

The future of surveys for official statistics

Jelke Bethlehem

Statistics Netherlands
Methodology Department



"Opinion poll!
Answer!"

Overview

Some history of survey research

- The ever changing landscape of survey research.

Current challenges in official statistics

- More statistics with less money.

Future trends

- The conquest of the web
- The quest for representativity
- From fixed to flexible survey design

Directions for Blaise

- How to prepare Blaise for the future?



Overview

Some history of surveys research

- The ever changing landscape of survey research

Current challenges for official statistics

- More statistics with less money

Future trends

- The conquest of the web
- The quest for representativity
- From fixed to flexible survey design

Directions for Blaise

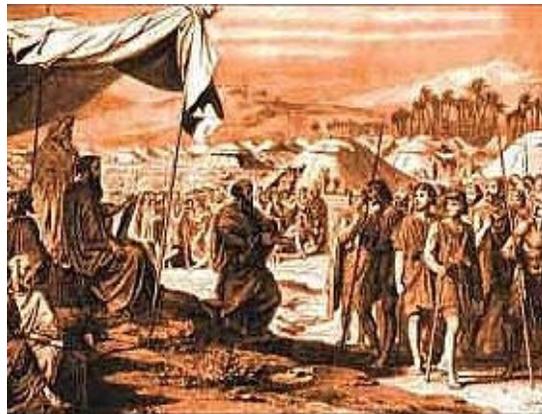
- How to prepare Blaise for the future?



Surveys through the ages

A long history

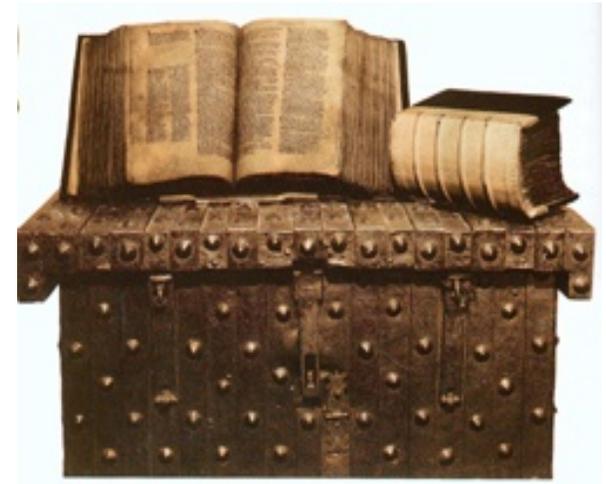
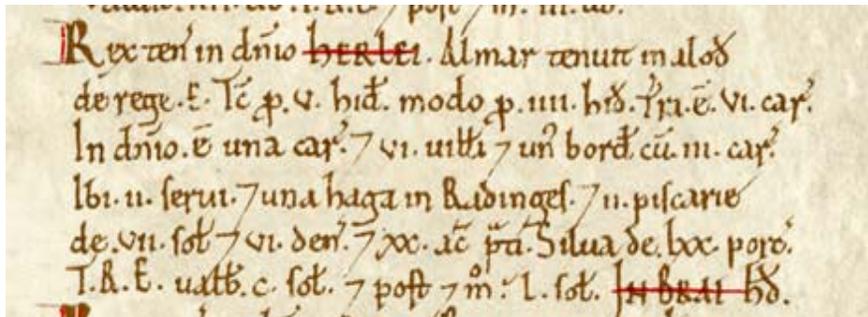
- There have always been statistical overviews.
- Complete enumeration, no samples.
- China and Egypt, 1000 BC: to determine tax and military strength.
- Roman Empire: counts of people and properties, for tax and military obligations.



Surveys through the ages

The Domesday Book (1086)

- By order of William the Conqueror.
- Data about 13.000 manors and villages.
- 10,000 facts per county.
- Data about ownership, value, +free man, slaves, woodland, pasture, meadow, mills, fishponds, ...



Surveys through the ages

The Quipucamayoc

- Statistician in the Inca empire (1200-1500).
- Counts of people, houses, llamas and young men.
- Recorded on a quipu.
- Knots in coloured ropes.
- Decimal system.
- RAPI: Rope Assisted Personal Interviewing.



The first modern censuses

Jean Talon (1666)

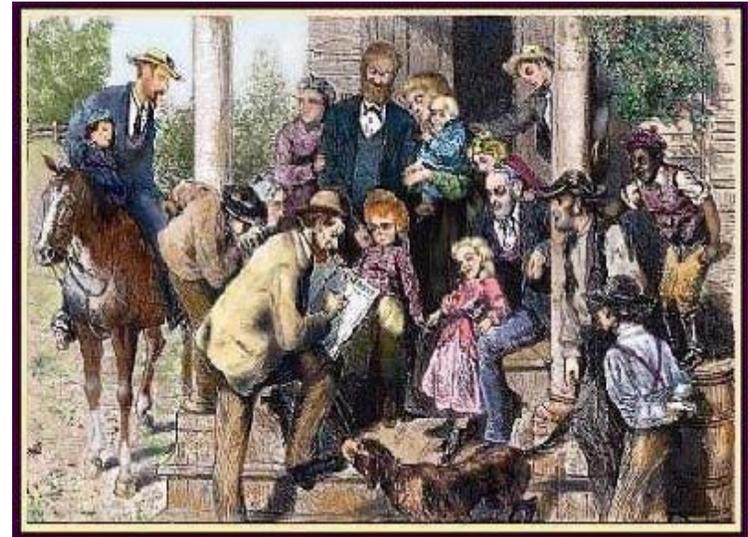
- Count of the people in New France (Canada)
- $N = 3215$

Scandinavia

- Sweden (1748): counts of men that could be enlisted, members of Lutheran church.
- Denmark (1769)

Netherlands (1795)

- Batavian Republic
- New election districts



The period until 1895

No scientifically based sampling

- It is not proper to replace people by calculations.

Partial investigations

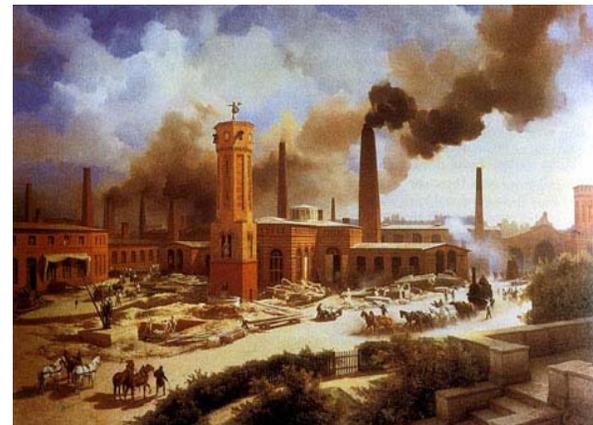
- Data on only part of the population.
- Selection mechanism unclear.

Monograph studies

- Investigation of only 'typical' representatives of the population.

The dawn of a new era

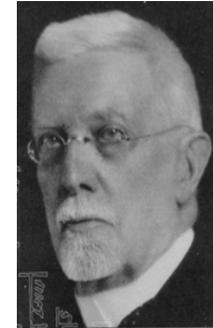
- Centralised government
- Industrialisation



The rise of sample surveys

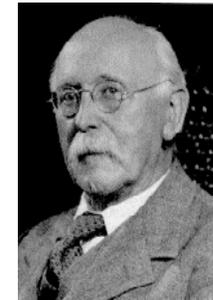
Ander Kiaer (1895)

- Representative Method.
- Miniature of population.
- Unable to compute accuracy.



Arthur Bowley (1906)

- Draw random sample.
- Probability theory can be applied.
- Computation of variance.



Jerzy Neyman (1934)

- Introduces confidence interval.
- Purposive sampling does not work.



The increasing role of the computer

Tabulation

- Hollerith machine (1890)

Analysis, editing

- Mainframe computers (1970)

Computer assisted interviewing

- CATI (1970s)
- CADI / CAPI (1980s)
- CASI / CASAQ (1980s)

The Internet

- Web surveys / web panels (1990s)
- Back to purposive sampling?



The challenges of official statistics

Costs

- Budget cuts.
- Survey costs must be reduced.

Response burden

- Companies complain about administrative burden.
- Less questionnaire forms.

More data

- Demand for more data.
- More regional statistics.

Solutions

- Smaller samples?
- Use of register data?
- Cheaper surveys: web surveys



The conquest of the web

Advantages

- Easy: simple access to large group of potential respondents
- Cheap: No interviewers, no printing, no mailing.
- Fast: surveys can be launched very quickly.
- Attractive: Use of sound, pictures, animation, movies, ...

Methodological disadvantages

- Under-coverage.
- Self-selection
- No interviewers

Question

- Can web surveys be used for official statistics?



Under-coverage

Internet access by households in Europe (2007)

Source: Eurostat

Country	Internet access	broadband
Netherlands	83%	74%
Sweden	79%	67%
Denmark	78%	70%
...		
Greece	25%	7%
Romania	22%	8%
Bulgaria	19%	15%
EU	54%	42%

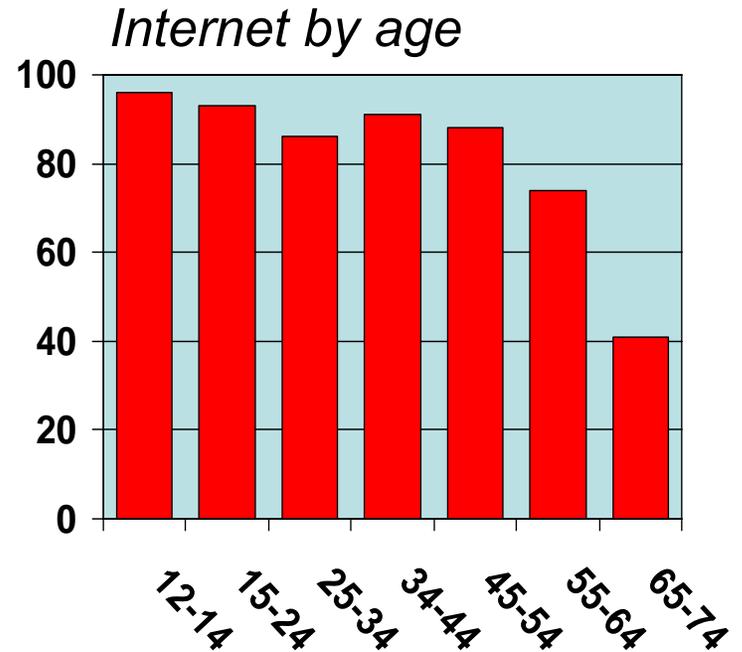
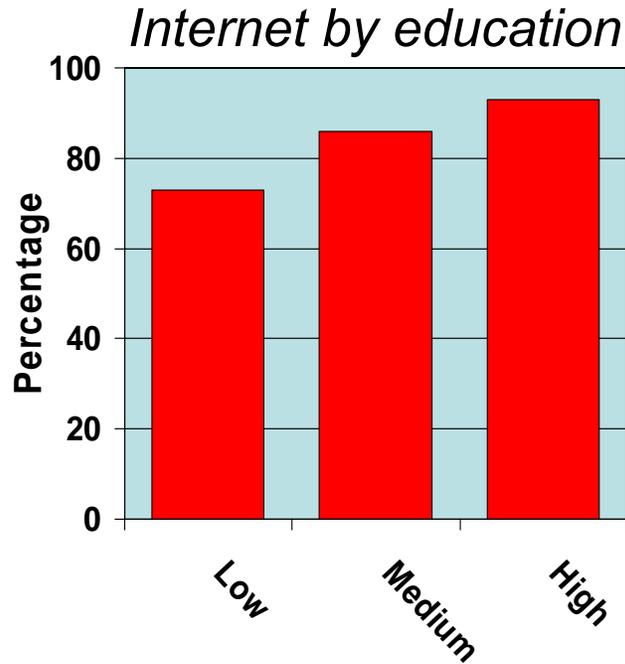
Note: Percentage of households with a listed phone number in The Netherlands is 67%.



Under-coverage

Under-represented groups

- Elderly
- Low-educated
- Ethnic minority groups



Effects of under-coverage

Under-coverage can lead to biased estimates

$$B(\bar{y}) = E(\bar{y}) - \bar{Y} = \bar{Y}_I - \bar{Y} = \frac{N_{NI}}{N} (\bar{Y}_I - \bar{Y}_{NI})$$

Bias of estimates is the product of two factors:

- Relative size of the group of people without Internet
- Difference (on average) between people with internet and people without internet

Developments

- First factor decreases as Internet access increases.
- Second factor may increase as remaining group without internet may be more and more different.



Self-selection

In theory ...

- The sample must be selected from a sampling frame using a probability sample with known selection probabilities.

In practice ...

- Self-selection of respondents: only those people respond who happen to visit the website and decide to participate.
- Selection probabilities are unknown. Therefore it is impossible to construct unbiased estimates.
- Specific groups can even attempt to influence the composition of the sample.



Self-selection - example

Parliamentary election 2006, opinion polls

Seats in parliament (total =150)

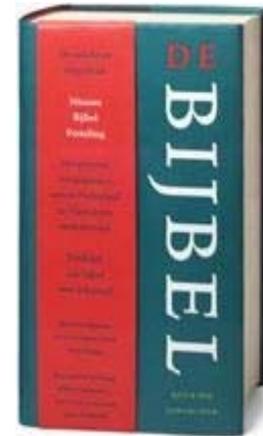
Party	Election result	Politieke Barometer	Peil.nl	De Stemming	DPES 2006
Sample size		1,000	2,500	2,000	2,600
CDA	41	41	42	41	41
PvdA	33	37	38	31	32
VVD	22	23	22	21	22
SP	25	23	23	32	26
GL	7	7	8	5	7
D66	3	3	2	1	3
CU	6	6	6	8	6
SGP	2	2	2	1	2
Animals	2	2	1	2	2
Wilders	9	4	5	6	8
Other	-	2	1	2	1
MAD		1.27	1.45	2.00	0.36



Self-selection - example

2005 Book of the Year award

- Web survey to selected one of the nominated books or suggest another book.
- 90,000 people participated.
- The winner was a non-nominated book: New Inter-confessional Bible Translation (72% of votes).
- Campaign by Bible societies, Christian radio/tv station, and Christian newspaper.



Self-selection

Bias due to self-selection

$$B(\bar{y}) = E(\bar{y}) - \bar{Y} \approx \frac{C_{\rho Y}}{\bar{\rho}} = \frac{R_{\rho Y} S_{\rho} S_Y}{\bar{\rho}}$$

Maximum absolute bias

$$|B(\bar{y})| \leq B_{max} = S_Y \sqrt{\frac{1}{\bar{\rho}} - 1}$$

Example

- CAPI survey, response rate 70%: $B_{max} = 0.65 S_Y$
- Web survey ($n = 170,000$) from Dutch population ($N = 12,800,000$): $B_{max} = 8.61 S_Y$
- The bias of the web survey can be 13 times as large!

Does weighting help?

Weighting techniques

- Post-stratification
- Calibration estimation
- Propensity weighting

Required: auxiliary variables:

- Measured in survey.
- Population distribution, or individual values for non-participants.
- Correlated with survey variables and / or response behaviour.
- Such variables are very often not available



Does weighting help?

Reference survey

- Define your own weighting variables.
- Estimate population distribution in different survey: reference survey.
- Other mode of data collection, e.g. CAPI or CATI.
- No non-response, or ignorable non-response.
- Examples of weighting variables: 'webographics', 'psychographics' or 'lifestyle' variables.

Problems

- Expensive.
- Bias reduced at the cost of a loss of precision.
- Measure problems for attitudinal variables
- Why do a web survey at all?



Mixed-mode data collection

Approaches

- Concurrent approach: best mode for each group. For example: CAPI for the elderly, CAWI for the young.
- Sequential approach (costs): cheapest mode first, for example: CAWI - CATI - CATI.
- Sequential approach (response): best mode first, for example: CAPI - CATI - CAWI - PAPI.
- Respondents select preferred mode themselves. This may not work well in practice.

Problem

- Mode effects: same question may be answered differently in different mode.



Mixed-mode data collection

Mode effects

- Presence of interviewers leads to more socially desirable answers.
- Presence of interviewers leads to acquiescence: increased tendency to agree.
- Interviewers can see to it that respondents understand the question.
- CAWI/PAPI: preference for first answer in list of answers to closed question (primacy effect).
- CATI: preference for last answer in list of answers to closed question (recency effect).
- Treatment of “don’t know”: offer explicitly or not?



Mixed-mode data collection

Reducing mode effects – solution 1

- Different formats for the same question in different modes.
- Question should measure the same thing in each mode: cognitively equivalent.
- Different Blaise instrument for each mode

Reducing mode effects – solution 2

- Dillman's unimode approach: guidelines for question design that reduce mode effects:
- All answer options the same across modes;
- Reduce number of answer options as much as possible;
- Include answer options in question text;
- Randomize answer options;
- Same routing structure in each mode. Difficult for PAPI.



Mixed-mode data collection

Reducing mode effects – Practical solution

- Some questions may be transformed into unimode, for others it may not be possible.

```
Question_A
IF Mode = CAPI THEN
    Question_B1
ELSEIF Mode = CATI THEN
    Question_B2
ELSEIF Mode = CAWI THEN
    Question_B3
ENDIF
Question_C
```

Prerequisite for mixed-mode data collection

- General case management system that covers all modes.
- Cases must have proper mode at proper time.
- Cases may not disappear.
- There may not be duplicate cases



Mixed-mode data collection

Other issues

- CAPI and CATI are interviewer-assisted. Web surveys are not. Effect on quality?
- CAPI and CATI instruments contain extensive error checking. Can this be implemented for web surveys?
Dillman: *error message are the death of web surveys!*
- Giving respondents a choice of mode seems to reduce response rates.

Web surveys in official statistics?

- In the end, we have to.
- More research on how to improve data quality.



The quest for representivity

The problem of nonresponse

- Response rates are decreasing
- Therefore, quality of the survey outcomes decrease
- Its more costly to fight nonresponse

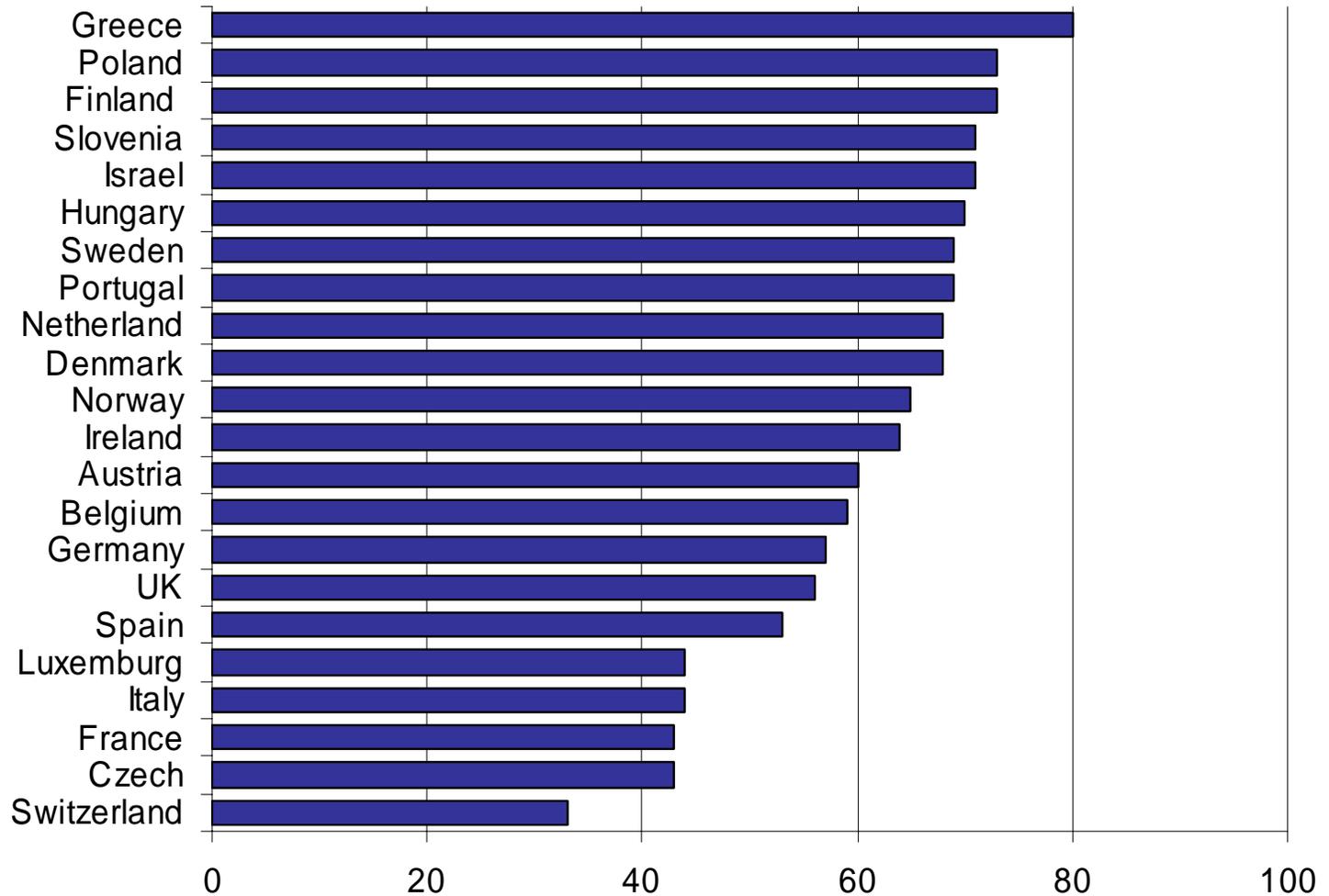
What we need?

- Quality indicators
- More effective fieldwork procedures



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.



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 9e 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 ρ_k of responding when selected in the sample.



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

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

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)

Computation

- Response probabilities are unknown
- Therefore, they have to be estimated
- Fit logit model for response probabilities.
- Other models: probit or linear
- Required: auxiliary variables
- Values must be available for both respondents and non-respondents
- $R(\rho)$ can be estimated.



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

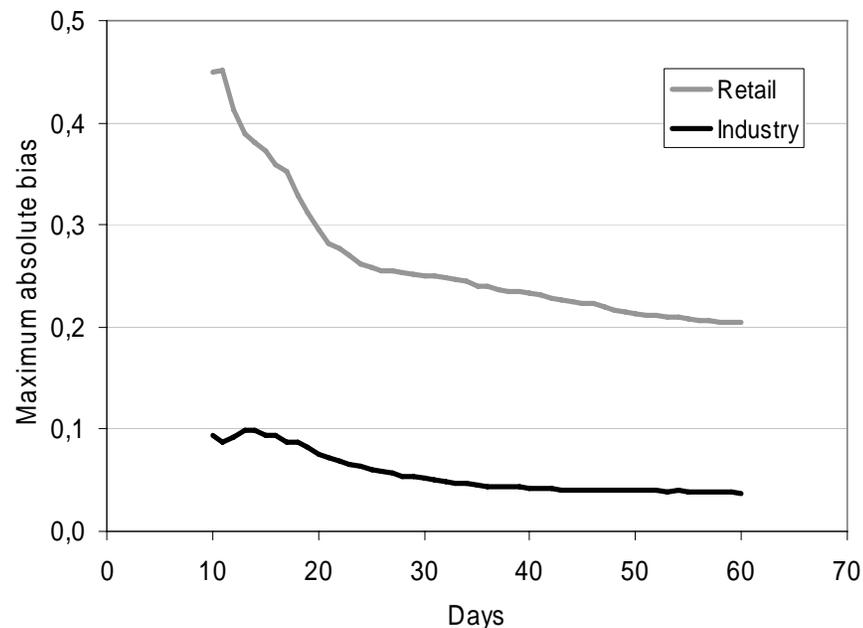
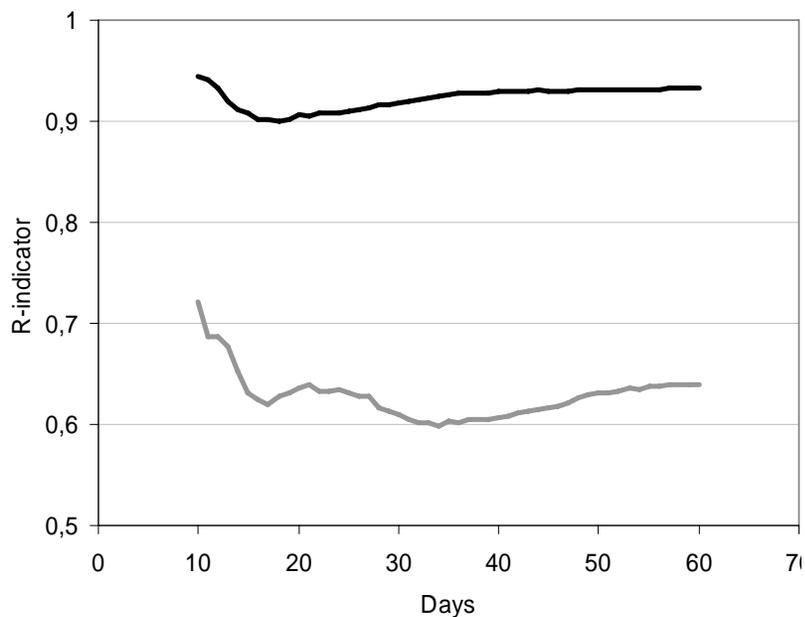
- Survey: Dutch Labour Force Survey.
- Experiment: Can response rates be improved by using incentives
- Three groups: no incentives, 5 stamps, 10 stamps

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.

Application 3

- 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.



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:

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

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.risq-project.eu



Response survey design

Fixed or flexible survey design?

- Traditional surveys have fixed designs. All is decided in the design phase: sampling design, mode of data collection, number of call-back, etc.
- It may turn out during data collection that design decisions were not the best ones.
- To improve quality and reduce costs, it may be better to change the design on the run.
- Groves & Heeringa (2006): Responsive design
- Compare Deming (1986): Inspection of the final product to improve quality is too late, ineffective and costly. Quality must be build controlled during the process.
- Required: quality control indicators.
- Required: process data (paradata)



Response survey design

Implementing responsive survey design

- Divide fieldwork into a number of data collection phases. Each phase has may have own characteristics.
- Identify survey design features that may effect costs and quality. For example, mode of data collection, sample size, number of call-backs.
- Define indicators that measure these design features. For example: response rate, R-indicator.
- Decide at the end of each stage to change the design features of the next stage, using the values of the indicators.
- At the end of the fieldwork period, combine the data from the various stages to obtain estimators for population characteristics.



Blaise

Some consequences for Blaise

- More attention for implementing state-of-the-art web surveys.
- More support for mixed-mode data collection, such as a general case management system.
- Integrated survey data and paradata.
- Support for computing (possibly complex) indicators. Manipula may not be sufficiently powerful.
- Better links to analysis packages, like SPSS.
- Link to R language?

