Small Numbers, Big Problems:

Constructing Life Tables for Danish Municipalities

Ilya Kashnitsky // Statistics Denmark // ikx@dst.dk

2025-09-24, Voorburg UN City Group Meeting, Copenhagen

Context / Relevance

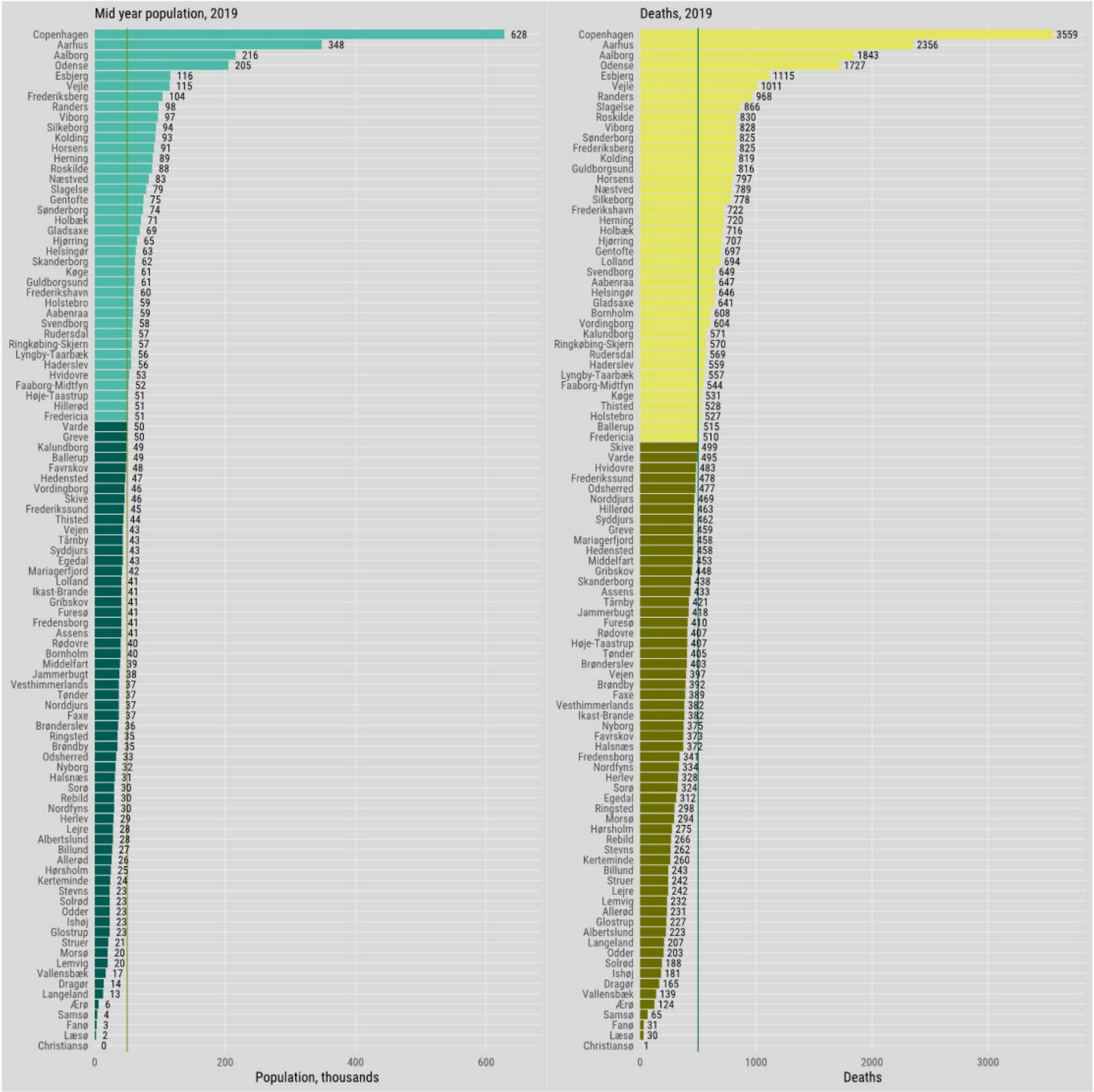
– Denmark has 98 municipalities, with a median size of less than 50,000 inhabitants

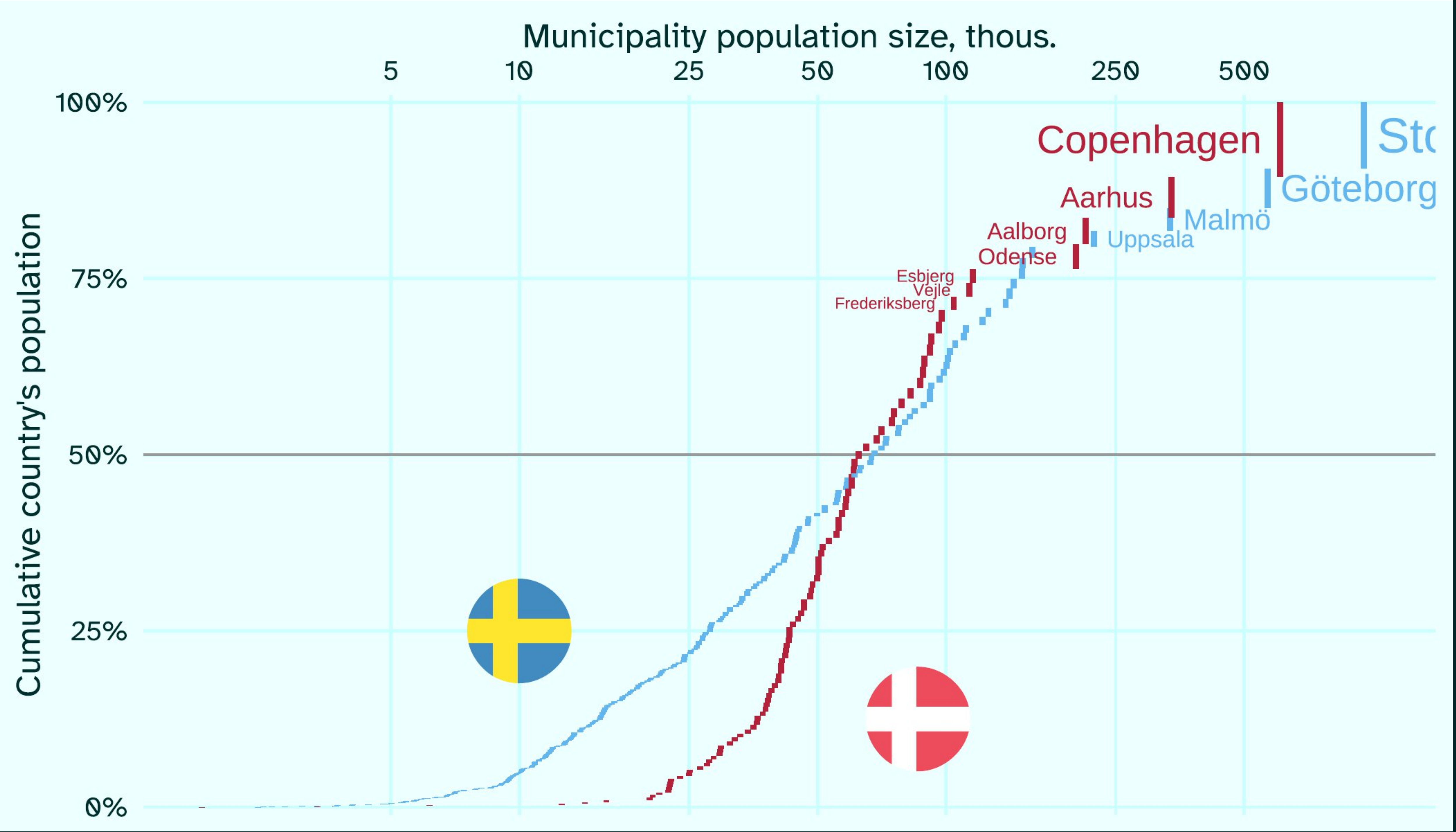
– Urgent need for coherent demographic estimates and forecasts for planning and policy purposes – but they are difficult to obtain due to the small population counts

Danish municipalities are small

About half of them are less than **50K** people

and see less than **500** deaths per year





Municipality population size, thous.
5
10
25
50
100
250
500
100%
75%
50%
25%
0%
Cumulative country's population
Copenhagen
Sto
Göteborg
Aarhus
Malmö
Aalborg
Uppsala
Odense
Esbjerg
Vejle
Frederiksberg

The main message of Figs. 9 and 10 lies in the very high overlap between the posterior distributions of many of the microregions. These are not especially small areas: Livramento do Brumado was the least-populous microregion in Bahia, but it had a total population of 97,786 in 2010, and the median total population of Bahian microregions was slightly below 290,000.

Despite these fairly large areas, however, uncertainty dominates most pairwise comparisons. It is clear that at this geographic level, researchers and policy makers should not rely on point estimates to distinguish high- and low-mortality areas—especially if differences in best-guess estimates of median e_0 are less than one year. That result applies even more strongly to smaller areas, such as municipalities.

Schmertmann, C. P., & Gonzaga, M. R. (2018). Bayesian Estimation of Age-Specific Mortality and Life Expectancy for Small Areas With Defective Vital Records. ***Demography***, 55(4), 1363–1388. https://doi.org/10.1007/s13524-018-0695-2

TOPALS to construct life tables

De Beer, J. (2011). A new relational method for smoothing and projecting age-specific fertility rates: TOPALS. Demographic Research, 24, 409–454. doi.org/10.4054/DemRes.2011.24.18

Table 2: Goodness of fit (measured by root mean square error, RMSE) of the logarithms of age-specific probabilities of death in 26 European countries, 2006

	Men			Women		
	TOPALS	Heligman-Pollard	Brass	TOPALS	Heligman-Pollard	Brass
	RMSE (x 10^{-3})					
Austria	191	184	197	223	255	232
Belarus	144	255	284	214	192	249
Belgium	157	154	164	208	248	218
Bulgaria	164	145	206	255	184	353
Czech Republic	145	260	167	231	244	229
Denmark	201	190	207	232	234	271
Estonia	319	355	347	427	418	417
Finland	334	283	343	307	319	296
France	67	136	133	95	212	135
Germany	91	95	117	100	176	132
Hungary	143	433	324	232	327	302
Ireland	275	247	297	341	306	330
Italy	105	108	118	106	172	112
Latvia	364	343	408	411	369	419
Lithuania	188	201	308	222	222	265
Netherlands	148	110	222	173	192	184
Norway	268	268	291	360	362	350
Poland	67	171	162	116	170	159
Portugal	128	170	200	210	214	247
Russia	72	126	255	93	88	246
Slovakia	232	236	246	262	235	322
Spain	129	131	130	106	214	142
Sweden	188	292	221	228	298	229
Switzerland	196	217	214	274	299	282
Ukraine	126	102	251	122	92	253
United Kingdom	76	91	94	90	116	117
Average	174	204	227	217	237	250

TOPALS to construct life tables

Denecke, E., Grigoriev, P., & Rau, R. (**2023**). Evaluation of small-area estimation methods for mortality schedules (arXiv:2302.01693). **arXiv**. http://arxiv.org/abs/2302.01693

De Beer, J. (**2011**). A new relational method for smoothing and projecting age-specific fertility rates: TOPALS. *Demographic Research*, 24, 409–454. doi.org/10.4054/DemRes.2011.24.18

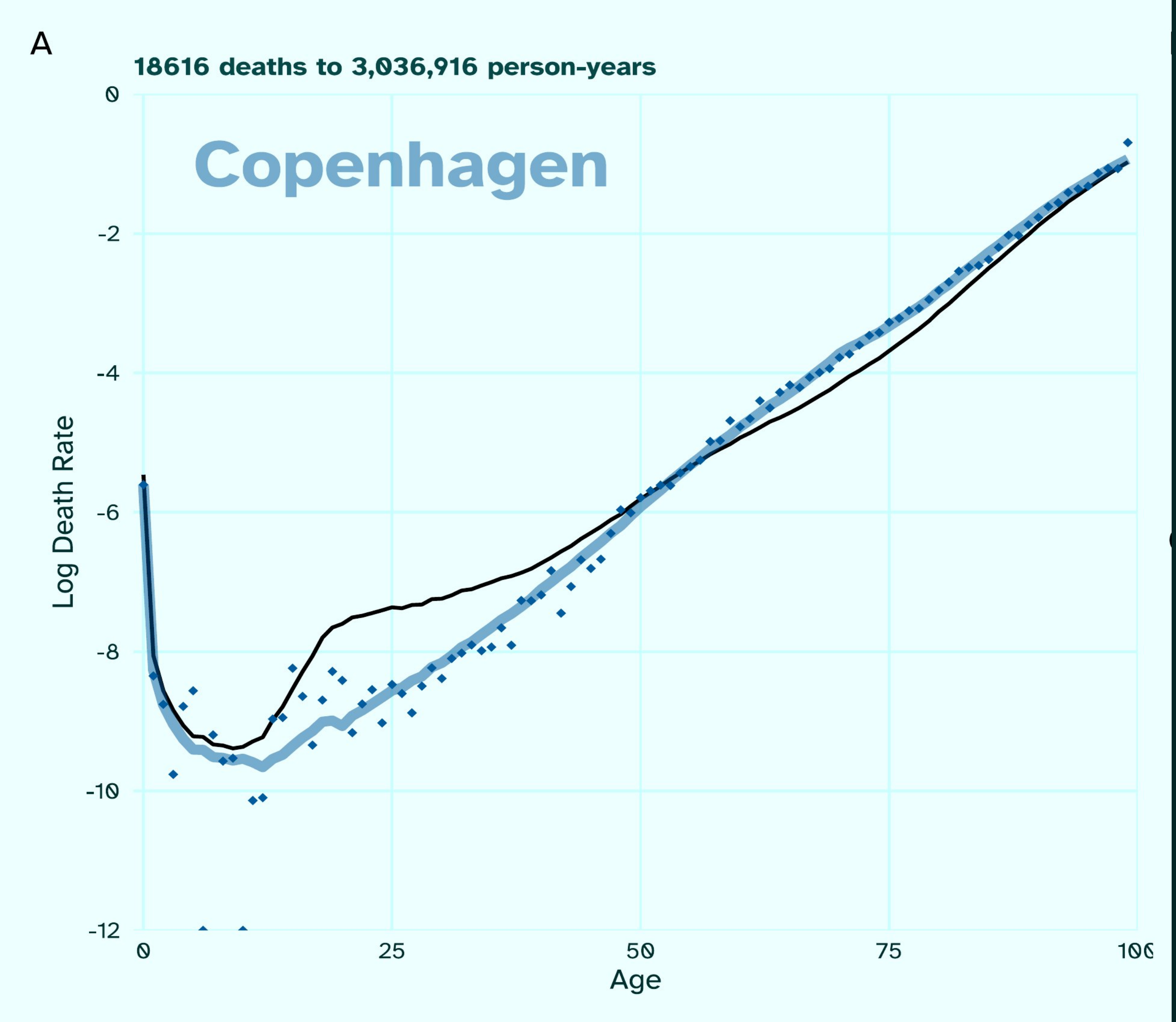
We find that recent demographic methods differ in their data requirements. As such, frequentist TOPALS regression models and D-splines estimate mortality rates in different regions/subpopulations and over time independently of each other while one Bayesian versions of TOPALS pools across space and the SVD-model across both space and time. This is also reflected in the data requirements and ease of use. Once code is available, (frequentist) TOPALS and D-splines are simple and quick to use. In the simulation study, we reveal that averaging performance measures over several regions or ages may mask underlying variability. As such, we show that all methods are somewhat sensitive to exposure size and the incorporated demographic knowledge across different regions but that there is large variation. While bias tends to be larger and coverage of uncertainty intervals lower for younger ages, there is no clear pattern across methods and regions. In the results presented, D1 and DLC generally exhibit lower variability than D2 and TOPALS which are more accurate on average. Thus, DLC and D1 tend to exhibit more bias which translates into the estimates of life expectancy. For larger exposures this bias for life expectancy tends to diminish. The coverage of uncertainty intervals varies over the ages but tends to be best around the age of 50. At some ages, especially the youngest, coverage may be zero. Good coverage (i.e. around 0.95), especially at younger ages tends to come with very wide uncertainty intervals. This may be considered a desired feature: There are many zero death counts at these ages such that we really do not know much.

Based on the presented results, we would advise against using the D2 estimator of D-Splines in (very) small populations as the resulting mortality schedules can be implausible. TOPALS and the SVD-model showed reasonable results for life expectancy while the DLC and D1 estimator tend to be more biased at (very) small populations. Using the raw data to estimate life expectancy seems to be a viable alternative with slight overestimation and huge variability in small populations (see Eayres and Williams (2004) for more on the overestimation of life expectancy). However, we want to draw attention to two further aspects. First, we want to caution against overinterpreting point estimates. In small populations there is considerable variability in the results and there is likely a limit to

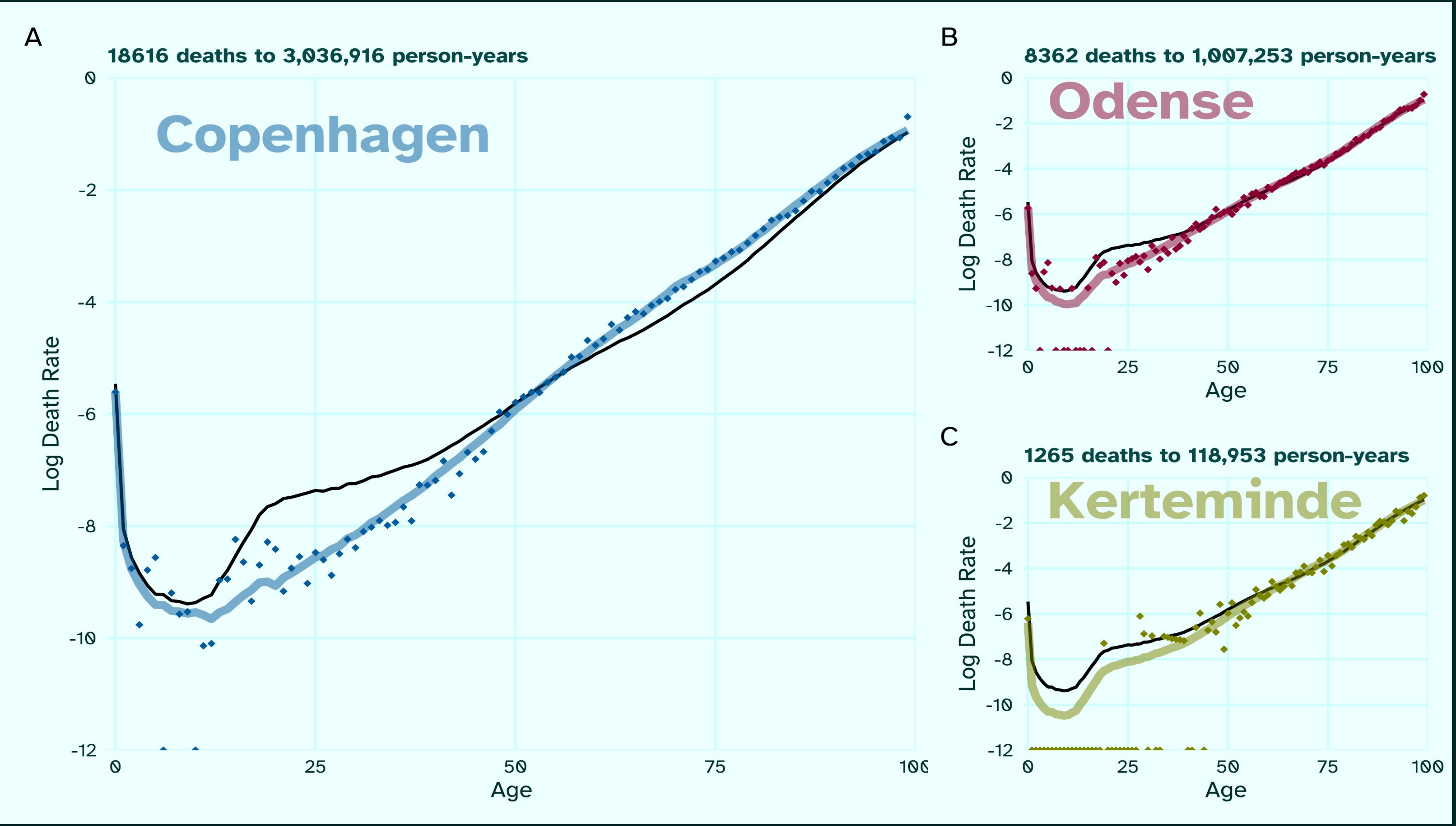
Table 2: Goodness of fit (measured by root mean square error, RMSE) of the logarithms of age-specific probabilities of death in 26 European countries, 2006

	Men			Women		
	TOPALS	Heligman-Pollard	Brass	TOPALS	Heligman-Pollard	Brass
	RMSE (x 10^{-3})					
Austria	191	184	197	223	255	232
Belarus	144	255	284	214	192	249
Belgium	157	154	164	208	248	218
Bulgaria	164	145	206	255	184	353
Czech Republic	145	260	167	231	244	229
Denmark	201	190	207	232	234	271
Estonia	319	355	347	427	418	417
Finland	334	283	343	307	319	296
France	67	136	133	95	212	135
Germany	91	95	117	100	176	132
Hungary	143	433	324	232	327	302
Ireland	275	247	297	341	306	330
Italy	105	108	118	106	172	112
Latvia	364	343	408	411	369	419
Lithuania	188	201	308	222	222	265
Netherlands	148	110	222	173	192	184
Norway	268	268	291	360	362	350
Poland	67	171	162	116	170	159
Portugal	128	170	200	210	214	247
Russia	72	126	255	93	88	246
Slovakia	232	236	246	262	235	322
Spain	129	131	130	106	214	142
Sweden	188	292	221	228	298	229
Switzerland	196	217	214	274	299	282
Ukraine	126	102	251	122	92	253
United Kingdom	76	91	94	90	116	117
Average	174	204	227	217	237	250

TOPALS – intuition for the method



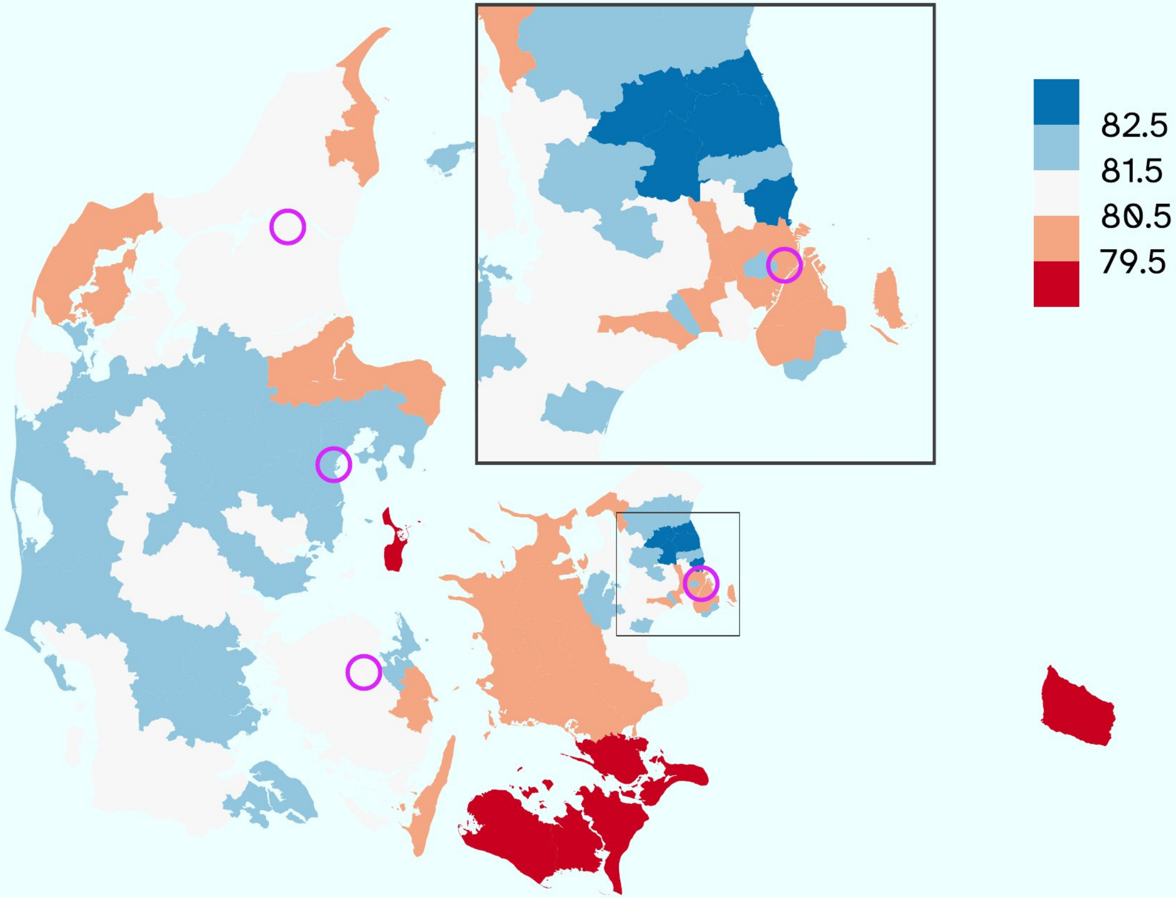
TOPALS – intuition for the method



A

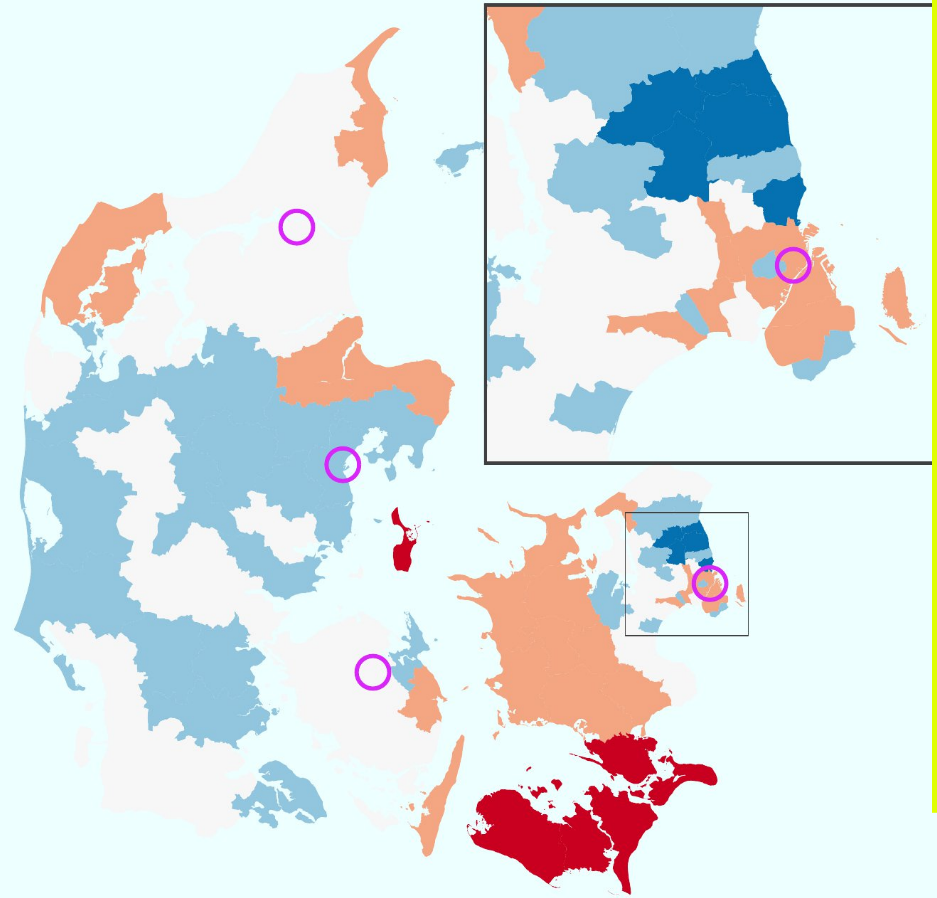
Life expectancy at birth

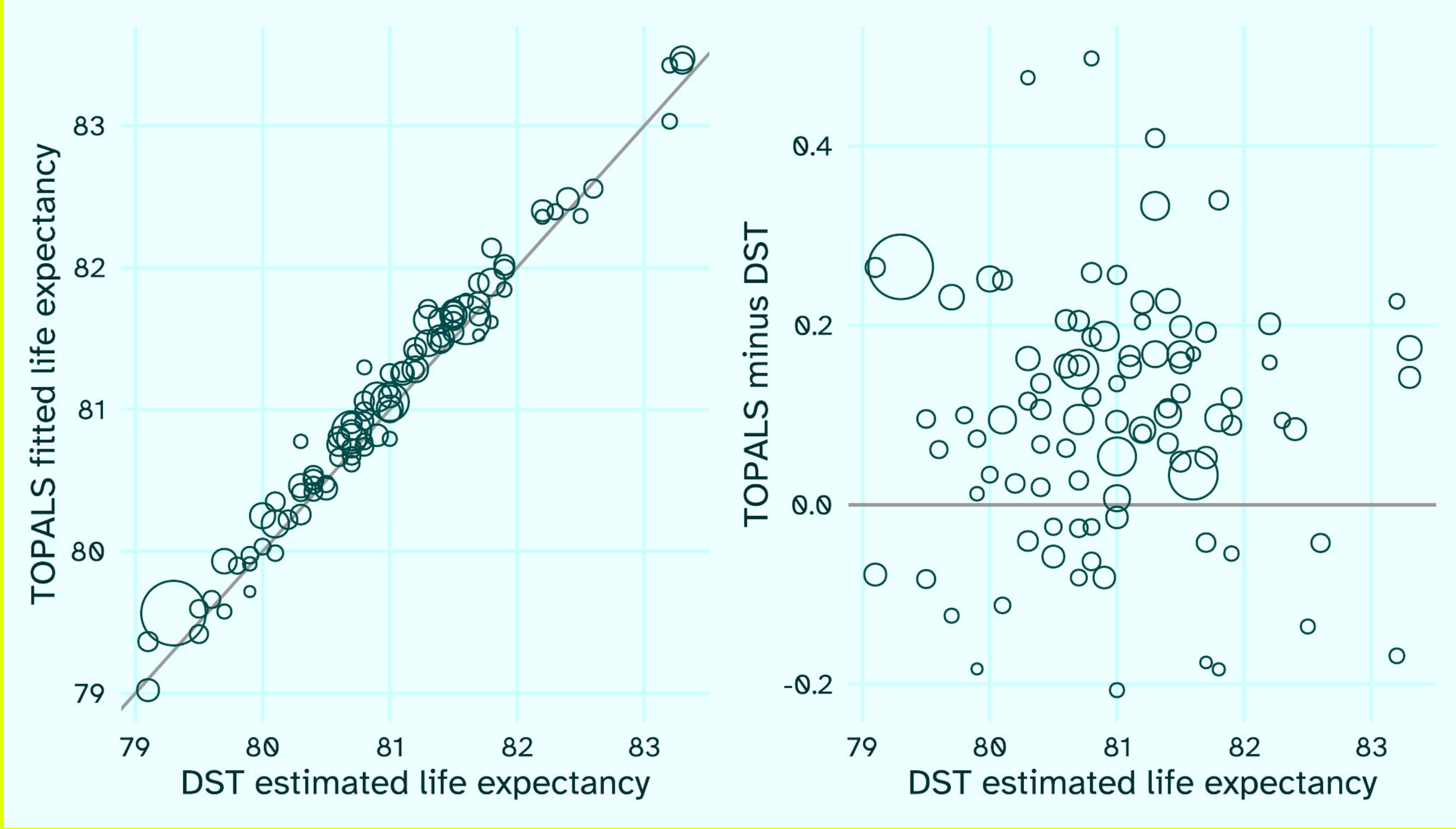
TOPALS model, pooled data 2015-19
National life expectancy at birth is 81 years

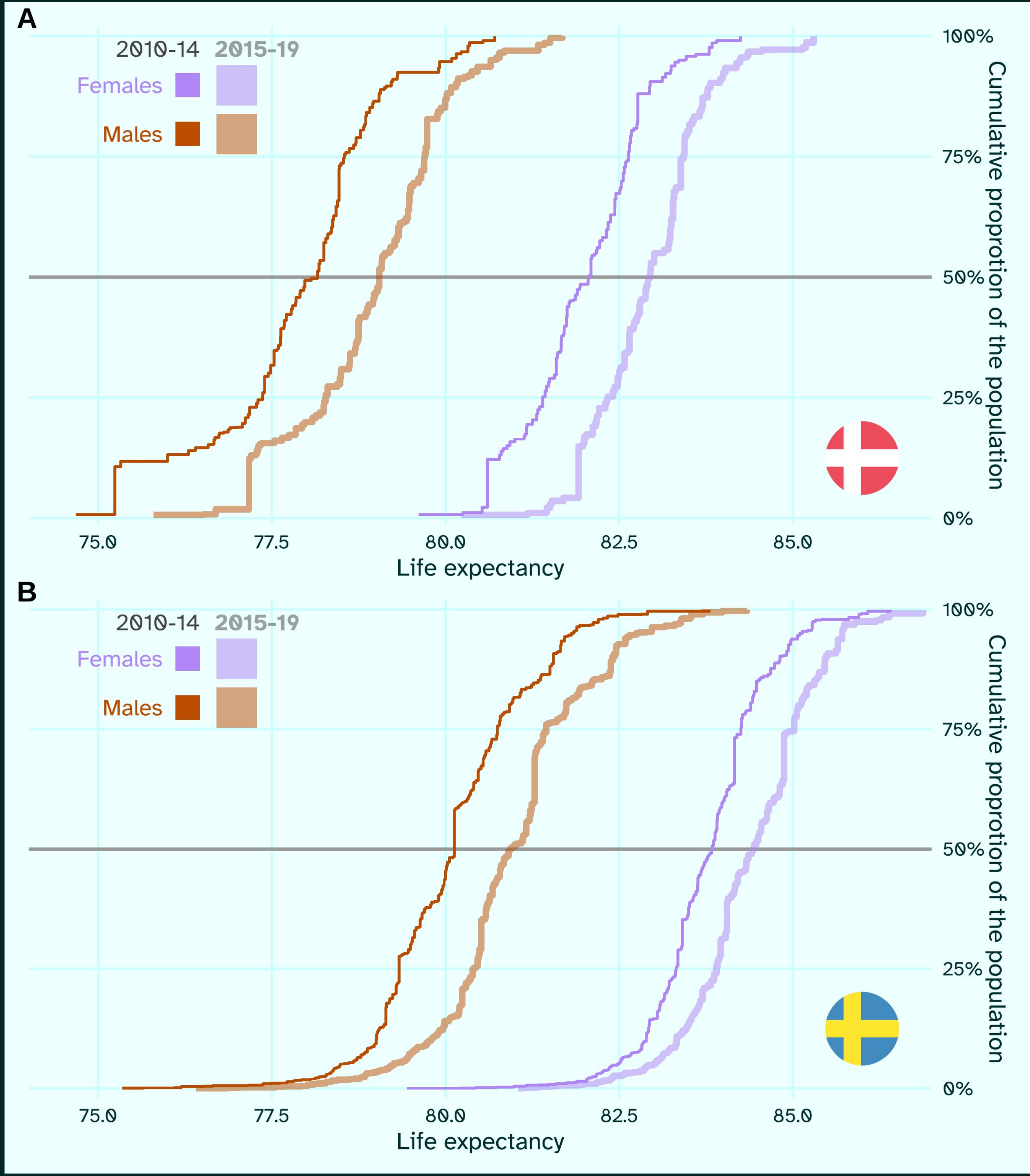


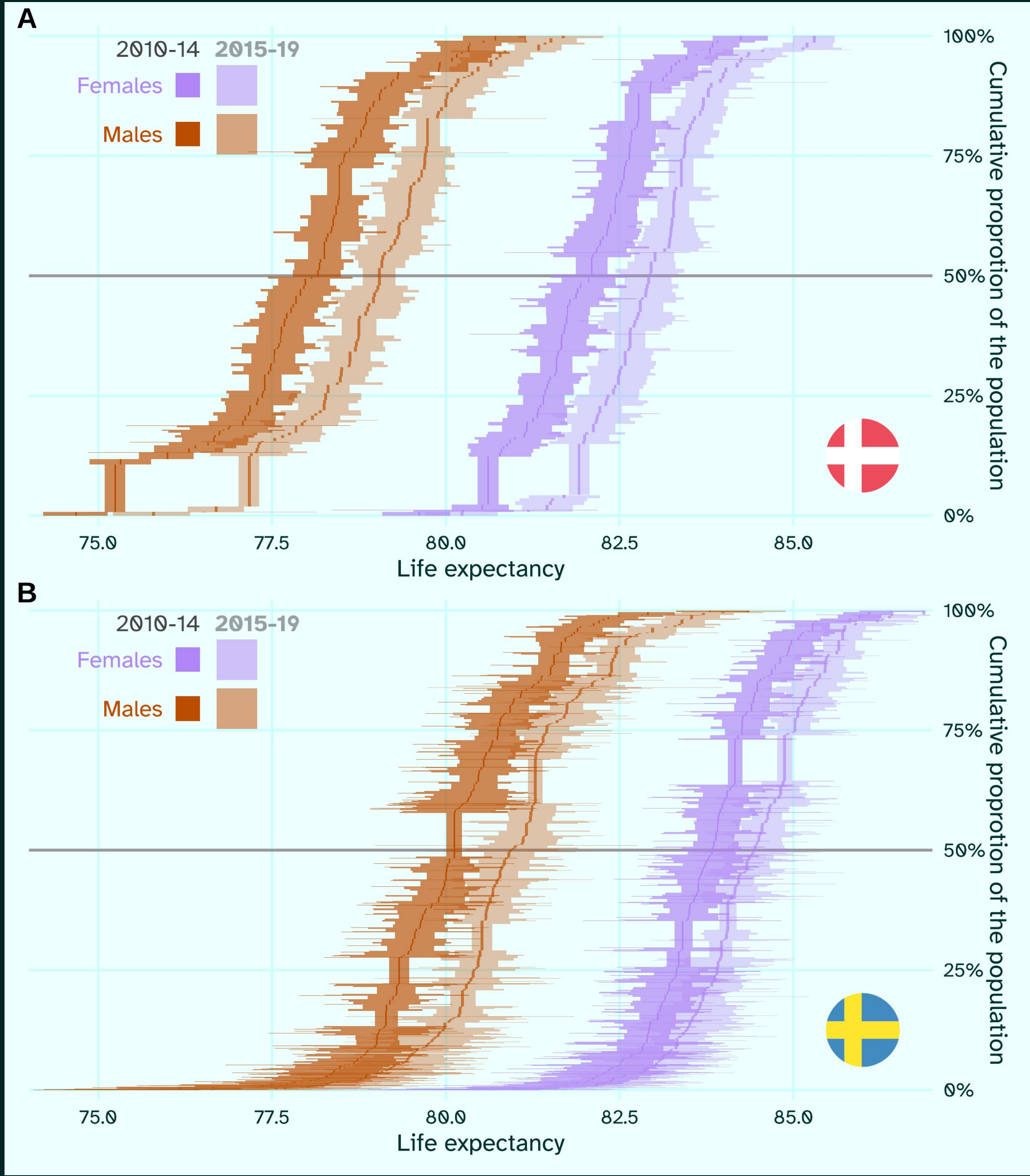
Life expectancy at birth

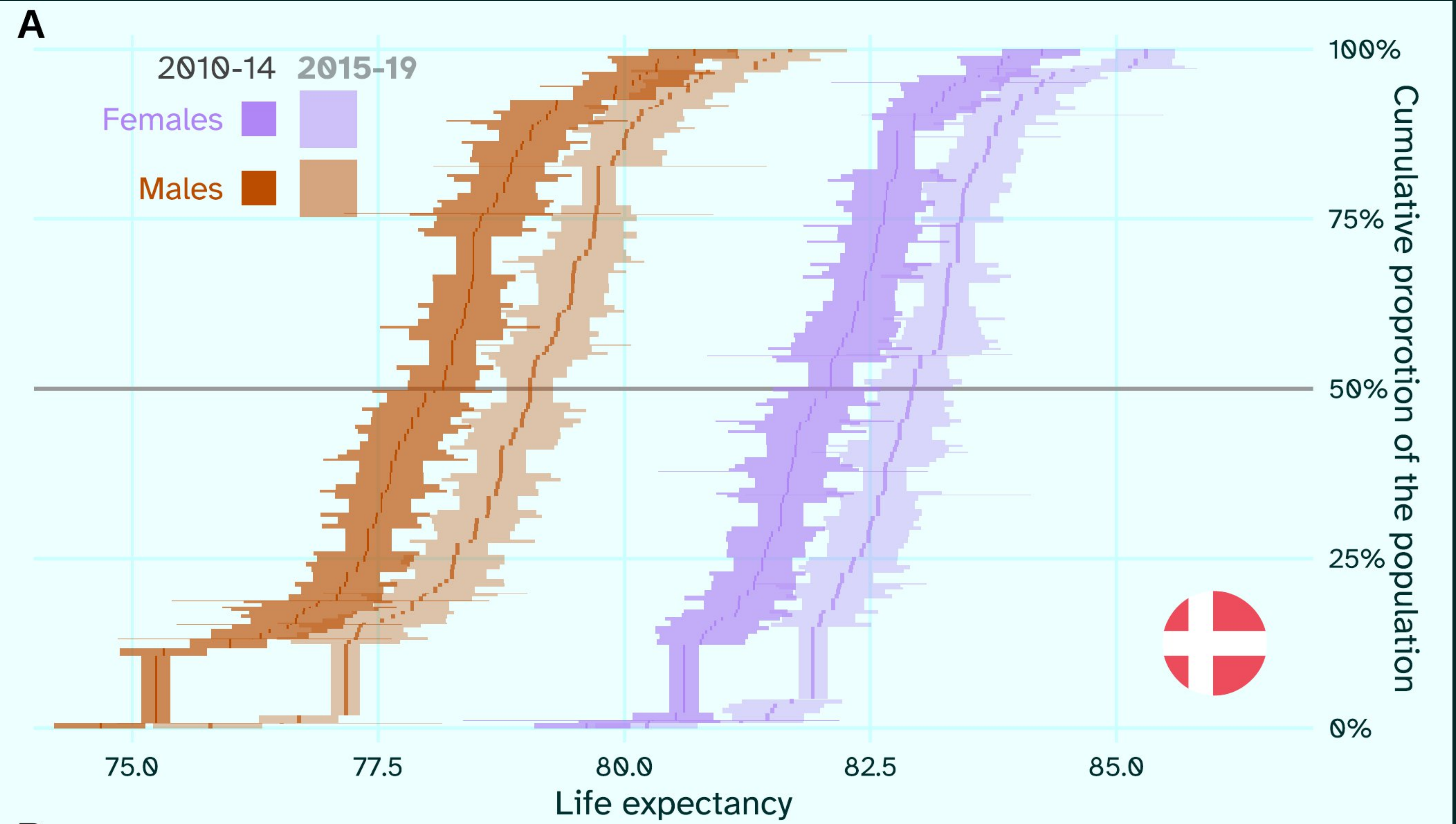
TOPALS model, pooled data 2015-19
National life expectancy at birth is 81 years






A
2010-14
2015-19
Females
Males
100%
75%
50%
25%
0%
Cumulative proportion of the population
75.0
77.5
80.0
82.5
85.0
Life expectancy
B
2010-14
2015-19
Females
Males
100%
75%
50%
25%
0%
Cumulative proportion of the population
75.0
77.5
80.0
82.5
85.0
Life expectancy


A
2010-14
2015-19
Females
Males
100%
75%
50%
25%
0%
Cumulative proportion of the population
75.0
77.5
80.0
82.5
85.0
Life expectancy
B
2010-14
2015-19
Females
Males
100%
75%
50%
25%
0%
Cumulative proportion of the population
75.0
77.5
80.0
82.5
85.0
Life expectancy


A
2010-14
2015-19
Females
Males
100%
75%
50%
25%
0%
Cumulative proportion of the population
75.0
77.5
80.0
82.5
85.0
Life expectancy

Population, Space and Place

WILEY

RESEARCH ARTICLE

Non-Survival to Pension Age in Denmark and Sweden: A Sub-National Investigation

Ilya Kashnitsky

Data Science Lab, Statistics Denmark, Copenhagen, Denmark

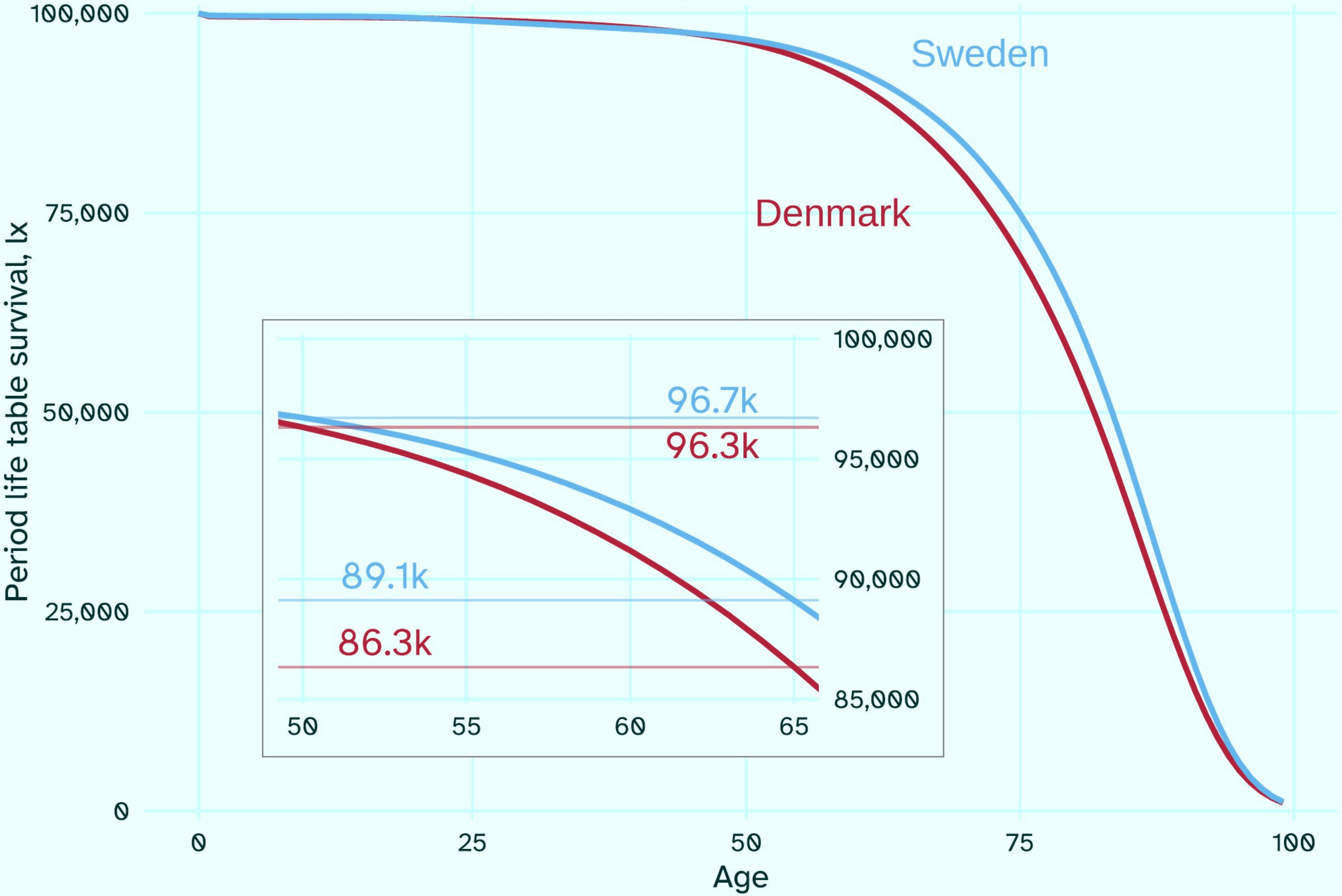
Correspondence: Ilya Kashnitsky (ikx@dst.dk; ilya.kashnitsky@gmail.com)

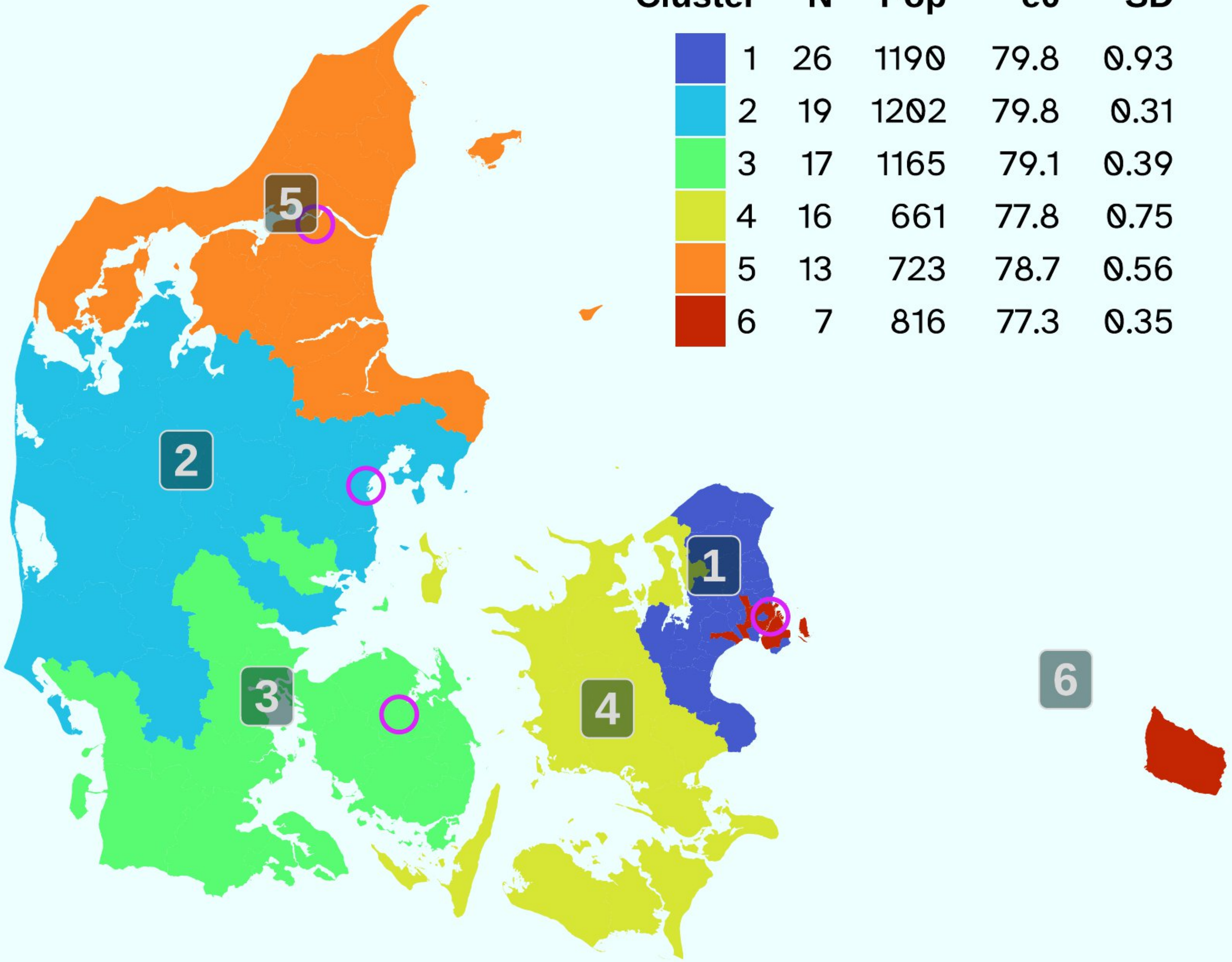
Received: 21 April 2023 | **Revised:** 28 April 2025 | **Accepted:** 29 April 2025

Funding: The research leading to this publication was part of a project that has received funding from the ROCKWOOL Foundation, through the research project 'Challenges to Implementation of Indexation of the Pension Age in Denmark'.

Keywords: adult mortality | life tables | municipalities | small-area estimation

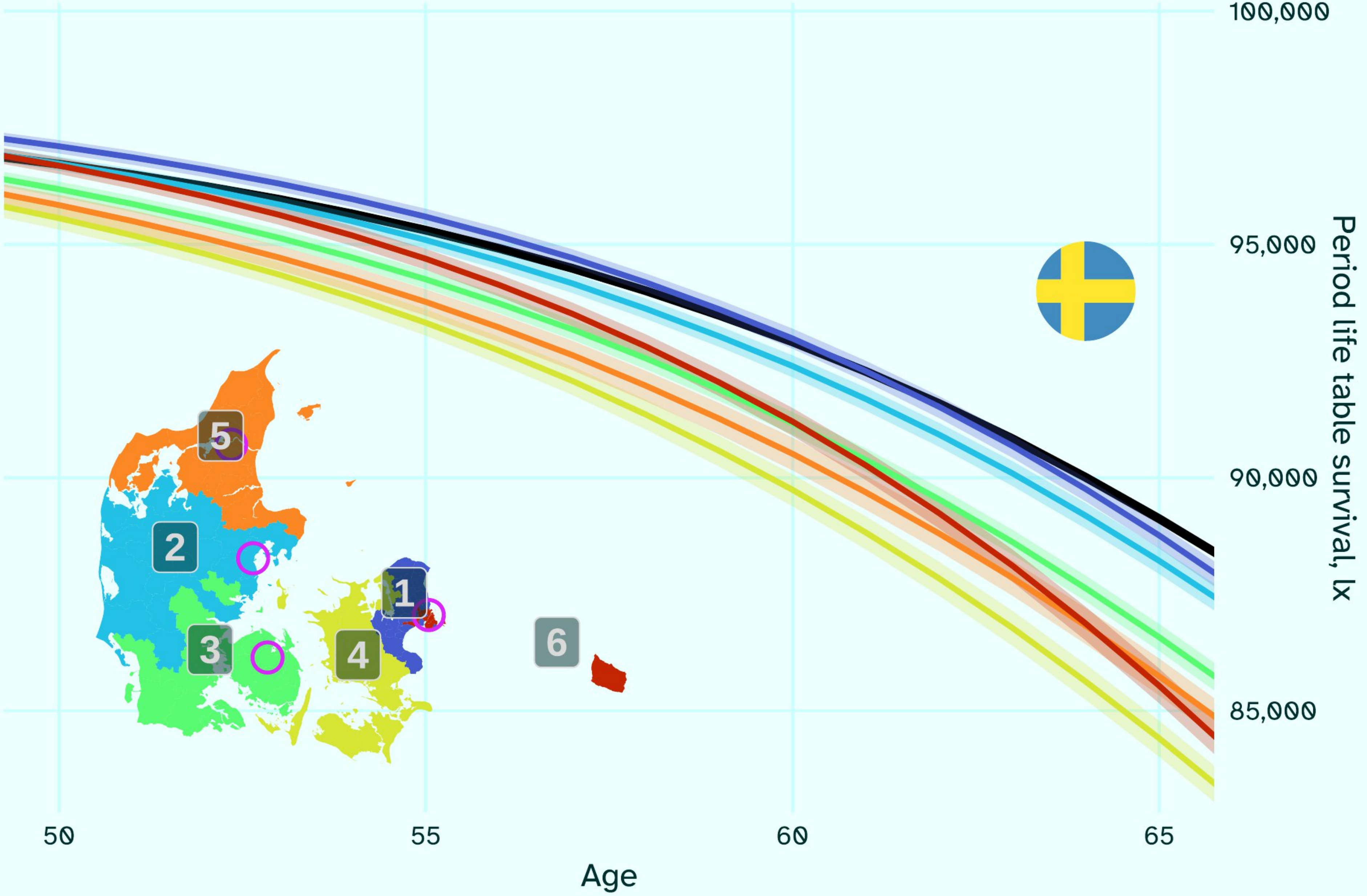
Survival of males, 2015-19





Cluster	N	Pop	e0	SD
1	26	1190	79.8	0.93
2	19	1202	79.8	0.31
3	17	1165	79.1	0.39
4	16	661	77.8	0.75
5	13	723	78.7	0.56
6	7	816	77.3	0.35

Survival of males, 2015-19




2010-14
2015-19
Sex gap in life expectancy
-3
-4
-5
-6
-7
76
78
80
82
76
78
80
82
Male life expectancy

Prospects

How to best estimate municipal demographic trends?

- exploratory statistics for DST
- comprehensive comparison with the current method

—> StatBank

How to best forecast **mortality** and **fertility**, **internal** and **external migration** across the municipalities?

- coherent population projection for the municipalities
- possibilities for extensive analysis at municipal level

Thank you!

Ilya Kashnitsky

ikx@dst.dk

@ikashnitsky.phd

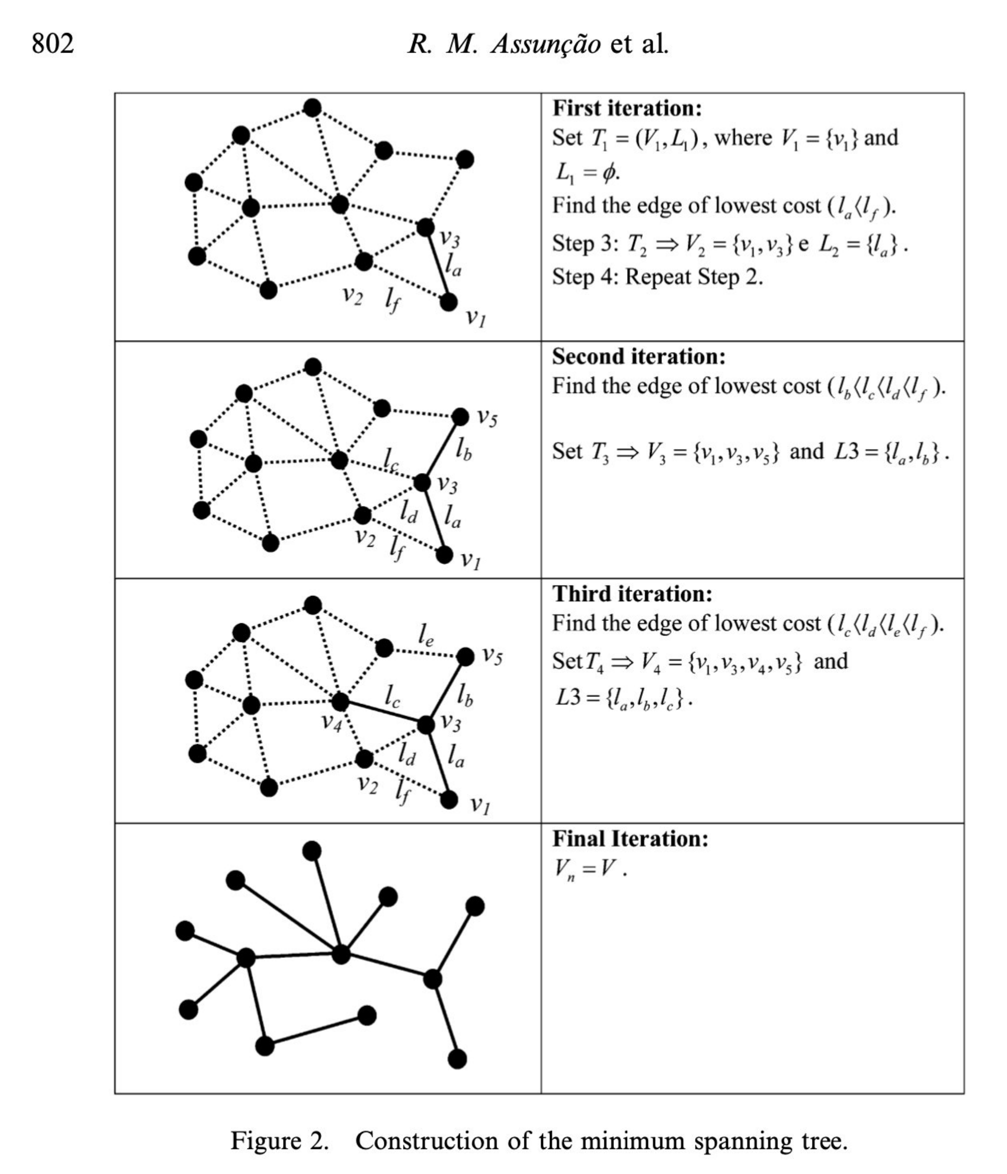


Figure 2. Construction of the minimum spanning tree.

SKATER spatial clustering

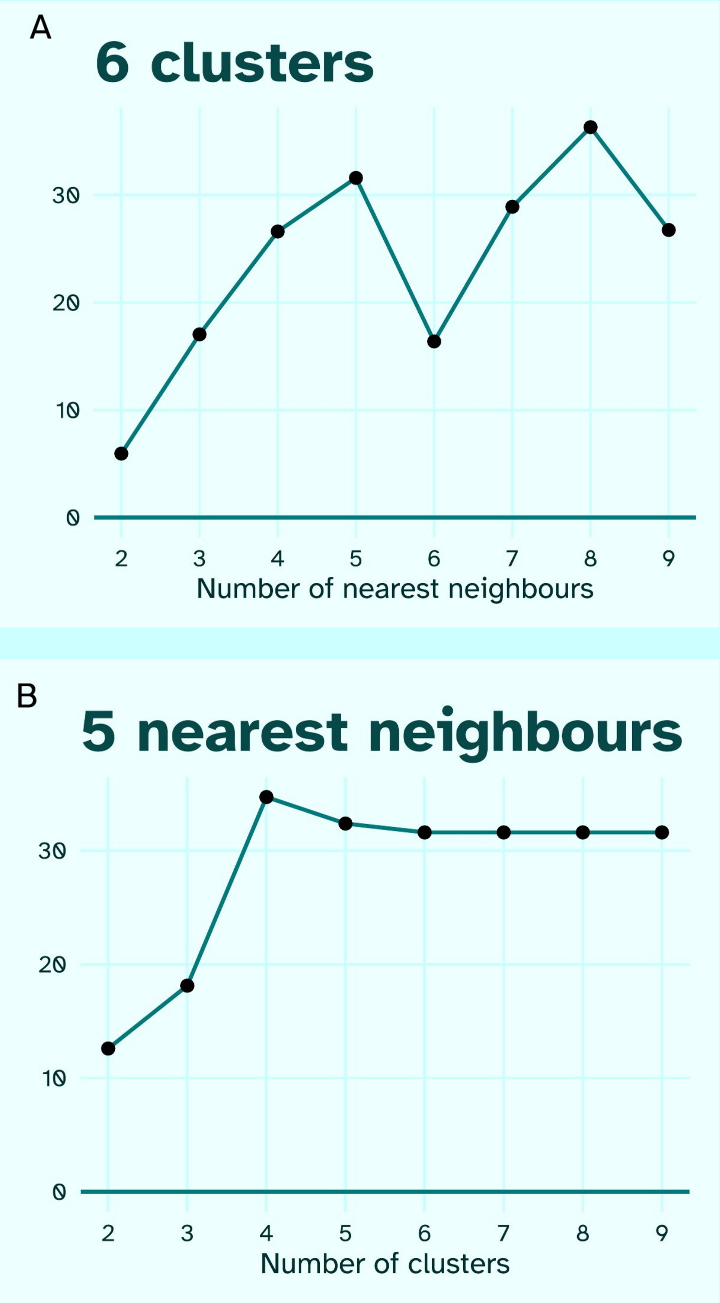
AssunÇão, R. M., Neves, M. C., Câmara, G., & Da Costa Freitas, C. (2006). Efficient regionalization techniques for socio-economic geographical units usaing minimum spanning trees. International Journal of Geographical Information Science, 20(7), 797–811. https://doi.org/10.1080/13658810600665111

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. Communications in Statistics, 3(1), 1–27. https://doi.org/10.1080/03610927408827101

If k, the number of clusters, is not known, we proceed as follows: first we take k=2, then k=3, and so on. At each stage we find "the best sum of squares split" of the dendrite, for which we calculate not only the (minimum) WGSS, but also the (maximum) BGSS and the variance ratio criterion

$$\boxed{\mathrm{VRC} = \frac{\mathrm{BGSS}}{k-1} \,/\, \frac{\mathrm{WGSS}}{n-k}\,.} \tag{2}$$

We suggest the application of (2) as an informal indicator for the "best number" of groups. It is evident that this criterion is analogous to the F-statistic in univariate analysis. In fact it has already been used by Edwards and Cavalli-Sforza (1965, p. 374) as an F-test in a multivariate cluster analysis.



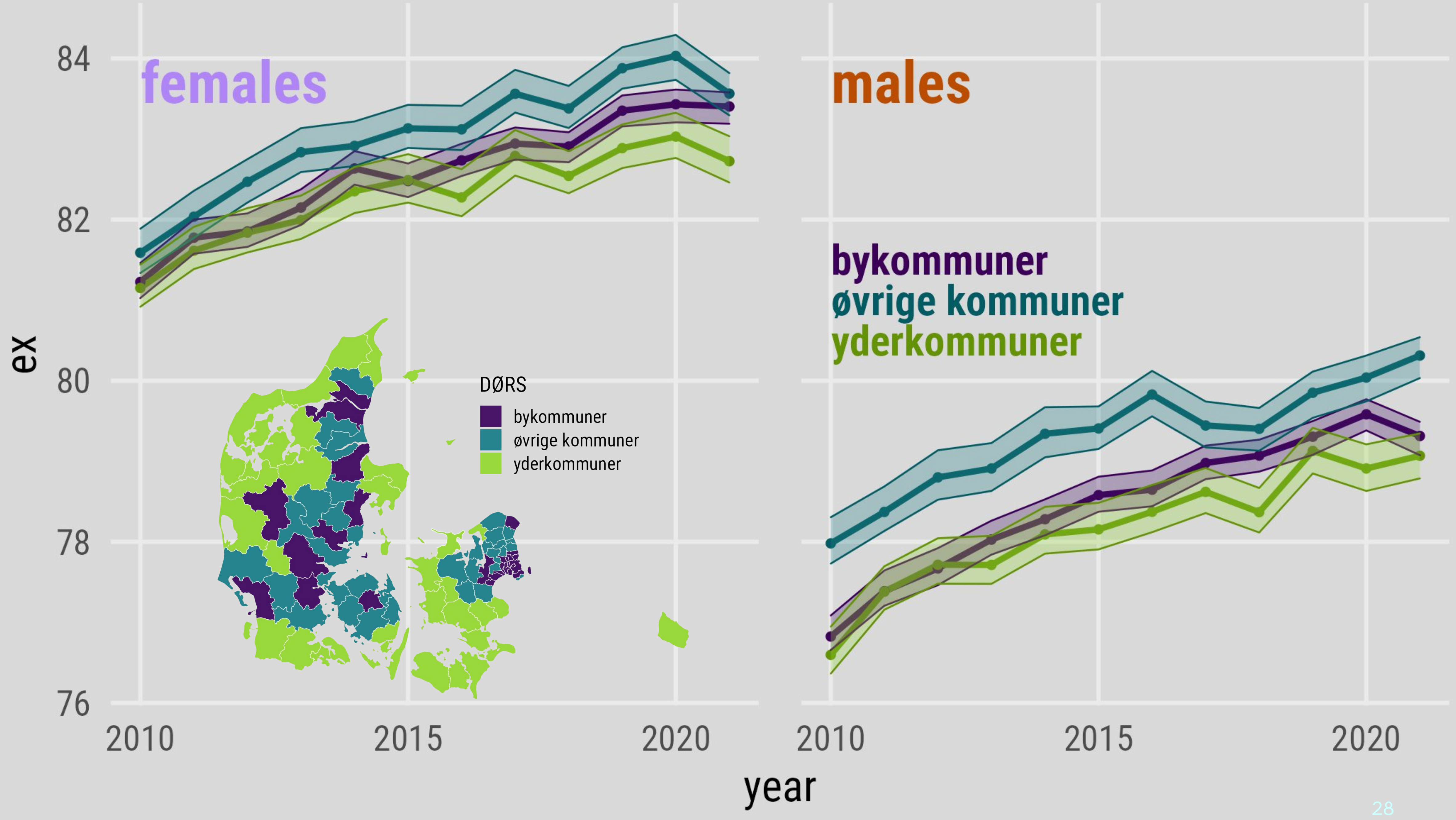
Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. Communications in Statistics, 3(1), 1–27. https://doi.org/10.1080/03610927408827101

CALIŃSKI AND HARABASZ

If k, the number of clusters, is not known, we proceed as follows: first we take k=2, then k=3, and so on. At each stage we find "the best sum of squares split" of the dendrite, for which we calculate not only the (minimum) WGSS, but also the (maximum) BGSS and the variance ratio criterion

$$\boxed{\mathrm{VRC} = \frac{\mathrm{BGSS}}{k-1} \Big/ \frac{\mathrm{WGSS}}{n-k}\ .} \qquad (2)$$

We suggest the application of (2) as an informal indicator for the "best number" of groups. It is evident that this criterion is analogous to the F-statistic in univariate analysis. In fact it has already been used by Edwards and Cavalli-Sforza (1965, p. 374) as an F-test in a multivariate cluster analysis.



females
males
bykommuner
øvrige kommuner
yderkommuner
DØRS
bykommuner
øvrige kommuner
yderkommuner
84
82
80
78
76
ex
2010
2015
2020
2010
2015
2020
year

No story

¯_(ツ)_/¯

No story

¯_(ツ)_/¯

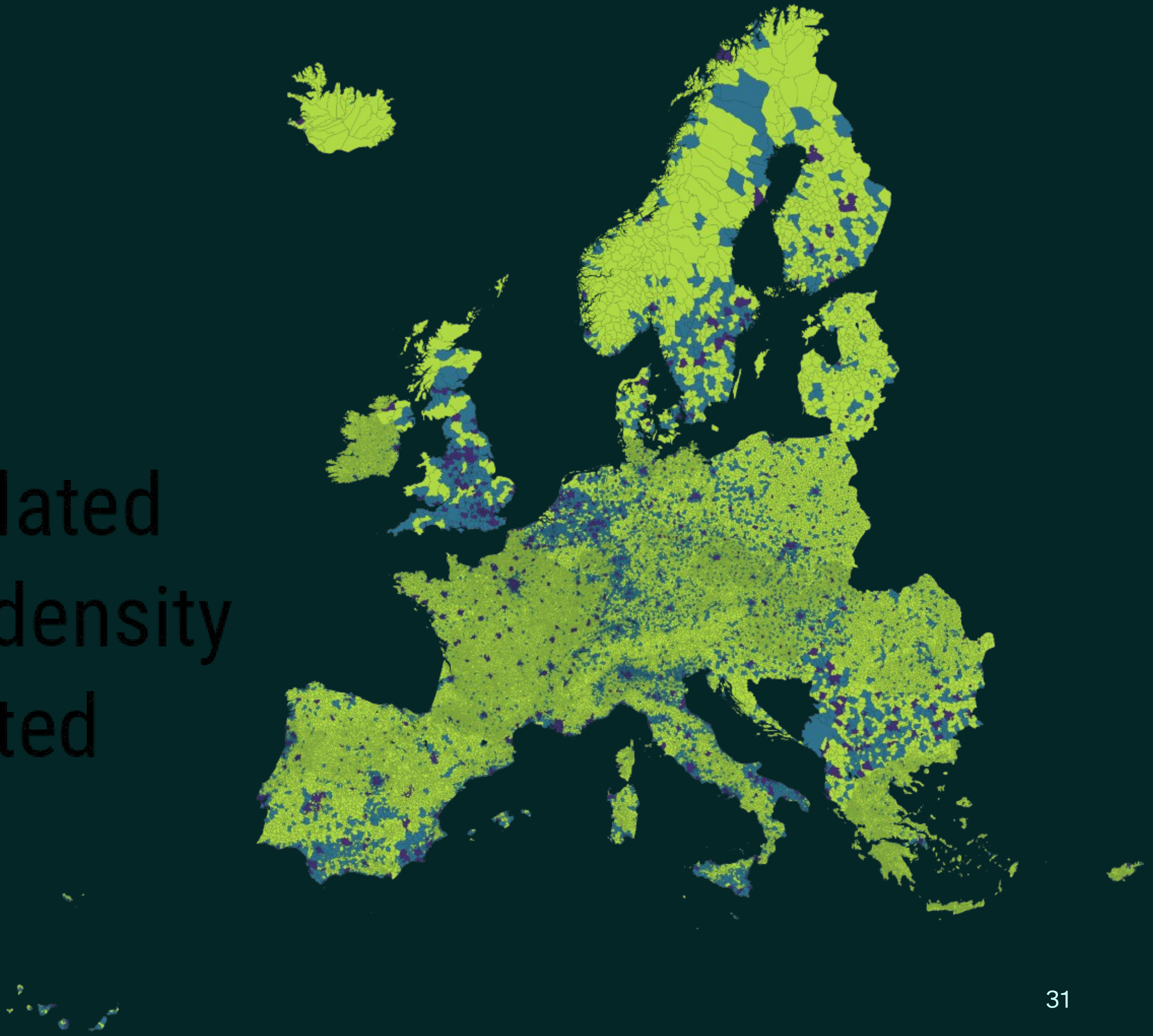
Maybe other countries?

DEGURBA classification

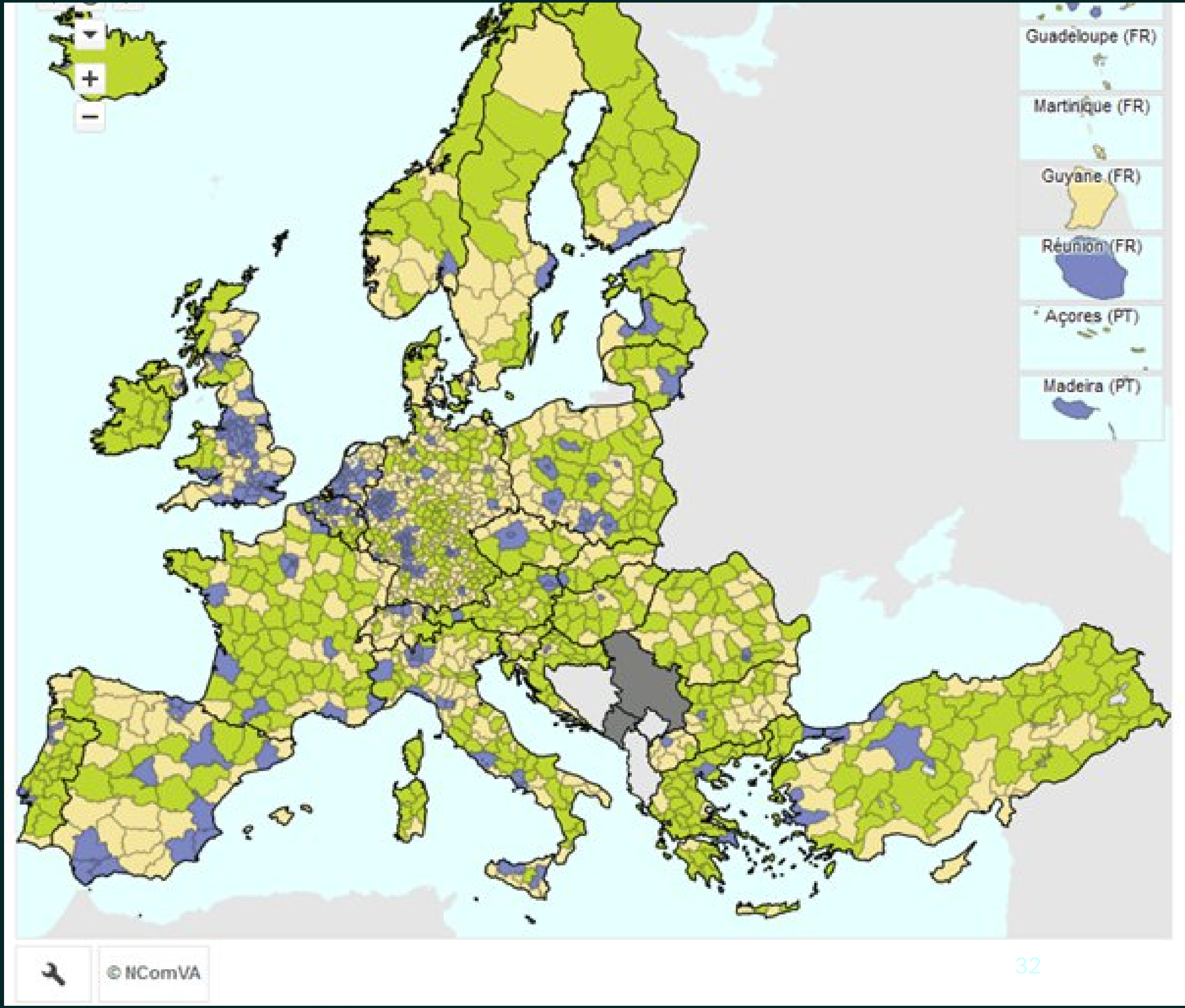
Densely populated

Intermediate density

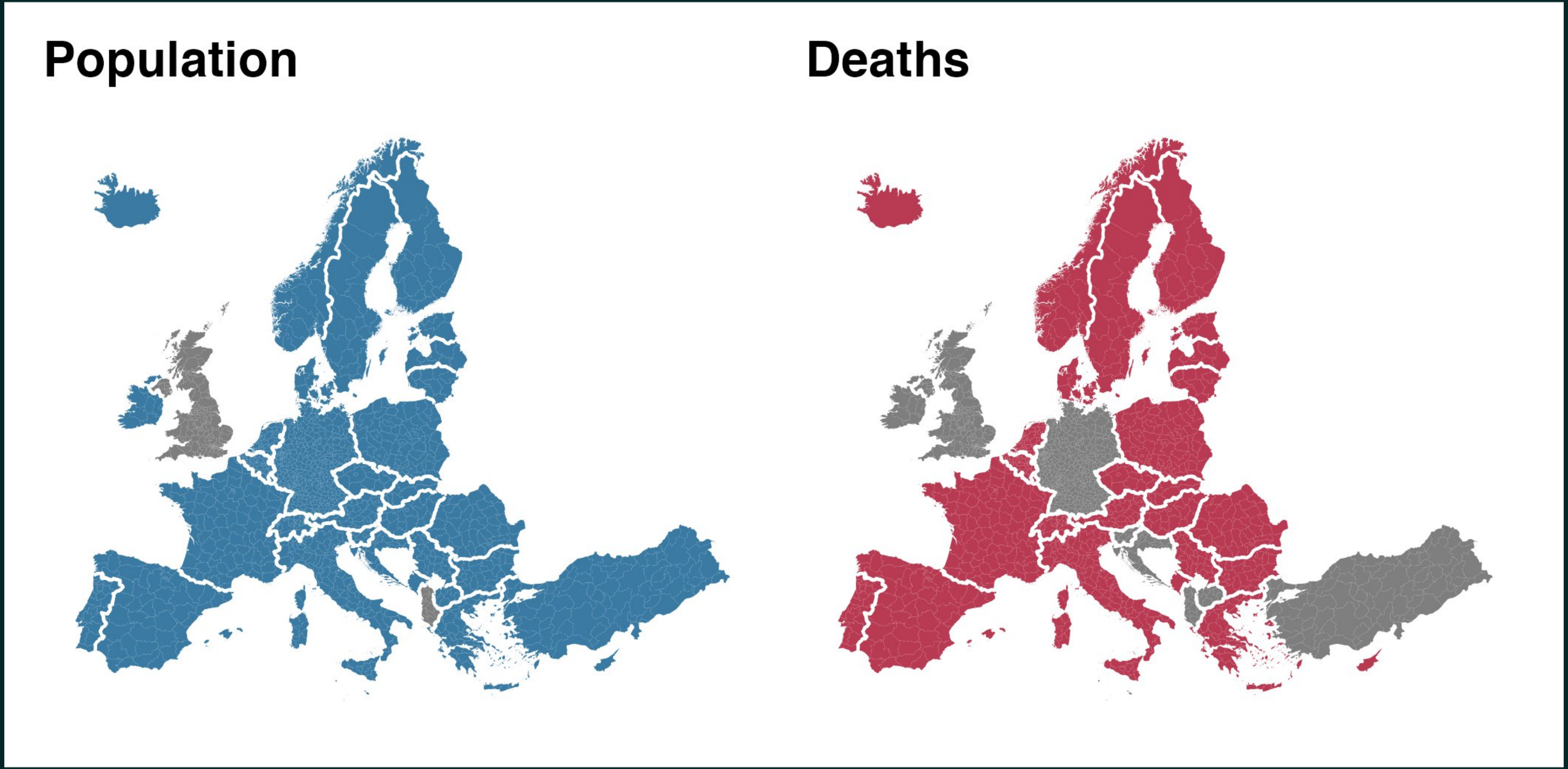
Thinly populated

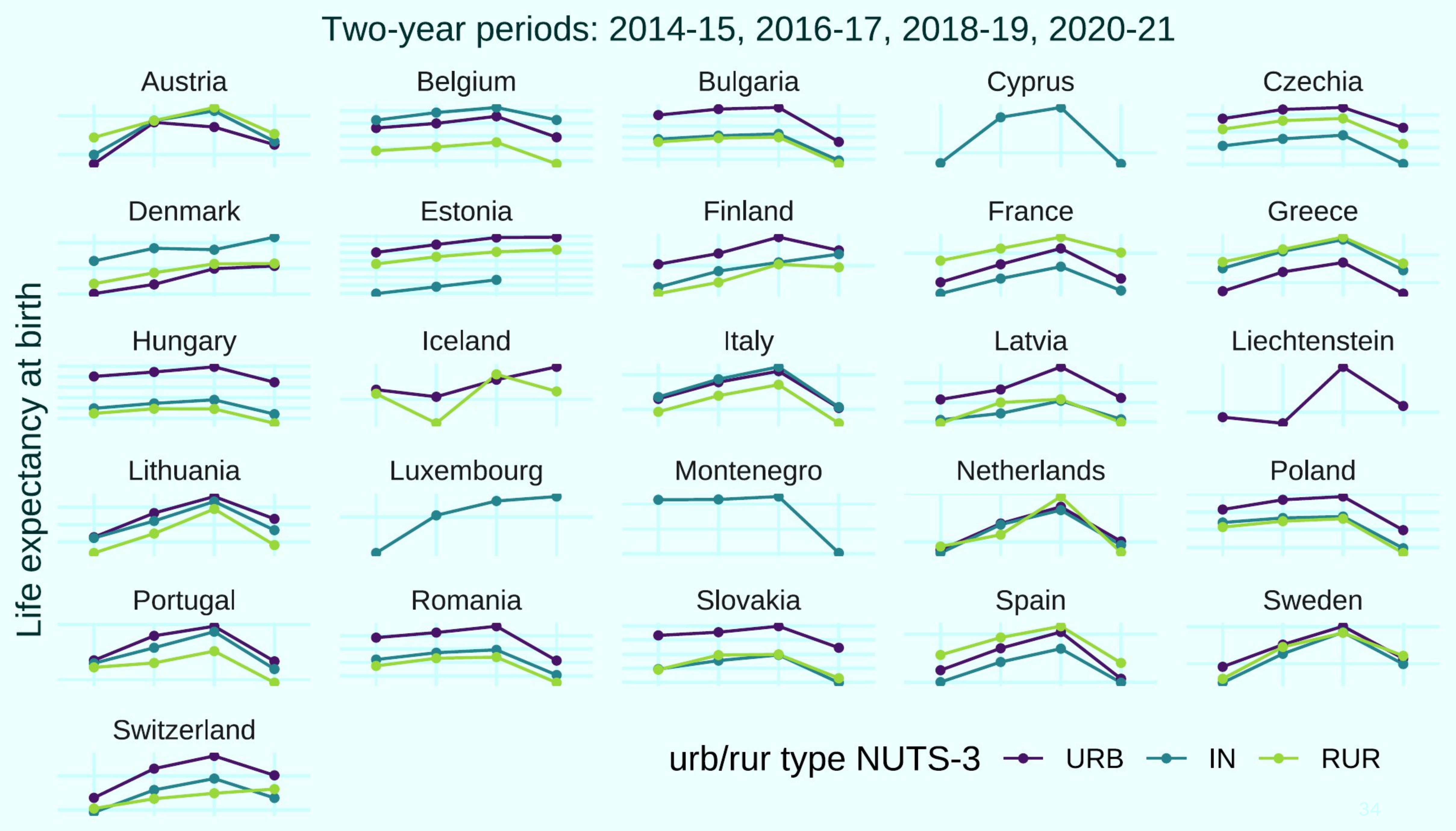


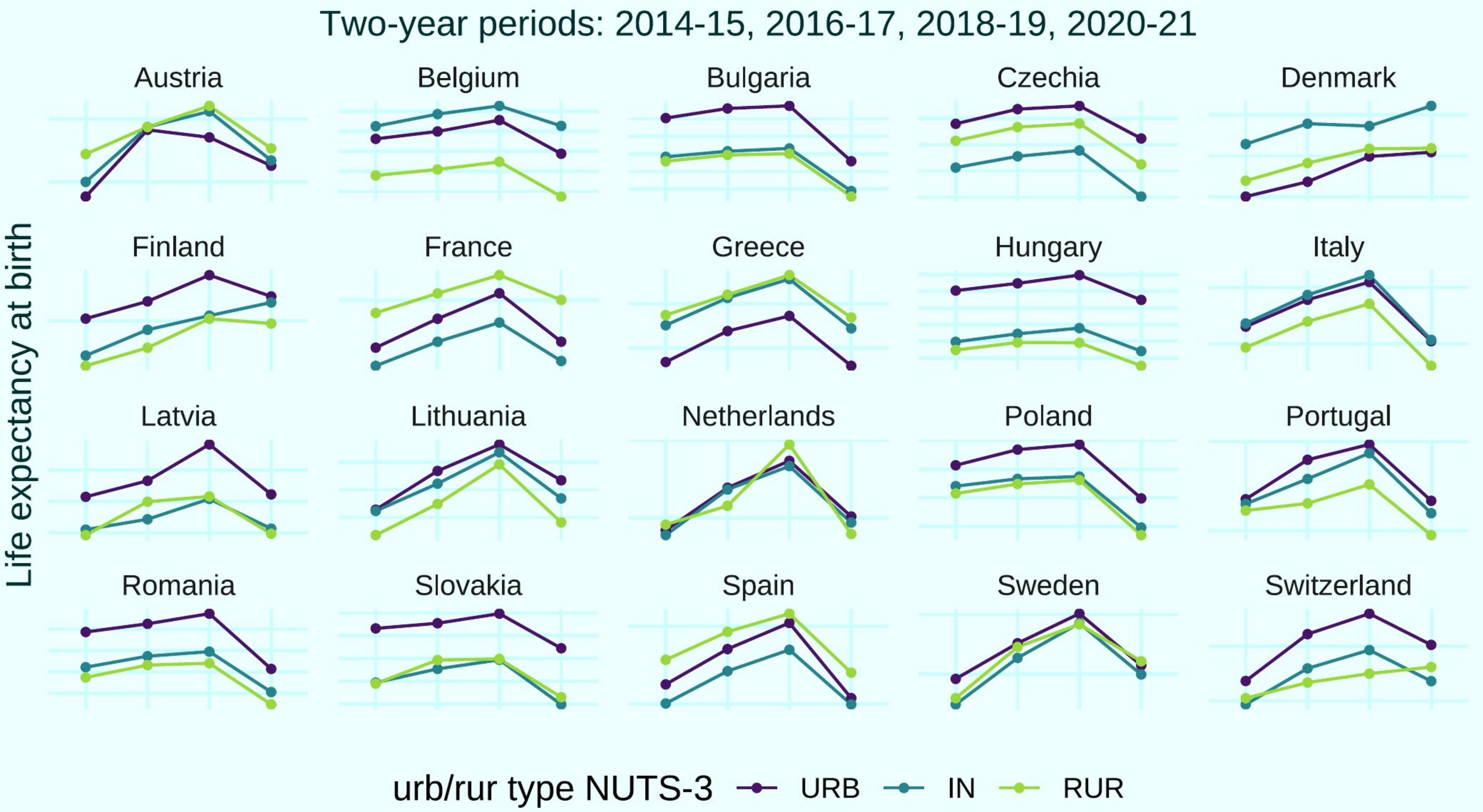
NUTS-3 classification



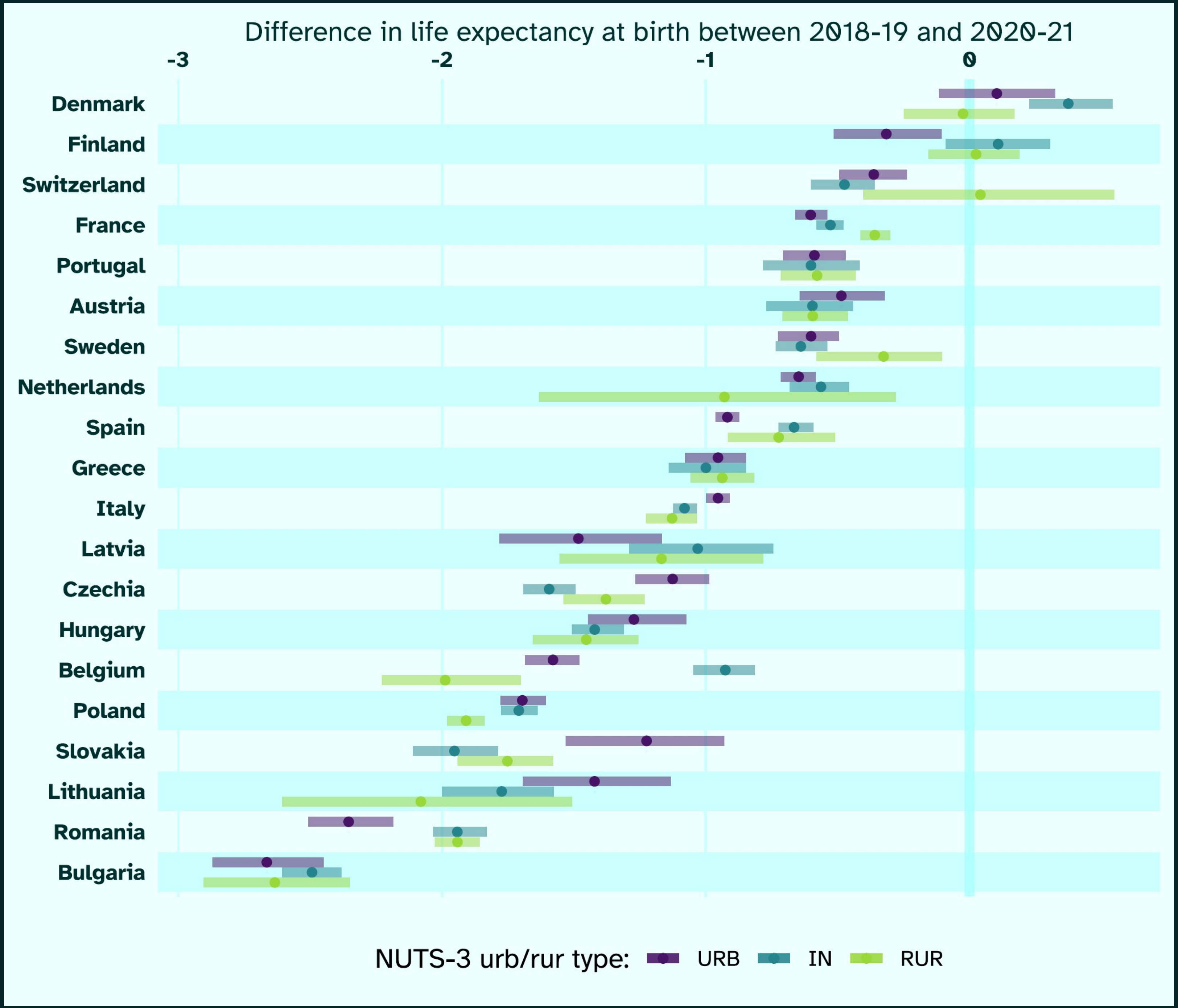
Matching the available data

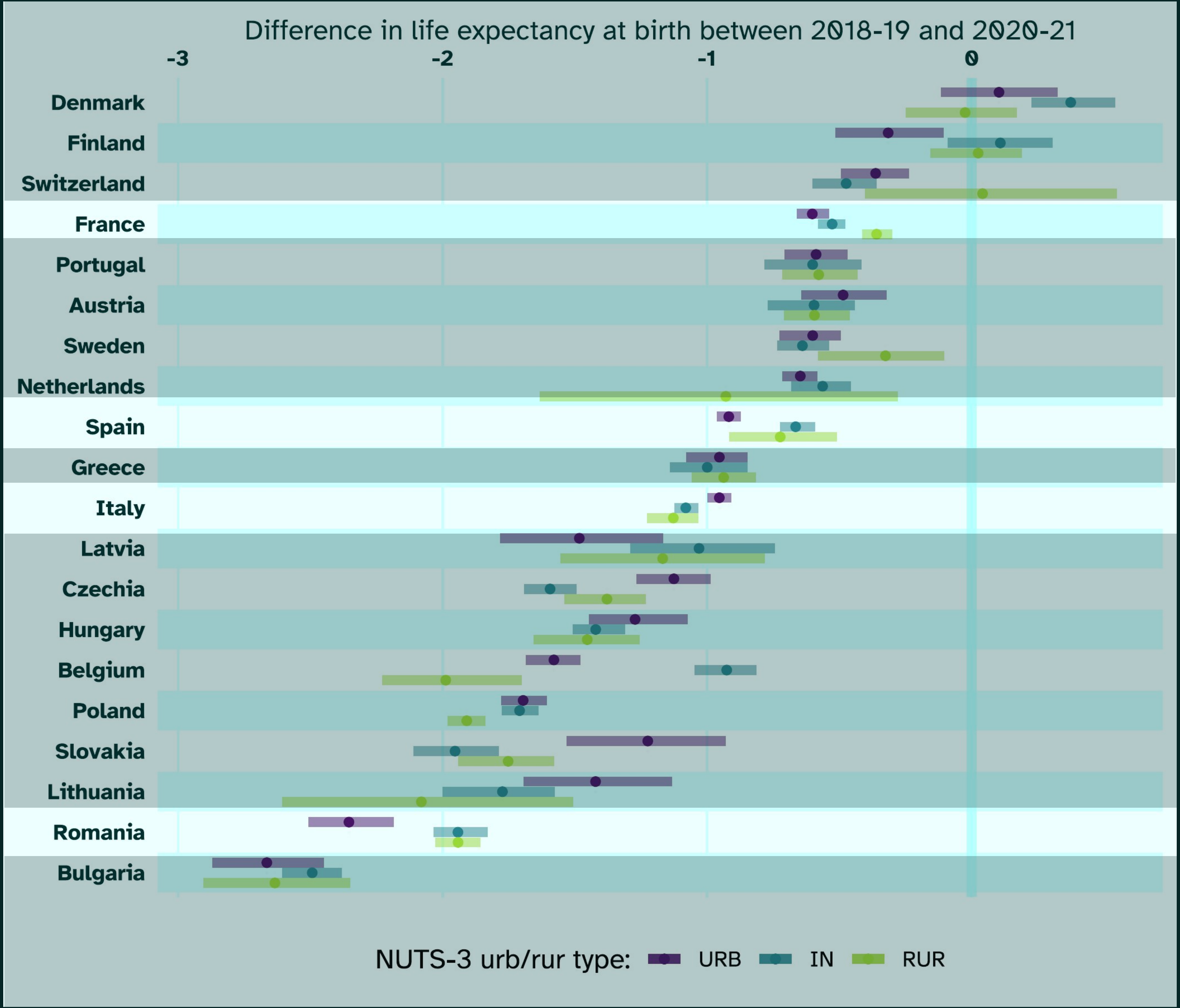



Two-year periods: 2014-15, 2016-17, 2018-19, 2020-21
Life expectancy at birth
Austria
Belgium
Bulgaria
Cyprus
Czechia
Denmark
Estonia
Finland
France
Greece
Hungary
Iceland
Italy
Latvia
Liechtenstein
Lithuania
Luxembourg
Montenegro
Netherlands
Poland
Portugal
Romania
Slovakia
Spain
Sweden
Switzerland
urb/rur type NUTS-3
URB
IN
RUR



Two-year periods: 2014-15, 2016-17, 2018-19, 2020-21
Austria
Belgium
Bulgaria
Czechia
Denmark
Finland
France
Greece
Hungary
Italy
Latvia
Lithuania
Netherlands
Poland
Portugal
Romania
Slovakia
Spain
Sweden
Switzerland
Life expectancy at birth
urb/rur type NUTS-3
URB
IN
RUR


Difference in life expectancy at birth between 2018-19 and 2020-21
-3
-2
-1
0
Denmark
Finland
Switzerland
France
Portugal
Austria
Sweden
Netherlands
Spain
Greece
Italy
Latvia
Czechia
Hungary
Belgium
Poland
Slovakia
Lithuania
Romania
Bulgaria
NUTS-3 urb/rur type:
URB
IN
RUR


Difference in life expectancy at birth between 2018-19 and 2020-21
-3
-2
-1
0
Denmark
Finland
Switzerland
France
Portugal
Austria
Sweden
Netherlands
Spain
Greece
Italy
Latvia
Czechia
Hungary
Belgium
Poland
Slovakia
Lithuania
Romania
Bulgaria
NUTS-3 urb/rur type:
URB
IN
RUR