IMPROVING PRODUCTIVITY IN AN ANNUAL SURVEY OF PHYSICIANS

Sara L. Thran, American Medical Association, Theodore Downes-Le Guin, Griggs Anderson Research, and Sandra Berry, RAND Corporation

Sara L. Thran, American Medical Association, 515 N. State Street, Chicago, IL 60610

KEY WORDS: physician surveys, nonresponse effects, efficiency of data collection

This study examines factors influencing the probability of response and level of effort required to obtain a response in the American Medical Association's annual telephone survey of physicians' practice characteristics. In addition, completion rates and refusal rates by time of day are examined. The goal is to determine an optimal calling schedule to minimize the number of calls required and thus conserve resources to optimize response. The results of the analyses should be of interest to researchers involved in surveying physicians and other busy professionals.

DESCRIPTION OF SURVEY

The survey examined in this paper is the 1993 Socioeconomic Monitoring System (SMS) survey. The SMS is the American Medical Association's survey program for collecting information on medical practice characteristics. SMS is a series of annual telephone surveys of a representative sample of non-federal, patient care physicians excluding residents.

Each year, about 4,000 physicians respond to the survey; in the past few years, the survey response rates have averaged between 65 and 70 percent. The surveys are conducted using Computer Assisted Telephone Interviewing (CATI) and are about 25 minutes long, on average.

The survey is used to collect factual information about physicians' practices. Many of the questions on the survey are repeated from year to year. For instance, each survey includes questions on:

- basic practice characteristics such as employment status, number of physicians in the practice, number of nonphysician personnel, and involvement with managed care organizations;
- number of hours spent and patients seen in different practice setting;
- volume and fees for selected procedures; and
- income and expenses from medical practice.

In addition, about one-third of the survey each year is devoted to collecting information on special topics of interest to the Association. These topics have included: use of electronic billing, defensive medicine, self-referral, provision of charity care, physician advertising, and provision of in-office laboratory testing.

The sample for each survey is selected from the AMA Physician Masterfile, an enumeration of all physicians in the U.S. Each survey includes a panel component. About one-third of the physicians surveyed in any year responded to the survey in the previous year. The remainder of the physicians surveyed are being contacted for the first time. Physicians are surveyed no more than two consecutive years.

Data collection efforts used to maximize survey response include:

- an advance mailing to sampled physicians including a letter from AMA's Executive Vice President, an endorsement letter from the physicians's specialty society, and a brochure describing the survey and including telephone numbers to call for more information;
- numerous callbacks;
- refusal conversion attempts; and
- allowing the use of proxy respondents, designated by the physician.

SURVEY OBJECTIVES

The primary objectives of the survey program are:

- to collect trend data on key indicators of medical practice;
- to collect data for use in AMA's long term health policy research projects; and
- to collect information needed for AMA policy development and advocacy.

Trends in medical practice indicators are summarized in several AMA publications which are available for purchase. In addition, the AMA sells public use tapes approximately two years after completion of the survey; these tapes are sold to researchers at consulting firms, universities, and government organizations. AMA is now widely recognized as the source of reliable, accurate, regular data on physicians' practice characteristics.

GOALS OF THE ANALYSIS

One of the primary goals of this research is to provide guidelines to those planning physician surveys as to expectations for response rates in various subpopulations (e.g., specialty group, years of

experience, race, and census division). Having such information available may help to more accurately forecast survey costs and plan appropriate sample sizes.

In addition, the analysis of completion rates and refusal rates by time of day will be useful in improving efficiency of operations in future years of this particular survey. This information should be useful to others conducting physician surveys (or surveys of other professionals whose time is valuable).

LITERATURE REVIEW

Research on Nonresponse Effects

Although there is a clear consensus among survey researchers that high response rates are critical, the studies that have been conducted on non-response patterns in physician surveys are encouraging to researchers concerned about bias. Studies have been conducted comparing survey respondents and non-respondents and difficult/late vs. easy/early respondents to determine whether bias due to nonresponse is a measurable phenomenon.

Berk (1985) examined results from the 1977 Physician Practice Cost Survey to see whether single variable estimates would have been different if the survey period had lasted 2 months or 4 months instead of the actual 6 months. The findings suggested that the addition of late responders to the sample did not affect most estimates of key demographic variables. When Berk compared the estimates after 4 months of the field period, when the response rate was at 49%, to the final survey estimates, after a 74% response rate was reached, only 3 of the 14 estimates for key provider and practice characteristics changed more than 5%. Seven of the estimates had changed less than 1% as the response rate moved from 49% to 74%.

With the SMS survey, Thran, et al. (1986) found that means and regression coefficients were not affected by the addition of late respondents and other difficult-to-interview sample members. More recent studies of other physician surveys by Guadagnoli and Cunningham (1989) and Sobal and Ferentz (1989) similarly found that the addition of late respondents did not lead to a more representative sample.

Descriptive studies which have compared the demographic characteristics of non-responding and responding physicians generally have found that they are similar for most important demographic characteristics (Goodman and Jensen, 1981; Berk and Myers, 1980; Kaspar, 1979; and Loft, 1981).

Marder and Thran (1989) examined the impact of survey nonresponse in the SMS survey. A representative subsample of nonrespondents to the 1988 SMS survey was recontacted several months later and asked to complete an abbreviated telephone survey to collect information on key survey variables, The response rate for the subsample was raised from 69% to 86%. The results from respondents to the special survey were generally quite similar to those of the regular survey respondents. Thus, Marder and Thran concluded that additional survey efforts designed to increase the response rate did not appear to be warranted.

Research on Efficiency of Data Collection

An optimal calling schedule minimizes the number of callbacks required and conserves labor, time, and money. Previous research on this topic has been relatively sparse, and most of the previous literature on improving efficiency of survey operations has focused on surveys of household populations, rather than physicians.

Weeks, et al (1987) presented an analysis of time-of-day and day-of week effects in a large national telephone survey of adult males. First-call results were presented, as well as an analysis of second-call results on first-call no-answers. In addition, answered outcomes as well as interviewed outcomes were examined. Key findings were that, for the survey population included in this study, the chances of obtaining an answer and conducting an interview on the first call were much better on weekday evenings and on weekends than during weekday daytime hours. Moreover, this finding also applied to second calls made to first call no-answers. Time-of-day and day-of week effects on first call outcomes were generally consistent with respect to both contacting and interviewing, although Sunday had a higher interview response rate for answered calls than the other calling periods examined.

Vigderhous (1981) examined the optimal time for household telephone interviews, in terms of minimizing the probability of temporary nonresponse and in terms of maximizing the probability of completing the interview. Completion ratios by hour were examined; the completion ratio was defined as the ratio of complete dialings to incomplete dialings. A dialing was considered complete if the interviewer was able to present all the survey questions to the respondent. Incomplete dialings were those in which the respondent could not be reached or did not respond to the survey. With a few exceptions, the completion ratio was the highest between 6:00 p.m. and 7:00 p.m. over the 12 months. A log-linear technique was also used to examine day effects and day-hour interactions; separate analyses were done for each month.

METHODOLOGY

In light of the limited literature on ways to optimize

efficiency in telephone interviews of physicians, we decided to examine two factors that affect many decisions in such telephone interviews. The first factor is the schedule (time of day and day of week) under which calls are placed. The hypothesis is that certain calling periods have a higher probability of completed interviews (or other final dispositions). Since the interviewing schedule can be controlled by the researcher, knowing the optimal schedule for calling presents a clear opportunity for improving productivity in this survey -- just as it does for household surveys.

The second factor is physician characteristics, which we also hypothesized have an impact on the likelihood of obtaining a completed interview. In contrast to interview scheduling, the characteristics of potential respondents are generally not within the researcher's control (except insofar as the sample design can be changed). However, by knowing which respondents have a lower probability of completing an interview, the researcher can take proactive measures. For example, specialties which typically exhibit lower response rates can be targeted with prenotification letters or given special attention in refusal conversion. Using data from the 1993 survey, we analyzed both these factors to determine their role in determining interview outcomes.

DESCRIPTION OF DATASETS

Several different datasets were merged to build one complete file with attributes for both survey respondents and nonrespondents.

- The 1993 SMS Contact History File - this file contains summary information on the status of each physician in the survey sample. Variables include final status (complete, final refusal, partial, ineligible, etc)., total number of calls made, date of interview, and type of physician (initial vs. panel).
- AMA Physician Masterfile - this file contains basic demographic information for all physicians. Variables include specialty, Census region, year of graduation from medical school, age, etc.
- 1993 SMS Scheduler File - this file is maintained by the survey contractor and includes data on each call made in the 1993 survey. Variables include call outcome (final refusal, no answer, soft appointment, hard appointment, interview, etc.), interviewer ID, date/time/day of week, etc.

RESULTS

What predicts completion?

We examine the final disposition for the eligible initial sample and panel sample. (The initial sample consists of those physicians surveyed for the first time in 1993, while the panel sample physicians were surveyed in 1992 and reinterviewed in 1993). As expected, considerably more physicians in the initial sample than the panel sample were found to be ineligible. Among the eligible sample, the overall response rate was 64.6% -- 70.6% for the panel sample and 62.0% for the initial sample. The refusal rate, excluding ineligibles, was slightly lower for the panel cases than for the initial cases (13.2% vs 13.6%). The initial sample had considerably more cases in the max calls/field period ended category; however, this was at least partly due to release of some of the initial sample near the end of the field period.

Next, we examine demographic and practice characteristics by sample disposition. The physician characteristics examined are: sex, race, census region, years since graduation from medical school, and specialty group. We combined various specialties to derive categories reflecting differences in practice styles. For example, both surgeons and anesthesiologists have rigid schedules dictated by appointments for surgical procedures, while radiologists and pathologists tend to be hospital-based and typically do not see patients. Thus, in our analysis, specialty serves as a proxy for variations in schedule and practice style, including volume of patient contacts and location of practice.

Among the initial sample, the significant differences in response rates were found by race, specialty, and years in practice. For the eligible initial sample, response rates ranged from 55.5% for physicians in internal medicine to 78.3% for those in radiology or pathology. Physicians who have been in practice between 11 and 30 years were less likely to respond than those in earlier and later stages of their careers. Physicians whose race is unknown (i.e., they did not respond to the question in the AMA census which collects basic demographic information) were the least likely to respond to the SMS survey. In general, the same pattern of results was found for the panel cases and for the entire sample.

Next, we examined the probability of response among the eligible sample in a multivariate context, using logistic regression. Physician characteristics were included as the independent variables in a predictive model of nonresponse. Because the dependent variable as we have defined the model is dichotomous (complete/incomplete), the model was estimated using logit parameters. Based on examination of the contingency tables, the variables hypothesized to be most influential were included in the model. In separate analyses, not presented here, we also examined response rates by specialty-sex groupings and specialty-race groupings; the response rates by specialty- race groups were significantly different. Thus, those interaction terms are included in the second model presented.

Table 1 presents the model without interaction terms, including estimated coefficients for predictors. Respondent race was recoded as a variable with three values (white, non-white, and unknown), with white as the reference cell, because of the small cell sizes for the non-white categories. As described above, specialties were grouped according to practice differences, with internal medicine as the reference category. Dummy variables for the experience categories were also included. Finally, interview type was included in the model, with the reference cell being initial cases. For each independent variable, the largest absolute category is used as the reference cell.

The final model in Table 1 corresponds closely to our hypothesis and to the findings from the contingency tables. Calls to female physicians are somewhat more likely to result in completed interviews. Compared with internal medicine, all specialty groups are more likely to complete the interview. Panel cases are more likely to result in complete interviews. The main effects for race are strong, with both non-white as well as physicians whose race is unknown less likely to respond than white physicians. Physicians with less than ten or greater than thirty years of experience are significantly more likely to complete the interview than those with ten-twenty years. The Pearson chi-square for the model, which tests the hypothesis that all coefficients in the model except the intercept are equal to zero, is 228.5 (df=14, p=0.001).

We also examined a logit model not presented here, including dummy variables for race by specialty combinations. The reference cell for the race-specialty interactions is white, internal medicine. The only race-specialty group significantly less likely to respond is race unknown, internal medicine; there are, however, several combinations more likely to respond. Parameter estimates for the other independent variables are the same as in the model without interaction terms. The Pearson chi-square for the model, which tests the hypothesis that all coefficients in the model except the intercept are equal to zero, is 241.40 (df=28, p=.001).

What patterns of completion do we see?

Total number of calls made for eligible cases by final disposition were examined, but are not presented here. An interesting finding was that more calls were required for the panel cases than for the initial cases (mean number of calls to panel physicians was 16.3, compared to 14.9 for initial cases). This may have been due to timing of sample release or the longer field period for the panel sample -- during the first few weeks of the field period, interviewers were calling only the panel sample -- or due to interviewer expectations that the panel cases were more likely to complete the interview. The number of calls required to obtain a final disposition of the case ranged from 1 to 51. More calls were required to reach a final status of refusal on a case than to complete an interview, particularly among the panel cases, reflecting refusal conversion efforts.

We examined number of calls made for complete and all eligible cases by various demographic and practice characteristics. As might be expected given our earlier findings on response rate differences, physicians in internal medicine required a greater number of calls on average to complete the interview than did any other specialty group, while physicians in radiology or pathology required fewer calls. Similarly, physicians with 11-30 years experience required more calls than did those with less than 11 years or more than 30 years experience.

We also examined number of calls made and completion ratios for the eligible sample by respondent's time of day for each of the specialty groups. As defined by Vigderhous (1981), the completion ratio is the number of calls resulting in completed or partial interviews divided by the number of calls not resulting in an interview. Day of week findings are not presented, because very few differences were found. Since similar patterns were observed for the initial and panel samples, only results for the total eligible sample are presented.

Some differences in completion ratios by time of day are notable. Afternoon hours have somewhat higher completion ratios, dropping off greatly after 5 p.m. A disproportionately high number of calls seem to be made late in the day.

Optimal scheduling does vary by specialty. For instance, in certain morning hours, physicians in emergency medicine, radiology, or pathology have high completion ratios. For surgeons/anesthesiologists, completion ratios are higher in the afternoon hours. For other specialties, such as internal medicine, time of day does not appear to be a significant predictor of completion rate.

We also examined number of calls made and completion ratios by respondent's time zone and time of day. It should be noted that the telephone center was typically open for interviewing between 6 a.m. and 6 p.m. Pacific time. In the Pacific time zone, the completion ratio before 8 a.m. was startlingly high and, in general, morning hours appear to be more productive times to call. In the Eastern time zone, 8-9 a.m. was a very productive time to call (a 9% completion ratio) - although given that only 24 calls were made in this time slot, the finding may be anomalous. For all time zones except Pacific, afternoon regular business hours had slightly higher completion ratios than morning hours. Finally, a disproportionately high number of calls were

made to the Eastern time zone after regular business hours.

DISCUSSION

Apparently, efficiency could be improved somewhat by having physician specialty included in the scheduler algorithm. For example, physicians in emergency medicine, radiology, or pathology should have a higher probability of selection in the morning, while surgeons/anesthesiologists should have a higher probability of selection in the afternoon.

The most significant finding is that a disproportionately high number of calls were placed late in the day, particularly to physicians in the Eastern time zone. It appears that reducing interviewing staff late in the day could significantly improve interviewing efficiency and/or reduce survey costs. It also appears that the few staff who make calls late in the day should focus their efforts on the Pacific and Mountain time zones (except for hard appointments to the other areas).

In improving efficiency of data collection, there are other factors to be considered besides the optimal timing of calls, such as the time available for data collection, the availability of qualified interviewers, the capacity of the interviewing facility, and the need to provide coverage of nonoptimal hours in order to make appointments at the respondent's convenience and receive incoming calls. In spite of these constraints, however, it should be possible to improve efficiency by modest changes in the scheduling of calls and interviewer work shifts.

References

Berk, M. Interviewing Physicians: The effect of improved response rate. American Journal of Public Health, Vol. 75, No. 11, 1985.

Berk, M. and S. Meyers. Reasons for nonresponse on the Physicians' Practice Survey. Proceedings of the Social Statistics Section. American Statistical Association, 1980.

Goodman, L. and L. Jensen. Economic survey of medical practice: AMA's Periodic Survey of Physicians, 1966-1978. In: Proceedings of the Third Biennial Conference of Health Survey Research Methods. Rockville, MD: National Center for Health Services Research, 1981.

Guadagnoli, E. and S. Cunningham. The effects of nonresponse and late response on a survey of physician attitudes. Evaluation and the Health Professions, 12:318-328, 1989.

Kaspar, J. Problem respondents. In: Andersen, R., et al., Total Survey Error. San Francisco, CA: Josey Bass, 1979.

Marder, W. and S. Thran. Are nonrespondents different? Results of an experimental design. Proceedings of the Survey Research Section, American Statistical Association, 1989.

Sobal, J. and K. Ferentz. Comparing physician responses to the first and second mailing of a questionnaire. Evaluation and the Health Professions, 12:329-339, 1989.

Thran, S., W. Marder and R. Willke. Probability of response: A multivariate probit analysis. Proceedings of the Survey Research Section, American Statistical Association, 1986.

Vigderhous, G. Scheduling telephone interviews: A study of seasonal patterns. Public Opinion Quarterly, 45:250-259, 1981.

Weeks, M. et al. Optimal call scheduling for a telephone survey. Public Opinion Quarterly, 51:540-549, 1987.

Table 1. **Logistic Regression Results on Probability of Completing Interview**

Variable	Parameter Estimate	Pr > Chi-Square	Odds Ratio
INTERCPT	0.34	0.0001	1.40
FEMALE	0.14	0.0739	1.16
RACENWH	-0.22	0.0175	0.81
RACEUN	-0.40	0.0001	0.67
SPGP	0.14	0.1300	1.15
SPSURAN	0.27	0.0003	1.31
SPPDOB	0.36	0.0001	1.43
SPPATRA	1.10	0.0001	3.00
SPPSYCH	0.34	0.0060	1.41
SPEM	1.05	0.0001	2.86
SPOTH	0.31	0.0172	1.36
YEARPRAC	0.00	0.4404	1.00
PANEL	0.36	0.0001	1.44