

# Indicators for the Representativity of Survey Response

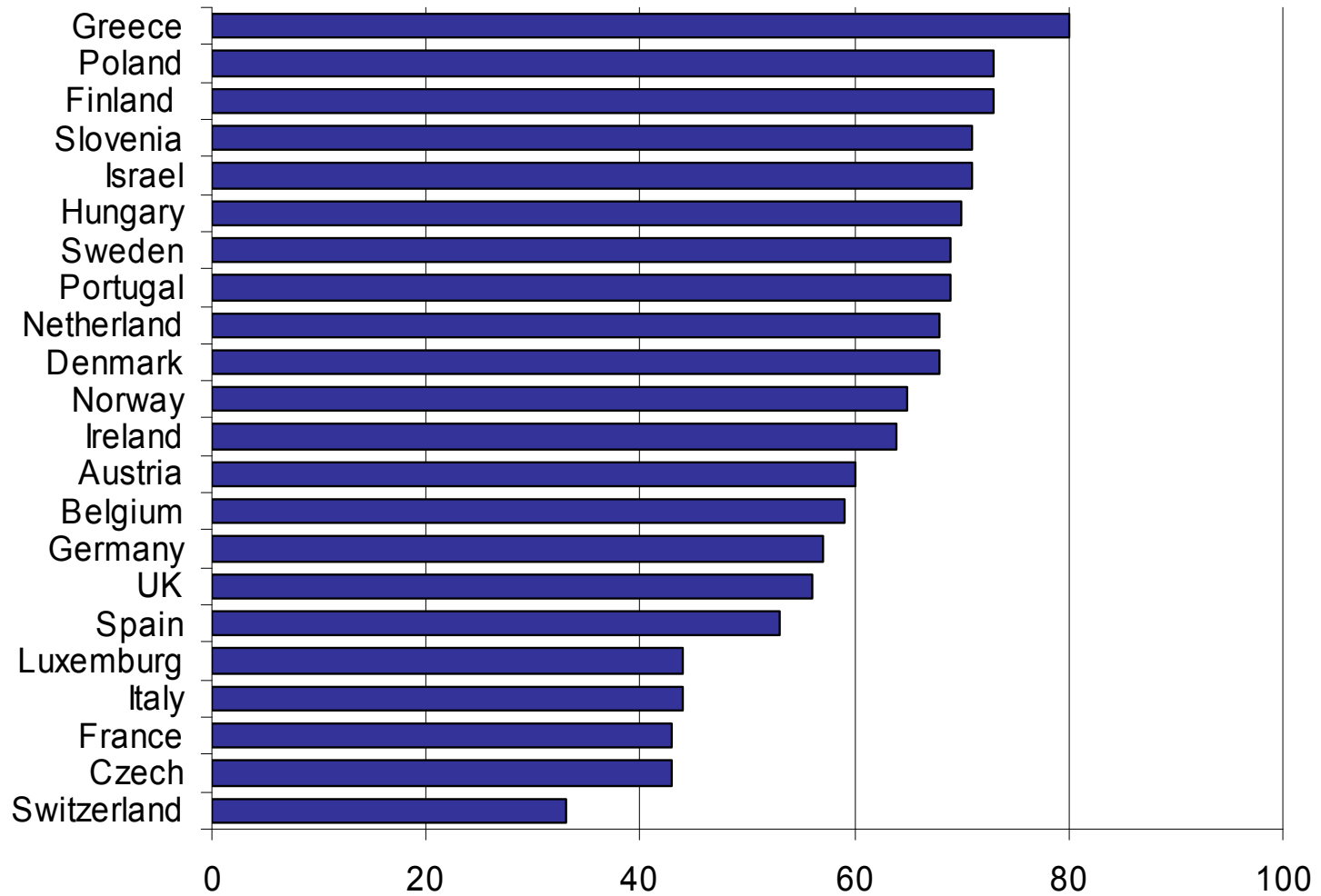
Jelke Bethlehem, Fannie Cobben &  
Barry Schouten

Statistics Netherlands  
Methodology Department



# Is the response rate a good quality indicator?

Response rates in the first round of the European Social Survey



## Is more response always better?

- The accuracy of survey estimates is determined by the precision (variance) and the bias of estimators.
- A higher response rate is only better if the bias is smaller. This is not always the case!

### Integrated Survey on Household Living Conditions 1998

Response	Response after 1 month	Response after 2 months	Complete sample
% social allowance	10.5 %	10.4 %	12.1 %
% non-natives	12.9 %	12.5 %	15.0 %

The composition of the sample deteriorated in month 2.



# Is more response always better?

## Other examples

- The composition of the response deteriorates after a call-back survey among non-respondents.
- The composition of the response may deteriorate if incentives are used to increase response.



## The concept of representativity

- A better quality indicators should reflect how well the composition of the survey response reflects the population (or complete sample).
- These indicators are based on the concept of representativity. Therefore they are called **R-indicators** (short for: **Representativity Indicators**).
- Representativity is not well-defined. See the nine definitions by Kruskal & Mosteller (1979).
- Here, representativity is defined as the absence of selective forces.
- Every element  $k$  in the population is assumed to have an (unknown) probability  $\rho_k$  of responding when selected in the sample.



## The concept of representativity (continued)

### Strong representativity

- The response to a survey is **strongly representative** if all response probabilities are equal.
- This implies Missing Completely at Random (MCAR).
- Cannot be established in practice.

### Weak representativity

- The response to a survey is **weakly representative** with respect to a variable  $X$  if the average response probability is the same for each category of  $X$ .
- Respondents cannot be distinguished from non-respondents just using the values of  $X$ .
- Hopefully, the sample is weakly representative with respect to many auxiliary variables.



## An example of an R-indicator

- The bias (due to non-response) of the response mean is equal to

$$\frac{Cor(Y, \rho) \times S(Y) \times S(\rho)}{\bar{\rho}}$$

- $Y$  is the survey variable and  $\rho$  is the response probability.
- $Cor(Y, \rho)$  is the correlation between  $Y$  and  $\rho$ .
- $S(Y)$  and  $S(\rho)$  are the standard deviations of  $Y$  and  $\rho$ .
- The bias vanishes if all response probabilities are equal. Then

$$S(\rho) = \frac{1}{N-1} \sum_{k=1}^N (\rho_k - \bar{\rho})^2 = 0.$$



## An example of an R-indicator (continued)

- Definition of an indicator:

$$R(\rho) = 1 - 2 \times S(\rho)$$

- $R(\rho) = 1$ : All response probabilities are equal.
- $R(\rho) = 0$ : Maximum possible deviation from representativity.
- Definition does not involve target variables of the survey.

### Applications

- Compare surveys of over space and time.
- Monitor data collection process.
- Processing register data.
- Household and business surveys.





## An example of an R-indicator (continued)

### Case 1: response probabilities are known

- $R(\rho)$  can be computed.
- $R(\rho) = 1$  implies that the response is strongly representative.

### Case 2: Response probabilities are unknown

- Response probabilities have to be estimated
- Fit logit (or other) model for response probabilities.
- Required: auxiliary variables (response and non-response)
- $R(\rho)$  can be estimated.
- Comes down to testing weak representativity.



## Application 1

- Survey: Dutch Labour Force Survey.
- Sample of non-respondents re-approached (call-back) with complete questionnaire.
- Sample of non-respondents re-approached with small questionnaire (basic question approach).

Response	Rate	$R(\rho)$
LFS	62 %	0.79
LFS + Basic question approach	76 %	0.77
LFS + Call-back approach	77 %	0.85

- The composition of the response improves after the call-back approach.
- The composition of the response does not improve after the basic question approach.



## Application 2

- Survey: Dutch Labour Force Survey.
- Experiment: Can response rates be improved by using incentives
- Three groups: no incentives, 5 stamps, 10 stamps

Response	Rate	$R(\rho)$
No stamps	67 %	0.86
5 stamps	72 %	0.82
10 stamps	74 %	0.84

- Incentives do not improve the composition of the response.



## Research issues

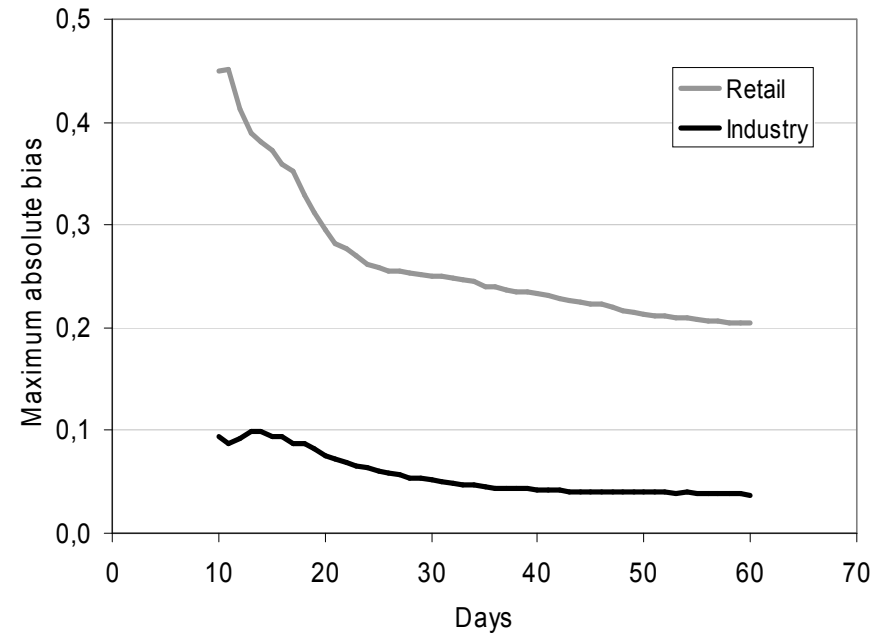
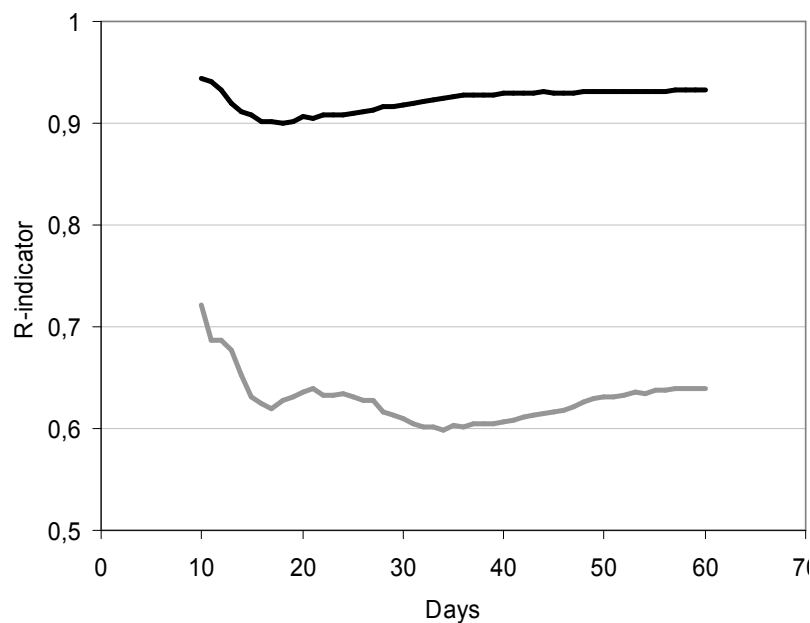
- R-indicator is based on variance of response probabilities. Other possibility: see Särndal & Lundström (2008)
- Dependency on sample size.
- Dependency on auxiliary variables.
- Estimation of response probabilities if only population distribution of auxiliary variables is available.
- Development of partial R-indicators to identify groups at risk.
- Use of paradata (fieldwork data) in response probability models.
- Relationship between R-indicator and maximum possible bias:

$$B(\bar{y}) \leq \frac{(1 - R(\rho)) \times S(y)}{2\bar{\rho}}$$



## Another example

- Use of the R-indicator to monitor fieldwork of business surveys
- The representativity of the short term statistics for industry and retail trade by number of fieldwork days.





## The RISQ Project

### Project

- Further development of R-indicators
- Financed by the 7th Framework Programme of the EU.

### Project partners:

- Statistics Netherlands (Netherlands, co-ordinator)
- University of Southampton (UK)
- Statistics Norway (Norway)
- University of Leuven (Belgium)
- Statistical Office of the Republic of Slovenia (Slovenia)

### Website

- [www.r-indicator.eu](http://www.r-indicator.eu)



Statistics Netherlands

