UNRELATED QUESTION RANDOMIZED RESPONSE TECHNIQUES WITH TWO TRIALS PER RESPONDENT 1/

A. L. Gould and B. V. Shah, Research Triangle Institute
and
J. R. Abernathy, University of North Carolina

1. Introduction

Surveys aimed at obtaining information about the incidences of abortions in a population, the proportion of births occurring to unwed mothers, etc., usually are complicated by the fact that respondents are often reluctant to truthfully answer direct questions about such sensitive subjects. In an effort to circumvent this problem, and elicit a greater degree of cooperation from the respondents, researchers have developed randomized response techniques [1-4]. With such techniques, interviewed individuals are required to answer a randomly chosen question, with only the answer and not the question answered being known to the interviewer. However, even with randomized response procedures, respondents may refuse to cooperate. The purpose of the research reported here is to develop some models for the behavior of respondents when the unrelated question randomized response technique is used, and to apply these models to the results from surveys aimed at obtaining information about sensitive subjects; in particular we shall consider the results from a survey aimed at obtaining information about the proportion of households in North Carolina in which an illegitimate birth occurred [3].

Randomized response techniques originate from the work of Warner [4], who proposed that the sensitive question be presented to a respondent in a negative and a positive form, with the respondent randomly choosing one of the two forms and (truthfully) answering the chosen question without telling the interviewer which question was being answered. According to Warner, in view of the possibility of untruthful answers to direct questions, appreciable increases in efficiency are realizable with the randomized response technique.

The unrelated question randomized response variation of Warner's model was first reported by Horvitz, Shah, and Simmons [3], and its theoretical framework has been discussed by Greenberg, Abul-Ela, Simmons, and Horvitz [2]. The purpose of the variation was to increase the cooperation of respondents and the veracity of their responses beyond what the Warner technique might accomplish by, for example, removing the "heads I win, tails you lose" impression the structure of the Warner procedure might give. The unrelated question procedure poses two questions to the respondent, one dealing with the sensitive subject and the other being unrelated and innocuous; a randomizing device determines which question is to be answered, and the respondent (hopefully truthfully) answers the specified question without telling the interviewer which question is being answered. Horvitz, Shah and Simmons considered the situations where the respondent uses the device twice, giving two (assumed) independent answers, or once, giving one answer. They developed the formulas when one and two independent samples are interviewed, with the probability that the sensitive question will be presented to the respondent being different in the two samples, and applied their results to data from a study on the proportion of households in which an illegitimate birth occurred. The predicted proportions of households with illegitimate births differed from what would be expected, and they discussed some reasons for the discrepancies, observing that alternative models for the behavior of the randomizing devices and respondents might be appropriate. Such alternative models form the subject matter of this paper, and the data reported by Horvitz, Shah, and Simmons will be used subsequently for application of the various models to be discussed.

2. Preliminary Considerations

The unrelated question randomized response model assumes that the randomizing device presents the sensitive question to a respondent with the nominal probability, and that each respondent truthfully answers the presented question. In practice, these assumptions probably do not hold because of the idiosyncrasies of human beings and the imperfection of mechanical devices.

For example, respondents may not understand the procedure even after having it explained to them, and instead of answering the presented question, may randomly answer affirmatively or negatively. Alternatively, respondents in the sensitive group may, with a certain probability, answer the sensitive question untruthfully when it is presented, and this probability may depend (if each respondent takes two trials at using the device) on what had transpired previously. (That is, whether this was the first or second time the device was being used or the sensitive question was being presented, and if the second time, what had occurred on the first trial.) As still another possibility, a respondent in the sensitive group may refuse with some probability to truthfully answer any question requiring an affirmative answer, figuring perhaps that any positive answer might be stigmatizing. Or, possibly, this might be the attitude of respondents even in the non-sensitive group. Yet another possibility is that, when presented with the sensitive question, a respondent might decide to (possibly, but not necessarily, truthfully) answer the non-sensitive question.

Besides all the multitude of behavioral variations on the part of respondents, the randomizing devices may not be presenting the two questions with the nominal (assumed) probabilities. Horvitz, Shah, and Simmons discussed a number of possible explanations for this situation.

A general model from which particular behavioral models may be obtained by introducing sets of assumptions turns out to be a most practical vehicle for considering the

1/ Research supported by the National Institutes of Health, Contract No. 325-NIH-918A.

possible variations in behavior in a unified way. To specify such a model, some basic assumptions about the behavior of respondents are required; for the model to be considered, only four assumptions are required:

(1) For the two trial situation, a respondent who does (not) understand the procedure on the first trial will (not) understand the procedure on the second trial;

(2) Any respondent will always answer a question truthfully if the true answer is the negative (that is, the questions are defined so that a negative answer can never be stigmatizing);

(3) The randomizing device is such that the probability either question is selected at any trial does not depend upon which questions were selected at a previous trial (that is, separate trials with the randomizing device are independent and identically distributed).

(4) The innocuous question is unrelated to the sensitive question (i.e., the probability that an individual is in one or the other of the groups specified by the innocuous question does not depend upon the individual's possessing or not possessing the sensitive attribute, and vice versa).

To develop the probabilities of the various responses an individual may make, no other basic assumptions are required.

3. The General Model

We shall consider the situation in which each respondent uses the randomizing device twice. The form of the model which applies when there are two trials per respondent is given in Figure 1. The various probabilities appearing in Figure 1 are conditional upon events which precede them along the flow chart (which starts in the diamond labelled "comprehend?"). Aside from the parameter p which characterizes the behavior of randomizing device (and of which there are as many values as there are samples), the double trial model contains forty-two paramaters. Not all of these parameters are simultaneously estimable; indeed, given two samples in the two-trial case, at most six parameters can be estimated. However, by judiciously selecting the paramters to be estimated and the assumptions (fixed values, equality relations, etc.) to be applied to the remaining parameters, a wide variety of models for which the parameters can be estimated may be defined.

For an individual from whom two responses are elicited there are four possible responses: "yes, yes", "yes, no", "no, yes", "no, no". By following along the flow chart in Figure 1 the expressions for the various response probabilities may be obtained; these are given as Equations (1).

The parameter p appearing in these expressions denotes the probability with which the randomizing device presents the sensitive question to a respondent. If two samples from the population are drawn, different randomizing devices, with different values of p, will generally be used for the two samples. In what follows, p_1 will denote the value of p for the randomizing devices used in the first sample, and p_2 will denote the value of p for the randomizing devices used in the second sample. The parameter π_1 denotes the proportion of individuals in the population who possess the sensitive attribute, the parameter π_2 denotes the proportion possessing the innocuous attribute, and the parameter π_3 denotes the proportion who do not understand the procedure and answer at random, independently of the attributes they do or do not possess (with the ρ's denoting the probabilities of the various answers for

EQUATIONS (1)

$$\begin{aligned}
Pr\{yes, yes\} = \Big\{ & p^2\{\pi_1\xi_1\xi_2\sigma_1\sigma_2 + \pi_1\pi_2[\xi_1(1-\xi_2)\sigma_1\sigma_3 + (1-\xi_1)\xi_4\mu_1\mu_2 + (1-\xi_1)(1-\xi_4)\mu_1\mu_3]\} + p(1-p)\pi_1\pi_2\{\xi_1\sigma_1\sigma_4 \\
& + (1-\xi_1)\mu_1\mu_4 + \xi_7\nu_1\nu_2 + (1-\xi_7)\nu_1\nu_3\} + (1-p)^2\{\pi_1\pi_2\nu_1\nu_4 + (1-\pi_1)\pi_2\eta_1\eta_3\}\Big\}\pi_3 + \rho_1\rho_2(1-\pi_3)
\end{aligned}$$

$$\begin{aligned}
Pr\{yes, no\} = \Big\{ & p^2\{\pi_1\xi_1\xi_2\sigma_1(1-\sigma_2) + \pi_1\pi_2[\xi_1(1-\xi_2)\sigma_1(1-\sigma_3) + (1-\xi_1)\xi_4\mu_1(1-\mu_2) + (1-\xi_1)(1-\xi_4)\mu_1(1-\mu_3)] \\
& + \pi_1(1-\pi_2)\xi_1(1-\xi_2)\sigma_1\} + p(1-p)\{\pi_1\pi_2[\xi_1\sigma_1(1-\sigma_4) + (1-\xi_1)\mu_1(1-\mu_4) + \xi_7\nu_1(1-\nu_2) + (1-\xi_7)\nu_1(1-\nu_3)] \\
& + \pi_1(1-\pi_2)\xi_1\sigma_1 + (1-\pi_1)\pi_2\eta_1\} + (1-p)^2\{\pi_1\pi_2\nu_1(1-\nu_4) + (1-\pi_1)\pi_2\eta_1(1-\eta_3)\}\Big\}\pi_3 + \rho_1(1-\rho_2)(1-\pi_3)
\end{aligned}$$

$$\begin{aligned}
Pr\{no, yes\} = \Big\{ & p^2\{\pi_1(1-\sigma_1)\sigma_5\xi_1\xi_3 + \pi_1\pi_2[\xi_1(1-\xi_3)(1-\sigma_1)\sigma_6 + (1-\xi_1)\xi_5(1-\mu_1)\mu_5 + (1-\xi_1)(1-\xi_5)(1-\mu_1)\mu_6] \\
& + \pi_1(1-\pi_2)(1-\xi_1)\xi_6\mu_8\} + p(1-p)\{\pi_1\pi_2[\xi_1(1-\sigma_1)\sigma_7 + (1-\xi_1)(1-\mu_1)\mu_7 + \xi_8(1-\nu_1)\nu_5 + (1-\xi_8)(1-\nu_1)\nu_6] \\
& + \pi_1(1-\pi_2)\xi_9\nu_8 + (1-\pi_1)\pi_2\eta_2\} + (1-p)^2\{\pi_1\pi_2(1-\nu_1)\nu_7 + (1-\pi_1)\pi_2(1-\eta_1)\eta_4\}\Big\}\pi_3 + (1-\rho_1)\rho_3(1-\pi_3)
\end{aligned}$$

$$Pr\{no, no\} = 1 - Pr\{yes, yes\} - Pr\{yes, no\} - Pr\{no, yes\}$$

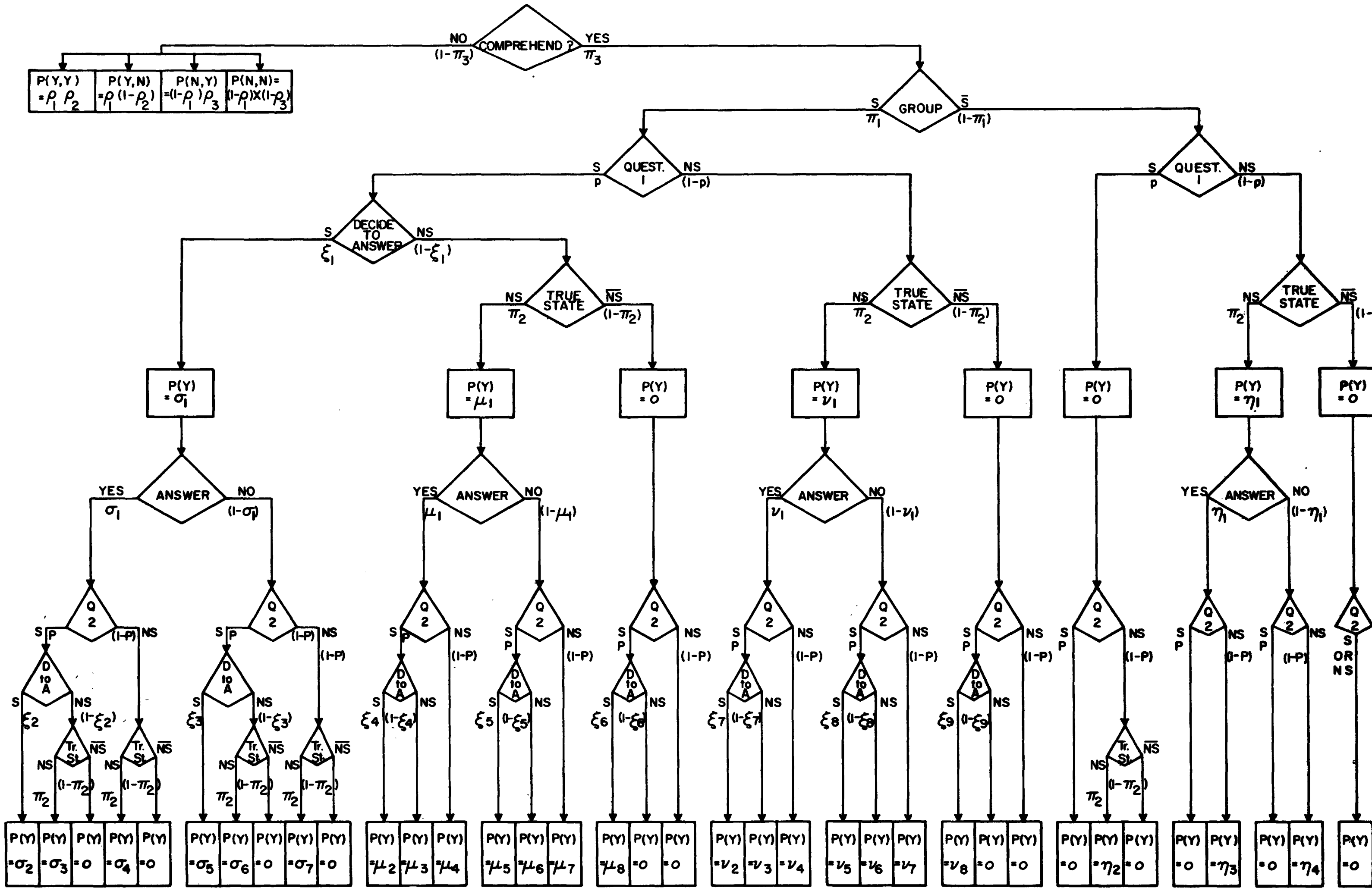

Figure 1: Flow Diagram of the Behavioral Model. (S means the sensitive question or group, NS means the nonsensitive question or group, and a bar over S or NS means the group in the population not possessing the attribute specified by the symbol.)

these individuals). The ξ parameters denote probabilities that an individual possessing the sensitive attribute will, when presented with the sensitive question, choose to consider instead the innocuous question. The remaining parameters denote the probabilities of telling the truth in various situations. If the responses of different individuals are independent, then the set of response probabilities (1) constitute the parameters of a multinomial distribution. Table 1 contains the probability functions, with the conbinational factors omitted, appropriate to the two-trial design for one and two samples.

The complicated functional forms of the response probabilities for the models render attempts to obtain explicit functional forms for estimators of the parameters impractical, if not impossible. Even for relatively simple behavioral models involving only a few of the parameters, explicit solutions for the parameters may not be unique. In order to make consideration of a variety of models convenient, the parameters were estimated by directly maximizing the likelihood function, using a direct search technique, with the parameters to be estimated and the various constraints specified in the input. Unfortunately, a direct maximization of the likelihood function does not lead to an expression of the covariance matrix of estimators as a by-product, as, for example, Gauss-Newton iteration does. Another approach to obtaining the covariance matrix of estimates is required; although results are not yet available, the technique of pseudoreplication will be used to obtain the estimates of the variances and covariances of the estimators. In view of the desirability of looking at the variation in values of the likelihood function [5], this technique has considerable merit. It proceeds by randomly allocating the units of samples into subsets, and estimating the parameters separately for each subset. The variances and covariances of the estimators obtained from the whole sample are then estimated using the corresponding quantities computed for the estimates obtained separately for each of the subsets.

4. Behavioral Models Obtained from the General Model.

The behavioral models to be applied to the illegitimacy data are described in Table 2. The general model is so formulated that, except for the p's, π's, ξ's, and ρ's, all the parameters are probabilities that a "yes" answer will be given when it is the truth; a "no" answer is assumed to be always given when it is the truth. In performing the calculations, all parameters which do not have specified fixed values, are being solved for, are not otherwise equal to other parameters, or do not satisfy other constraints, are assumed to have a value equal to unity.

TABLE 1

PROBABILITY FUNCTIONS OF THE POSSIBLE RESPONSES WHEN RESPONSES OF DIFFERENT INDIVIDUALS ARE ASSUMED INDEPENDENT; TWO TRIALS PER RESPONDENT.*)

Structure of Survey	Notation	Probability Function, Combinatorial Factors Omitted
Two trials per respondent, one sample	n_{11} = no. of "yes, yes" responses n_{12} = no. of "yes, no" responses n_{01} = no. of "no, yes" responses	$\alpha^{n_{11}} \beta^{n_{10}} \gamma^{n_{01}} (1-\alpha-\beta-\gamma)^{n-n_{11}-n_{10}-n_{01}}$
Two trials per respondent, two samples	n_{11} = no. of "yes, yes" responses in sample 1 n_{10} = no. of "yes, no" responses in sample 1 m_{11}, m_{10}, m_{01} are the corresponding quantities for sample 2	$\alpha_1^{n_{11}} \beta_1^{n_{10}} \gamma_1^{n_{01}}$ $\cdot\ (1-\alpha_1-\beta_1-\gamma_1)^{n-n_{11}-n_{10}-n_{01}}$ $\cdot\ \alpha_2^{m_{11}} \beta_2^{m_{10}} \gamma_2^{m_{01}}$ $\cdot\ (1-\alpha_2-\beta_2-\gamma_2)^{m-m_{11}-m_{10}-m_{01}}$

*) α = Pr{yes, yes}, β = Pr{yes, no}, γ = Pr{no, yes}; the subscripts of α, β, γ denote the sample to which the values of the probabilities apply.

TABLE 2

SOME BEHAVIORAL MODELS WHICH MAY BE OBTAINED FROM THE GENERAL MODEL

Model Number	Verbal Description	Parameters to be Estimated	Fixed Values Not Equal to Unity	Equalities	Other Constraints
1	All respondents comprehend, and answer the presented question truthfully	π_1, π_2	$p_1 = .7$ $p_2 = .3$	$\rho_2 = \rho_1$ $\rho_3 = \rho_1$	
2	Not all respondents comprehend; those who do answer the presented question truthfully; those who don't are equally likely to answer "yes" or "no"	π_1, π_2, π_3	$p_1 = .7$ $p_2 = .3$ $\rho_1 = .5$		
3	Not all respondents comprehend; those who do answer the presented question truthfully; those who don't answer "yes" with unknown probability	π_1, π_2, π_3 ρ_1	$p_1 = .7$ $p_2 = .3$	$\rho_2 = \rho_1$ $\rho_3 = \rho_1$	
4	Same as for Model 1	π_1, π_2, p_1			$p_2 = .42857\ p_1$
5	Same as for Model 1	π_1, π_2, p_1			$p_2 = p_1 - .4$
6	Same as for Model 1	π_1, π_2, p_1, p_2			
7	All respondents comprehend, respondents in the sensitive group might lie when presented with the sensitive question, and this doesn't depend upon any previous behavior; otherwise, the presented question is answered truthfully	π_1, π_2, σ_1	$p_1 = .7$ $p_2 = .3$	$\sigma_2 = \sigma_1$ $\sigma_5 = \sigma_1$ $\nu_2 = \sigma_1$ $\nu_8 = \sigma_1$	
8	All respondents comprehend; respondents in the sensitive group might lie (with unknown probability) when presented with the sensitive question the first time; if presented with the sensitive question a second time, they lie with probability 1 if they lied the first time, and with unknown probability (possibly different from the first) if they told the truth the first time; otherwise, the presented question is answered truthfully)	$\pi_1, \pi_2, \sigma_1, \sigma_2$	$\sigma_5 = 0$ $p_1 = .7$ $p_2 = .3$	$\nu_2 = \sigma_1$ $\nu_8 = \sigma_1$	
9	Same as for 8, except that respondents in the sensitive group may answer the sensitive question truthfully if it is presented a second time and they lied the first time it was presented	$\pi_1, \pi_2, \sigma_1,$ σ_2, σ_5	$p_1 = .7$ $p_2 = .3$	$\nu_2 = \sigma_1$ $\nu_8 = \sigma_1$	

Table 2 (cont'd)

Model Number	Verbal Description	Parameters to be Estimated	Fixed Values Not Equal to Unity	Equalities	Other Constraints
10	All respondents comprehend; when respondents in the sensitive group are presented with the sensitive question, they might lie; when respondents are required to answer the nonsensitive question affirmatively, they might lie, possibly with a different probability	π_1, π_2, σ_1, σ_4	$p_1 = .7$ $p_2 = .3$	$\sigma_2 = \sigma_1$ $\sigma_5 = \sigma_1$ $\nu_2 = \sigma_1$ $\nu_5 = \sigma_1$ $\nu_1 = \sigma_4$ $\nu_4 = \sigma_4$ $\nu_7 = \sigma_4$ $\eta_1 = \sigma_4$ $\eta_2 = \sigma_4$ $\eta_3 = \sigma_4$ $\eta_4 = \sigma_4$	
11	All respondents comprehend; any respondent faced with giving an affirmative answer to a presented question might lie	π_1, π_2, σ_1	$p_1 = .7$ $p_2 = .3$	$\sigma_2=\sigma_4=\sigma_5=\sigma_7$ $=\nu_1=\nu_2=\nu_4=\nu_5$ $=\nu_7=\eta_1=\eta_2=\eta_3$ $=\eta_4=\nu_8=\sigma_1$	
12	All respondents comprehend; respondents in the sensitive group, when presented with the sensitive question, might decide instead to truthfully answer the nonsensitive question; the nonsensitive question is always truthfully answered	π_1, π_2, ξ_1	$p_1 = .7$ $p_2 = .3$	$\xi_2 = \xi_3 = \xi_4$ $= \xi_5 = \xi_6$ $= \xi_7 = \xi_8$ $= \xi_9 = \xi_1$	

5. Application of the Models to Data from a Study on the Proportion of Households in North Carolina in which an Illegitimate Birth Occurred.

A summary of the data from the illegitimacy study reported by Horvitz, Shah, and Simmons [3], is presented in Table 3.

For all of the surveys, p_1 equals 0.7 and $p_2 = 0.3$. A summary of the results of the calculations is presented in Table 4. Table 5 contains a summary of the results of Horvitz, Shah, and Simmons.

TABLE 3

FREQUENCIES OF THE VARIOUS RESPONSES FROM THE ILLEGITIMACY STUDY. THE OBJECTIVE OF THE STUDY WAS TO DETERMINE THE PROPORTION OF HOUSEHOLDS IN NORTH CAROLINA REPORTING A BIRTH TO AN UNWED MOTHER. TWO RANDOMIZING DEVICES WERE USED. SEE [3] FOR DETAILS.

POPULATION SUBSET	Sample 1, p = .7				Sample 2, p = .3			
	n_{11}	n_{10}	n_{01}	n	m_{11}	m_{10}	m_{01}	m
White households (randomizing device is deck of cards)	137	271	253	1227	512	291	215	1340
Nonwhite households (same randomizing device)	29	52	45	223	124	54	61	298
White households (randomizing device is bead box)	37	55	61	320	141	67	48	375
Nonwhite households (same randomizing device)	16	21	22	117	25	13	5	67

TABLE 4.

SUMMARY OF CALCULATIONS FOR THE ILLEGITIMACY DATA FROM [3]*

Model Number (See TABLE 3)

		1	2	3	4	5	6	7	8	9	10	11	12
No. Parameters		2	3	4	3	3	4	3	4	5	4	3	3
White HH. Card Deck	Log Likelihood	-3368.42	-3346.40	-3345.83	-3357.82	-3367.56	-3344.47	-3352.64	-3362.34	-3351.51	-3348.65	-3364.36	-3368.42
	Chi-Square	64.88	19.28	17.92	43.96	63.23	15.63	33.82	52.92	31.65	24.17	56.51	64.88
	d.f. (χ^2)	4	3	2	3	3	2	3	2	1	2	3	3
	$\hat{\pi}_1$	.02824	.004041	.01872	.00211	.02005	.002749	.9482	.08408	.2181	.3931	.05869	.02865
	$\hat{\pi}_2$	.8616	.8514	.8925	.8391	.8531	.8637	.8134	.8427	.8074	.8296	.8854	.8618
Non-White HH. Card Deck	Log Likelihood	-685.84	-682.02	-683.01	-683.94	-685.76	-681.53	-685.72	-684.87	-682.88	-681.55	-685.28	-685.84
	Chi-Square	10.00	2.23	2.20	6.16	9.79	1.24	9.76	8.00	4.00	1.30	8.97	10.00
	d.f. (χ^2)	4	3	2	3	3	2	3	2	1	2	3	3
	$\hat{\pi}_1$	.04299	.02163	.01989	.01995	.02053	.01057	.04805	.09615	.2322	.5264	.05727	.04299
	$\hat{\pi}_2$	.8981	.9073	.9078	.8781	.8877	.8979	.8958	.8837	.8540	.8553	.9088	.8978
White HH. Bead Box	Log Likelihood	-879.56	-878.16	-877.90	-877.61	-879.29	-876.88	-879.46	-879.52	-879.45	-879.51	-879.54	-879.55
	Chi-Square	9.33	6.06	5.66	5.48	8.89	3.81	9.12	9.25	9.11	9.14	9.20	9.30
	d.f. (χ^2)	4	3	2	3	3	2	3	2	1	2	3	3
	$\hat{\pi}_1$	.05434	.01999	.02052	.02024	.02024	.02038	.05830	.05767	.05943	.06752	.06533	.05446
	$\hat{\pi}_2$	.7646	.7478	.7326	.7446	.7518	.7502	.7631	.7631	.7624	.7688	.7702	.7646
Non-White HH. Bead Box	Log Likelihood	-232.76	-231.76	-231.76	-231.09	-231.99	-230.93	-230.80	-232.07	-230.79	-231.31	-232.77	-232.75
	Chi-Square	7.28	4.78	4.79	4.32	6.22	3.70	3.55	6.25	3.54	4.31	7.26	7.27
	d.f. (χ^2)	4	3	2	3	3	2	3	2	1	2	3	3
	$\hat{\pi}_1$	.08512	.06436	.02098	.004756	.04941	.004050	.5374	.1594	.3059	.3305	.08888	.08596
	$\hat{\pi}_2$	.7525	.7730	.6758	.7048	.7188	.7079	.6464	.7120	.6438	.6639	.7532	.7517

* The sensitive attribute is occurrence of an illigitimate birth in the household during the past year; the nonsensitive attribute is having been born in North Carolina.

TABLE 5

SUMMARY OF RESULTS OF HORVITZ, SHAH AND SIMMONS

Population	Expected $\hat{\pi}_1$	Model 1 Estimates $\hat{\pi}_1$	Model 1 Estimates $\hat{\pi}_2$	Model 4 Estimate $\hat{\pi}_1$
White HH. Card Deck	.002	.141	.755	-.0003
Nonwhite HH. Card Deck	.034	.151	.805	.027
White HH. Bead Box	.002	.122	.704	-.0007
Nonwhite HH. Bead Box	.034	.180	.648	.020

6. Discussion and Remarks

The estimates obtained by Horvitz, Shah and Simmons differ from those obtained through the use of the model because the Horvitz, Shah and Simmons estimators are moment estimators, and those obtained for the model are obtained by a search for a set of values maximizing the likelihood function. In discussing the large deviation of the estimate of π_1 from its expected value, Horvitz, Shah and Simmons considered a number of explanations relevant to the modeled behavior, in particular the possibility that the realized p values were different from the nominal ones, and that there may have been some confusion on the part of the respondents. They considered also the possibility that misreading of the sensitive question, or the way the procedure was described to the respondents, may have led to an excess of positive answers. The possibility in the general model which comes closest to describing this source of error is model 3, although the situation is not exactly that of confusion on the part of the respondents. Deliberate untruthfulness on the part of the respondents does not appear to be a relevant explanation in the present case. From the results of fitting the models, it appears that models 3 and 6 do the best job in describing the data. The estimated parameter values and predicted frequencies for these models are displayed in Table 6. We note that the predicted proportions of households with an illegitimate birth are considerably closer to what would be expected than was the case for Horvitz, Shah

TABLE 6

PARAMETER ESTIMATES AND PREDICTED

FREQUENCIES FOR MODELS 3 AND 6

Population	MODEL 3 $\hat{\pi}_1$	MODEL 3 $\hat{\pi}_2$	MODEL 3 $\hat{\pi}_3$	MODEL 3 $\hat{\rho}_1$	MODEL 6 $\hat{\pi}_1$	MODEL 6 $\hat{\pi}_2$	MODEL 6 $\hat{p}_1$	MODEL 6 $\hat{p}_2$
White HH. Card Deck	.01872	.8925	.7179	.4102	.002749	.8637	.6270	.3373
Nonwhite HH. Card Deck	.01989	.9078	.7462	.5000	.01057	.8979	.6208	.3241
White HH. Bead Box	.02052	.7326	.8448	.5780	.02038	.7502	.6222	.2995
Nonwhite HH. Bead Box	.02098	.6758	.7509	.5898	.004050	.7090	.5548	.2719

Population	Sample	Observed YY	Observed YN	Observed NY	Observed NN	Model 3 YY	Model 3 YN	Model 3 NY	Model 3 NN	Model 6 YY	Model 6 YN	Model 6 NY	Model 6 NN
White HH. Card Deck	1	137	271	253	566	143	246	246	592	150	247	247	582
	2	512	291	215	322	493	269	269	310	510	258	258	314
Nonwhite HH. Card Deck	1	29	52	45	97	31	45	45	102	31	47	47	99
	2	124	54	61	59	120	61	61	57	124	58	58	58
White HH. Bead Box	1	37	55	61	167	39	53	53	175	39	56	56	170
	2	141	67	48	119	136	62	62	117	141	58	58	117
Nonwhite HH. Bead Box	1	16	21	22	58	17	19	19	61	17	20	20	59
	2	25	13	5	25	23	11	11	22	25	9	9	23

and Simmons' results; this may be due to the fact that the present estimators are maximum likelihood, rather then moment estimators. The relatively close agreement of the predicted frequencies with the observed frequencies does not, of course, mean that the models are adequate representations of reality: models 3 and 6 describe different types of behavior. However, the models are relatively consistent insofar as their predictions of the values of π_1 are concerned. At least for the population considered, the percentage of households in which an illegitimate birth occurred appears to be approximately 2%, for both white and non-white households. However, the precisions of the estimates remain to be determined.

There is certainly a need for considerable empirical experience in applying the models to surveys employing the unrelated question randomized response technique. Such experience is necessary for determining which of the particular models obtainable from the general model will be useful in various situations, and which of the models will prove to be of limited or negligible utility. Moreover, empirical experience is also necessary to determine if there are behavioral models not obtainable from the general model which are useful for describing the behavior of respondents.

Acknowledgement

It is a pleasure to acknowledge the illuminating comments and discussion of Dr. D. G. Horvitz in bringing the problem to our attention, and during the course of carrying out the research.

7. References

[1] Abul-Ela, A. A., Greenberg, B. G., and Horvitz, D. G. (1967). A multi-proportions randomized response model. Jour. Am. Stat. Assoc. 62, 999-1008.

[2] Greenberg, B. G., Abul-Ela, A. A., Simmons, W. R., and Horvitz, D. G. (1969). The unrelated question randomized response model: theoretical framework. Jour. Am. Stat. Assoc. 64, 520-539.

[3] Horvitz, D. G., Shah, B. V., and Simmons, W. R. (1967). The unrelated question randomized response model. Social Statistics Section of the American Statistical Association, 65-72.

[4] Warner, S. L. (1965). Randomized response: a survey technique for eliminating evasive answer bias. Jour. Am. Stat. Assoc. 60, 63-69.

[5] Sprott, D. A., and Kalbfleisch, J. D. (1969). Examples of likelihoods and comparison with point estimates and large sample approximations. Jour. Am. Stat. Assoc. 64, 468-484.