

Survey Practice This Month

Tuesday, April 28, 2009, 12:04:40 PM | Editor

This month, Survey Practice brings back the Ask the Experts column. Aaron Maitland wrote a short review of the recent literature on labeling scale points for attitude questions.

Ineke Stoop presents some evidence from surveys where attempts to increase response were successful. The paper shows that data quality can be improved with targeted efforts to improve response but that simply increasing response does not necessarily improve the data.

In mail surveys, a particular challenge is to create mail materials that are opened by the sample member. In the article by Emily McFarlane and her colleagues, they show that neither using a stamp instead of metered mail, nor using personalized stickers had any effect on response rates.

We also include in this issue, a short summary of changes to Survey Practice that the editors and the Survey Practice Advisory Board are considering.

John Kennedy Diane O'Rourke
David Moore Andy Peytchev

survprac@indiana.edu

Posted in Monthly Summary [Comments: 0](#)

 [Comments](#)

Should I label all scale points or just the end points for attitudinal questions?

Tuesday, April 28, 2009, 12:03:37 PM | Editor

Aaron Maitland
National Center for Health Statistics*

The decision about whether to label all response scale points or just the end points for attitudinal questions can be an important and vexing decision for question designers. This short article will first discuss theoretical and practical considerations that should help guide the decision about how to label response scale points. Next, I will discuss some of the important empirical findings on this question from the literature. I will also provide some advice about how to evaluate verbal labels.

The amount of clarity that the labels add to the response scale is the most important consideration in the decision to label scale points. The approach to labeling should be the one that most clearly defines the response scale for respondents. One might argue that labeling of all scale points might offer an advantage in this regard. Several authors concede that it is probably more natural for a person to express his or her opinion using words (Fowler 1995; Krosnick and Fabrigar 1997). However, there is inherent ambiguity in the verbal labels that are frequently used with response scales. For example, people might have different interpretations of what it means to "somewhat favor" a public policy. Conversely, one might argue that even though numbers might be more abstract for most respondents, they might also be more accurate. For example, it has been noted that numbers in response scales at least convey the idea of equal intervals between points on a response scale (Krosnick and Fabrigar 1997). However, respondents can also vary in how they interpret numbers. Schwarz et al. (1991) present evidence that respondents interpret 10 point scales from -5 to +5 quite differently than 10 point scales ranging

from 1 to 10. They found that negative numbers imply the opposite of something, whereas the low end of a scale with all positive numbers merely implies the absence of something.

There are many other aspects of the research design that might influence the decision to label scale points. For example, the length of the scale will determine the feasibility of labeling all points. It will be much easier to create labels for 5 point scales than it will be for 11 point scales. As Fowler (1995) writes, "it is difficult to think up adjectives for more than 5 or 6 points along most continua." The mode of the interview also influences whether or not all scale points can be labeled. Generally, the use of telephone interviewing encourages shorter scales so that the respondent does not have to listen to a long list of response options before answering a question. Furthermore, when longer scales are used in telephone surveys it is more common to label only the endpoints of the scale (Dillman, Smith, and Christian 2008). Data collection methodologies that utilize visual modes of communication can more easily incorporate a full set of labels for response scales.

There are particular challenges with using verbal labels in cross-cultural research. Adequately translating fully labeled verbal scales into other languages is extremely difficult and can require considerable resources. For this reason, some (e.g., Fowler 1995) have suggested that numeric scales might be easier to translate and thus better suited for cross-cultural surveys since the researcher would only need to translate anchors at the ends of a scale. However, there is a dearth of evidence that numeric scales create more comparable survey data and cultural factors can also influence the interpretation of the numbers in a scale (Smith 2003).

Several empirical studies (Krosnick and Fabrigar 1997; Alwin 2007; Saris and Gallhofer 2007) have examined the effect of fully labeling scales versus partial labeling of scales on the quality of the resulting data. The general consensus from these studies is that fully labeled scales produce better data quality than partially labeled scales. Krosnick and Fabrigar (1997) came to this conclusion based on a review of several studies that examined the reliability and validity of fully versus partially labeled scales. Alwin (2007) compared the longitudinal reliability of 26 seven-point scales with only the endpoints labeled with 11 fully labeled seven-point scales from the National Election Studies. This study found that the fully labeled scales had a reliability of .719, whereas the scales with only the endpoints labeled had a reliability of .506. Saris and Gallhofer (2007) also concluded that labeling all points had a positive effect on reliability based on a meta-analysis of 1023 survey questions from 87 multi-trait multi-method experiments. Labeling did not have a significant effect on validity according to their meta-analysis.

There are a couple of other reasons to believe that labeling might lead to better data quality. First, improvements in both reliability and validity tend to be the greatest amongst respondents with lower levels of education (Krosnick and Fabrigar 1997). This is a group that frequently encounters comprehension problems in surveys and question designers are often looking for design strategies to improve data quality among this group. Second, fully labeling scales seem to reduce the effects of question features that are unrelated to the response task. For example, a study by Tourangeau, Couper, and Conrad (2007) conducted experiments using Web surveys that varied the color of the shading on response scales and the use of labels. They found that the effect of color on respondents answers - something designers would not intend - disappeared when the response scales were fully labeled. Hence the literature reviewed in this article seems to suggest that fully labeling has a number of advantages over labeling only the endpoints.

The decision about which labels to use is just as important as the decision about whether or not to use the labels at all. Saris and Gallhofer's (2007) meta-analysis offers some guidelines for selecting labels. Generally, symmetrical labels tend to yield higher reliability and validity. Verbal labels that match the numbers on the scale lead to higher reliability. For example, it might be better to have bipolar scales with negative labels matching negative numbers and positive labels matching positive numbers. Finally, if one does decide to label only the endpoints, these labels should represent fixed reference points (e.g. completely dissatisfied - completely satisfied, instead of dissatisfied-satisfied) so that the end of the scale is well defined.

Even though there are some guidelines, it is always necessary to evaluate a question to determine the best labeling. There are a number of question evaluation methods that might be useful for designing scale labels. Qualitative techniques such as

cognitive interviewing that make use of probing or thinkaloud methods will help to understand how respondents assign meaning to the labels on a response scale and whether that meaning matches the survey designer's intended meaning. Item response theory modeling approaches provide a useful quantitative tool to assess whether the pattern of responses matches the theoretical underpinnings of the response scale. For example, one can assess whether the points on the scale are increasingly difficult to endorse as one expresses more extreme attitudes. It is always good practice to use a combination of qualitative and quantitative methodologies such as these to design good survey questions.

In conclusion, the approach to labeling should be the one that most clearly defines the response scale for respondents. The existing empirical literature generally suggests that fully labeled scales have an advantage over partially labeled scales according to a number of criteria. However, practical considerations such as the mode of the interview, number of scale points, and the population under study can play an important role in deciding the extent of labeling that should be used. Additionally, question designers should rely on appropriate question evaluation methods to determine which labeling approach is best for a specific research design.

** The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.*

References

Alwin, D.F. Margins of Error: A Study of Reliability in Survey Measurement. New York, NY: John Wiley and Sons, Inc. 2007.

Alwin, D.F., and Krosnick, J.A. "The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes." *Sociological Methods and Research* 20 (1991): 139-181.

Dillman, D.A., Smith, J.D., and Christian, L.M. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. New York, NY: John Wiley and Sons, Inc. 2008.

Fowler, F.J. *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage

Krosnick, J.A. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5 (1991): 201-219.

Krosnick, J.A., and Fabrigar, L.R. "Designing Rating Scales for Effective Measurement in Surveys." *Survey Measurement and Process Quality*. Ed. Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N., Trewin, D. New York, NY: John Wiley and Sons, Inc. 1997. 141-164.

Saris, W.E. and Gallhofer, I.N. *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. John Wiley and Sons, Inc. 2007.

Schwarz, N., Knauper, B., Hippler, H.J., Noelle-Neumann, E., and Clark, F. "Rating Scales May Change the Meaning of Scale Labels." *Public Opinion Quarterly* 55 (1991): 618-630.

Smith, T. "Developing Comparable Questions in Cross-National Surveys." *Cross-Cultural Survey Methods*. Harkness, J., Van de Vijver, F., Moher, P. New York, NY: John Wiley and Sons, Inc. 2003. 69-91.

Posted in Ask the Experts [Comments: 5](#)

 [Comments](#)

A few questions about nonresponse in the Netherlands

Ineke Stoop
Netherlands Institute for Social Research

In the recent past the Netherlands has stood out as a country with rather low response rates; not a good starting position in a period when response rates are going down (De Heer, 1999; De Leeuw and De Heer, 2002). This paper seeks to examine a number of assumptions on non-response in the Netherlands. Is the situation really so bad compared to other countries, and are response rates really going down? The answer to these questions will turn out to be negative. This leads to other questions: how were higher response rates achieved in the past few years? Can this level be maintained? And finally: do higher response rates result in a final sample that better represents the target population?

[Figure 1](#) shows that the Netherlands (NL) is doing fairly well in the European Social Survey. Response rates are not as high as in the Scandinavian countries (NO, SE, FI) or Poland, but are higher than in Switzerland (CH), France (FR) and Germany (DE). Low response rates in the Netherlands need not be a fact of life.

[Figure 2](#) shows that response rates in a series of major Dutch face-to-face surveys were much higher in the past than they are today, but also that response rates have been going up again in recent years. This raises three questions. Why were response rates so high in the past? Why did they go up again? And can they be maintained at their present level?

The first question is why response rates were so high in the past. This could be due to the fact that people were less used to - and more interested in - surveys than nowadays, or that they were more 'open' to authority (these are mostly government surveys) or were more accustomed to people knocking on their door with all kinds of requests. It could also be that the procedures for calculating response rates were far less standardized than today. In many of the surveys no call records were kept; substitution must have taken place to some extent; and the distinction between ineligibility and noncontact may not always have been clear (or at least less clear than it is today). So in fact response rates may not have been as high as they appear.

Even if high response rates in the past may have been doctored a little, it is clear they went down initially. This was most likely due to lack of field efforts, a failure to adapt to the increasing difficulty of obtaining response. The increase in later years was due to a combination of extended field efforts (more contact attempts at different times of the day and on different days of the week and in some cases (AVO) extensive refusal conversion attempts and successes), better training and closer monitoring of interviewers, and the use of small incentives and informative advance letters and brochures. Compared to the past, obtaining fairly high response rates today takes a lot of work, a lot of time and a lot of money. If these investments are not made, response rates will be low. If they are made, Dutch results look quite good, including when compared with other European countries.

Despite these good results, it may still be difficult to maintain the present response rates. Extended field efforts may have held off the decrease in response rates, but they may not be able to counteract the continuing downward trend. Indeed, there is evidence that response rates are slowly declining again despite the efforts. The negative consequences of this can be deduced from studying the positive effect of the previous increase in response rates in the AVO (Stoop, 2004) and DTUS (Van Ingen et al., 2009), two long and time-consuming SCP surveys. These effects were studied partly out of methodological interest, and partly because these studies are designed to measure changes over time and the 'time' effect had to be distinguished from the 'method' effect.

In the AVO survey, as in many other studies, people living in urban areas, persons who pursue activities outside the home and persons who are part of smaller households are more difficult to catch at home. Increasing the number of calls, especially in the evening, resulted in higher contact rates and a better representation of those who are hard to reach. This is likely to reduce nonresponse bias, as the hard to reach participate more in cultural activities, for instance, even when corrected for background variables. In this survey, a large proportion of the respondents initially refused and had to be converted. Among these initially reluctant respondents were relatively few men and few singles. Rather than concluding that

men and singles immediately cooperated, it turned out that they were less willing to participate at the first request and were also less susceptible to refusal conversion attempts by the interviewer. The same is true for people with a higher education level. After correcting for socio-demographic factors, there were very few differences between cooperative and initially reluctant respondents.

The final nonrespondents who cooperated in an 80% response follow-up survey of refusers more often partook in popular culture (movies, pop concerts, clubbing), played fewer sports, less often had a PC - but when they did have one they used it more for games and chatting on the Internet - and less often had a religious affiliation. Although the differences were not large and were limited to a small number of variables, they were not similar to the initially reluctant respondents in the main survey. In this study refusal conversion mainly brought in additional respondents who were very similar to those who were overrepresented anyway, and who were not similar to the final refusals.

In the DTUS, increasing the contact rate resulted in a larger share of respondents with paid jobs and a higher education level, whereas the proportion of elderly respondents decreased considerably. This makes sense as the former spend less time at home and the latter more. However, the employed, younger people and the better educated were overrepresented from the beginning (and the elderly underrepresented). The hard-to-reach respondents spent more time on commuting, which is likely due to their more often being employed. As compared to the refusers (or at least the 31% of the refusers who completed a short doorstep questionnaire), the respondents more often read newspapers, watched television and were interested in politics, played sports more frequently, traveled more during the week and were engaged in volunteer work more often. Contrary to expectations, there is no evidence, either in terms of actual (demands of work, traveling, household chores, etc.) and perceived time pressure, that busy people cooperate less. This confirms earlier findings that busyness is not a cause of nonparticipation in time use surveys, and that time use on voluntary work may be overestimated (Stoop, 2005; Abraham et al., 2006).

The results of the two studies are both reassuring and disconcerting. Reassuring because response rates can be enhanced substantially. We do not have information on the exact costs of improved sample and fieldwork designs and the higher response rates, partly because so many things change when a survey is conducted every 4 or 5 years that straightforward price comparisons are difficult. Also reassuring is that, after correcting for background variables, the effect of the increased response rates on survey estimates is limited. This means that time series are not damaged by method effects. The limited effect is also disconcerting, however. Firstly, because substantial funds have been spent on survey improvement whereas there is little evidence that these efforts manage to bring in respondents who would not have cooperated otherwise. The additional respondents belong to socio-demographic groups that would have been well-represented with smaller efforts. And secondly, they are disconcerting because the results from the AVO follow-up study and the DTUS basic questions suggest that the respondents who were hard to contact or initially reluctant differ to a certain extent from the final non-respondents on substantive survey outcomes, again after correcting for background variables. Enhancing response rates may thus not minimize bias.

The present results suggest that, rather than spending funds on increased fieldwork efforts to indiscriminately haul in additional respondents, it might be better to focus field efforts on target respondents who are unlikely to participate. Ideally this strategy would be adapted throughout the fieldwork process (Groves and Heeringa, 2006). In addition auxiliary information should be collected on respondents and non-respondents from rich sample frames, interviewer observations, call records and reasons for refusal. In addition, high response basic question approaches or follow-up surveys of refusals could show us what we are missing or what is being distorted. High response rates are a **means toward** data quality. Low non-response bias is an **indicator** of data quality.

References

Abraham, Katharine G., Aaron Maitland, and Suzanne M. Bianchi (2006) Nonresponse in the American Time Use Survey: Who Is Missing from the Data and How Much Does It Matter? *Public Opinion Quarterly*, Vol. 70, pp. 676-703.

De Heer, Wim (1999) International Response Trends, Results of an International Survey. *Journal of Official Statistics*, Vol. 15,

N° 2, pp. 129-142.

De Leeuw, Edith, and Wim de Heer (2002) Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In: Robert M. Groves, Don A. Dillman, John L. Eltinge and Roderick J.A. Little (eds.) *Survey Nonresponse*. New York: Wiley, pp. 41-54.

Groves, Robert M., Steven Heeringa (2006) Responsive design for household surveys: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society Series A: Statistics in Society*,

Stoop, Ineke A.L. (2004) Surveying Nonrespondents. *Field Methods*, Vol. 16, nr. 1, pp. 23-54.

Stoop, Ineke A.L. (2005) *The Hunt for the Last Respondent*. The Hague: SCP.

Van Ingen, Erik, Ineke Stoop, and Koen Breedveld (2009) Nonresponse in the Dutch Time Use Survey: Strategies for Response Enhancement and Bias Reduction. *Field Methods*, Vol.21, No. 1, pp. 69-90.

Note : This article is based on:

Stoop, Ineke (2008) Niet thuis voor de enquêteur. *Facta*, nr. 16 (3), pp. 12-17.

Stoop, Ineke, Josine Verhagen and Erik van Ingen (2007) Increased fieldwork efforts, enhanced response rates, inefficient samples. In: M. Trotman et al. (eds.) *The Challenges of a Changing World*. Proceedings of the 5th International Conference of the Association for Survey Computing. University of Southampton, September 2007, pp. 167-178.

Posted in [Methods](#) [Comments: 0](#)

 [Comments](#)

Does the use of stamps and/or stickers increase physician participation in mail surveys?

Tuesday, April 28, 2009, 11:59:52 AM | Editor

Emily McFarlane, RTI International
Joe Murphy, RTI International
Murrey Olmsted, RTI International

Over the years, a number of methods have been investigated to increase response rates to mail surveys. These methods include pre-notification letters, incentives, reminders, survey form design, endorsement letters, sending additional surveys, shortening survey length, and delivery method. These methods have been demonstrated to have a positive impact on the level of participation in surveys with the general population (Fowler, 1993; Dillman, 2000) as well as with many specialized populations such as physicians (Kellerman and Herold, 2001). Despite these efforts, response rates for mail questionnaires, and in particular mail questionnaires of physicians have been declining over the past decade (Cull, Karen, O'Connor, Sharp, and Tang, 2005). As a result, researchers must continue to explore possible methods for increasing survey response.

Some research with physician populations has shown that the use of stamps can be effective compared to metered or business reply envelopes for return mail (Streiff, 2001; Kellerman and Herold, 2001; Urban, Anderson, and Tseng, 1993; Fox, 1998; Armstrong and Lusk, 1987). However, there is no current research that examines the use of stamps on outgoing mail. Seven studies conducted in the 1970's found no difference in response rates for mailings that used a stamped outgoing envelope versus a metered outgoing envelope (Edwards, 2007; Gullen, 1973). Yet, it is unclear what the effect would be today. It is generally recommended to use first-class mail over bulk-mail to improve response (Dillman, 2000), but it is unknown whether the use of first-class stamps compared to first-class metered postage will have a similar effect on

response rates. Armstrong and Lusk (1987) posit that stamps might be expected to increase response rates compared to metered mail because respondents perceive stamps as more personal. While this did not prove to be true several decades ago, it may have an effect now. In a letter to *The Lancet* in 2000, one physician noted that he received approximately 122kg of medical junk mail in one year (Montauk, 2000). The use of stamps may distinguish a survey request from what is perceived as unwanted medical junk mail. In addition to using stamps on the outgoing envelopes, there may be other ways to personalize the appearance of the envelope encouraging recipients to open and respond to the mail. In addition, most research comparing the results of the stamped return envelopes versus business reply mail were conducted on either the first mailing or the study overall. What impact on response rates, if any, do stamped envelopes have on nonresponse follow-up mailings?

In an effort to add to this body of literature, this paper discusses two experiments focused on improving physician survey cooperation by altering the appearance of the outgoing and return survey envelopes. The experiments were conducted within a national survey of board certified physicians by researchers at RTI International as part of the "America's Best Hospitals" project conducted for *U.S. News & World Report*.

In the first survey mailing, the authors tested the effect of postage stamps compared to metered mail on the outgoing envelope. The authors also tested the effect of including a small sticker on the outgoing envelope as well. The sticker was intended to have the same personalizing effect of the stamp while also identifying the purpose of the envelope. The two methods were combined into a 2 by 2 factorial design. The stickers indicated that the material in the letter was from the "America's Best Hospitals" project and highlighted the sampled physician's medical specialty. The authors hypothesized that both the stamps and the stickers would be viewed as more personal by respondents. Therefore more respondents would open and ultimately respond to the survey.

In order to test the effect of stamped return envelopes on the follow-up mailings, an additional experiment was conducted on the third, fourth, and fifth nonresponse follow-up mailings to the physicians. The authors tested the effect of business reply mail compared to stamped return envelopes, and the use of two different types of stamps on the return envelopes: American flag stamps compared to breast cancer research stamps. Breast cancer research stamps were chosen because they donate six cents per stamp to breast cancer research and this might appeal to physicians. Previous research by Choi (1990) showed that commemorative stamps on return envelopes increased response compared to regular stamps or business reply mail.

METHODS

Since 1990, *U.S. News & World Report* has assessed the quality of hospitals in the United States annually in the form of lists collectively titled "America's Best Hospitals." Each year, the magazine identifies hospitals of exceptional quality from over 5,000 hospitals in the United States across a variety of medical specialties. Hospitals are assigned a composite score and ranked based on data from multiple sources. One of the primary sources of data is a survey of board-certified physicians asking them to nominate the "best hospitals" in their medical specialty.

The sample for the 2007 physician survey consisted of 3,400 board-certified physicians selected from the American Medical Association (AMA) Physician Masterfile. Stratifying by census region (Midwest, Northeast, South, West) and medical specialty, we selected a random probability sample of 200 physicians (50 from each region) from each of the 17 specialty areas. The 17 medical specialties represented in the sample included the following: Cancer; Digestive Disorders; Ear, Nose, and Throat; Geriatrics; Gynecology; Cardiology; Hormonal Disorders; Kidney Disease; Neurology and Neurosurgery; Ophthalmology; Orthopedics; Pediatrics; Psychiatry; Rehabilitation; Respiratory Disorders; Rheumatology; and Urology.

The physician survey mailings were conducted in stages during several weeks. The initial mailing was sent via United States Postal Service (USPS) first-class metered mail. It included a cover-letter, survey, business-reply envelope and a \$2 bill as an incentive. Three weeks after the initial survey mailing, a thank-you/reminder letter was sent to the sampled physicians with a replacement survey and business reply envelope. Two weeks following the reminders, we sent a USPS Priority mailing to nonresponders, along with another copy of the questionnaire, a new cover letter, and a business reply envelope. Two weeks

after the priority survey was sent, a fourth survey mailing was sent overnight via Federal Express to the remaining nonresponders; the packet included the questionnaire, a cover letter, and a business reply envelope. A final mailing was sent via USPS first-class mail approximately 4 weeks later. This mailing included the questionnaire and a personalized letter with a handwritten note and signature. (See **Table 1** for a schedule of the physician survey mailing.)

Table 1: Physician Survey Mail Schedule

<i>Mailing</i>	<i>Sent via</i>	<i>Date</i>	<i>Experiments</i>
1	USPS, first-class letter	Sep. 15, 2007	Metered mail versus postage stamps on outgoing mail and stickers versus no stickers on outside envelope
2	USPS, first-class letter	Oct. 6, 2007	No experiments
3	Priority mail	Oct. 20, 2007	Business reply mail versus two types of stamped return envelopes
4	Federal Express	Nov. 3, 2007	Business reply mail versus two types of stamped return envelopes
5	USPS, first-class letter	Dec. 8, 2007	Business reply mail versus two types of stamped return envelopes

Experimental Conditions

To evaluate methods for improving the overall level of participation in the survey, two experiments were conducted to see if subtle changes in the appearance of the outgoing and return envelopes would have any impact on physician participation in the survey. The experiments are described below.

Experiment 1: Appearance of outgoing envelope

In the first survey mailing, we tested the effect of using postage stamps on the outgoing envelopes compared to traditional metered mail. The authors also tested the use of stickers on the outgoing envelope. The stickers were bright blue and indicated that the material in the letter was from the “America’s Best Hospitals” project and highlighted the sampled physician’s medical specialty. The two alterations were combined into a 2 by 2 factorial design, where physicians in each region/specialty combination were randomly assigned to one of four experimental conditions such that 850 physicians or 25% of the sample fell into each of the 4 conditions (See **Table 2**). This allowed the authors to assess the individual and joint effects of the experimental conditions. The results were analyzed using an two-way analysis of variance test. The sticker-stamp experiment was not conducted on any of the follow-up mailings.

Table 2. Sample distribution for postage and sticker experiment

	<i>Postage stamp</i> <i>n (%)</i>	<i>Metered postage</i> <i>n (%)</i>
Sticker	850 (25%)	850 (25%)
No sticker	850 (25%)	850 (25%)

Experiment 2: Return envelope postage

On the third, fourth, and fifth nonresponse follow-up mailings to physicians, an additional experiment was conducted exploring different types of postage used on the return envelope. The authors compared three types of postage: (1) traditional postage-paid business reply mail, (2) a standard 39-cent first-class postage stamp of an American flag, and (3) a 45-cent breast cancer research stamp, which donated 6 cents per stamp to breast cancer research. Of the 2,450 nonresponders after the first and second mailing, 50% were randomly assigned to receive a postage-paid business reply envelope. The remaining nonresponders were randomly assigned to receive either the American flag stamp or the breast cancer research stamp. See **Table 3** for approximate sample distributions. Chi-square tests were used to analyze the impact of these methods on response rates.

Table 3. Approximate sample distribution for postage and sticker experiment

<i>Experimental Group</i>	<i>Distribution</i> <i>n (%)</i>
Business reply envelope	1226 (50%)
American flag stamped return envelope	612 (25%)
Breast cancer stamped return envelope	612 (25%)

RESULTS

Data for the physician survey were collected from September 15, 2006 to January 31, 2007. The survey achieved an overall response rate of 47.2% as calculated using American Association for Public Opinion Research (AAPOR) standard response rate 6 (standard definitions are located on the Web at www.aapor.org/pdfs/standarddefs_ver3.pdf), which treats undeliverables as ineligible cases. The results of the two experiments are reported below.

Experiment 1: Appearance of outgoing envelope

Chi-square tests show that the respondents in each of the four groups were not significantly different with regard to several demographic variables that may have an effect on survey response age, region, gender, years since graduation (McFarlane, 2007).

A total of 35 surveys were considered ineligible because they were undeliverable or because the sample member was deceased or no longer practicing medicine. Out of the 3,365 eligible surveys, 648 surveys were received during the first three weeks of data collection, before the second mailing was sent, and were included in the analyses for experiment 1. Table 4

shows the response rate for the four experimental conditions and for the use of stamps overall and the use of stickers overall.

Table 4. Response Rates for Sticker-Stamp Experiment

	<i>Stamp</i>	<i>No Stamp</i>	<i>Total</i>
<i>Sticker</i>	19.1%	19.9%	19.5%
<i>No Sticker</i>	19.9%	18.2%	19.0%
<i>Total</i>	19.5%	19.0%	19.3%

The response rate data were analyzed to determine if the different treatment groups resulted in significant response rate differences and what, if any, are the main effects or interactions of using stickers

An analysis of variance did not reveal a significant effect of treatment on response rate ($F = 0.38$, $df_s = 2$ and, $p = .7679$). The test results did not show statistically significant separation between groups that received stickers and groups that did not receive stickers ($F = 0.13$, $df = 1$, $p = .7218$). The second factor analyzed the impact on survey response rates for groups that received stamped envelopes versus those that received metered envelopes. The test results do not show a significant main effect on response rates for stamps ($F = .10$, $df =$, $p = .7483$). Finally, the interaction between the two independent variables was not statistically significant ($F = .91$, $df =$, $p = .3411$). However, the group that received mailings with neither a sticker or stamp had the lowest response rate.

Experiment 2: Return envelope postage

Chi-square tests show that the respondents in each of the three groups were not significantly different with regard to several demographic variables that may have an effect on survey response age, region, gender, years since graduation.

The analyses for the second experiment were limited to the 2,450 physicians who did not respond to the first or second mailings. The nonresponders were randomly assigned to one of the three groups. The three groups were not statistically different with respect to region, sex, age and years since graduation. Up to three additional mailings were sent to these nonresponders, which we will call mailings 3, 4, and 5. Table 5 shows the cumulative response rates for the three groups after each mailing. Chi-square analyses did not reveal a significant effect of treatment group on response rate for any of the mailings.

Table 5. Response Rates for Return Envelope Postage Experiment

<i>Data Collection Condition</i>			
Data Collection Phase	<i>No Stamp</i>	<i>Traditional Stamp</i>	<i>Breast Cancer Stamp</i>
Mailing 3			
Completes	126	59	55
Response Rate	10.3%	9.8%	9.1%
		<i>p</i> = .6930	
Mailing 4			
Cumulative Completes	272	131	136
Cumulative RR	22.8%	22.2%	23.0%
		<i>p</i> = .7285	
Mailings 5			
Cumulative Completes	300	151	152
Cumulative RR	25.5%	26.0%	26.3%
		<i>p</i> = .9051	

It was also hypothesized that the breast cancer stamps might have a larger impact with women compared to men and with oncologists compared to other specialists. To compare differences between men and women, we limited the analyses to only those sample members who received a stamped return envelope (n = 1,224). An analysis of variance showed no differences between the groups for males versus females ($F = 0.21$, $df_s = 3$ and $p = .8925$).

We then limited the analyses to sample members who were oncologists to determine if there was a difference in response for this group between the two types of stamps. However, only 74 oncologists were included in the follow-up mailings. Table 6 shows the cumulative response rate for the three follow-up mailings for oncologists by the two experimental groups.

Table 6. Response Rates for Stamp Experiment on Oncologists

	<i>Breast Cancer Stamp</i>	<i>American Flag Stamp</i>
Completes	15	9
Response Rate	44.1%	24.3%
	<i>p</i> = .0823	

There was a marginally significant difference between response rates for those who received the breast cancer stamps and those who received the American flag stamps. Response rates were considerably higher for physicians who received breast cancer stamps compared to American flag stamps.

SUMMARY AND DISCUSSION

This study found that surveys sent using first-class metered postage achieved a comparable response rate among physicians to surveys sent using first-class stamps. Tailored stickers applied to the outside of the envelope were also not effective in increasing response rates. While not statistically significant, the group that received neither a stamp nor a sticker on the outgoing envelope had the lowest response rate. These findings suggest that including something on the outgoing envelope may be better than nothing. Further research investigating different approaches for altering the appearance of the outgoing envelope is needed.

This study also found that neither the use of stamped return envelopes nor the type of stamp (American Flag versus Breast Cancer research) affected the overall physician response rates in follow-up survey mailings. As a result, stamped return envelopes were also not cost-effective compared to business reply mail. However, this experiment was not conducted on the first survey mailing, so it is impossible to tell if stamped return envelopes are ineffective in this population or ineffective for follow-up mailings only. In the 2008 survey, the authors are going to continue the stamp experiment on the first mailing.

The use of breast cancer research stamps did not improve response rates over standard stamps or business reply mail for the entire sample of physicians. However when the analyses were limited to oncologists only, the results showed that breast cancer research stamps marginally improved response rates compared to American flag stamps. The results of this analysis are extremely limited, however, due to the small sample size of 74. The findings suggest that breast cancer stamps may be a promising approach to increasing response rates specifically among oncologists. Further research investigating the use of breast cancer research stamps with oncologists is needed.

REFERENCES

- Armstrong, J. S., & Lusk, E. J. (1987). Return postage in mail survey: A meta analysis. *Public Opinion Quarterly*, 51, 233-248.
- Asch, D. A., Christakis, N. A., & Ubel, P. A. (1998). Conducting physician mail surveys on a limited budget: A randomized trial comparing \$2 bill versus \$5 bill incentives. *Medical Care*, 36, 95-99.
- Choi, B. C., Pak, A. W., Purdham, J. T. (1990). Effects of mailing strategies on response rate, response time, and cost in a questionnaire study among nurses. *Epidemiology*, 1(1), 72-4.
- Cull, W., O'Connor, K.G., Sharp S., and Tang, S. S. (2005). Response Rates and Response Bias for 50 Surveys of Pediatricians. *Health Services Research*, 40, 213-225.
- Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method*. 2nd ed. New York, NY: John Wiley & Sons.
- Edwards P, Roberts I, Clarke M, DiGiuseppi C, Pratap S, Wentz R, Kwan I, Cooper R. (2007). Methods to increase response rates to postal questionnaires. *Cochrane Database of Systematic Reviews*, Issue 2. Art. No.: MR000008. DOI: 10.1002/14651858.MR000008.pub3
- Fowler, FJ, Jr. (1993). *Survey research methods*, 2nd ed. Newbury Park, CA: Sage Publications, Inc.
- Fox R. J., Crask M. R., & Kim, J. (1988). Mailed survey response rate: A meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly*, 52, 467-491.
- Gullen W. H, & Garrison G. E. (1973). Factors influencing physicians' response to mailed questionnaires. *Health Services Rep*, 88, 510-4.
- Kellerman, S. E., & Herold, J. (2001). Physician Response to Surveys: A Review of the Literature. *American Journal of*

Preventive Medicine, 20, 61-67.

Montauk, I. (2000). Medical junk mail, *Lancet*, **356**, 344.

McFarlane, E.M., Olmsted, M. G., & Murphy, J. (2007). Nonresponse Bias in a Mail Survey of Physicians. *Evaluation and the Health Professionals*, 30(2), 170-185.

Streff M. B, Dundes L., & Spivak J. L. (2001). A mail survey of United States hematologists and oncologists: A comparison of business reply versus stamped return envelopes. *Journal of Clinical Epidemiology*, 54, 430-2.

Shiono P. H, & Klebanoff, M. A. (1991). The effect of two mailing strategies on the response to a survey of physicians. *American Journal of Epidemiology*, 134(5), 539-42.

Urban N., Anderson G. L. , Tseng A. (1993). Effects on response rates and costs of stamps vs business reply in a mail survey of physicians. *Clinical Epidemiology*, 46(5), 455-9.

For more information on the "America's Best Hospitals" project:

Website: www.RTI.org/BestHospitals

Email: BestHospitals@RTI.org

A version of this article was presented at the Southern Association for Public Opinion Research (SAPOR) meeting on *October 4, 2007*

Posted in [Methods](#) [Comments: 0](#)

 [Comments](#)

Changes to Survey Practice

Tuesday, April 28, 2009, 11:59:10 AM | Editor

Survey Practice has been issued monthly since August 2008. We planned that during the first two years, we would experiment using different formats, publication schedules, content, and the number of articles. We believed that the flexibility offered by an Internet publication could allow us to adjust the publication as it develops. The SP editors are now considering a number of changes.

Our first change is to a bimonthly publication schedule. We will publish the next issue in June. Instead of 3-4 articles in each monthly issue, we will publish 6-7 in each bimonthly issue. We are doing this to reduce the number of email announcements and to provide a wider range of topics in each issue.

During the AAPOR meeting in May, we will be discussing possible changes with the SurveyPractice Advisory Committee. We also hope to discuss the changes informally with others who attend the meeting.

If you have thoughts on changes or improvements to SP, please contact us through email or talk to us during breaks and meals at the AAPOR meeting. We very much welcome your thoughts on how to make SP better and more useful to you.

John Kennedy
David Moore

Diane O'Rourke
Andy Peytchev

survprac@indiana.edu

Posted in Monthly Summary [Comments: 0](#)

 [Comments](#)