

Asking Sensitive Questions in Online Surveys

An Experimental Comparison of the Randomized Response Technique and the Crosswise Model

Andreas Diekmann (ETH Zurich)

Marc Höglinger (ETH Zurich)

Ben Jann (University of Bern)

GOR 2012, Mannheim, March 5-7, 2012

Outline

- Misreporting of sensitive issues in (online) surveys
- Some indirect approaches to elicit truthful answers
 - The Randomized Response Technique (RRT)
 - The Crosswise Model (CM): A new alternative to RRT
- Experimental comparison of the different approaches: an online survey on student cheating and plagiarism
- Conclusion

Misreporting of sensitive issues – a pervasive problem

- Survey respondents might not tell the truth if asked questions on sensitive issues such as norm violations or deviant behavior. This leads to distorted results.
- Considerable share of ‘liars’ (respondents with a false negative response) in surveys that use direct questioning (estimates from validation studies):
 - Penal conviction: 42.5% (F2F, Wolter & Preisendörfer 2011)
 - Welfare and unemployment benefit fraud: 75% (F2F, van der Heijden et al. 2000)
 - Driving under influence: 54% (P&P, Locander et al. 1976)
 - Bankruptcy: 32% (Ibid.)

Misreporting of sensitive issues – a problem also in online surveys

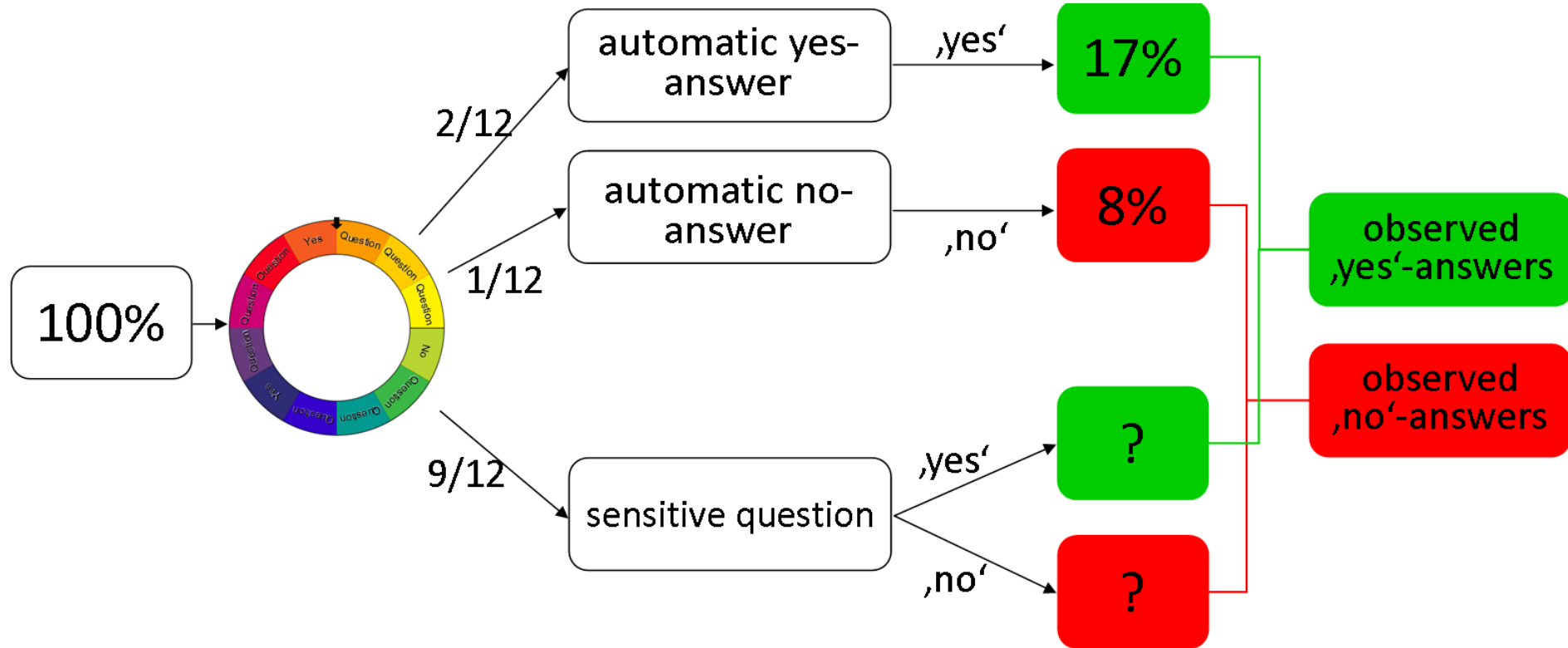
- Online surveys offer more anonymity and privacy than interviewer-administered surveys.
- Decrease in the misreporting of sensitive information in online mode compared to CATI (Kreuter, Presser and Tourangeau 2008).
- However, a substantial amount of misreporting remained (Ibd.).
 - falsely denying of poor grade point average:
83% CATI vs. 62% online mode
 - falsely denying having received an unsatisfactory grade:
33% CATI vs. 20% online mode

The Randomized Response Technique (RRT)

(Warner 1965; Fox and Tracy 1986)

- Main principle: privacy protection through randomization (i.e. add random noise to the answers)
- A randomizing device, the outcome of which is only known to the respondent, decides whether...
 - the sensitive question has to be answered
 - an automatic 'yes' or 'no' has to be given or a surrogate question has to be answered
- Since only the respondent knows the outcome of the randomization device, a 'yes' cannot be interpreted as an admission of guilt.
- However, with knowledge of the properties of the randomizing device, a prevalence estimate for the sensitive question can be derived.

RRT example (forced response design)



- Prevalence estimate (π):

$$Pr(\text{observed yes}) = Pr(\text{sensitive question}) * \pi + Pr(\text{automatic yes})$$

$$\pi = \frac{Pr(\text{observed yes}) - P(\text{automatic yes})}{Pr(\text{sensitive question})}$$

The Crosswise Model (CM): a new alternative to RRT

(Yu, Tian, and Tang 2008)

- Simple idea: Ask a sensitive question and a non-sensitive question and let the respondent indicate whether ...
 - the answers to the questions are the **same** (both 'yes' or both 'no')
 - the answers to the questions are **different** (one 'yes', the other 'no')

		<i>non-sensitive question</i>	
		No	Yes
<i>sensitive question</i>	No	same	different
	Yes	different	same

- Note: Questions must be uncorrelated and the probability of answering 'yes' to the non-sensitive question must be unequal 0.5.

The Crosswise Model (CM): a new alternative to RRT

(Yu, Tian, and Tang 2008)

- Prevalence estimate (π):

$$Pr(\text{same}) = (1 - \pi) * (1 - Pr(\text{nonsensitive yes})) + \pi * Pr(\text{nonsensitive yes})$$

$$\pi = \frac{Pr(\text{same}) + Pr(\text{nonsensitive yes}) - 1}{2 * Pr(\text{nonsensitive yes}) - 1}$$

- Note: CM is formally identical to Warner's original RRT model.

Performance of RRT in online mode

- RRT does not seem to work well in in online surveys. Results so far showed for the RRT...
 - no difference in or even lower prevalence estimates for socially undesirable behavior compared to direct questioning (Coutts et al. 2011 , Coutts & Jann 2011, Peeters 2006, Snijders & Weesie 2008)
 - unrealistically high prevalence estimates for voting (Holbrook & Krosnick 2010)
 - exception: higher prevalence estimates with the RRT in a survey on adult entertainment desires (de Jong, Pieters and Fox 2010)
- However, RRT implementations so far were often not well suited to online mode.
 - randomizing device not at respondents' immediate reach
 - randomizing device not trustworthy

Performance of the Crosswise Model

- The Crosswise Model seems to be a promising alternative
 - higher prevalence estimates than with direct questioning in a p&p survey on plagiarism (Jann, Jerke, Krumpal 2011)
 - however, no empirical application in online mode so far
- Advantages of the Crosswise Model over RRT
 - easier to understand
 - no need for a randomizing device
 - respondent is not forced into giving a ‘false’ automatic response
 - no obvious self-protective answering strategy (e.g. always tick ‘no’)

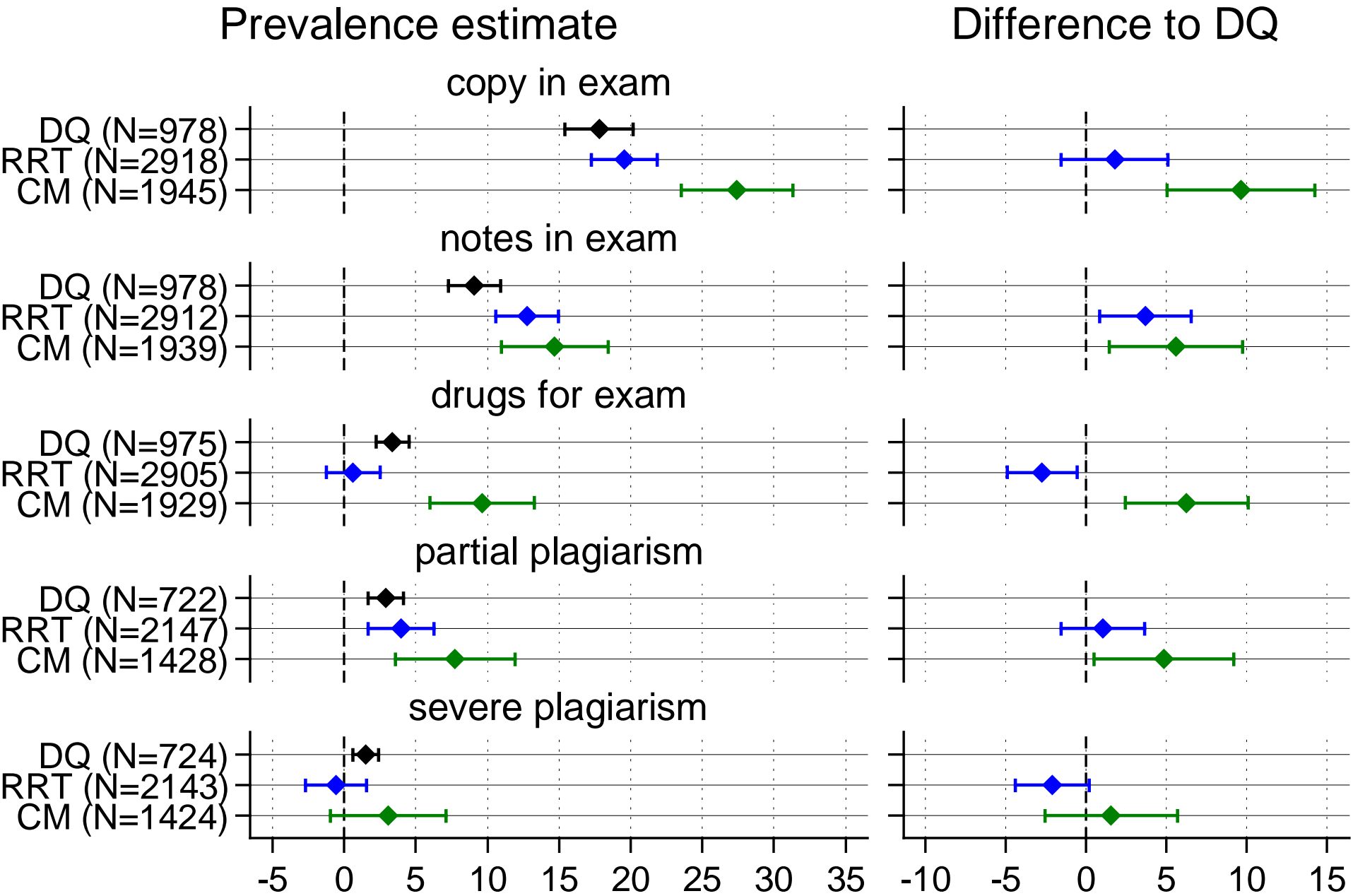
Our study: survey on student cheating and plagiarism

- Web survey among students of the University of Bern and the ETH Zurich in spring 2011
- Response rate 33%, 6'494 completed interviews
- Sensitive questions on
 - copying from other students in exam (copy)
 - using crib notes in exam (notes)
 - taking drugs to enhance exam performance (drugs)
 - partial paper plagiarism (partial)
 - severe paper plagiarism (severe)
- Comparing direct questioning (DQ) to three variants of RRT and two variants of the Crosswise Model (CM)
- Aprox. 1'000 randomly assigned respondents in each experimental condition

Experimental conditions: 6 different implementations of the sensitive questions

- **DQ**: direct questioning
≥
- **RRT wheel**: forced response RRT using virtual random wheel
≥
- **RRT pick**: forced response RRT using ‘Pick a number’ method
≥
- **RRT Benford**: RRT using Benford distribution and unrel. questions
≥, >>
- **CM unr. quest.**: Crosswise Model using unrelated questions
≥
- **CM pick**: Crosswise Model using ‘Pick a number’ method
≥

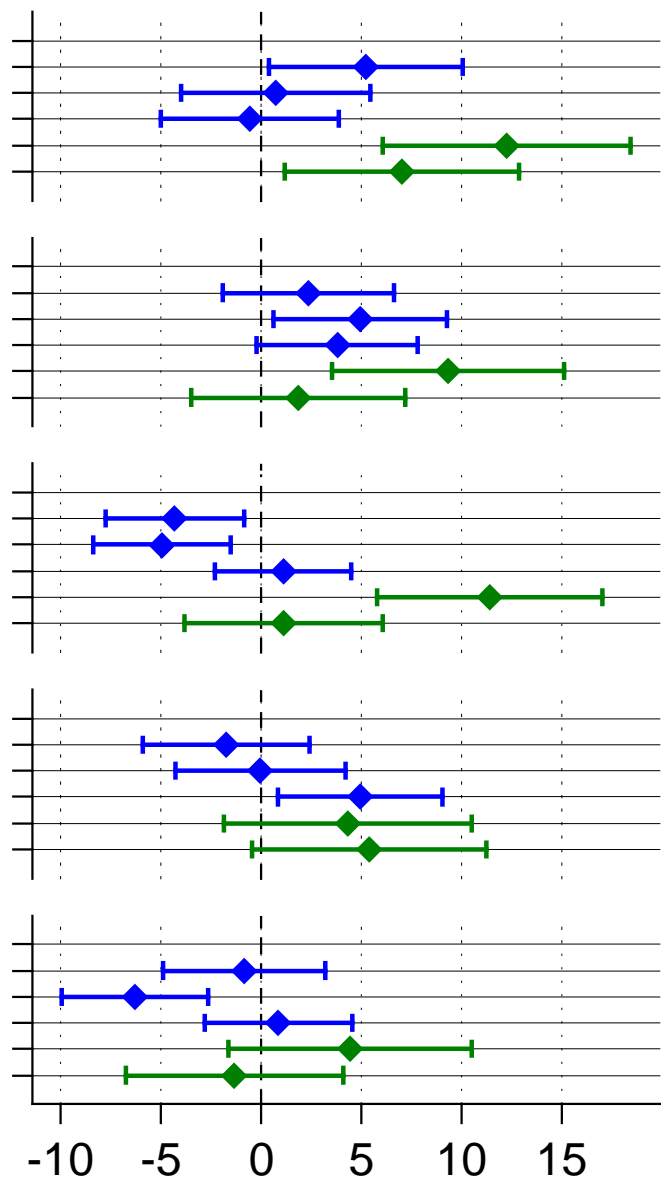
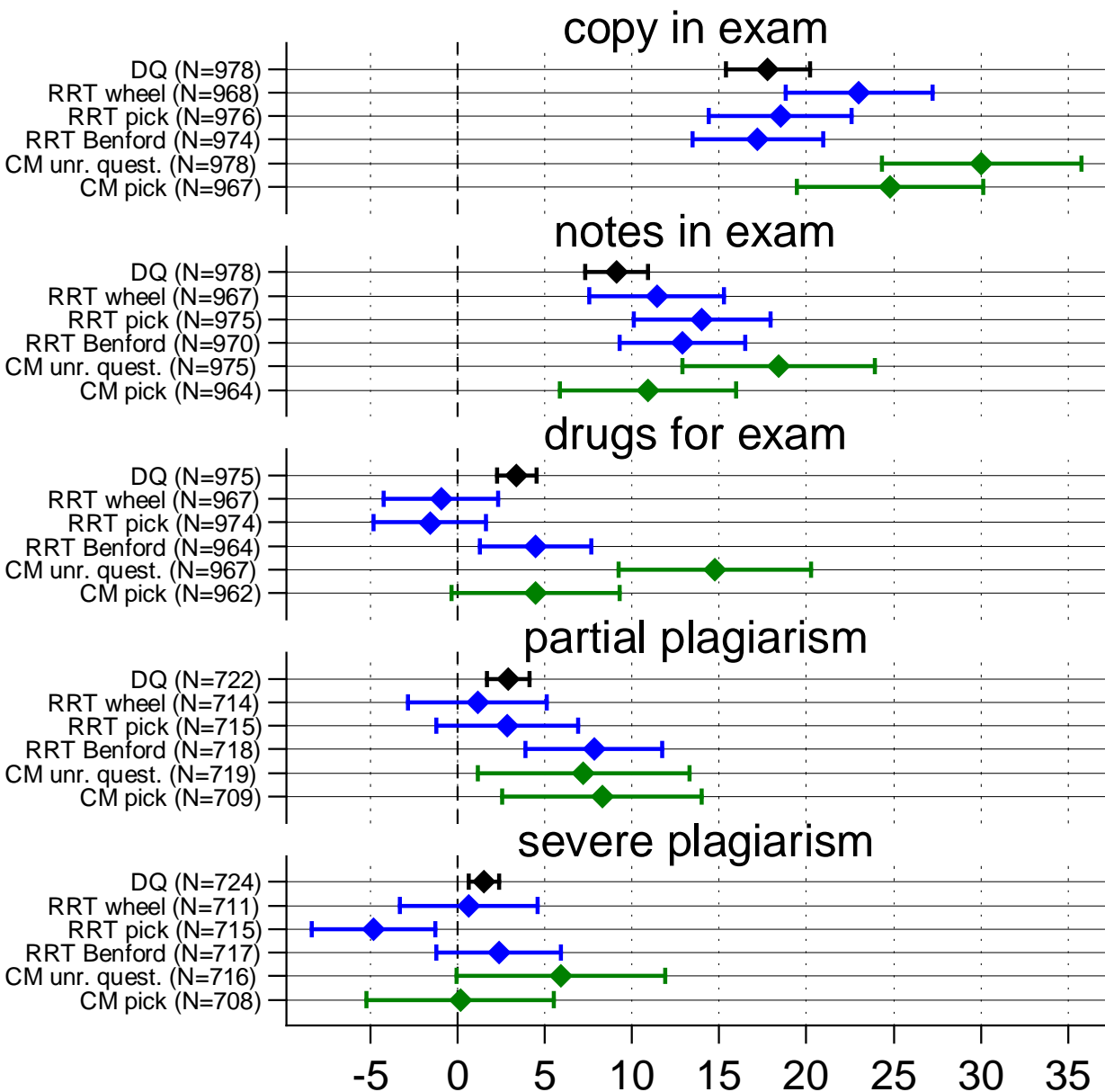
Prevalence estimates (with 95%-ci) by technique



Prevalence estimates (with 95%-ci) by implementation

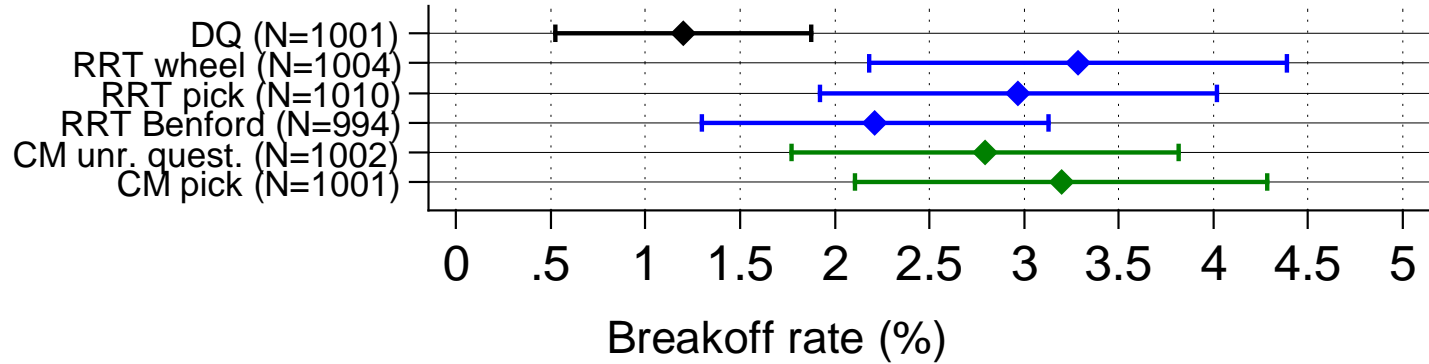
Prevalence estimate

Difference to DQ

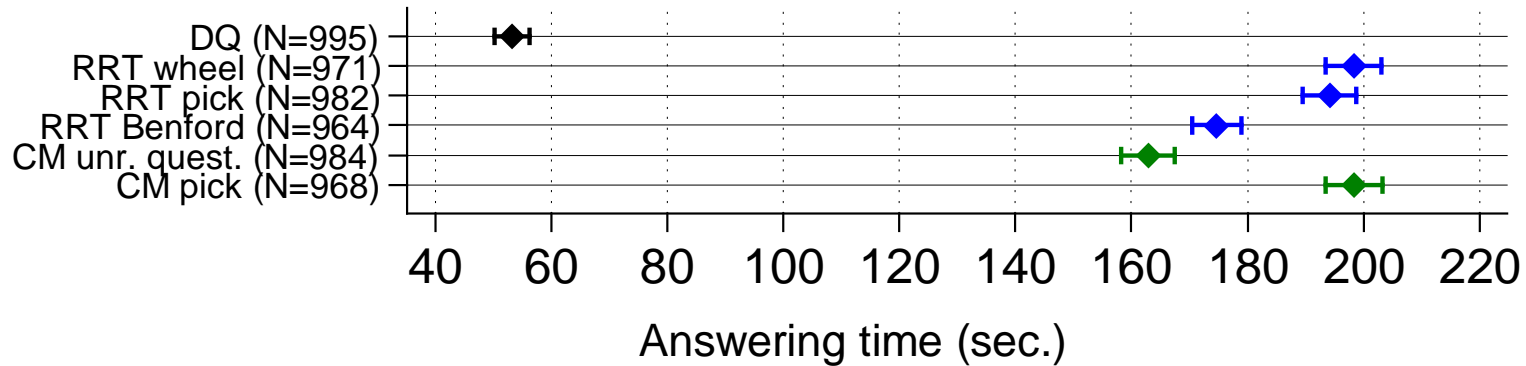


Breakoff rates and response time by implementation

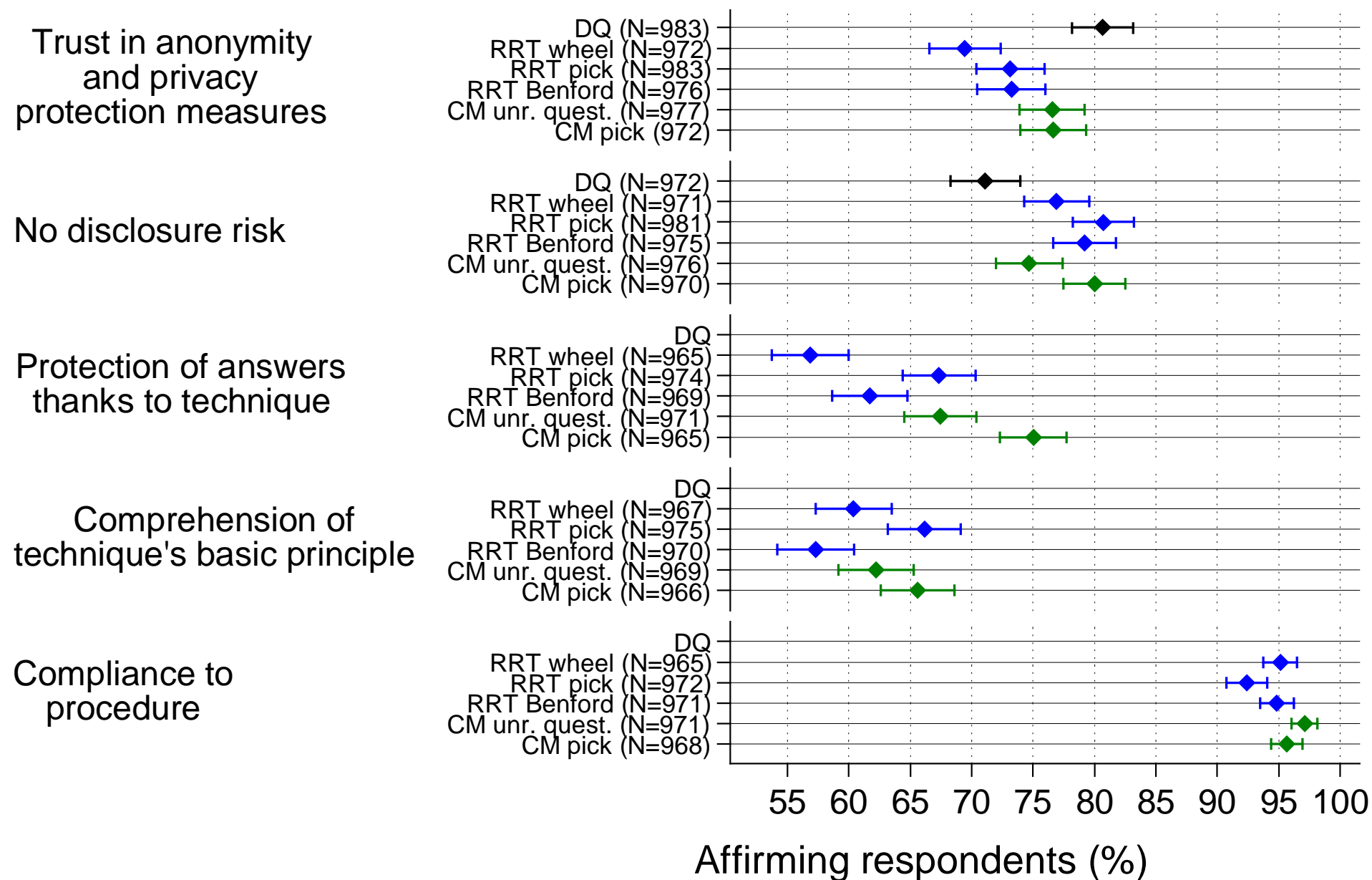
Breakoff after start sensitive items



Time to answer sensitive items



Respondents' evaluation by implementation



Conclusions

- RRT does not seem to be a good method for online surveys.
 - Although a lot of effort has been put into pretesting and finding good implementations, no convincing evidence could be found that RRT yields more valid estimates than DQ.
 - With RRT ‘Benford’ performing somewhat better than the other RRT implementations...
- The Crosswise Model produced significantly higher prevalence estimates than DQ for four of the five sensitive items. If one accepts the ‘more-is-better’ assumption, the CM succeeds in eliciting more truthful answers to sensitive questions and, hence, produces the most valid estimates.
- Of course, the use of the CM is still in an early phase and the method certainly needs further testing to draw a final conclusion about its validity.

Thank you!

References I

- Coutts, E. und B. Jann. 2011. Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT), *Sociological Methods & Research*, Vol. 40, 196-193.
- Coutts, Elisabeth, Benn Jann, Ivar Krumpal, and Anatol-Fiete Näher. 2011. "Plagiarism in Student Papers: Prevalence Estimates Using Special Techniques for Sensitive Questions." *Journal of Economics and Statistics* 231:749-760.
- de Jong, Martijn G., Rik Pieters, and Jean-Paul Fox. 2010. "Reducing Social Desirability Bias Through Item Randomized Response: An Application to Measure Underreported Desires." *Journal of Marketing Research (JMR)* 47:14-27.
- Edgell, S. E., S. Himmelfarb und K. L. Duchan. 1982. Validity of forced responses in a randomized-response model, *Sociological Methods & Research*, Vol. 11, 89-100.
- Fox, J. A. und P. E. Tracy. 1986. *Randomized response: A method for sensitive surveys*, Newbury Park, CA: Sage.
- Holbrook, A. L. und J. A. Krosnick. 2010. Measuring Voter Turnout By Using The Randomized Response Technique: Evidence Calling Into Question The Method's Validity, *Public Opin Q*, Vol. 74, 328-343.
- Jann, B., J. Jerke und I. Krumpal. 2011. Asking Sensitive Questions Using the Crosswise Model: Some Experimental Results, *Public Opinion Quarterly*.
- Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. "Social Desirability Bias in CATI, IVR, and Web Surveys." *Public Opinion Quarterly* 72:847-865.

References II

- Lensvelt-Mulders, G. J. L. M. und H. R. Boeije. 2007. Evaluating compliance with a computer assisted randomized response technique: a qualitative study into the origins of lying and cheating, *Computers in Human Behavior*, Vol. 23, 591-608.
- Locander, W., S. Sudman und N. Bradburn. 1976. An Investigation of Interview Method, Threat and Response Distortion, *Journal of the American Statistical Association*, Vol. 71, 269-275.
- Peeters, C. F.W. 2006. Measuring politically sensitive behavior. Using probability theory in the form of randomized response to estimate prevalence and incidence of misbehavior in the public sphere: a test on integrity violations. Amsterdam: Faculty of Social Sciences, Vrije Universiteit Amsterdam.
- Snijders, C., and J. Weesie. 2008. "The online use of randomized response measurement." Paper presented at General Online Research 2008, Hamburg, Germany.
- van der Heijden, P. G. M., G. van Gils, J. Bouts und J. J. Hox. 2000. A Comparison of Randomized Response, Computer-Assisted Self-Interview, and Face-to-Face Direct Questioning. Eliciting Sensitive Information in the Context of Welfare and Unemployment Benefit, *Sociological Methods and Research*, Vol. 28, 505–537.
- Warner, S. L. 1965. Randomized-response: A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, Vol. 60, 63-69.
- Wolter, F., and P. Preisendörfer. 2011. "Asking Sensitive Questions: An Evaluation of the Randomized Response Technique vs. Direct Questioning Using Individual Validation Data." unpublished.
- Yu, J.-W., G.-L. Tian, M.-L. Tang (2008). Two new models for survey sampling with sensitive characteristic: design and analysis. *Metrika* 67(3): 251-263.

Appendix: Items

Wording sensitive items

	Item	Wording (translated from German)
1	copying from fellow students in exam	Have you ever copied from fellow students in exam during your studies?
2	using crib notes in exam	Have you ever illicitly used crib notes in exam during your studies (also notes on mobile phones, calculators or similar)?
3	taking drugs to enhance exam performance	Have you ever used drugs only available on prescription to enhance your exam performance during your studies?
4	partial paper plagiarism	Have you ever deliberately taken a whole passage from another source without marking it as a quote in a submitted paper during your studies?
5	severe paper plagiarism	Have you ever had someone else writing the bulk of a submitted paper or have you handed in someone else's paper as your own during your studies?

Wording evaluation questions

Zum Abschluss interessiert uns noch Ihre Einschätzung dieser Umfrage:

Bitte ganz ehrlich: Wie stark vertrauen Sie unseren Massnahmen zur Anonymität und zum Persönlichkeitsschutz der Teilnehmenden bei dieser Umfrage?

gar nicht



eher nicht



teils, teils



eher stark



sehr stark



Für wie wahrscheinlich halten Sie es, dass aufgrund dieser Umfrage nachvollzogen werden kann, ob eine bestimmte Teilnehmerin/ein Teilnehmer dieser Umfrage eines der erfragten heiklen Verhalten (Abschreiben, Spickzettel, Plagiate etc.) begangen hat?

unmöglich



sehr un-
wahrscheinlich



eher un-
wahrscheinlich



eher
wahrscheinlich



sehr
wahrscheinlich



Wording evaluation questions

Um Ihren Persönlichkeitsschutz bedingungslos sicherzustellen, haben wir bei einzelnen heiklen Fragen eine spezielle Befragungsmethode benutzt:

Wie umständlich war für Sie das Befolgen dieser speziellen Befragungsmethode?

sehr umständlich eher umständlich teils, teils eher nicht umständlich überhaupt nicht umständlich

Denken Sie, dass Sie die spezielle Befragungsmethode jeweils korrekt befolgt haben?

(mit Ausnahme der Beispielfrage zum Schwarzfahren)

bestimmt nicht eher nicht teils, teils eher ja ja, ganz bestimmt

Was ist Ihre persönliche Einschätzung:

Schützt die verwendete spezielle Befragungsmethode Ihre Antworten auf die heiklen Fragen zu 100%?

bestimmt nicht eher nicht teils, teils eher ja ja, ganz bestimmt

Für wie sinnvoll halten Sie den Einsatz dieser Befragungsmethode, um die Antworten der UmfrageteilnehmerInnen auf heikle Fragen zu schützen?

gar nicht sinnvoll eher wenig sinnvoll teils, teils eher sinnvoll sehr sinnvoll

Können Sie nachvollziehen, weshalb die verwendete spezielle Befragungsmethode Ihre Antworten zu 100% schützt?

nein, überhaupt nicht eher nicht teils, teils eher ja ja, ganz bestimmt

Appendix: additional tables

Prevalence estimates by technique

	copy	notes	drugs	partial	severe
Level					
DQ	17.8 (1.2)	9.1 (0.9)	3.4 (0.6)	2.9 (0.6)	1.5 (0.5)
RRT	19.6 (1.2)	12.8 (1.1)	0.6 (1.0)	3.9 (1.2)	-0.6 (1.1)
CM	27.4 (2.0)	14.7 (1.9)	9.6 (1.9)	7.8 (2.1)	3.1 (2.1)
Difference					
RRT - DQ	1.8 (1.7)	3.7* (1.4)	-2.7* (1.1)	1.0 (1.3)	-2.1 (1.2)
CM - DQ	9.6*** (2.3)	5.6** (2.1)	6.3** (2.0)	4.8* (2.2)	1.5 (2.1)
N	5841	5829	5809	4297	4291

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

Prevalence estimates by implementation

	copy	notes	drugs	partial	severe
Level					
Direct questioning	17.8 (1.2)	9.1 (0.9)	3.4 (0.6)	2.9 (0.6)	1.5 (0.5)
RRT random wheel	23.0 (2.1)	11.4 (2.0)	-0.9 (1.7)	1.2 (2.0)	0.7 (2.0)
RRT pick a number	18.5 (2.1)	14.0 (2.0)	-1.6 (1.6)	2.9 (2.1)	-4.8 (1.8)
RRT Benford	17.2 (1.9)	12.9 (1.8)	4.5 (1.6)	7.8 (2.0)	2.4 (1.8)
CM unrelated quest~n	30.0 (2.9)	18.4 (2.8)	14.8 (2.8)	7.2 (3.1)	5.9 (3.1)
CM pick a number	24.8 (2.7)	10.9 (2.6)	4.5 (2.5)	8.3 (2.9)	0.2 (2.7)
Difference					
RRT rand. wheel - DQ	5.2* (2.5)	2.3 (2.2)	-4.3* (1.8)	-1.8 (2.1)	-0.8 (2.1)
RRT pick number - DQ	0.7 (2.4)	4.9* (2.2)	-5.0** (1.7)	-0.1 (2.2)	-6.3*** (1.9)
RRT Benford - DQ	-0.6 (2.3)	3.8 (2.0)	1.1 (1.7)	4.9* (2.1)	0.8 (1.9)
CM unr. quest. - DQ	12.2*** (3.2)	9.3** (3.0)	11.4*** (2.9)	4.3 (3.1)	4.4 (3.1)
CM pick number - DQ	7.0* (3.0)	1.8 (2.7)	1.1 (2.5)	5.4 (3.0)	-1.3 (2.8)
Observations	5841	5829	5809	4297	4291

Breakoff rates, response time and respondents' evaluation by implementation

	N	Breakoff	Time	Anonym.	NoRisk	Protect	Compreh.	Compl.
Direct questioning	1001	1.2 (0.3)	53.1 (1.5)	80.7 (1.3)	71.1 (1.4)			
RRT random wheel	1004	3.3 (0.6)	198.3 (2.4)	69.4 (1.5)	76.9 (1.4)	56.9 (1.6)	60.4 (1.6)	95.1 (0.7)
RRT pick a number	1010	3.0 (0.5)	194.1 (2.4)	73.1 (1.4)	80.7 (1.3)	67.4 (1.5)	66.2 (1.5)	92.4 (0.9)
RRT Benford	994	2.2 (0.5)	174.7 (2.2)	73.3 (1.4)	79.2 (1.3)	61.7 (1.6)	57.3 (1.6)	94.9 (0.7)
CM unrel. question	1002	2.8 (0.5)	162.8 (2.3)	76.6 (1.4)	74.7 (1.4)	67.5 (1.5)	62.2 (1.6)	97.1 (0.5)
CM pick a number	1001	3.2 (0.6)	198.4 (2.5)	76.6 (1.4)	80 (1.3)	75.0 (1.4)	65.6 (1.5)	95.7 (0.7)

SE in parenthesis.

N: Number of assigned respondents

Breakoff: % who did not complete survey after reaching the sensitive questions

Time: Av. time (seconds) to answer the sensitive questions (highest 2.5 percentiles excluded)

Anonym.: % who trust in anonymity and privacy protection measures

NoRisk: % who think there is no disclosure risk

Protect: % who think their answers are protected thanks to RRT/CM

Compreh.: % who think they comprehend why RRT/CM protects their answers

Compl.: % who think they complied with RRT/CM procedure