Asking Sensitive Questions in Online Surveys: An Experimental Comparison of the Randomized Response Technique and the Crosswise Model

Ben Jann

University of Bern, jann@soz.unibe.ch

In collaboration with Andreas Diekmann and Marc Höglinger (ETH Zurich)

University of Vienna
Department of Methods in the Social Sciences
January 30, 2013

Sensitive Questions in Survey Research

- Population surveys using standardized interviews containing are one of the most important instruments of social science research.

- Such surveys work well in many cases. A fundamental problem, however, is how to obtain valid information from respondents on topics they don't want to talk about.

Sensitive Questions in Survey Research

- What happens if you ask people directly?

Karl-Theodor zu Guttenberg, February 2011
(German Minister of Defense at that time)

Sensitive Questions in Survey Research

`http://www.youtube.com/watch?v=FmKtFCvC8uQ`

1218 Plagiatsfragmente aus 135 Quellen
auf 371 von 393 Seiten (94.4%)
in 10421 plagiierten Zeilen (63.8%)
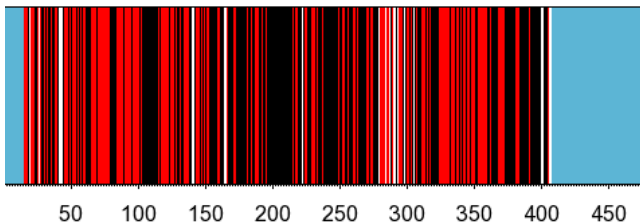
Stand: 03.04.2011 11:55 Uhr

■ Seiten, auf denen Plagiate gefunden wurden

■ Seiten mit Plagiaten aus mehreren Quellen

☐ Seiten, auf denen bisher keine Plagiate gefunden wurden

▨ Das Inhaltsverzeichnis (Seiten 1-14) und die Anhänge (ab Seite 408)
wurden nicht bei der Berechnung des Prozentualwertes mit einbezogen

http://de.guttenplag.wikia.com

135 Quellen auf 371 Seiten mit eigener Farbe
1218 Fragmente mit 10421 Zeilen (63.8%)
Stand: 03.04.2011 11:55 | http://de.gutenberg-wikis.com/wiki/Benutzer:Userff

Sensitive Questions in Survey Research

- Misreporting of sensitive issues is a pervasive problem: Survey respondents might not tell the truth if asked about sensitive topics 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.)

Sensitive Questions in Survey Research

- Online surveys offer more anonymity and privacy to the respondents than interviewer-administered surveys. Yet, misreporting of sensitive issues is a problem also in online surveys.

- For example, Kreuter, Presser and Tourangeau (2008) report a decrease in the misreporting of sensitive information in online mode compared to CATI.

- However, there was still a substantial amount of misreporting in their study:
  - false denial of poor grade point average:
    - 83% CATI vs. 62% online mode
  - false denial of having received an unsatisfactory grade:
    - 33% CATI vs. 20% online mode

Alternatives to Direct Questioning

- Because direct questioning does not work, various alternatives have been proposed in the literature:

Asking the Embarrassing Question

BY ALLEN H. BARTON

University of Chicago

THE POLLSTER's greatest ingenuity has been devoted to finding ways to ask embarrassing questions in non-embarrassing ways. We give here examples of a number of these techniques, as applied to the question, "Did you kill your wife?"

(POQ 22/1958: 67-68)

Alternatives to Direct Questioning

1. The Casual Approach:
    "Do you happen to have murdered your wife?"

2. The Numbered Card:
    Would you please read off the number on this card which corresponds to what became of your wife?" (HAND CARD TO RESPONDENT)
    1. Natural death
    2. I killed her
    3. Other (What?)
    (GET CARD BACK FROM RESPONDENT BEFORE PROCEEDING!)

3. The Everybody Approach:
    "As you know, many people have been killing their wives these days. Do you happened to have killed yours?"

Alternatives to Direct Questioning

4. The "Other people" Approach:
   - (a) "Do you know any people who have murdered their wives?"
   - (b) "How about yourself?"

5. The Sealed Ballot Technique:

   In this version you explain that the survey respects people's right to anonymity in respect to their marital relations, and that they themselves are to fill out the answer to the question, seal it in an envelope, and drop it in a box conspicuously labelled "Sealed Ballot Box" carried by the interviewer.

Alternatives to Direct Questioning

A.
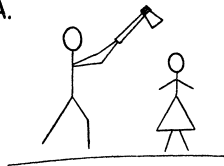
B.

C.
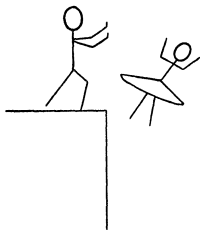
D.
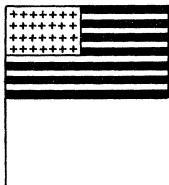
6. The Projective Technique:
   "What thoughts come to mind as you look at the following pictures?"

   (Note: The relevant responses will be evinced by picture D.)

Alternatives to Direct Questioning

7. The Kinsey Technique:

    Stare firmly into respondent's eyes and ask in simple, clearcut language such as that to which the respondent is accustomed, and with an air of assuming that everyone has done everything,

        "Did you ever kill your wife?"

8. Putting the question at the end of the interview.

Alternatives to Direct Questioning

- Gaining valid answers to sensitive questions is difficult. People may underreport sensitive behavior (and overreport socially desirable behavior).

  - *"A question is sensitive when it asks for a socially undesirable answer, when it asks, in effect, that the respondent admits he or she has violated a social norm"* (Tourangeau/Yan 2007: 860).

- Various techniques have been developed to guarantee anonymity and minimize the respondent's feelings of jeopardy, so that more honest answers can be expected.

- Three such "dejeopardizing" techniques are:
  - The randomized response technique (RRT)
  - The item count technique (ICT)
  - The crosswise model (CM)

Outline

- Introduction
- Sensitive question techniques
  - The randomized response technique (RRT)
  - The item count technique (ICT)
  - The crosswise model (CM)
- Three preliminary studies on plagiarism
- Main study on cheating in exams and plagiarism
- Conclusions

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
  - or 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, if the properties of the randomizing device are known, a prevalence estimate for the sensitive question can be derived.

Example (forced response RRT)
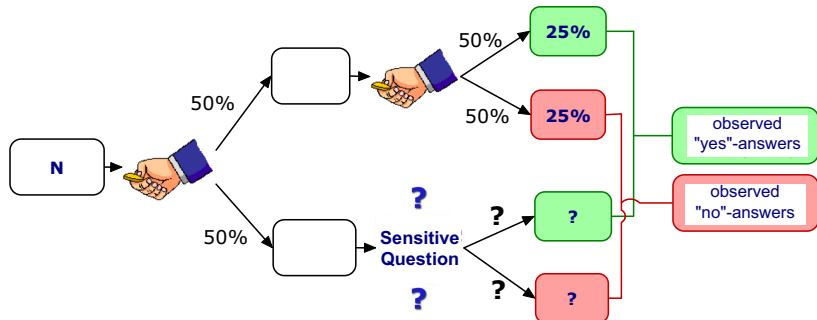
$$\Pr(\text{observed yes}) = \Pr(\text{sensitive question}) \cdot \pi + \Pr(\text{surrogate yes})$$

$$\Rightarrow \quad \pi = \frac{\Pr(\text{observed yes}) - \Pr(\text{surrogate yes})}{\Pr(\text{sensitive question})}$$

Warner's RRT (Warner 1965)
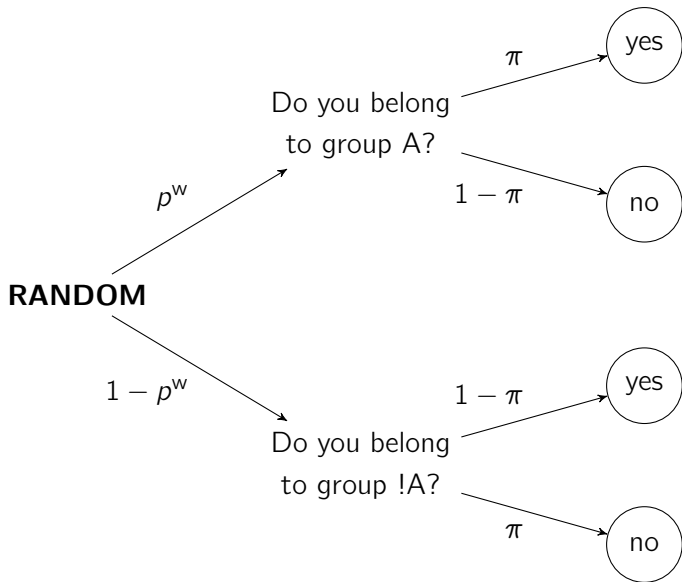
Warner's RRT (Warner 1965)

- "Group A" is the sensitive group, i.e. belongig to group A is equivalent to answering "yes" to the sensitive question ($SQ = 1$).

- Point estimate for $\pi = \Pr(\text{"belongs to group A"}) = \Pr(SQ = 1)$?

$$\Pr(\text{"yes"}) = \lambda = p^w \pi + (1 - p^w)(1 - \pi)$$

$$\pi = \frac{\lambda + p^w - 1}{2p^w - 1}, \quad p^w \neq 0.5$$

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} y_i \quad \text{where} \quad y_i = \begin{cases} 1 & \text{if "yes"} \\ 0 & \text{if "no"} \end{cases}$$

$$\hat{\pi} = \frac{\hat{\lambda} + p^w - 1}{2p^w - 1}$$

Warner's RRT (Warner 1965)

Sampling variance of $\hat{\pi}$?

Delta method:

$$\text{Var}\{f(x)\} = \left(\frac{df(x)}{dx}\right)^2 \text{Var}(x)$$

if $f(x)$ is a linear transformation.

$$f(\hat{\lambda}) = \frac{\hat{\lambda} + p^w - 1}{2p^w - 1} \quad \Rightarrow \quad f' = \frac{1}{2p^w - 1}$$

$$\widehat{\text{Var}}(\hat{\lambda}) = \frac{\hat{\lambda}(1 - \hat{\lambda})}{n}$$

$$\widehat{\text{Var}}(\hat{\pi}) = \frac{\hat{\lambda}(1 - \hat{\lambda})}{n(2p^w - 1)^2} = \frac{\hat{\pi}(1 - \hat{\pi})}{n} + \frac{p^w(1 - p^w)}{n(2p^w - 1)^2}$$

Forced Response RRT (Boruch 1971)
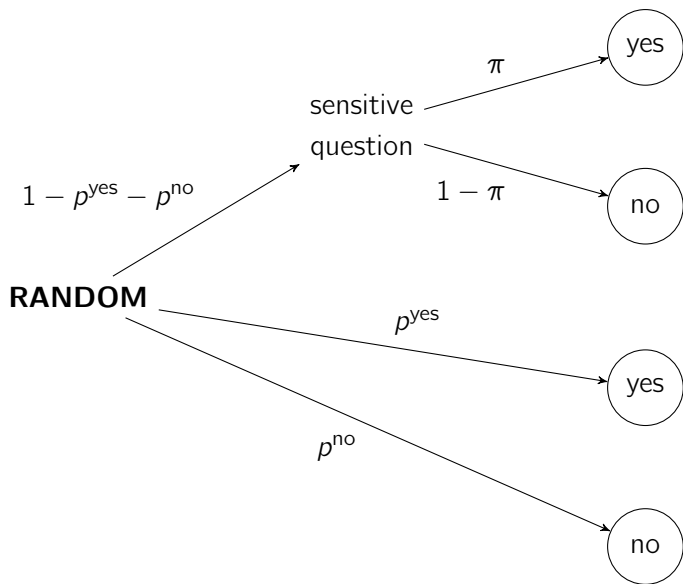
Forced Response RRT (Boruch 1971)

$$\Pr(\text{"yes"}) = \lambda = (1 - p^{\text{yes}} - p^{\text{no}})\pi + p^{\text{yes}}$$

Hence

$$\hat{\pi} = \frac{\hat{\lambda} - p^{\text{yes}}}{1 - p^{\text{yes}} - p^{\text{no}}}$$

and

$$\widehat{\text{Var}}(\hat{\pi}) = \frac{\hat{\lambda}(1 - \hat{\lambda})}{n(1 - p^{\text{yes}} - p^{\text{no}})^2}$$

Unrelated Question RRT (Horvitz et al. 1967)
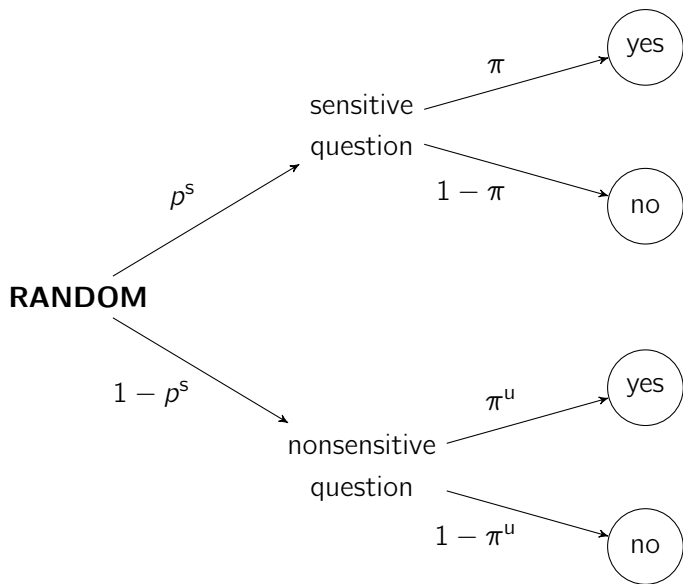
Unrelated Question RRT (Horvitz et al. 1967)

$$\Pr(\text{``yes''}) = \lambda = p^s \pi + (1 - p^s)\pi^u$$

Let

$$p^s = 1 - p^{yes} - p^{no}, \quad \pi^u = \frac{p^{yes}}{p^{yes} + p^{no}}$$

then

$$\lambda = (1 - p^{yes} - p^{no})\pi + (1 - (1 - p^{yes} - p^{no}))\frac{p^{yes}}{p^{yes} + p^{no}}$$

$$= (1 - p^{yes} - p^{no})\pi + p^{yes}$$

Unrelated Question RRT (Horvitz et al. 1967)

- Hence, if π^u is known, the Unrelated Question RRT is formally equivalent the Forced Response RRT with

$$p^{yes} = (1 - p^s)\pi^u, \quad p^{no} = (1 - p^s)(1 - \pi^u)$$

- If π^u is unkown, it has to be estimated from a control sample. This does not change the formula for the point estimate, but it has consequences for the sampling variance (increase). Use bootstrap for variance estimation in this case.

- Alternatively, here's the variance formula (assuming that π^u is estimated using an independent sample):

$$\hat{\pi} = \frac{1}{p^s}\hat{\lambda} - \frac{1 - p^s}{p^s}\hat{\pi}^u \Rightarrow \widehat{Var}(\hat{\pi}) = \left(\frac{1}{p^s}\right)^2 \widehat{Var}(\hat{\lambda}) + \left(\frac{1 - p^s}{p^s}\right)^2 \widehat{Var}(\hat{\pi}^u)$$

Generalized regression estimator for RRT

- Let

  Y_i response ($Y_i = 1$ if "yes" in RRT or "A" in CM, else $Y_i = 0$)

  λ_i probability of $Y_i = 1$

  π_i (unknown) prevalence of sensitive item

  p_i^{w} probability of the non-negated question in Warner's RRT (prevalence of nonsensitive item in CM)

  p_i^{yes} probability of a forced "yes"

  p_i^{no} probability of a forced "no"

- Then

  $$\lambda_i = (1 - p_i^{yes} - p_i^{no})p_i^{w}\pi_i + (1 - p_i^{yes} - p_i^{no})(1 - p_i^{w})(1 - \pi_i) + p_i^{yes}$$

  and hence

  $$\pi_i = \frac{\lambda_i - (1 - p_i^{yes} - p_i^{no})(1 - p_i^{w}) - p_i^{yes}}{(2p_i^{w} - 1)(1 - p_i^{yes} - p_i^{no})}$$

Generalized regression estimator for RRT
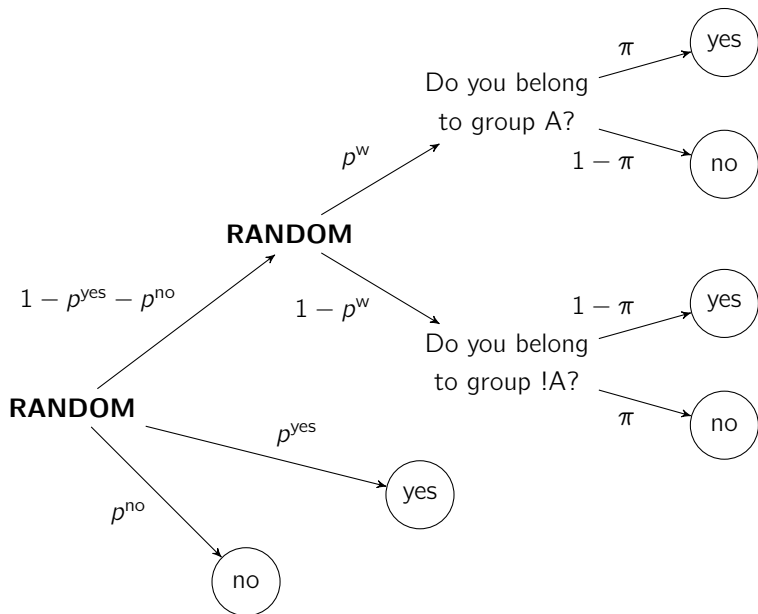
Generalized regression estimator for RRT

- By parametrizing π_i we can formulate regression models.
- For example, assuming $\pi_i = X_i'\beta$, we can estimate β by applying least squares regression to a transformed response variable

$$\tilde{Y}_i = \frac{Y_i - (1 - p_i^{\text{yes}} - p_i^{\text{no}})(1 - p_i^{\text{w}}) - p_i^{\text{yes}}}{(2p_i^{\text{w}} - 1)(1 - p_i^{\text{yes}} - p_i^{\text{no}})}$$

- This is because

$$E(\text{SQ} = 1 | X_i) = \frac{E(Y_i | X_i) - (1 - p_i^{\text{yes}} - p_i^{\text{no}})(1 - p_i^{\text{w}}) - p_i^{\text{yes}}}{(2p_i^{\text{w}} - 1)(1 - p_i^{\text{yes}} - p_i^{\text{no}})}$$

Generalized regression estimator for RRT

- More resonable might be to assume a functional form such as $\ln(\pi_i/(1 - \pi_i)) = X_i'\beta$ (logit), i.e. $\pi_i = e^{X_i'\beta}/(1 + e^{X_i'\beta})$.

- In this case, we can derive the log likelihood as

$$\ln L = \sum_{i=1}^{n} \left[Y_i \ln(\lambda_i) + (1 - Y_i)\ln(1 - \lambda_i) \right]$$

$$= \sum_{i=1}^{n} \left[Y_i \ln(R_i) + (1 - Y_i)\ln(S_i) - \ln(1 + e^{X_i\beta}) \right]$$

with

$$R_i = c_i + q_i e^{X_i'\beta} \qquad c_i = (1 - p_i^{\text{yes}} - p_i^{\text{no}})(1 - p_i^{\text{w}}) + p_i^{\text{yes}}$$

$$S_i = (1 - c_i) + (1 - q_i)e^{X_i'\beta} \quad q_i = (1 - p_i^{\text{yes}} - p_i^{\text{no}})p_i^{\text{w}} + p_i^{\text{yes}}$$

and estimate β using maximum likelihood methods.

Two Stata commands

- Least-squares estimation with $\pi_i = X_i'\beta$ (Jann 2008):

  rrreg *depvar* $\left[\textit{indepvars}\right]$ $\left[\textit{if}\right]$ $\left[\textit{in}\right]$ $\left[\textit{weight}\right]$ $\left[\,,\textit{regress\_options}\right.$
        $\left.\texttt{pwarner}(\#|\textit{varname})\ \texttt{pyes}(\#|\textit{varname})\ \texttt{pno}(\#|\textit{varname})\,\right]$

- Maximum likelihood estimation with $\pi_i = e^{X_i'\beta}/(1 + e^{X_i'\beta})$ (Jann 2005):

  rrlogit *depvar* $\left[\textit{indepvars}\right]$ $\left[\textit{if}\right]$ $\left[\textit{in}\right]$ $\left[\textit{weight}\right]$ $\left[\,,\textit{logit\_options}\right.$
        $\left.\texttt{pwarner}(\#|\textit{varname})\ \texttt{pyes}(\#|\textit{varname})\ \texttt{pno}(\#|\textit{varname})\,\right]$

- rrlogit may make more sense in terms of functional form. However, rrreg is more robust, especially if there is noncompliance with the RRT procedure.

The Item Count Technique (ICT)

(see, e.g., Dalton et al. 1994, Raghavarao and Federer 1979)

- Given a list of statements, respondents report how many of them are true, but not which ones. For some respondents the list contains the sensitive item, for others not (randomized).

- Example: "How many of the following statements apply to you?"

Group A (short list)	Group B (long list)
I have a cat.	I have a cat.
I have blue eyes.	I have blue eyes.
I like country music.	I like country music.
	I use drugs.

- Advantage: Requires no randomization device.

Analysis of Item Count Data

- Estimate of the probability of the sensitive item $\pi = \Pr(SQ = 1)$?

- Mean difference between the two groups:

$$\hat{\pi} = \bar{y}^{LL} - \bar{y}^{SL} = \frac{1}{n^B} \sum_{i \in B} y_i - \frac{1}{n^A} \sum_{i \in A} y_i$$

- Variance of $\hat{\pi}$?

$$\text{Var}(\hat{\pi}) = \text{Var}(\bar{y}^{LL}) + \text{Var}(\bar{y}^{SL})$$

Analysis of Item Count Data

- Double list design:
  - ▶ Both groups answer to two sets of items. In one group, the sensitive item is paired with the first set of nonsensitive items, in the other group the sensitive item is paired with the second set of nonsensitive items.

Set 1:	group A	group B
	nonsensitive item 1	nonsensitive item 1
	nonsensitive item 2	nonsensitive item 2
	nonsensitive item 3	nonsensitive item 3
		sensitive item

Set 2:	group A	group B
	nonsensitive item 4	nonsensitive item 4
	nonsensitive item 5	nonsensitive item 5
	nonsensitive item 6	nonsensitive item 6
	sensitive item	

Analysis of Item Count Data

$$\hat{\pi}_1 = \bar{y}^{LL1} - \bar{y}^{SL1}, \quad \hat{\pi}_2 = \bar{y}^{LL2} - \bar{y}^{SL2}$$

$$
\begin{aligned}
\hat{\pi} = \frac{\hat{\pi}_1 + \hat{\pi}_2}{2} &= \frac{(\bar{y}^{LL1} - \bar{y}^{SL1}) + (\bar{y}^{LL2} - \bar{y}^{SL2})}{2} \\
&= \frac{(\bar{y}^{LL1} - \bar{y}^{SL2}) + (\bar{y}^{LL2} - \bar{y}^{SL1})}{2} \\
&= \frac{\frac{1}{n^B}\sum_{i \in B}(y_{1i} - y_{2i}) + \frac{1}{n^A}\sum_{i \in A}(y_{2i} - y_{1i})}{2}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(\hat{\pi}) &= \frac{\mathrm{Var}(\hat{\pi}_1) + \mathrm{Var}(\hat{\pi}_2) - 2\mathrm{Cov}(\hat{\pi}_1, \hat{\pi}_2)}{4} \\
&= \frac{\mathrm{Var}(\bar{y}^{LL1} - \bar{y}^{SL2}) + \mathrm{Var}(\bar{y}^{LL2} - \bar{y}^{SL1})}{4}
\end{aligned}
$$

Analysis of Item Count Data

- Regression model for single list design:
  - Estimate β by applying least-squares regression (with robust standard errors) to

  $$Y_i = (LL_i \cdot X_i)'\beta + X_i'\gamma + \epsilon_i$$

  (For more sophisticated approaches see Glynn 2010, Imai 2010, Blair and Imai 2012.)

- Regression model for double list design:
  - Approach 1: estimate separate models (as above) for Y_1 and Y_2, combine estimates using suest to obtain joint variance matrix, compute average coefficients using lincom
  - Approach 2: estimate a system of equations (e.g. using sureg) for Y_1 and Y_2 with the contraint that the coefficients are the same

Crosswise Model (Yu et al. 2008)

- Ask a sensitive question and a nonsensitive question and let the respondent indicate whether . . .

    A the answers to the questions are the same (both "yes" or both "no")

    B the answers are different (one "yes", the other "no")

		nonsensitive question	
		no	yes
sensitive question	no	**A (same)**	**B (different)**
	yes	**B (different)**	**A (same)**

  ▸ Assumtion: The two questions are uncorrelated.
  ▸ $p = \Pr(\text{"yes"})$ of the nonsensitive question must not be 0.5.

Crosswise Model (Yu et al. 2008)

- Prevalence estimate:

$$Pr(A) = \lambda = (1 - \pi) \cdot (1 - p) + \pi \cdot p$$

$$\Rightarrow \quad \hat{\pi} = \frac{\hat{\lambda} + p - 1}{2 \cdot p - 1}$$

- The Crosswise Model is formally equivalent to Warner's RRT with $p^w = p$.

Crosswise Model (Yu et al. 2008)
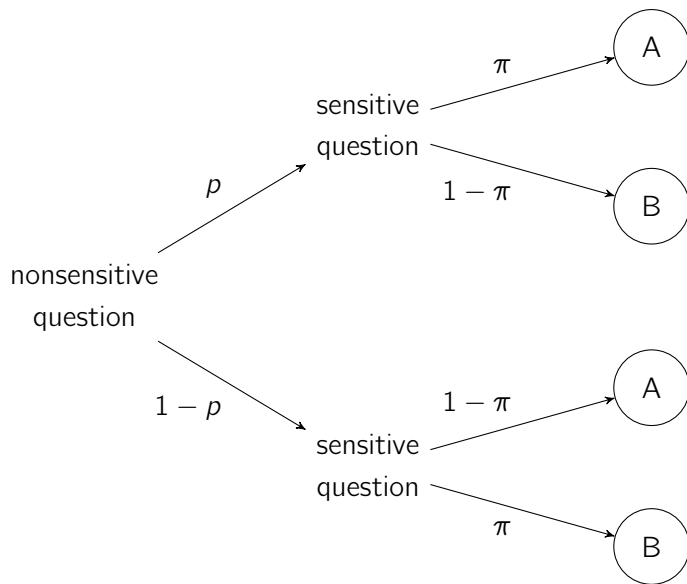
The Crosswise Model: Let's practice

- Two questions:

    1. Is your mother's birthday in January or February?

    2. Did you ever falsify your data or results?

        (e.g. edit data points or delete observations so that hypothesis is confirmed, falsify entire dataset, invent or manipulate reported results)

- Compare your answers: Are they the same or different?

    ▸ Write "**A**" if they are **the same** (both Yes or both No)

    ▸ Write "**B**" if they are **different** (one Yes, the other No)

Three preliminary studies on plagiarism

(Jann/Jerke/Krumpal 2012, Coutts/Jann/Krumpal/Näher 2011)

- Study 1
    - Web-Survey among student of ETH Zurich in 2005
    - Response rate 33 Percent
    - Comparing direct questioning (DQ) to forced response RRT
- Study 2
    - Web-Survey among students of the University of Konstanz in 2009
    - Response rate 24 Percent
    - Comparing direct questioning (DQ) to the doule-list Item Count Technique (ICT)
- Study 3
    - Classroom P&P survey at ETH Zurich, University of Leipzig, and LMU Munich in 2009
    - Comparing direct questioning (DQ) to the Crosswise Model (CM)

Study 1: Using RRT to Measure Plagiarism

Nimm bitte eine Münze 🪙 zur Hand und führe einen Münzwurf durch. Beantworte gemäss ⬚ Ergebnis die entsprechende Frage:

Hast du <u>Kopf</u> geworfen, dann beantworte bitte die folgende Frage:

Hast du in einer dieser Arbeiten (Semester-, Bachelor-,
Master- oder Diplomarbeit) schon einmal bewusst ein Zitat
nicht gekennzeichnet?

Hast du <u>Zahl</u> geworfen, dann beantworte bitte die folgende Frage:

Bitte nimm **nochmals** die Münze 🪙 zur Hand und führe
einen **Münzwurf** durch. Ist das Ergebnis 'Kopf' so beantworte
die Frage mit Ja. Im anderen Fall beantworte die Frage mit
Nein.

○ Ja ○ Nein

Study 2: Using ICT to Measure Plagiarism

Nachfolgend finden Sie vier Gruppen mit verschiedenen Aussagen. Zwei der Gruppen enthalten je eine Aussage, zu der man vielleicht nur ungern Auskunft gibt.

Zählen Sie deshalb bitte für jede Gruppe nur, wie viele der Aussagen Sie bejahen würden. Diese Zahl geben Sie dann für die entsprechende Gruppe an. Wenn zum Beispiel in einer Gruppe mit insgesamt fünf Aussagen drei auf Sie zutreffen, geben Sie für diese Gruppe als Antwort "3" an.

Diese Befragungsmethode garantiert Ihre Anonymität, da für uns nicht ersichtlich ist, welche der einzelnen Aussagen auf Sie zutreffen. Mit Hilfe der Wahrscheinlichkeitsrechnung ist es uns aber möglich, eine Häufigkeit für die Gesamtheit aller Befragten zu berechnen.

Gruppe 1:
- Ich bin ein sehr spontaner Mensch und manchmal auch ein bisschen chaotisch.
- Die Wahl der Uni fiel mir leicht, da ich mich einfach den Entscheidungen meiner Freunde anschloss.
- Ich spiele regelmässig Schach.
- Ich bin meistens sehr pünktlich.
- Beim Schreiben einer Hausarbeit (z.B. Seminararbeit, Semesterarbeit, Abschlussarbeit, etc.) habe ich schon einmal bewusst eine Textpassage aus einem fremden Werk übernommen, ohne diese als Zitat zu kennzeichnen.

Anzahl Aussagen, die Sie in dieser Gruppe mit "Ja" beantworten würden:

☐

Study 3: Using the Crosswise Model to Measure Plagiarism

Block 1

1. Frage: *Hat Ihre Mutter in den Monaten Januar, Februar oder März Geburtstag?*

2. Frage: *Haben Sie beim Schreiben einer Hausarbeit (z.B. Seminararbeit, Semesterarbeit, Abschlussarbeit, etc.) schon einmal bewusst eine Textpassage aus einem fremden Werk übernommen, ohne diese als Zitat zu kennzeichnen?*

Wie lauten Ihre Antworten auf die beiden Fragen?

☐ **(A) auf beide Fragen *Nein* oder auf beide Fragen *Ja***
☐ **(B) auf genau eine der beiden Fragen *Ja* und auf die andere *Nein***

Results (prevalence of plagiarism in percent)

Study 1	DQ	RRT	Δ
in major papers	12.0 (2.0)	3.7 (4.0)	−8.3 (4.4)
	N = 266	N = 495	
in other papers	19.4 (1.4)	17.6 (2.4)	−1.8 (2.8)
	N = 826	N = 1521	
Study 2	DQ (N = 396)	ICT (N = 846)	Δ
partial plagiarism	8.1 (1.4)	9.0 (4.0)	0.9 (4.2)
severe plagiarism	2.0 (0.7)	−4.0 (4.4)	−6.0 (4.5)
Study 3	DQ (N = 96)	CW (N = 310)	Δ
partial plagiarism	7.3 (2.7)	22.3 (5.5)	15.0 (6.1)
severe plagiarism	1.0 (1.0)	1.6 (5.0)	0.6 (5.1)

Summary of preliminary studies

- RRT and ICT do not seem to work well

- In particular, with the RRT, estimates of plagiarism are even lower than with direct questioning

- Reasons for the failure of RRT
  - difficulties understanding RRT, no trust in RRT
  - "self-protective no" bias
    - ★ respondents who are not guilty are reluctant to give a "yes" answer and, hence, do not comply with the instructions
    - ★ in RRT there is a "dominant strategy": say "no", no matter what

- The Crosswise Model works better
  - easier to understand
  - no obvious self-protective answering strategy

Performance of RRT in online surveys

- Results from studies in which RRT was applied in online surveys indicate that RRT might not work well in this mode.
  - 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, the used RRT implementations were not well suited for online mode.
  - randomizing device not at respondents' immediate reach
  - randomizing device not trustworthy
- The Crosswise Model seems more promising, but there are no online applications yet.

Main study on cheating in exams and plagiarism

- Web survey among students of University of Bern and 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 performance on exam (drugs)
  - partial plagiarism (partial)
  - severe plagiarism/ghostwriting (severe)
- Comparing direct questioning to three variants of RRT and two variants of the Crosswise Model
- Approximately 1000 randomly assigned respondents in each experimental condition.

Question wording

Item	Frageformulierung
Abschreiben	Haben Sie während Ihrem Studium jemals während einer Prüfung von Mitstudierenden abgeschrieben?
Spicken	Haben Sie während Ihrem Studium jemals unerlaubterweise einen Spickzettel (auch Handy-, Taschenrechner-Notizen und Ähnliches) in einer Prüfung verwendet?
leistungsfördernde Substanzen („Doping")	Haben Sie während Ihrem Studium jemals rezeptpflichtige Substanzen/Medikamente eingenommen, um Ihre Leistung an Prüfungen zu steigern?
Teilplagiat	Haben Sie während Ihrem Studium jemals bei einer eingereichten Arbeit bewusst eine ganze Textpassage aus einem fremden Werk übernommen, ohne diese als Zitat zu kennzeichnen?
Vollplagiat	Haben Sie während Ihrem Studium jemals einen Grossteil einer Arbeit durch eine andere Person schreiben lassen oder eine fremde Arbeit als Ihre eigene ausgegeben?

Comparison of 6 experimental conditions
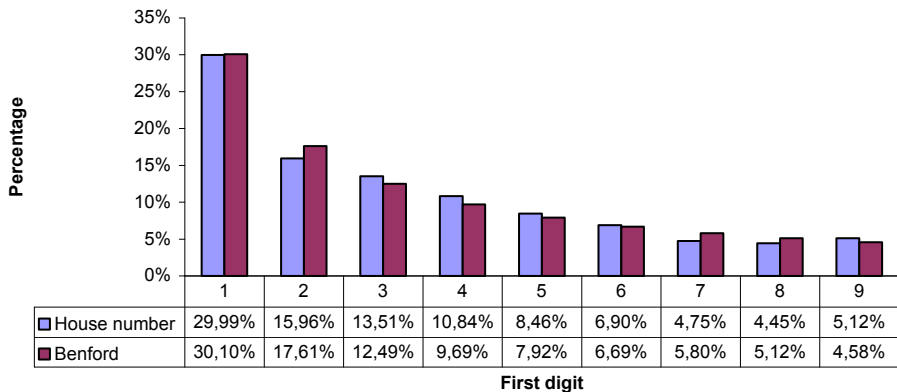
- Direct questioning
  - *example*
- forced response RRT using virtual random wheel
  - *example*
- forced response RRT using "pick a number" method
  - *example*
- RRT using Benford distribution and unrelated questions
  - *example part 1*
  - *example part 2*
- Crosswise Model using unrelated questions
  - *example*
- Crosswise Model using "pick a number" method
  - *example*

Why Benford?

Benford's Law: Pr(d) = log10(1 + 1/d) with d in {1,2,...,9}

House numbers collected from the telephone directory of Zurich

| First digit	1	2	3	4	5	6	7	8	9
House number	29,99%	15,96%	13,51%	10,84%	8,46%	6,90%	4,75%	4,45%	5,12%
Benford	30,10%	17,61%	12,49%	9,69%	7,92%	6,69%	5,80%	5,12%	4,58%

Why Benford?

Denken Sie an eine Ihnen bekannte Person, von der Sie die Wohnadresse wissen. Bitte geben Sie die **erste Ziffer** der Hausnummer der Adresse dieser Person an.

The answers follow Benford's Law! (chi2 = 6.2, 8 df, p = 0.62)
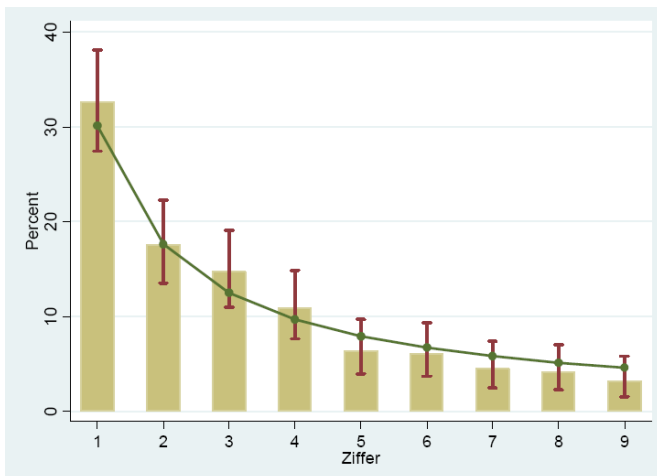
Results: Prevalence estimates by technique
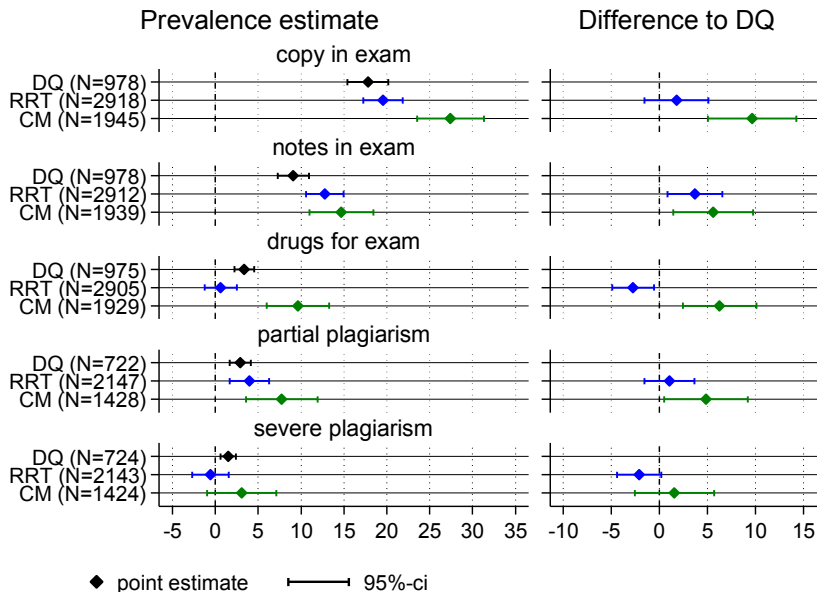
Results: Prevalence estimates by technique

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

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Results: Prevalence estimates by implementation
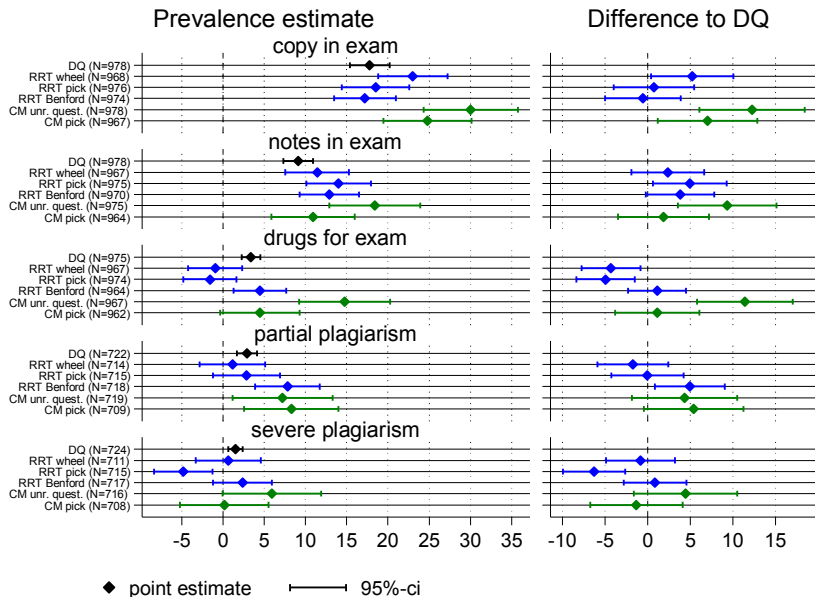
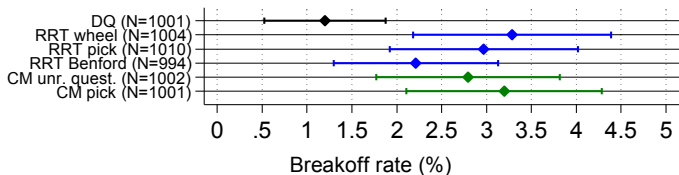# Results: Prevalence estimates by implementation

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

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001
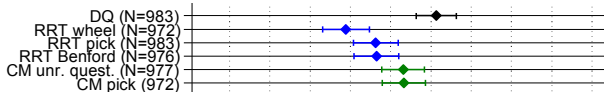
Results: Break-off rates and response times
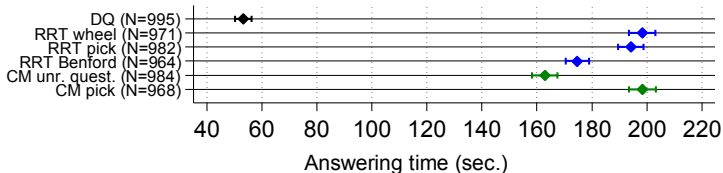
Results: Respondents' opinion

Results

	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

Results: Determinants of sensitive behavior

(Randomized response logistic regression)

	copy	notes	drugs	partial	severe
ETH (ref. UniBE)	-0.150	0.192	-0.577*	0.386	1.032
	(0.104)	(0.140)	(0.264)	(0.308)	(0.697)
Semester (log)	0.072	0.298	-0.026	-0.120	-0.589
	(0.163)	(0.211)	(0.355)	(0.323)	(0.523)
Nbr. exams/papers (log)	0.596***	0.298*	-0.264	0.393*	0.208
	(0.111)	(0.135)	(0.211)	(0.196)	(0.398)
Perceived risk	-0.014***	-0.024***		-0.008	-0.009
	(0.004)	(0.006)		(0.006)	(0.013)
Risk attitude	0.066**	0.095**	0.184**	0.144	-0.037
	(0.025)	(0.032)	(0.066)	(0.078)	(0.132)
Procrastination	0.202***	0.213**	0.189	0.275	0.001
	(0.049)	(0.066)	(0.128)	(0.152)	(0.320)
Stress at exams	0.104*	0.084	0.462***		
	(0.050)	(0.066)	(0.134)		
Constant	-4.615***	-5.202***	-5.791***	-5.843***	-3.706*
	(0.414)	(0.596)	(0.934)	(1.065)	(1.521)
N	5713	5713	5761	4224	4221

Standard errors in parentheses
Additional controls for experimental conditions
* p<0.05, ** p<0.01, *** p<0.001

A little bit of magic: Cheating correction in RRT

- In many RRT designs, the "self-protective no" bias can occur.
- In these designs, some of the respondents are instructed to answer "yes" by the randomization device, even though the sensitive item does not apply to them.
- There is evidence that these respondents often deviate from the instructions and answer "no".
- Such non-compliance introduces a large bias to RRT estimates. It is noteworthy that this bias does not come from respondents who did commit the sensitive behavior and want to conceal it. It comes from respondents who did not and don't want it to look like they did.
- In a standard design, it is not possible to account for such "cheaters". However, if the RRT design parameters are variied, this variation can be used to identify the proportion of cheaters and correct the estimates.

A little bit of magic: Cheating correction in RRT
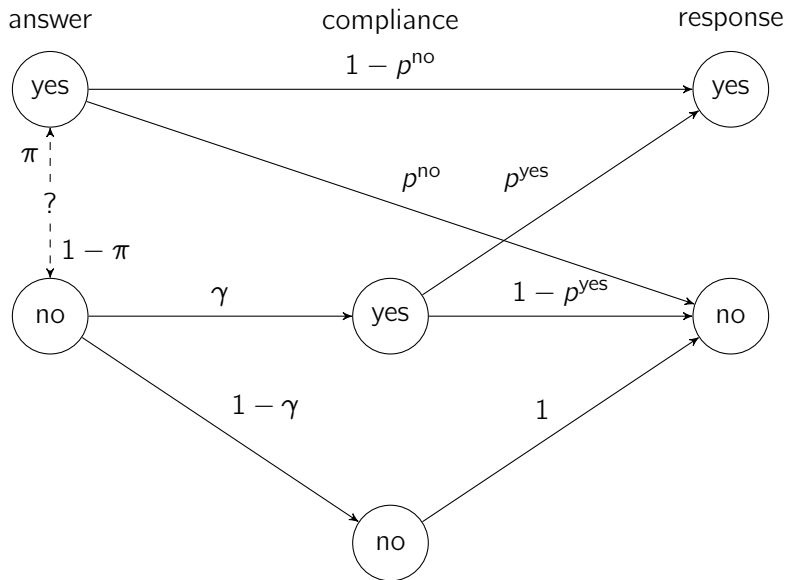
A little bit of magic: Cheating correction in RRT

- Assumptions:
  - There is random variation in p^{yes} and p^{no} between respondents.
  - π and γ do not depend on p^{yes} and p^{no} (which may be justified if the variation in p is small)
  - Respondents do not say "yes" if instructed to say "no" by the randomization device.

- π and γ can then be estimated using the following log likelihood:

$$\ln L = \sum_{i=1}^{n} Y_i \ln(\ell_i) + (1 - Y_i) \ln(1 - \ell_i)$$

with

$$\ell_i = \pi_i(1 - p_i^{\text{no}} - \gamma p_i^{\text{yes}}) + \gamma p_i^{\text{yes}}$$

A little bit of magic: Analysis

```
program define rrcheat_lf
    args lnf theta1 cheat
    local p1 $rrcheat_pyes
    local p2 $rrcheat_pno
    quietly replace `lnf´ = cond($ML_y1, ///
        ln(`theta1´ * (1 - `p2´ - (1-`cheat´)*`p1´) + (1-`cheat´)*`p1´), ///
        ln(1 - (`theta1´ * (1 - `p2´ - (1-`cheat´)*`p1´) + (1-`cheat´)*`p1´)))
end
forv i = 1/5 {
    local depvar: word `i´ of $sqvar
    global rrcheat_pyes pyesQ`i´
    global rrcheat_pno  pnoQ`i´
    ml model lf rrcheat_lf (`depvar´: `depvar´ = ) /cheat if RRT==1
    ml maximize
    eststo `depvar´
}
esttab, nonumb  nostar mti se b(1) transform(100*@ 100) ///
    eqlab(none) coef(main:_cons "RRT adjusted" cheat:_cons "Cheaters")
```

A little bit of magic: Results

	copy	notes	drugs	partial	severe
RRT adjusted	17.9	12.0	16.7	14.3	6.7
	(6.5)	(6.1)	(5.6)	(6.6)	(5.9)
Cheaters	-9.5	-3.6	88.9	54.3	36.1
	(36.1)	(31.9)	(36.9)	(40.1)	(31.8)
N	2855	2855	2849	2105	2104

Standard errors in parentheses

Unadjusted results for comparison:

	copy	notes	drugs	partial	severe
DQ	17.5	8.8	3.4	2.5	1.5
	(1.2)	(0.9)	(0.6)	(0.6)	(0.5)
RRT	19.6	12.7	0.6	4.2	-0.6
	(1.2)	(1.1)	(1.0)	(1.2)	(1.1)
CM	27.2	15.0	9.9	8.2	3.0
	(2.0)	(1.9)	(1.9)	(2.1)	(2.0)

Summary of results

- The Crosswise Model produced significantly higher prevalence estimates and, therefore, clearly outperformed direct questioning (if we are ready to accept the "more-is-better assumption").
  - An exception is the last item (severe plagiarism), where prevalence is very low for all techniques.
- RRT, on the other hand, did not yield higher estimates than direct questioning (even lower and sometimes negative estimates).
  - One reason might be the "self-protective no" bias, which prevents respondents to say "yes" if instructed to do so by the randomizing device.

Methodological conclusions

- The Randomized Response Technique does not seem to be a good method for self-administered surveys. Although we put a lot of effort into pretesting and finding good implementations, no convincing evidence could be found that RRT yields more valid estimates than direct questioning.

    - However, RRT Benford performed somewhat better than the other RRT implementations.

- The Crosswise Model is a promising alternative, since it does not suffer from some of the deficiencies of the RRT.

- Improvement of RRT estimates is possible by correcting for cheating respondents who do not comply with the instructions. Such estimates, however, have low efficiency.

Substantive conclusions

- A substantial proportion of students have cheated on an exam (copying: 20 to 25 percent, crib notes: around 15 percent)
- Using drugs to enhance performance on exams is not uncommon (about 10 percent)
- Rates for partial plagiarism (using a passage from someone else's work without providing proper citation) are about 8 percent. The prevalence of severe plagiarism (hand in someone else's work) is about 3 percent.
- These numbers may not seem too high, but we have to keep in mind that they most likely still underestimate the true prevalence.

References I

- Barton, Allen H. 1958. Asking the Embarrassing Question. Public Opinion Quarterly 22:67-68.

- Blair, G., K. Imai. 2012. Statistical Analysis of List Experiments. *Political Analysis* 20:47-77.

- Boruch, R.F. 1971. Assuring Confidentiality of Responses in Social Research: A Note on Strategies. *The American Sociologist* 6(4): 308-311.

- 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 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.

- Dalton, Dan R., James C. Wimbush, Catherine M. Daily. 1994. Using the unmatched count technique (UCT) to estimate base rates for sensitive behavior. Personnel Psychology 47:817-828.

References II

- 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.

- Droitcour, J., R.A. Caspar, M.L. Hubbard, T.L. Parsely, W. Vischer, T.M. Ezzati. 1991. The item count technique as a method of indirect questioning: A review of its development and a case study application. Pp. 185-210 in: P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, S. Sudman (eds.), Measurement errors in surveys. New York: Wiley.

- Fox, J. A. und P. E. Tracy. 1986. Randomized response: A method for sensitive surveys, Newbury Park, CA: Sage.

- Glynn, A.N. 2010. What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment. Harvard University.

- 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 Opinion Quarterly 74:328-343.

- Horvitz, D.G., B.V. Shah, W.R. Simmons. 1967. The unrelated question randomized response model. Pp. 65-72 in: Proceedings in the Social Science Section. Baltimore: American Statistical Association.

References III

- Imai, K. 2011. Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association* 106:407–16.

- Jann, B. 2005. rrlogit: Stata module to estimate logistic regression for randomized response data. Statistical Software Components S456203, Boston College Department of Economics, revised 12 May 2011. Available from http://ideas.repec.org/c/boc/bocode/s456203.html

- Jann, B. 2008. rrreg: Stata module to estimate linear probability model for randomized response data. Statistical Software Components S456962, Boston College Department of Economics, revised 12 May 2011. Available from http://ideas.repec.org/c/boc/bocode/s456962.html

- Jann, B., J. Jerke und I. Krumpal. 2012. Asking Sensitive Questions Using the Crosswise Model: Some Experimental Results, Public Opinion Quarterly 71: 32-49.

- Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. Social Desirability Bias in CATI, IVR, and Web Surveys. Public Opinion Quarterly 72:847-865.

- Locander, W., S. Sudman und N. Bradburn. 1976. An Investigation of Interview Method, Threat and Response Distortion, Journal of the American Statistical Association 71:269-275.

References IV

- 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.

- Raghavarao, Damaraju, Walter T. Federer. 1979. Block total response as an alternative to the randomized response method in surveys. Journal of the Royal Statistical Society Series B (Statistical Methodology) 41:40-45.

- Snijders, C., and J. Weesie. 2008. The online use of randomized response measurement. Paper presented at General Online Research 2008, Hamburg, Germany.

- Tourangeau, Roger, Ting Yan. 2007. Sensitive questions in surveys. Psychological Bulletin 133:859-883.

- 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 28:505–537.

- Warner, S. L. 1965. Randomized-response: A survey technique for eliminating evasive answer bias, Journal of the American Statistical Association 60:63-69.

References V

- 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 67:251-263.