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
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
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: $\sim 1/\sqrt{n}$

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.
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:

\[
\text{MSE}(p) = \text{Var}(p) + \text{Bias}^2(p)
\]

Instead of sampling variance we use MSE for interval estimates.
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

<table>
<thead>
<tr>
<th>Interviewer involvement</th>
<th>Survey mode</th>
<th>CASIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper and pencil</td>
<td>Paper and pencil (face-to-face) interviewing (PAPI)</td>
<td>CAPI, CASI, Audio/Video CASI</td>
</tr>
<tr>
<td></td>
<td>Paper assisted telephone interviewing (PATI)</td>
<td>CATI, CAVI (computer assisted video interviewing)</td>
</tr>
<tr>
<td>No interviewer</td>
<td>Self-administered paper questionnaires (mail questionnaires)</td>
<td>Web/mobile CSAQ, TDE, IVR, Virtual interviewer, …</td>
</tr>
</tbody>
</table>
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.
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

Solicitation

Contact with a respondent
mail, telephone, personal

Surveying
face-to-face, mail, CATI, web

Survey administration
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 across the surveys and questions, there is not much evidence that surveys/questions with high response rates would have lower non-response 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);

\[ \text{Bias}_{NR}(y) = Wn \times (Yn - Yr) \]

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

Similar is also true for non-coverage bias.
Rates vs. Biases

Response rate vs. non-response bias

Non-response bias vs. Response rate
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)

- Case studies: 276
- Mode comparisons: 239
- Nonresponse: 230
- General & metastudies: 150
- Noncoverage & sampling: 140
- Technology: 138
- Questionnaire design: 131
- Measurement: 108
- Costs: 83
- Solicitations & incentives: 74
- Weighting & imputation: 10
- Internet access Panels: 3
Cost model

General model for estimation of costs:

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

- number of solicitation waves \((K)\)
- number of modes within the \(k\)-th wave \((M)\)
- fixed costs \((c_0, c_{0km}, a_{0km})\)
- 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 each 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
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 units):
• 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

<table>
<thead>
<tr>
<th></th>
<th>Web options (B)</th>
<th>Telephone (A)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Web / Mail</td>
<td>Mail, no web</td>
</tr>
<tr>
<td>No incentive</td>
<td>200</td>
<td>100</td>
</tr>
<tr>
<td>Non-monetary</td>
<td>100</td>
<td>/</td>
</tr>
<tr>
<td>Monetary (5€)</td>
<td>100</td>
<td>/</td>
</tr>
</tbody>
</table>
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.
<table>
<thead>
<tr>
<th>Group</th>
<th>Response rate</th>
<th>Costs (€)</th>
<th>( \bar{X} )</th>
<th>Bias</th>
<th>MSE</th>
<th>MSExCosts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mail/web, no incentive</td>
<td>29%</td>
<td>8,966</td>
<td>28.56</td>
<td>2.49</td>
<td>6.6</td>
<td>59172</td>
</tr>
<tr>
<td>Mail/web, wallet</td>
<td>35%</td>
<td>13,909</td>
<td>32.24</td>
<td>-1.19</td>
<td>1.8</td>
<td>25034</td>
</tr>
<tr>
<td>Mail/web, 5€ in cash</td>
<td>73%</td>
<td>10,211</td>
<td>28.71</td>
<td>2.34</td>
<td>5.6</td>
<td>57181</td>
</tr>
<tr>
<td>Mail-only, no incentive</td>
<td>23%</td>
<td>11,509</td>
<td>30.24</td>
<td>0.81</td>
<td>1.3</td>
<td>14961</td>
</tr>
<tr>
<td>Web/CATI, no incentive</td>
<td>32%</td>
<td>9,394</td>
<td>35.10</td>
<td>-4.05</td>
<td>17.0</td>
<td>159694</td>
</tr>
<tr>
<td>Web/CATI, wallet</td>
<td>30%</td>
<td>18,077</td>
<td>29.90</td>
<td>1.15</td>
<td>1.7</td>
<td>30730</td>
</tr>
<tr>
<td>Web/CATI, 5€ in cash</td>
<td>51%</td>
<td>15,382</td>
<td>29.90</td>
<td>1.15</td>
<td>1.5</td>
<td>23075</td>
</tr>
<tr>
<td>CATI-only, no incentive</td>
<td>55%</td>
<td>3,409</td>
<td>35.15</td>
<td>-4.1</td>
<td>16.9</td>
<td>57615</td>
</tr>
<tr>
<td>Mobile-only, no incentive</td>
<td>30%</td>
<td>6,300</td>
<td>27.04</td>
<td>4.01</td>
<td>16.2</td>
<td>102060</td>
</tr>
<tr>
<td>CATI &amp; Mobile, no incentive</td>
<td>46%</td>
<td>4,091</td>
<td>32.47</td>
<td>-1.42</td>
<td>2.2</td>
<td>9000</td>
</tr>
<tr>
<td>F2F*, no incentive</td>
<td>73%</td>
<td>12,697</td>
<td>31.05</td>
<td>0.00</td>
<td>0.2</td>
<td>2540</td>
</tr>
<tr>
<td>Group</td>
<td>Response rate</td>
<td>Initial sample size</td>
<td>p</td>
<td>Bias</td>
<td>MSE</td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>---------------</td>
<td>---------------------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>Mail/web, no incentive</td>
<td>29%</td>
<td>717</td>
<td>0.895</td>
<td>-0.13</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>Mail/web, wallet</td>
<td>35%</td>
<td>346</td>
<td>0.886</td>
<td>-0.12</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td>Mail/web, 5€ in cash</td>
<td>73%</td>
<td>218</td>
<td>0.958</td>
<td>-0.19</td>
<td>0.036</td>
<td></td>
</tr>
<tr>
<td>Mail-only, no incentive</td>
<td>23%</td>
<td>732</td>
<td>1.000</td>
<td>-0.23</td>
<td>0.053</td>
<td></td>
</tr>
<tr>
<td>Web/CATI, no incentive</td>
<td>32%</td>
<td>568</td>
<td>0.900</td>
<td>-0.13</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>Web/CATI, wallet</td>
<td>30%</td>
<td>279</td>
<td>0.833</td>
<td>-0.06</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Web/CATI, 5€ in cash</td>
<td>51%</td>
<td>184</td>
<td>0.860</td>
<td>-0.09</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>CATI-only, no incentive</td>
<td>55%</td>
<td>1072</td>
<td>0.810</td>
<td>-0.04</td>
<td><strong>0.002</strong></td>
<td></td>
</tr>
<tr>
<td>Mobile-only, no incentive</td>
<td>30%</td>
<td>1063</td>
<td>0.840</td>
<td>-0.07</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>CATI &amp; Mobile, no incentive</td>
<td>46%</td>
<td>1069</td>
<td>0.820</td>
<td>-0.05</td>
<td><strong>0.003</strong></td>
<td></td>
</tr>
<tr>
<td>F2F*, no incentive</td>
<td>73%</td>
<td>198</td>
<td>0.770</td>
<td>0.00</td>
<td><strong>0.001</strong></td>
<td></td>
</tr>
<tr>
<td>Variable: Mobile only</td>
<td>Response rate</td>
<td>Initial sample size</td>
<td>Costs = 2000 €</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------</td>
<td>---------------------</td>
<td>----------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Group</strong></td>
<td><strong>Rate</strong></td>
<td><strong>Sample Size</strong></td>
<td><strong>p</strong></td>
<td><strong>Bias</strong></td>
<td><strong>MSE</strong></td>
<td></td>
</tr>
<tr>
<td>Mail/web, no incentive</td>
<td>29%</td>
<td>717</td>
<td>0.255</td>
<td>-0.12</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td>Mail/web, wallet</td>
<td>35%</td>
<td>346</td>
<td>0.206</td>
<td>-0.07</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>Mail/web, 5€ in cash</td>
<td>73%</td>
<td>218</td>
<td>0.219</td>
<td>-0.08</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>Mail-only, no incentive</td>
<td>23%</td>
<td>732</td>
<td>0.100</td>
<td>0.04</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>Web/CATI, no incentive</td>
<td>32%</td>
<td>568</td>
<td>0.200</td>
<td>-0.06</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Web/CATI, wallet</td>
<td>30%</td>
<td>279</td>
<td>0.069</td>
<td>0.07</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Web/CATI, 5€ in cash</td>
<td>51%</td>
<td>184</td>
<td>0.160</td>
<td>-0.02</td>
<td><strong>0.002</strong></td>
<td></td>
</tr>
<tr>
<td>CATI-only, no incentive</td>
<td>55%</td>
<td>1072</td>
<td>0.050</td>
<td>0.09</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>Mobile-only, no incentive</td>
<td>30%</td>
<td>1063</td>
<td>0.300</td>
<td>-0.16</td>
<td>0.027</td>
<td></td>
</tr>
<tr>
<td>CATI &amp; Mobile, no incentive</td>
<td>46%</td>
<td>1069</td>
<td>0.133</td>
<td>-0.02</td>
<td><strong>0.001</strong></td>
<td></td>
</tr>
<tr>
<td>F2F*, no incentive</td>
<td>73%</td>
<td>198</td>
<td>0.136</td>
<td>0.00</td>
<td><strong>0.001</strong></td>
<td></td>
</tr>
</tbody>
</table>
Comments

1. Important limitations/specifcics 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


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