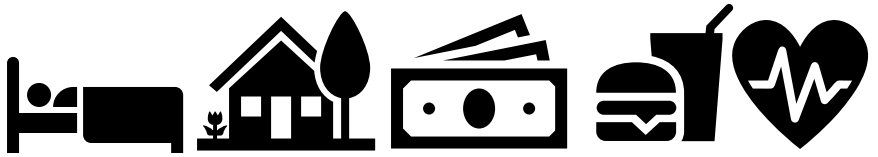
Statistics Finland

From household level weights to individual level weights: rethinking weighting adjustment in the context of EU-SILC

Oona Pentala-Nikulainen

BaNoCoSS June 2019

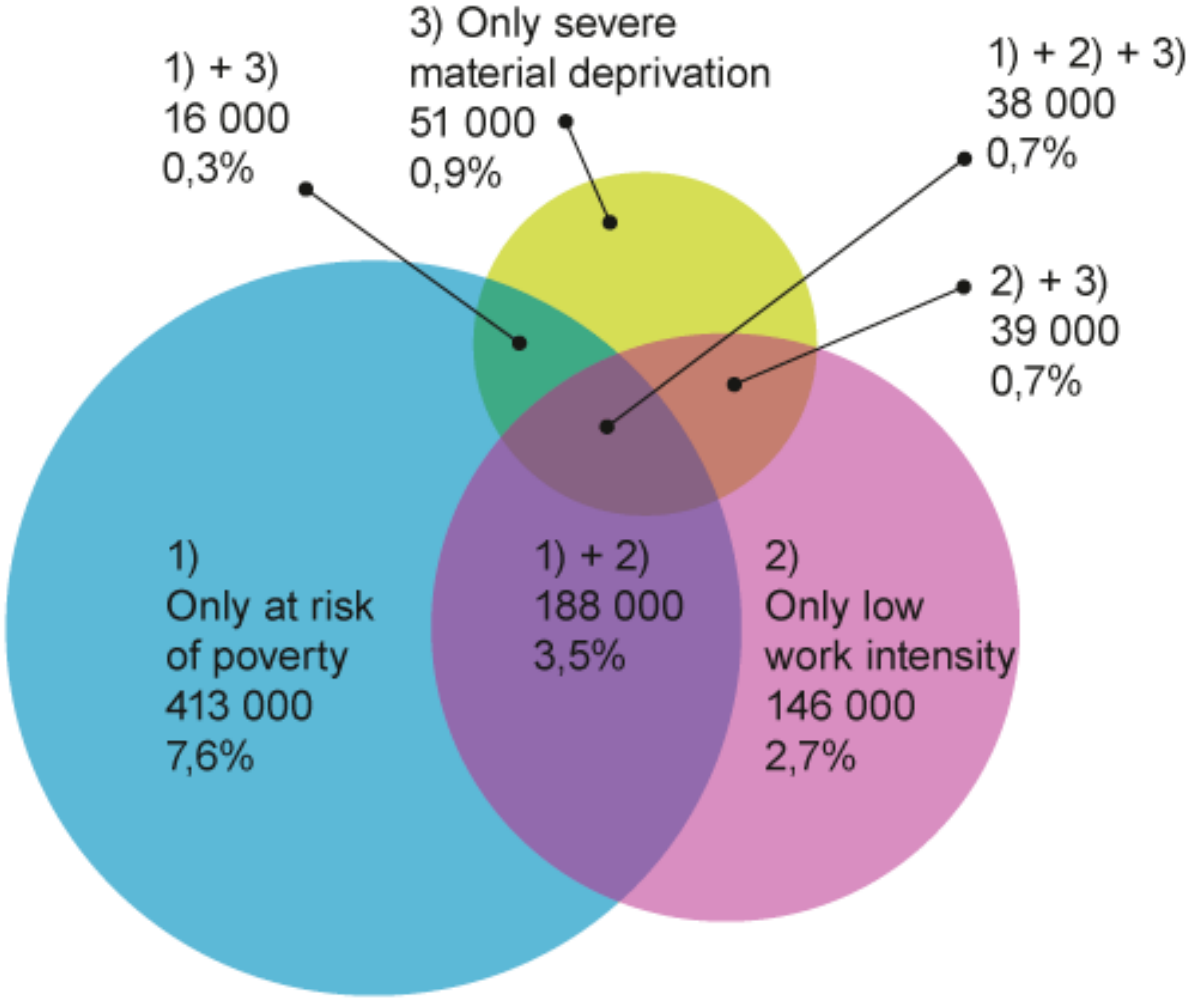
Statistics on living conditions



- The statistics on living conditions describe the living conditions of the household population from different perspectives, such as risk of poverty or social exclusion, subjective well-being and livelihood, health and housing by population group in Finland.
- Data are collected yearly with the income and living conditions survey. Some of the data is collected on household level and some on sample person level. Sample size is 10 000 households.
- These sample data are also used for the production of the income distribution statistics and Finland's data for Eurostat's EU-SILC survey.

Persons at risk of poverty or social exclusion by dimensions of risk in 2017

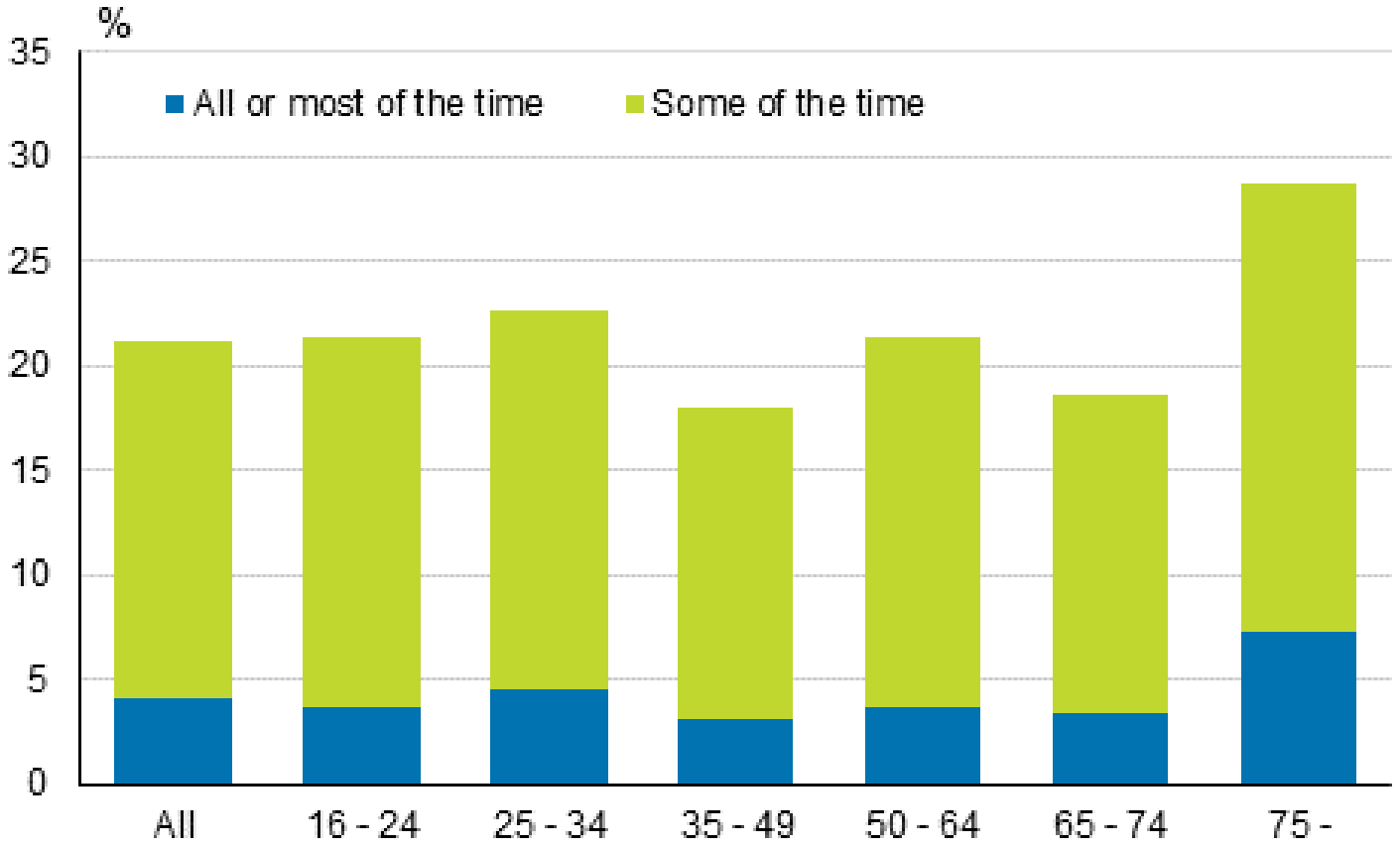
Most of the data depicting persons and households (e.g. level of education, marital status, income bracket) are derived from registers.



Statistics Finland

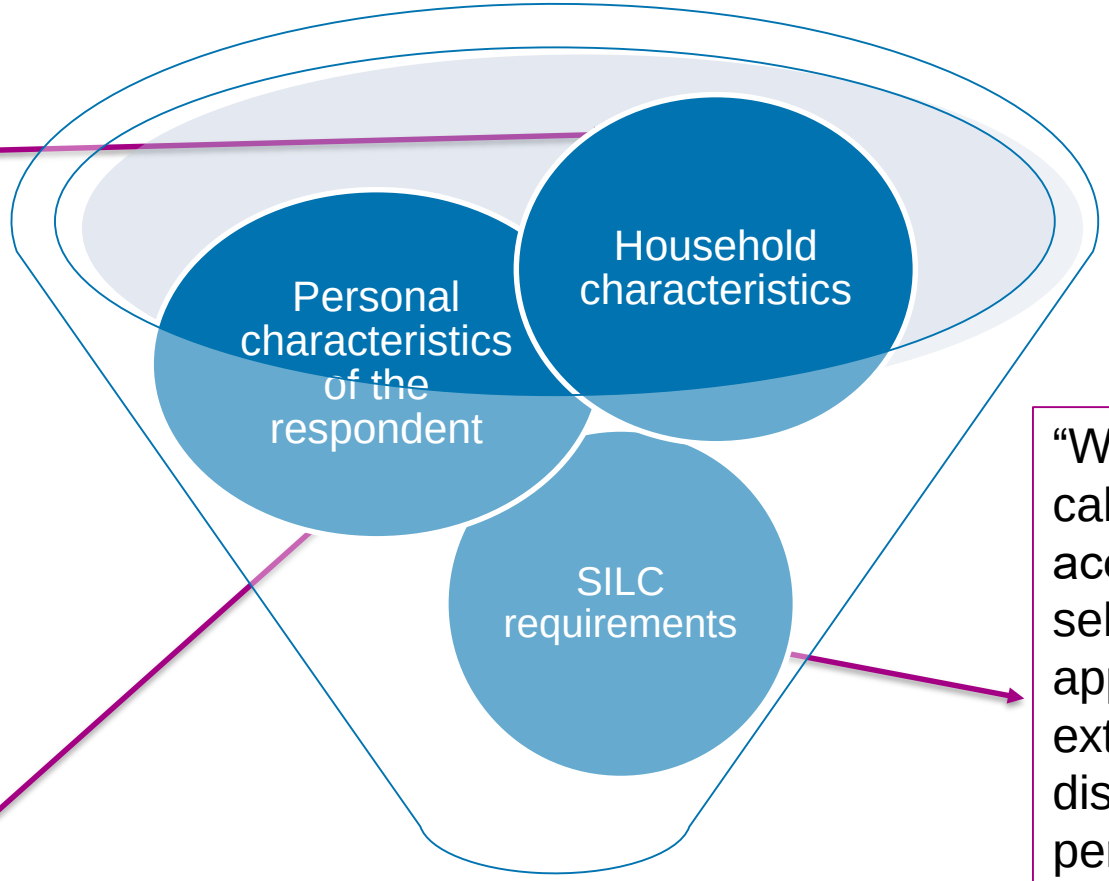
Frequency of feeling lonely in the past four weeks by age in 2018

But some data is collected from sample persons via CATI survey



Statistics Finland

Characteristics especially related to living conditions and household size



Characteristics especially related to quality of life and health

Adjusted personal level cross-sectional weights

"Weighting factors shall be calculated as required to take into account the units' probability of selection, non-response and, as appropriate, to adjust the sample to external data relating to the distribution of households and persons in the target population, such as by sex, age (five-year age groups), household size and composition and region (NUTS II level), or relating to income data from other national sources."

Design weights, nonresponse adjusted weights and scaling

Possible unequal selection probabilities are compensated

- Base weights/ Design weights
- Household level

The second stage is to compensate for the total nonresponse

- Total nonresponse adjusted weights
- Household level

Final nonresponse adjustments with calibration

- Final weights
- Individual level

- After the second stage the hh level weights have to be scaled to individual level for calibration
- Two possibilities:
  - Using k_i which is the number of persons in the hh
  - Using k^* which is the Total population size/ number of hh

Inclusion probabilities

nk_i/N or $\left(\frac{k_i}{k^*}\right)\frac{n}{N}$

Calibration adjustment

- The key feature of this step is the modification of the total nonresponse adjusted weights to reproduce characteristics from the sample population, namely totals and category frequencies
- More precisely, suppose that there exist J auxiliary variables $x_1 \ldots x_j \ldots x_J$ called calibration variables, with known population totals or marginal counts.
- We seek new weights that are "as close as possible" (as determined by a certain distance function) to the initial weights.
- These new weights are calibrated on the totals X_j of the J auxiliary variables; in other words they verify the calibration equations:

$$\forall j = 1 \ldots J \quad \sum_{k \in s} w_k \times x_{jk} = X_j$$

$$where\ w_k = g_k \times total\ nonresponse\ corrected\ weight$$

Choice of (new) calibration variables

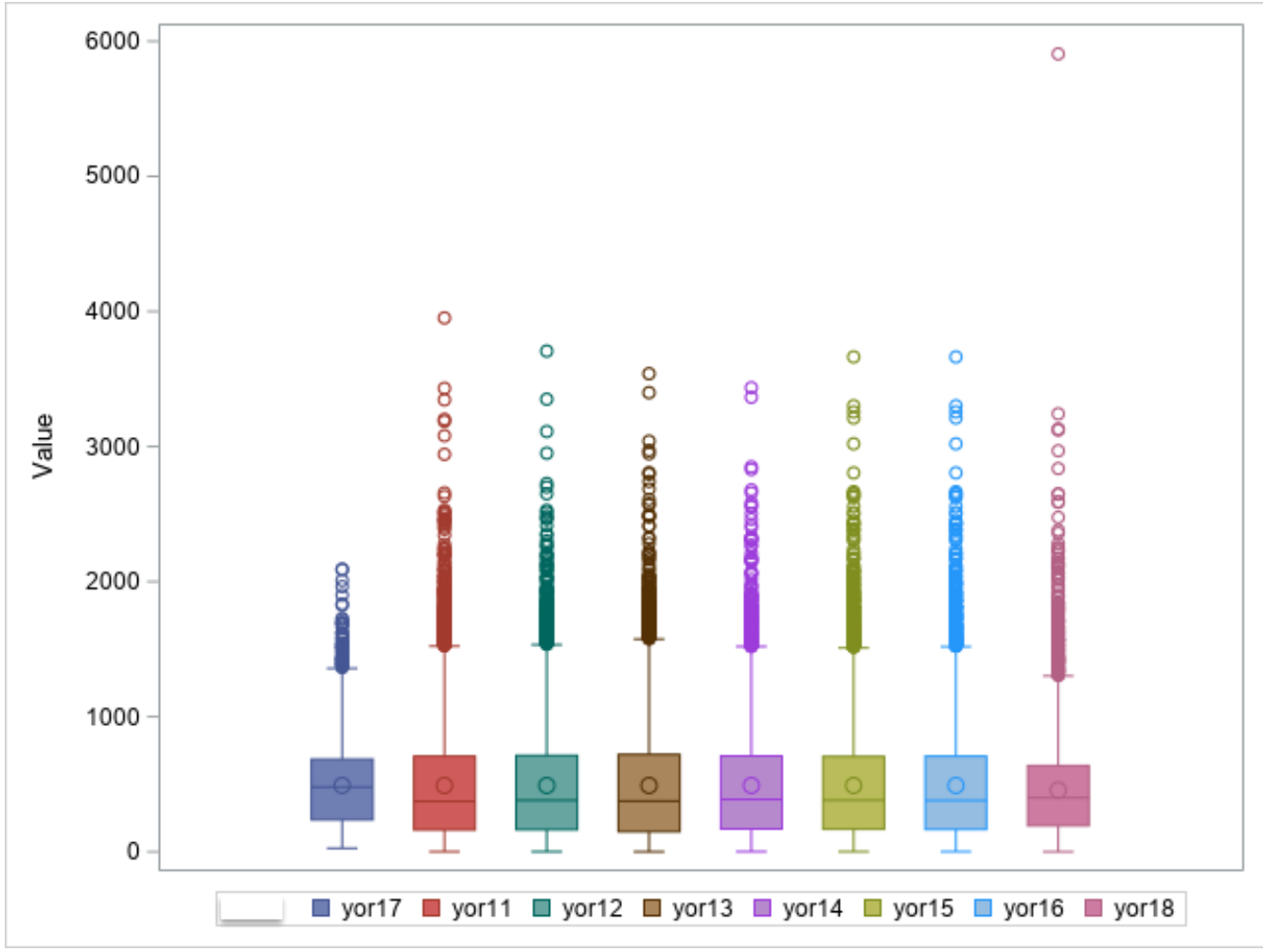
- The variables used in calibration should be related to sample design, nonresponse and the main research phenomenon
- We started from trying to preserve the previous hh based model and started adding new variables
- Our administrative sources were quite rich and we had a choice e.g. between
  - Various income variables such as beneficiary group, income quintile, socioeconomic status
  - Sociodemographic variables such as age, gender, marital status
  - Geographic variables such as municipality or county

Possible calibration variables

D/T	Variable	ykor11	ykor12	ykor13	ykor14	ykor15	ykor16	final model	EU-SILC pb060
D	Region	*	*	*	*	*	*	*	*
D	Degree of urbanisation	*	*	*	*	*	*	*	
D	Household composition	*	*	*	*	*	*	*	*
D	Marital status	*	*	*					
T	Gender x agegroup	*	*	*	*	*	*	*	*
T	Low education level	*	*	*	*	*	*	*	*
T	Upper educaiton level	*	*	*	*	*	*	*	*
T	High education level	*	*	*	*	*	*	*	*
T	Wage earners	*	*	*	*	*	*		*
T	Unemployment benefit earners	*	*	*	*	*	*		*
T	Pension earners	*	*	*	*	*	*		*
T	Income less than 60% of population median	*				*			
D	Income decile			*			*		
T	Small income group	*	*	*	*	*	*	*	
D	Socioeconomic status							*	
T	Lowest income quintile							*	

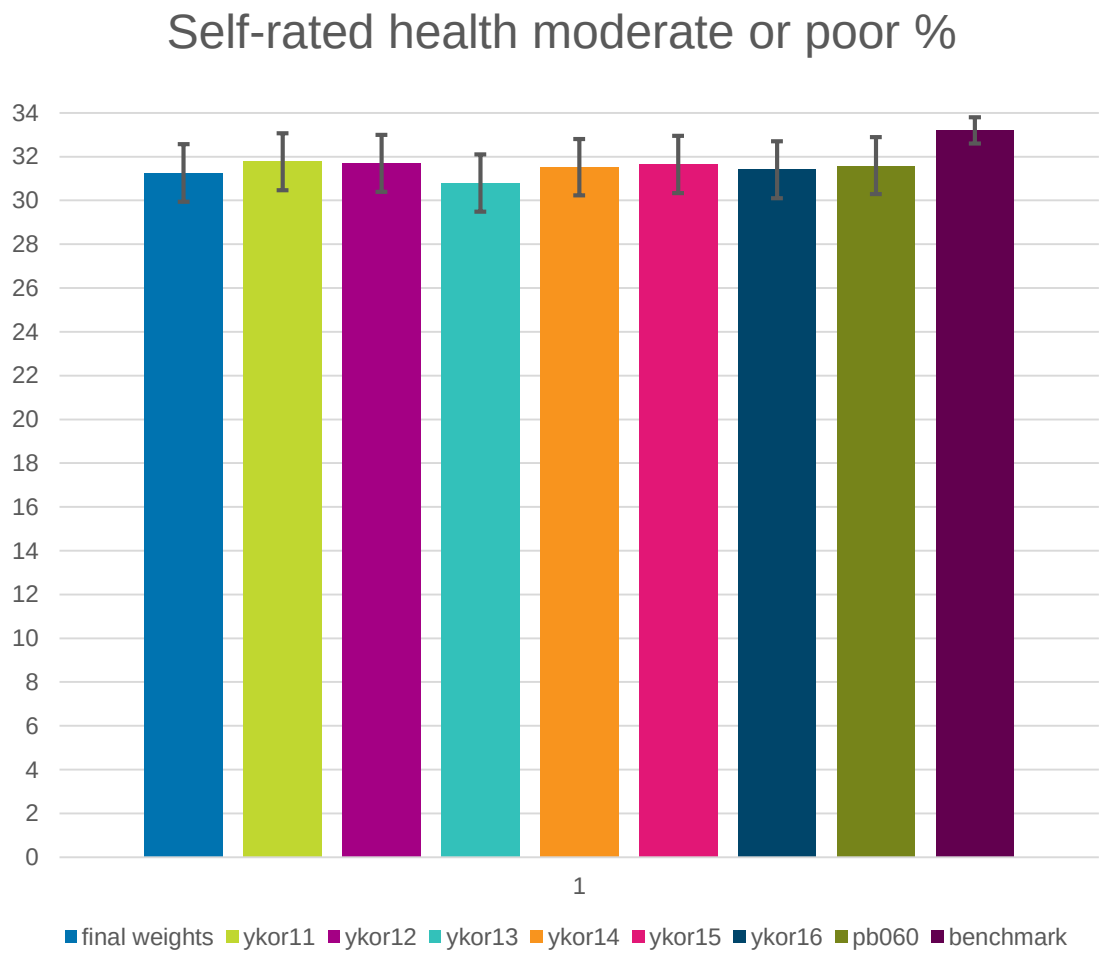
Comparing calibrated weights' descriptives

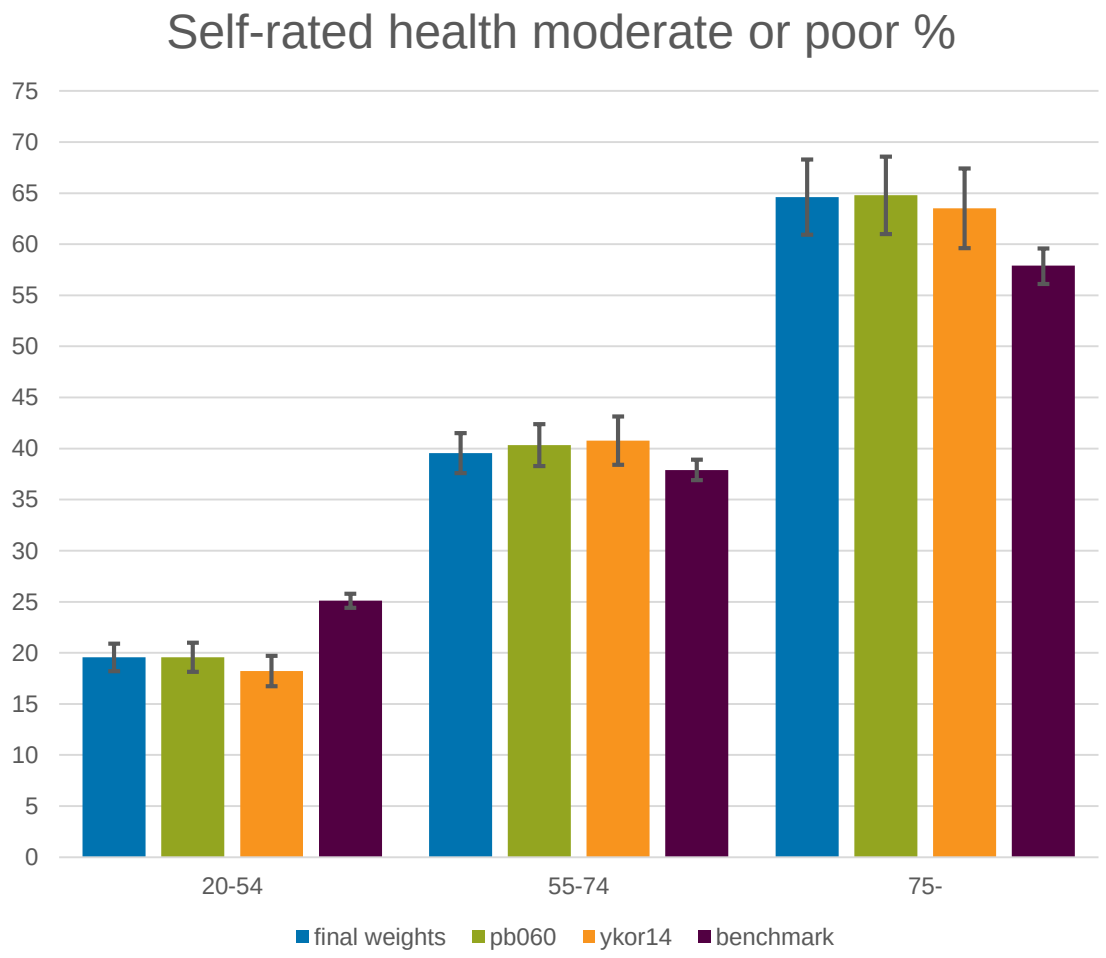
Variable	Median	Mean	Coeff of Variation	Min	Max	Range
Final weight	479.3	491.0	59.5	26.2	2092.9	2066.8
ykor11	372.9	490.4	87.5	0.9	3951.7	3950.8
ykor12	381.3	490.5	85.1	1.0	3706.5	3705.5
ykor13	375.6	490.6	88.1	0.4	3539.0	3538.7
ykor14	387.6	490.5	83.3	1.3	3435.1	3433.8
ykor15	381.4	490.4	85.9	1.2	3663.1	3661.9
ykor16	379.8	490.6	85.3	1.2	3663.1	3661.9
pb060	399.8	456.1	74.0	0.3	5906.3	5906.0



Statistics Finland

Results with differenct weights





Statistics Finland

Persons at risk of poverty or social exclusion by age and gender (%)

Age group	Gender	Register	New weights	EU-SILC Pb060
16-19	Men	13.5	13.6	16.1
	Women	20.7	20.8	19.9
20-29	Men	22.3	26.8	23.8
	Women	21.1	22.1	19.0
30-39	Men	11.9	10.0	12.6
	Women	8.8	11.7	8.2
40-49	Men	7.8	6.1	7.4
	Women	7.8	3.9	7.9
50-59	Men	12.9	16.5	13.0
	Women	7.1	9.3	6.9
60-69	Men	8.0	9.8	7.8
	Women	6.2	7.1	7.1
70-79	Men	7.5	5.6	7.8
	Women	15.3	13.0	14.9
80+	Men	18.2	14.9	19.0
	Women	32.2	24.3	32.7
All	All	12.7	12.8	12.7

Discussion & conclusion

- Creating new weights was found **useful** and made using the data **easier** since the selected weighting model created **more stable weights**
- New weights did not significantly change any opinion based results
  - these weights were sent to Eurostat as an optional national weights for the 2018 ad-hoc module
  - The results were in accordance to the national income distributin statistics as well
- **Some changes** were observed **when comparing** weighted results **with registers**
- **The weighting procedure should always start from defining the purpose**; this makes choosing the calibration variables easier

Thank You!

Statistics Finland

Oona Pentala-Nikulainen, oona.pentala-nikulainen@stat.fi

BaNoCoSS 2019, Örebro

References

- Deville, J.-C. and Särndal, C.-E (1992). "Calibration estimation in survey sampling". Journal of the American Statistical Association, 87, n°418, pp. 375-382.
- European Comission. COMMISSION REGULATION (EU) 2017/310. Official Journal of the European Union. 2017 (1).
- European Comission. METHODOLOGICAL GUIDELINES AND DESCRIPTION OF EU-SILC TARGET VARIABLES. 2017 (2).
- Official Statistics of Finland (OSF): Statistics on living conditions [e-publication]. 2017. Helsinki: Statistics Finland [referred: 24.5.2019]. Access method: http://www.stat.fi/til/eot/2017/eot_2017_2019-05-24_tie_001_en.html