SELECTING OPTIMAL VALUES FOR Π_Y IN THE UNRELATED QUESTION RANDOMIZED RESPONSE MODEL

Lisa A. Kammerman, Westat, Inc.;
Bernard G. Greenberg, Dana Quade, University of North Carolina

1. Introduction

In surveys containing direct questions on sensitive issues, individuals may refuse to participate or they may choose to give incorrect answers, resulting in a biased estimate of the population proportion of individuals who possess a sensitive characteristic. In order to elicit higher response rates and better cooperation among participants, Warner (1965) proposed the use of a randomized response design, which gives participants some protection from being perceived as possessing a sensitive attribute. The single unrelated question randomized response model, a variation of Warner's model, constituted the next major development of randomized response models (Abul-Ela 1966; Horvitz, Shah, and Simmons 1967; and Greenberg et al. 1969). Like Warner's model, this model provides the respondent a choice of two questions, one of which concerns the sensitive characteristic; however, the other concerns an innocuous or unrelated characteristic. This model, which assumes that respondents are more likely to cooperate if they are given a chance to select a question that concerns a non-sensitive characteristic, is described as follows.

Assume that the population is divided into two groups, which may intersect:

Group A: members with the sensitive characteristic (A) and

Group Y: members with the unrelated characteristic (Y).

Let Π_A = proportion of population with A

and Π_Y = proportion of population with Y.

If Π_Y is known, then only one simple random sample is required to obtain the estimate of Π_A. Assuming full cooperation among respondents, let

p = the probability of selecting the sensitive question,

$1-p$ = the probability of selecting the unrelated question,

λ = the population proportion of "Yes" responses

$$= p(\Pi_A-\Pi_Y)+\Pi_Y \;, \qquad (1.1)$$

$\hat{\lambda}$ = the observed proportion of "Yes" responses, and

n = size of sample.

The true proportion of the individuals with the sensitive characteristic is designated in the single unrelated question model (Π_Y known) as $(\Pi_A|\Pi_Y)_{U1}$:

$$(\Pi_A|\Pi_Y)_{U1} = \frac{\lambda - \Pi_Y(1-p)}{p} \;. \qquad (1.2)$$

Then, the single unrelated question (Π_Y known) estimator for Π_A, $(\hat{\Pi}_A|\Pi_Y)_{U1}$, is obtained by substituting $\hat{\lambda}$ for λ.

The optimal values for p and Π_Y, which are preselected, are not obvious. For example, increasing the value of p decreases the variance of the estimate but increases the likelihood that a respondent will be perceived as possessing the sensitive attribute when giving a "Yes" answer; this situation could result in nonresponse or nontruthful answers, thus increasing the bias of the estimate. Similarly, decreasing the value of Π_Y results not only in a smaller variance but also in an increase in the bias of the estimate.

These examples illustrate the central question to selecting the parameters: can a proper selection of parameters, which actually determines the degree of protection, increase response rates and truthful responses while also yielding reasonable estimates as measured by bias and variance? Only a few studies have examined the relationship between respondent risks or jeopardies and the parameters in the randomized response models. Among them, Lanke (1975) explored the relationship between p and Π_Y through $P(A|\text{Yes})$, the conditional probability that the respondent belongs to the A group given that his/her response to the question selected by the randomizing response device is "Yes." Lanke argued that this "risk of suspicion" should be bounded by some constant θ: $P(A|\text{Yes}) \le \theta$.

The rate of cooperation decreases beyond θ. In fact, Lanke postulated that potential respondents may use $P(A|\text{Yes})$ as a basis for deciding on whether or not to participate or give a truthful response. Our model incorporates these concepts.

Greenberg et al. (1977) examined the risk to A respondents who give a "Yes" response and to those respondents without A, whom we term "a" respondents. The limited hazard for A individuals was defined as

H_A^* = P(A is perceived as A when answering Yes)

$$= \quad P(Yes|A)P(A|Yes)$$

$$= \quad [p+(1-p)\Pi_Y]P(A|Yes) \qquad (1.3)$$

A similar definition holds for the a individuals. A reduction in limited hazard for the A respondents necessitates an increase in limited hazard among the a respondents, which may lead to reduced cooperation among the a respondents. The authors recommended relating the bias $P(A|Yes)$ in order to show the influence of p, Π_Y, and Π_A on the bias, and that the best strategy for choosing p and Π_Y would involve minimizing the mean squared error (MSE) of Π_A.

We will use this approach to produce recommended values for Π_Y in the unrelated question randomized response model (Π_Y known).

2. Sensitivity of Question

2.1 Introduction

Two of the few studies that have attempted to measure the respondent's perceived level of protection or to assess the sensitivity of questions, provide the basis for much of the research presented in this manuscript. Soeken and Macready (1982), in a study designed to assess a respondent's perceived level of protection and willingness to cooperate, suggested that perceived protection is related to the sensitivity of the topic under study. They recommended further research of this relationship.

Along those lines, Himmelfarb and Lickteig (1982) proposed a social desirability scale which measured the sensitivity of a direct question and possible direction of bias by comparing responses to a direct question with those to a randomized response question.

In a separate study, Greenberg et al. (1977) postulated a mathematical relationship between $P(A|Yes)$ and the probability of answering truthfully among the A individuals.

In combining and extending these three studies, we assume that the relationship between the probability of a truthful response (T) and hazard is question-specific or sensitivity-specific. Thus, through the use of a measurement instrument such as Himmelfarb and Lickteig's, we suggest rating and categorizing the sensitivity (S) of a topic on an ordinal scale of 0 to 10, where 10 denotes the most sensitive and threatening topic.

For each of the eleven categories, we will use an ogive to generate a curve that depicts a relationship between T and hazard as postulated by Greenberg et al. (1977). Each curve would correspond to a different threshold value (θ), such that the topics with S=10 would have a small θ, and those with S=0 would have an extremely high θ. For example, a large S would correspond to a small θ, and T would remain close to unity over only a small range of hazard. Tables of optimal values of $\Pi_Y(\Pi_{Y_{opt}})$ will be determined for each S by minimizing the MSE for fixed p and presumed Π_A.

2.2 A Family of Relationships Between and Hazards

In order to generate the family of eleven curves, the beta survival function was used since $0 \le x,y \le 1$ and the shape is in the form of an ogive.

Let X denote a random variate such that $0 \le x \le 1$. Let $\alpha > 0$ and $\beta > 0$ denote shape parameters. Then, X follows a beta distribution with the following cumulative distribution, also known as the incomplete beta function:

$$F(x:\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\int_0^x t^{\alpha-1}(1-t)^{\beta-1}\,dt \qquad (2.1)$$

where $F(x:\alpha,\beta)$ is increasing in x over $0 \le x \le 1$ and $B(\alpha,\beta)$ represents the beta function, with parameters α and β:

$$B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt. \qquad (2.2)$$

The beta survival function is simply

$$S(x:\alpha,\beta) = 1 - F(x:\alpha,\beta) \qquad (2.3)$$

where $S(x:\alpha,\beta)$ is decreasing in x over $0 \le x \le 1$, and $S(0:\alpha,\beta)=1$ and $S(1:\alpha,\beta)=0$ for all α and β. If we set $S(x:\alpha,\beta)=T$ for a given limited hazard denoted by X and for a given α and β, then we can generate a family of curves that relates T with the limited hazards by varying α and β as follows.

Define S = sensitivity of the threatening question, such that $S=0,1,\ldots,9,10$ as defined in 2.1. Then, letting $\beta = 2S+1$ and $\alpha = 22-\beta=21-(2S+1)=21-2S$ gives $(\alpha,\beta) = (21,1), (19,3), \ldots, (3,19), (1,21)$ for $S = 0,1, \ldots, 9,10$. Expression (2.3) can be rewritten as

$$S(x:21-2S,2S+1) = 1 - F(x:21-2S,2S+1) \qquad (2.4)$$

The use of $0 \le x \le 1$ for each S generated the family of curves depicted in Figure 1.

3. An Unrelated Question Randomized Response Model that Allows Nontruthful Reporting Among A & a Individuals

We now develop a general model that estimates Π_A in the presence of nontruthful reporting among both the A and a respondents. This model assumes that individuals possess their own tolerance of suspicion, namely θ. If the individual selects a question requiring a "No" response, then a truthful answer is given. However, if the question requires a "Yes" response, the individual calculates or estimates in some way the probability that a person who answers "Yes" will be perceived as having the stigmatizing characteristic. Suppose the respondent arrives at the value Q; then a "Yes" response is given if Q is less than the tolerance for suspicion. The tolerance θ has a probability distribution (tolerance distribution) over the population. The shape of the tolerance distribution depends on how stigmatizing the characteristic is; the more stigmatizing, the lower the average tolerance. The distribution may also differ for people who have the sensitive characteristic from those who do not. However, we assume that an individual's tolerance is fixed before the question is selected, and is independent of whether the individual has the innocuous characteristic or not.

3.1 An Estimator for Π_A and its Properties

The mathematical formulation for the model is the following. For any individual, let

θ = tolerance of suspicion, and

Q = the calculated value that a person who answers "Yes" will be perceived as possessing the stigmatizing characteristic.

Note that θ is a random variable and Q is a value fixed in advance of answering "Yes". An individual who is confronted with a question requiring a "Yes" response uses the following rule to decide whether to give a truthful answer:

1. If $Q \leq \theta$, give a true answer, i.e., "Yes"
2. If $Q > \theta$, give a false answer, i.e., "No".

Let $F_1(Q)$ be the tolerance distribution among persons who have the stigmatizing characteristic and $F_0(Q)$ among those who do not have it, where $F(Q) \equiv \Pr(\theta \leq Q)$ and $1-F(Q) \equiv \Pr(\theta > Q)$.

Among the population of A individuals, the proportion $1-F_1(Q)$ would give truthful answers to questions requiring a "Yes" response and among the a population, the proportion $1-F_0(Q)$ would give a truthful response. If we assume that the tolerance distribution for the A individuals is the same as that for the a individuals, then $F_1 = F_0 = F$. Using the above notation, the unrelated question model with Π_Y known yields the following estimator of Π_A. Let

λ' = expected proportion of "Yes" responses, and

$\hat{\lambda}'$ = observed proportion of "Yes" responses.

Then $n\hat{\lambda}'$ is distributed as Binomial(n,λ') and $E(\hat{\lambda}')=\lambda'$. Then,

$$\lambda' = (1-F(Q))(p\Pi_A + (1-p)\Pi_Y) \qquad (3.1)$$

Define the following expression:

$$(\hat{\Pi}_A|\Pi_Y)'_{U1} = \frac{\hat{\lambda}' - \Pi_Y(1-p)}{p} \qquad (3.2)$$

which we will use as an estimator of Π_A, because its expectation is $(\Pi_A|\Pi_Y)_{U1}$ which equals Π_A under 100 percent truthful reporting. It can be shown that the bias is then:

$$\text{Bias}(\hat{\Pi}_A|\Pi_Y)'_{U1} = \frac{\lambda' - \lambda}{p} \quad . \qquad (3.3)$$

Then, by substitution of (1.1) and (3.1), (3.3) becomes

$$\text{Bias}(\hat{\Pi}_A|\Pi_Y)'_{U1} = [-F(Q)][\Pi_A + \frac{(1-p)}{p}\Pi_Y] \qquad (3.4)$$

Because $n\hat{\lambda}'$ is distributed as Binomial(n,λ'), it can be shown that the variance of the estimate is

$$\text{Var}(\hat{\Pi}_A|\Pi_Y)'_{U1} = \frac{[(1-F(Q))(p\Pi_A+(1-p)\Pi_Y)]}{np^2} \times [1 - ((1-F(Q))(p\Pi_A+(1-p)\Pi_Y)] \qquad (3.5)$$

and the mean squared error is

$$\text{MSE}(\hat{\Pi}_A|\Pi_Y)'_{U1} = \{\text{Bias}(\hat{\Pi}_A|\Pi_Y)'_{U1}\}^2 + \text{Var}(\hat{\Pi}_A|\Pi_Y)'_{U1} \qquad (3.6)$$

3.2 Defining Q and F(Q)

In order to implement the model, an approximation for Q must be defined. Two very different approaches to arriving at Q might exist. A sophisticated respondent might conceivably calculate $Q = P(A|\text{Yes})$, by estimating Π_A, Π_Y, and p and then applying probability theory. However, most respondents will follow cruder methods for which setting $Q = P(A|\text{Yes})$ might be a reasonable approximation, the approach used here. Further, $Q = P(A|\text{Yes})$ will be calculated assuming that some respondents might give non-truthful responses, since some respondents will realize that other respondents will have incentives to be untruthful also.

So, under the assumption that $F_1 = F_0 = F$,

$$Q = P(A|\text{Yes}) = \frac{P(A \cap \text{Yes})}{P(\text{Yes})}$$

$$= \frac{\Pi_A p + \Pi_A \Pi_Y (1-p)}{\Pi_A p + \Pi_Y (1-p)} \qquad (3.7)$$

The eleven curves depicting T as a function of limited hazard (Figure 1) will be used as the tolerance distributions represented by F(Q); they are cumulated from the right instead of the left. These distributions are described by the relationship $F(Q) = S(Q:\alpha,\beta)$.

3.3 Choosing Π_{Yopt}

Selecting the Π_Y which minimizes $MSE(\hat{\Pi}_A|\Pi_Y)'_{U1}$ is equivalent to minimizing (3.6) with respect to Π_Y for fixed p. To obtain a general result for Π_{Yopt} requires differentiating (3.6), which is very difficult. Thus, Π_{Yopt} were obtained through the use of numerical techniques.

First, we specified the values of p, Π_A, and n for which to obtain the Π_{Yopt}. We chose three levels of p (1/2, 2/3, and 3/4), five presumed values of Π_A(.01, .02, .05, .10, and .25) and four values of n (100, 500, 1,000, and 10,000).

Further, a unique Π_Y is determined for each Q, when n, p and Π_A are known, by solving for Π_Y in (3.7). Thus, for fixed p and Π_A, we need only to specify many values of Q to obtain the corresponding Π_Y. The values of F(Q) were generated by specifying Q and then using the eleven curves depicted in Figure 1.

4. Results from the Investigations of the Optimal Π_Y

4.1 Investigation of Π_{Yopt}

We note that topics of high sensitivity yield a minimum MSE, with respect to y, for which $|Bias\ (\hat{\Pi}_A|\Pi_Y)_{U1'}| \geq \Pi_A$; see Table 1 for an example. (Note that we present here tables and figures containing results for n=1000 only.) Further, when n=100 and Π_A=.01, it is always true that $|Bias(\hat{\Pi}_A|\Pi_Y)_{U1'}| > \Pi_A$; and when Π_A=.02, only the least sensitive categories contain any recommended values for Π_{Yopt}. For these topics more protection, resulting in a reduction in bias, might be given to the respondent either by reducing p to a value less than one-half or by increasing Π_Y to one. Indeed, inspection of the bias when Π_Y=1 showed that, in most instances, the $|Bias(\hat{\Pi}_A|\Pi_Y)_{U1'}| < \Pi_A$ and that the variance is only slightly higher than the variance corresponding to the minimum MSE. However, reducing p to 1/4 or 1/3 resulted in Π_{Yopt}=0 for many of the highly sensitive topics. Secondly, Greenberg et al. (1969) recommended that Π_Y should be selected in the neighborhood of the presumed or postulated Π_A. The results here for topics of modest sensitivity ($1 \leq S \leq 3$) and where Π_A ranges from .02 to .10 show that the approximation was a reasonable one.

For topics of no or little sensitivity (S=0,1) and for which Π_A=.02 or .05, these results suggest Π_{Yopt} is close to zero. When Π_{Yopt}=0 (a direct question), a "No" response is always required when the unrelated question is selected; thus, a "Yes" response to a randomized response question would identify the respondent as possessing the sensitive attribute. Since a randomized response model contributes an additional cost in variance over a direct question survey, the use of Π_{Yopt}=0 for mildly sensitive topics may more than offset an increase in the squared bias from a direct question with a greater decrease in the variance.

These results also show that the Π_{Yopt} in adjacent columns tend to be very similar. Also, the Π_{Yopt} tend to increase as the sensitivity increases. This is again what one would expect -- higher sensitivity requires greater protection for the respondent. When Π_A is fairly common (Π_A=.25), a higher proportion of the optimal values are Π_{Yopt}=1 unless p=1/2. Increasing the sample size tends to increase Π_{Yopt}. Concomitantly, increasing the sample size tends to decrease the variance while the bias remains independent of sample size. Recall that, in general, the bias tends to decrease as more protection is given to the respondent through an increase in the choice of Π_Y. Thus, for larger n, the contribution of the variance to the MSE becomes smaller, and we can afford to increase Π_Y. Also, as n increases, the values of S for which we can obtain Π_{Yopt} also increase. And when n=10,000, S can go as high as 9 when Π_A=.01 or .02.

In order to graphically depict some of these findings and to suggest a possible aid for selecting Π_{Yopt} for a sample design, we have plotted Sensitivity vs. Π_{Yopt} for p=1/2, 2/3, 3/4, Π_A=.05, and n=1000 (Figure 2).

5. Discussion

This work has introduced the notion of categorizing social topics into eleven categories of sensitivity, and has identified a possible measure of social desirability. While several authors have suggested that the sensitivity of a social issue affects the level of cooperation among respondents, no other study to our knowledge has assumed that a feasible instrument exists for determining the sensitivity of topics. However, more research is required to fully develop the suggested measure or to formulate a different one. Ideally, a proposed measure would be tested on a well-defined sample of individuals for which randomized response techniques are of potential usefulness.

The proposed model is based on the assumption that each individual possesses his/her own tolerance of suspicion, θ, which is assumed fixed before the question is asked. The results here reflect the assumption that A and a individuals possess the same tolerance distributions, and did not take into account that the tolerance distribution for an A individual may differ from that for an a individual.

For topics of high sensitivity, no $\Pi_{Y\text{opt}}$ were recommended because the magnitude of the bias for $\Pi_{Y\text{opt}}$ obtained through the use of MSE criteria was equal to or greater than the estimated value of Π_A. In most of these cases a value of $\Pi_A=1.0$ yielded an acceptable level of bias.

Topics of slight sensitivity (S=0,1) call for $\Pi_{Y\text{opt}}=0$, which is equivalent to the direct question setting in which a "Yes" response identifies the participant as possessing the sensitive attribute. For topics categorized as mildly sensitive it is possible that certain individuals may be willing, without protection, to give truthful responses for a topic that they may feel is not sensitive for themselves. This latter finding leads to a relationship which assumes that some respondents are willing to give a truthful response without any protection or when their limited hazard is equal to one. Such a relationship may be depicted as shown in Figure 3.

Finally, adjacent categories tended to contain similar $\Pi_{Y\text{opt}}$. Thus, a topic with an unknown sensitivity could be classified into a range of sensitivity categories instead of a specific one. Since the $\Pi_{Y\text{opt}}$ for the adjacent categories fall into a small range, the selected $\Pi_{Y\text{opt}}$ would be close to the ideal one.

The model requires estimates of Π_A and yields recommended values of $\Pi_{Y\text{opt}}$. However, Π_A is not known exactly; if it were then the survey would be both redundant and unnecessary. Further, the value of $\Pi_{Y\text{opt}}$ actually used in the survey for $\Pi_{Y\text{opt}}$ may not be the recommended value. Thus, each of Π_A and $\Pi_{Y\text{opt}}$ is a potential source of nonsampling error.

This research has considered only the unrelated question randomized response model, $\Pi_{Y\text{opt}}$ known. It is possible that other models may contain respondent behavior curves that differ from those presented here. For example, a model offering the respondent a choice of one sensitive question and two unrelated questions may have curves that give higher probabilities of telling the truth because the respondent may feel more protected. This research has focused on only the respondents and has not considered the effects of the model design on nonresponse. Finally, empirical research documenting the behavior curves and providing validation of randomized response models in general would be an important contribution to this area.

BIBLIOGRAPHY

Abul-Ela, A.A. (1966), Randomized response models for sample surveys on human populations, Ph.D. dissertation, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Greenberg, B.G., Abul-Ela, A.A., Simmons, W.R., and Horvitz D.G., (1969), "The unrelated question randomized response model: theoretical framework," Journal of the American Statistical Association, 64, 520-539.

Greenberg, B.G., Kuebler, R.R., Abernathy, J.R., and Horvitz, D.G. (1977), "Respondent hazards in the unrelated question randomized response model," Journal of Statistical Planning and Inference, 1, 53-60.

Himmelfarb, S. and Lickteig, C. (1982), "Social desirability and randomized response technique," Journal of Personality and Social Psychology, 43, 710-717.

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

Lanke, J. (1975), "On the choice of the unrelated question in Simmons' version of randomized response," Journal of the American Statistical Association, 70, 80-83.

Soeken, K.L. and Macready, G.B. (1982), "Respondents' perceived protection when using randomized response," Psychological Bulletin, 92, 487-489.

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

Acknowledgements. This research was part of the principal author's doctoral dissertation at the University of North Carolina-Chapel Hill. Much appreciation is given to Drs. James R. Abernathy, Daniel G. Horvitz, William D. Kalsbeek, and Berton H. Kaplan for the time, contributions, and support they gave to this project. We would like to thank Mary Lou Pieranunzi and Jackie Swope for typing the manuscript.
A COMPLETE VERSION OF THIS PAPER IS AVAILABLE FROM DR. KAMMERMAN.

REFERENCES

Bonham, G. S. and L. S. Corder. 1981. NMCES Household Interview Instruments. National Health Care Expenditures Study. National Center for Health Services Research and National Center for Health Statistics. Instruments and Procedures Series #1. Department of Health and Human Services Publication No. (PHS) 81-3281.

Cohen, S. B., and W. D. Kalsbeek. 1981. NMCES Estimation and Sampling Variances in the Household Survey. National Health Care Expenditures Study. National Center for Health Services Research. Instruments and Procedures Series #2. Department of Health and Human Services Publication No. (PHS) 81-328.

Eckhart, K. W., and D. E. Wenger. 1975. "Respondent Coding of Occupations." Public Opinion Quarterly 40:245-255.

Finch, R. W., S. A. Mueller, and L. Godiksen. 1969. "The Reliability of Respondent-Coded Occupational Prestige." American Sociological Review 34:245-251.

Taylor, D. G. 1976. "The Accuracy of Respondent-Coded Occupation." Public Opinion Quarterly 40:245-255.

Chart 1

Group 1.
PEOPLE WHO OPERATE FARMS.
Some examples: owner or tenant farmers
farm managers

Group 2.
PEOPLE WHO DO OTHER FARM WORK.
Some examples: farm foremen
farm workers

Group 3.
PEOPLE WHO PRACTICE PROFESSIONS ON TECHNICAL SPECIALTIES.
Some examples: engineers
teachers
social workers
clergymen
physicians
registered nurses and therapists
writers and editors
artists and entertainers
computer programmers
laboratory technologists
analysts and researchers
counselors and consultants
draftsmen
airplane pilots

Group 4.
PEOPLE WHO ARE MANAGERS OR ADMINISTRATORS IN BUSINESSES, ORGANIZATIONS, OR GOVERNMENT.
Some examples: sales managers
bank officers and financial managers
public officials and administrators
union officials
business executives
restaurant managers
office administrators
shopkeepers

Group 5.
PEOPLE WHO SELL THINGS.
Some examples: insurance salesmen
real estate agents
sales representatives
sales clerks

Group 6.
PEOPLE WHO DO OFFICE OR CLERICAL WORK.
Some examples: postal clerks and mail carriers
bookkeepers
secretaries
telephone operators
cashiers
stock clerks

Group 7.
PEOPLE WHO PRACTICE SKILLED TRADES OR CRAFTS.
Some examples: carpenters
machinists
printers
heavy equipment operators such as cranemen
foremen
mechanics and repairman

Group 8.
PEOPLE WHO HELP MANUFACTURE OR PROCESS THINGS.
Some examples: meat cutters and butchers
assemblers
welders
lathe and milling machine operators
sewers and stitchers
packers and wrappers
checkers and inspectors
mine workers
clothing ironers and pressers

Group 9.
<u>PEOPLE WHO OPERATE OR SERVICE VEHICLES.</u>
Some examples: deliverymen
truck, bus, and taxi drivers
fork lift operators
railroad switchmen
garage workers and gas station attendants

Group 10.
<u>PEOPLE WHO PROVIDE SERVICES.</u>
Some examples: policemen and firefighters
practical nurses
guards and watchmen
cooks and chefs
waiters
hairdressers and barbers
custodians
maids
nurses aides, orderlies, and attendants

Group 11.
<u>PEOPLE WHO DO HEAVY PHYSICAL WORK.</u>
Some examples: construction workers
freight or stock handlers
gardeners and groundskeepers
vehicle washers
garbage collectors

Table 1. Percent agreement between professionally coded and respondent coded occupation, by self or proxy coder. (NMCES: United States, 1977)

Professional occupational classification of sample Individuals as:	Total Population		Self coders only			Proxy-coders only	
	Percent	n	Percent	n		Percent	n
"White-collar"							
Professional	73	(192)	76	(134)		71	(58)
Manager Administrator	58	(105)	53	(57)		65	(48)
Sales	69	(72)	67	(46)		73	(26)
Clerical	76	(199)	80	(135)	*	67	(64)
"Blue Collar"							
Crafts	69	(97)	63	(40)		74	(57)
Operatives	71	(111)	78	(55)		64	(56)
Transportation operatives	70	(37)	89	(9)		64	(28)
Services	82	(151)	82	(89)		84	(62)
Laborers	46	(24)	43	(7)		47	(17)
Total	72	(988)	74	(572)		70	(416)

*P less than or equal to .05.

SOURCE: National Medical Care Expenditure Study, National Center for Health Services Research.

Table 1. Optimal values[1] for Π_Y given Π_A, p, and degree of sensitivity of stigmatizing question; n=1,000

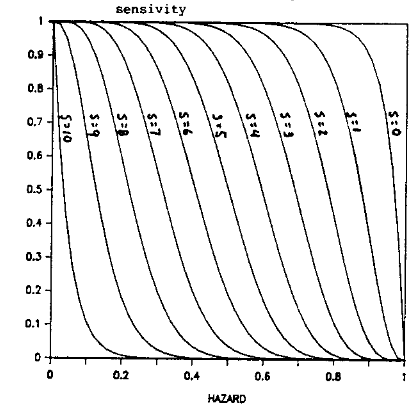
p	0	1	2	3	4	5	6	7	8	9	10
					DEGREE OF SENSITIVITY $(=1/2(\beta-1))$						
					$\Pi_A = .01$						
1/2	.001[2]	.003	.005	.008	.012	.018	.027	*	*	*	*
2/3	.002	.006	.011	.017	.025	.037	.058	.097	*	*	*
3/4	.004	.010	.017	.026	.039	.058	.089	.149	*	*	*
					$\Pi_A = .02$						
1/2	.003	.007	.012	.018	.027	.041	.065	.113	*	*	*
2/3	.006	.014	.025	.038	.058	.087	.138	.240	*	*	*
3/4	.009	.022	.038	.059	.088	.133	.210	.370	*	*	*
					$\Pi_A = .05$						
1/2	.009	.021	.036	.056	.086	.134	.226	.476	*	*	*
2/3	.019	.044	.075	.117	.179	.282	.484	1.000	*	*	*
3/4	.029	.068	.115	.179	.275	.432	.751	1.000	*	*	*
					$\Pi_A = .10$						
1/2	.022	.051	.088	.141	.226	.393	1.000	1.000	*	*	*
2/3	.047	.107	.184	.297	.483	.897	1.000	*	*	*	*
3/4	.072	.164	.281	.453	.743	1.000	1.000	*	*	*	*
					$\Pi_A = .25$						
1/2	.083	.201	.397	1.000	1.000	1.000	*	*	*	*	*
2/3	.178	.437	1.000	1.000	1.000	*	*	*	*	*	*
3/4	.275	.688	1.000	1.000	1.000	*	*	*	*	*	*

[1] Optimal value is based upon minimum MSE which, in turn, is derived from probability of answering truthfully according to the hazard presented by having to answer "Yes" to either question (or both).

[2] When Π_{Yopt} = .000, a direct question design is recommended.

* $|Bias| \geq \Pi_A$ for this cell. Π_Y = 1.00 is the recommended value to be used in a survey; however, the |Bias| may not necessarily be reduced to < Π_A.

Figure 1. Relationship between probability of a truthful response (T) and limited hazard for eleven categories of sensivity



S = sensitivity of stigmatizing question

$= \frac{1}{2} (\beta-1)$,

where $\alpha+\beta = 22$, $\beta = 1,3,\ldots,19,21$, and α and β are parameters of the beta probability density function.

Figure 2. Model II: Plots of Π_{Yopt} vs. sensitivity for p=1/2, 2/3 and 3/4; Π_A=.05; n=1000

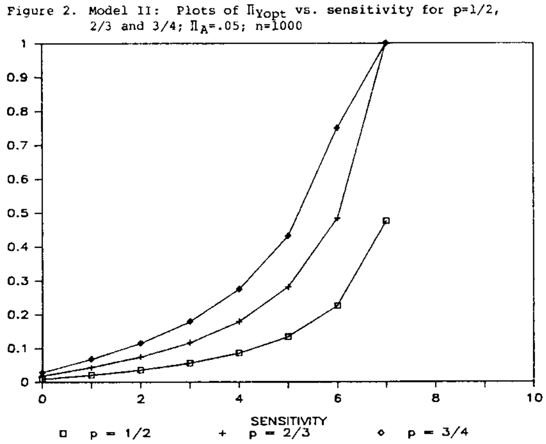


Figure 3. The postulated relationships between the probability of giving a truthful response (T) and the limited hazard for eleven categories of sensitivity