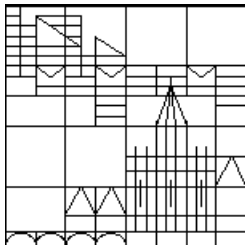


Inaccuracy of Birthday Respondent Selection Methods in Mail and Telephone Surveys

Rainer Schnell, Sonja Ziniel¹ und Elisabeth Coutts²

Center for Quantitative Methods and Survey Research
University of Konstanz



Structure of the talk

Selection for surveys within households

Previous research

Methodological problems of previous research

A national CATI survey with Next-Birthday selection

Registry based study: Mail Survey

- Design of the study

- Partly misleading first summary

- Qualitative Study

- Detailed analysis

Summary

References

Selection for surveys within households

- ▶ Many surveys of the general population use sampling frames based on houses or households.
- ▶ In order to draw a random sample, a procedure to sample from eligible household members is needed.
- ▶ Gaziano (2005) lists 16 different procedures for this problem. the best-known are
 - ▶ "Kish-Table"
 - ▶ "Last-Birthday"-selection
 - ▶ "Next-Birthday"-selection

Previous research

- ▶ CATI surveys nearly always use a Last- or next-Birthday-procedure, since there is – limited – evidence for higher refusal rates due to Kish-tables
- ▶ Very few – and only american – studies have been published. Gaziano (2005) listed 16.
- ▶ Most studies compare marginal distributions of demographic variables of subsamples or compare marginals of samples with population data.
- ▶ Gaziano (2005) summarizes: “(. . .) little systematic, accessible evidence exists to guide choice of respondent selection method.”

Methodological problems of previous research

1. Since all studies are american CATI surveys, we don't know anything on cultural effects on selection error.
2. The probability of a selection error in the american studies varies between 9 - 35%. The causes of variation are unknown in detail.
3. Comparing distributions of demographic variables gives no insight into the mechanisms of selection error.
4. No study is based on a respondent independent process of determining the selection error.
5. Therefore, we conducted 3 studies:
 - ▶ A national CATI-survey
 - ▶ A community mail survey based on registries
 - ▶ A qualitative survey of a random sample of households with selection error in the mail survey (Ziniel 2003)

A national CATI survey with next-birthday selection

- ▶ Between January und February 2002, 1615 respondents were selected by Next-Birthday using a national random sample of phone numbers.
- ▶ At the end of the interview, for each person between 18 and 65 years day and month of the birthday was asked.
- ▶ The following analyses are based on the sample after exclusion of 28.4% one person households.
- ▶ 1.9% (n=42) refused the month of birth. They were excluded.
- ▶ Of the remaining 1114 persone, 167 (15%) persons refused the day of birth. This persons were classified as unclear and not excluded.

Error rate

Depending on the unclear cases, the error rate was 10.1% resp. 29.5%.

Base	n	Error
unambiguously classified cases	947	11.8%
assuming no errors among the unclear cases	1114	10.1%
assuming 50% errors among the unclear cases	1114	17.5%
assuming 100% errors among the unclear cases	1114	29.5%

However, with the exception of the last estimate these estimates are at the lower bound of the US results.

Demographic variables and error rates

Variable	χ^2	df	p
Number of adults	21.01	3	0.000
Gender	0.83	1	0.363
Age	0.21	3	0.976
Education	3.62	3	0.306

- ▶ Only the number of adults has a significant effect. The effect is not linear: 9.5% (a=2), 22.8% (a=3), 19.2% (a=4) 12.5% (a=5).
- ▶ Neither CART nor logistic regressions reveal significant interactions.
- ▶ No interviewer-effects were found.

Registry based study

- ▶ Registries allow the independent assessment of selection errors.
- ▶ A special registry based study for CATI surveys would be useful, but we have no suitable sampling frame in Germany and many other countries.
- ▶ Such frames are available for mail surveys.
- ▶ Since the availability of new geo-data bases, a mail survey using only household names and not individual names as addressee is becoming a viable option to town registries.
- ▶ Therefore, we decided to do a registry based study of respondent selection in mail surveys.

Design of the study

- ▶ The town registry of a university town was used as sampling frame. All adults, which had their birthday within 10 days of january (immediately preceeding the survey) were considered eligible.³
- ▶ Faculty members and their relatives were excluded from the frame.
- ▶ A random sample of $n=2000$ was drawn.
- ▶ A TDM-survey was implemented.
- ▶ The letters were adressed to "Family XXX".

³This idea to use the registry despite the data protection obstacles is due to Sonja Ziniel.

Response rate

- ▶ After three months, we received 1081 questionnaires. 11 respondents removed the ID, 1 questionnaire was answered by two respondents. The remaining 1069 questionnaires resulted in a response rate (AAPOR (2000, definition 2) of 54.05%.
- ▶ For each designated respondent, the correctness of selection was determined by comparing gender and age of respondent and age of respondent and designated respondent.

Results

Birth year	Gender				Total
	Agree		Disagree		
Agree	798	(75.1%)	21	(2.0%)	819
Disagree	68	(6.4%)	175	(16.5%)	243
Total	866		196		1062

Data editing

- ▶ Exclusion of 235 households with only one adult
- ▶ Exclusion of 64 households with missing data on household composition variables
- ▶ Exclusion of 4 persons in nursing homes
- ▶ Exclusion of 74 persons in flat sharing communities
- ▶ 685 persons remaining

Results after editing

Selection status	Percentage
Correct	67.45 (n=462)
Incorrect	32.55 (n=223)
Total	(n=685)

Characteristics	Percentage of selection inaccuracy
Number of adults	$\chi^2=23.14, df=3, p=.000$
Two	29.4
Three	33.6
Four	65.8
Five	50.0
Respondent gender	$\chi^2=.002, df=1, p=.962$
Male	32.7
Female	32.5
Respondent age	$\chi^2=9.43, df=5, p=.093$
<30 years	32.1
30 to 39 years	25.0
40 to 49 years	30.2
50 to 65 years	34.3
60 to 69 years	41.6
70 years and older	36.8
Respondent education	$\chi^2=2.14, df=2, p=.344$
< 10 years	36.8
10 years	31.2
13 years	30.4

Characteristics	Percentage of selection inaccuracy
Designated respondent age	$\chi^2=29.03, df=5, p=.000$
< 30 years	51.8
30 to 39 years	27.3
40 to 49 years	31.3
50 to 59 years	20.7
60 to 69 years	31.3
70 years and older	36.8
Designated respondent gender	$\chi^2=.7.26, df=1, p=.007$
Male	37.7
Female	28.5

Partly misleading first summary

- ▶ Neither gender, education or age of respondents seem to influence the probability of a selection error.
- ▶ Gender of designated respondent has a significant effect: The probability of a selection error for female designated respondents is lower.
- ▶ Age of the designated respondent has a significant effect: Very young or very old respondents have higher probabilities of selection error.
- ▶ Only one of the 19 dependent variables (fear of bag snatching) showed a significant difference between households with and without selection error.
- ▶ Despite the large amount of selection error the effect on dependent variables seems to be small.

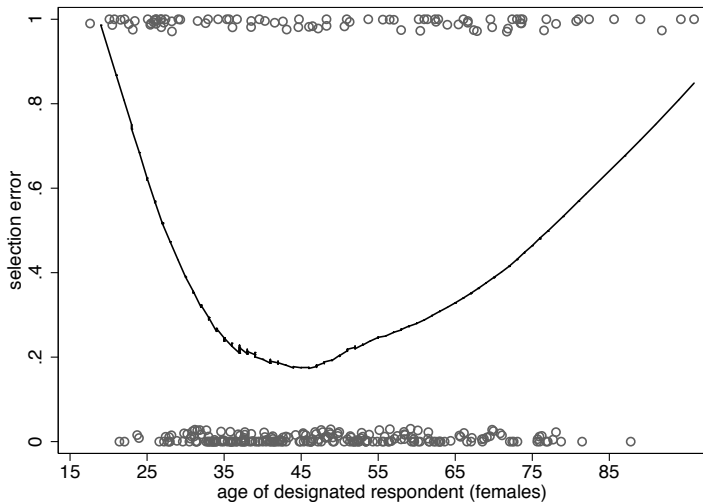
Qualitative Study

- ▶ Ziniel (2003) did qualitative interviews with 30 of 48 randomly selected households with selection error.
- ▶ Some persons responded, because the designated respondent refused and the survey seems to be important.
- ▶ In most cases the selection error was due to household specific division of labor and situational demands.
- ▶ For example, even if labor was divided according to education the actual task was allocated according to available time.
- ▶ Finally, there seems to be effects of cohort specific division of labor.

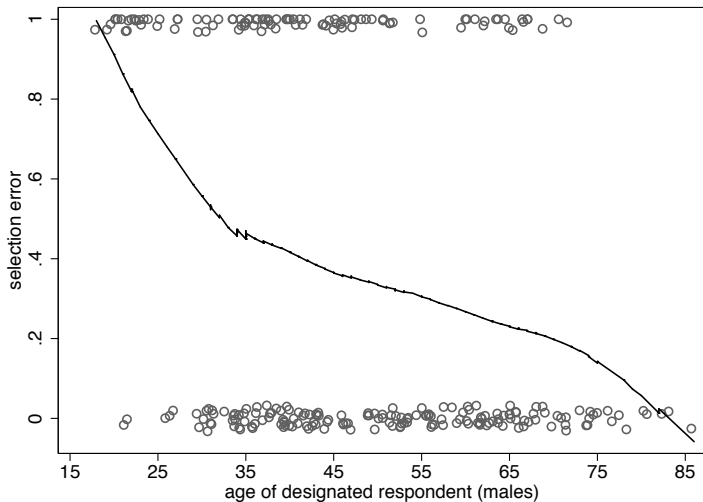
Detailed analysis

- ▶ The probability of a selection error is highest for the youngest designated respondents (< 30 years)
- ▶ Among the 59 households with a designated respondent < 30 and a selection error, 72.9% of the age differences between respondent and designated respondent is > 17 years.
- ▶ In these cases a parent may have answered the questionnaire.
- ▶ Neither the standard analysis nor the qualitative study showed any evidence for that.
- ▶ Although a gender and age effect was evident from the standard analysis, this complicated interaction effect is surprising.

Women: Selection error by age of designated respondent(lowess)



Men: Selection error by age of designated respondent(lowess)



- ▶ Among the women as designated respondent, a U-shaped relationship between age and probability of selection error was observed.
- ▶ Among the men as designated respondent the probability of a selection error decreases with age.
- ▶ Therefore, expecting a high number of interaction effects we preferred to estimate different models for the subgroups.

Logistic regression of selection error for male designated respondents older than 29 years

	Coeff.	SE	z	p
DR age	-.604	.207	-2.92	0.003
Rr age	.253	.205	1.23	0.217
number of adults	.532	.248	2.14	0.032
DR education	-.298	.220	-1.35	0.177
R education	.392	.228	1.72	0.086
constant	-.877	.945	-0.93	0.354

n=240, LR $\chi^2 = 20.95$, df=5, p= 0.001, Pseudo $R^2 = 0.069$

Logistic regression of selection error for female designated respondents older than 29 years

	Coeff.	SE	z	p
DR age	-.070	.199	-0.35	0.727
R age	.569	.208	2.73	0.006
number of adults	.396	.217	1.82	0.068
DR education	.476	.244	1.95	0.051
R education	-.729	.219	-3.32	0.001
constant	-3.483	1.014	-3.43	0.001

n=276, LR $\chi^2 = 39.93$, df=5, p= 0.000, Pseudo $R^2 = 0.131$

Summary

- ▶ Analysis of selectivity bias have to consider household composition.
- ▶ Selection within households is due to different processes.
- ▶ The effects of these processes may enhance or compensate the effects of each other.
- ▶ The effect of selection error therefore will depend on topic and population of interest.
- ▶ Comparing demographical variables of respondents with population distributions is insufficient.
- ▶ Qualitative studies of subsamples may be used to find possible explanations of erros or residuals.

References

- ▶ Gaziano,C. (2005): Comparative Analysis of Within-Household Respondent Selection Techniques, in: Public Opinion Quarterly, 69, S.124-157
- ▶ Rizzo,L./Brick,J.M./Park,I. (2004): A Minimally Intrusive Method for Sampling Persons in Random Digit Dial Surveys; in: Public Opinion Quarterly, 68, S.267-274
- ▶ Ziniel, Sonja (2003): The external validation of the Last-Birthday Method as respondent selection method in mail surveys. Master's thesis, Universität Konstanz, Department of Public Policy, Germany.