Determinants of survey response in Croatian EVS 2008.

IVAN RIMAC
LEA ZOREC
JELENA OGRESTA

FACULTY OF LAW, UNIVERSITY OF ZAGREB

EVS 2008

- EVS 2008 wave introduced many methodological demands in order to improve representativeness and comparability of data.
- The most of these improvements are part of struggle for probabilistic type of sampling in survey research
- To mention the most pronounced ones:
  - More strictly defined sampling frame
  - Probabilistic sampling procedure
  - No substitution of respondents that have refused to participate
  - Obvious personal first contact
  - 4 visits to those who are not reachable on first visit
  - …

Goals

- Aim of all these improvements (or complications) is to make our survey more representative to population (in terms of lower bias) and, in the same time, (more) invariant in terms of comparison to other samples.
- But in different countries there are different conditions in which survey has to be planed and conducted.
- More or less each country implements these methodological demands in a different way.
- Subsequent analyses could lead to different conclusions.
- Usually in surveys, when fieldwork is finished, researchers turn to results, forget what they have done in methodology and think about it only if data did not confirm their theory.

Aim of presentation

- Aim of this presentation is to analyse methodology used in Croatian EVS wave 2008 and to make some conclusions about introduced changes.

Sampling frame

Date: **11.04.2008**

Target Population, Population coverage	All persons aged 18 and over living in private households in Croatia.

Sampling frame	Census data from 2001 for PSU. There is no sampling frame for further stages of sampling procedure.

Sampling procedure overview (2)

Multistage probability sampling in 3 stages

1st stage

- PSU=settlement
- PSUs are ordered by County, Municipality and number of inhabitants age 18+
- PPS sampling – systematic sample, cumulative size method (Lohr, 1999)
- Implicit stratification on region and size of settlement
- To each PSU 20 questionnaires are allocated
- This stage has properties of self-weighted sample

Sampling procedure overview (3)

Multistage probability sampling in 3 stages

2nd stage

- SSU=dwellings*
- Random walk method**
- Starting point for random walk was determined by simple random sampling from telephone book – fix lines. Starting point is not included in sample.
- Maximal sampling step in random procedure is 10. Step less than 10 will be applied in very small settlements in which the sampling step of 10 will lead to exhaustion of all possible dwellings.

*Dwellings are selected as SSU as they are less confusing for interviewers in on-field sampling (it's easy to connect with physical existence of entry door and it is easy to enumerate without proxy interviewing). All persons aged 18+ in all households are included in 3rd stage sampling.

*This stage of sampling combines two sampling strategies – as random walk step cannot cover whole settlement this method can be considered as cluster sampling (settlement is divided into x imaginary clusters of same size – number of clusters can be obtained by dividing number of inhabitants of settlement with number of persons accountable for the 3rd stage). Group of accountable persons for 3rd stage can be treated as cluster because random walk does not cover whole settlement. As picking the starting point is random, sample is probabilistic. Including relatively close subjects into sample has properties of cluster sampling, and demands checking cluster effect by interclass correlation.

- Sample weights need only clustering effect. Sample is self-weighted in this phase – implicitly same sized clusters within PSU are drawn by simple random sampling.

Sampling procedure overview (4)

Multistage probability sampling in 3 stages

3rd stage

- TSU=individual
- Last birthday method
- As interviewer's diary includes information of number of household members (here defined as dwelling) it is possible to calculate design weights for this stage.

Problems of implementation

- As procedure is different from usual fieldwork practice we expected that not all elements will go smooth in application.
- Conclusion was that only full control over process and good control of fieldwork should provide (at least evidence) how good fieldwork was.

- In organizational structure this leads us to conclusion that we have to build our own organization and do the fieldwork by our selves.
- Why:
  - If we transfer our demands to someone else (agency) strictness of demands will motivate agency to put all the dirt under carpet and never say to us what was really happening in the fieldwork.
  - We have enough experience in fieldwork to organize it.

Where we expected problems (1)

- 2nd stage: Random walk
  - It is usual to do the random walk in process or interviewing. Who can control if some interviewer tries one household and after refusal goes to another one and put the second one on random walk list as primary selection?
  - Interviewers in Croatia usually work in fieldworks that allow substitution with next door neighbour. How can we be convincing to prevent them not to do so (we fight against impression of good practice)?
  - Interviewers are paid by item (successful one) so they will obviously return with list of interviewed persons in order to enable control of their work (and payment in consequence), but would they bring us a list of non-respondents?

Where we expected problems (2)

- 3nd stage: Last birthday random selection in household
  - This is a usual practice, but do they count only present members or all members of household (second definition is in some cases blurred by members who are absent for more than a week – dislocated work, students studying in other town and live there in student homes…)?
  - Do they think that this procedure ruins their success of entering the household (passing the doorway) and are they keen to interview someone else instead of person that is the right choice but it is hot present in the moment?

Scientific reaction

- Scientific reaction is not in accordance to human reaction! Put more propositions in order to make sure that things are going against usual practice. Because methodological knowledge is our knowledge in which we differentiate our proof model from usual practice of making conclusions.
- And we did the same mistake again.
- We put some more control in process in order to force interviewers to do that right way.

More control

- 2nd stage – Do your random walk:
  - from the point we choose (random pick from telephone book),
  - do it in the way that we can count households on the list (advance always in same direction – direction of larger household numbers) and
  - make your random walk list day before you start the interviewing job and send it to us before you start (you have to delivery advance letter to your household in order to inform them that you'll come – activity that was mostly proven to be ineffective)
- 3rd stage – Choose your respondent after you:
  - put birthdates of all household members in your contact form and round your choice.
- In order to prevent interviewers from being *lazy* in asking open ended questions (like profession where in 1999 the dominant profession was "clerk in office") we provide them with the list of codes of National classification of professions to force them to comply to that level of precision.

Consequences

- Fieldwork lasts forever.
  - Start of fieldwork May 1st 2008
  - Temporary slowed down Aug 10th 2008 – Sep 1st 2008
  - New start of interviewing on field Sep 1st 2008
  - Closed Oct 31st 2008

  - From 130 locations – 14 were repeated once due to either low response rate or withdrawal of interviewer, and 2 were repeated ones more.

Resume of fieldwork

- Among 139 engaged interviewers (to cover 130 locations)
- 31 withdraw from engagement
- 26 showed some *inconsistencies* in random walk
- 31 broke the rules of last birthday at least once
- 56 did not return all contact forms

- Do we have to forgot all these things after we finish the fieldwork?

It is opportunity to test some things

- Test of fieldwork variations and its influence on data is facilitated by the fact that scientific evidence have plenty of experimental trials of different conditions of fieldwork, but there are very few analysis about concurrent influence of several factors in the real conditions.
- We were attracted by nice criteria variable not present in usual fieldwork research practice that can be used in regression model.

Regression analysis

- Entities
  - Interviewers (N=139)
- Criteria
  - Response (rate) for each interviewer – each interviewer had opportunity to contact 20 respondents and we have their number of successful trials (as some of them withdraw before they try all possible respondents we use relative response rate)
- Predictive variables (blocks of variables)
  - breaking the rules of fieldwork
  - interviewer's characteristics
  - location characteristics
  - characteristics of contact and respondent

Fieldwork rules

- Variables
  - days of work,
  - deviations from random walk (dichotomous 0=no deviations, 1=registered irregularities)
  - deviations form last birthday procedure (dichotomous 0=no deviations, 1=evidence of breaking the rules)
  - No contact forms (dichotomous, 0=all contact forms , 1=some contact forms are missing)
  - Withdrawal of engagement (dichotomous, 0=finished job on location, 1=withdrawal before the end of job)

Results

Criteria – response rate	B	std. error of B	Beta	t	p
Intercept	**,661**	,039		16,890	< 0,01
Deviation from random walk	,071	,059	,101	1,205	> 0,05
Days of work	**-,004**	,001	-,218	-2,786	< 0,01
deviations form last birthday procedure	**,432**	,176	,204	2,456	< 0,05
No contact forms	**,028**	,011	,191	2,442	< 0,05
Withdrawal of engagement	**,143**	,051	,217	2,830	< 0,01

R =	,482				
R2 =	,233	F = 8,068	df1=5	df2=133	p<0,01

Conclusions

- Deviation from random walk does not influence response rate (- in consequence locking of procedure supposed to be good – our detections of deviations are more or less false alarms).
- Prolonged work is due to problems of contacting the less approachable respondents. More days of work do not improve response rate (- do more visits improve our job? It is out of scope of this analysis.)
- Deviations of last birthday procedure (mostly), no contact forms and withdrawal are coping mechanisms to improve response rate (- in consequence to keep work profitable to interviewer).

Other predictors were added in regresion in blocks

- Personal characteristics of interviewer
  - Gender (male=0, female=1),
  - Year of birth,
  - Education level (elementary=1, vocational=2, Matura level =3, higher=4),
  - Years of experience in interviewing,
  - Location of work is familiar to interviewer (no=0, yes=1)
- Characteristics of location
  - Size of settlement (eight categories)
  - Proportion of 60+ persons in population
- Characteristics of contact and respondent
  - Reduced to factor scores

More analysis – factor analysis of contact form data

	vx1	vx2	vx3	vx4	vx5
No of contacts	,098	,103	,145	-,158	**,808**
Average no of contacts for successful interview	-,026	-,048	-,076	,106	**,856**
Number of male respondents	**,767**	,247	,071	-,095	,063
Number of employed respondents	**,797**	-,108	,167	**,422**	-,008
No of temporarely absent	**,359**	**,794**	-,053	-,244	-,033
No of present members of household	,056	-,049	**,888**	-,172	,088
No of respondents in single member households	-,004	**,649**	,197	**,474**	,027
No of respondents age 18-34	**,382**	-,054	**,816**	,074	,009
No of respondents age 35-64	**,842**	,207	-,019	,009	-,052
No of respondents age 65+	,063	**,896**	-,109	,037	,071
No of respondents age with les than elementary school	,146	**,666**	-,057	**-,427**	,000
No of respondents age with elementary school education	,125	**,380**	**,310**	**-,649**	-,030
No of respondents age with vocational school education	**,855**	,074	,143	-,048	,088
No of respondents with matura level education	-,061	,017	**,561**	**,322**	-,025
No of respondents age with high education	,187	,011	,196	**,820**	-,066

Factors

- Varimax factors are either approximation of dominant demographical segments (imprecise ones due to reduction of dimensionality) or content of contact characteristics.
  - (vx1) male employed respondent segment
  - (vx2) respondents form single member elderly households
  - (vx3) young respondents
  - (vx4) highly educated respondents
  - (vx5) number of contacts

Final stepwise solution

Criteria – response rate	B	std. error of B	Beta	t	p
Intercept*	8,262	2,389		3,458	<0,01
Days of work	-,001	,001	-,030	-,721	> 0,05
Withdrawal	**,215**	,027	,324	7,917	<0,01
Deviation from last birthday procedure	,157	,087	,074	1,815	> 0,05
No contact forms	**,068**	,006	,469	11,093	<0,01
Year of birth of interviewer	**-,004**	,001	-,127	-3,199	<0,01
Settlement size	-,006	,005	-,067	-1,241	> 0,05
Vx 1 male employed respondents	**,175**	,013	,633	13,627	<0,01
Vx 2 single member elderly households respondents	**,102**	,011	,370	8,990	<0,01
Vx 3 young respondents	**,105**	,011	,379	9,402	<0,01
Vx 4 highly educated respondents	**,033**	,013	,121	2,476	< 0,05
R =	,902				
R2 =	,813	F=55,669	df1=10	df2=128	p<0,01

Conclusions

- Entering of characteristics of contact and respondents tear of some of other determinants. It is expected that difficulties of approach some parts of population are the main reason of non-response and bias in coverage.
- But *knocking out* cheating in last birthday procedure leads to hypothesis that these cheats are motivated by difficulty of finding and convincing less available respondents to participate.
- Older interviewers have batter chances to higher response rate.
- Withdrawal from the assignment is still ultimate act of protection of financial interests of interviewer.
- Absence of characteristics of settlement (urbanization and elderly index) shows that non-response is individual respondent characteristics, not environmental factor. That is in opposition with usual conclusions that different response rate is consequence of rural or urban social environment.
- Frequently tested characteristics of contacts do not show regular influence, or at least it is outfitted by other determinants.

Further directions

- Refinement of statistical analysis -> multilevel modelling offers possibility to test interaction effects and test impact of these determinants on data.
- Focusing on specific hypothesis -> testing the idea of influence of economic model of paying of interviewers on their behaviour and consequences on data quality, testing other versions of restrictions and fieldwork control in order to reduce fieldwork bias.

Source (or advertising?)

- Data used in this presentation is availiable in:
  - Rimac, I., Zorec, L., Ogresta, J. (2010) Analiza determinanti odaziva u anketnom istraživanju Europske studije vrijednosti, *Društvena istraživanja*, vol. 19(1-2), 105.
- Downloadable from:

  http://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=79786
- Why on Croatian:
  - Visibility of new and better methods in domestic publications is obligatory for improvements in means of ESOMAR evaluation measures (domestic practice is standard of estimation).

Determinants of survey response in Croatian EVS 2008.

Thank you