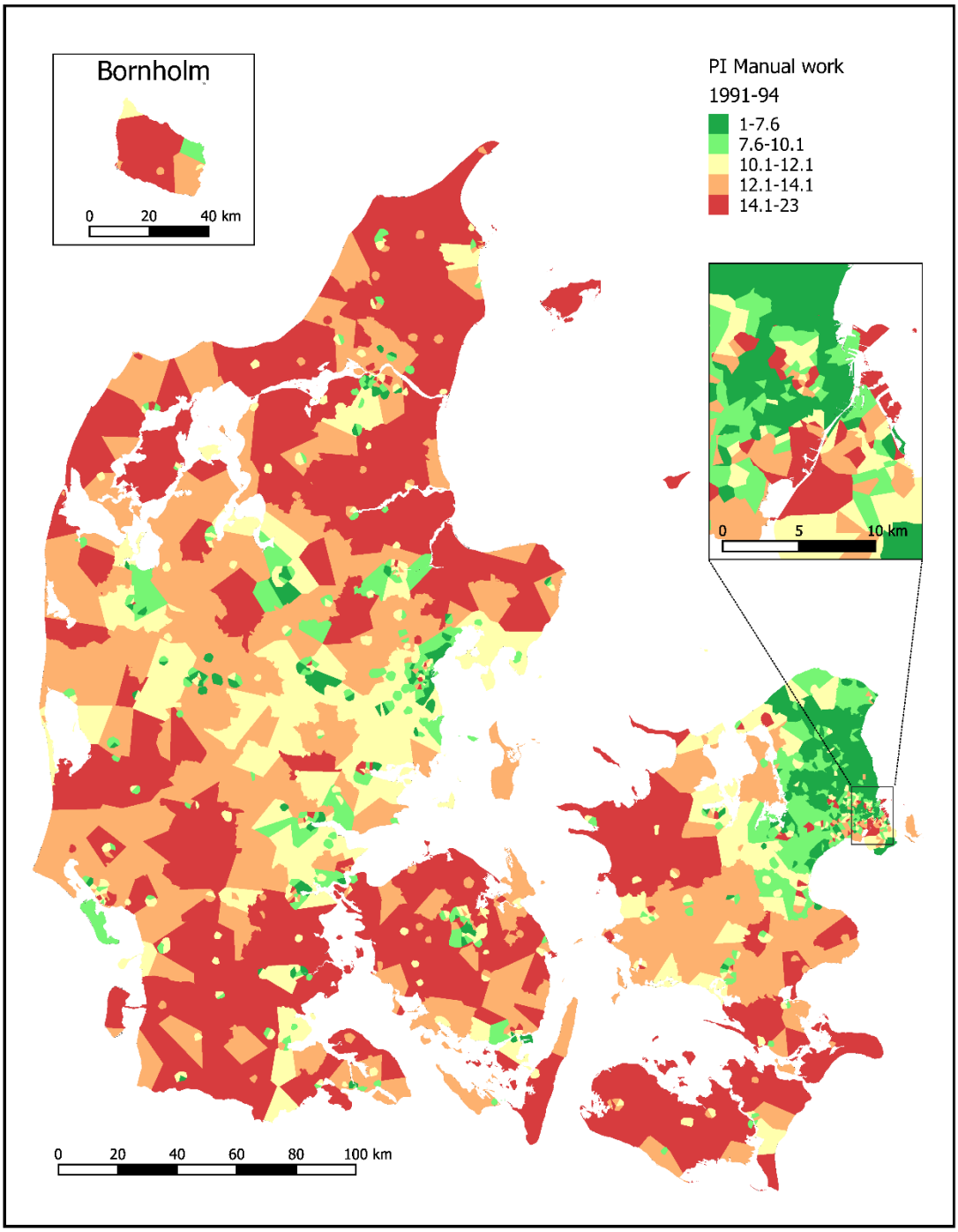
eFigure 2B: Material deprivation: Short education



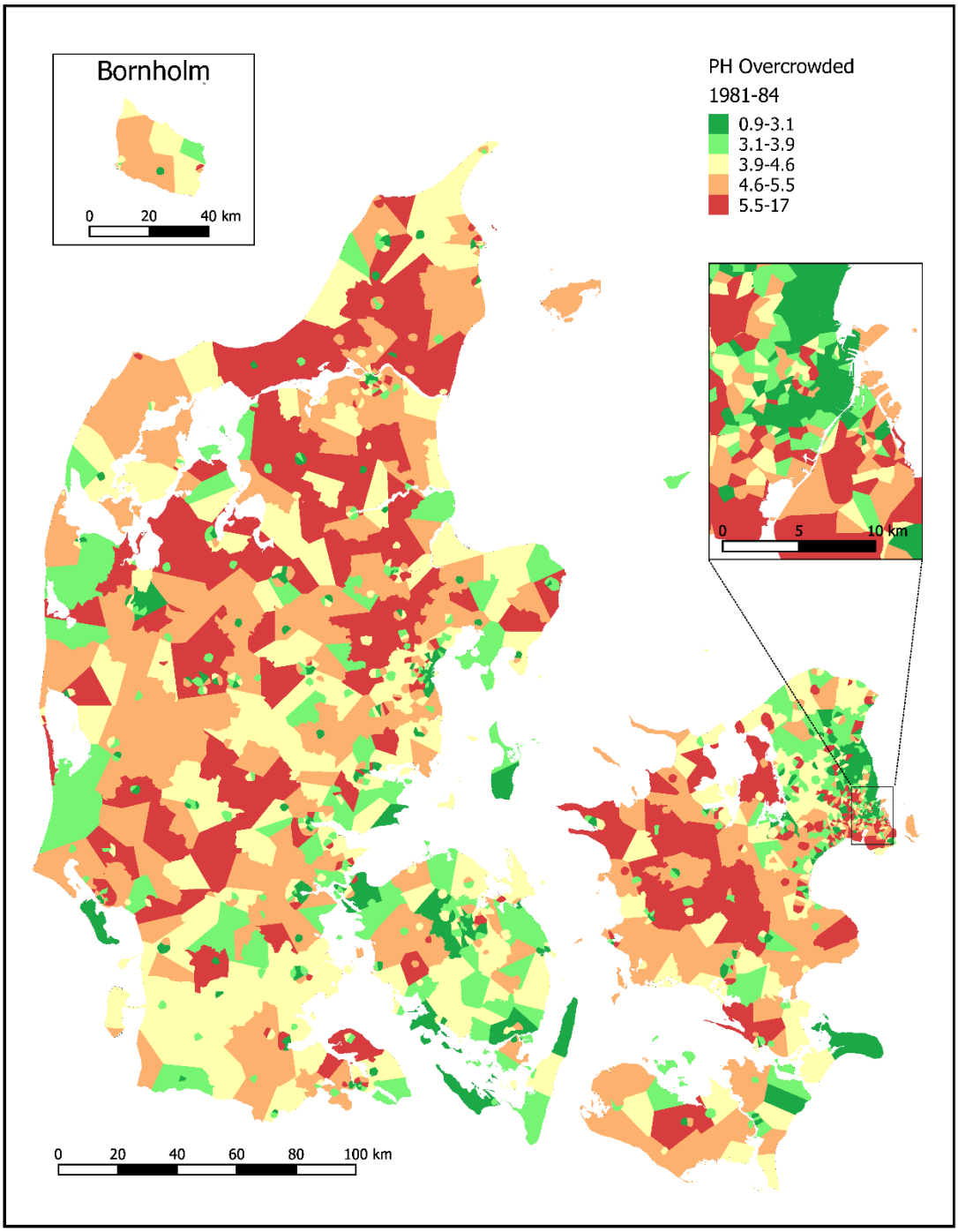
eFigure 2C: Material deprivation: Not employed



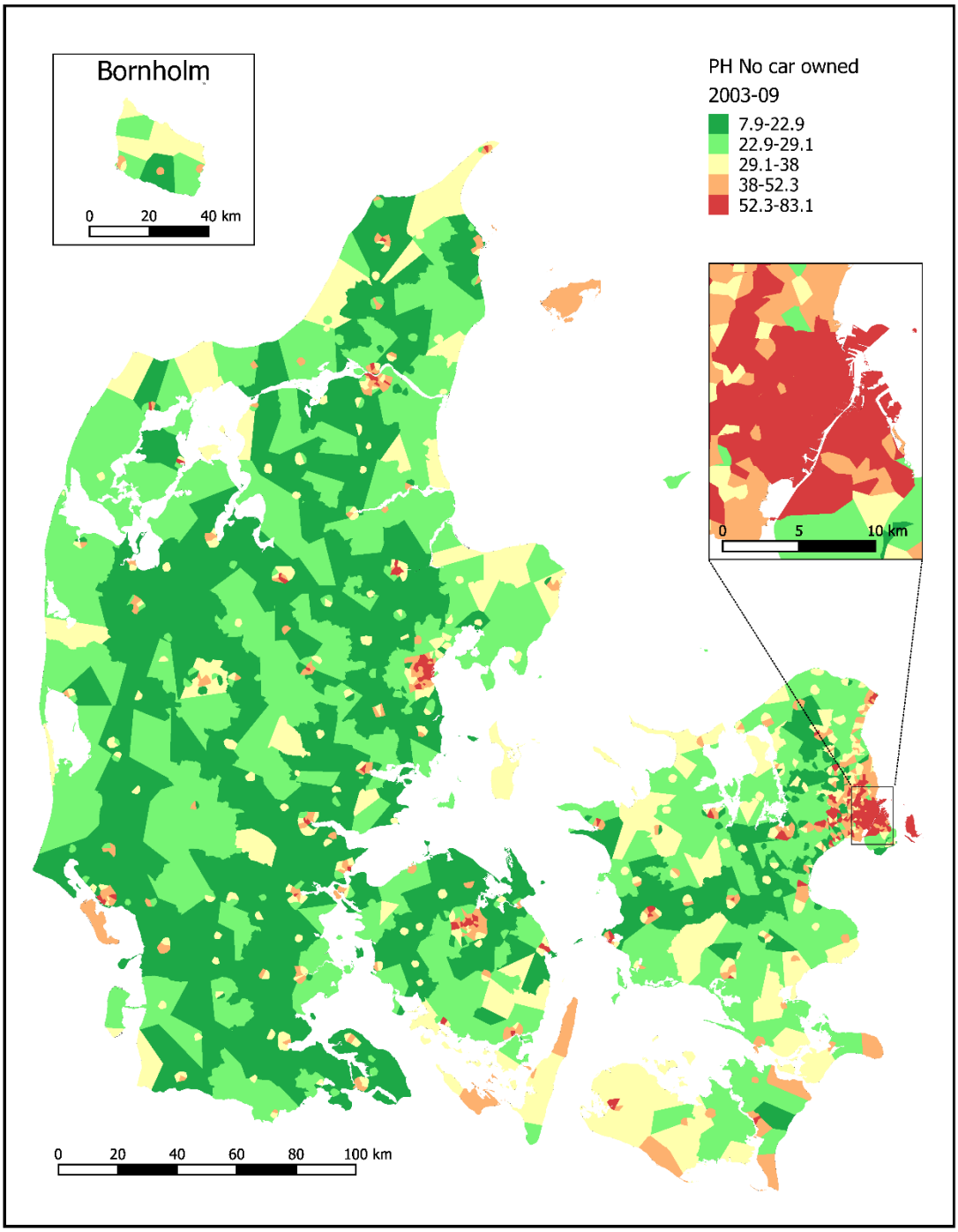
eFigure 2D: Material deprivation: Manual work



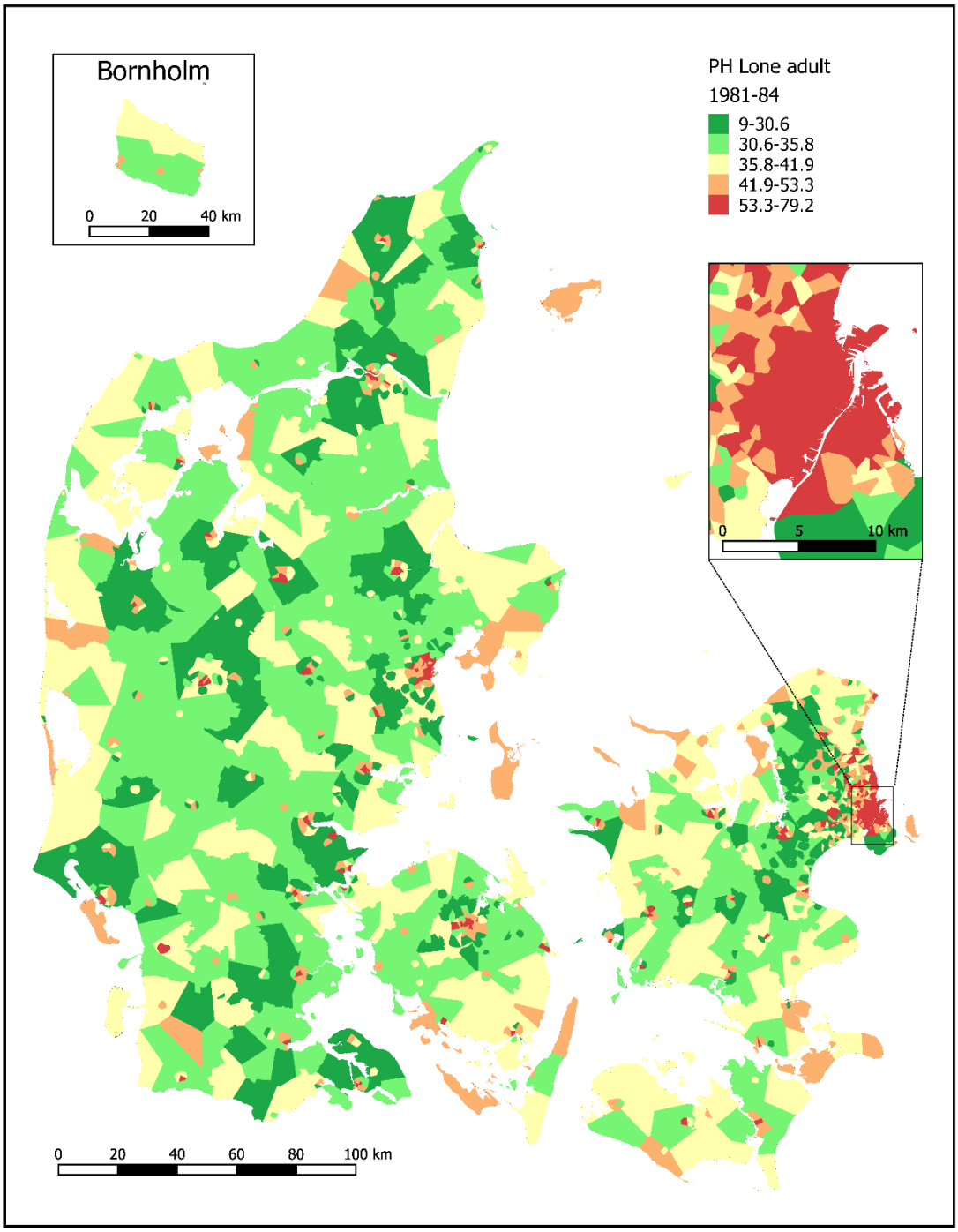
eFigure 2E: Material deprivation: Overcrowded household



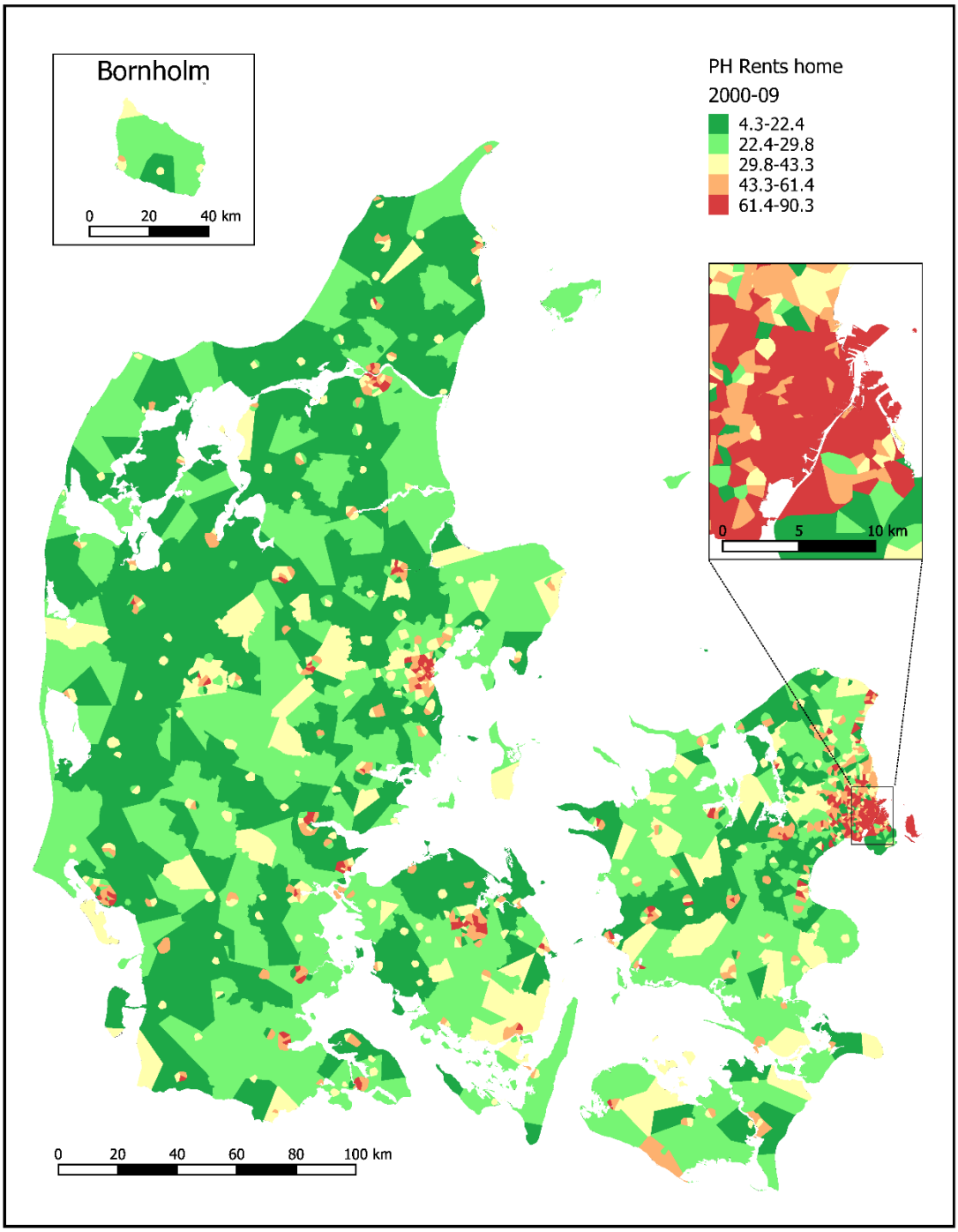
eFigure 2F: Material deprivation: No car owned in household



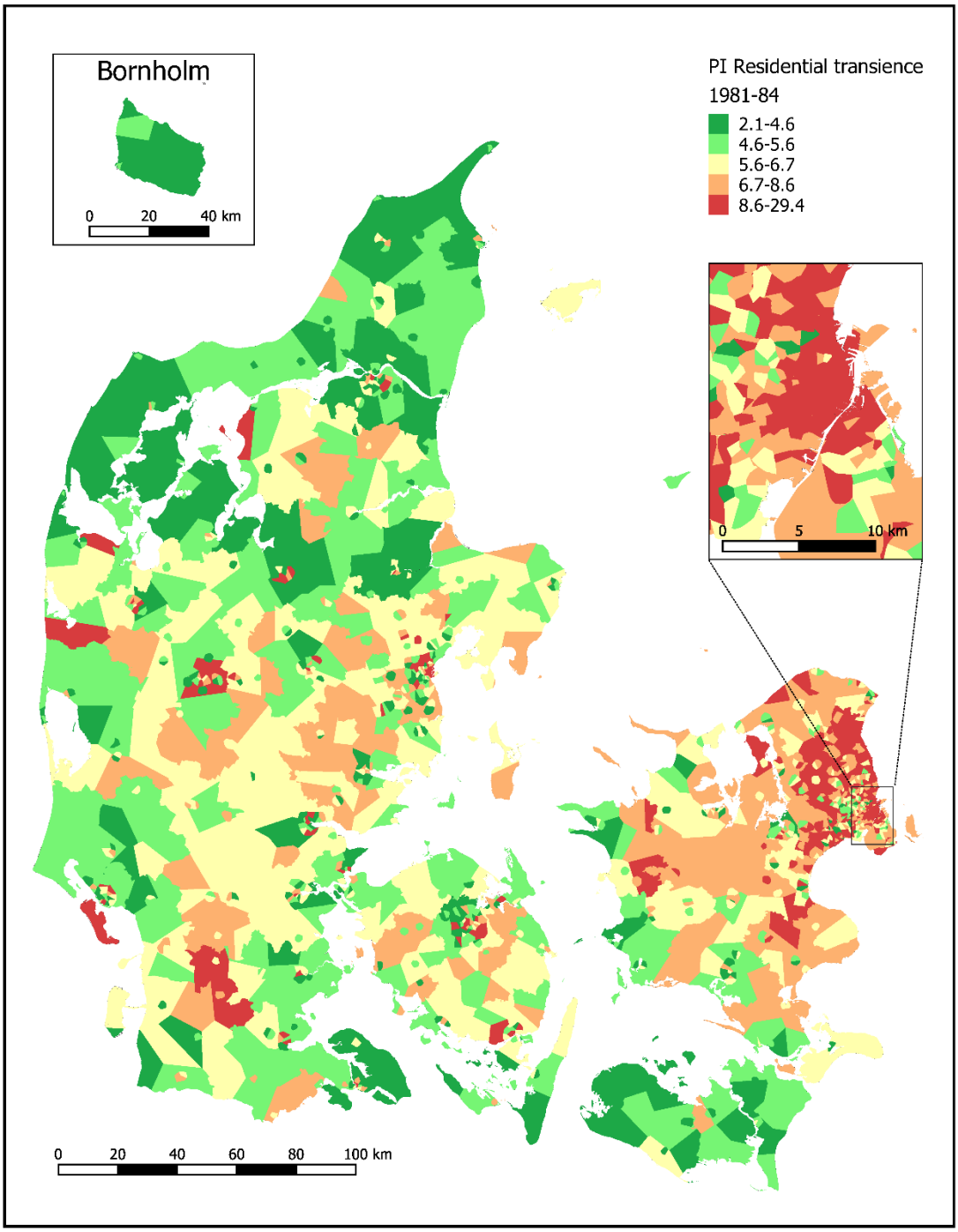
eFigure 2G: Social fragmentation: Lone adult household



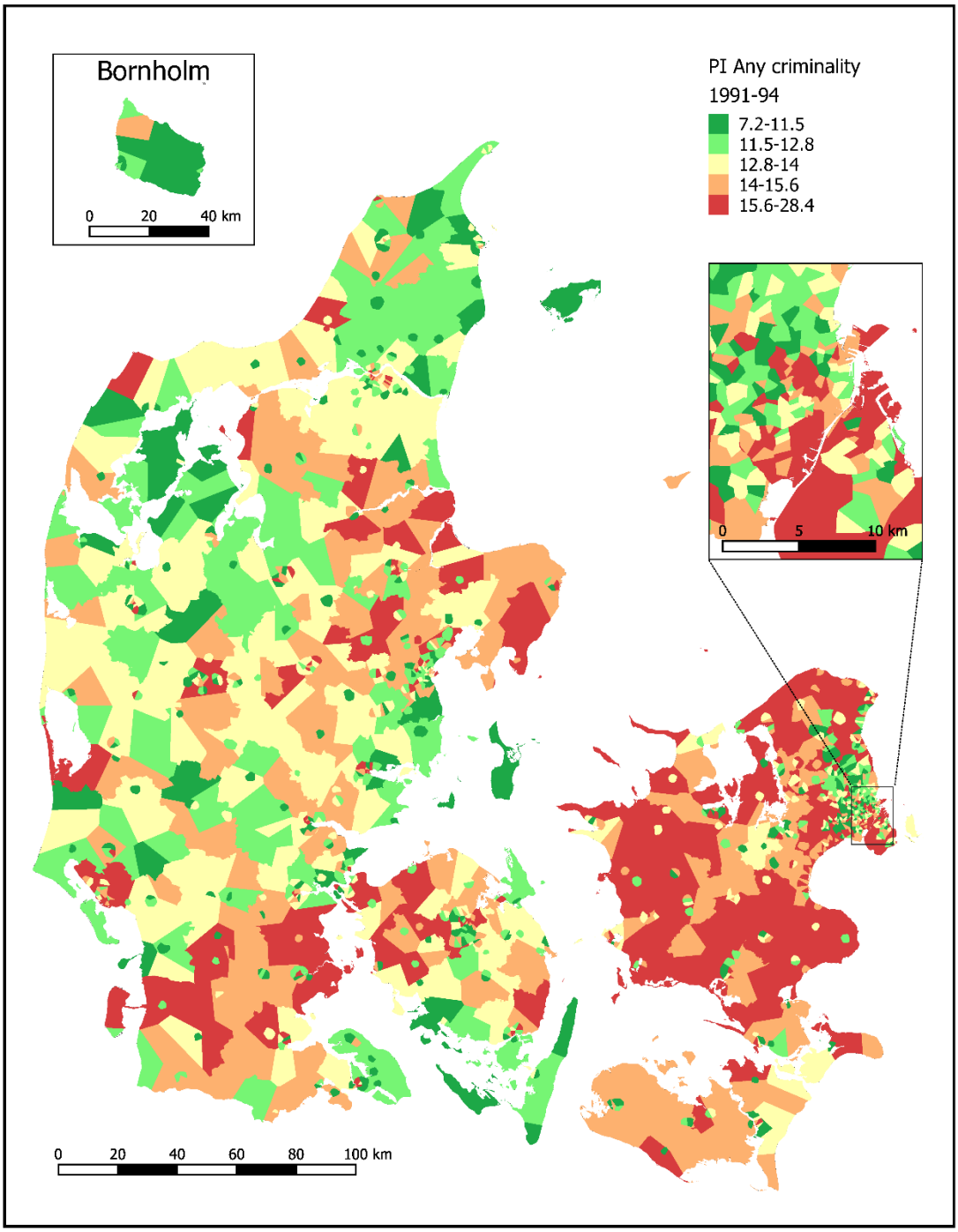
eFigure 2H: Social fragmentation: Rents home



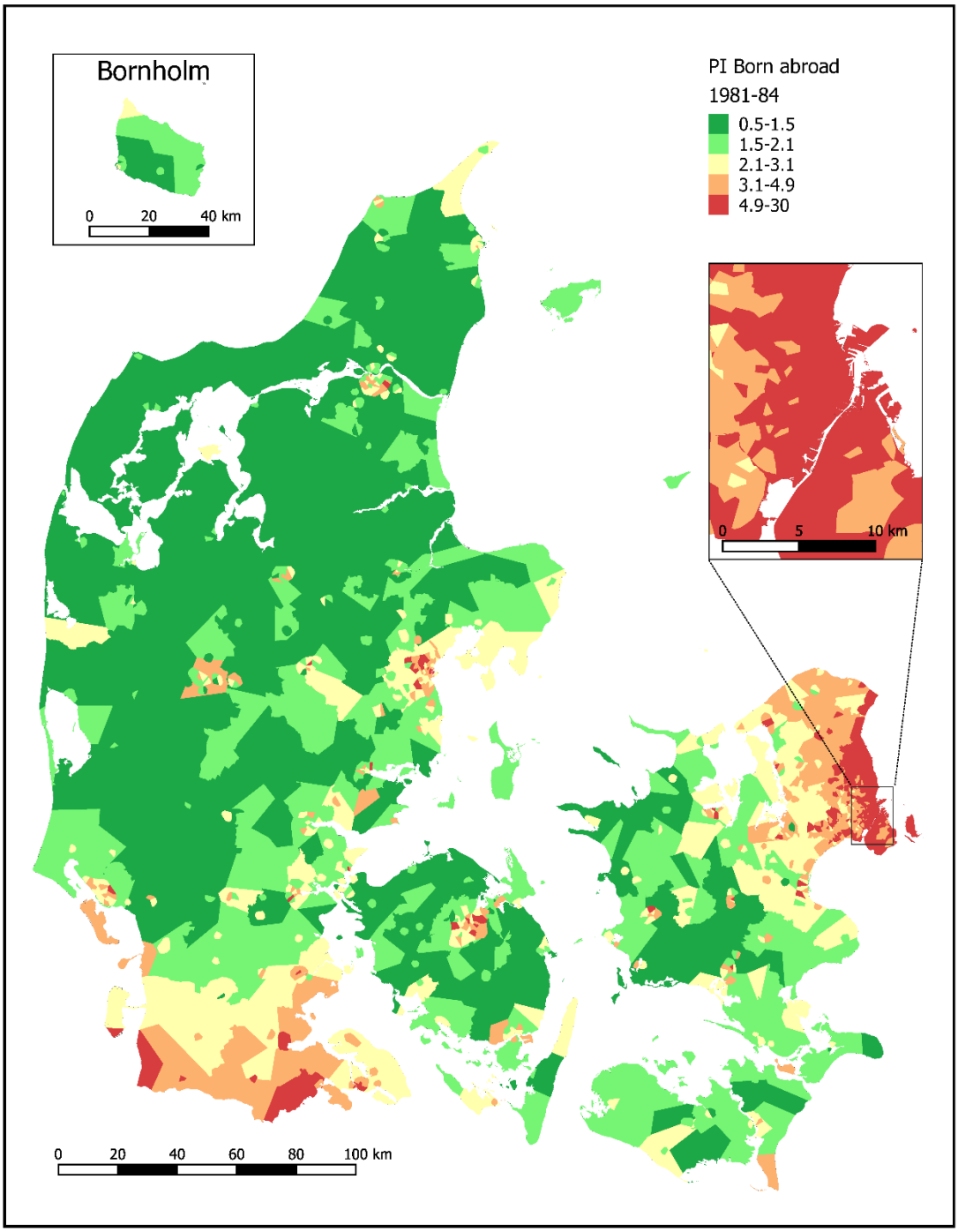
eFigure 2I: Social fragmentation: Residential transience



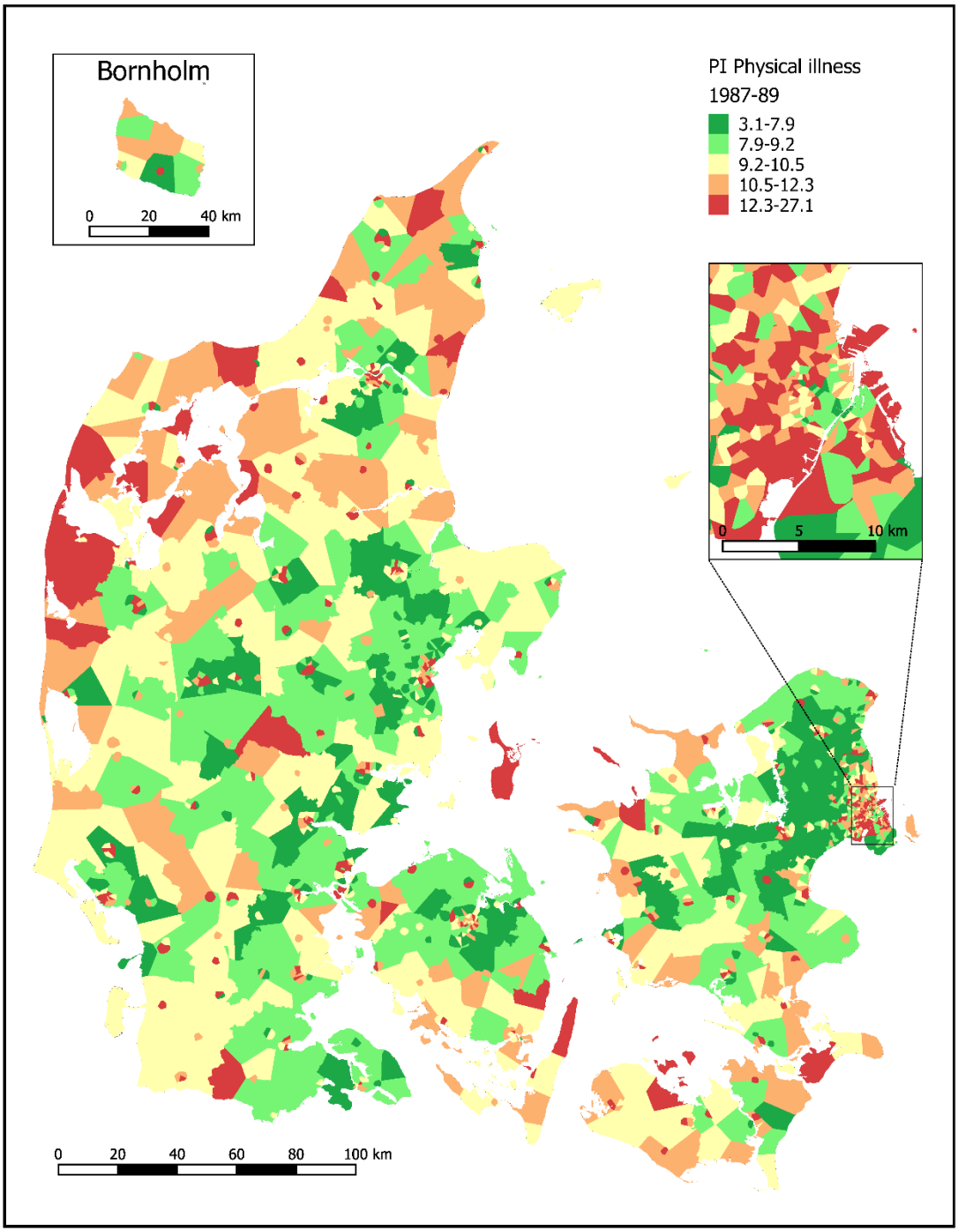
eFigure 2J: Social marginalization: Any criminality



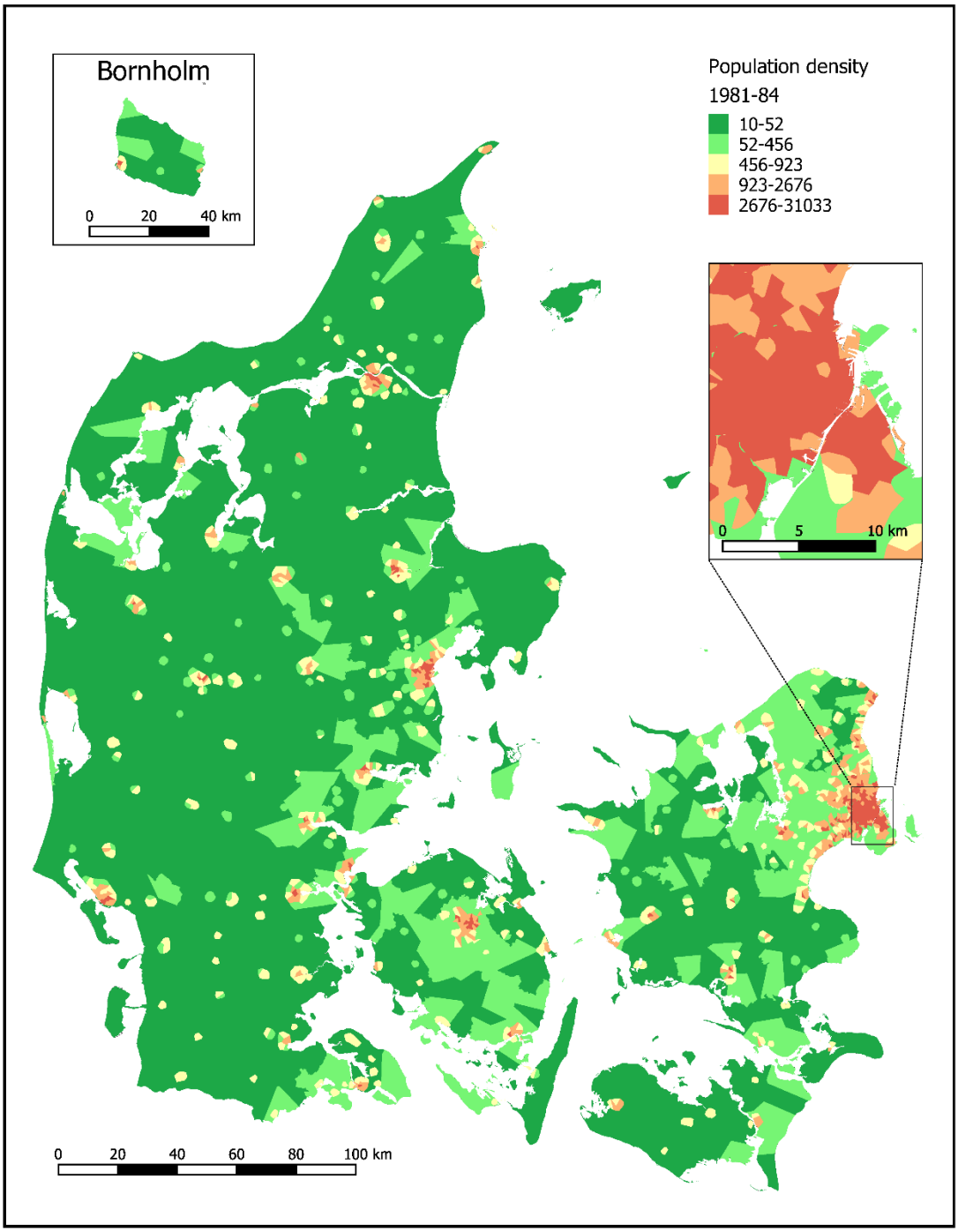
eFigure 2K: Social marginalization: Born abroad



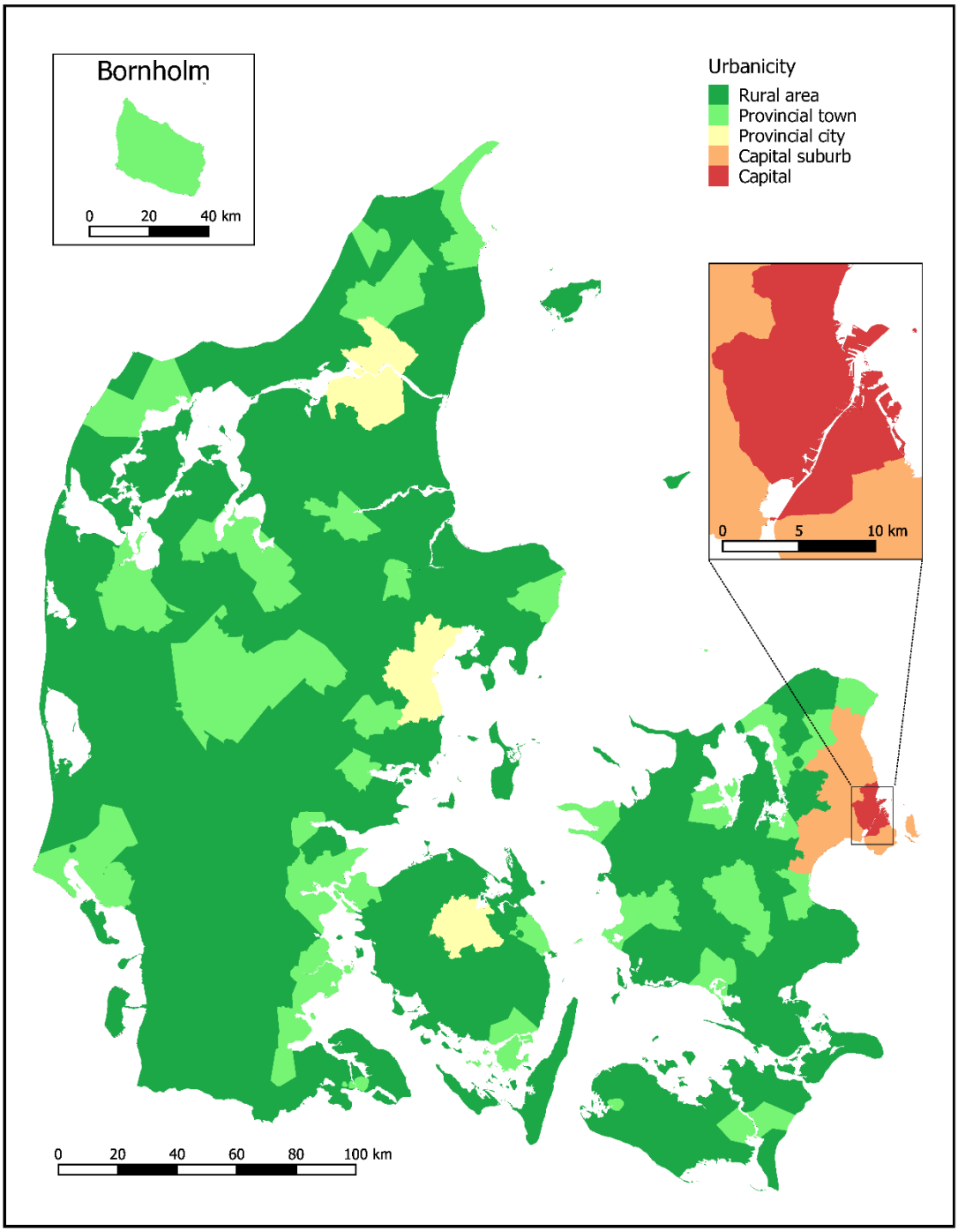
eFigure 2L: Physical illness



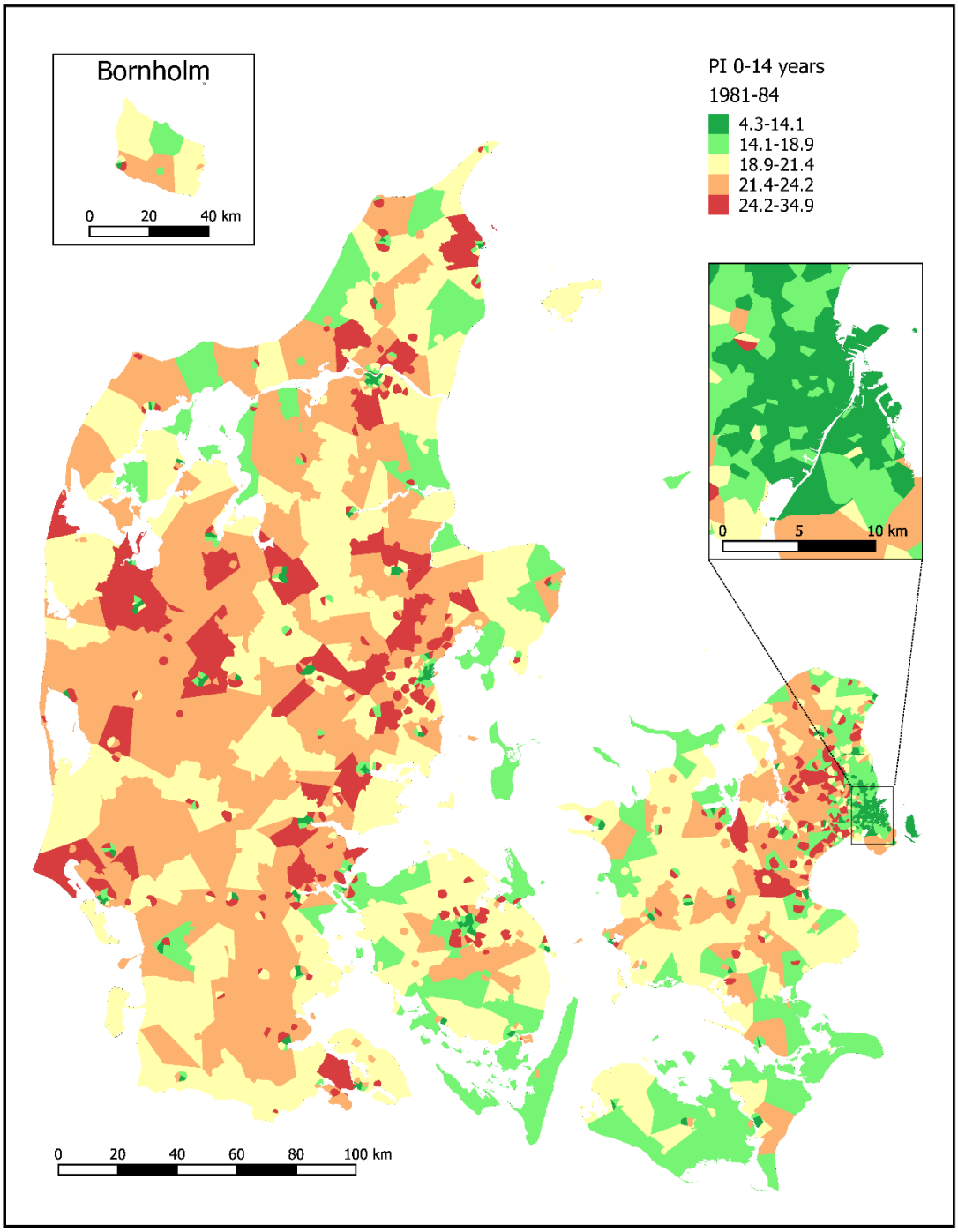
eFigure 2M: Population density



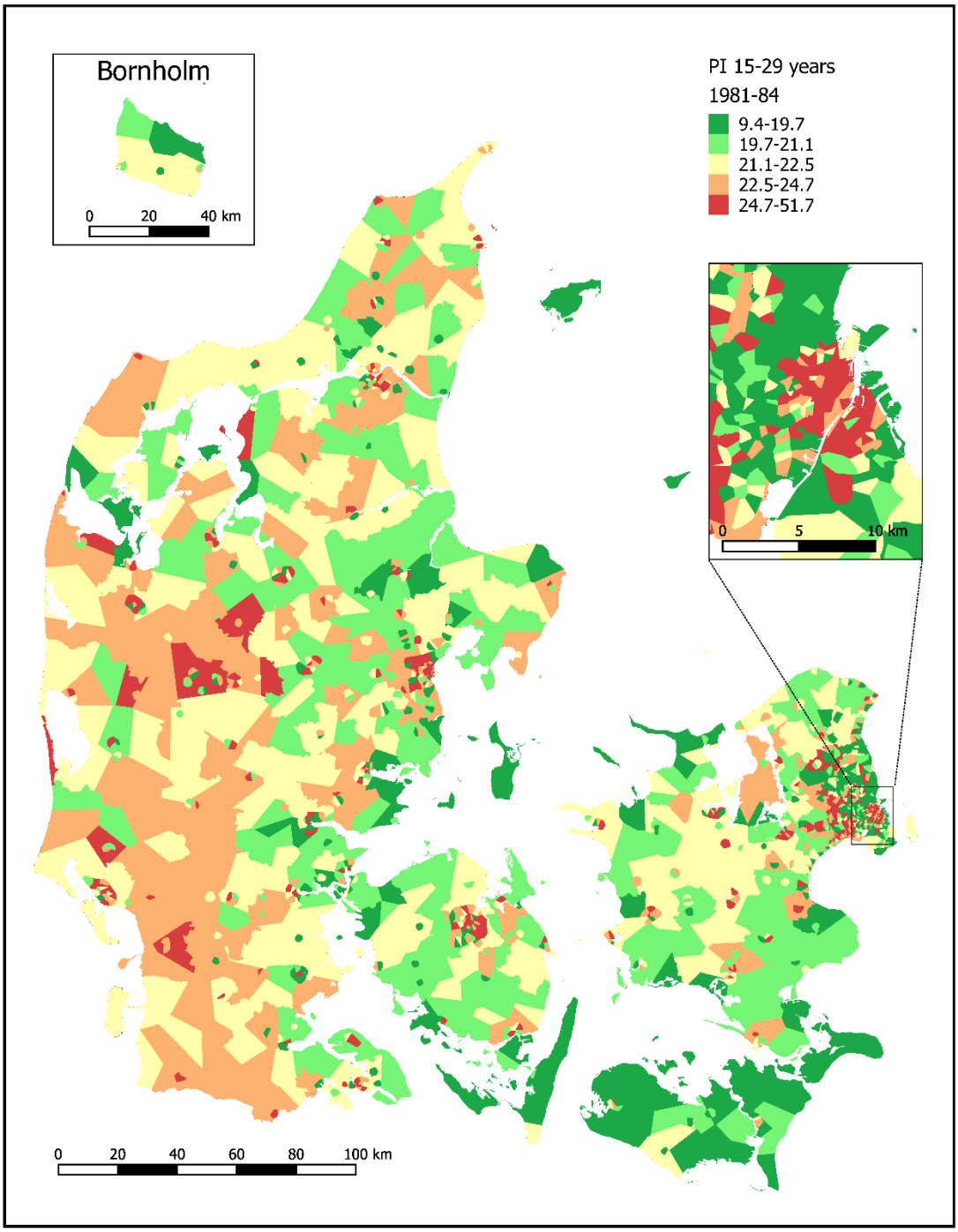
eFigure 2N: Urbanicity



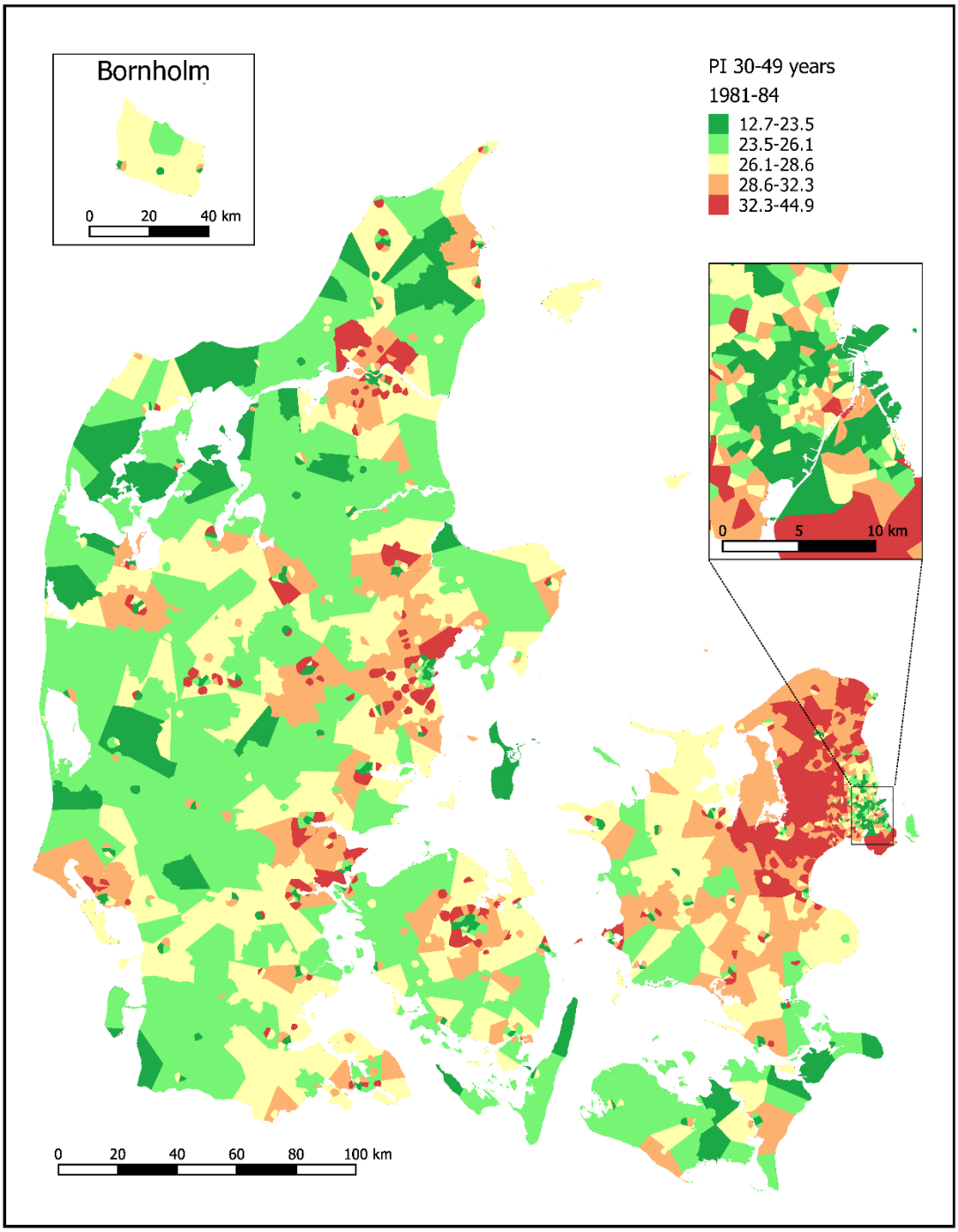
eFigure 2O: Age distribution (0-14 years)



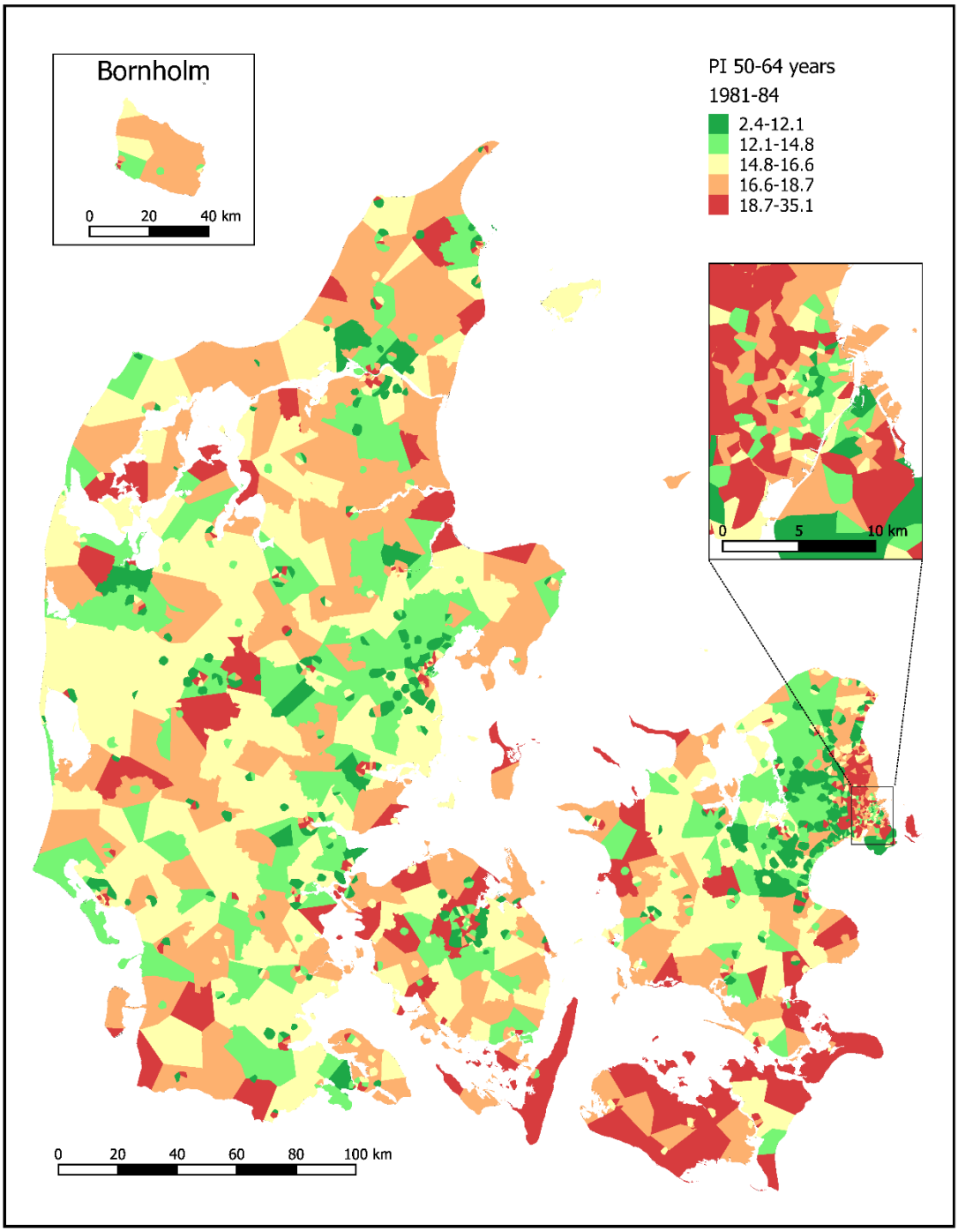
eFigure 2P: Age distribution (15-29 years)



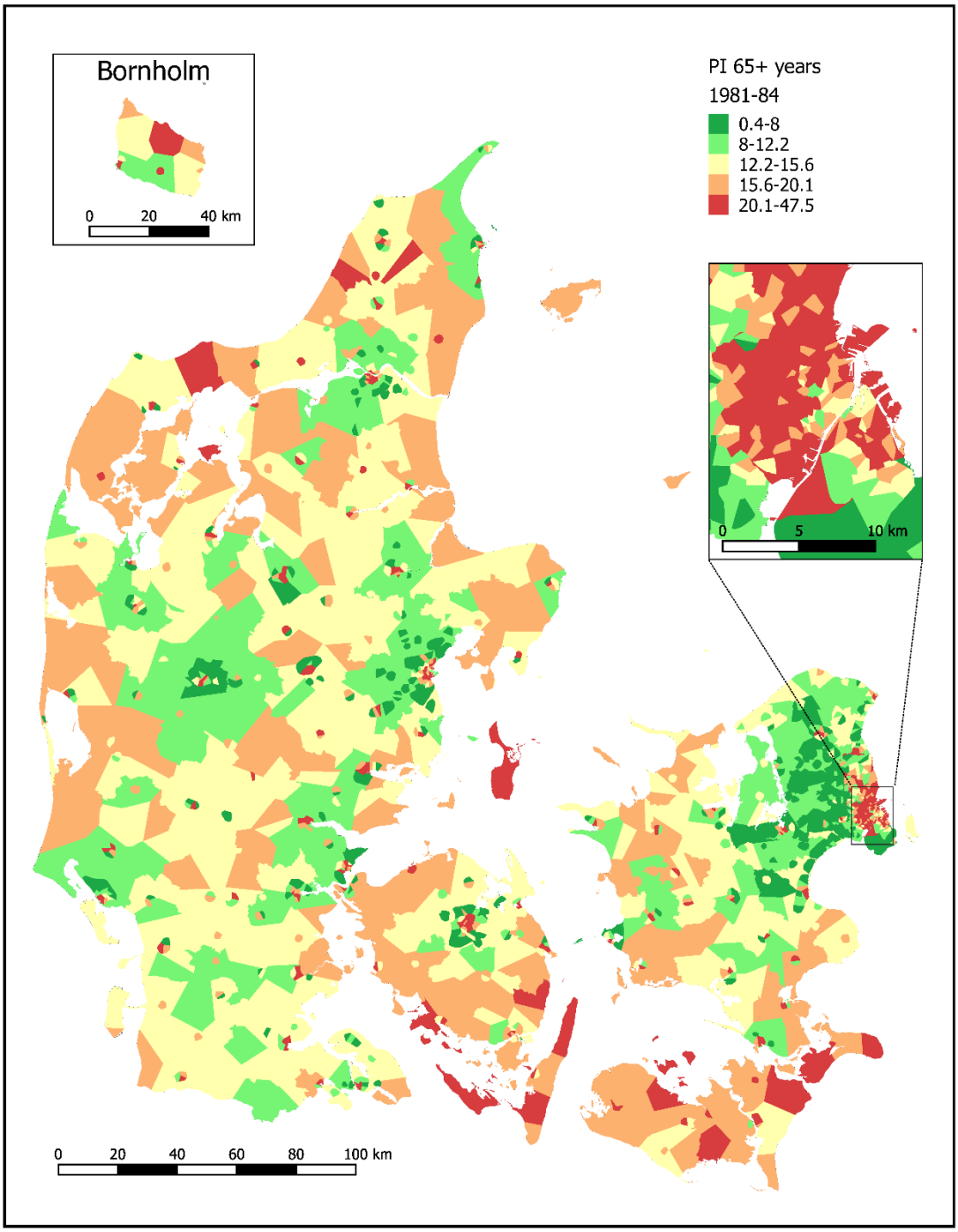
eFigure 2Q: Age distribution (30-49 years)



eFigure 2R: Age distribution (50-64 years)



eFigure 2S: Age distribution (65+ years)



Online references

1. Ministeriet for Videnskab -Teknologi og Udvikling [Ministry of science - technology and development]. Digitalt atlas over Danmarks historisk-administrative geografi [Digital map of historical administrative boarders in Denmark]. Ministeriet for Videnskab, Teknologi og Udvikling [Ministry of science, technology and development]. www.digdag.dk. Published 2021. Accessed2021.
2. The Danish Civil Registration Office. Kommunesammenlægninger / opdeling [Danish municipality reforms]. Published 2007. Accessed2021.
3. Flowerdew R, Manley DJ, Sabel CE. Neighbourhood effects on health: does it matter where you draw the boundaries? *Social science & medicine (1982).* 2008;66(6):1241-1255.
4. Sabel CE, Kihal W, Bard D, Weber C. Creation of synthetic homogeneous neighbourhoods using zone design algorithms to explore relationships between asthma and deprivation in Strasbourg, France. *Social science & medicine (1982).* 2013;91:110-121.
5. Martin D, Nolan A, Tranmer M. The Application of Zone-Design Methodology in the 2001 UK Census. *Environment and Planning A: Economy and Space.* 2001;33(11):1949-1962.
6. Statistics Denmark. Guidelines for transferring aggregated results from Statistics Denmark. Statistics Denmark. https://www.dst.dk/ext/3477468153/0/forskning/Guidelines-for-transferring-aggregated-results-from-Statistics-Denmark--pdf. Published 2021. Accessed2021.
7. Danish Health Data Authority. Guidelines for transferring aggregated results from the Danish Health Data Authority. Danish Health Data Authotrity. https://sundhedsdatastyrelsen.dk/-/media/sds/filer/forskerservice/retningslinjer-for-hjemsendelse-af-analyseresultater_2018_07_03.pdf?la=da. Published 2021. Accessed2021.
8. Pedersen CB, Gotzsche H, Moller JO, Mortensen PB. The Danish Civil Registration System. A cohort of eight million persons. *Danish medical bulletin.* 2006;53(4):441-449.
9. Statistics Denmark. *IDA - en integreret database for arbejdsmarkedsforskning (The Integrated Database for Longitudinal Labour Market Research).* Copenhagen: Statistics Denmark Press; 1991.
10. Mok PL, Webb RT, Appleby L, Pedersen CB. Full spectrum of mental disorders linked with childhood residential mobility. *J Psychiatr Res.* 2016;78:57-64.
11. Mok PL, Antonsen S, Pedersen CB, Appleby L, Shaw J, Webb RT. National cohort study of absolute risk and age-specific incidence of multiple adverse outcomes between adolescence and early middle age. *BMC Public Health.* 2015;15(1):920.
12. Cantor-Graae E, Pedersen CB. Full spectrum of psychiatric disorders related to foreign migration: a Danish population-based cohort study. *JAMA psychiatry.* 2013;70(4):427-435.
13. Thygesen SK, Christiansen CF, Christensen S, Lash TL, Sorensen HT. The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. *BMC medical research methodology.* 2011;11:83.
14. Rothman J, Greenland S, Lash TL. *Modern epidemiology, 3.rd edition.* Philadelphia: Lippinicott-Raven Publishers, 1998.; 1998.