Mode Effect or Question Wording?
Sources of Measurement Error in Mixed Mode Surveys

Edith de Leeuw & Joop Hox

Department of Methodology & Statistics, Utrecht University

Annette Scherpenzeel & Stephanie Stam

CenterData, University of Tilburg

Robbert Zandvliet TNS-NIPO

Mixing Modes

❑ Mixed-Mode Surveys combine data, but

   ❑ Are responses between modes comparable?

❑ Modes differ on many mode-inherent factors, e.g.

   ❑ Information transmission

      ❑ Audio vs visual vs both

   ❑ Privacy

      ❑ Interviewer present or not

      ❑ Computer vs paper

   ❑ Interactivity

      ❑ Interviewer help or not

      ❑ Feedback

Question Design

❏ Numerous studies have shown that modes differ very slightly when comparable questions are used, but

❏ Question format influences answers given

  ❏ Response options, question wording, etc

    ❏ Early overview Sudman et al 1974

    ❏ Effects found even within one mode

❏ Different modes have different traditions

  ❏ Unfolding or two-step (telephone)

  ❏ Complete question one-step (web, mail)

  ❏ Fully labeled (face-to-face, mail)

  ❏ Endpoint labeled (telephone)

  ❏ Explicit do-not-know (mail) vs not offered (interview)

❏ May data be combined?

Mode or Question?

- Web surveys hybrid of paper mail & interview
  - Self-administered, but interactive
    - Help may be offered, feedback given
  - Computer-assisted
    - More question formats possible
- Great opportunity to investigate question formats in different modes: CATI-CAWI
  - Inspired by Dillman, Christian & Smyth
    - Question effect
      - Compare different formats within mode
    - Mode effect
      - Compare different formats across modes
    - Robustness
      - Which questions give the smallest effects across modes

Mode Experiments

- Series of Mode Experiments in Spring 2009
  - CATI vs CAWI
  - Members Dutch LISS-panel (CenTERdata)
    - Probability based panel, nationwide, established in 2007
      - Recruitment based on random sample of addresses (Statistics Netherlands)
      - Recruitment interviews face-to-face or telephone
      - Internet access no prerequisite
        - Free SimPC and web access offered to those with no access
  - Random assignment to mode
    - CATI: assigned 2000, responded 1207 (60%)
    - CAWI: assigned 6134, responded 4003 (65%)

Experiment A
CATI probes vs emulating Interviewer Probes Online
inspired by Wine 2006

- ❑ Six questions on use of embryos in biomedical research ("Are they used for research on…")
  - ❑ Response options yes/no/(DK)
  - ❑ Questions used earlier, elicited many DK responses
- ❑ Two experimental conditions
  - ❑ DK not offered (but possible)
  - ❑ DK explicitly offered
- ❑ After a DK response, *always* a probe
  - ❑ "Thank you. I have noted down your response. Could you possibly tell us a preference for Yes or No…"
  - ❑ Response options Yes/No/DK

Results In Sum
See Appendix 1 for detailed Results

❑ Dependent variables: number of DK responses on the six questions, before/after probing

❑ Before: Effect of mode, of explicit offering, and interaction!

    ❑ If DK is not offered, this results in more non-informative answers in CATI than CAWI (average CATI 0.29 vs CAWI .05)

    ❑ If DK is explicitly offered, more non-informative answers in CAWI than CATI (average CATI 0.91 vs CAWI 1.32)

❑ After: Effect of mode, probing, and interaction with mode

    ❑ If DK is not offered, probing reduces non-informative answers (average CATI now 0.12 vs CAWI 0.02)

    ❑ If DK is explicitly offered, strong reduction by probing (average non-informative answers CATI 0.28 vs CAWI 0.39)

❑ Probing Effective in Web!

Experiment B
Questions & Format

- Eight questions on acceptability of usage of advanced medical technology
  - "If it will save a live, everything is permitted"
    - Five point scale: Totally Agree….Totally Disagree
    - Do-not-Know not offered, but accepted when given
- Two question format experiments
  - Unfolding (two step) vs Direct Question (one step)
  - Fully Verbally Labeled vs Endpoints Only Labeled Verbally
- Dependent variables
  - Mean score and Extremeness

Results In Sum
See Appendix 2 for detailed Results

- ❑ Mean Score
  - ❑ Significant but small effect of mode and unfolding
    - ❑ Web and complete question slightly higher (more acceptable)
  - ❑ No effect of labeling,
  - ❑ No Interactions Format and Mode!
- ❑ Extremeness
  - ❑ Overall significant (small) effects of
    - ❑ Mode: CATI more extremes
    - ❑ Unfolding: Unfolding more extremes
    - ❑ Endpoint labeling: Endpoint-labeling more extremes
  - ❑ No interactions
  - ❑ Smallest effects endpoint labeling

Experiment C
Recency and Primacy

- ❑ Two questions on 'information seeking'
    - ❑ "When a newspaper or journal publishes an article on medical topics, how often do you read it?"
    - ❑ When a tv program is broadcasted on medical topics … how often..?
    - ❑ Seven point scale: Always….Never
        - ❑ Do-not-Know not offered, but accepted when given
- ❑ Question format experiment
    - ❑ Order Always-Never vs Never-Always
        - ❑ CAWI: horizontal vs vertical visual presentation
    - ❑ Dependent variable: Mean score

Experiment C
Theoretical Background

I. Questions mildly socially desirable

   ➢ State that you do read /watch are 'information seeking'

   ➢ Expect stronger tendency in CATI

      ➢ Literature: interviewer present results in more soc.des

II. Primacy vs recency

   ➢ Expect more primacy CAWI and more recency CATI

      ➢ Visual presentation vs aural presentation

         ➢ CAWI: Vertical stronger than horizontal (top is best)

Experiment C
Expectations

- Expect Interaction of Mode and Response Category Order
  - Offer ALWAYS first:
    - the social desirability should be counteracted by recency in CATI
  - Offer NEVER first:
    - the social desirability should be counteracted by primacy in CAWI

Results Exp 3
Contrary to Expectations

❑ Primacy effect also in CATI!

    ❑ CATI more socially desirable in Always-Never

    ❑ CATI less socially desirable in Never-Always

❑ Primacy effect in CAWI Vertical

    ❑ Vertical display reinforces primacy ONLY

    ❑ No effect with horizontal display

❑ WHY???

Results Experiment 3

	CATI	CAWI-HOR	CAWI-VERT	Overall mean
Q1 A–N	4.70	4.48	4.60	4.58
Q1 N–A	4.39	4.45	4.17	4.32
Q2 A–N	4.52	4.34	4.41	4.41
Q2 N–A	4.33	4.32	4.08	4.23

★ Significant primacy effects

Note: Scale 1-7, Recoded so 1 is N & 7 is A

Discussion

- Question indeed are slightly sensitive
  - Checked
    - Correlations with social desirability scale are in expected direction
- Literature primacy/recency inconclusive
  - Effects in long UNORDERED lists
- Here clear scale: always-never
  - Easier to keep in short term memory?
  - Plus trained respondents

Mode Effect or Question Wording? Appendix I

Results of First Mode Experiment:
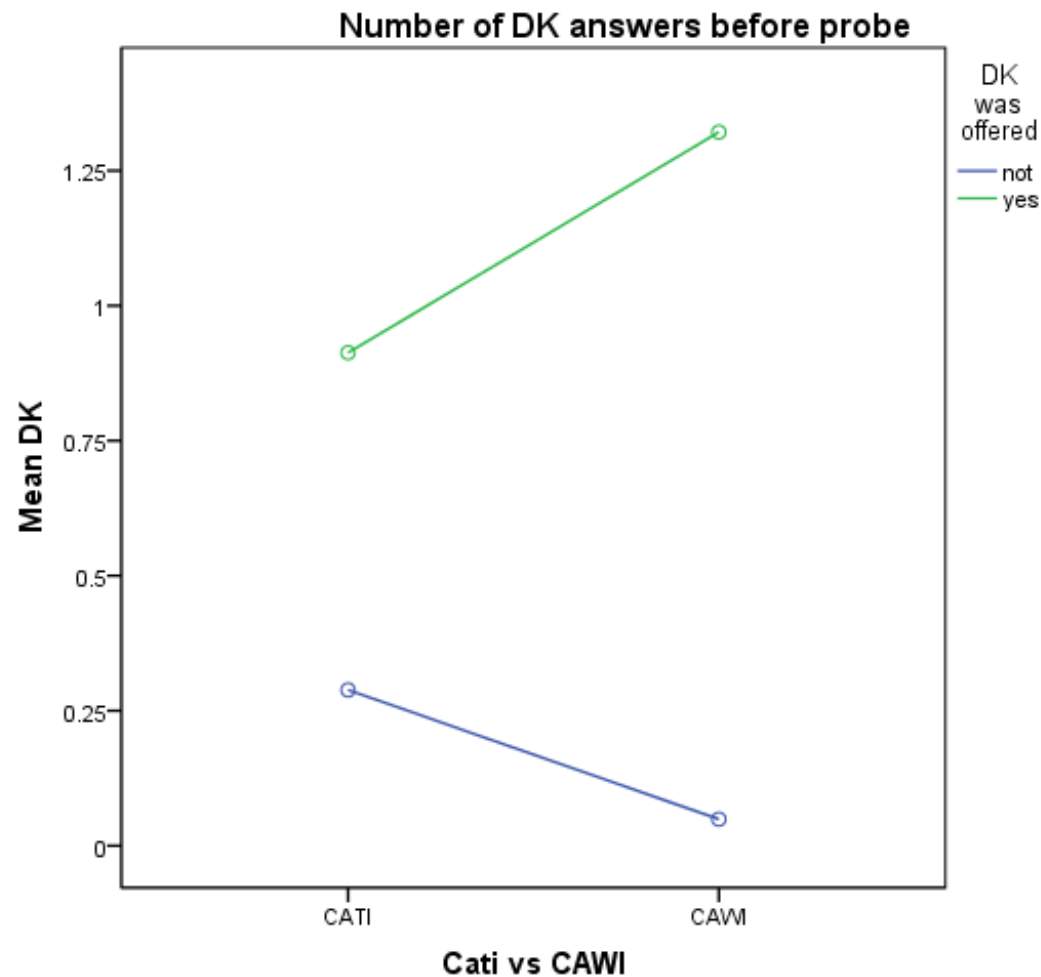Explicit offering of Do-not Know or not
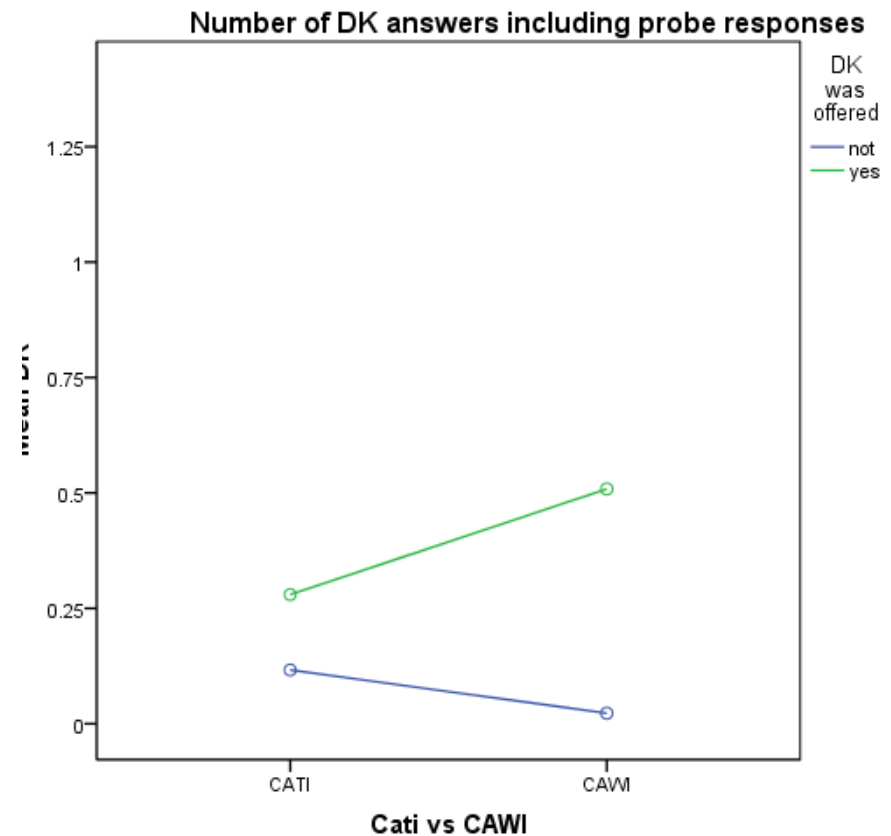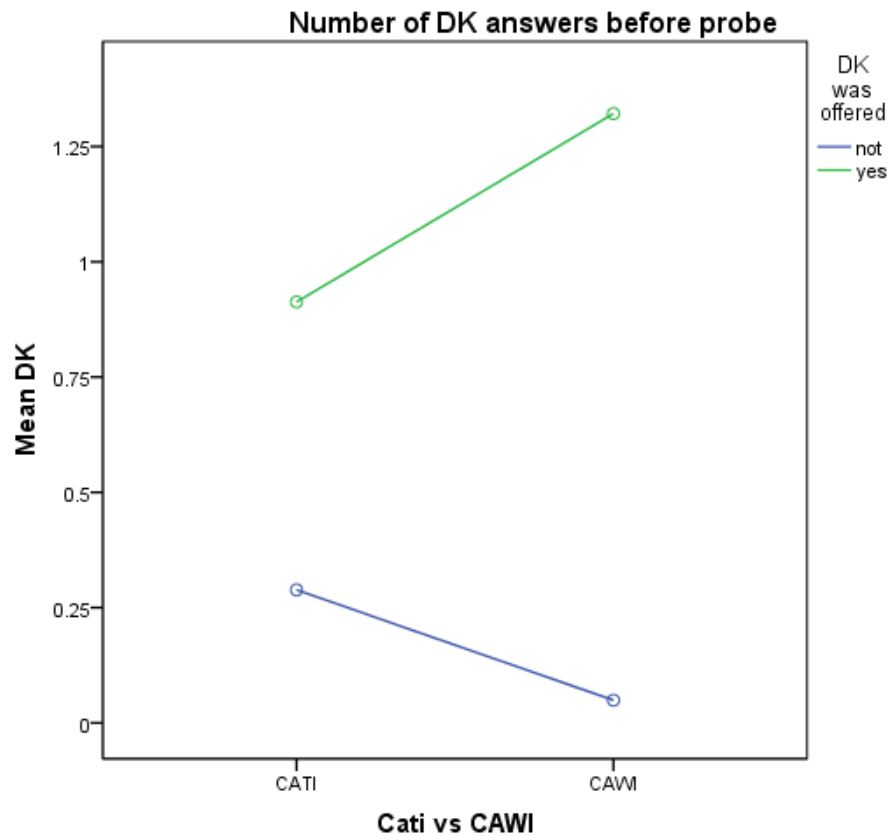Probing or Not (in both CATI and CAWI)

Don't Know & Probing

- Don't Know (DK) option in surveys is controversial
  - But generally offered if respondents may legitimately not know the answer
  - Offering DK is known to produce more DK responses
- Q1: do CATI and CAWI differ in this respect?
- Q2: does following DK with a probe help?
  - How much?
- Q3: does the effect of a probe differ between CATI and CAWI

Q1: Do CATI & CAWI differ in the effect of offering DK? Yes!

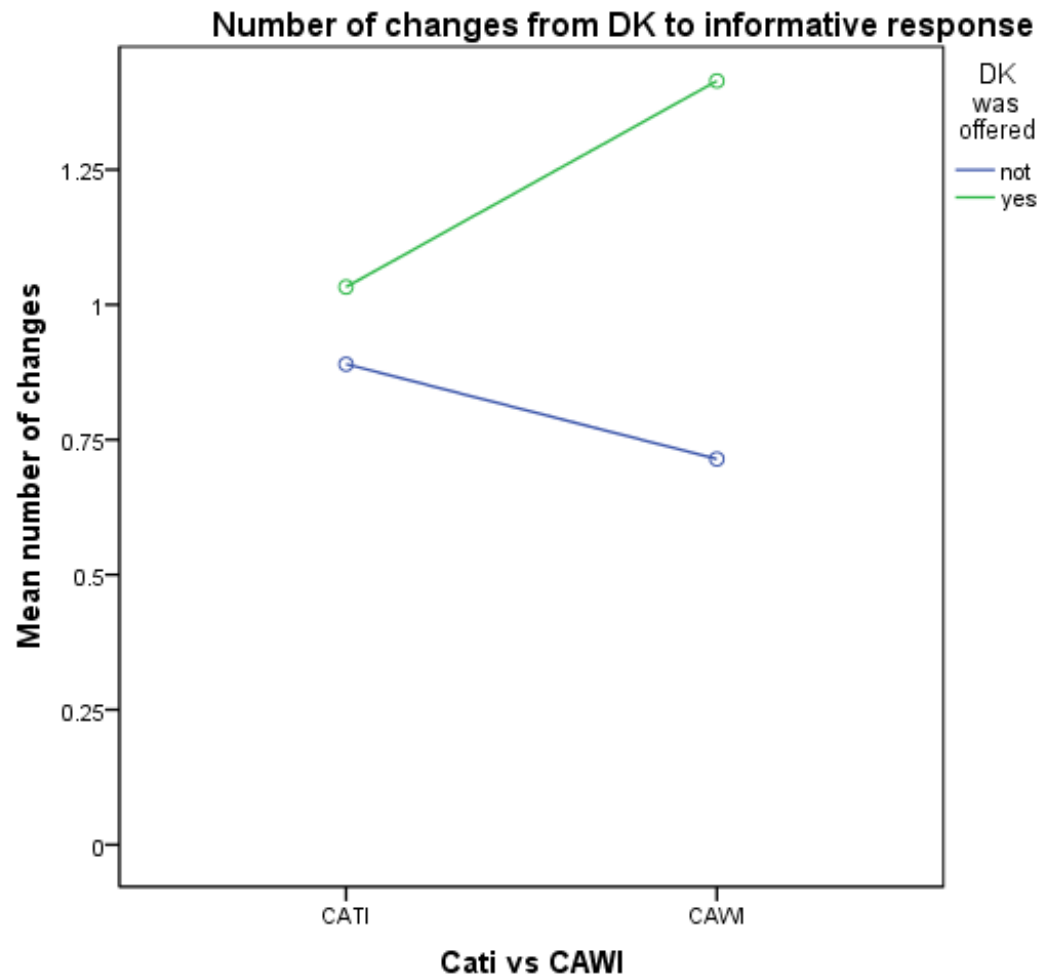# Q2: Does following DK with a probe help? Yes!

Number of DK answers before probe

Number of DK answers including probe responses

Q3: Does the effect of a probe differ between CATI and CAWI? Yes!

Number of changes from DK to informative response

Mode experiments

- Series of Mode Experiments in Spring 2009
  - CATI vs CAWI
  - Members Dutch LISS-panel (CenterDATA)
    - Probability based panel, established in 2007
      - Recruitment based on random sample of addresses (Statistics Netherlands)
      - Recruitment interviews face-to-face or telephone
      - Internet access no prerequisite
        - Free SimPC and internet offered to those with no access
  - Random assignment to mode
    - CATI: assigned 2000, responded 1207 (60%)
    - CAWI: assigned 6134, responded 4003 (65%)

Don't Know Experiment

- Six questions on use of embryos in biomedical research ("Are they used for research on…")
  - Response options yes/no/(DK)
  - Questions used earlier, elicited many DK responses
- Two experimental conditions
  - DK not offered (but possible)
  - DK explicitly offered
- After a DK response, *always* a probe
  - "Thank you. I have noted down your response. Could you possibly tell us a preference for Yes or No…"
  - Response options Yes/No/DK

Don't Know Experiment

❑ Dependent variables: number of DK responses on the six questions, before/after probing

❑ Main analysis:

    ❑ CATI vs CAWI $\times$ No DK vs DK offered

❑ Follow-up analysis: effect on data quality

Q1: Do CATI & CAWI differ in the effect of offering DK?   Yes!

Nr of DKs first time question asked

	No DK	Explicit DK	Mean
CATI	0.29	0.91	0.60
CAWI	0.05	1.32	0.69
Mean	0.17	1.12	

Mode p=.07; DK p<.00; Interaction p<.00

Q2: Does following DK with a probe help? Yes!

Nr of DKs first time question asked

	No DK	Explicit DK	Mean
CATI	0.29	0.91	0.60
CAWI	0.05	1.32	0.69
Mean	0.17	1.12	

Mode p=.07; DK p<.00; Interaction p<.00

Nr of DKs following probe question

	No DK	Explicit DK	Mean
CATI	0.12	0.28	0.20
CAWI	0.02	0.51	0.27
Mean	0.07	0.39	

Mode p=.05; DK p<.00; Interaction p<.00

Q3: Does the effect of a probe differ between CATI and CAWI? Yes!
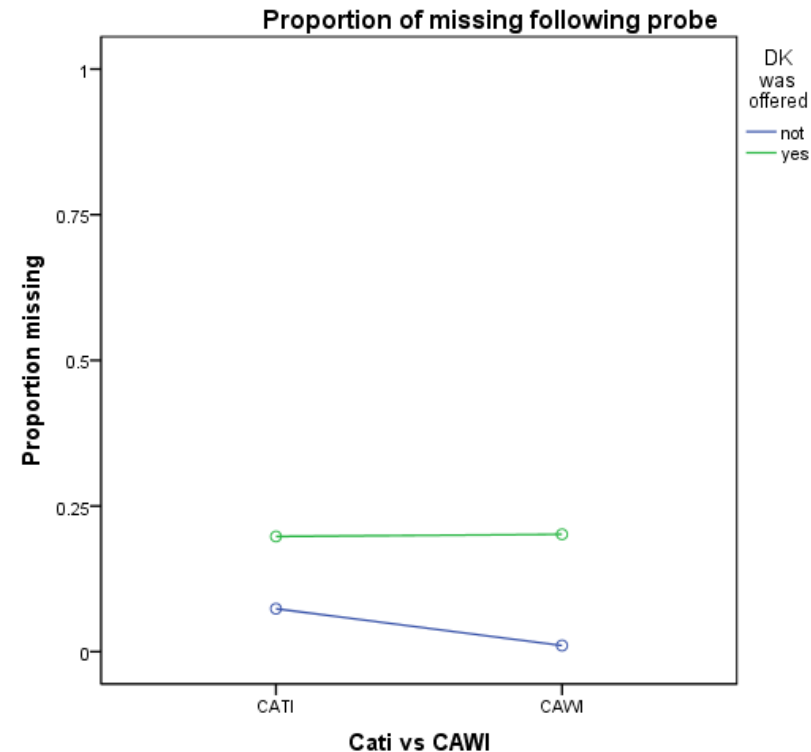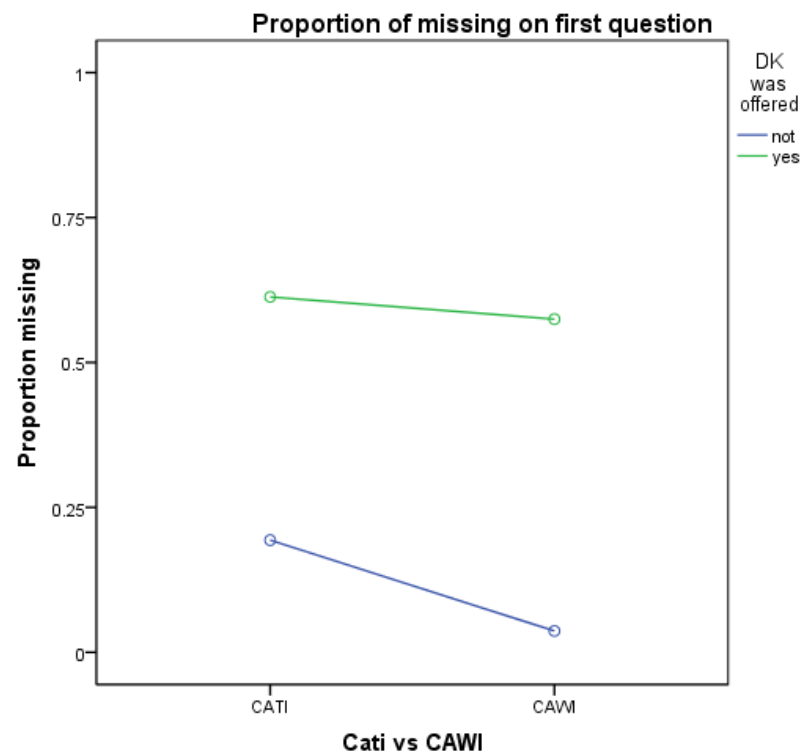
Nr of changes from DK to informative answer

	No DK	Explicit DK	Mean
CATI	0.12	0.28	0.20
CAWI	0.02	0.51	0.27
Mean	0.07	0.39	

Mode p=.42; DK p<.00; Interaction p<.03

Data quality: missing values

- ❏ Assume DK recoded to missing value
- ❏ Assume a multivariate analysis on the six questions using listwise (casewise) deletion
- ❏ What is the proportion of cases that is deleted?

Data quality: missing values

Proportion of cases deleted in listwise deletion,
first time question asked

	No DK	Explicit DK	Mean
CATI	0.20	0.61	0.40
CAWI	0.04	0.58	0.31
Mean	0.12	0.59	

Mode p<.00; DK p<.00; Interaction p<.03

Proportion of cases deleted in listwise deletion,
following probe question

	No DK	Explicit DK	Mean
CATI	0.07	0.20	0.14
CAWI	0.01	0.20	0.11
Mean	0.04	0.20	

Mode p<.00; DK p<.00; Interaction p<.01

Data quality: reliability?

Reliability six questions, first time asked (N=1117)

	No DK	Explicit DK	Mean
CATI	0.45	0.56	0.49
CAWI	0.74	0.69	0.72
Mean	0.65	0.65	

Mode p<.00; DK p<.00; Interaction p<.03

Reliability six questions, after probe (N=2894)

	No DK	Explicit DK	Mean
CATI	0.44	0.50	0.47
CAWI	0.74	0.66	0.71
Mean	0.64	0.61	

Mode p<.00; DK p<.00; Interaction p<.03

Conclusions

- ❏ When it is necessary to offer a DK option, following up DK with a probe repairs much of the damage

- ❏ Especially needed in Web interview

- ❏ Converting DK into informative answers increases useful sample size *much*

- ❏ Converting DK into informative answers does no harm to reliability of responses

Mode Effect or Question Wording?
Appendix II

Results of Second Mode Experiment:
Two-step (unfolding) vs One-step (direct) Question
Endpoint Labeled vs Fully Labeled

Mode Experiments

- Series of Mode Experiments in Spring 2009
  - CATI vs CAWI
  - Members Dutch LISS-panel (CenTERdata)
    - Probability based panel, established in 2007
      - Recruitment based on random sample of addresses (Statistics Netherlands)
      - Recruitment interviews face-to-face or telephone
      - Internet access no prerequisite
        - Free SimPC and web access offered to those with no access
  - Random assignment to mode
    - CATI: assigned 2000, responded 1207 (60%)
    - CAWI: assigned 6134, responded 4003 (65%)

Questions & Format

- Eight questions on acceptability of usage of advanced medical technology
  - If it will save a live, everything is permitted
  - It is not desirable to utilize every medical invention, just because it is technologically possible
    - Five point scale: Totally Agree….Totally Disagree
    - Do-not-Know not offered, but accepted when given
- Two question format experiments
  - Unfolding (two step) vs Direct Question (one step)
  - Fully Verbally Labeled vs Endpoints Only Labeled Verbally
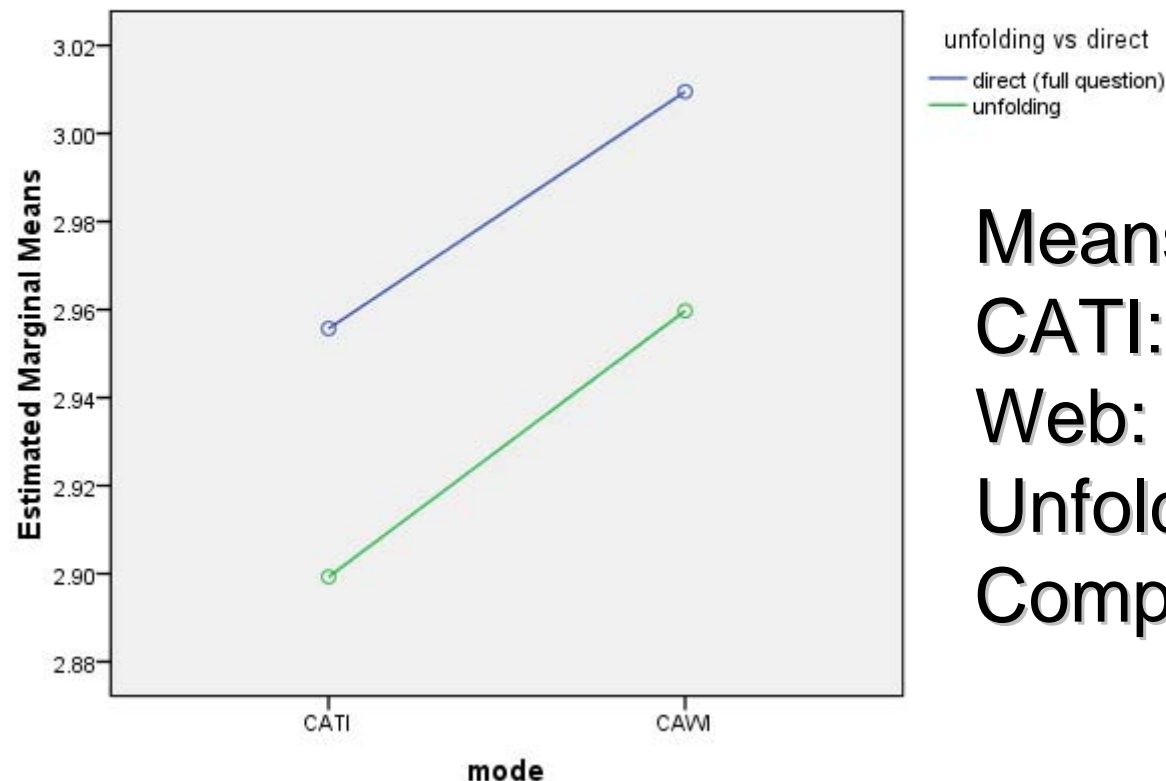
Analyses

- ❑ **Variables of Interest**
  - ❑ Independent variables (dichotomous)
    - ❑ Mode: CATI vs WEB
    - ❑ Format
      - ❑ Unfolding vs Complete question
      - ❑ Fully labeled vs End-point labeled
  - ❑ Dependent variables
    - ❑ Mean score (Conceptual of interest: what do people think about the topic)
    - ❑ Extremeness: Proportion extreme answers 1 & 5
    - ❑ Response distribution: proportion 1, proportion 2, proportion 3, proportion 4, proportion 5
  - ❑ Analyses done after appropriate recoding of responses
    - ❑ 1: most conservative in acceptance modern technology
    - ❑ 5: very prone to accept new medical technology

Mean Score on 8 Questions
Effect of Mode and Unfolding

Estimated Marginal Means of mean v3-v10 with 3, 5, 7, 10 recoded

Means Overall
CATI: 2.93
Web: 2.99
Unfolding: 2.93
Complete Q: 2.98

Significant but small effect of mode and unfolding: R^2= .006
No effect of labeling, no Interactions Format and Mode!

Differences in Means: Why?

- ❑ Two factors affect mean score (and thereby conclusions on acceptance of medical technology)
    1. Mode
    2. Unfolding
        - ➢ Note: no effect labeling & no interactions
- ❑ What may cause these shifts in mean?
    - ❑ Mean score sensitive for extremes!
    - ❑ More extreme scores?
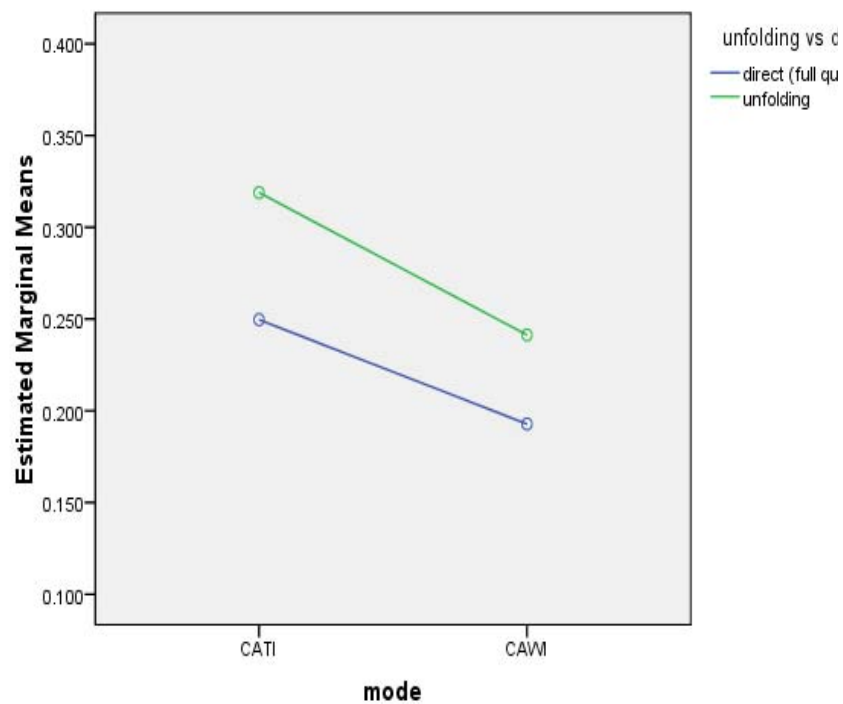   - ❑ Let's find out

Proportion Extreme Answers

❑ Extreme response

    ❑ Either 1 or 5 on five point scale

❑ Proportion 1 or 5 over 8 questions

❑ Overall significant (small) effects of

    ❑ Mode: CATI more extremes

    ❑ Unfolding: Unfolding more extremes

    ❑ Endpoint labeling   Endpoint-labeling more extremes

        ➢ All effects small (.05-.06) difference in mean proportion extreme answers, R^2=.008

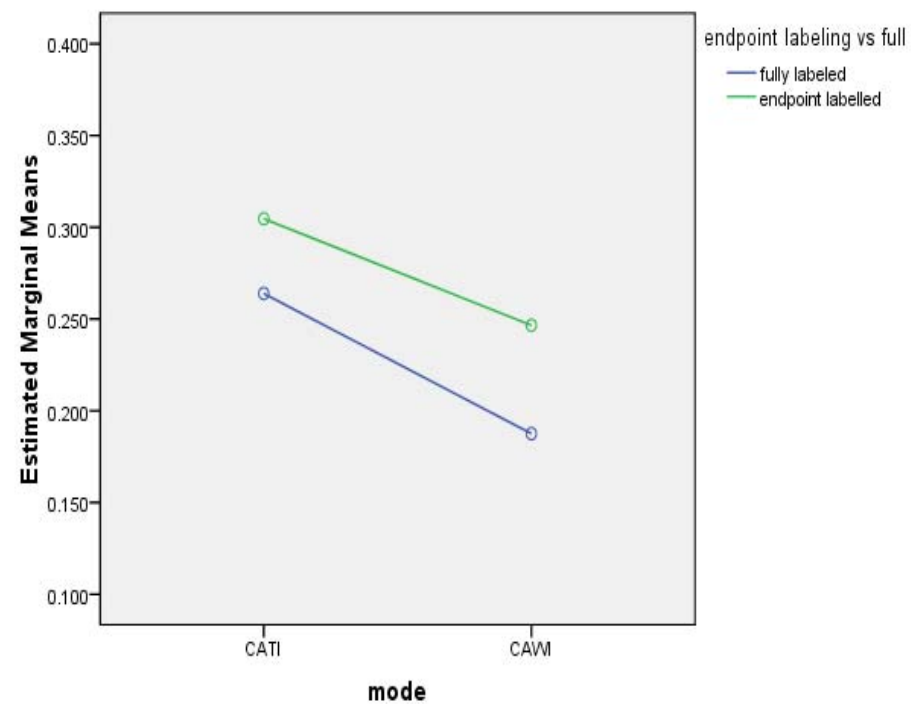   ➢ No interactions!

Extreme Answers

Estimated Marginal Means of prop. extreme values in 3-10

Estimated Marginal Means of prop. extreme values in 3-10

Shift In Responses

- Higher Overall Means:
  - CAWI
  - Complete Question (one step question)
- More Extreme Answers
  - CATI
  - Unfolding (two step question)
  - Endpoint Labeling
- What is happening in response distributions?
- Let's find out

Response Distribution

Effect of Mode, Unfolding, Endpoint Labeling Difference in proportion: Five Point Scale			
	CATI - CAWI	Full Q.- Unf	Fully - Enp
Proportion 1	+ 0.02	- 0.06	- 0.03
Proportion 2	+ 0.07	+ 0.02	+ 0.00
Proportion 3	- 0.13	+ 0.04	+ 0.04
Proportion 4	+ 0.04	+ 0.06	+ 0.01
Proportion 5	- 0.00	- 0.05	- 0.02

➤ Smallest effects with fully labeled vs endpoint labeled

➤Mode has effect, Question Format has also effect

➤if you use unfolding in CATI and complete question in CAWI: OEPS

In Sum / Conclusion

- ❑ Mode has effect
- ❑ Question Format has additional effects
  - ❑ Smallest effects endpoint labeling
- ❑ Do not UNFOLD!
- ❑ This study 5-point scales
  - ❑ How about larger scales 7-point or 9 point
    - ❑ Then telephone one step+completely labeled not feasible?
    - ❑ One step + Endpoint labeling?
    - ❑ Replication and extension in Germany (Uni Bremen)
- ❑ Advice: If possible incorporate small mode experiment to collect data for adjustment