

# Rational Interviewer Behavior and Data Quality

Josef Brüderl  
Bernadette Huyer-May  
Claudia Schmiedeberg

VIU-Seminar 2012

November 28, 2012



## Introduction

- Rational interviewer behavior and data quality
  - Hypothesis: data quality is negatively affected
  - Not much studies in the literature
    - Schnell (2012) has one page on this topic
- Lessons from the German Family Panel (pairfam)
  - Usually survey researchers have no interest to talk about problems with their data produced by rational interviewers
- Contents of the talk
  - Number of alteri generated by name generators
  - Consent for interviewing secondary respondents
  - Questions on fertility intentions

- Piece-rate pay
  - Commercial survey agencies in Germany pay interviewers per interview completed
  - E.g., 30,- € in the Family Panel (including contacting and traveling)
- This creates incentives to do it the “rational way”
  - To fake the interview
    - Until now no faked interviews have been found in the Family Panel
  - To shorten the interview
    - Answer filter questions so as to skip follow-up questions
    - Abbreviate question texts

- Annual panel survey
  - 14 waves 2008 - 2021
  - 3 birth cohorts: 1971-73, 1981-83, 1991-93
  - Random sample from the population registers
  - N ~ 4000 for each cohort
  - One hour CAPI/CASI interviews
- Network module in waves 2 and 4
  - 10 minutes module at the end of the questionnaire
- Secondary respondents (partners, parents, children)
  - Primary respondents: anchors
  - Consent for sending anchor's parents a questionnaire is asked for since wave 2
- Each wave there is a fertility module

Number of interviewers in wave 2	341
Females	43%
Experience from wave 1	83%
Mean age	58.9 years (range: 24-83)
Mean number of interviews per interviewer	26.6 (range: 1-137)

- Name generators
  - With whom do you share personal thoughts and feelings or discuss things that you would not discuss with just anyone?
  - Who do you meet regularly for activities, e.g. sports, when you go out (cinema, dancing), or when you just want to talk to someone?
  - Who helps you whenever you need information or concrete advice in practical matters?
  - With whom do you occasionally have quarrels or conflicts?
- Rules for the name generators
  - Per name generator max. 30 alteri can be named
  - Each alter can be named in all 4 generators
    - Max. 120 network relationships
- For 8 randomly picked alteri descriptor-questions are asked

# The Name Generators in Detail

With whom do you share personal thoughts and feelings or discuss things that you would not discuss with just anyone?

Already named: Linda, David, Lea

Name Person Nr.4:

- No other person
- Don't know
- No answer

Whom of the named people do you meet regularly for activities, e.g. sports, when you go out (cinema, dancing), or when you just want to talk to someone?

- Linda
- David
- Lea
- Tim
- None of them
- Don't know
- No answer

Which other people do you meet regularly for activities, e.g. sports, when you go out (cinema, dancing), or when you just want to talk to someone?

Already named: Linda, David, Tim

-> Int.: Personen müssen mindestens 10 Jahre alt sein!

Name Person Nr.5:

- No one
- Don't know
- No answer

...

# Interviewer Effects (in Wave 2)

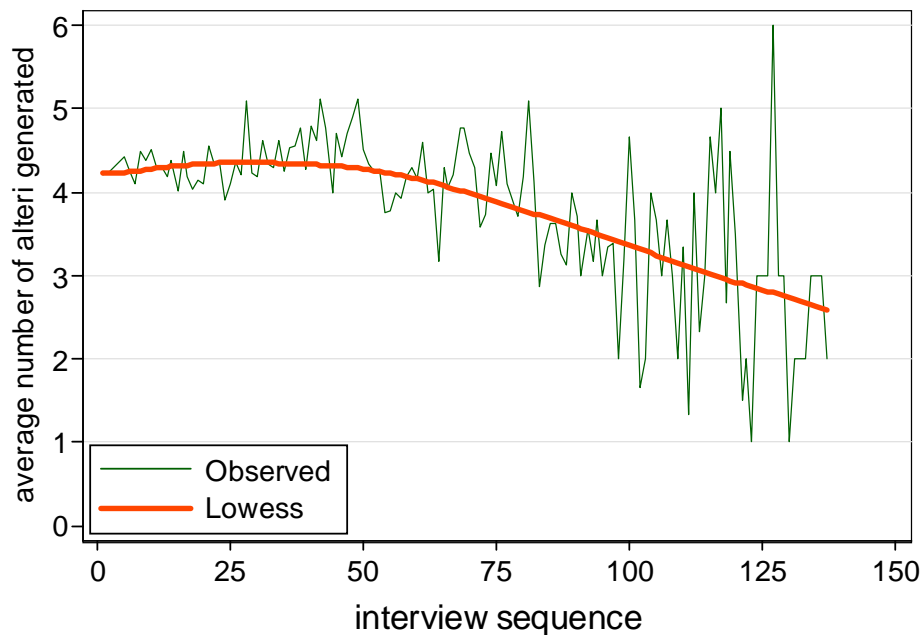
- Rational interviewer behavior
  - Collecting the names of alteri is time consuming
  - For max. 8 alteri descriptor questions follow in a second round
  - **Too few alteri/relationships generated**
- The numbers generated seem to be very low

	Range	Mean
Alteri	0-30	4.3
Relationships	0-75	9.1

- Interviewer effects in number of alteri generated
  - Bivariate intra-class correlation (ICC)
    - Van Tilburg (1998): 0.21
    - Marsden (2003): 0.15
    - pairfam: **0.40**

# Interviewer Learning?

- Do interviewers learn how to avoid work?



This replicates a finding by Matschinger et al. 2005

# Interviewer Learning?

## Multi-level linear regression on number of alteri

	(1) bivariate	(2) controlling for age
Interview sequence number	-0.004* (-2.28)	0.004* (2.23)
Age of anchor		-0.052*** (-16.70)
<i>N</i>	9069	9069

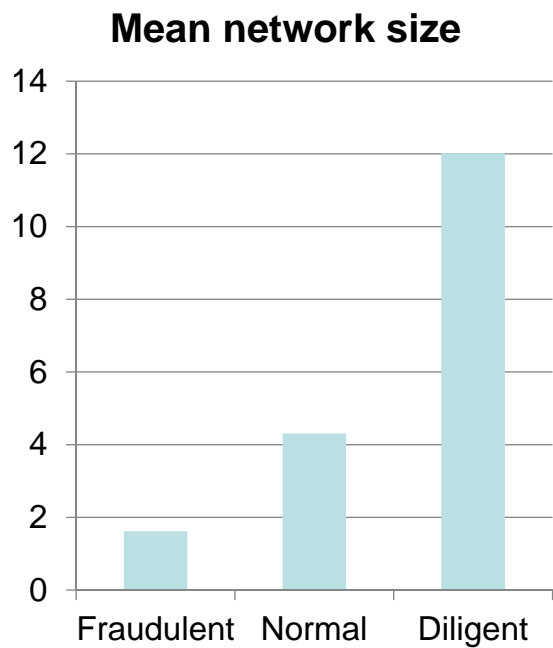
*t* statistics in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Reason: The younger anchors are easier to contact, and are therefore interviewed first. Later on older anchors with smaller networks dominate.

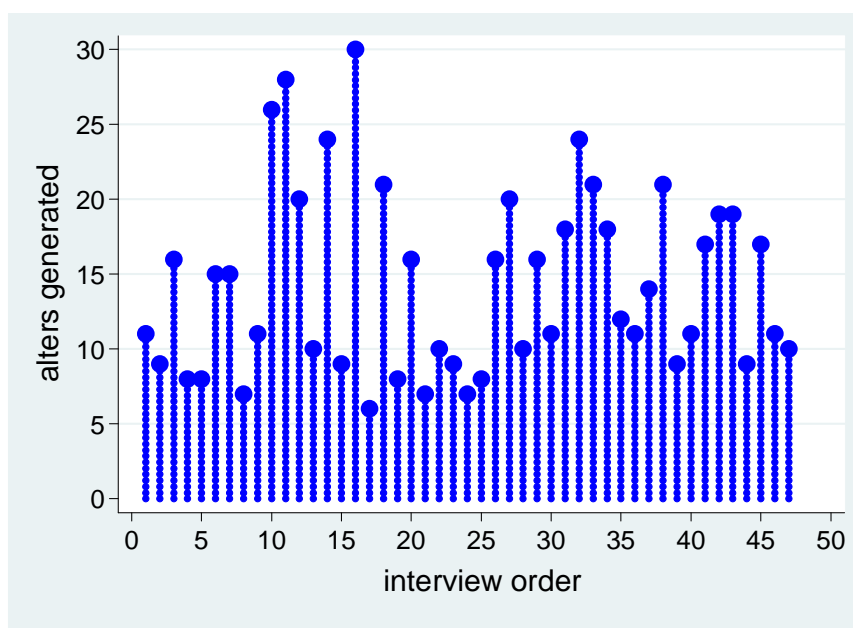
## Different Interviewer Types?

- Who are the rational interviewers?
  - Jackknifing the ICC
    - Leaving out all interviews of one interviewer and register how the ICC changes
  - 5% interviewers, who affect the ICC most
    - Below average networks („fraudulent“) [N=7]
    - Above average networks („diligent“) [N=6]
  - 95% rest („normal“) [N=328]



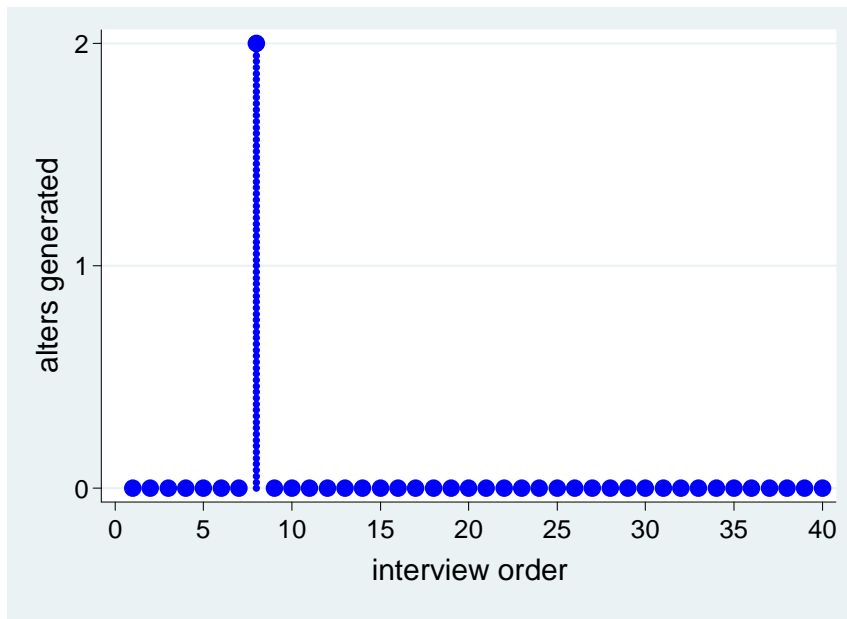
## Different Interviewer Types

A diligent interviewer (No. 1300)



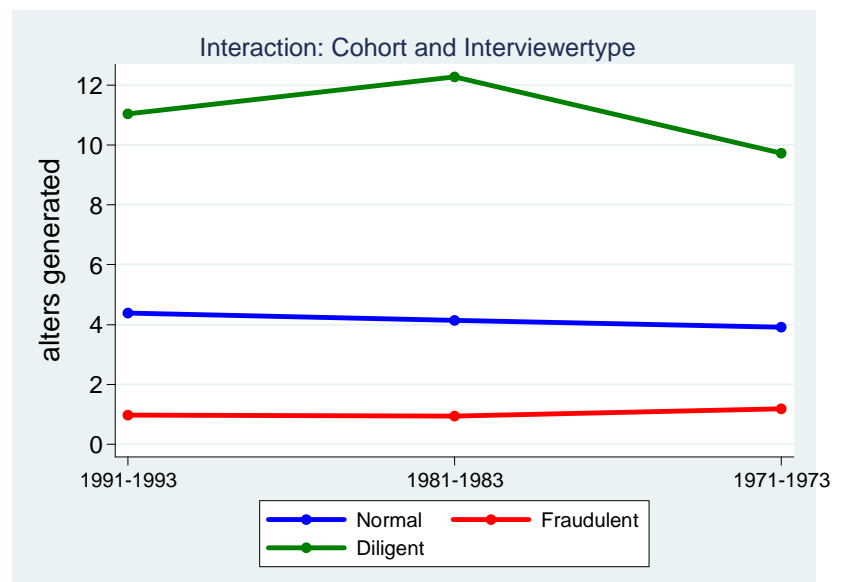
# Different Interviewer Types

A fraudulent interviewer (No. 1099)



# Does it Matter?

- Does it matter in substantive analyses?
  - Are effect estimates on network size affected?
- Example: age effect
  - Regression with interaction terms cohort x inttype
  - Diligent: inverted U-shaped pattern of the age effect
  - Normal/fraudulent: negative age effect



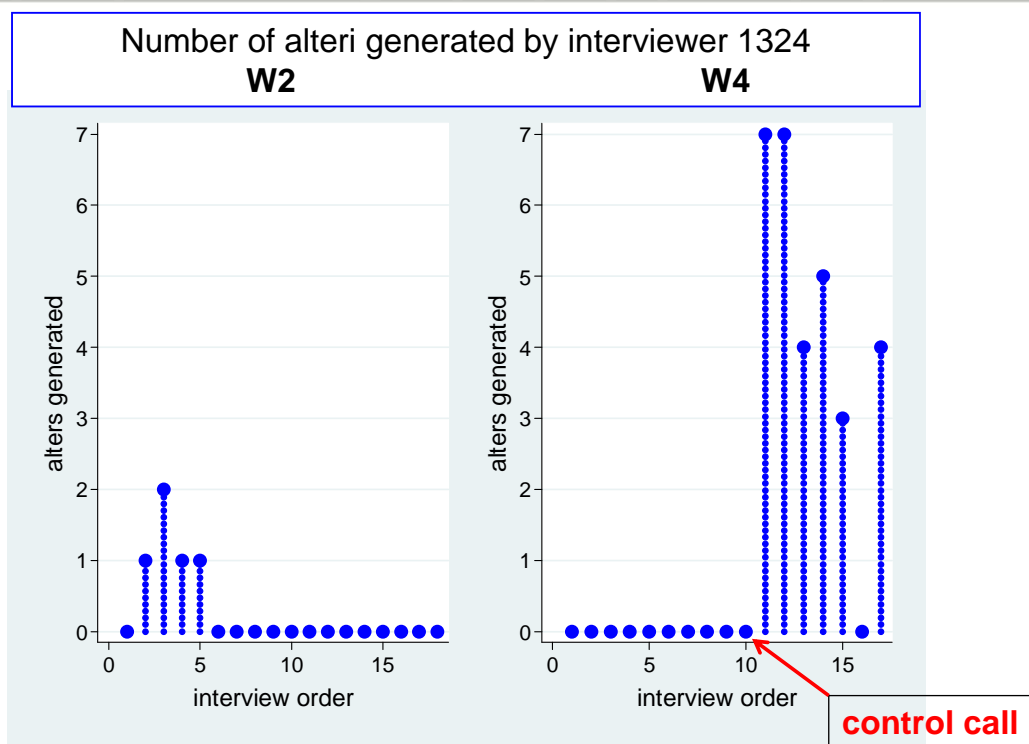
## Did It Work in Wave 4?

- Remedies taken in wave 4
  - Instructions emphasizing the importance of the network module in the interviewer manual
  - Electronic message to all interviewers communicating the importance of collecting complete network data in the beginning of the field period
  - Control mechanism during the field time: Interviewers who entered 0-1 alters several times were contacted and asked for explanation

- Did it work?

	Mean network size	ICC
Wave 2	4.3	0.40
Wave 4	4.9	0.39

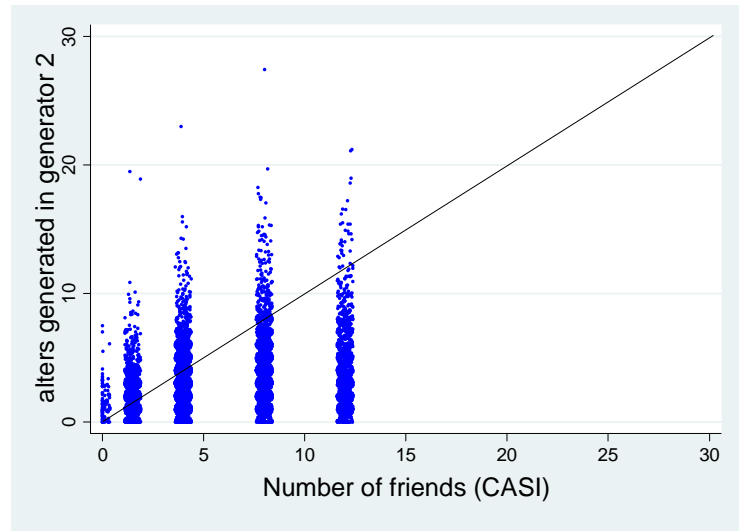
## Sometimes It Worked





## Most Time It Did Not Work

- Control question on the number of friends in the CASI module
  - No automatic checking in the network module
  - Ca. 50% of all networks have less alters than friends named in the control question
  - 15% more, 35% equal size



## Anchors' Consent to Parents Survey

- Since wave 2 anchors' are asked for consent to send a questionnaire to their (step-)parents (max. 3 parents)
  - If the anchor says yes
    - Parents' address has to be noted
    - If parents live in anchors' HH no address has to be noted

Mean duration in W2 in minutes	
No consent	0.65
Consent to 1 parent	2.08
Consent to 2 parents	2.36
Consent to 3 parents	3.91

- Time consuming

→ Interviewers have an incentive to skip the module

→ A low consent rate follows

→ Consent rate is highest for parents living in anchors' HH

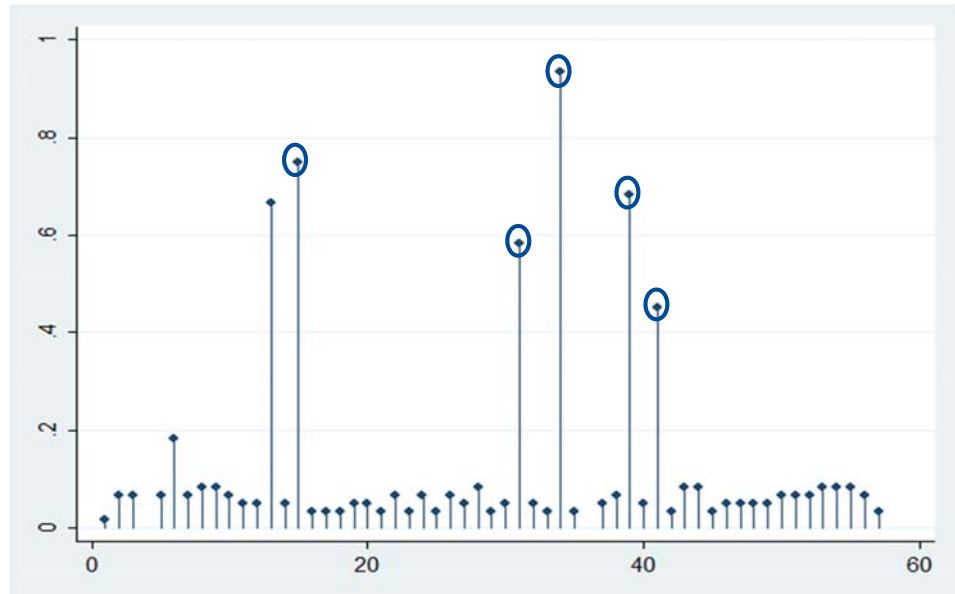
## anchors' Consent to Parents Survey

- Again we see "fraudulent" interviewers
  - Interviewer No. 1110 (W3)

55 interviews

5 times consent

Mean module duration 7.5 sec.



## anchors' Consent to Parents Survey

- Very low consent rates

	wave 2	wave 3	wave 4
Mother	55%	46%	44%
Father	46%	38%	35%
Step father	37%	28%	27%
Step mother	21%	17%	12%

- Consent rates are double as high for cohabiting parents

	Parent living in the same house	
	yes	no
wave 2	67%	34%
wave 3	57%	28%
wave 4	55%	27%

- Measures to achieve higher consent rates
  - W2-W4 2 € per questionnaire sent back (did not work)
  - Interviewer control since W4 (did not work)

# Fertility Intentions Wave 1

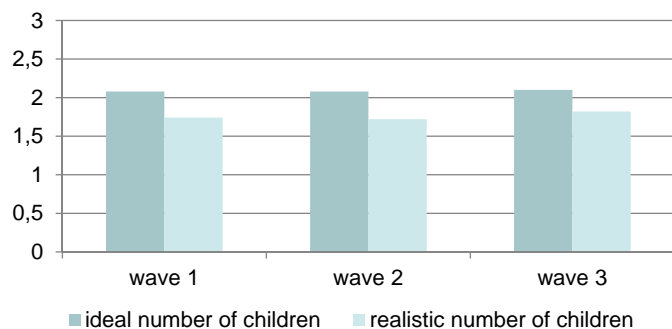
- „Ideal“ number of children
  - Wenn Sie einmal alle Hindernisse außer acht lassen: Wie viele Kinder würden Sie im Idealfall insgesamt gerne haben?
  - Int.: Es geht um die Gesamtzahl, einschließlich bereits vorhandener Kinder.
- „Realistic“ number of children
  - Wenn Sie einmal realistisch über eigene (weitere) Kinder nachdenken: Wie viele (weitere) Kinder denken Sie, werden Sie haben?
  - Int.: Gemeint sind weitere Kinder außer den schon vorhandenen.

Question text  
Interviewer Instruction

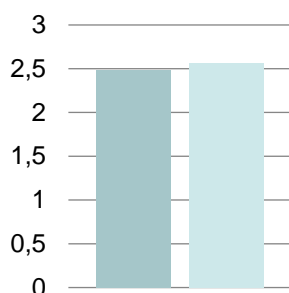
# Fertility Intentions Wave 1

- For childless anchors the questions worked well
  - Realistic number is lower

respondents without children



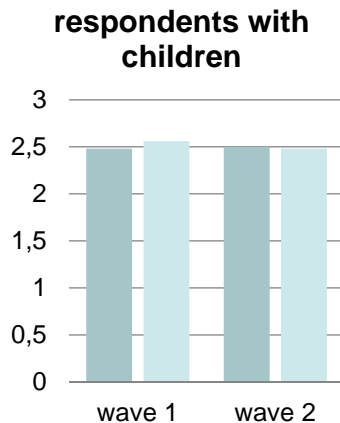
respondents with children



- Not so for anchors with children
  - Realistic number is even higher!
- What happened?
  - Anchors overheard the „weitere“
    - Because in the question before the „total number“ was asked for, this stayed the frame
  - Interviewers didn't spell out the instruction

## Fertility Intentions Wave 2

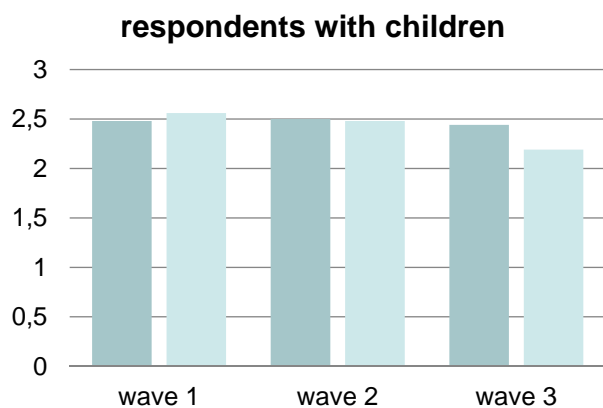
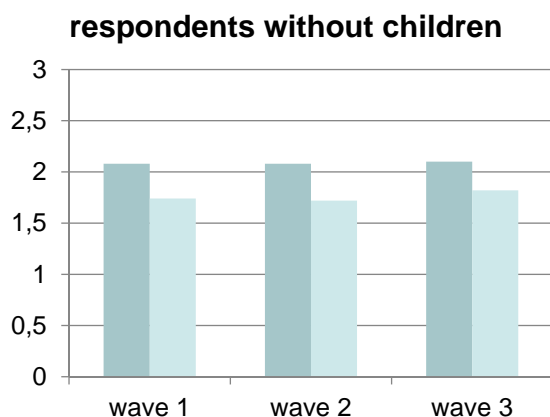
- „Realistic“ number of children in wave 2
  - Wenn Sie einmal realistisch über (weitere) Kinder nachdenken: Wie viele (weitere) Kinder denken Sie, werden Sie haben? Gemeint sind weitere Kinder außer den bisherigen.



- Again it did not work
- What happened?
  - Anchors again overheard the „weitere“
    - Framing effect
  - Interviewers did not read out the second sentence

## Fertility Intentions Wave 3

- „Realistic“ number of children in wave 3
  - For those with children a filter construction was implemented
    - Wenn Sie einmal realistisch über weitere Kinder nachdenken: Denken Sie, dass Sie weitere Kinder zusätzlich zu Ihren bisherigen Kindern haben werden?
    - If yes: Wie viele weitere Kinder zusätzlich zu Ihren bisherigen Kindern denken Sie, werden Sie haben?



- Be careful when analyzing network data
- Rational interviewer behavior has to be taken into regard when selecting/training/controlling interviewers
  - Use only diligent interviewers
    - Do the survey yourself
    - Invest in the motivation of the interviewers
  - Set the incentives right
    - Pay interviewers by the hour
    - Control and sanction the interviewers
  - Do a telephone survey where interviewers are paid by the hour and can be controlled easily
- Rational interviewer behavior has to be taken into regard when constructing a questionnaire
  - Foolproof your questions so that they can be understood without interviewer instruction and in one sentence