Mobile phone surveys in mixed mode environment: Balancing costs and errors

Vasja Vehovar, University of Ljubljana, Slovenia

London, 17th February, 2009

Lessons from the first day

- mixed modes are not used much in marketing yet
- mobile modes are not used much yet either
- all presentations were focused on mobile CSAQ (web?)
- is mobile research the same as mobile CSAQ?
- mobile IVR mentioned, mobile CATI not mentioned much
- mobile CSAQ very useful for specific needs (fast, pictures,..)
- all empirical examples were recruiting from some web panel
- technical problems due to devices will need years
- certain lack of commercial interest from clients

Structure

1. Sample survey research
2. Survey errors
3. Computer assisted data collection
4. Non-probability samples
5. Mixed mode context
6. Survey costs
7. Cost – error optimization
8. Empirical example
9. Conclusions

1. Sample survey research

Big picture

In general, we conduct sample survey data collection to infer about the entire target population.

We would like to perform this with:

1. highest possible methodological quality and with
2. lowest possible spending of resources and other inconveniences.

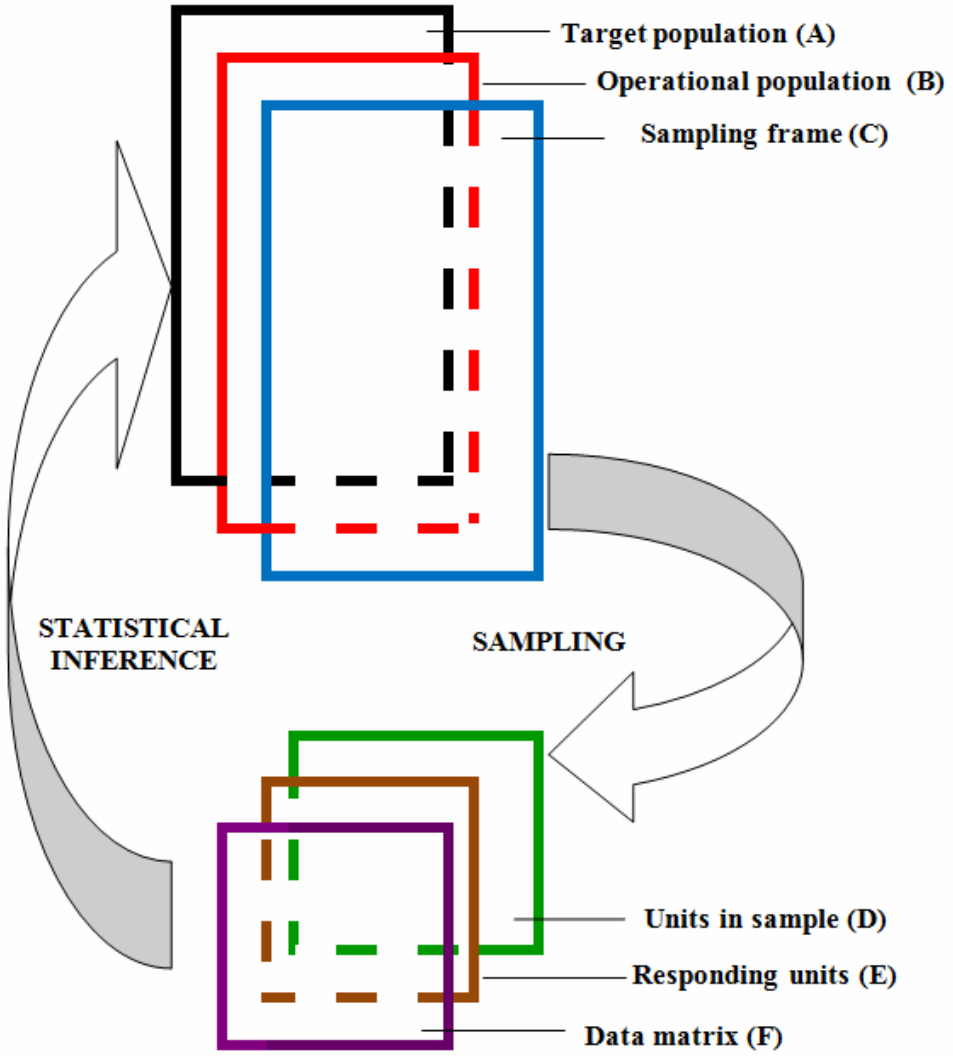
Questions

- How we perform the inference?
- How we measure the two conflicting aspects, i.e. the survey errors and the survey costs?
- How we balance and optimize?

Trends in survey data collection

1. Trend towards paper-less and people-less data collection
2. Trend towards non-probability samples
3. Trend of mixing survey modes

2. Survey errors


Target population (A)
Operational population (B)
Sampling frame (C)
STATISTICAL INFERENCE
SAMPLING
Units in sample (D)
Responding units (E)
Data matrix (F)

Sampling error

Sampling error is related to the essence of statistical inference.

It is a price for having only a sample, not the population.

Of course, we need probability samples to calculate it.

It is also the basis for confidence intervals, e.g. P = 20% ± 4%

In general, it decreases with square root of sample size: ~ $1/\sqrt{n}$

It is the only error we can calculate easily

Sampling error SE (p) is a square root of the sampling variance VAR(p)

Other random errors

There are other random errors, e.g.:

- reliability of measurement instrument,
- interviewer variability,
- respondents variability.

These are more complicated to compute and evaluate.

We usually believe that these errors are below sampling error.

Biases

Bias is systematic difference between our estimate and true value:

$$\text{bias}(p) = p - P$$

There are numerous sources of bias:

- noncoverage,
- nonresponse,
- sampling frame,
- respondent,
- measurement instrument...

We believe that noncoverage and nonresponse biases dominate.

Mean squared error (MSE)

Accuracy, measured by MSE:

$$MSE(p) = Var(p) + Bias^2(p)$$

Instead of sampling variance we use MSE for interval estimates.

Notion of data quality

The concept of data quality is much broader than the concept of total survey error or mean squared error.

Other dimension are also important: timing, validity, comparability, consistency, documentation, ….

.

3. Computer assisted survey data collection - CASIC

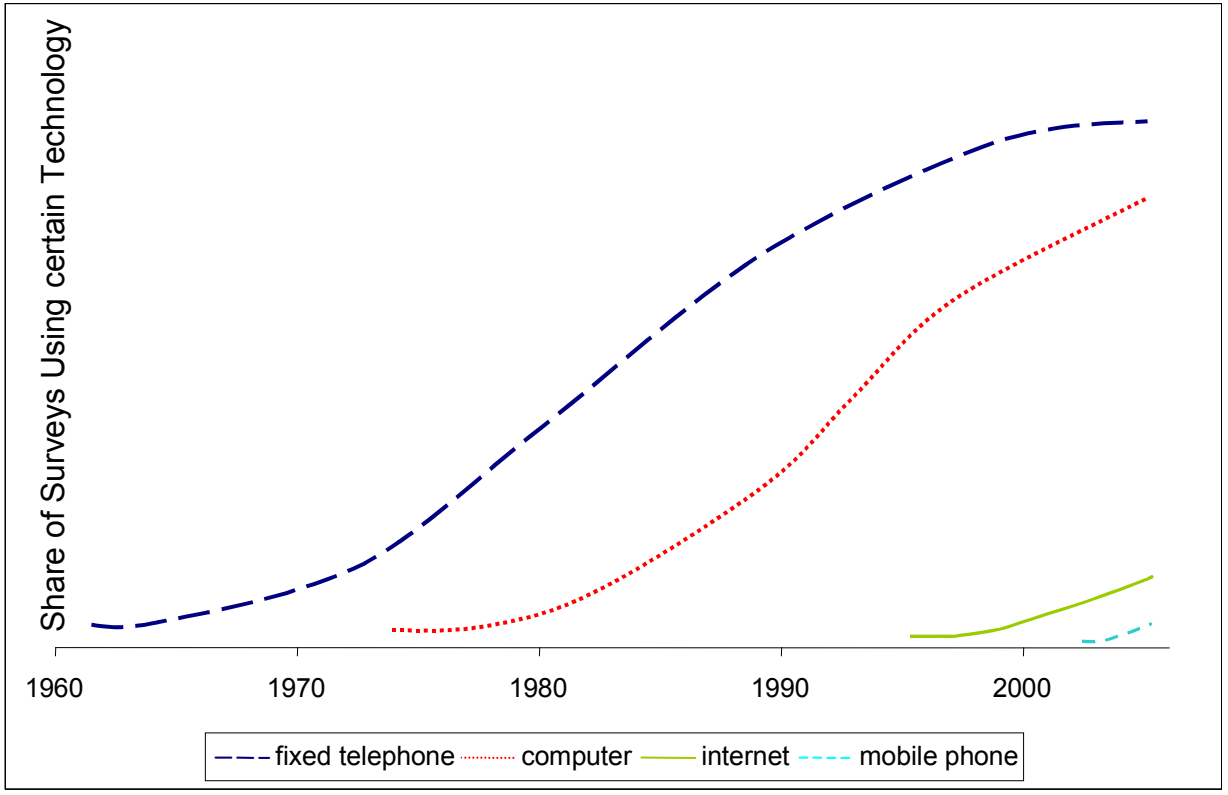
Emergency of CASIC options

New survey options are introduced increasingly:

- Due to new/improved communication protocols,
- Due to new devices,
- Due to changes in ICT prices,
- Due to new ICT applications and services that support and/or improve survey process.

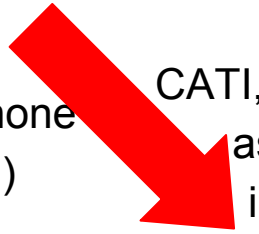
Are these survey modes, or, just technological options?

Devices in probability samples



Interviewer-less and paper-less surveys

Interviewer involvement	Survey mode	
	Paper and pencil	CASIC
Interviewer presence	Paper and pencil (face-to-face) interviewing (PAPI)	CAPI, CASI, Audio/Video CASI
Remote interviewer	Paper assisted telephone interviewing (PATI)	CATI, CAVI (computer assisted video interviewing)
No interviewer	Self-administered paper questionnaires (mail questionnaires)	**Web/mobile CSAQ, TDE, IVR, Virtual interviewer,…**



4. Non-probability samples

Non-probability surveys and panels

When we do not control the sample selection for all units of the population, we talk about non-probability samples.

Usually, we let the respondents to self-select them into the sample.

Why would we pay 10 times more for probability sample if we can get the same results much cheaper and much faster?

Large self-selected market research household panels have been around for decades and they have been serving well their purpose.

Internet (access) panel are only expanding this approach.

Improvements

More and more can be done to improve quality of these samples:

- ***Recruiting***: disperse marketing efforts.
- ***Sampling***: intelligent selection of units into the panel/survey.
- ***After data collection***:
  - weighting, calibration, propensity score weighting,
  - modeling, causal analysis, (multiple) imputation, matching.

Performance

These samples may work well for a lot of marketing purposes.

But how they behave when we need a reliable estimate for an unexplored phenomena in the general population?

There is mixed evidence about how these estimates work when we can control the results (e.g. elections).

There is a lot evidence about these samples being painfully wrong, but of course, they vary dramatically in their quality and in price.

Probability samples with low response

When the response rate (RR) destroys the probability nature?

- Old standards were 80%, then 70% and now we talk 60%, 50%.

- Is it enough to have RR of 20%, 30%, 40%?

- Most interestingly - what is the value around 1%, 5%, 10%?

What is better, good quality self-selected panel of probability sample with 2% or 12% response rate?

The art of non-probability samples

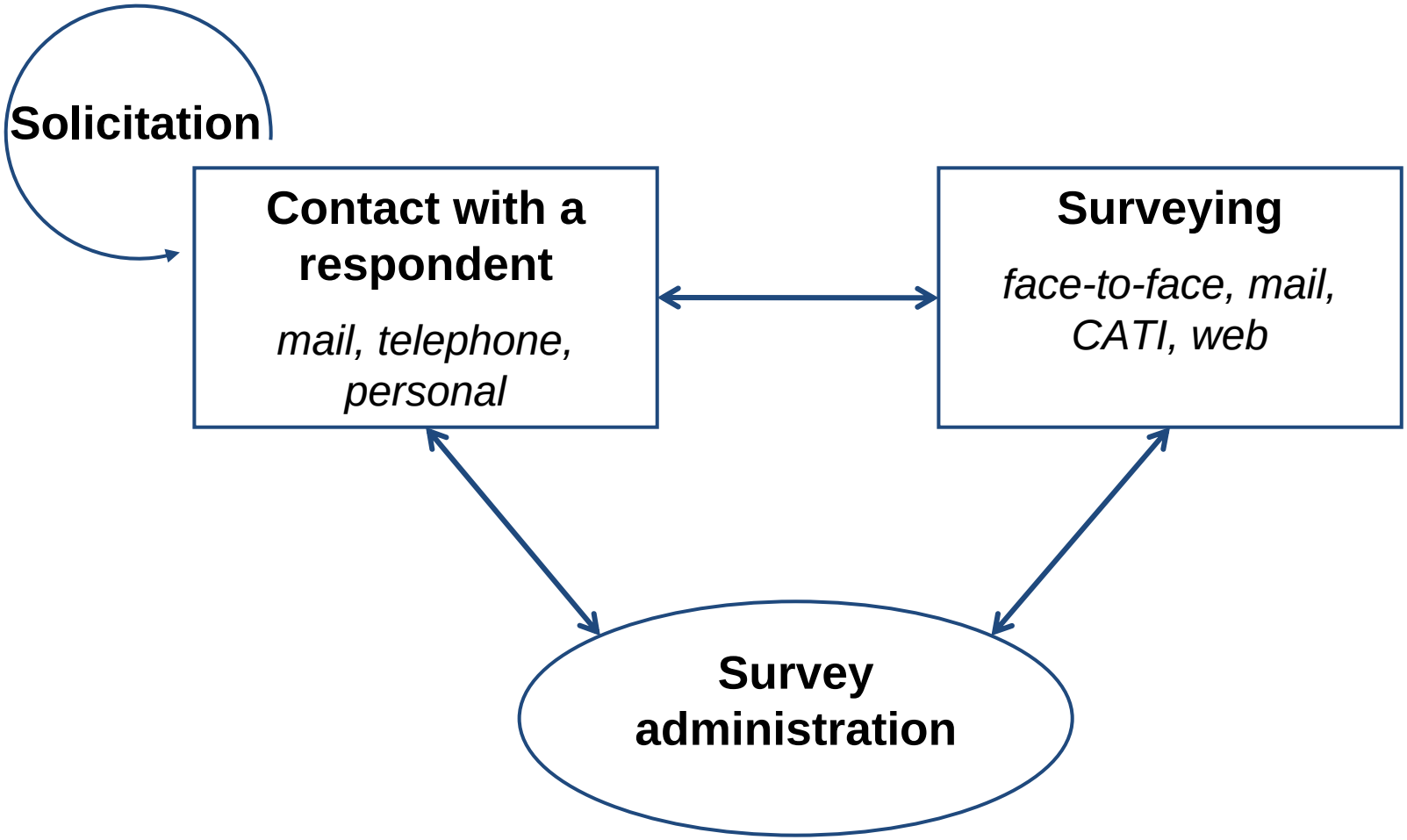
"…. it is not a scientific method with precise definition. It is more of an art practiced widely with very different skills and diverse successes by many people in different places.

There exist no textbooks on the subject to which we can refer to base our discussion. This alone should be a warning signal."

Leslie Kish on quota sampling, 1993

5. Mixed mode context

Mixed-mode designs



How we mix survey modes?

Three major approaches:

(A) give options to respondents (e.g. They can choose mail or web), what seems not to be very effective,

(B) contact the non-respondents with different (sharper) mode, e.g. email invitation to web is followed by telephone call,

(C) use different modes for different population segments, which may overlap or not (dual frames)

Why we mix survey modes?

First, with mixing modes we combine different solicitation and data collection modes (e.g. Mail invitation to web CSAQ, or, SMS invitation to mobile IVR, etc).

Second, with that we often hope to:

1. increase response and/or coverage rates (and thus lower the corresponding biases):
    - follow-up mode may convert the non-respondents (e.g. unsuccessful mail attempt is followed with telephone one);
    - additional frame may increase the coverage of the target population (e.g. mobile phone combined with face-to-face);
2. lower the costs (e.g. web, TDM mail)

Mixing modes to increase the rates

Most often we mix modes to increase the response and/or coverage rates.

But what is the relation between rates and biases?

It has been shown (Groves, POQ 2006, Gallup 2009) that **ACCROSS** the surveys and questions, there is not much evidence that surveys/questions with high response rates would have lower non-nonresponse bias.

But here, of course, we do not have any controls neither insight into numerous other interfering factors.

Mixing modes to increase the rates

Of course, **WITHIN** each survey this relation does exist.

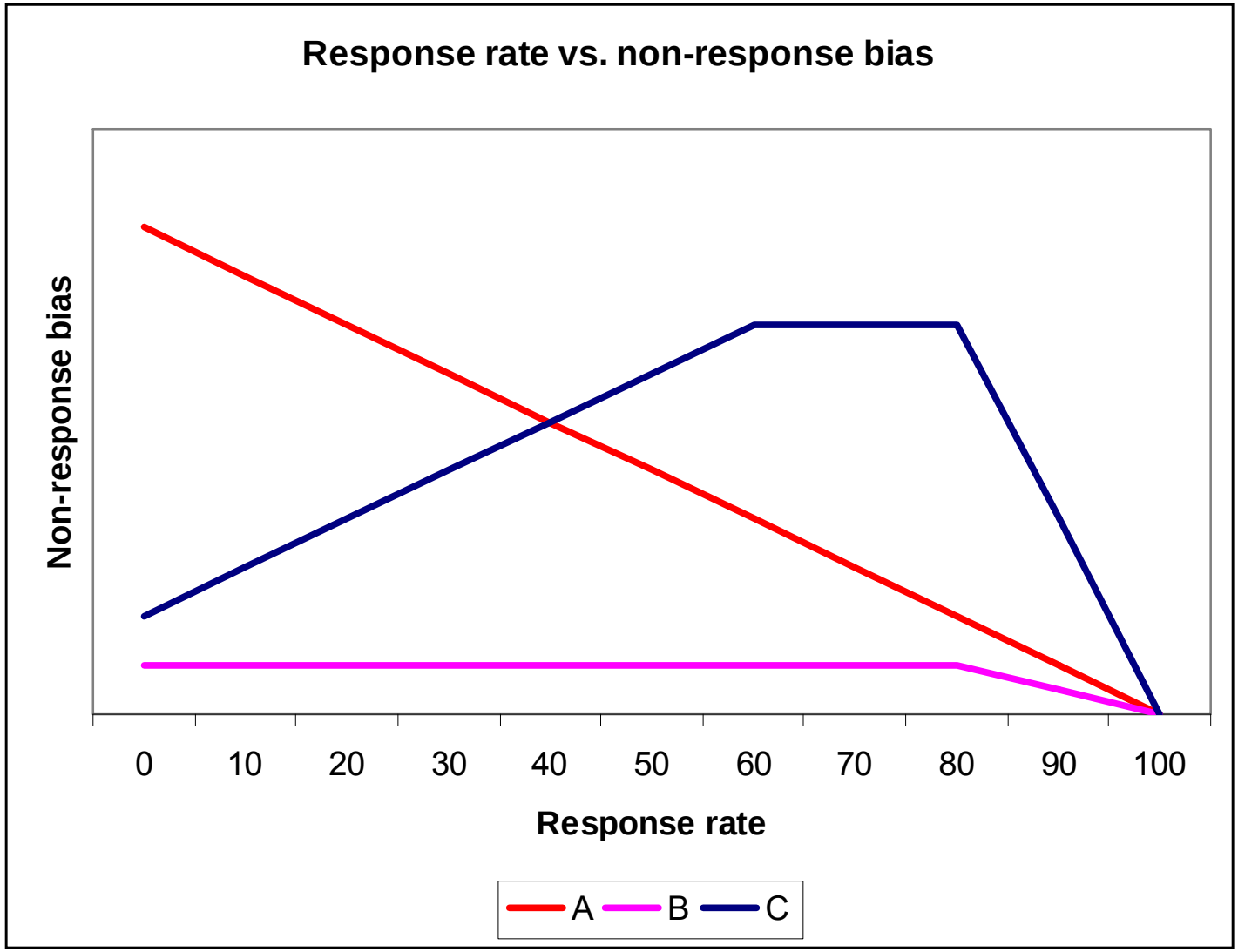
Well-known formulae (Kish 1965);

$$Bias_{NR}(y) = Wn * (Yn-Yr)$$

Obviously, no non-response (Wn=0) → no bias.

Similar is also true for non-coverage bias.

Rates vs. Biases



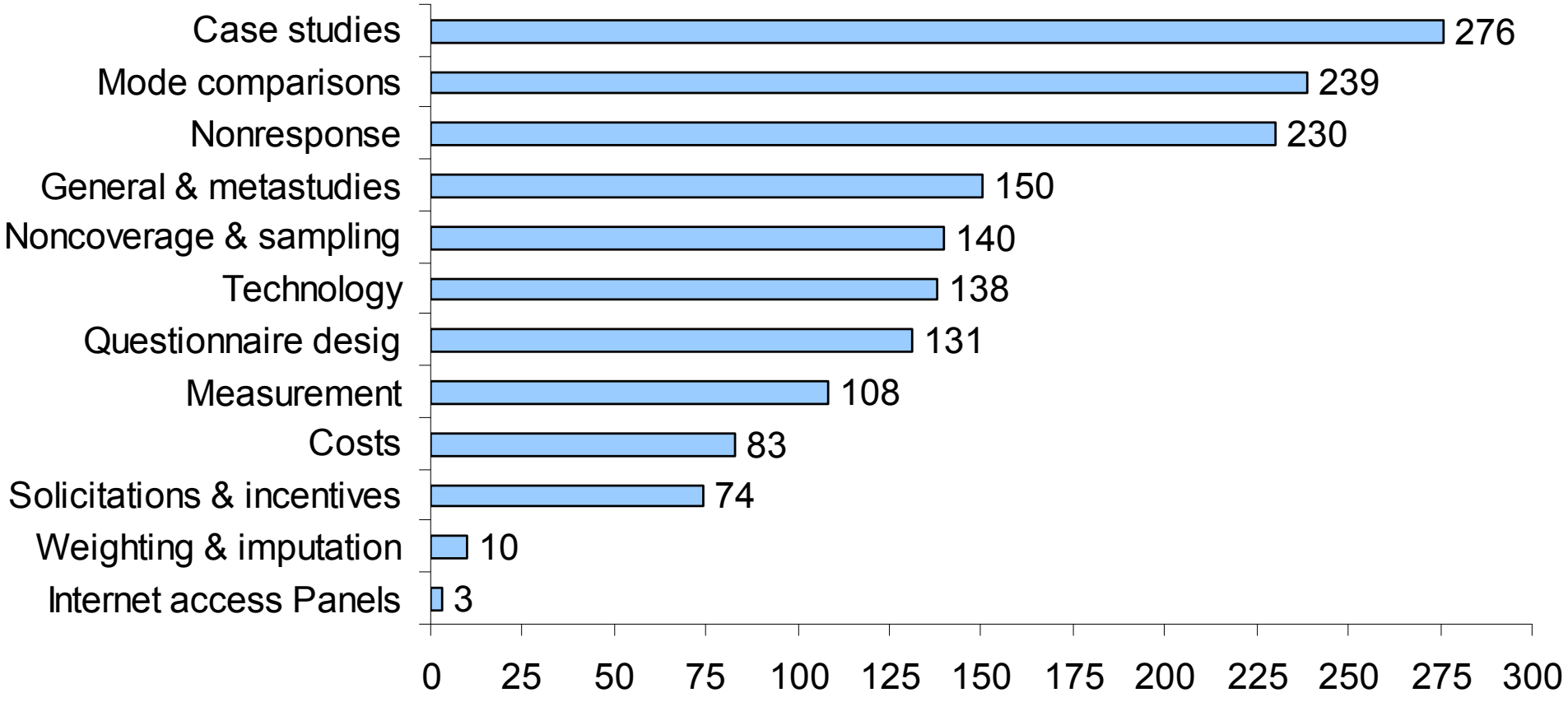
6. Survey costs

Literature on survey costs

- There is almost no literature specialized on this issue.
- When response rates of different modes are compared, costs are almost never involved, what is unfair for cheaper modes.
- Two possible explanations:
  - researchers do not think properly about costs.
  - businesses do not write much about how they handle costs.

Number of journal papers in selected categories

(Web Survey Methodology, http://WebSM.org)



Cost model

General model for estimation of costs:

$$C = c_0 + \sum_{k=1}^{K}\sum_{m=1}^{M}(c_{0_{km}} + c_{km} \cdot n_{km}) + \sum_{k=1}^{K}\sum_{m=1}^{M}(a_{0_{km}} + a_{km} \cdot r_{km})$$

solicitation

data collection

- number of solicitation waves (K)
- number of modes within the k-th wave (M)
- fixed costs (c_0, $c_{0_{km}}$, $a_{0_{km}}$)
- per-unit variable costs (c_{km}, a_{km})
- can also add stages, strata, phases,...

7. Optimisation

Mixing modes to optimize the costs

With our money we would like to buy the best information, i.e. the survey data with lowest survey error.

We should thus minimize the product:

Survey Cost * Survey Errors

Estimating survey errors

Problems with MSE as the estimate for errors:

- Is MSE exhaustive enough for survey errors?
- How to estimate the unknown population value of the variable P, *so* to calculate the bias = (P-p) ?
- Which are the key variables to be used? (As e*ach variable may have a unique optimization)*.

Estimating survey costs

- No conceptual problems,
- Just practical issues related to book-keeping and desegregation,

Approaches to the problem

- Analytical solutions for optimization
- Simulation studies
- Web application
- Case study

8. Empirical example

Case study survey description

EU survey on ICT usage 2008 (households):

- an official Eurostat survey;
- in Slovenia:
  - conducted by the Statistical Office of the Republic of Slovenia;
  - face-to-face and CATI;
  - general population, 10-74 years
  - Central Register of Population as sampling frame
  - 44 questions

Experimental design

Part by the Statistical Office (SORS), split sample (total 2000 unites):

- half F2F, half CATI (plus F2F follow up for non-respondents);
- both recruited from the register of population, up to 5 contacts

Part by the Faculty of Social Sciences (FSS), cells of 100 units:

- 7 mixed-mode experimental cells (B type) with the web (initial mail contact was based on register of population)
- 2 mixed mode experimental cells (C type) with telephone (CATI frame - telephone directory; mobile – RDD)
- Plus simulation (again C type) for 2/3 CATI and 1/3 mobile dual frame sample;
- only individuals 10-50 years old, up to 3 contacts

Pilot experimental cells

	Web options (B)			Telephone (A)	
	Web / Mail	**Mail, no web**	**Web / CATI**	**Mobile**	**CATI**
No incentive	200	100	100	100	100
Non-monetary	100	/	100	/	/
Monetary (5€)	100	/	100	/	/

Target variables

Target variables used for illustrative calculations:

- use of the Internet in the last three months
- age
- mobile only persons (no fixed telephone in the household)

True population value: Assumed to be obtained by F2F mode.

Comparisons

We analyzed all cells for fixed (equal) effective sample sizes (n=1000).

We used the parameters from real data to recalculate the figures.

We present here only the variable AGE.

.

Group	*Response rate*	*Costs (€)*	$\bar{X}$	Bias	MSE	MSExCosts
Mail/web, no incentive	29%	8,966	28.56	2.49	6.6	59172
Mail/web, wallet	35%	13,909	32.24	-1.19	1.8	25034
Mail/web, 5€ in cash	73%	10,211	28.71	2.34	5.6	57181
Mail-only, no incentive	23%	11,509	30.24	0.81	1.3	14961
Web/CATI, no incentive	32%	9,394	35.10	-4.05	17.0	159694
Web/CATI, wallet	30%	18,077	29.90	1.15	1.7	30730
Web/CATI, 5€ in cash	51%	15,382	29.90	1.15	1.5	23075
CATI-only, no incentive	55%	3,409	35.15	-4.1	16.9	57615
Mobile-only, no incentive	30%	6,300	27.04	4.01	16.2	102060
CATI & Mobile, no incentive	46%	4,091	32.47	-1.42	2.2	9000
F2F*, no incentive	73%	12,697	31.05	0.00	0.2	2540

Variable: Internet Group	*Response rate*	*Costs* = 2000 €			
		Initial sample size	*p*	Bias	MSE
Mail/web, no incentive	29%	717	0.895	-0.13	0.016
Mail/web, wallet	35%	346	0.886	-0.12	0.014
Mail/web, 5€ in cash	73%	218	0.958	-0.19	0.036
Mail-only, no incentive	23%	732	1.000	-0.23	0.053
Web/CATI, no incentive	32%	568	0.900	-0.13	0.017
Web/CATI, wallet	30%	279	0.833	-0.06	0.005
Web/CATI, 5€ in cash	51%	184	0.860	-0.09	0.010
CATI-only, no incentive	55%	1072	0.810	-0.04	0.002
Mobile-only, no incentive	30%	1063	0.840	-0.07	0.005
CATI & Mobile, no incentive	46%	1069	0.820	-0.05	0.003
F2F*, no incentive	73%	198	0.770	0.00	0.001

Variable: Mobile only Group	Response rate	Initial sample size	Costs = 2000 € p	Bias	MSE
Mail/web, no incentive	29%	717	0.255	-0.12	0.015
Mail/web, wallet	35%	346	0.206	-0.07	0.006
Mail/web, 5€ in cash	73%	218	0.219	-0.08	0.008
Mail-only, no incentive	23%	732	0.100	0.04	0.002
Web/CATI, no incentive	32%	568	0.200	-0.06	0.005
Web/CATI, wallet	30%	279	0.069	0.07	0.005
Web/CATI, 5€ in cash	51%	184	0.160	-0.02	0.002
CATI-only, no incentive	55%	1072	0.050	0.09	0.008
Mobile-only, no incentive	30%	1063	0.300	-0.16	0.027
CATI & Mobile, no incentive	46%	1069	0.133	-0.02	0.001
F2F*, no incentive	73%	198	0.136	0.00	0.001

Comments

1. Important limitations/specifics of our results/research:
   - data were not weighted yet (intentionally),
   - we did not observe attitudes but facts; attitudes behave “better” (they are more robust).

2. Different criteria give us different solution for the optimal combination of survey modes.

3. Bias dominates over sampling error, even in case of small samples.

4. Every variable may suggest different optimization.

9. Conclusions

Well, where are the mobile phones?

1. When talking about modern survey data collection, mobile phones are but one element in broad range of options.

2. If we talk about mobile CATI, this is increasingly important option in surveys of general population.

3. If we talk about mobile CSAQ, the usage for general population surveys can be foreseen only for future years (unless we provide mobile devices to respondents).

Mixed mode options

... within respondent (option B), two waves:

- Mail-Web →Mail-Mail (with various combinations)
- In case of a panel or register:
  - SMS-Mobile Web → Email-Web
  - Email-Web (or Mail-Web) → Fixed/Mobile
  - Fixed/Mobile→F2F
  - Mail-Web →Mail-Mail → Phone/Mobile → F2F

... in dual frames (option C):

- Mobile & Fixed
- Mobile & F2F
- Fixed & F2F
- Mail-Web & Fixed/mobile

Summary

1. Are we explicit what we optimize? Response rates? Coverage rates? Sum of squared discrepancies for socio-demographic variables? Costs? Biases? MSEs? Or, product MSE*Costs?

2. Cost-error issues in mixed mode surveys are very complex to process intuitively. Each variable may behave differently.

3. There is no general solution for our specific cost-error problem. We need more analysis of our past costs and biases. We need more experiments for better decisions in the future.

4. It is very hard to beat the face-to-face option (bias dominates!).

Future: probability panels?

Probability based panels:

- F2F or telephone recruiting
- optional supply of PC (mobile device?)
- lot of incentives (initial one plus monthly ones, all monetary).

LISS panel in Netherlands: RR around 50%, cost/minute around 1€.

In future a range (price, quality) of supply to appear on the market:

- entirely self-selected (with wide range of quality),
- commercial probability panels with up to 10% response rates,
- advanced probability panels with response rate around 50%.

More

http://WebSM.org