

Indicators for representativeness of survey response

UNECE workshop on Measuring poverty in pandemic times

Barry Schouten (Statistics Netherlands and Utrecht University)

March 26, 2021



Academy

Survey nonresponse

- Nonresponse occurs in every survey;
- Nonresponse is result of lack of contact, not being able physically or in terms of language, lack of time and refusal;
- Nonresponse causes estimates to be biased;
- Nonresponse leads to smaller samples;
- Nonresponse problem seems to increase, i.e. more effort/budget is needed to get the same response rates;

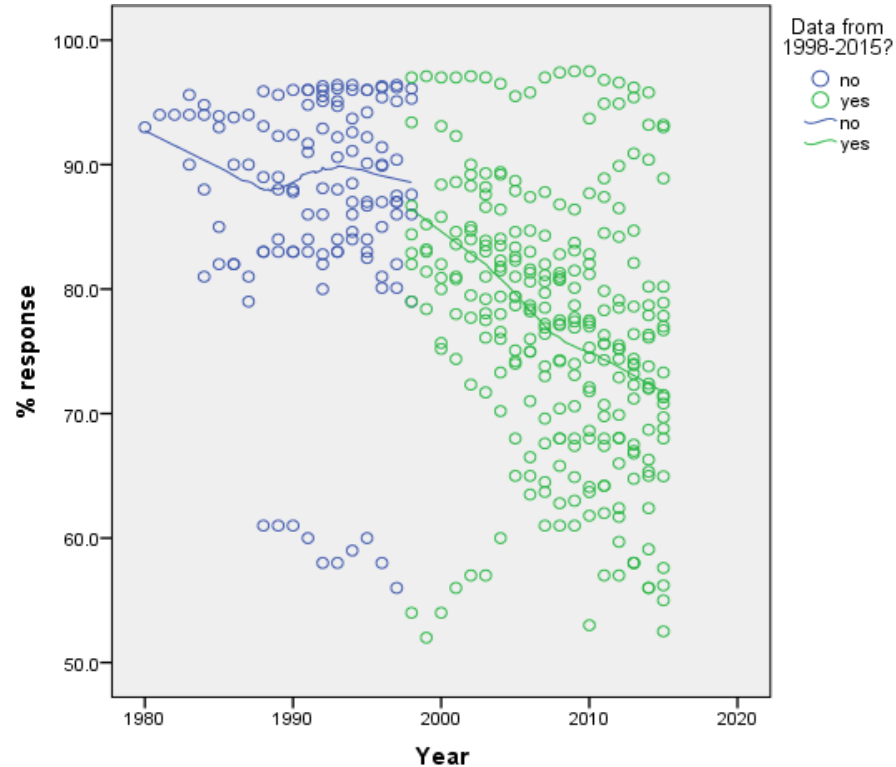
Options:

- Prevent it from happening (reduction);
- Adjust afterwards through a statistical model (adjustment);
- Hybrid approach: adjust by design (adaptive survey designs)



Trends in response rates

Labour Force Surveys response rates 1980 – 2015 in various countries



Nonresponse analysis

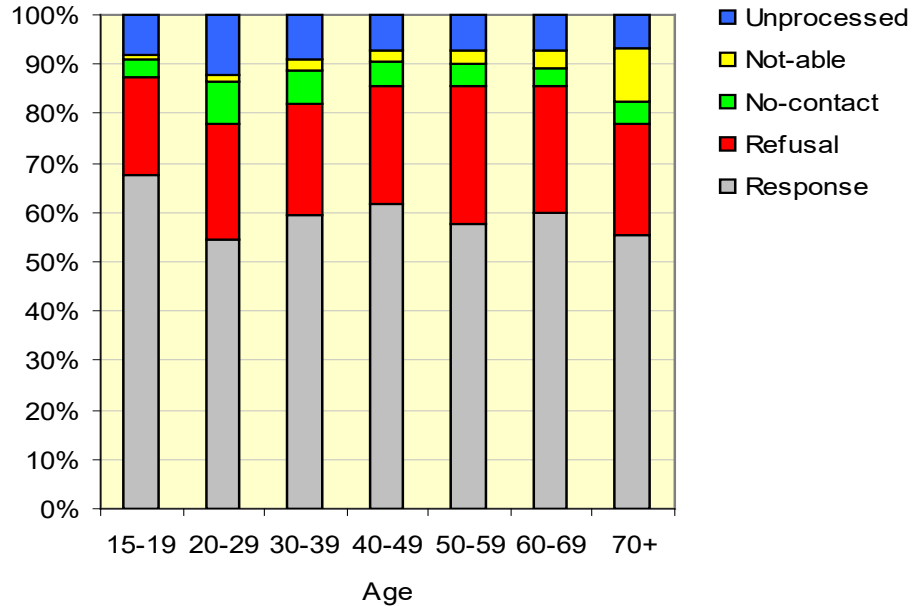
Key component is auxiliary information:

- Information available in the sampling frame
- Information that can be linked from administrative sources
- Information for which population totals are available
- Information that is collected by interviewers for the whole sample



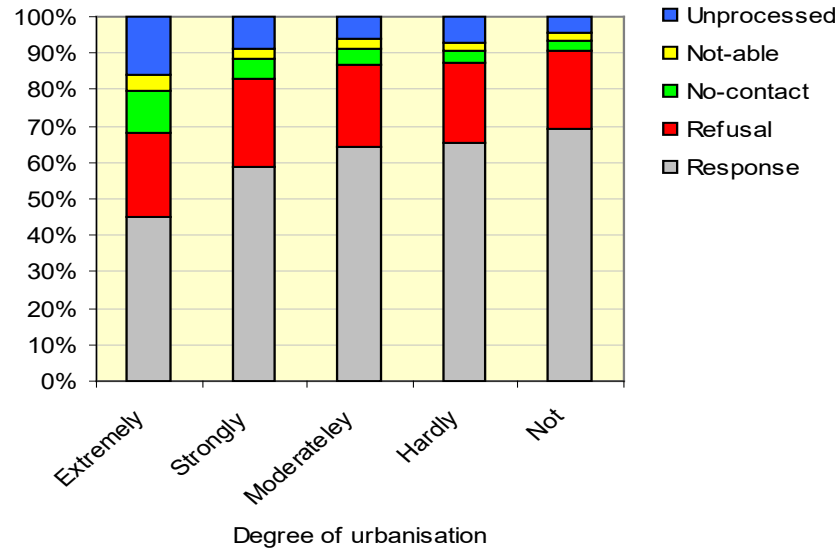
Global nonresponse analysis

- Age



Global nonresponse analysis

- Degree of urbanisation



R-indicators

- R-indicator is based on variation in individual response propensities

$$R(\rho) = 1 - 2S_\rho$$

- Two types:
 - Sample-based: Response is compared to sample totals
 - Population-based: Response is compared to population totals
- At www.risq-project.eu code in SAS and R plus manual and test data set



R-indicators

- Nonresponse bias of response mean

$$|B(\bar{y}_R)| = \frac{|R_{\rho Y}| S_{\rho} S_Y}{\bar{\rho}} \leq \frac{S_{\rho} S_Y}{\bar{\rho}} = \frac{(1 - R(\rho)) S_Y}{2\bar{\rho}}$$

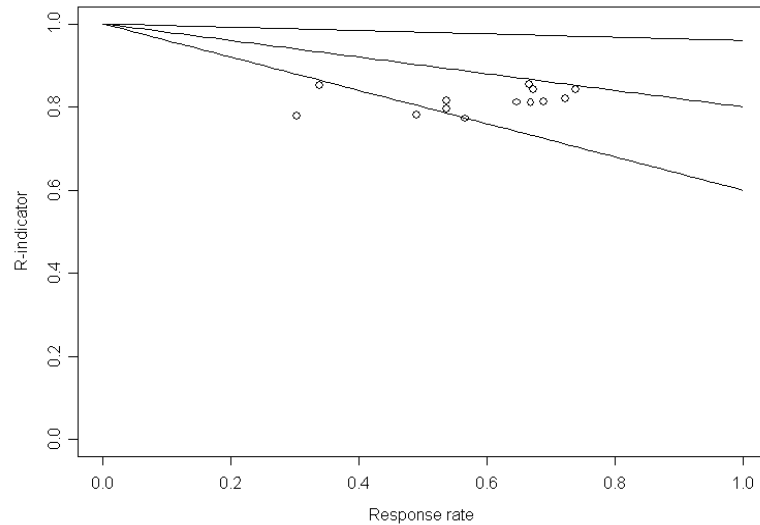
- Bounding R-indicators: response-representativity plots

$$\frac{|B(\bar{y}_R)|}{S_Y} < \gamma \longrightarrow R(\rho) > 1 - 2\bar{\rho}\gamma$$



R-indicators

- Examples of response rates and R-indicators (including three curves $\gamma = 2\%, 10\%, 20\%$)



Example 1 – Various ESS surveys

X = gender, age, urbanization

	Sample size	Response rate	R-indicator
Health Survey 2005 (Holland)	15,411	67.3%	0.832
ESS 2006 (Belgium)	2,927	61.4%	0.807
ESS 2006 (Norway)	2,673	65.6%	0.762
Level of Living 2004 (Norway)	4,837	69.1%	0.872
LFS Quarter 3 – 2007 (Slovenia)	2,219	70.1%	0.854
LFS Quarter 4 – 2007(slovenia)	2,215	69.3%	0.807



Example 2 - Survey on Informal Economy

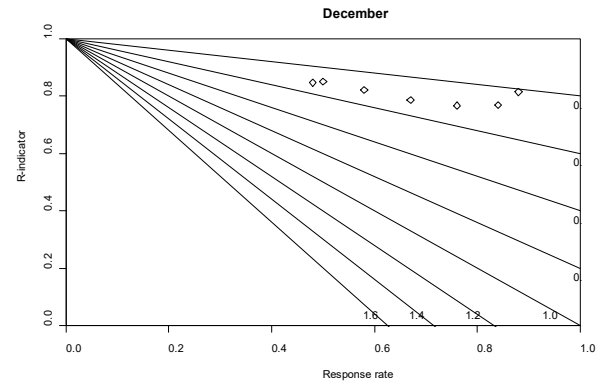
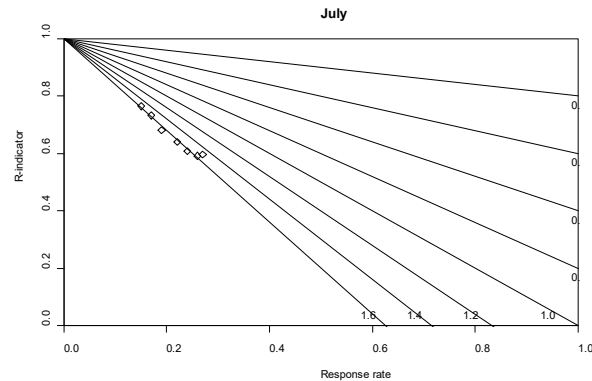
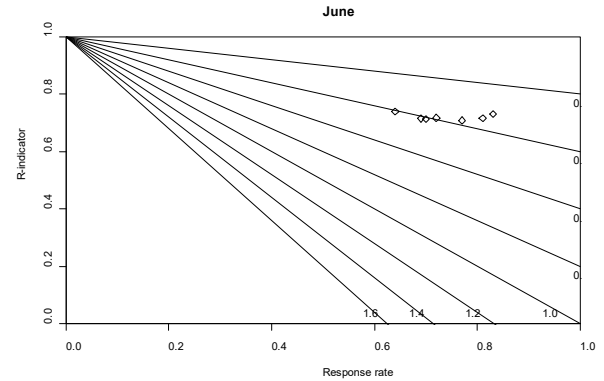
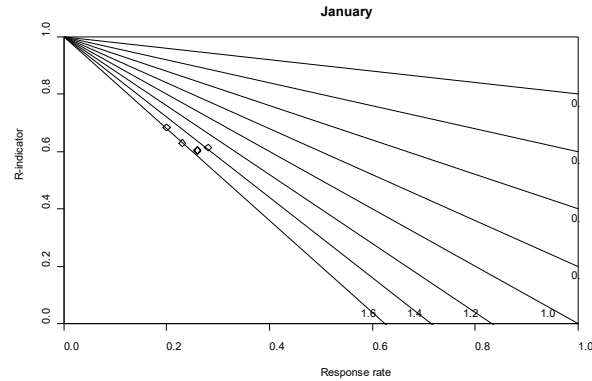
X = age, house value, ethnicity, type of household, employment, urban

Response group	Response rate	Representativity measure R	Confidence interval	CV
Face-to-face	56.7%	77.8%	74.4% - 81.3%	0.102
Web/paper	33.9%	86.3%	83.1% - 89.4%	0.112
Web/paper + phone	49.0%	79.3%	75.6% - 83.0%	0.113



Example 3 - Business survey

$X = wages(t), NACE, VAT(t-12) \times size$



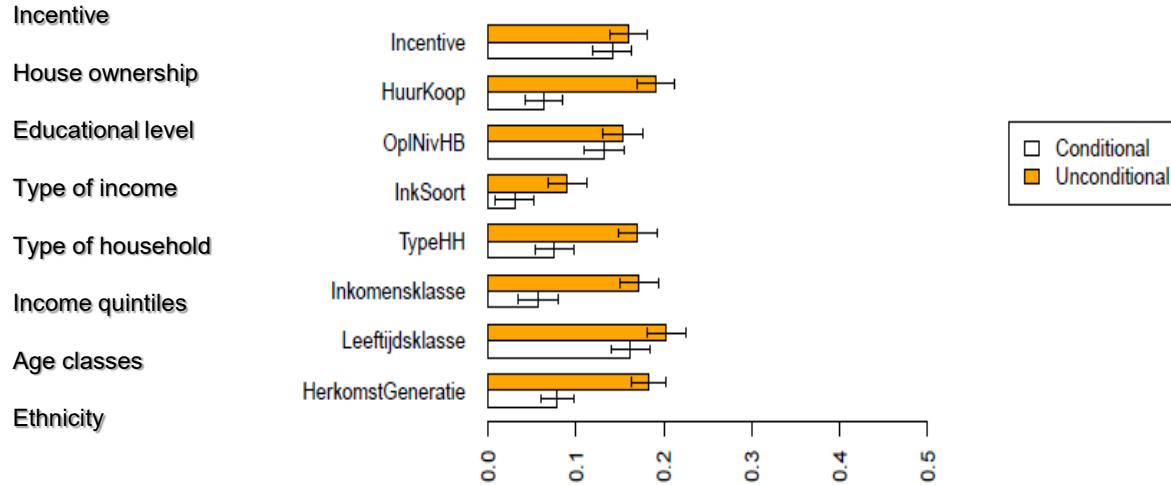
Partial R-indicators

- Partial R-indicators decompose R-indicator based on the impact of single variables
total variance = between variance + within variance
- Unconditional partial R-indicator for a single variable Z: *the between variance of response propensities*
- Conditional partial R-indicator for a single variable Z given X: *the within variation in response propensities given a stratification on X*
- Both type of indicators should ideally be close to 0 and allow for monitoring of data collection and resource allocation



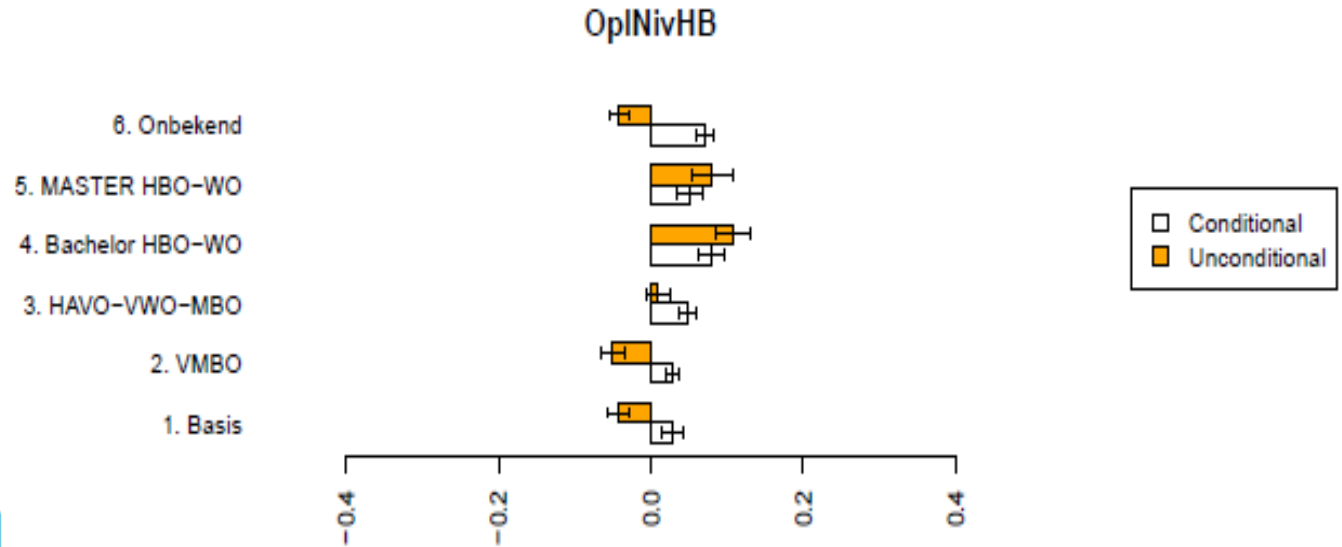
Example EU-SILC

variable-level partial CV



Example – EU-SILC

Partial R-indicators at the category level for educational level



Summary

- Indicators have been developed to monitor surveys during data collection and in time.
- Indicators can be used to adapt and tailor fieldwork strategies
- Key ingredient are auxiliary variables

