False Positives and the “More-is-Better” Assumption in Sensitive Question Research: New Evidence on the Crosswise Model and the Item Count Technique

Felix Wolter & Andreas Diekmann

(University of Konstanz; University of Leipzig and ETH Zürich)

November 17, 2020

felix.wolter@uni-konstanz.de
andreas.diekmann@soz.gess.ethz.ch

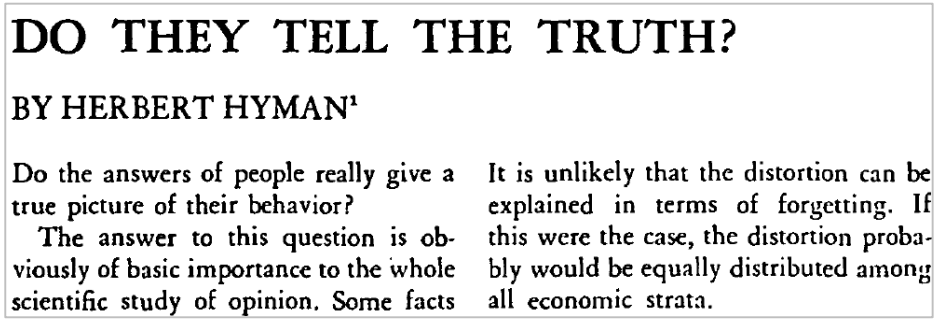
Outline

1. Background
2. State of Research
   - Empirical Evidence 1: Enthusiasm
   - Empirical Evidence 2: Spoiling the Party
3. Our Study
4. Results
5. Discussion

Background: Sensitive Questions in Surveys

- E.g., delinquency, substance abuse, health issues, sexuality, xenophoby, homophoby, voting, green behavior.

- Main problem = misreporting (e.g., Kuhn & Vivyan 2020; Locander et al. 1976; Parry & Crossley 1950; Wolter & Preisendörfer 2013):
  - Under-reporting of negatively connoted traits (e.g., shoplifting).
  - Over-reporting of positively connoted traits (e.g., voting in elections).



DO THEY TELL THE TRUTH?

BY HERBERT HYMAN[1]

Do the answers of people really give a true picture of their behavior?

The answer to this question is obviously of basic importance to the whole scientific study of opinion. Some facts

It is unlikely that the distortion can be explained in terms of forgetting. If this were the case, the distortion probably would be equally distributed among all economic strata.

Hyman 1944 in Public Opinion Quarterly.

Background

- Survey methodologists have proposed several special questioning techniques in order to tackle misreporting on sensitive questions:
  - Randomized response technique (RRT, Warner 1965).
  - Crosswise model (CM, Yu et al. 2008).
  - Item count technique (ICT, Droitcour et al. 1991; aka list experiment, unmatched count technique).

- Our hope is that the enhanced anonymity and reduced embarrassment induced by these techniques make repondents answer truthfully.

Background

- Survey methodologists have proposed several special questioning techniques in order to tackle misreporting on sensitive questions:
  - Randomized response technique (RRT, Warner 1965).
  - Crosswise model (CM, Yu et al. 2008).
  - Item count technique (ICT, Droitcour et al. 1991; aka list experiment, unmatched count technique).
- Our hope is that the enhanced anonymity and reduced embarrassment induced by these techniques make repondents answer truthfully.

The Crosswise Model (CM)

- **Question A:**
  Is your mother‘s birthday in January or February: yes or no?
  (*If you do not know, please choose the birthday of someone elso you know).*

- **Question B:**
  Have you ever taken cocaine: yes or no?

- Compare your responses to questions A and B: Are they equal or unequal?

  ❒ – equal (both „yes“ or both „no“)
  ❒ – unequal (once „yes“ and once „no“)

The Item Count Technique (ICT)

- Random split of the sample into (at least) two groups, a short-list group and a long-list group.
- Respondents indicate only the number of items that apply.
- A prevalence estimate of the sensitive item can be calculated by subtracting the mean of the short list from the mean of the long list.

Short-list group

- Have you ever been abroad?
- Have you ever used a taxi?
- Have you been using a plane this week?
- Did you wash your car this week?

Long-list group

- Have you ever been abroad?
- Have you ever used a taxi?
- Have you been using a plane this week?
- Did you wash your car this week?
- Have you ever been driving a car although you had drunk too much alcohol?

(Wolter & Laier 2014)

Outline

1. Background
2. **State of Research**
   - **Empirical Evidence 1: Enthusiasm**
   - **Empirical Evidence 2: Spoiling the Party**
3. Our Study
4. Results
5. Discussion

Empirical Evidence 1

- Main question: Do these questioning techniques actually work?

- Main approach: Experimental „more-is-better“ (MiB) studies:
  - Randomization: Direct questioning (DQ) vs. CM/ICT.
  - „More is better“: Higher estimates for undesirable traits are taken as more valid („better“).
  - Example (Wolter & Laier 2014):

	% DQ	% ICT
Fare dodging	66	80
Driving w/o licence	34	49
Drunk driving	46	68

- Also: External validation studies with known true values (very rare).

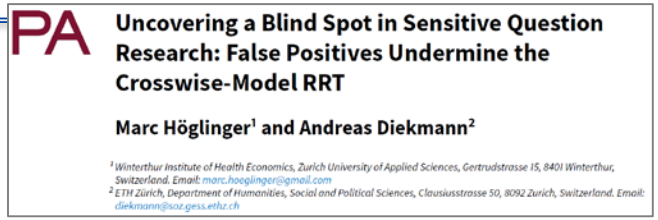
Empirical Evidence 1

- Meta-studies show encouraging results in the sense that, using the MiB criterion, ICT outperforms DQ (Blair et al. 2019, Ehler et al. 2020, Li & Noortgate 2019).

- No meta-study exists for CM, but several studies have found remarkably positive results regarding its performance, e.g:
  - "[CM] appears to be a very promising indirect questioning technique that can be used to successfully control for social desirability on surveys of sensitive behavior" (Hoffmann et al. 2015: 409).
  - "[CM] seems to be a promising alternative to conventional RRT variants" (Jann et al. 2012: 183).
  - [CM] offers a valid and useful means for achieving the experimental control of social desirability" (Hoffmann & Musch 2016: 1042).

- So, all in all, the community is quite enthusiastic.

Empirical Evidence 2: Spoiling the Party


PA

Uncovering a Blind Spot in Sensitive Question Research: False Positives Undermine the Crosswise-Model RRT

Marc Höglinger[1] and Andreas Diekmann[2]

[1] Winterthur Institute of Health Economics, Zurich University of Applied Sciences, Gertrudstrasse 15, 8401 Winterthur, Switzerland. Email: marc.hoeglinger@gmail.com
[2] ETH Zürich, Department of Humanities, Social and Political Sciences, Clausiusstrasse 50, 8092 Zurich, Switzerland. Email: diekmann@soz.gess.ethz.ch

- Core objection:
  - CM and ICT generate false positive estimates.
  - That is, respondents ***not*** having engaged in socially undesirable behavior are wrongly estimated as having done so.

- Höglinger and Diekmann (2017) show that CM estimates for zero-prevalence placebo items amount to:
  - 8 % for having received an engrafted organ.
  - 5 % for having suffered from „Chagas disease“ („Schlafkrankheit“).

- Höglinger & Jann (2018):
  - False-positive rate of 10 % for CM.


More is not always better: An experimental individual-level validation of the randomized response technique and the crosswise model

Marc Höglinger[1,2*], Ben Jann[2*]

1 Zurich University of Applied Sciences, Winterthur Institute of Health Economics, Winterthur, Switzerland,
2 University of Bern, Institute of Sociology, Bern, Switzerland

Empirical Evidence 2: Spoiling the Party

Likewise for ICT:

- Riambau & Ostwald (2020): ICT estimate of 12 % for a placebo item on having been invited to dinner with the Prime Minister of Singapore.

- Kuhn & Vivyan (2020):
  - Full individual validation design on reported and actual voter turnout.
  - Among real voters, 6 to 7 % are estimated as non-voters by ICT.


Political Science Research and Methods (2020), page 1 of 8
doi:10.1017/psrm.2020.18

RESEARCH NOTE

Placebo statements in list experiments: Evidence from a face-to-face survey in Singapore*

Guillem Riambau[1]* and Kai Ostwald[2]

Empirical Evidence 2: Spoiling the Party

- The (bitter) consequence is clear-cut:
  - If CM and ICT generate false positives, comparisons based on the MiB assumption are misleading.
  - Apparently higher (better) estimates are produced by false positives, i.e., by counting „innocent“ respondents als „guilty“ ones.
  - Hence CM and ICT would not improve measurement validity, but do the very reverse.

- A decade-old research field would be a dead-end street!
  - Every single MiB study is questionable!
  - Every single field application using CM or ICT is questionable!

Outline

1. Background
2. State of Research
   - Empirical Evidence 1: Enthusiasm
   - Empirical Evidence 2: Spoiling the Party
3. **Our Study**
4. Results
5. Discussion

Open Issues in this Debate

- Replication and extension: More material advisable before totally abandoning this research and these methods.
- Compare DQ with CM and ICT in one setting.
- Investigate the causes for the emergence of false positives.

Study Design

Two online surveys („Environment, Health, and Organ Donation“), November & December 2019)**:**

- German student survey:
  - all students of the University of Mainz (JGU) invited (N = 29,826).
  - n = 2,607.
- Swiss survey:
  - Commercial online access panel (respondi), German-speaking Switzerland.
  - n = 3,203.
- Main strategy: Test for false-positive estimates using zero-prevalence/placebo items (adopting the strategy of Höglinger & Diekmann 2017).

Study Design

- Four experimental groups:
  - DQ (p = 0.13)
  - CM (p = 0.3)
  - ICT short list (p = 0.283)
  - ICT long list (p = 0.283)

- Five sensitive questions in each group:
  - Blood donation
  - Excessive drinking
  - Dengue fever
  - Engrafted organ
  - Chagas disease

  } (virtually) zero prevalence in reality

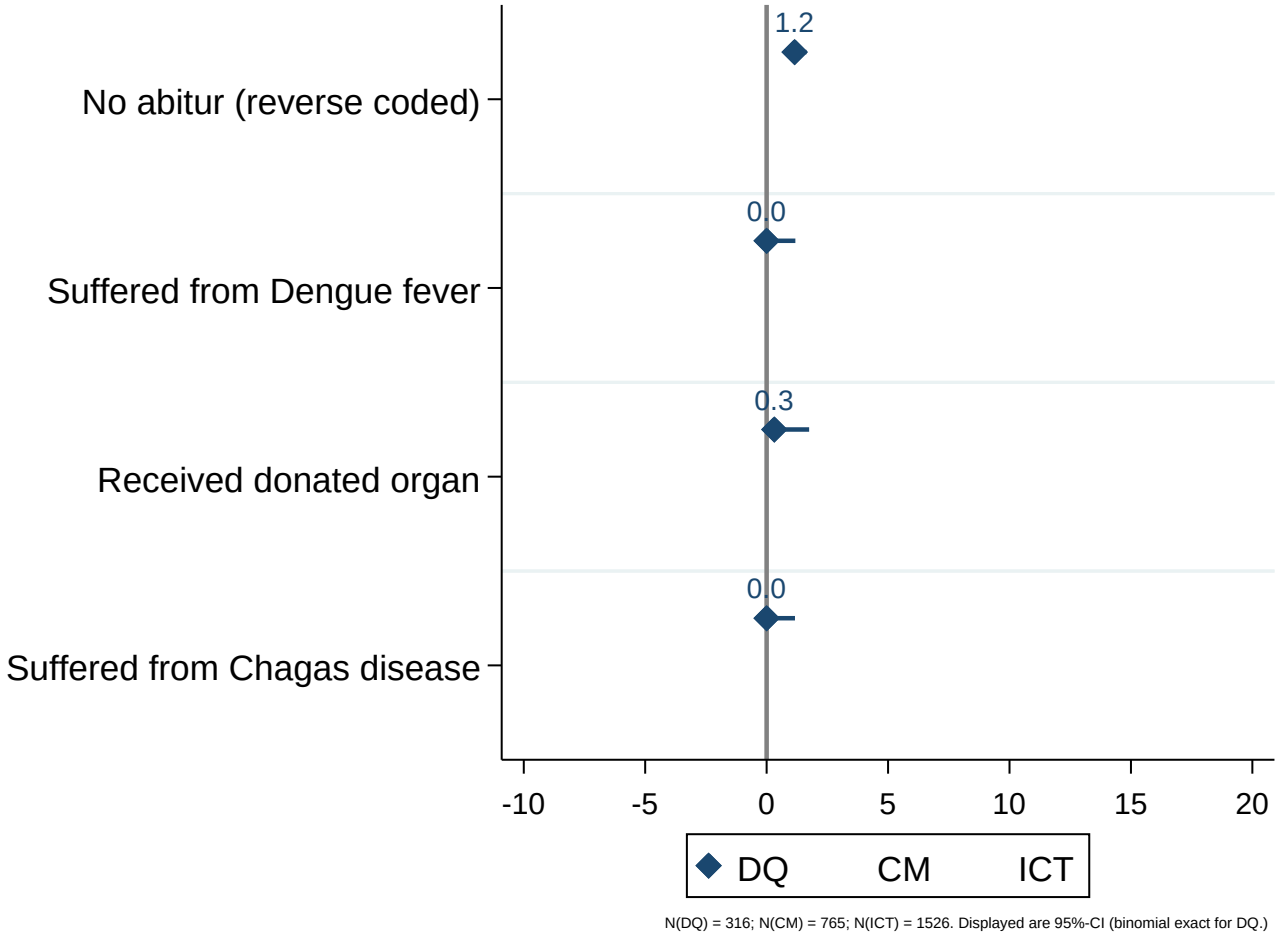
  - Plus: Item on having obtained the „Abitur“ (student survey; DQ & CM).

Outline

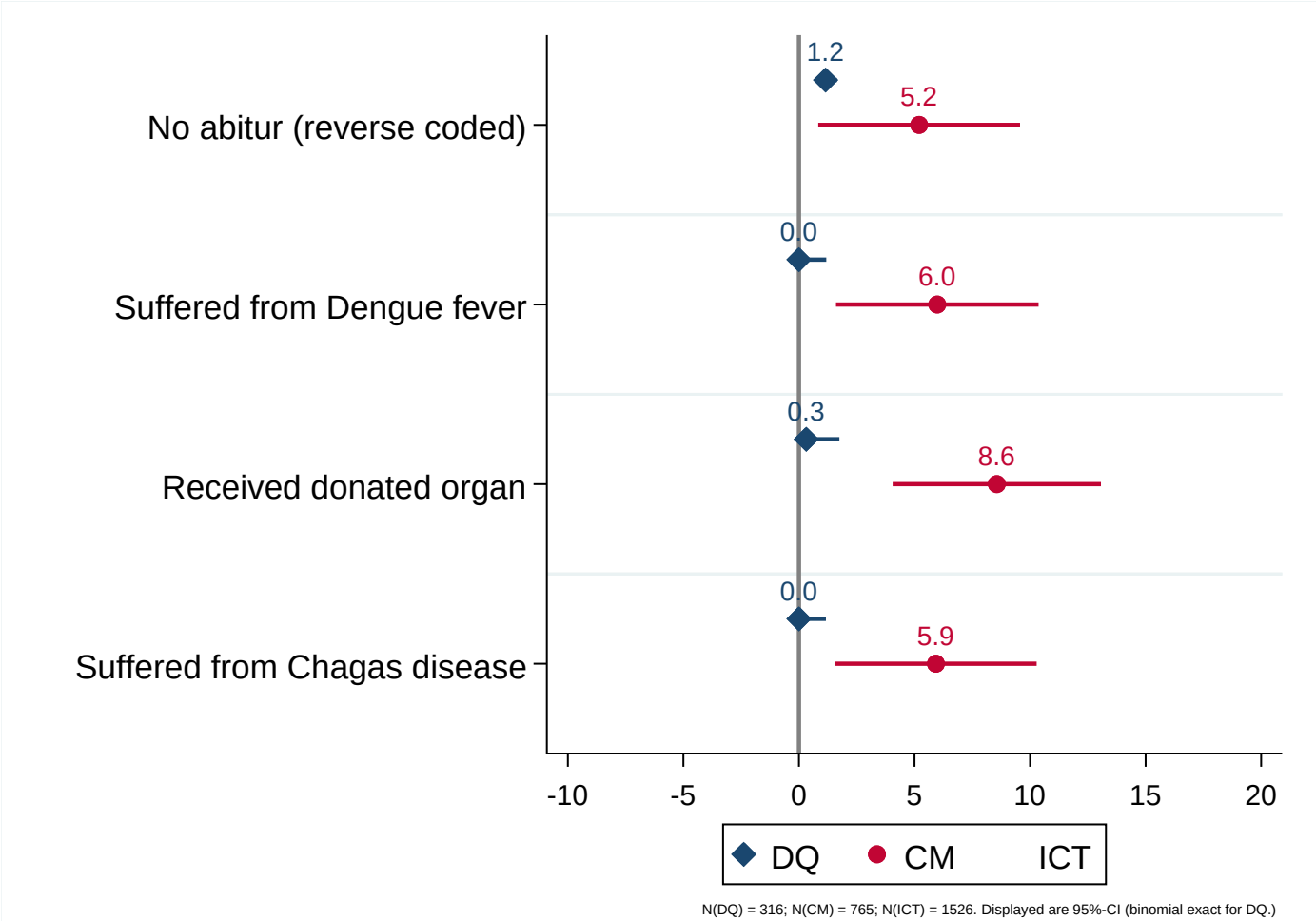
1. Background
2. State of Research
   - Empirical Evidence 1: Enthusiasm
   - Empirical Evidence 2: Spoiling the Party
3. Our Study
4. **Results**
5. Discussion

False Positives by Question Format: GER JGU Survey

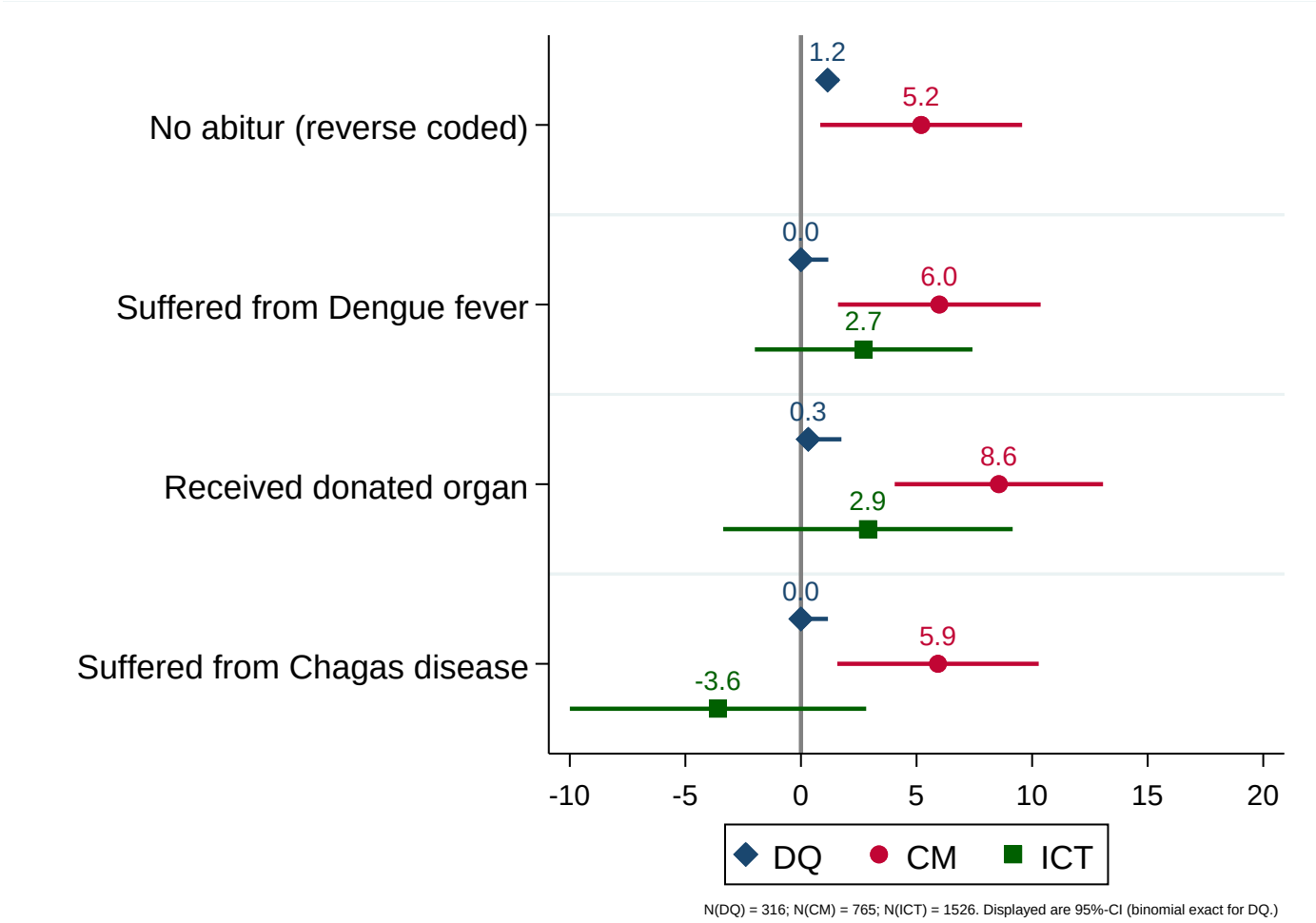
False Positives by Question Format: GER JGU Survey



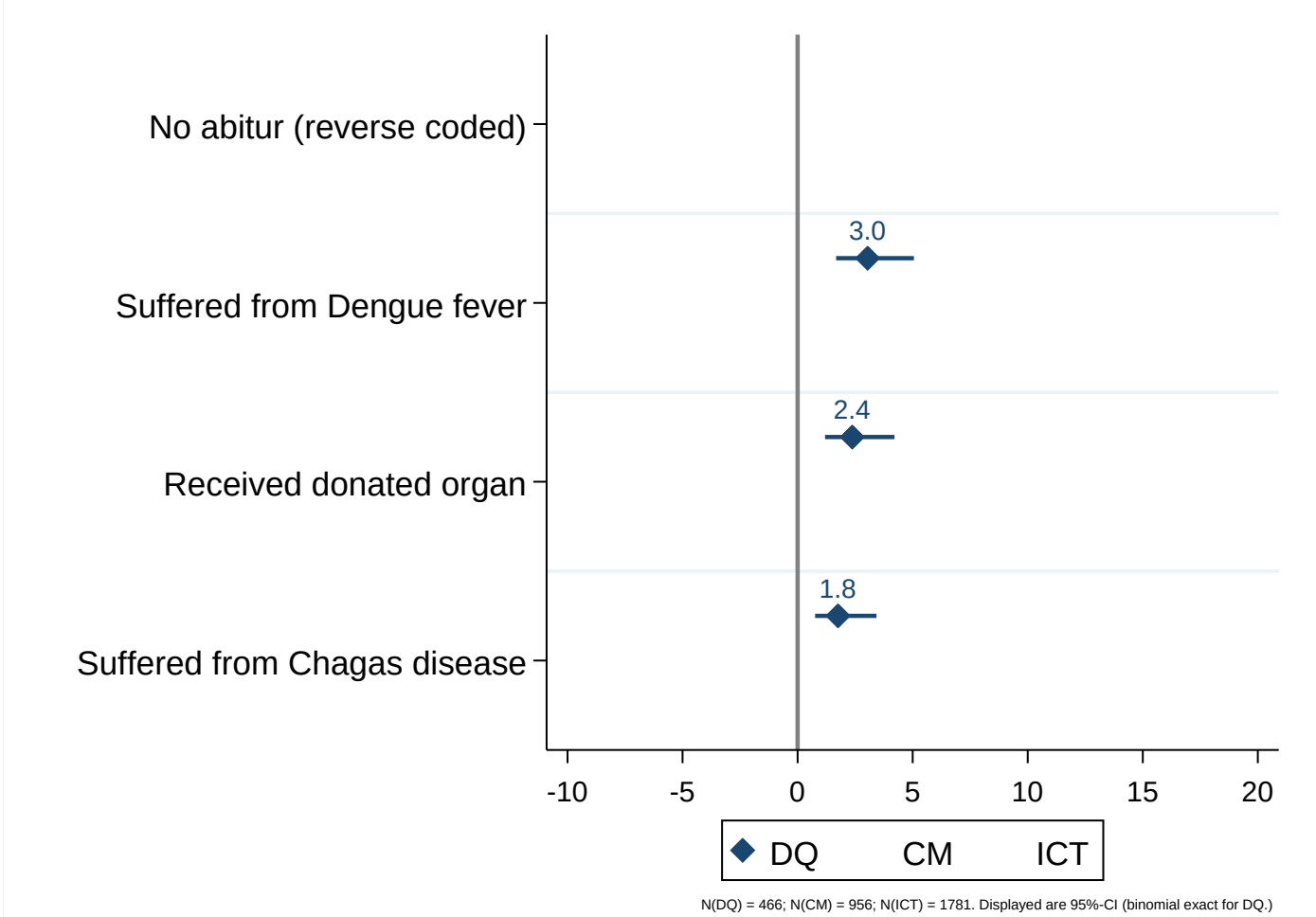
False Positives by Question Format: GER JGU Survey



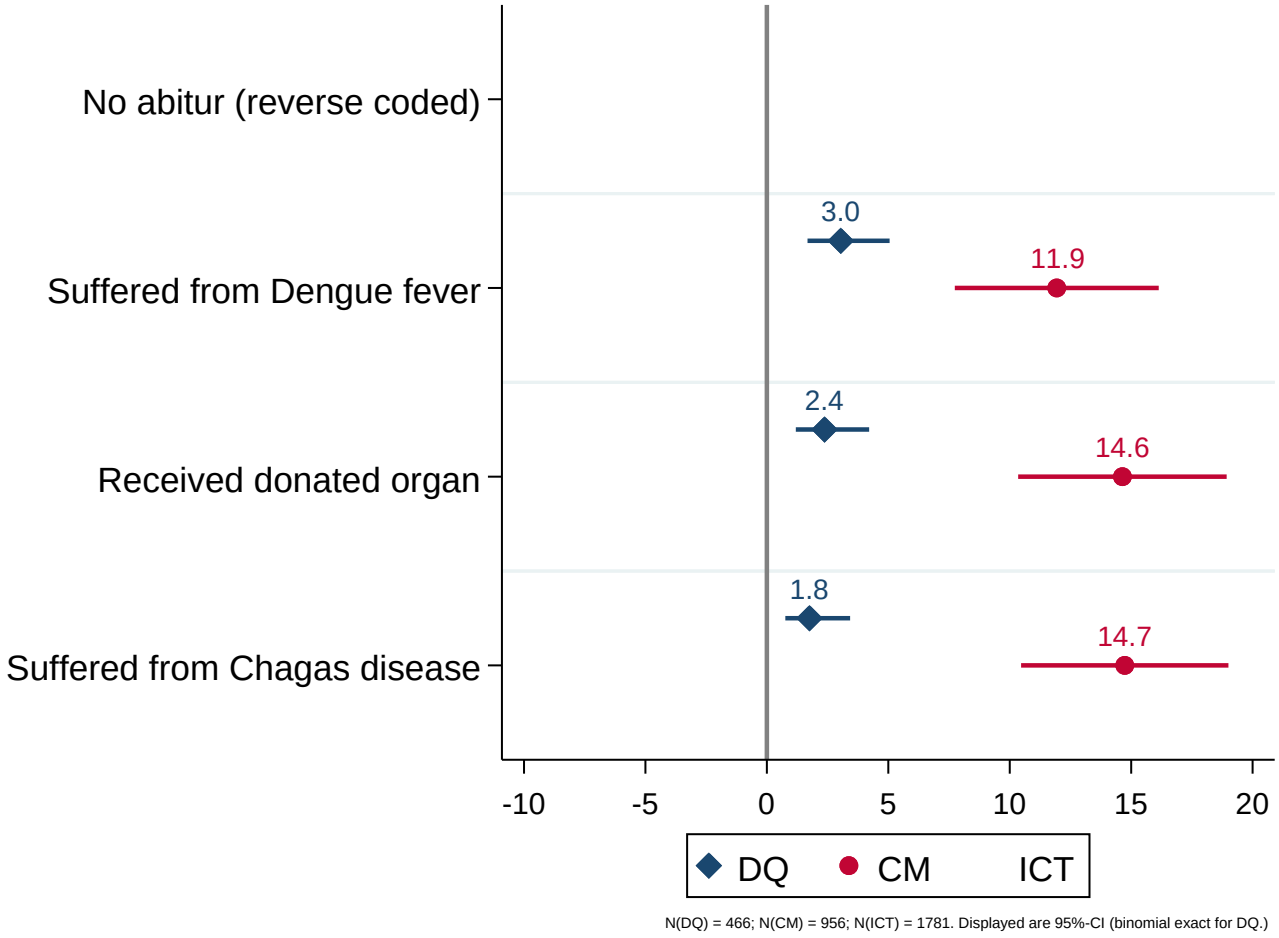
False Positives by Question Format: GER JGU Survey



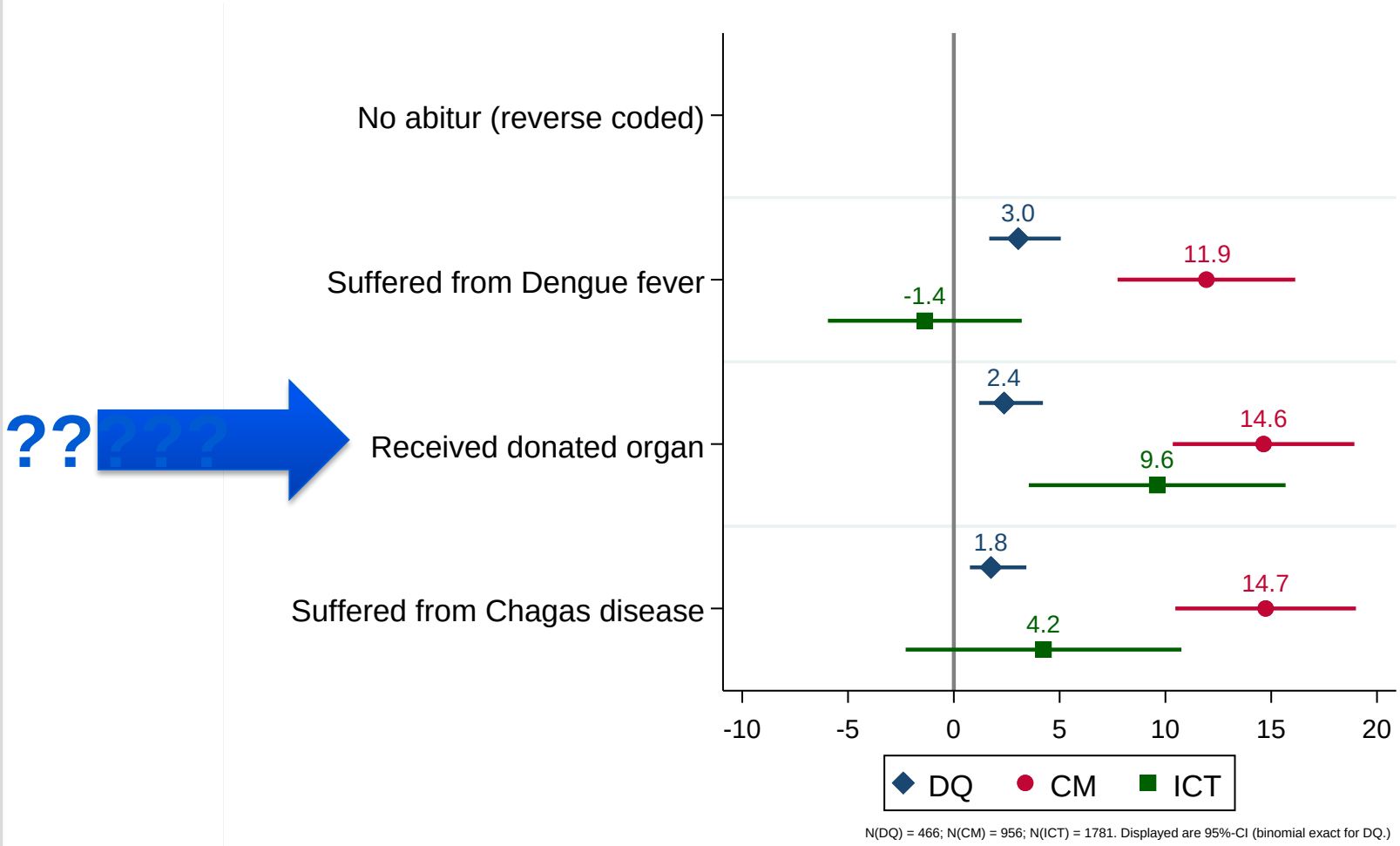
False Positives by Question Format: Swiss Survey



False Positives by Question Format: Swiss Survey



False Positives by Question Format: Swiss Survey



??

False Positives by Question Format: Overall Test

- Overall test with the joint data (both surveys):
  - Test all three zero-prevalence items at once against zero.
  - Stack data into long format.
  - Account for repeated measurements (clustered data).
- This is the test with the largest power (bear in mind that it would be naive to test against zero with small sample sizes in this context).

- Results:

	DQ	CM	ICT
	Estimate (SE)	*Estimate (SE)*	*Estimate (SE)*
Overall test: 3 items at once	1.466 *** (0.319)	10.665 *** (1.015)	3.000 (1.937)
N	2319	5105	9855

CM: Causes of False Positives (I/II)

	Student survey (N from 2182 to 2283)	Swiss survey (N from 2676 to 2822)	Both surveys (N from 4858 to 5105)
Design variables:			
URQ on father (1 = yes, 0 = other)	−0.009	0.003	−0.002
URQ on house number (1 = yes, 0 = other)	0.036	0.020	0.027
URQ on birthday (1 = yes, 0 = other)	0.003	−0.014	−0.006
URQ on month of birth (1 = yes, 0 = other)	−0.039	−0.008	−0.023
With “don’t know” answer option (1 = yes)	0.037	−0.050	−0.012
Response option order (1 = eq./uneq. first)	−0.040	0.016	−0.009
High prevalence of URQ (1 = >0.8)	0.068 *	0.159 ***	0.119 ***

*Note: Average marginal effects (reporting binary change from 0 to 1 for dummy variables) from bivariate binary logistic regressions. Significance tests based on robust standard errors adjusted for the clustering of items in respondents. CM = crosswise model. URQ = unrelated question. * p < 0.05; *** p < 0.001.*

CM: Causes of False Positives (II/II)

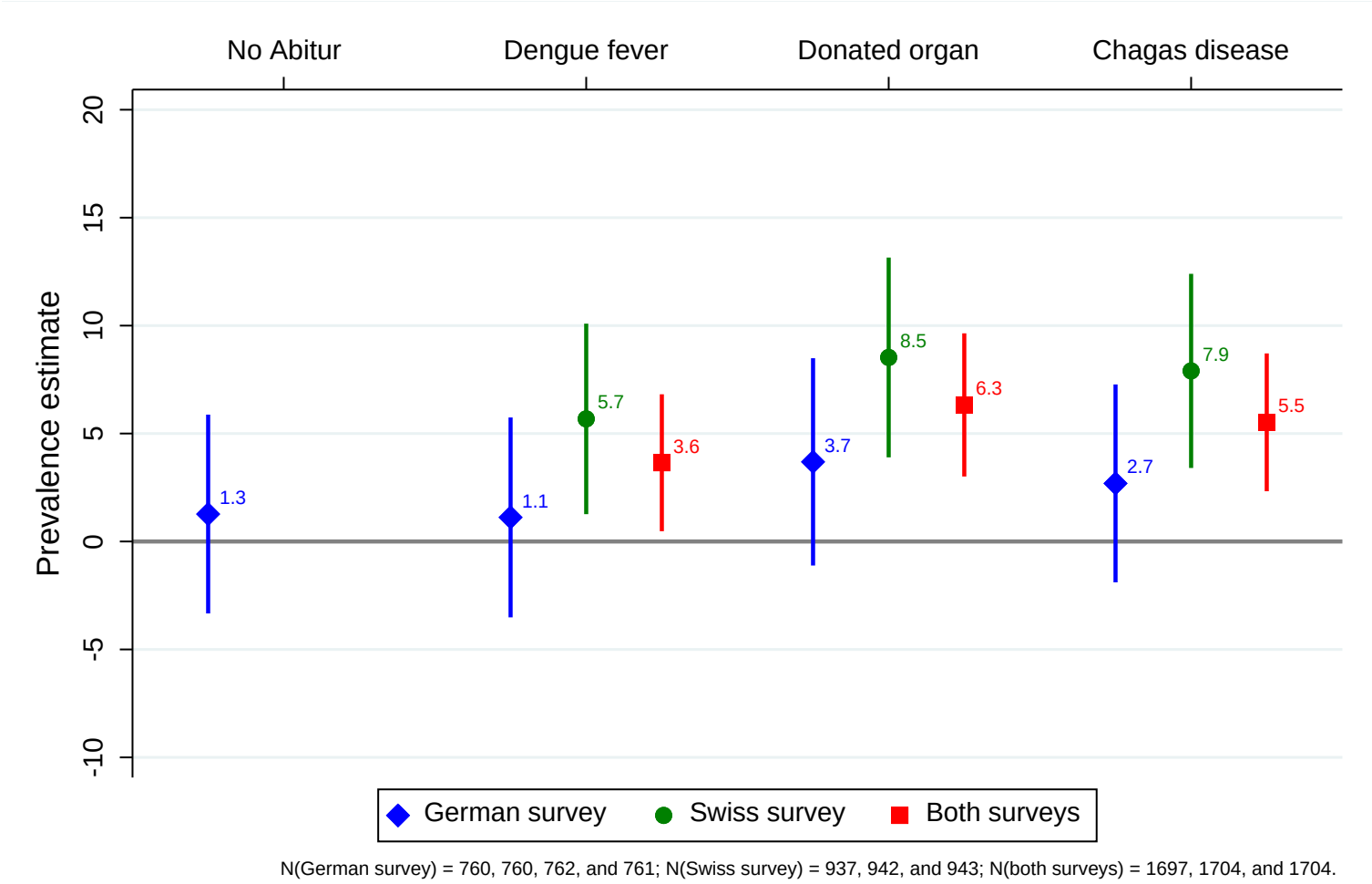
	Student survey (N from 2182 to 2283)	Swiss survey (N from 2676 to 2822)	Both surveys (N from 4858 to 5105)
Personal variables:			
Speeding in whole survey (1 = yes)	n.a.	0.115	0.160 *
Speeding on CM intro screen (1 = yes)	−0.003	0.131 ***	0.092 ***
Speeding on CM items (1 = yes)	−0.003	0.123 ***	0.092 ***
Understanding of CM procedure	0.006	−0.051 ***	−0.032 ***
Need for approval	0.008	−0.011	0.002
Gender (1 = female, other)	0.052	−0.038	−0.016
Age	0.002	−0.001	0.001
No “Abitur” diploma (1 = no Abitur)	n.a.	0.046	

*Note: Average marginal effects (reporting binary change from 0 to 1 for dummy variables) from bivariate binary logistic regressions. Significance tests based on robust standard errors adjusted for the clustering of items in respondents. No “speeding in whole survey” and “Abitur” effect for the student survey because of too little variance in this variable. CM = crosswise model. URQ = unrelated question. * $p < 0.05$; *** $p < 0.001$.*

CM: The Role of the Unrelated Questions (URQ)

- Remember: Consistent estimation of CM estimates require that the distribution of the URQ is known on the aggregate level and correct.
- What if our theoretically assumed values differ from the empirical ones?
- Empirical investigation indeed shows some deviations (see appendix table).

CM: Estimates Adjusted for Empirical Prevalence of URQs



N(German survey) = 760, 760, 762, and 761; N(Swiss survey) = 937, 942, and 943; N(both surveys) = 1697, 1704, and 1704.

Outline

1. Background
2. State of Research
   - Empirical Evidence 1: Enthusiasm
   - Empirical Evidence 2: Spoiling the Party
3. Our Study
4. Results
5. **Discussion**

Summary: Main Results

- CM is highly problematic with respect to false positives.
- ICT is less problematic; apart from one outlier, we do not find estimates significantly different from zero.
- We can identify the two main causes for the failure of CM:
  - Random clicking (speeding through the survey), at least partially caused by a poor survey quality.
  - URQ that deviate empirically from their expected values.

Consequences

- All studies on CM unanimously corroborate the problems with false positives. Hence:
  - Use CM only with the greatest care (and better not at all).
  - Always build designs that allow checking for false positives.
  - Always measure empirical prevalence rates for the URQs.

- Regarding ICT, our findings contradict those from the literature (Riambau/Ostwald 2020; Kuhn/Vivyan 2020). Hence:
  - We need more empirical material to clarify if and when false positives are an issue.
  - As with CM: Build designs that allow checking for false positives.

- Do not use CM and ICT and the „more-is-better“ assumption naively.

Thank you for listening!

felix.wolter@uni-konstanz.de

andreas.diekmann@soz.gess.ethz.ch

Literature

- Barton, Allen H. 1958. Asking the Embarrassing Question. *Public Opinion Quarterly 22 (1):67–68.*
- Blair, Graeme, and Kosuke Imai. 2012. Statistical Analysis of List Experiments. *Political Analysis 20 (1):47–77.*
- Blair, Graeme, Alexander Coppock, and Margaret Moor. 2019. *When to Worry About Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments: Unpublished Manuscript, https://graemeblair.com/papers/sensitivity-bias.pdf.*
- Bradburn, Norman M., and Seymour and Associates Sudman. 1979. *Improving Interview Method and Questionnaire Design. Response Effects to Threatening Questions in Survey Research. San Francisco: Jossey-Bass.*
- Droitcour, Judith, Rachel A. Caspar, Michael L. Hubbard, Teresa L. Parsley, Wendy Visscher, and Trena M. Ezzati. 1991. The Item Count Technique as a Method of Indirect Questioning: A Review of its Development and a Case Study Application. In *Measurement Errors in Surveys, edited by P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz. and S. Sudman. New York: Wiley.*
- Ehler, Ingmar, Felix Wolter, and Justus Junkermann. 2020. Sensitive Questions in Surveys: A Comprehensive Meta-Analysis of Experimental Survey Studies on the Performance of the Item Count Technique. *under review at Public Opinion Quarterly.*
- Hadaway, C. Kirk, Penny Long Marler, and Mark Chaves. 1993. What the Polls don't Show: A Closer Look at U.S. Church Attendance. *American Sociological Review 58 (6):741–752.*
- Höglinger, Marc, and Andreas Diekmann. 2017. Uncovering A Blind Spot in Sensitive Question Research: False Positives Undermine the Crosswise-Model RRT. *Political Analysis (25):131–137.*
- Höglinger, Marc, and Ben Jann. 2018. More Is Not Always Better: An Experimental Individual-Level Validation of the Randomized Response Technique and the Crosswise Model. *PLoS ONE 13 (8):https://doi.org/10.1371/journal.pone.0201770.*

Literature

- Hoffmann, Adrian, Birk Diedenhofen, Bruno Verschuere, and Jochen Musch. 2015. A Strong Validation of the Crosswise Model Using Experimentally-Induced Cheating Behavior. *Experimental Psychology 62 (6):403–414.*
- Hoffmann, Adrian, and Jochen Musch. 2016. Assessing the Validity of two Indirect Questioning Techniques: A Stochastic Lie Detector Versus the Crosswise Model. *Behavior Research Methods 48 (3):1032–1046.*
- Hyman, Herbert. 1944. Do They Tell the Truth? *Public Opinion Quarterly 8 (4):557–559.*
- Jann, Ben, Julia Jerke, and Ivar Krumpal. 2012. Asking Sensitive Questions Using the Crosswise Model: An Experimental Survey Measuring Plagiarism. *Public Opinion Quarterly 76 (1):32–49.*
- Johnson, Timothy P., Michael Fendrich, and Mary Ellen Mackesy-Amiti. 2012. An Evalutation of the Validity of the Crowne-Marlowe Need for Approval Scale. *Quality & Quantity 46 (6):1883–1896.*
- Li, Jiayuan, and Wim van den Noortgate. 2019. A Meta-analysis of the Relative Effectiveness of the Item Count Technique Compared to Direct Questioning. *Sociological Methods and Research (online first).*
- Kuhn, Patrick, and Nick Vivyan. 2020. The Misreporting Trade-Off Between List Experiments and Direct Questions in Practice: Partition Validation Evidence from Two Countries. *Unpublished Manuscript, July 13, 2020.*
- Locander, William, Seymour Sudman, and Norman Bradburn. 1976. An Investigation of Interview Method, Threat and Response Distortion. *Journal of the American Statistical Association 71 (354):269–275.*

Literature

- Parry, Hugh J., and Helen M. Crossley. 1950. Validity of Responses to Survey Questions. *Public Opinion Quarterly 14 (1):61–80.*
- Riambau, Guillem, and Kai Ostwald. 2020. Placebo Statemens in List Experiments: Evidence from a Face-to-Face Survey in Singapore. *Political Science and Research Methods (online first):1–8.*
- Schnapp, Patrick. 2019. Sensitive Question Techniques and Careless Responding: Adjusting the Crosswise Model for Random Answers. *methods, data, analyses (mda) 13 (2):307–320.*
- Tourangeau, Roger, and Ting Yan. 2007. Sensitive Questions in Surveys. *Psychological Bulletin 133 (5):859–883.*
- Warner, Stanley L. 1965. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association 60 (309):63–69.*
- Wolter, Felix, and Bastian Laier. 2014. The Effectiveness of the Item Count Technique in Eliciting Valid Answers to Sensitive Questions. An Evaluation in the Context of Self-Reported Delinquency. *Survey Research Methods 8 (3):153–168.*
- Wolter, Felix, and Peter Preisendörfer. 2013. Asking Sensitive Questions: An Evaluation of the Randomized Response Technique versus Direct Questioning Using Individual Validation Data. *Sociological Methods and Research 42 (3):321–353.*
- Wolter, Felix, and Peter Preisendörfer. 2020. Let's Ask About Sex: Methodological Merits of the Sealed Envelope Technique in Face-to-Face Interviews. In *Devianz und Subkulturen. Theorien, Methoden und empirische Befunde, edited by I. Krumpal and R. Berger. Wiesbaden: Springer VS.*
- Yu, Jun-Wu, Guo-Liang Tian, and Man-Lai Tang. 2008. Two New Models for Survey Sampling with Sensitive Characteristic: Design and Analysis. *Metrika 67 (3):251–263.*

Appendix and Backup

The Problem: Sensitive Questions in Surveys

Wie angenehm oder unangenehm wären Ihnen Angehörige dieser Gruppen als Nachbarn?

> Der Skalenwert -3 bedeutet „wäre mir sehr unangenehm“,
> der Skalenwert +3 bedeutet „wäre mir sehr angenehm“.

Mit den Werten dazwischen können Sie Ihre Meinung abstufen.

⇨ *Vorgaben bitte vorlesen! Zusätzlich Liste 49 vorlegen!*

Wie angenehm oder unangenehm wäre Ihnen -

	Wäre mir sehr unangenehm -3	-2	-1	0	+1	+2	Wäre mir sehr angenehm +3	KA
ein Italiener als Nachbar?	❍	❍	❍	❍	❍	❍	❍	❍
ein deutschstämmiger Aussiedler aus Osteuropa als Nachbar?	❍	❍	❍	❍	❍	❍	❍	❍
ein Asylbewerber als Nachbar?	❍	❍	❍	❍	❍	❍	❍	❍
ein Türke als Nachbar?	❍	❍	❍	❍	❍	❍	❍	❍
ein Jude als Nachbar?	❍	❍	❍	❍	❍	❍	❍	❍
ein Pole als Nachbar?	❍	❍	❍	❍	❍	❍	❍	❍

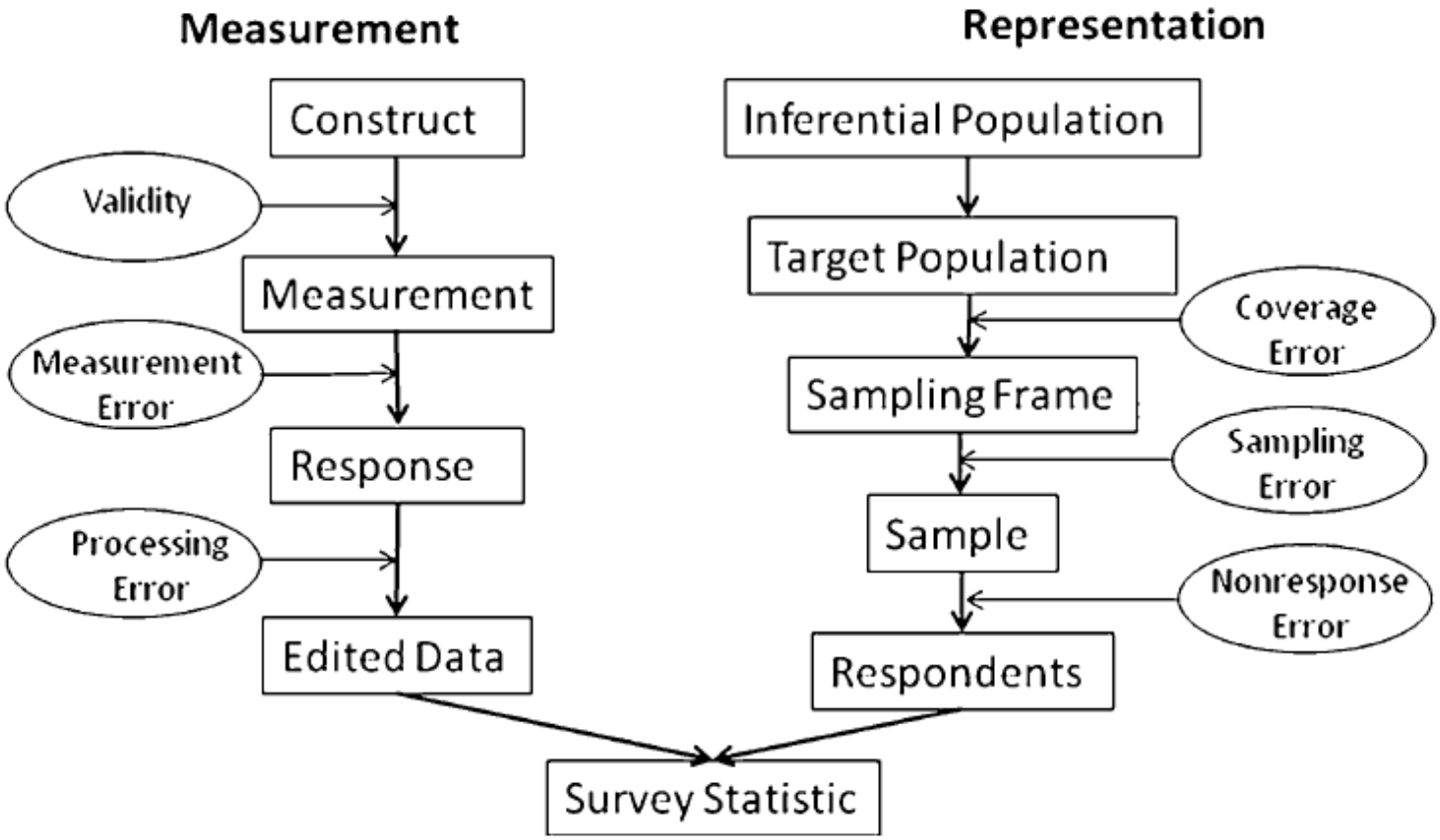
Source: ALLBUS 2016.

The Problem: Sensitive Questions in Surveys

Item	**Wording**
Sexual partners (lifetime)	"How many different sexual partners have you had in your lifetime so far? If you do not remember this exactly, please try to give an estimation."
Age of first sex	"At which age did you have sex for the first time?"
Coital frequency (last 4 weeks)	"In the last four weeks, how many times have you had sexual intercourse? If you do not remember this exactly, please try to give an estimation."
One-night stand (ever)	"Concerning your sexual experiences, have you ever had a one-night stand?" (with answers yes/no)
Sexual infidelity in current/last relationship (ever)	"In the time of your current (last) partnership, have you ever had an outside sexual affair? An outside sexual affair means that you had sex with a person other than your current (last) life partner" (with answers yes/no)
Homosexual contact (ever)	"Have you ever had a homosexual sexual contact?" (with answers yes/no)
Masturbation (often, very often)	"Although scientific studies show that more than 80 percent of all men and more than 60 percent of all women masturbate, most people don't like to talk about this. Nevertheless, we would like to know: How often do you masturbate?" (with answers very often, often, sometimes, seldom, never)

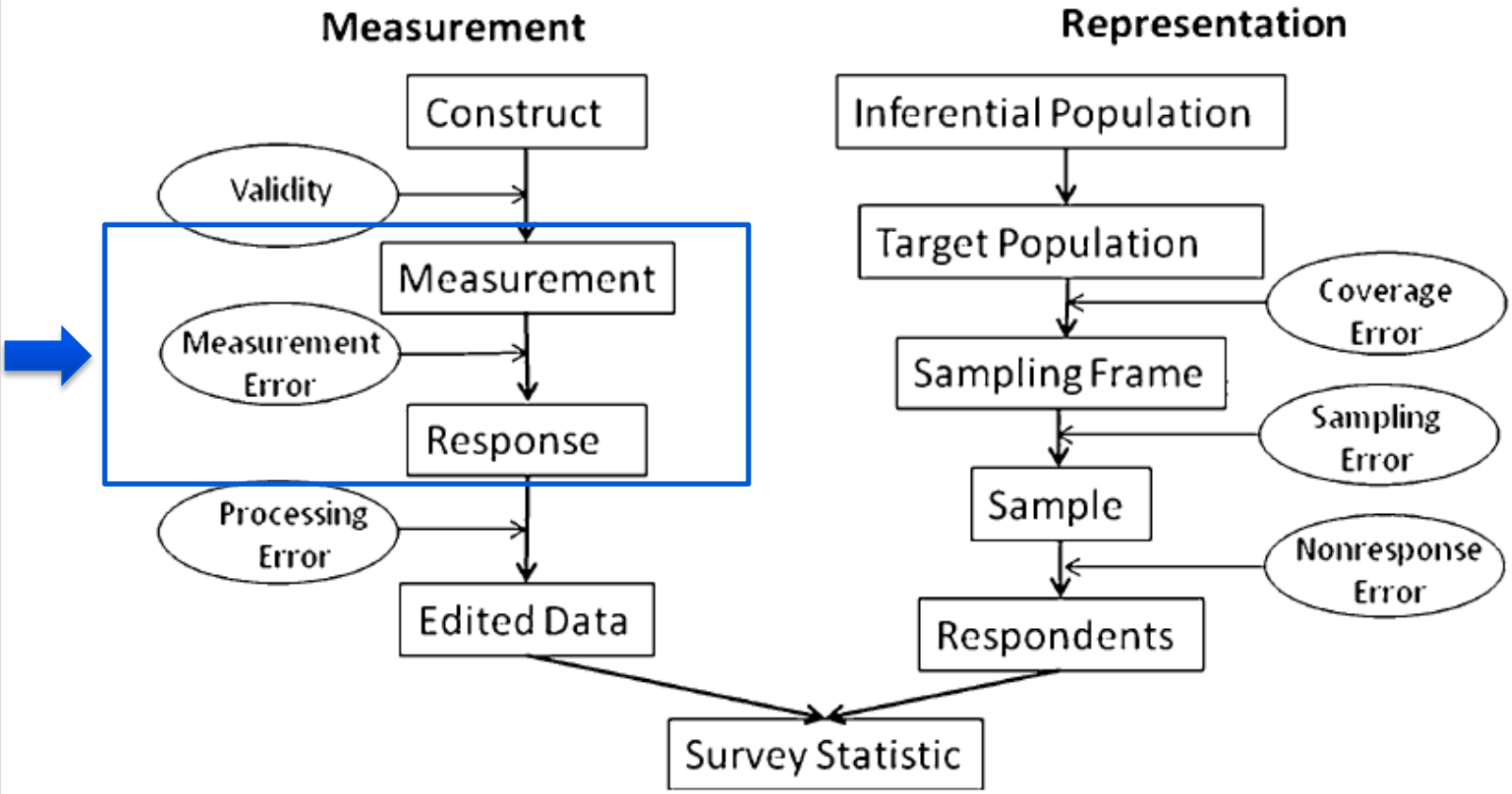
Source: Wolter & Preisendörfer 2020: 130.

Theory: Groves et al.: „Total Survey Error“



Quelle: Groves/Lyberg 2010: 856; Groves et al. 2004: 48.

Theory: Groves et al.: „Total Survey Error“



Quelle: Groves/Lyberg 2010: 856; Groves et al. 2004: 48.

Theory: Rational-Choice-Theory of Response Behavior

- RCT in a nutshell:

  „A respondent will answer correctly if the subjectively expected net utility of giving a truthful answer is higher than that of an edited or false answer“ (Preisendörfer/Wolter 2014: 128).

  „Respondents weigh the risks and benefits of responding truthfully” (Tourangeau et al. 2000: 14).

- See also:
  - Esser 1986, 1991: „Können Befragte lügen“?
  - Stocké 2004, 2007.

Theory: Rational-Choice-Theory of Response Behavior

- Cost factors:
  - Social desirability (SD; need for approval, SD-belief)
  - Identity management („internal“ costs)
  - Immediate, „objective“ costs (e.g., go into prison)
  - Cognitive effort

- Benefit factors:
  - Survey affinity
  - Personal or general contribution to science

- And:
  - Framing
  - Nonattitudes
  - Recall errors
  - Satisficing

Theory: Rational-Choice-Theory of Response Behavior

- Cost factors:
  - Social desirability (SD; need for approval, SD-belief)
  - Identity management („internal“ costs)
  - Immediate, „objective“ costs (e.g., go into prison)
  - **Cognitive effort**

- Benefit factors:
  - Survey affinity
  - Personal or general contribution to science

- And:
  - Framing
  - Nonattitudes
  - Recall errors
  - **Satisficing**

Satisficing with CM and ICT

- CM:
  - It can be shown that if respondents click randomly on CM questions, the CM prevalence estimate will be biased toward a prevalence rate of 50 percent.
  - Given a zero-prevalence item, the upward bias b is ½ of r, the proportion of respondents answering randomly (Höglinger & Diekmann 2017, online appendix). Thus, a proportion of, for example, 10 percent of random clickers yields a false positive rate of 5 percent.

- ICT:
  - Random clicking on the ICT answer scale might be guided by list length.
  - If answers are shifted to the right merely due to list length, false-positives occur.

The Crosswise Model (CM)

- The respondent‘s individual answer to both questions is not disclosed to anybody.

- Calculation of prevalence estimates for the sensitive item is still possible, because the distribution of the unrelated question (URQ) ist known on the aggregate level.

- Calculation of prevalence estimates of sensitive item $\hat{\pi}_{CM}$:

$$\hat{\pi}_{CM} = \frac{\hat{\lambda} + p - 1}{2p - 1}; p \neq 0.5$$

…where $\hat{\lambda}$ is the observed fraction of „equal“ answers and *p* the known prevalence of the URQ.

The Item Count Technique (ICT)

- Calculation of the prevalence estimate for the sensitive item ($\hat{\pi}_{ICT}$):

  $\hat{\pi}_{ICT} = \bar{x}_{LL} - \bar{x}_{SL}$, with $\bar{x}_{SL}$ = mean of the short list
  $\bar{x}_{LL}$ = mean of the long list

- Assumption: Independence of the subsamples.

- Other estimators are available (Blair & Imai 2012).

The Item Count Technique (ICT)

- Calculation of the prevalence estimate for the sensitive item ($\hat{\pi}_{ICT}$):

$\hat{\pi}_{ICT} = \bar{x}_{LL} - \bar{x}_{SL}$, with $\bar{x}_{SL}$ = mean of the short list
$\bar{x}_{LL}$ = mean of the long list

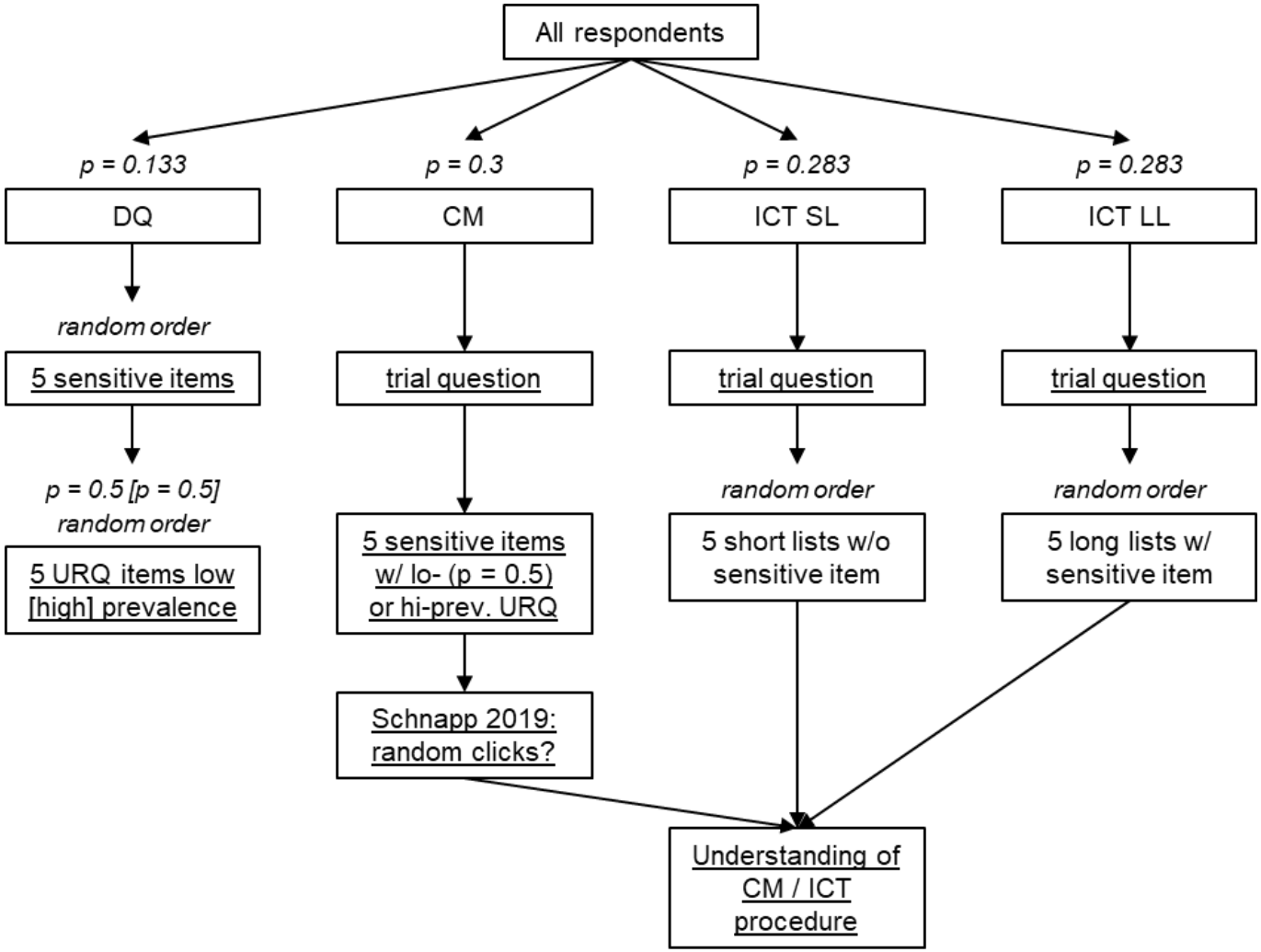
- Sampling variance:

$$Var(\hat{\pi}_{ICT}) = Var(\bar{x}_{LL}) + Var(\bar{x}_{SL})$$

- Assumption: Independence of the subsamples.

- Other estimators are availble (NLS, ML, see Blair & Imai 2012).

Study Design



Example: CM Procedure

- **Frage A:**
  Hat Ihre Mutter in den Monaten Januar oder Februar Geburtstag: Ja oder Nein?
  (*Falls Sie dies nicht wissen, nehmen Sie eine andere Ihnen bekannte Person, deren Geburtstag Sie kennen.)*

- **Frage B:**
  Sind Sie jemals an dem Dengue-Fieber erkrankt bzw. haben sich mit dem Dengue-Virus infiziert: Ja oder Nein?

- Vergleichen Sie Ihre Antworten auf die Fragen A und B. Sind diese gleich oder unterschiedlich?

  ❒ – gleich (beide „Nein“ oder beide „Ja“)
  ❒ – unterschiedlich (einmal „Nein“ und einmal „Ja“)

CM Procedure: Unrelated Questions (URQ)

- Two groups, URQ with low or high prevalence.

Item	Wording
URQ 2	“Is the birthday of your mother between the 1st and up to and including the 6th [the 7th and up to and including the 31st] of a month?”
URQ 3	“Is the birthday of your father in January or February [March to December]?”
URQ 4	“Is the birthday of your father between the 1st and up to and including the 6th [the 7th and up to and including the 31st] of a month?”
URQ 5	“Is the first digit of your physical address’s house number either 7, 8, or 9 [1, 2, 3, 4, 5, or 6]?”
URQ 6	“Is the first digit of your mother’s physical address’s house number either 7, 8, or 9 [1, 2, 3, 4, 5, or 6]?”

Example: ICT Procedure

Wie viele der folgenden Fragen treffen auf Sie zu?

- Haben Sie aktuell ein Haustier oder hatten Sie früher eines: Ja oder Nein?
- Besitzen Sie ein Auto: Ja oder Nein?
- Sind Sie Brillenträger/-in: Ja oder Nein?
- Haben Sie selbst ein Spenderorgan erhalten (Niere, Herz, Teile der Lunge oder Leber, Bauchspeicheldrüse): Ja oder Nein?
- Sind Sie Mitglied in einem Fußballverein: Ja oder Nein?

Anzahl Ihrer „Ja“-Antworten insgesamt:

❒	❒	❒	❒	❒
1	**2**	**3**	**4**	**5**

Question wording: Sensitive Items

Item	Wording
Dengue fever	“Have you ever suffered from Dengue fever or have you ever been infected with the Dengue virus?”
Received donated organ	“Have you received a donated organ (kidney, heart, part of a lung or liver, pancreas)?”
Chagas disease	“Have you ever suffered from Chagas disease (Trypanosomiasis)?”
Blood donation	“Have you ever donated blood?”
Excessive drinking	“In the last two weeks, have you consumed five or more alcoholic drinks in a row (e.g., glasses of wine, bottles of beer etc.)?”
Abitur diploma (CM only)	“Have you obtained the Abitur?” Swiss survey: “Have you obtained the Maturitätsprüfung / Matura / Abitur?”

Descriptive Statistics (Mean Values)

	German student survey	Swiss access panel survey
Response time (median)	14.667	13.417
Speeding in whole survey (1 = yes)	0.002	0.054
Speeding on CM intro screen (1 = yes)	0.272	0.579
Speeding on CM items (1 = yes)	0.200	0.671
Understanding of technique:		
CM	4.680	4.538
ICT	4.707	4.643
Need for approval	4.102	4.709
Gender (1 = female, other)	0.725	0.523
Age	24.937	47.646
“Abitur” diploma	0.989	0.260

Descriptive Statistics (Mean Values)

	German student survey	Swiss access panel survey
Response time (median)	14.667	13.417
Speeding in whole survey (1 = yes)	0.002	0.054
Speeding on CM intro screen (1 = yes)	0.272	0.579
Speeding on CM items (1 = yes)	0.200	0.671
Understanding of technique:		
CM	4.680	4.538
ICT	4.707	4.643
Need for approval	4.102	4.709
Gender (1 = female, other)	0.725	0.523
Age	24.937	47.646
“Abitur” diploma	0.989	0.260

Schnapp 2019 Proposition

- „Random clickers“ yield false positives in the CM procedure.
- Empirically adjust for these „random clickers“.
- Estimate proportion of random answers by asking respondents:
  „Have you answered carefully or have you given a random answer?“

Schnapp 2019 Procedure

Vielen Dank.

Die bisherige Forschung hat herausgefunden, dass die gerade verwendete Spezialtechnik viele Befragte verwirrt und die Fragen oft schwer zu beantworten sind.

Wir würden Sie daher bitten, unten ehrlich anzuklicken, ob Sie die fünf Fragen korrekt und gewissenhaft beantwortet haben, oder eine oder mehrere Fragen nur zufällig angeklickt haben.

Es ist nicht schlimm, wenn Sie zufällig geklickt haben, wir möchten nur eine ehrliche Antwort haben. Ihre Anonymität bleibt davon unberührt.

Blutspende:	❒ korrekte Antwort	❒ Zufallsklick
Dengue-Fieber:	❒ korrekte Antwort	❒ Zufallsklick
Spenderorgan erhalten:	❒ korrekte Antwort	❒ Zufallsklick
An Schlafkrankheit erkrankt:	❒ korrekte Antwort	❒ Zufallsklick
Alkohol konsumiert:	❒ korrekte Antwort	❒ Zufallsklick

Schnapp (2019) Estimate

- Prevalence estimate:

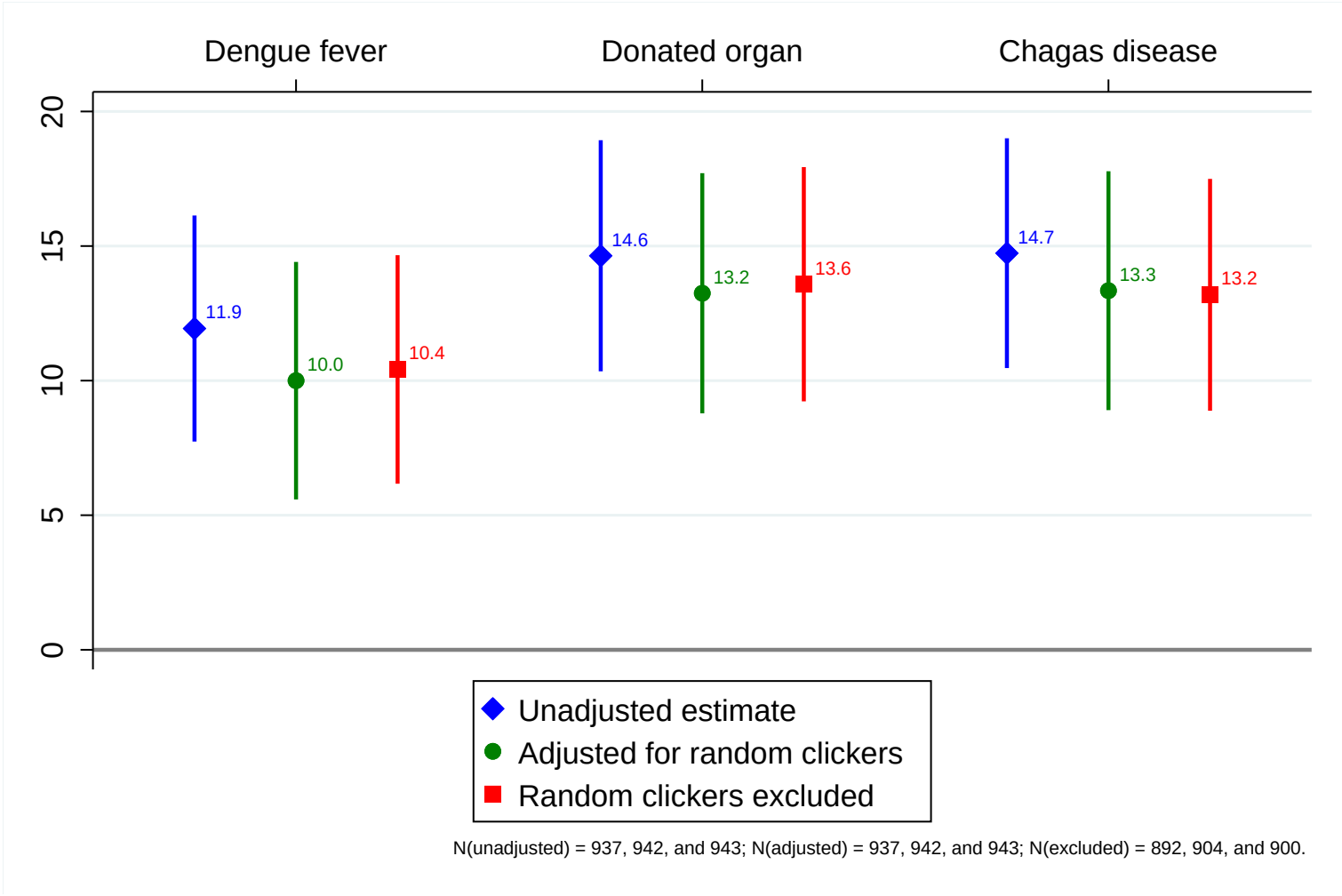
$$\hat{\pi}_{CMadj} = \frac{\hat{\pi}_{CM} - 0.5r}{1 - r}; r \neq 1$$

- With Variance:

$$Var\left(\hat{\pi}_{CMadj}\right) = \frac{Var(\hat{\pi}_{CM})}{(1 - r)^2}; r \neq 1$$

where r is the fraction of respondents having answered randomly to the CM question.

CM: Adjustment by the Schnapp (2019) Procedure



ICT: Non-Key Items

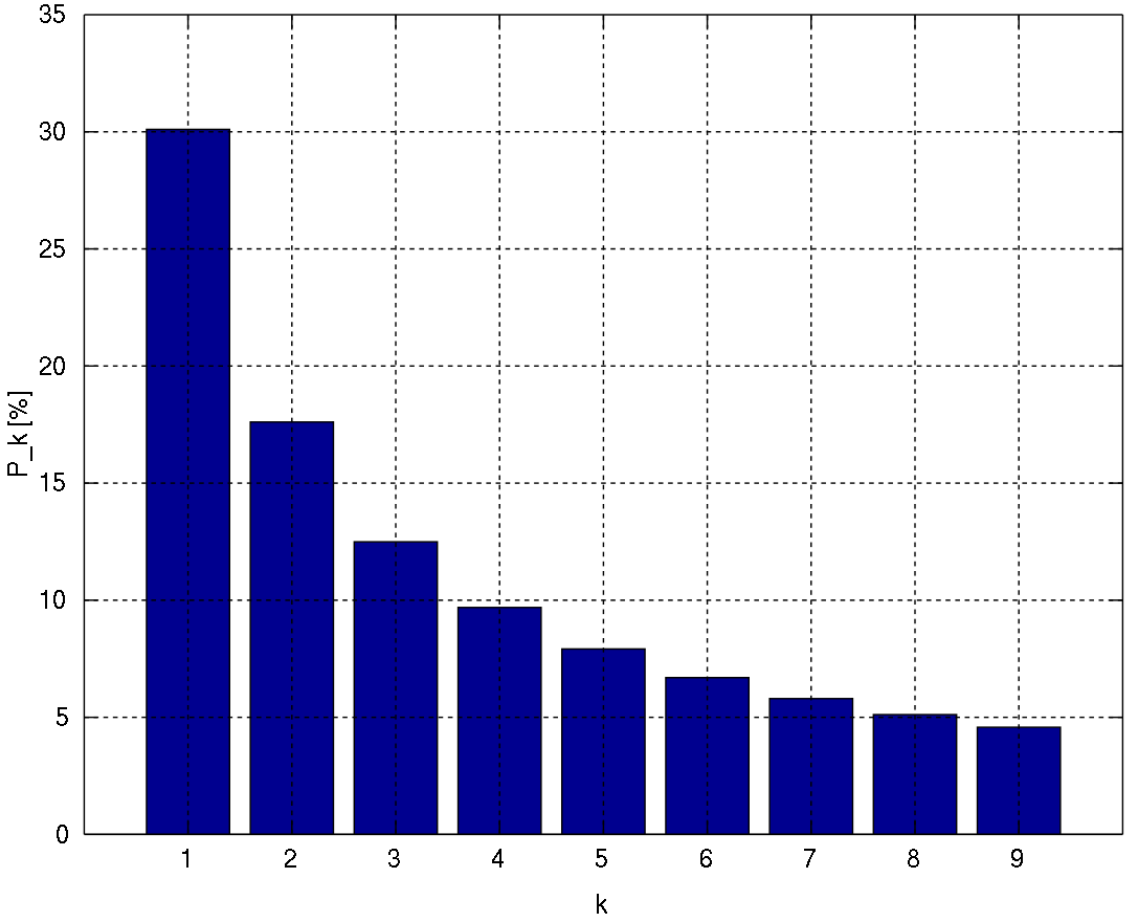
List for item	Non-key items
Dengue fever	Did you ever have a traffic accident: yes or no? Have you ever moved: yes or no? Does your house number start with the figure “8”: yes or no? Did you have a dentist’s appointment in the last five years: yes or no?
Received donated organ	Do you have a domestic animal now or have you had one before: yes or no? Do you own a car: yes or no? Are you a spectacles wearer: yes or no? Are you a member of a football [i.e., soccer] club: yes or no?
Chagas disease	Do you use aspirin regularly: yes or no? [Switzerland: Do you use pills against headaches regularly: yes or no?] Do you use an electric toothbrush: yes or no? Do you ride your bike regularly: yes or no? Have you ever been hospitalized: yes or no?

Other Variables

- Indicators for response latency:
  - Overall response time
  - Speeding in whole survey (1 = yes)
  - Speeding on CM intro screen (1 = yes)
  - Speeding on CM items (1 = yes)
- Subjective understanding of CM/ICT procedure
- SD: need for social approval (Crowne-Marlowe scale)
- Socio-demographics

- CM design (experimentally varied):
  - w/ versus w/o „don‘t know“ answer option
  - Response option order („equal“/“unequal“ versus „unequal“/“equal“)
  - High versus low prevalence of URQ

Benford Distribution



Source: Wikipedia.

Descriptive Statistics (Mean Values)

	German student survey	Swiss access panel survey
Response time (median)	14.667	13.417
Speeding in whole survey (1 = yes)	0.002	0.054
Speeding on CM intro screen (1 = yes)	0.272	0.579
Speeding on CM items (1 = yes)	0.200	0.671
Answered CM randomly (Schnapp):		
Dengue fever	0.008	0.048
Donated organ	0.005	0.038
Chagas disease	0.013	0.038
Understanding of technique:		
CM	4.680	4.538
ICT	4.707	4.643
Need for approval	4.102	4.709
Gender (1 = female, other)	0.725	0.523
Age	24.937	47.646
“Abitur” diploma	0.989	0.260

CM: The Role of the Unrelated Questions (URQ)

- Remarks:
  - We assume equally distributed birthdates across the year.
  - This is empirically not true.
  - However, no regular pattern exists across time periods (at least for Germany).
  - The house numbers are assumed to be Benford distributed.
  - We have no information whether this holds for Germany or Switzerland.
  - However, Höglinger/Jann (2018: 8) report an exact match between theoretic and empirical values.

CM: The Role of the Unrelated Questions (URQ)

	Theoretical value	Student survey (N from 143 to 170)	Swiss survey (N from 217 to 235)
Mother's birthday Jan–Feb	0.162	0.179	0.157
Mother's birthday 1st–6th	0.197	0.266 *	0.270 **
Father's birthday Jan–Feb	0.162	0.159	0.197
Father's birthday 1st–6th	0.197	0.269 *	0.296 ***
Own house number 7–9	0.155	0.159	0.103 *
Mother's house number 7–9	0.155	0.153	0.098 *
Mother's birthday Mar–Dec	0.838	0.818	0.782 *
Mother's birthday 7th–31st	0.803	0.794	0.775
Father's birthday Mar–Dec	0.838	0.839	0.798
Father's birthday 7th–31st	0.803	0.821	0.741 *
Own house number 1–6	0.845	0.792	0.752 ***
Mother's house number 1–6	0.845	0.725 ***	0.650 ***

** p < 0.05; ** p < 0.01; *** p < 0.001.*

go back