

gesis

Leibniz Institute
for the Social Sciences



The institutionalized population in survey research

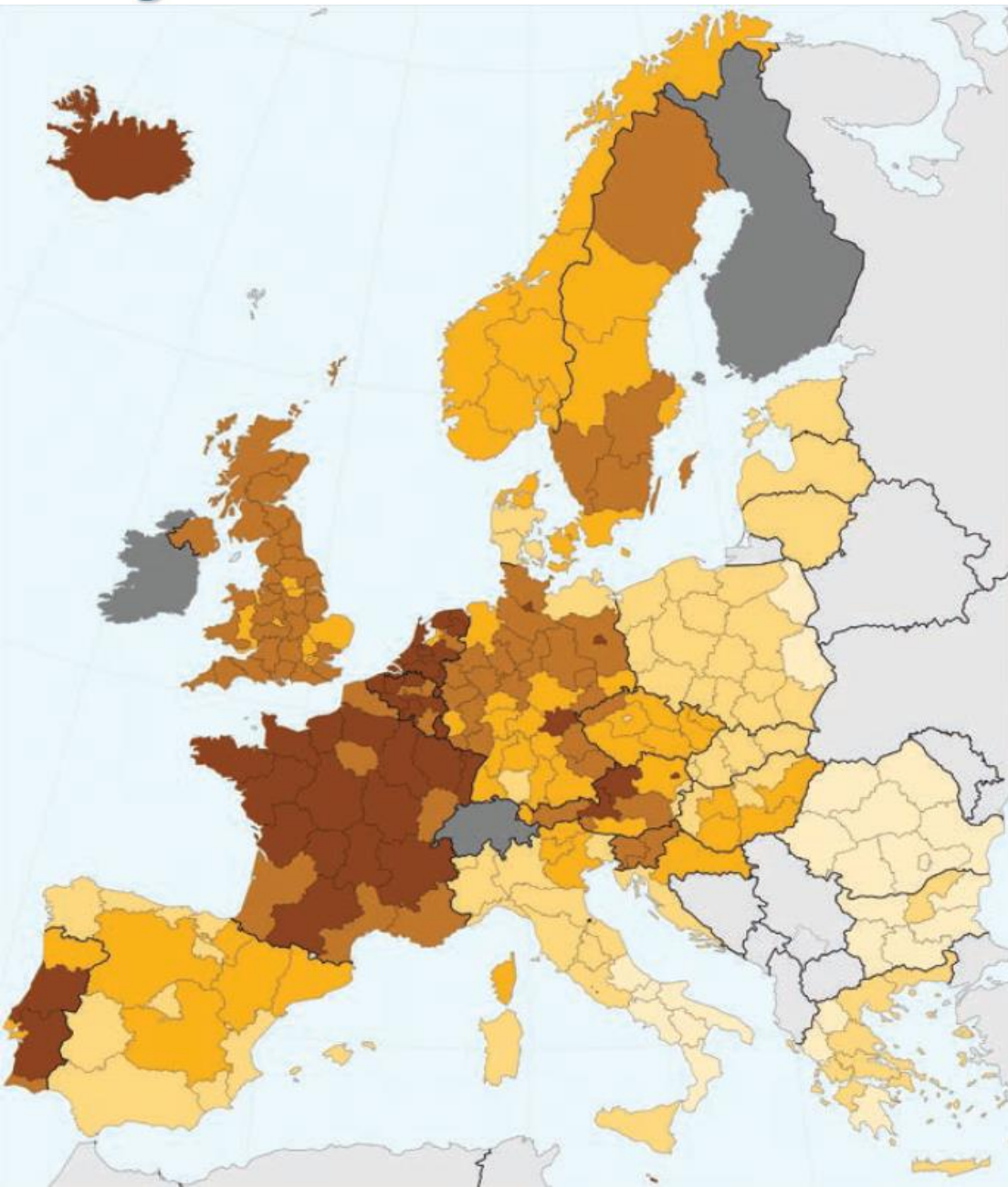
Assessing the peril of coverage bias with cross-national survey data

Jan-Lucas Schanze & Stefan Zins

ESRA conference, Lisbon (PT), 21st of July 2017

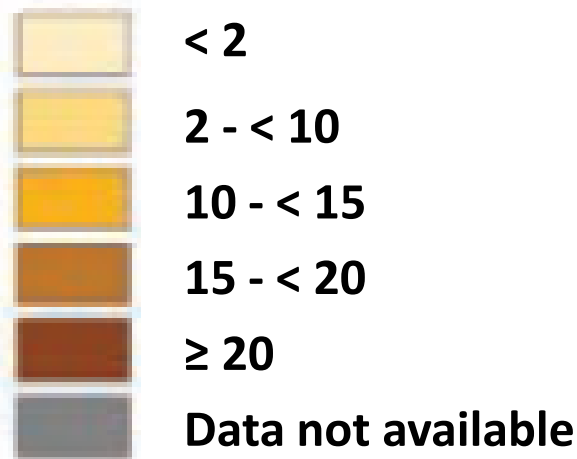


This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 654221.



Share of population aged 85 years and over living in an institutional household

(%) at NUTS 2



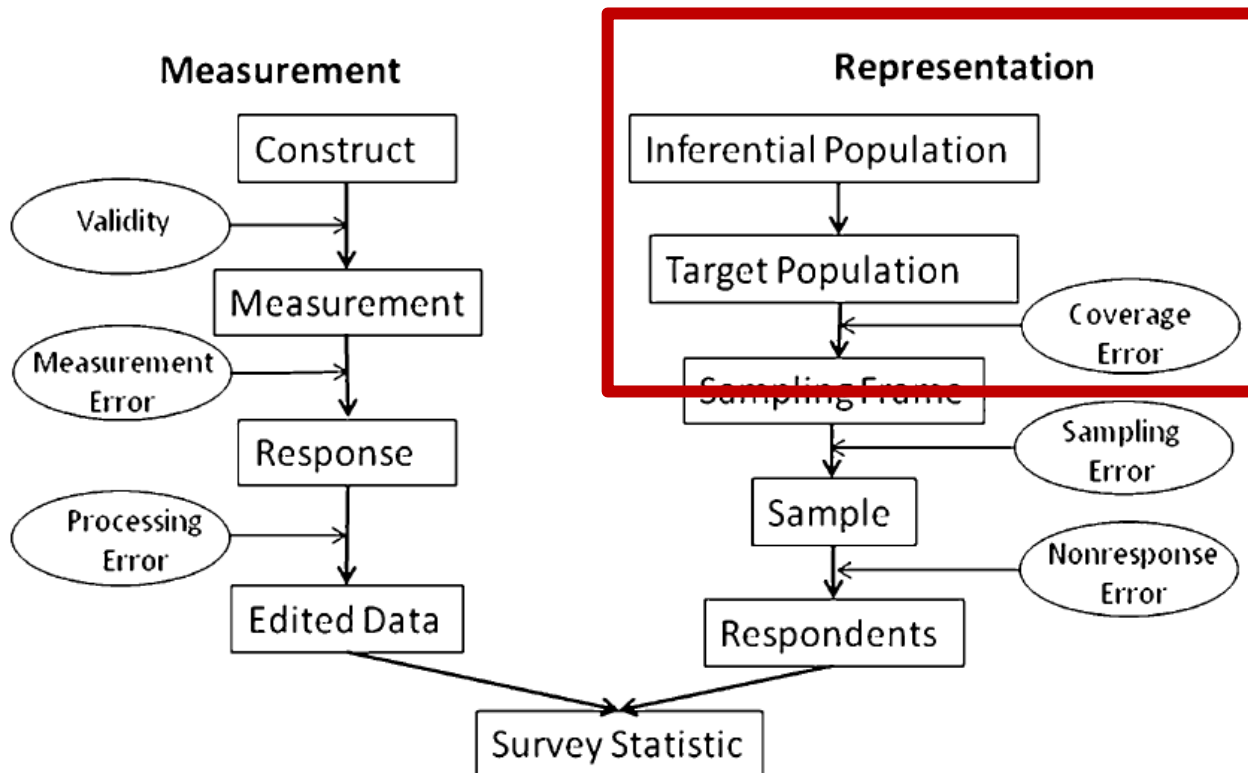
Source: Eurostat 2015. „People in the EU“

Research question

If social surveys infer their results from a sample of private households to the general population, ...

... to what extent do they suffer from coverage error due to the size and distinctiveness of the institutionalized population?

Total Survey Error Framework



Research findings: Significant differences

- Health-related variables
 - State of health, dementia, self-rated health
- Coping with daily living & housekeeping
 - Functional and/or cognitive impairments
- Demographic variables
 - Gender, age, ethnicity
- Social ties
 - Networks, marital status, parenthood, availability of informal caregiver
- Socio economic variables
 - Income, economic activity, housing, education
- Contextual variables
 - Social policies, demography of countries/regions

Hypotheses

Assuming bias in case of noncoverage or undercoverage of the institutionalized population

- H1: Increasing the coverage rate of institutionalized respondents will increase the aggregate mean of the number of limitations in activities of daily living (ADL)
- H2: Increasing the coverage rate of institutionalized respondents will decrease the aggregate mean of the self-perceived health perceptions

Data – Compiling our statistical population

- Five waves of the Survey of Health, Ageing and Retirement in Europe (SHARE)
 - Cross-national panel survey in Europe, beginning in 2004
 - The target population comprises the population aged 50 years or older living in private households *and institutions for the elderly*
- Pooling the data: Extracting the most recent observation for each SHARE respondent
 - Cross-sectional dataset with 91,671 cases without missing values in 15 countries
 - *Among these:* 2.0% (N = 1,840) identified as institutionalized residents

Our statistical population

(%)		Private hh.	Institutions
Gender	Male	45.2	35,5
	Female	54.8	64.5
Age	50 to 59	26.2	8.6
	60 to 69	34.4	14.7
	70 to 79	25.3	23.7
	80 to 89	12.6	37.0
	Older than 90	1.6	16.1
Marital status	Married	71.3	35.6
	Never married	5.6	11.2
	Divorced	8.7	10.7
	Widowed	14.4	42.5
Number of observations		89,831 (98%)	1,840 (2%)

Our statistical population (cont.)

(%)		Private hh.	Institutions	Total
Limitations with ADL	No limitations	87.2	51.1	86.5
	1-3 limitations	10.3	27.5	10.7
	>3 limitations	2.5	21.4	2.9
Self- perceived health	Poor	10.1	19.6	10.3
	Fair	27.6	38.3	27.9
	Good	36.2	30.2	36.1
	Very good	18.5	7.9	18.2
	Excellent	7.6	4.1	7.6
Number of observations		89,831	1,840	96,671

Monte Carlo Simulation

- 5,000 times simple random sampling with replacement of 3,000 observations with 5 different coverage rates

$$\frac{N_i}{N} (Y_p - Y_i)$$

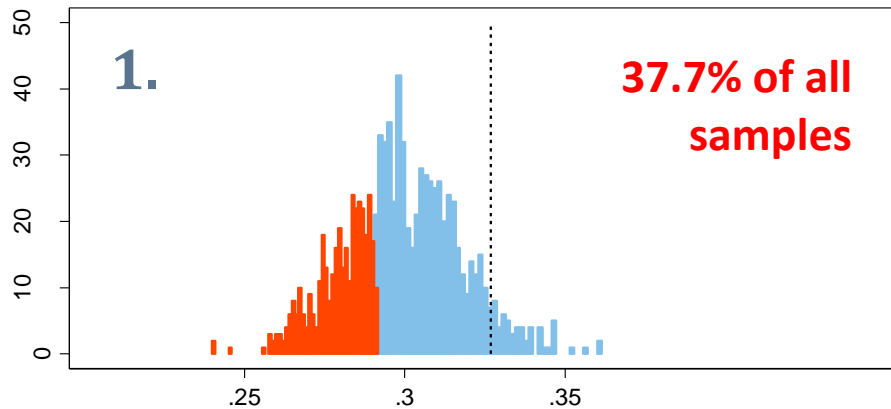
- **N** = Total number of residents
- **N_i** = Number of institutionalized residents
- **Y_p** = Mean of residents living in private households
- **Y_i** = Mean of institutionalized residents

	Degree of inclusion of N_i in target population				
	0%	25%	50%	75%	100%
Expected elements from N_i in samples of 3,000 obs	0	15	30	45	60

Weighting the samples

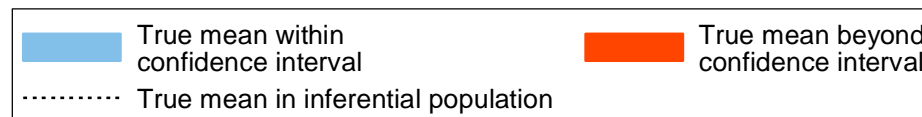
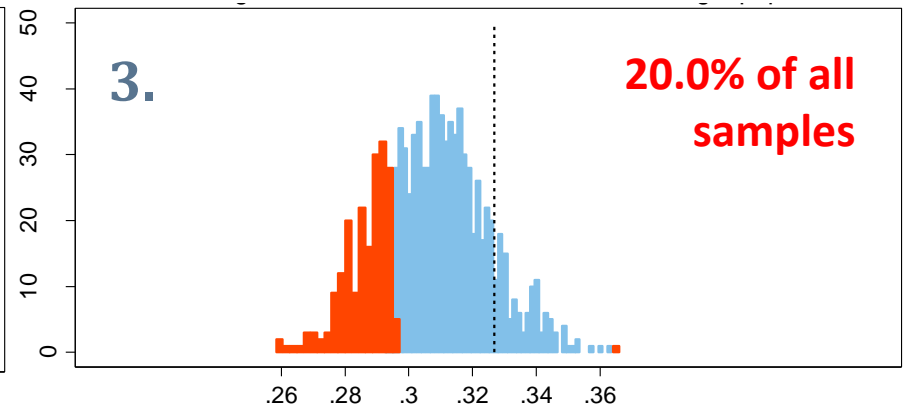
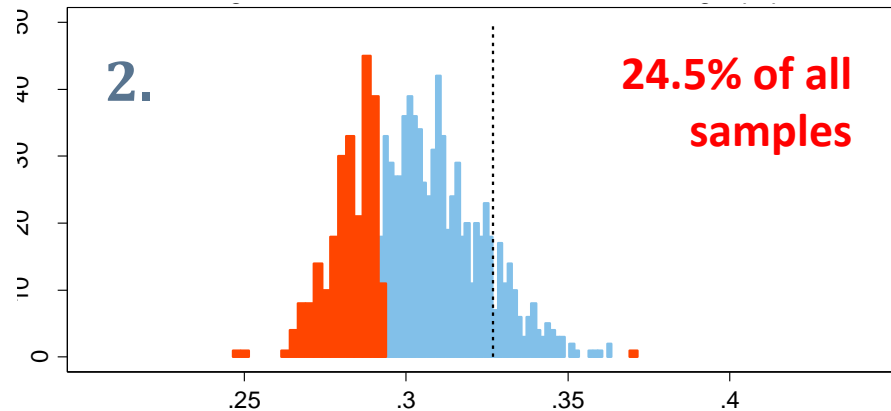
- Applying different weighting schemes
 1. Unweighted sample
 2. Traditional survey weights
 - Age, gender, country
 3. Multivariate control weights
 - Age, gender, country
 - Marital status, partner in household, origin, number of children, household size
 - Country-specific income percentiles
 - Limited in health, long-term illness

Mean of limitations in ADL

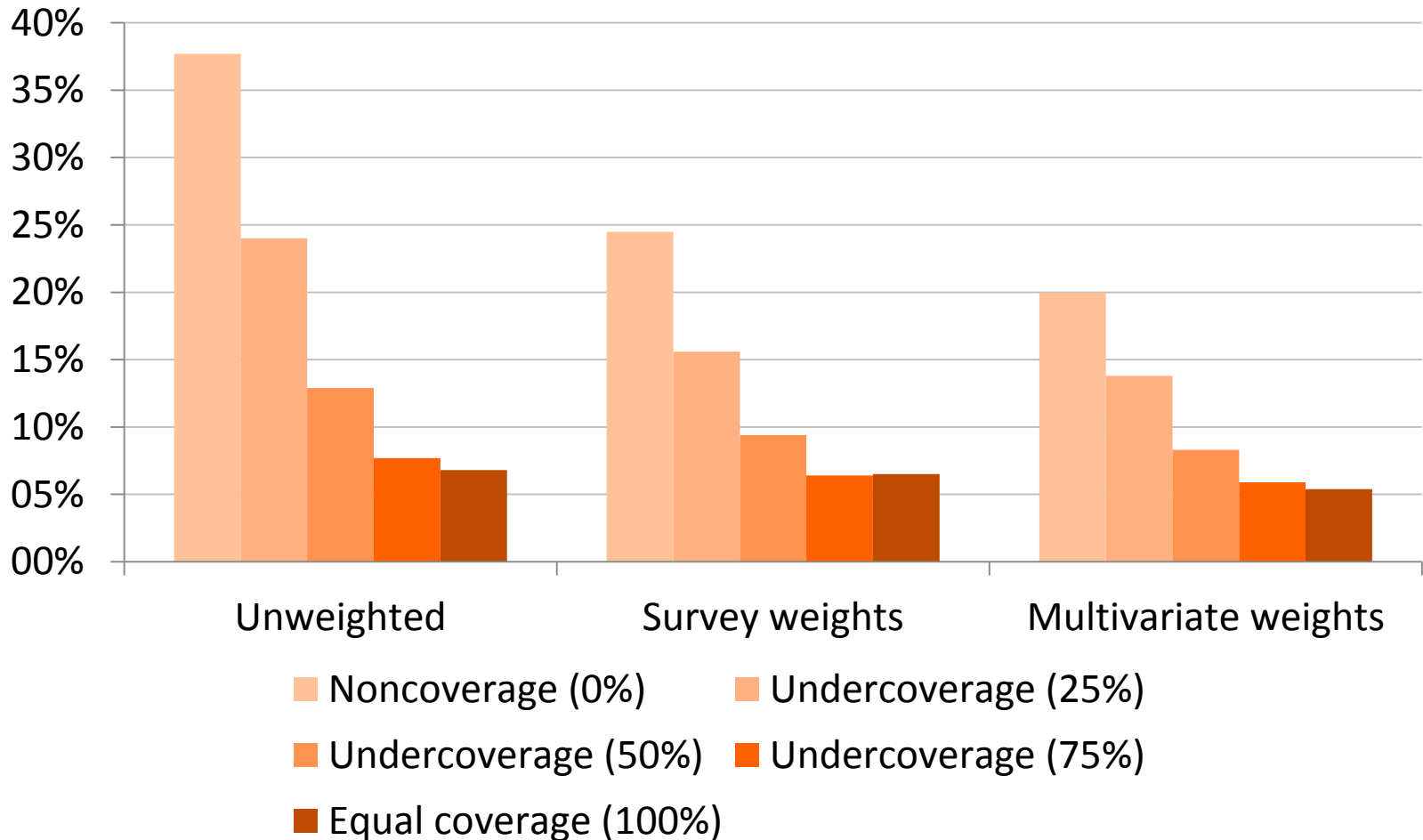


Noncoverage ($N_i = 0$)

1. Unweighted
2. Survey weights
3. Multivariate control weights

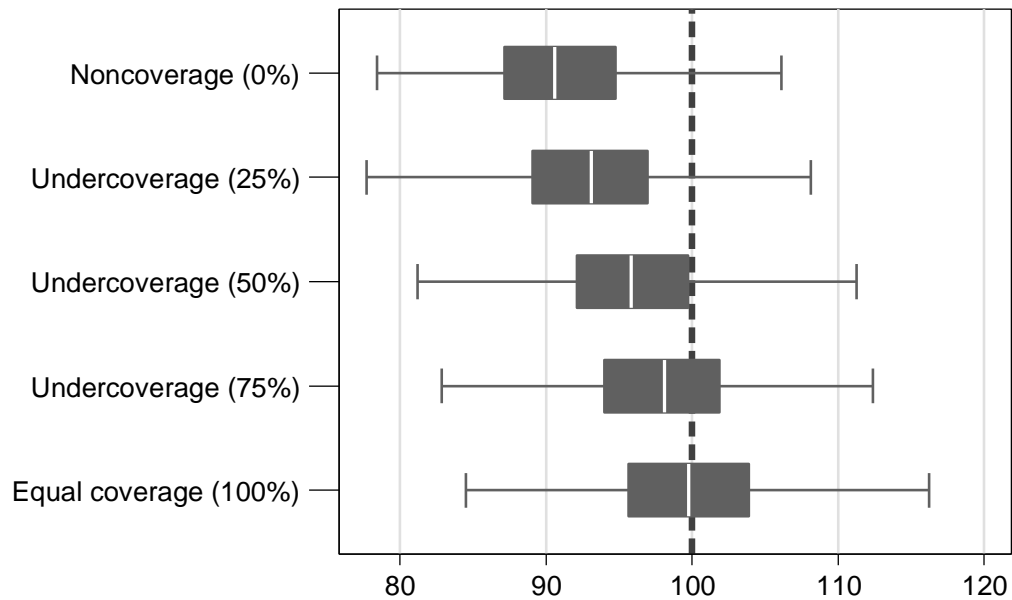


Mean of ADL: Biased samples

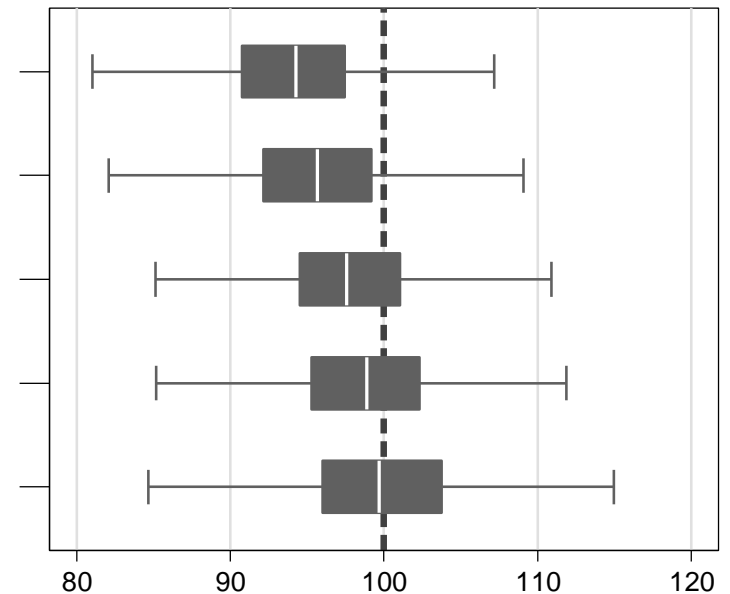


Deviation of means in ADL

1. Unweighted



3. Multivariate weights

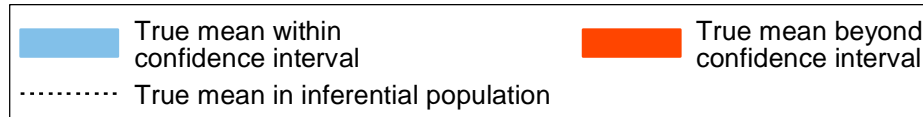
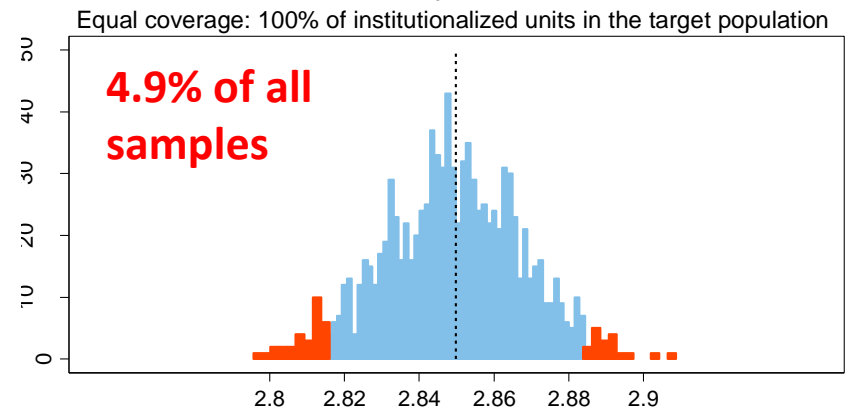
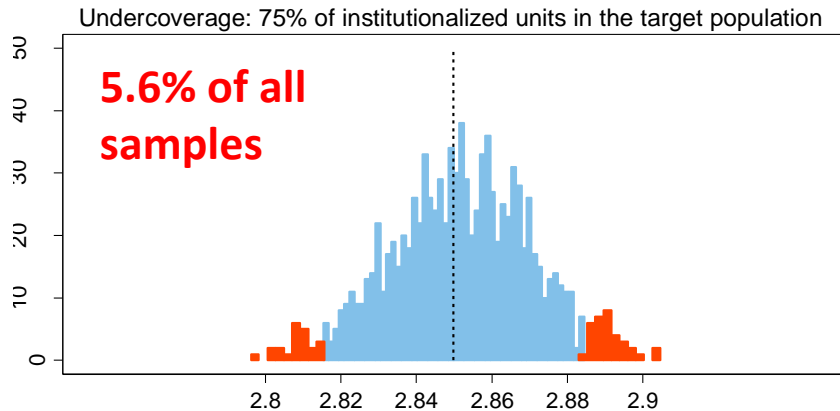
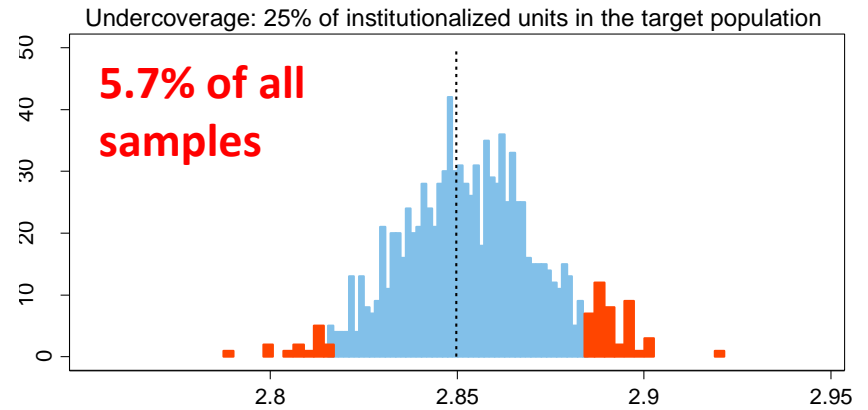
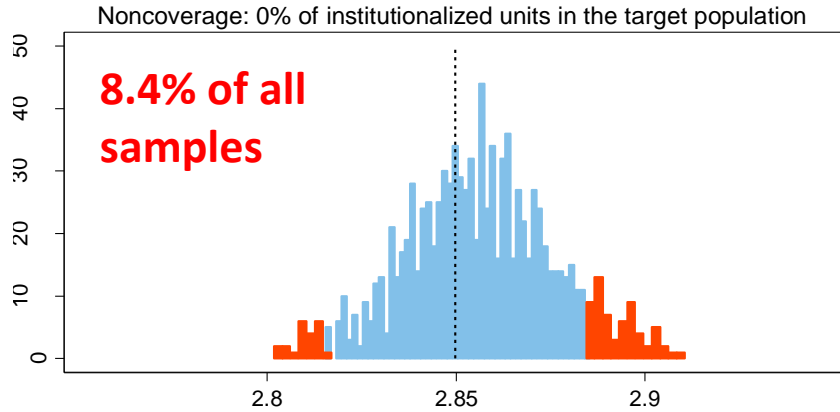


excludes outside values

True value in statistical population = 100%

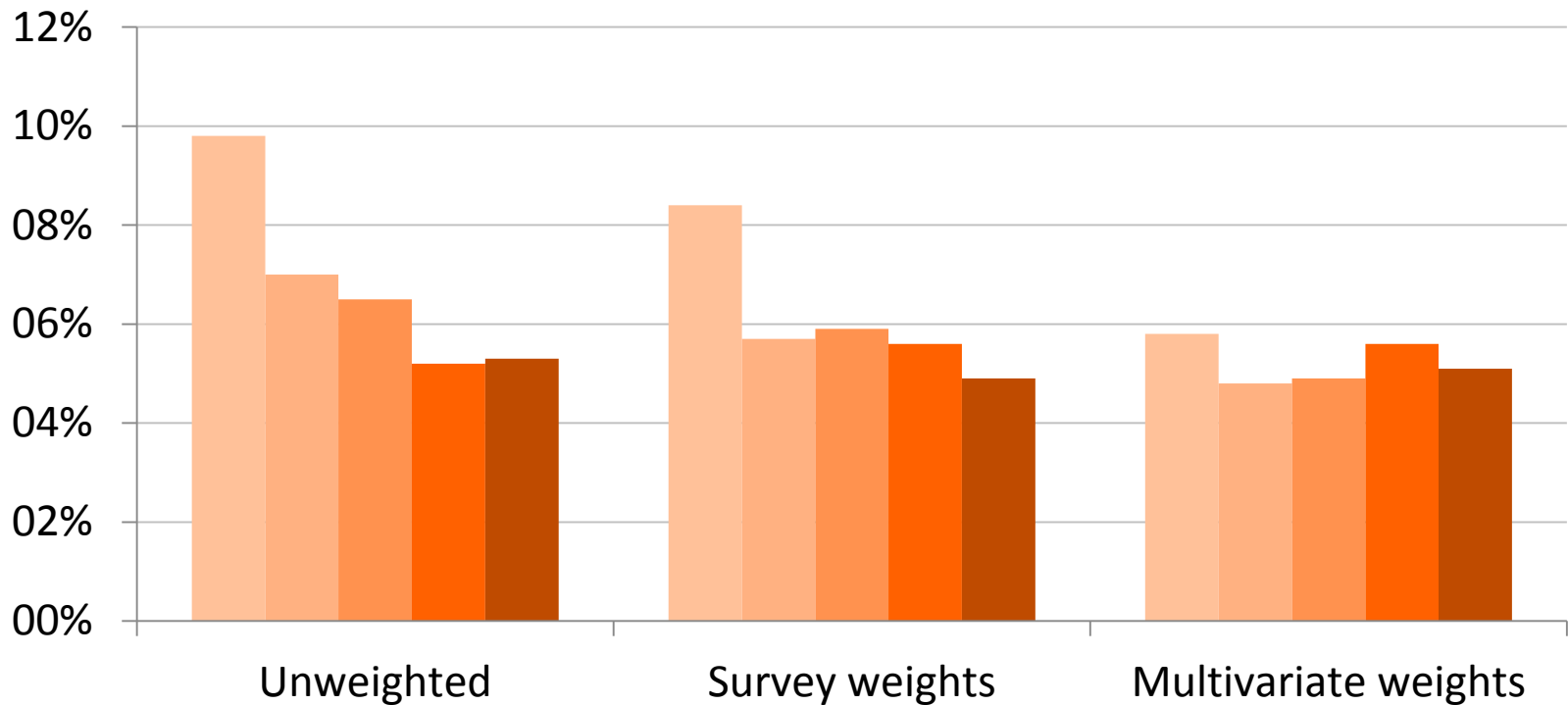
N of each category = 1000 samples of 3000 obs. each

Mean of self-perceived health



N = 1000 samples of 3000 obs. each; survey weights (gender, age)

Mean of health perception: Biased samples



- Noncoverage (0%)
- Undercoverage (25%)
- Undercoverage (50%)
- Undercoverage (75%)
- Equal coverage (100%)

Limitations

- Generalization of our simulation-based results for the universe of social surveys in Europe
- Survey data used for compiling our statistical population
 - Issue of undercoverage of the institutionalized population in some countries in the baseline wave
 - Institutionalization as a shock likely to increase noncontact and refusal of this group
 - Assumption of stability of variables over time (2004 to 2015) could be proved wrong in reality
- Analysis only comprises a small subset of variables and should be extended to further variables

Conclusion

- Noncoverage and undercoverage of the small group of institutionalized residents (2.0%) causes bias in some variables due to their statistical distinctiveness
 - Significant bias in health-related variables (ADL and self-perceived health)
- Weighting the samples for age and gender reduces the bias, but *cannot* eliminate it completely
 - In the ADL variable the bias even remains under control of multiple variables in case of noncoverage and undercoverage
- Undercoverage already improves the estimates compared to noncoverage, especially if weights are used!

Contact

Jan-Lucas Schanze

GESIS – Leibniz Institute for the Social Sciences
Survey Design and Methodology (SDM)

P.O. Box 12 21 55
68072 Mannheim (GER)

phone: 0049 (0)621 1246-405

mail: jan-lucas.schanze@gesis.org

References

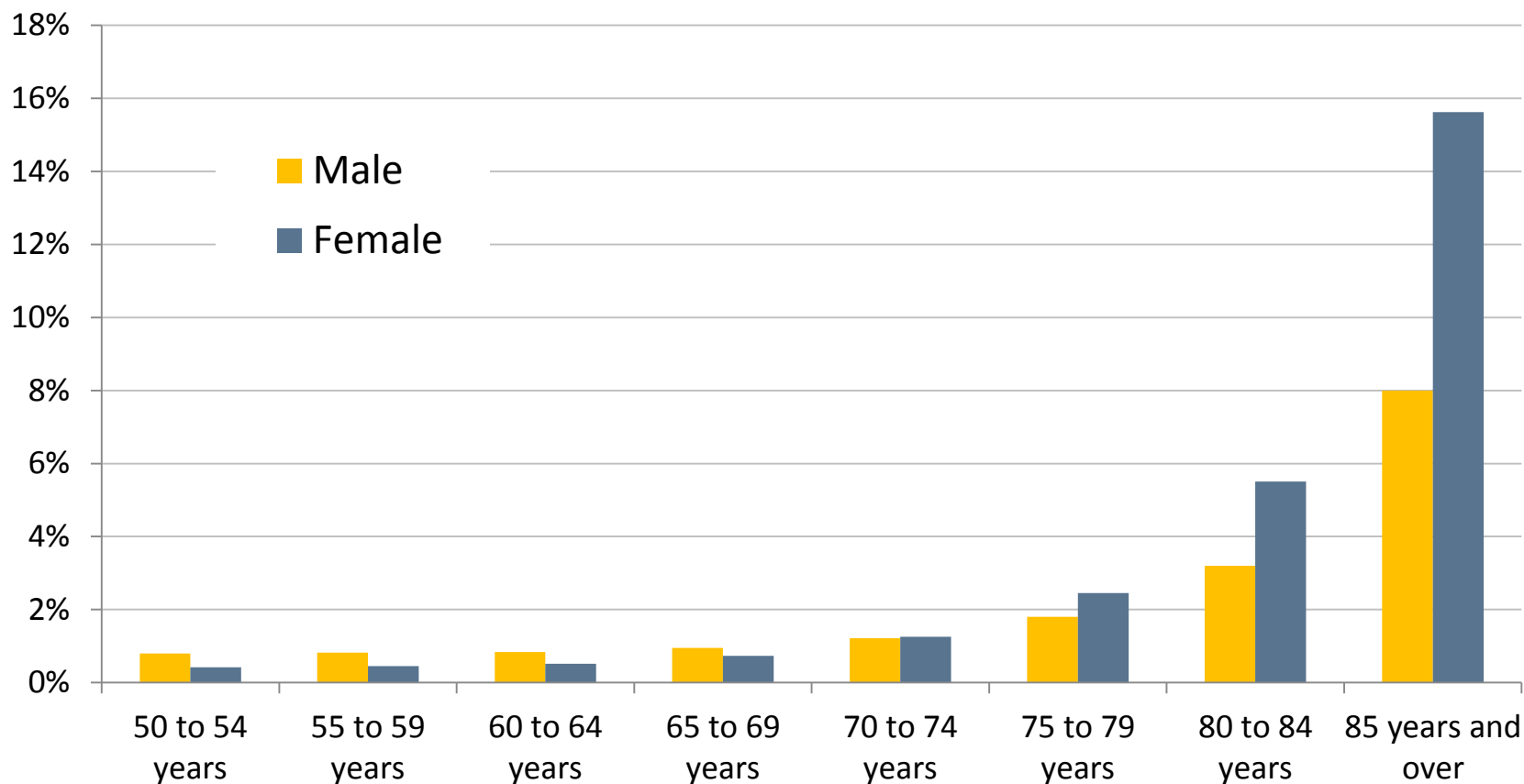
- Agüero-Torres, H. et al.** 2001. "Institutionalization in the elderly: The role of chronic diseases and dementia. Cross-sectional and longitudinal data from a population-based study" In: *Journal of Clinical Epidemiology*, Vol. 54 (8): pp.795–801.
- Asakawa, K. et al.** 2009. "Do the determinants of health differ between people living in the community and in institutions?" In: *Social Science and Medicine*, Vol. 69 (3): pp.345–353.
- Böckerman, P. et al.** 2012. "Institutionalisation and subjective wellbeing for old-age individuals: is life really miserable in care homes?" In: *Ageing and Society*, Vol. 32 (07): pp.1176–119.
- Börsch-Supan, A. et al.** 2013. "Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE)" In: *International Journal of Epidemiology*, Vol. 42 (4): pp.992–1001.
- Castora-Binkley, M. et al.** 2014. "Predictors of long-term nursing home placement under competing risk: Evidence from the health and retirement study". In: *Journal of the American Geriatrics Society*, Vol. 62 (5): pp.913–918.
- De Luca, G. et al.** 2015. "Sample design and weighting strategies in SHARE Wave 5". In: *SHARE Wave 5: Innovations & Methodology* MEA, ed. by Börsch-Supan and Malter, Munich: pp.75–84.
- Einio, E. et al.** 2012. "Gender Differences in Care Home Use Among Older Finns and Belgians" In: *Population*, Vol. 67 (1): pp.75–101.
- Eurostat.** 2015. "People in the EU - Who are we and how do we live" European Union: Luxembourg.
- Eurostat.** 2016. "European Statistical System - Census Hub" Online: <https://ec.europa.eu/CensusHub2>
- Gaugler, J. et al.** 2007. "Predicting nursing home admission in the U.S: a meta-analysis" In: *BMC Geriatrics*, Vol. 7 (13): pp.1–14.
- Groom, G. et al.** 2009. "Communal Establishments Survey - Findings of the Pilot Stage: Summary Report"
- Groves, R. et al.** 2004. "Survey Methodology" New York: Wiley.

References (cont.)

- Hancock, R. et al.** 2002. “The effect of older people’s economic resources on care home entry under the United Kingdom’s long-term care financing system”. In: *Journal of Gerontology: Social Sciences*, Vol. 57 (5): pp.285-293.
- Klevmarken, A. et al.** 2005. “The SHARE Sampling Procedures and Calibrated Design Weights”. In: *The Survey of Health, Ageing and Retirement in Europe – Methodology*, ed. by Börsch-Supan and Jürges, Mannheim: pp.5-28.
- Laferrère, A. et al.** 2013. “Entry into institutional care: predictors and alternatives” In: Börsch-Supan, A. et al. (eds.) *Active ageing and solidarity between generations in Europe*: pp.253-264.
- Luppa, M. et al.** 2008. “Prediction of institutionalisation in dementia: A systematic review”. In: *Dementia and Geriatric Cognitive Disorders*, Vol. 26 (1): pp.65–78.
- Lugtig, P.** 2014. “Panel Attrition: Separating Stayers, Fast Attriters, Gradual Attriters, and Lurkers” In: *Sociological Methods & Research*, Vol. 43 (4): pp.699-723.
- Luppa, M. et al.** 2010. “Prediction of institutionalization in the elderly. A systematic review” In: *Age and Ageing*, Vol. 39: pp.31-38.
- Nöel-Miller, C.** 2010. “Spousal Loss, Children, and the Risk of Nursing Home Admission” In: *Journal of Gerontology: Social Sciences*, Vol. 65 (3): pp.370-380.
- ONS, Social Survey Division.** 2002. “Survey of people living in communal establishments (pilot)” Technical report.
- Solé-Auró, A & Crimmins, E.** 2014. “Who cares? A comparison of informal and formal care provision in Spain, England and the USA” In: *Ageing Societies*, Vol. 34 (3): pp.495–517.
- Wolff, F.** 2013. “Well-Being of Elderly People Living in Nursing Homes: The Benefits of Making Friends” In: *Kyklos*, Vol. 66 (1): pp.153–171.

Appendix

Institutionalized population in EU-26*



Source: Eurostat 2016, CensusHub

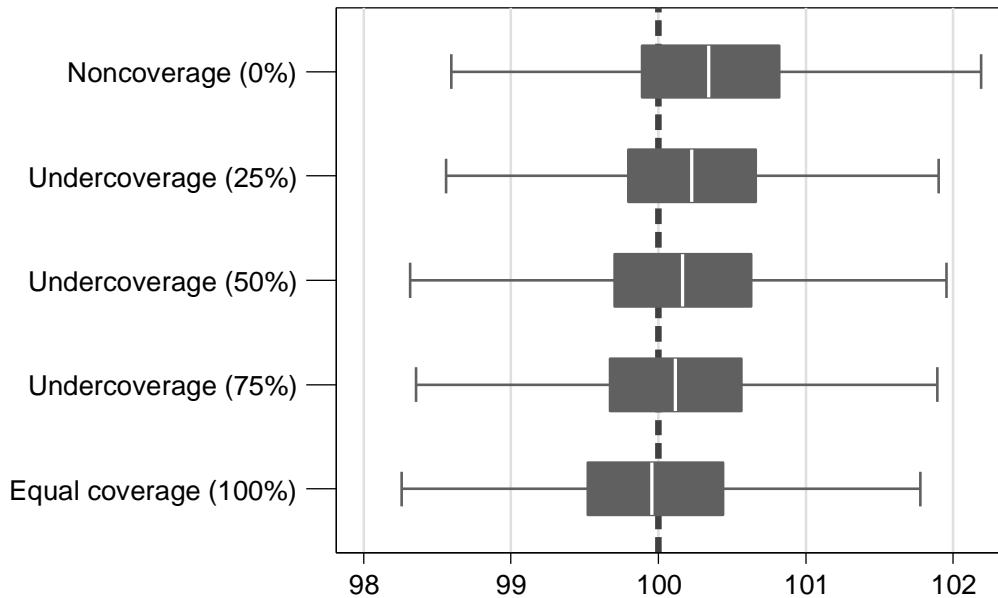
* Lithuania and Bulgaria missing

Variables

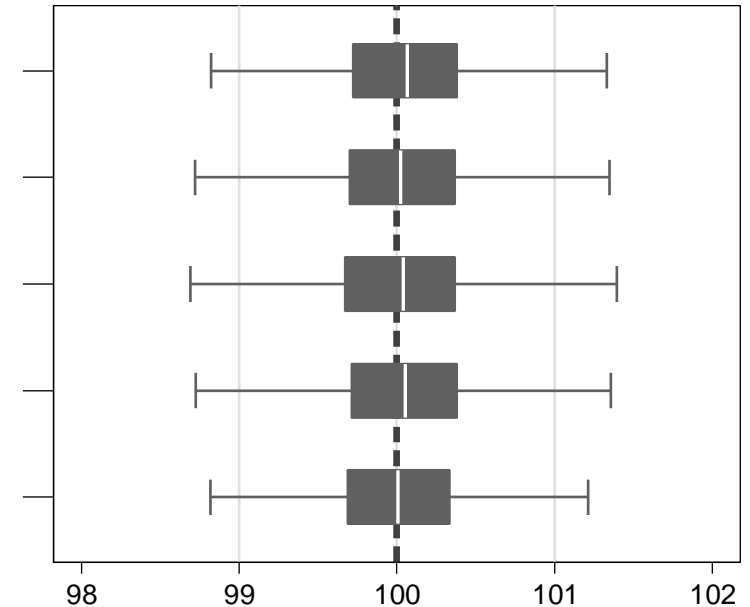
- Institutionalization of respondents
 - Interviewer observations (Living in a nursing home)
 - Self-reports of respondents
 - Living permanently in a nursing home
 - Information about the housing type
- Additive index of limitations with activities in daily living (ADL)
 - Count variable of 9 dichotomous variables (scale 0 to 9)
- Self-perceived health (U.S. version)
 - Ordered 5-level scale

Deviation of means in self-perceived health

1. Unweighted



3. Multivariate weights



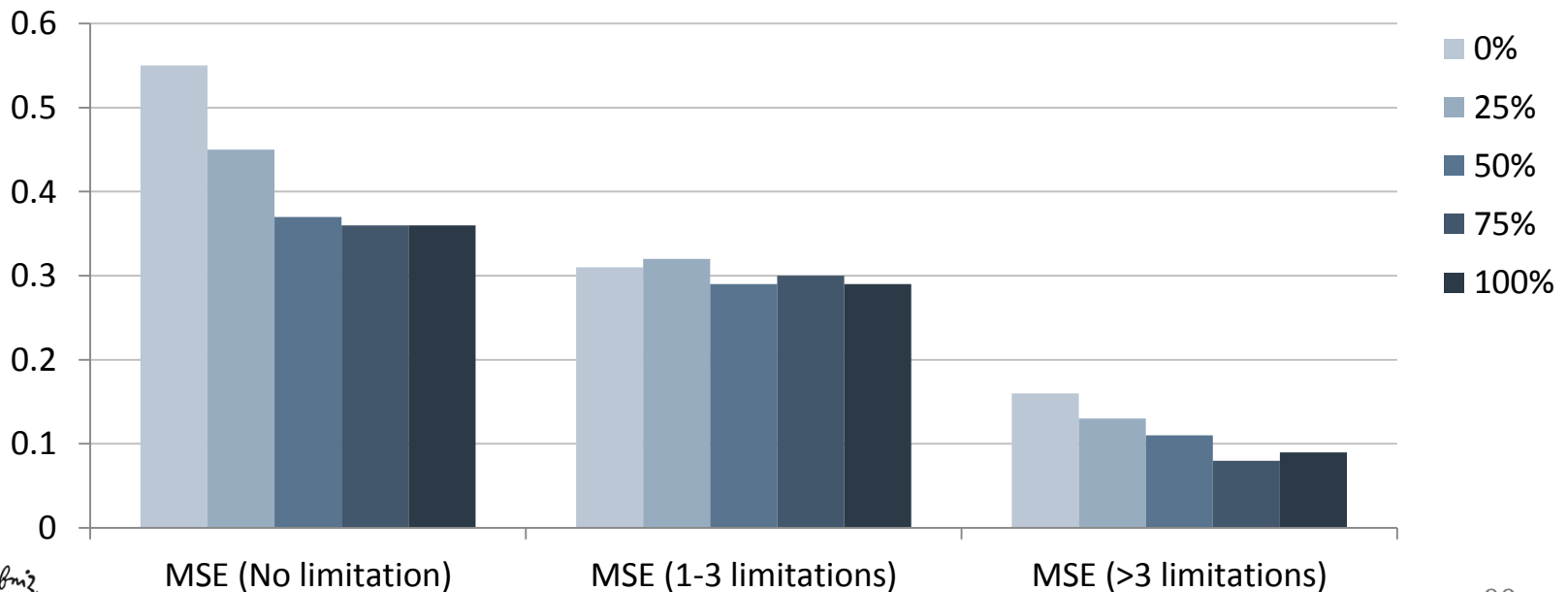
excludes outside values

True value in statistical population = 100%

N of each category = 1000 samples of 3000 obs. each

Mean square error within categories (1)

Limitations in ADL (%, weighted with survey weights)	Inclusion of institutionalized units					True value
	0%	25%	50%	75%	100%	
No limitation	86.94	86.81	86.67	86.57	86.46	86.46
1-3 limitations	10.47	10.54	10.58	10.60	10.67	10.66
>3 limitations	2.59	2.66	2.74	2.82	2.87	2.88



Mean square error within categories (2)

Self-perceived health (%, weighted with survey weights)	Inclusion of institutionalized units					True value
	0%	25%	50%	75%	100%	
Poor	10.16	10.21	10.22	10.24	10.26	10.28
Fair	27.71	27.76	27.76	27.78	27.9	27.85
Good	36.14	36.11	36.11	36.13	36.05	36.07
Very good	18.37	18.32	18.32	18.26	18.24	18.24
Excellent	7.62	7.60	7.58	7.60	7.55	7.57

