

## RESPONDENT REACTIONS TO ITEM COUNT LISTS AND RANDOMIZED RESPONSE

Michael L. Hubbard, Rachel A. Casper and Judith T. Lessler, Research Triangle Institute  
Michael L. Hubbard, P.O. Box 12194, Research Triangle Park, NC 27709

KEY WORDS: Non-Sampling Error, Laboratory Testing, Sensitive Issues

The research presented here was conducted at Research Triangle Institute as part of a contract from the National Center for Health Statistics and the Centers for Disease Control to plan and conduct the National Household Seroprevalence Survey (NHSS). Respondents in the NHSS are asked to donate a blood sample and complete a self-administered questionnaire about AIDS risk behaviors. A pilot study for the NHSS was completed this winter in Pittsburgh, PA with encouraging results: a response rate of approximately 85% was achieved. This high response rate was an important finding, since any estimates for rates of HIV infection in the general population may be biased to the extent that there exists a high rate of non-response (especially if non-response is disproportionately distributed within the population).

Despite the high response rate, there was still concern that respondents might not be giving entirely truthful answers to the sensitive AIDS risk questions. Although elaborate procedures were put in place to assure respondents that information they were providing would be kept anonymous (no names were taken; all address information was destroyed in the field after a completed questionnaire was obtained; the questionnaire was completely self-administered and it was sealed in a tamper-proof envelope which the respondent could mail himself or herself), we were concerned that respondents might be unwilling to disclose information about illegal or potentially stigmatizing behaviors.

In the next phase of the study, we decided to use additional indirect questioning techniques to give respondents an opportunity to give their answers in a more anonymous manner. We hoped to minimize response errors due to the denial of stigmatizing behavior by enhancing respondents' perception of the privacy of their answers. Two indirect questioning techniques from the survey research literature were selected for further study: Randomized Response and Item Count.

Randomized Response was originally formulated by Warner (1965) and alternate versions of the basic technique have been suggested by various authors including Horvitz, Shah and Simmons (1967). The technique consists of giving respondents two questions - the target question of interest and some irrelevant question - and having the respondent answer only one of them. The question to be answered is chosen using a random selection device such that only the respondent knows which question is being answered. For instance, a respondent might flip a coin to decide which question to answer. The non-target question is any irrelevant question for which answers have known probabilities (such as: "Was your mother born in April?"). Based on several reports in the literature (Folsom, 1974; Wiseman, Moriarty & Schafer, 1975-76; Bradburn & Sudman, 1979), we had questions about whether respondents would understand and accept Randomized Response questions.

The Item Count technique is a relative newcomer to the survey research literature, reported by Miller, Cisin & Harrell in 1986. Item Count questions consist of a list of behaviors the respondent may or may not have done. Respondents answer these questions by indicating the number of such behaviors they've done, but not which ones. Originally designed to be part of a wide-ranging, interviewer-administered survey, the Item Count technique avoids many of the pitfalls of Randomized Response: It is an unobtrusive questioning method (questions are embedded in a general survey) and no unusual procedures (like flipping a coin or using some other randomizing device) are required. Thus, using Item Count has not been found to arouse suspicion by drawing respondents' attention to the technique. Respondents need not be aware that a split sample design is being used wherein half of the respondents receive questions which include the target behavior in the list, while respondents in the other half are asked about the same list of non-target behaviors without the target behavior.

We had three basic concerns about using either of these two indirect questioning techniques in the next phase of the Seroprevalence Study. First, we wanted to know whether or not respondents could complete such questions in the context of a self-administered questionnaire on a single focused topic. Second, we wanted to design questions in such a way that useful statistical information could be derived from the answers. Third, we wanted to be sure that respondents found the questions acceptable and that they would answer them truthfully - more truthfully, we hoped, than the direct questions.

For Randomized Response, the main technical problem involves deciding what kind of random selection device is best for the interviewing situation. Whether the device involves flipping a coin, drawing a marble from an urn, or spinning a spinner, we wanted something which is easily understandable and no more obtrusive than necessary. The statistical considerations center on the selection of the alternate question: the better its statistical properties are known, the better the precision of the final prevalence estimates. The most crucial aspect of Randomized Response involves respondents' acceptance of the technique. We wanted to know how respondents react when presented with this rather unusual technique (for a survey) and whether or not they saw it as a way of protecting their privacy.

Item Count was a more unknown technique that would have to be adapted to a context different from the one for which it was originally designed. In terms of technical questions about the form of the questions, we had, as we saw it, two options. We could attempt to embed the Item Count within the rest of the questionnaire so that it would maintain its unobtrusive

characteristics. The alternative approach was to present Item Count as a separate, special way of asking questions which can offer additional privacy. The statistical issues with Item Count involve controlling the variability of the Item Count estimator. Since this variance is the sum of the variances of the individual items in the list, plus their covariances, it is desirable to choose a list which consists of a small number of items which are extremely low (or extremely high) in prevalence and which are not correlated with each other. If the prevalence of the non-target items in the list is higher than that of the target item, it is possible to estimate prevalences of the target behavior which are lower than those from direct questioning.

In terms of respondent acceptance of Item Count questions, Miller et al. state that questions work best when the items in them "fit together" or make sense to the respondent. In the present circumstances, this would mean that they would make sense in a self-administered questionnaire about AIDS. Items should also be non-verifiable behaviors and it should be highly unlikely that any respondent would have performed all of the behaviors described by the items. These features would be important in an embedded form of Item Count, where no attention is drawn to the technique. However, should Item Count be presented as a separate technique, then additional concerns would emerge - concerns similar to those regarding acceptance of Randomized Response. It would be important to know whether or not respondents see the technique as unusual, and whether or not they see it as providing enhanced privacy.

Our work in preparing to use indirect questioning methods involved designing appropriate questions and testing to see how they work in settings similar to those planned for the next phase of the Seroprevalence study. To accomplish this, we used the Cognitive Laboratory methods currently used at RTI. We designed a program of research which combined theory, methods and findings from cognitive sciences with information from respondents. This research tested hypotheses about various aspects of Randomized Response and Item Count questions in order to make a decision about using indirect questioning methods in the next phase of the NHSS project. We conducted a series of laboratory and field studies, going from observational studies with small samples to experimental studies testing hypotheses about factors which influence behavior using larger samples.

#### PROCEDURE

Since respondent acceptance was the problem for which we needed the most empirical information, our initial testing centered on respondents' reactions to the two techniques. We conducted two focus groups to try to ascertain the range of reactions and identify specific factors which affected those reactions. These groups were told the purpose of the indirect questioning techniques and both Item Count and several variants of Randomized Response (with different random selection devices and non-target questions) were demonstrated. Respondents in these groups were RTI employees - one group of non-college

educated data entry employees, the other group consisting of employees with college or post-graduate degrees.

Respondents' reactions to the indirect questioning techniques was mixed. Generally, Randomized Response was not well received. All respondents saw the technique as highly obtrusive and potentially disruptive. Generally, the response to answering Randomized Response questions was laughter: Using random selection devices like flipping coins or drawing marbles was seen as highly incongruent with a serious survey. While these respondents were not in the same position as a sampled person in the field, we found this reaction troubling. Comments indicated that the questions were seen as crazy, odd, weird, a joke, a game and not serious. Despite knowing the purpose of indirect questioning, they still found Randomized Response questions to be highly unusual.

Item Count questions were less troublesome to respondents. They understood how to answer them and their comments centered around the content of the non-target questions. Items completely different in content from the target item caused respondents to be suspicious of why they were being asked. Because respondents were aware that the purpose of the questions was to learn about AIDS-risk behaviors, they felt that admitting to more innocuous or positive behavior might be misconstrued as admissions of the more negative target behavior. Moreover, given that their attention was directed to the Item Count technique, a number of respondents wanted to know how information could be gained by asking questions in that way.

These results produced two decisions. First, the contracting agencies decided not to continue pursuing Randomized Response in the Seroprevalence Survey, feeling that too much time would be required to develop acceptable, working versions of Randomized Response questions. Second, we decided to implement a version of Item Count which was unobtrusive and embedded in the larger questionnaire.

The next round of testing was a series of fourteen cognitive interviews using RTI employees and gay men from the surrounding community as respondents. Respondents were given one of several versions of the self-administered materials to complete and asked to talk about their thought processes in answering the questions. We did not draw respondents' attention to the Item Count questions as they completed the questionnaire: The versions of Item Count they completed had minimal instructions for how to complete the questions and the questions were embedded in the larger questionnaire. The content of the questions involved items which we saw as related to AIDS and were perceived by respondents as fitting in with an AIDS questionnaire: The non-key items were behaviors involving travel to places where AIDS was more prevalent, blood contact, closeness to AIDS victims, sexual practices, and diseases mentioned in the Sample Person Questionnaire.

The results of this testing indicated that respondents found the Item Count questions embedded in the larger questionnaire to be

answerable and unobtrusive. There was even some evidence that at least one respondent felt more comfortable answering an Item Count version of a question than a direct question asking about the same content.

This version of Item Count was taken before the Dallas County Community Advisory Panel, an oversight committee which had to approve of all materials to be used in the next phase of the study, which was to take place in Dallas County, Texas. This group had serious reservations about the then current version of the Item Count questions. Their primary concern was with the content of the Item Count items. It was their belief that including items which described behaviors such as traveling, blood contact and proximity to AIDS victims would serve to perpetuate AIDS myths. Because the Seroprevalence Survey was to be a government sponsored survey about AIDS, they felt that asking about these activities would suggest to respondents that they are risk behaviors, when in fact they are not.

Furthermore, several staff members at NCHS and at RTI also had concerns about the current version of Item Count. They thought that Item Count required a clear perception on the part of respondents that their privacy is being protected. The way to enhance perceived privacy was thought to be by explicitly stating that Item Count questions protect privacy. Simply having respondents answer embedded Item Count questions would be no different than answering any other questions.

These criticisms meant that we could no longer use Item Count as an embedded, unobtrusive technique. Instead, we had to ask about these same behaviors in a context of items which clearly are not AIDS risk behaviors. This made the Item Count questions more obtrusive and left us uncertain about what type of item lists would be most acceptable and perceived as giving the most protection.

Our next step was to conduct the Item Count Rating Task: an experiment in which respondents were given different Item Count questions and asked to indicate which ones best protected their answers. We first constructed Item Count questions which varied on two dimensions: the type of non-target items (either stigmatizing or innocuous and, among the stigmatizing items, either similar or dissimilar in content to the target item) and the number of non-target items in the question (ranging from one to five low-prevalence items). In addition, we tested lists in which a single, non-stigmatizing, high-prevalence item was included with the various types of low-prevalence items. Testing was conducted in two stages: In the first stage, we constructed a 6 (Number of Items) X 3 (Type of Items) X 2 (Replications) factorial experiment in which 22 respondents made ratings of the protection given by the questions embodying the various combinations of number and type of items. In the second stage, 23 respondents made paired comparisons of the protection given by the most protective questions from Stage 1. Respondents for these experiments were RTI employees and members of the local community recruited as paid participants. The results of the rating task (Stage 1) can be seen in

Exhibit 1. There were significant main effects for Number of items ( $F(5,71) = 3.59, p < .05$ ) and Type of items ( $F(2,40) = 3.28, p < .05$ ). They indicate that respondents preferred longer lists (four or five non-target items) over short ones (one or two items) and questions containing all innocuous items over those containing any number of stigmatizing items (both comparisons significant at  $p < .05$  by Tukey's HSD test). The difference between lists which included a single high-prevalence item and those with all low-prevalence items was not significant. The paired comparison ratings from Stage 2 confirmed the preference for Innocuous items over Stigmatizing items (all  $p < .05$ ), while indicating that respondents did not strongly prefer having lists with one high-prevalence item over one with all low-prevalence items. Therefore these two types of lists were retained for further testing.

The final test we conducted was a four-way factorial field experiment using respondents who were more representative of the general population. The three Item Count factors (there was one additional factor related to the structure of the larger Sample Person Questionnaire) were:

- Item Content: Lists of four low-prevalence items vs. Lists of one high-prevalence and three low-prevalence items (all items described innocuous behaviors),
- Instruction Content: Short Instructions (containing a statement that Item Count protects privacy) vs. Long Instructions (containing an additional explanation of how Item count protects privacy),
- Instruction Form: Written (respondent-administered) vs. Oral (interviewer-administered) instructions.

Respondents in this large-scale field study were drawn from nearby community organizations - community college literacy classes, housing projects, rescue missions, and even some firemen from the Durham fire department. While not a random sample of the local community, these respondents did provide heterogeneity in terms of race, gender, marital status, and age groups. We made special efforts to include respondents with lower reading levels.

Each respondent was asked to complete both the larger sample person questionnaire and one version of the separate Item Count questions. We were interested in the effect of the experimental factors on several measures of how well the Item Count questions were working: First, we asked respondents whether they preferred answering Item Count questions or direct questions when being asked about risk behaviors; second, we looked at patterns of errors; and third, we examined written answers to a question about respondents' impressions of Item Count, as well as oral comments made during post-experimental debriefings.

Preference Ratings. Of a total of 61 respondents completing the Item Count questions, 27 preferred direct questions, 13 preferred Item Count questions and 21 expressed no preference, as can be seen in the upper panel of Exhibit 2. There was a stronger preference for Direct Questions with Oral Instructions than with

Written Instructions,  $\chi^2(2)=19.11$ ,  $p < .001$ . Preference for Direct questions over Item Count was not associated with Item Content

( $\chi^2(2)=0.15$ ,  $p > .90$ ) or to Instruction Content

( $\chi^2(2)=0.48$ ,  $p > .75$ ).

Errors. Item Count errors were (a) unanswered questions, (b) questions where individual items were checked off, rather than indicating the number of items in the space provided for answers, and (c) answers interpretable as indicating the performance of a given behavior rather than a count (for instance, when the number checked as the answer corresponds to the ordinal position of a high-prevalence item, which was always last in the list).

Eleven respondents had at least one error in completing Item Count questions, as seen in the lower panel of Exhibit 2. More errors were made in the Written Instructions condition than in the Oral Instruction condition (1 vs 10,

respectively,  $\chi^2(1) = 8.23$ ,  $p < .005$ ). More

respondents made errors when given lists containing a High Prevalence item than when given all Low Prevalence lists (7 vs. 4) and with Short vs. Long instructions (6 vs. 5). However, these differences were not

statistically significant,  $\chi^2(1) = 0.775$  and  $0.025$ , respectively, both  $p > .30$ . Nevertheless, we considered the higher number of errors with high-prevalence items lists unacceptable, and we chose to use only low-prevalence items in the final version of Item Count.

Respondent Comments. While respondents generally understood how to answer Item Count questions, several respondents commented in oral debriefings that they failed to see the technique as a privacy protection, despite explicit instructions aimed at emphasizing that fact. Most respondents thought that Item Count was unnecessary and that anyone who agreed to complete the questionnaire would be motivated to tell the truth and would prefer direct questions.

Other comments dealt with the items themselves, the technique in general, and respondents' preferences for direct questions over Item Count questions:

- Four respondents indicated that they did not like answering questions about non-AIDS activities.
- Four respondents wondered how anything could be learned by asking questions in this way.
- Six comments by respondents were expressions of preference for direct questions or general criticism of Item Count (Item Count is "tricky", "dumb", "too long", "beating around the bush").

Because they did not understand the purpose of Item Count, respondents sometimes attempted to impose some meaning of their own on the task: Several thought it might be a way of ascertaining social class. One respondent thought Item Count might be a personality test ("like the MMPI"). Several people also thought

that saying that they had done any behavior in a list with a risk item in it might be misconstrued as an admission of the risk behavior, making them trust the question less, not more.

Not all comments about Item Count indicated incomprehension or rejection. Nine respondents commented that Item Count hides answers. These comments can be interpreted as indicating that respondents recognize that Item Count protects an individual's privacy. However, these comments were usually couched as criticisms (Item Count is "imprecise", or "insensitive"; "You can't know which question is being answered"). Six of the nine still preferred direct questions in their preference ratings. This preference for Direct questions was not significantly different from the preferences of those who did not see Item Count as hiding

answers ( $\chi^2(2) = 2.27$ ,  $p > .25$ ).

Three respondents commented that they did see Item Count as providing additional privacy which could make people with something to hide feel safer. One respondent thought Item Count would lead to more truthful responding, while another thought that it might bring out more admissions from drug users. In the final case, the respondent, who was especially concerned with privacy, said that she liked Item Count because it protects privacy. Furthermore, as seen in the preference ratings, over 20% of respondents preferred Item Count questions.

#### DISCUSSION

Our original intention in this research was to decide whether two methods of asking indirect questions about sensitive AIDS-risk behaviors were acceptable for use in the National Seroprevalence survey. The two methods, Randomized Response and Item Count, had different strengths and weaknesses. We ended up adopting a sort of hybrid technique which combines aspects of both Randomized Response and Item Count. It follows the Item Count format of reporting the number of behaviors a person has enacted from a list. However, several undesirable aspects of Randomized Response were also present in our final version of Item Count. The technique is now obtrusive rather than unobtrusive: Rather than being embedded in a larger survey questionnaire, it is a separate document. Furthermore, the questions ask about behaviors totally irrelevant to the subject matter of AIDS, possibly arousing suspicion and confusion. The instructions that we are recommending for use in the next phase of the National Household Seroprevalence Survey attempt to highlight the privacy protection characteristics of Item Count. We have introduced an interviewer-guided practice question into our procedure to emphasize this privacy protection. However, these explicit instructions also serve to draw attention to the technique itself.

In an ideal study designed to ascertain whether an indirect questioning method yields more truthful answers from respondents, some means of validating answers is crucial. A study comparing direct and indirect questioning formats in a sample known to engage in the sensitive behavior would meet this ideal. Such

a study would also need a large enough sample size to overcome the statistical penalties imposed by the indirect methods. The studies presented today had a variety of limitations on how they were conducted which resulted in research which fell short of this ideal.

Primary among these limitations were political and procedural impediments. OMB guidelines limit the number of people who could see any version of the instruments without extensive and time-consuming clearance. Factors of community acceptability of question content restricted the type of questions which could be asked, and obviated the possibility of including people known to be at risk in the sample. Many problems are due to the context in which we were trying to develop the technique: for use in a survey with limited and sensitive subject matter, and with extensive assurances of confidentiality. Furthermore, financial and time constraints made it necessary to limit the range of options we explored and the type of studies we designed. In short, the real world made it impossible to conduct an ideal study.

We feel that, given the nature of the survey, the approach we are recommending has the best chance of success of all those we tested. We believe that both Item Count and Randomized Response are promising techniques for gaining information about sensitive behaviors, but that not enough is known about them. We don't know whether they work well or how respondents react to them. Basic research on these topics is still needed. The research presented today is a step in the right direction, but many more such steps are needed before we can confidently say that these techniques really lead respondents to feel more comfortable answering sensitive questions within a survey framework.

## REFERENCES

- Bradburn, N. M. and Sudman S. (1979). Improving Interview Method and Questionnaire Design. San Francisco: Jossey-Bass.
- Folsom, R. E. (1974). A randomized response validation study: Comparison of direct and randomized response reporting in DUI arrests. Research Triangle Institute report No. 254-807.
- Horvitz, D. G., Shah, B. V., and Simmons, W. R. (1967). The unrelated randomized response model. American Statistical Association, Proceedings of the Social Statistics Section, 65-72.
- Miller, J. D., Cisin, I. H. and Harrel, A. V. (1986). A new technique for surveying deviant behavior: Item-count estimates of marijuana, cocaine, and heroin. Paper presented the Annual Meeting of the American Association for Public Opinion Research, St. Petersburg, Florida.
- Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 60, 63-69.
- Wiseman, F. M. and Schafer, M. (1975-76). Estimating public opinion with the randomized response model. Public Opinion Quarterly, 39, 531-537.

Exhibit 1

First Stage of Item Count Rating Task: Mean Ratings (main effects)			
Number of Non-Target Items	Mean	Type of Non-Target Items	Mean
1 Low Prevalence	0.76	Innocuous	2.04
2 Low Prevalence	1.21	Stigmatizing/Dissimilar	1.35
3 Low Prevalence	1.35	Stigmatizing/Similar	1.32
4 Low Prevalence	1.63		
4 Low + 1 High Prevalence	1.93		
5 Low Prevalence	1.77		

Exhibit 2

	High Prevalence Item in List				No High Prevalence Item in List			
	Long Instructions		Short Instructions		Long Instructions		Short Instructions	
	Oral	Written	Oral	Written	Oral	Written	Oral	Written
Preferred Direct Q	6	1	5	1	4	2	6	2
No Preference	1	4	1	5	0	5	1	4
Preferred Item Count	4	2	2	2	2	0	2	2
Item Count Errors	0	2	1	4	0	3	0	1