

NTTS 2009 Session on Treatment of Nonresponse

**To Cope with Nonresponse and Untruthful Answering:
Different Questioning Designs for Different Variables**

Andreas Quatember
IFAS - Department of Applied Statistics
Johannes Kepler University Linz, Austria



1. Introduction

2. The Standardized Randomized Response Technique

3. Measures of Loss of Privacy

4. Optimum Questioning Designs for Different Variables

1. Introduction

U ... an universe of N population units

U_A ... a subset of N_A elements which belong to a class A of a categorical variable under study

U_A^C ... a group of $N - N_A$ elements which do not belong to U_A

Let

$$x_i = \begin{cases} 1 & \text{if } i \in U_A \\ 0 & \text{otherwise} \end{cases}$$

and the parameter of interest be

$$\pi_A = \frac{\sum_U x_i}{N} = \frac{N_A}{N}.$$

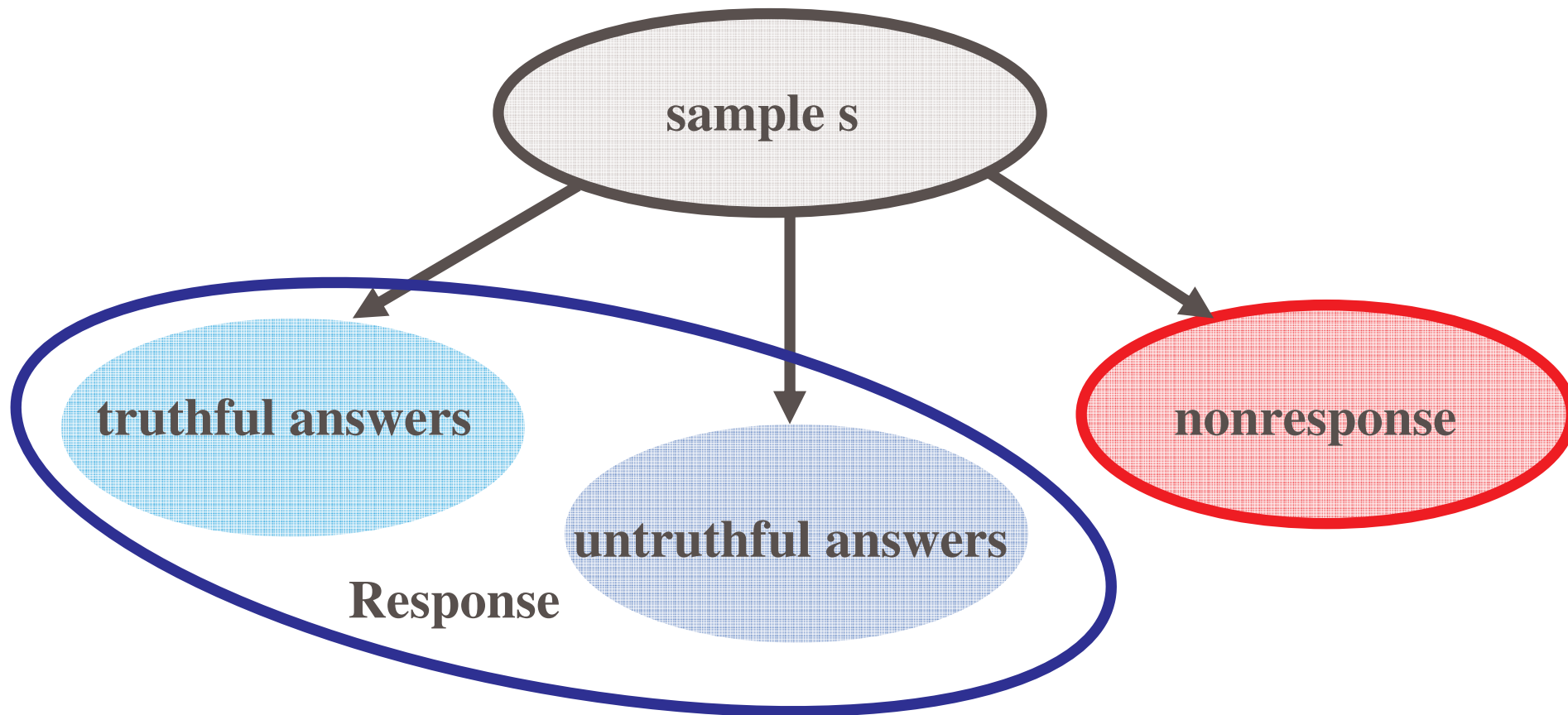
s ... a simple random sample of n elements from U (wr or wor)

Direct questioning (D) on the membership of U_A :

$$\hat{\pi}_A^D = \frac{\sum_{i \in \mathcal{S}} x_i}{n}.$$

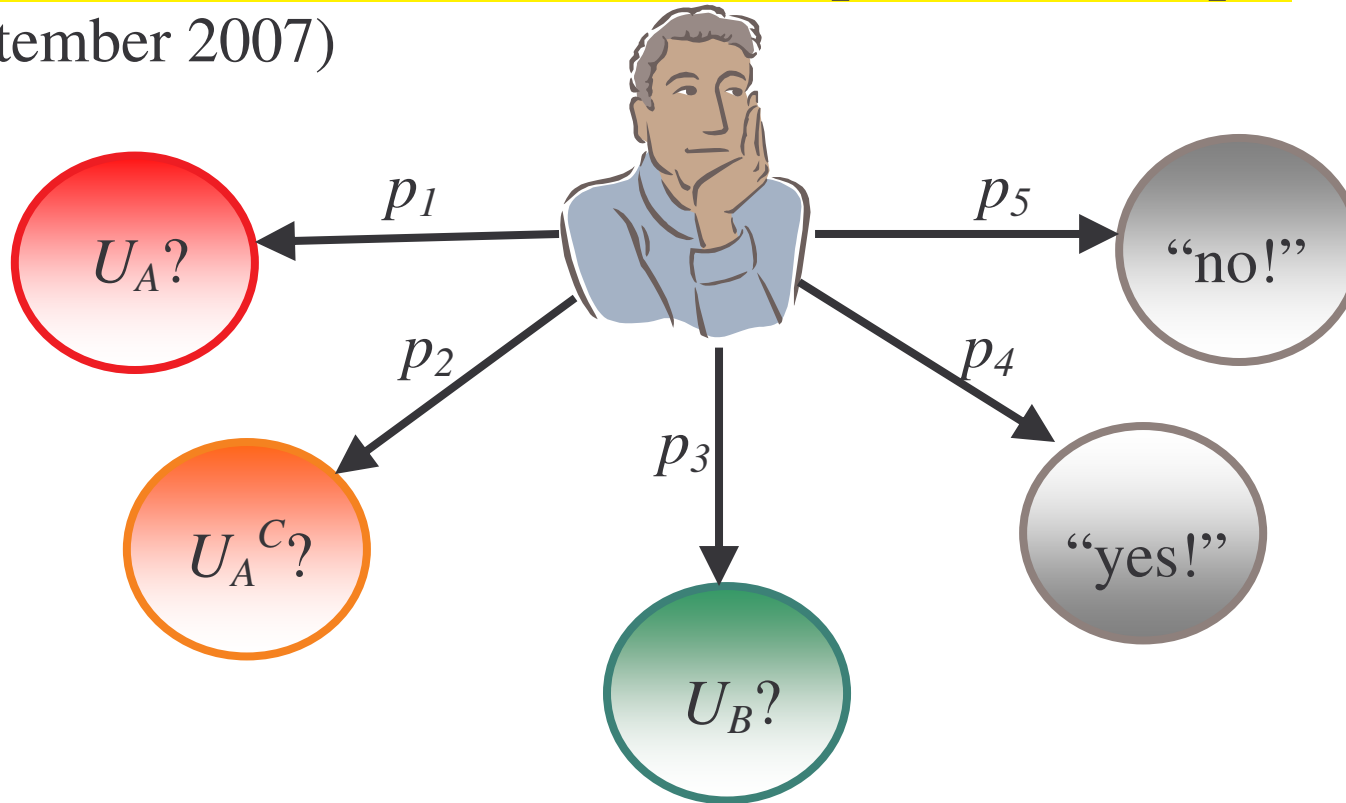


The Problem:



2. The Standardized Randomized Response Technique

(Quatember 2007)



($\sum_{i=1}^5 p_i = 1$, $0 \leq p_i < 1$ for $i = 1, 2, \dots, 5$). The relative size of group U_B in the population: $\pi_B = N_B/N$ (with $0 < \pi_B < 1$)

π_B and the probabilities p_1, p_2, \dots, p_5 are the *design parameters*

Let $y_k = \begin{cases} 1 & \dots \text{ if } k \text{ answers "yes"} \\ 0 & \dots \text{ otherwise.} \end{cases}$

With $a \equiv p_1 - p_2$ and $b \equiv p_2 + p_3 \cdot \pi_B + p_4$ the S -estimator for π_A ($a \neq 0$):

$$\hat{\pi}_A^S = \frac{\hat{\pi}_y - b}{a}.$$

With $c \equiv \frac{1}{n \cdot a} \cdot \left[\frac{b \cdot (1-b)}{a} + \pi_A \cdot (1-a-2b) \right] > 0$ its variances are given by:

$$V_{wor}(\hat{\pi}_A^S) = V_{wor}(\hat{\pi}_A^{dir}) + c$$

and

$$V_{wr}(\hat{\pi}_A^S) = V_{wr}(\hat{\pi}_A^{dir}) + c.$$

Questioning Design	Question on U_A^C	Question on U_B	Answer “yes”	Answer “no”
S1	✓			
S2		✓		
S3			✓	
S4				✓
S5	✓	✓		
S6	✓		✓	
S7	✓			✓
S8		✓	✓	
S9		✓		✓
S10			✓	✓
S11	✓	✓	✓	
S12	✓	✓		✓
S13	✓		✓	✓
S14		✓	✓	✓
S15	✓	✓	✓	✓

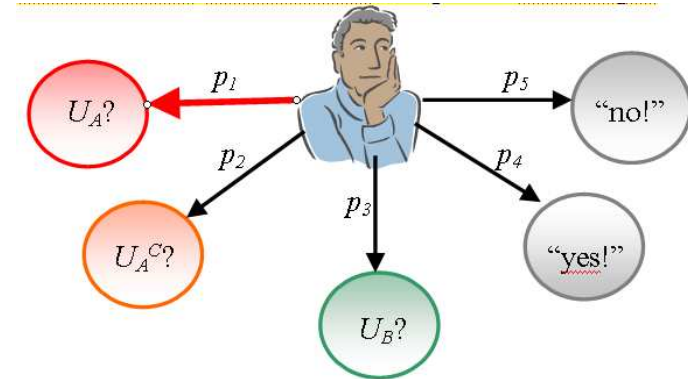


Table 1: The components of the special cases of technique S

Decreasing privacy protection → increasing rate of nonresponse and untruthful replies

3. Measures of Loss of Privacy

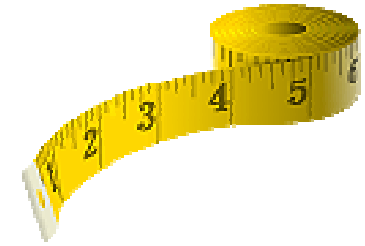
$$\lambda_i = \frac{\max(P(y_k = i|U_A), P(y_k = i|U_A^C))}{\min(P(y_k = i|U_A), P(y_k = i|U_A^C))}$$

$(1 \leq \lambda_i \leq \infty; i = 1, 0)$.

For questioning design S :

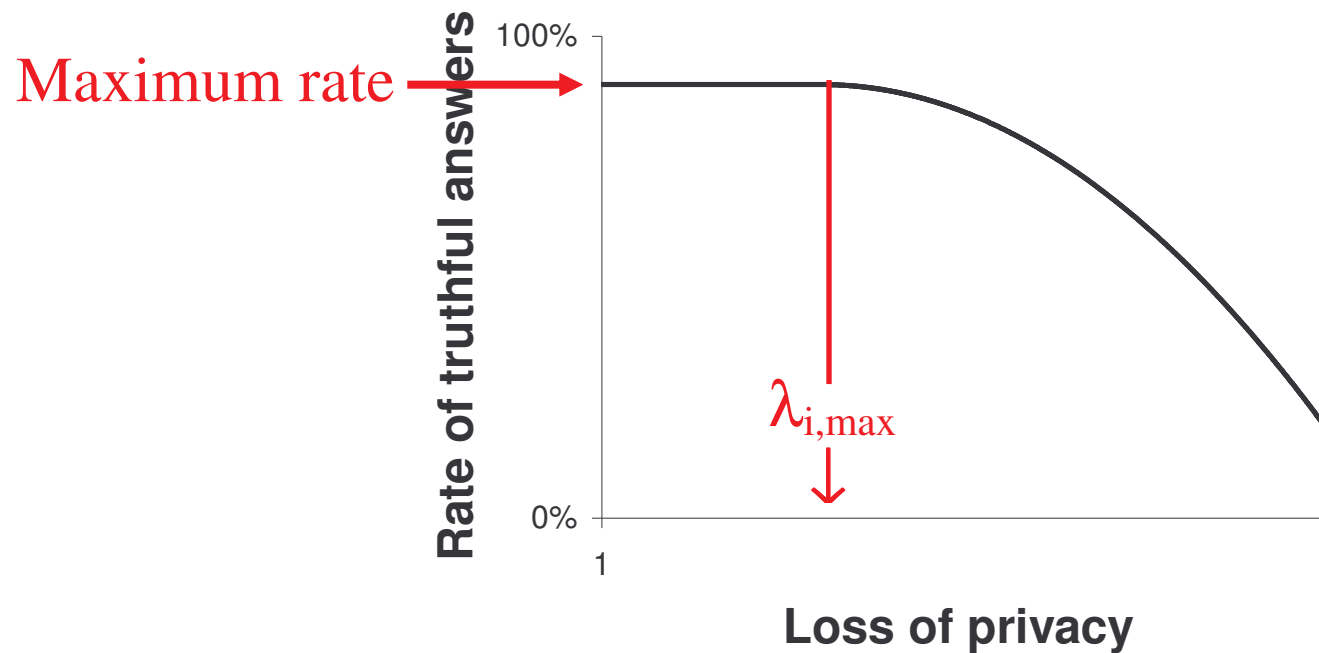
$$\lambda_1^S = \frac{\max(a+b, b)}{\min(a+b, b)} \quad \text{and} \quad \lambda_0^S = \frac{\max(1-a-b, 1-b)}{\min(1-a-b, 1-b)}$$

Designs with the same λ_1 and λ_0 are equally efficient



$\lambda_{1,\max}$, $\lambda_{0,\max}$: The maximum loss which just guarantees full, truthful response

Model of truthful answering rate with respect to answer “i”:

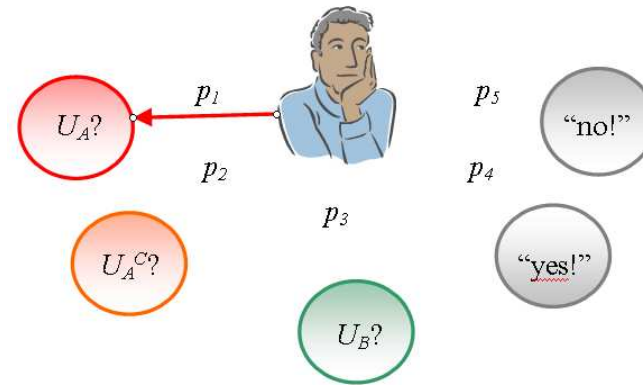


4. Optimum Questioning Designs for Different Variables

4.1 Variables with $\lambda_{1,max} = \lambda_{0,max} = \infty$

Examples: “gender”, “month of birth”

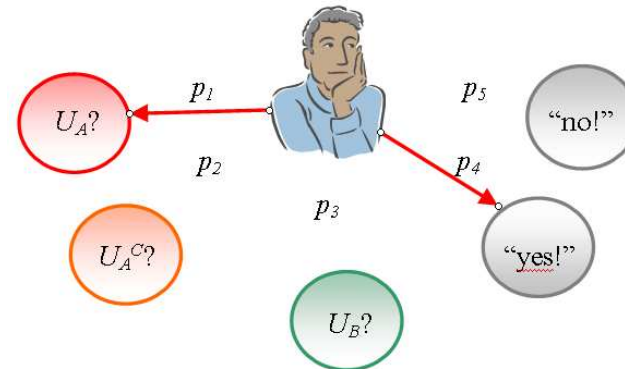
Best strategy: *D*



4.2 Variables with $\lambda_{1,max} < \lambda_{0,max} = \infty$

Examples: “theft at work”, “pretended sickness absence”

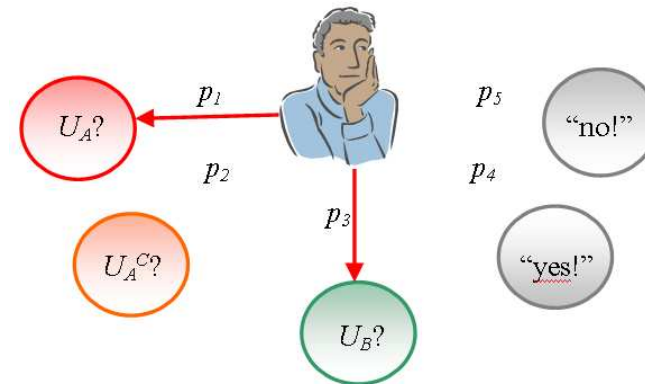
Best strategy: *S3*



4.3 Variables with $\lambda_{1,max} < \lambda_{0,max} < \infty$

Examples: “domestic violence”, “support of death penalty”

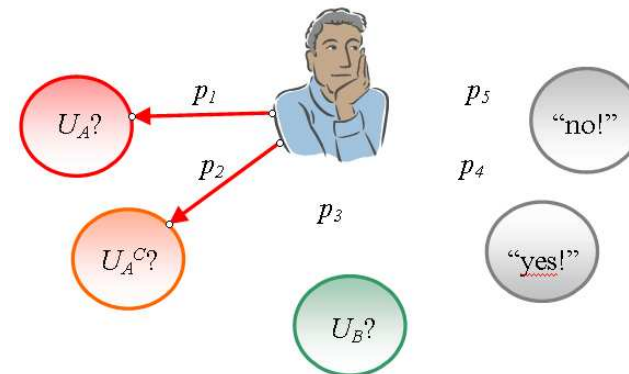
Best strategies: S2, S5, S6, S8-S15



4.4 Variables with $\lambda_{1,max} = \lambda_{0,max} < \infty$

Examples: “sexual behavior”, “tax fraud”

Best strategies: S1, S2, S5, S8-S15



References:

- Lensvelt-Mulders, G. J. L. M. et al. (2006) A validation of a computer-assisted randomized response survey to estimate the prevalence of fraud in social security, *Journal of the Royal Statistical Society, Series A*, 169(2), 305–318
- Quatember, A. (2007) A standardized technique of randomized response, *IFAS Research Paper Series*, 2008-28, www.ifas.jku.at/e2550/e2756/index (submitted)
- Quatember, A. (2008a) A Questioning Design to Identify the Proportion of Defrauders, *Communications in Dependability and Quality Management*, 11(4), 79-86
- Warner, S. L. (1965) Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias, *Journal of the American Statistical Association*, 60, 63–69

