Factors Influencing the Effectiveness of Mailed Health Surveys

GREGG H. GILBERT, DDS
JEFFREY LONGMATE, PhD
LAURENCE G. BRANCH, PhD

Dr. Gilbert and Dr. Longmate are with the University of Florida. Dr. Gilbert is an Assistant Professor in the College of Dentistry, Department of Oral and Maxillofacial Surgery, and the Claude D. Pepper Center for Research on Oral Health in Aging. Dr. Longmate is a Research Assistant Professor, Department of Statistics. Dr. Branch is Director of Long-term Care Research, Abt Associates, Cambridge, MA, and Professor, Department of Socio-Medical Sciences and Community Medicine, School of Medicine, Boston University.

The study was funded by grant IIR 88-026 from the Department of Veterans Affairs and by the National Institutes of Health, National Institute of Dental Research, under grant P50-DE-08845.

Tearsheet requests to Gregg H. Gilbert, DDS, University of Florida, Department of Oral and Maxillofacial Surgery, JHMHSC, Box 100416, Gainesville, FL 32610-0416; tel. (904) 392-6796; fax (904) 392-3070.

Synopsis

The authors investigated sources of bias in health surveys by examining responses to their 1989 questionnaire mailed to 1,255 Massachusetts men who were eligible for dental care provided by the Department of Veterans Affairs. After a maximum of three mailings and one telephone call to nonrespondents, a total of 1,049 veterans had responded out of 1,228 finally determined to be eligible, a response rate of 85 percent.

The investigators found that small differences in univariate estimates would have occurred had the field phase been terminated after the first mailing, which had a response rate of 61 percent. To evaluate multivariate distributions, they duplicated their previously published logistic regression model for sources of dental care, using only those who responded to the first and second mailings. Although model fits would have been substantively the same had the field phase been terminated after the first or the second mailings, analysis of parameter estimates and their statistical significances suggested bias that would have led to different substantive conclusions, in some instances.

Another potential source of bias in surveys was found to be item omission. Fifty-eight percent of respondents answered all 46 survey questions, and 90 percent answered at least 91 percent of the questions. Fewer questions were answered by those whose responses were received last, but trends regarding missing data by age or education were not statistically significant. Although the survey using this methodology met all objectives, subject nonresponses, the ineligibility of potential respondents, item nonresponses, and skewed distributions of outcome variables combined to reduce the statistical power to detect differences among groups or to alter the analysis of the differences. These factors need to be planned for by investigators undertaking similar surveys.

A HEALTH SURVEY of a group of men in Massachusetts who were eligible for Department of Veterans Affairs (VA) dental care was carried out in 1989 and the results reported (*1*). The survey strategies and the analysis of the factors involved in the survey have relevance for similar mailed surveys, particularly concerning the issue of the likelihood of bias in the sample.

The survey procedure consisted of mailed questionnaires with telephone followup to provide descriptive data on the issue of veterans' access to VA dental and medical care in general and to identify factors associated with veterans' choices of VA-sponsored care or non-VA care. The survey met the objectives of gathering a large amount of information from the population of interest, obtaining the information at a low cost compared to other surveys, providing a sample size of sufficient statistical power to detect differences between groups, and preserving the anonymity of the respondents. Bias from nonresponse was believed to be largely limited.

Many mailed surveys have initial response rates that are much lower than those observed in our survey, and consequently the potential for substantively important bias presumably would be greater. Nonresponse, ineligibility of potential respondents, item nonresponse, and unanticipated problems with

violations of logistic regression model assumptions combined to reduce statistical power to detect differences among groups or to alter the analysis of these differences.

Such factors need to be planned for by those undertaking similar surveys. We report our analysis of findings regarding response rate, missing data, and statistical power from mailed questionnaires, and how the methodologic findings can affect substantive conclusions in similar health policy investigations.

Mailed Questionnaire Bias

Gathering survey data by mailed questionnaires and telephone interviews has been advocated as a cost-effective alternative to in-person interviews. However, each strategy has sources of potential bias in common with alternative strategies, as well as unique sources of potential bias.

Response rate. Mailed surveys can be especially affected by bias resulting from low response rates. Bias can occur when different response rates are obtained among subgroups of the sample, or when nonresponse is correlated with the outcomes of interest and their explanatory correlates. For example, nonrespondents may have less interest in the subject matter of the survey than respondents, may be less educated, or may have other distinguishing characteristics that can bias results (*2–5*). In a mailed survey on oral and facial pain, Locker and Grushka (*6*) found that those who responded early and late in the response period differed in terms of sociodemographic variables and responses to items concerning pain.

Two methods of quantifying and evaluating survey nonresponse bias have been advocated. A comparison of the characteristics of respondents and nonrespondents can be made (*7–9*), but usually little or no information is known about nonrespondents. Hochstim (*10*) advocated an extrapolation technique for projecting the nonresponse bias of a study by comparing cumulative estimates from responses received in the early, middle, and late portions of the response period.

The best strategy to limit nonresponse bias is to maximize response rate. Kanuk and Berenson (*11*) classified efforts designed to increase response rate, grouping those having to do with timing, such as letters of introduction and followup efforts, and those having to do with technique, such as questionnaire length, modification of the return envelope and postage, personalized letters, the color of the survey instrument, anonymity, and monetary incentives. The success of such techniques has been reported (*11–17*).

In a telephone survey of farmers' attitudes regarding crop and livestock markets, Stinchcombe and coworkers (*9*) concluded that refusals create much more bias than inaccessibility; consequently, they suggested that more resources go into converting refusals into responses than into finding respondents who are difficult to reach. To estimate nonresponse bias accurately, they advocated distinguishing between refusal nonresponse and inaccessibility nonresponse. In the mailed survey strategy, the group analogous to the inaccessible is those for whom the investigator cannot locate the correct address; refusals are simply those whose address is correct but who do not return a completed questionnaire.

Item omission. Unanswered items can be additional sources of bias. O'Toole and coworkers (*18*) compared the quality of the data obtained from mailed questionnaire, telephone interview, and home interview techniques. Item omission was confined virtually to the mailed questionnaire technique, averaging 5.5 percent for the 84 questions assessed. Telephone interview techniques averaged 0.4 percent; home interview techniques averaged 0.2 percent. Respondents who return questionnaires with data missing may differ from those who answer all questions; consequently, imputation of missing values has been advocated. However, imputation requires a value predicted from the responses of persons or subgroups whose responses are complete (*19*). The use of predicted values assumes that relationships among measures are the same for responses with and without data missing. The presence of responses with data missing suggests that this assumption may not be appropriate (*20*).

Data quality. Differences in the reliability and validity of respondents' reports have been addressed (*21–26*). Investigators have measured how data quality may be affected by the age and health of the respondent. Andrews and Herzog (*27*) measured the quality of data obtained from 7,706 persons who responded to one of five national surveys or one local survey that were taken by telephone interview or in-person interview techniques. The older the respondents, the more the percentage of true score variance tended to decrease and the percentage of error variance tended to increase. Colsher and Wallace (*20*) interviewed noninstitutionalized Iowans in upper age groups. They found that

'The survey strategies and the analysis of the factors involved in the survey have relevance for similar mailed surveys, particularly concerning the issue of the likelihood of bias in the sample.'

the number of persons responding "don't know" increased with old age groups. They were found to be associated with poorer physical, cognitive, and psychological functioning than those whose responses were logically inconsistent. "Don't know" responses varied considerably with the topic of the question and were systematically related to health and psychobehavioral characteristics. Rodgers and Herzog (*28*) found no differences in the validity of respondents' reports by age. Similar findings were reported by Herzog and Dielman (*14*).

Methods

Sample selection and questionnaire development. The population of interest for our survey consisted of male veterans residing in Massachusetts who were eligible for VA outpatient dental care on a continuous basis and who consequently could choose between VA and non-VA dental care providers for their oral health care needs. We have described the dental care eligibility criteria (*1*). Briefly, we selected three groups of veterans eligible for such dental care. The first group was class I veterans, those who had a service-connected oral condition or disability. The second group was class IV veterans, those who received 100-percent compensation for one or more service-connected medical (nondental) disabilities. The third group was former prisoners of war (ex-POWs).

The population of interest was identified by selecting names and addresses from the March 1989 Compensation and Pension File of the VA Data Processing Center at Austin, TX. Because we wanted a sufficient sample size in each of the three classes to permit comparisons among classes, all of the class I veterans, one-half of the ex-POWs, and one-eighth of class IV veterans were selected using a random start within classes, after stratification by eligibility classification. This provided a total selected sample size of 1,255 veterans. We have described the details of the sample selection (*1*).

A 46-item survey instrument was pretested in 1989 by mail with telephone followup on two consecutive samples of 46 veterans and among a convenience sample of 19 veterans at a VA dental care facility (*1*). We calculated the reliability of the questions and measured the validity of selected responses by comparing the subjects' reports with data from VA eligibility files and dental records. We measured test-retest reliability and validity by calculating Spearman's correlation coefficients (*1*).

Mailings. After the sample selection and development of the questionnaire, 1,255 mailings were made, using bulk rate. Mailings consisted of a cover letter, the questionnaire with a self-addressed return envelope, and a self-addressed return postcard.

No identifying marks were placed on the questionnaire. We were able to identify those persons who had returned a questionnaire, but not the particular questionnaire that was returned, by asking the respondent to return the identified postcard and the unidentified questionnaire separately.

The first questionnaires returned on day 6 of the field phase. We had anticipated from 1 to 3 weeks between mailings; however, after monitoring the trend of daily returns, the second and third mailings were made on days 21 and 47. (A plot of the frequency distribution by day of return is available from the authors upon request.) After the three mailings, 10 persons were identified as dead, 15 were in a nursing home, 1 lived outside Massachusetts, and 1 was female. These 27 persons were classed as ineligible and as neither respondents nor nonrespondents.

A total of 745 persons returned a first mailing questionnaire, 218 returned a second mailing questionnaire, and 71 returned a third mailing questionnaire. Those who were identified as eligible, and who had not responded after three mailings, were called and given an abbreviated version of the questionnaire by telephone. Only 30 persons could be reached by telephone, of whom 14 reported that they had already returned a questionnaire, 15 cooperated for the telephone interview, and 1 refused. A total of 1,049 questionnaires or telephone interviews were completed. We classified a person as a respondent if he was not identified as ineligible and if he had returned a questionnaire that had at least one question answered, or if he participated in a telephone interview. A nonrespondent was any person who did not meet all the respondent or ineligible criteria.

After excluding ineligibles from the denominator, we calculated a response rate of 85 percent (1,049 completed responses ÷ 1,228 eligible recipients).

Table 1. Percent distribution of returns of 1,034 questionnaires, by selected questions and response time, in a survey of male veterans residing in Massachusetts and eligible for VA outpatient dental care, 1989

Question category or sample characteristics	Responding to first mailing (N = 745)	Responding to second mailing (N = 218)	Distribution after second mailing	Responding to third mailing (N = 71)	Distribution after third mailing
Knew they were eligible for VA dental care:					
Yes	69.4	71.7	70.0	76.9	70.4
No	17.8	16.5	17.5	14.1	17.3
Didn't know	12.8	11.7	12.5	9.0	12.3
Current source of dental care:[1]					
VA is only source	44.3	38.2	42.9	32.3	42.2
Primarily VA, sometimes private	5.3	7.4	5.8	10.3	6.1
Primarily private, sometimes VA	4.8	1.5	4.0	7.2	4.3
Don't use VA at all	18.1	22.9	19.2	32.7	20.2
Never have used VA	27.5	30.1	28.1	17.5	27.3
Current source of medical care:					
VA is only source	34.5	38.1	35.3	31.2	35.0
Primarily VA, sometimes private	25.3	19.8	24.1	31.2	24.6
Primarily private, sometimes VA	16.6	20.6	17.5	13.7	17.2
Don't use VA at all	12.4	11.0	12.1	16.0	12.4
Never have used VA	11.1	10.5	11.0	8.0	10.8
Last visit to dentist:					
Less than 6 months ago	42.5	34.1	40.6	30.4	39.8
6 months to 1 year ago	21.9	23.0	22.1	21.3	22.1
Between 1 and 2 years ago	10.8	14.5	11.7	20.9	12.4
2 to 5 years ago	12.5	12.7	12.6	11.4	12.5
More than 5 years ago	12.3	15.7	13.1	16.0	13.3
VA dental eligibility:[1,2]					
Ex-POWs	21.8 [78.4]	15.8 [16.0]	20.4	16.0 [5.7]	20.1
Class 1	13.1 [75.6]	10.2 [16.6]	12.5	13.6 [7.7]	12.6
Class 4	65.1 [70.0]	74.0 [22.6]	67.1	70.4 [7.5]	67.3
Age in years[3]	61.8 (12.7)	60.8 (16.3)	61.8 (13.3)	59.0 (16.8)	61.5 (13.9)

[1] $P < 0.001$.
[2] Percent distribution within the category is shown within brackets.
[3] Mean standard deviation is shown in parentheses.

NOTE: Chi-square test was used for categorical characteristics. Analysis of variance was used to test differences in ages by the mailing to which the veteran responded.

Among the nonrespondents were 6 persons who were hospitalized during the field phase and not followed further, 9 who returned blank questionnaires, 39 survey mailings that were undeliverable, and 1 person who refused because he had recently undergone eye surgery. The number of postcards returned from eligible persons equalled 99.5 percent of the number of returned and eligible questionnaires.

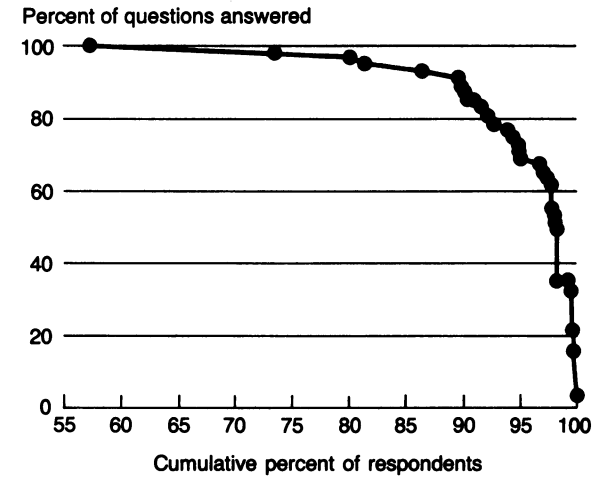
As expected, the response times, as measured by color of the questionnaire indicating the mailing responded to, and the dates of return were highly correlated (Spearman's $r = 0.77$). Thirty-one persons responded to the first mailing more than 5 days after the second mailing. Ten who responded to the second mailing returned the questionnaire more than 5 days after the third mailing. In the analyses having to do with response time that follow, we performed the same analyses using both measures. In no case did the substantive trends or conclusions regarding response time differ.

We tested several hypotheses regarding how selected univariate and multivariate characteristics of the sample varied with response time. We hypothesized that those likely to be interested in the subject matter of the survey would be more likely to respond early and that nonrespondents would differ from those who did respond. We hypothesized that later respondents would answer fewer of the items on the questionnaire. All analyses were made using the SAS microcomputer program, version 6.04 (*29*). Findings of statistical significance were based on probability values of less than 0.05. Except for instances where noted, all results presented in the remainder of this report are weighted by eligibility class because classes were sampled differentially. Weights were normalized to the mean weight to avoid inflating the sample size.

Results

Univariate characteristics by mailing. Because some reports have suggested that persons who have an interest in the subject matter of a survey are more likely to return a questionnaire early, we hypothesized that persons who were interested in either the

Percent of questions respondents answered, by cumulative distribution of respondents



VA, VA dental care, or dental care in general would likely be early responders. The distribution of these and other selected sample characteristics by response time are shown in table 1.

Because only 16 percent of the eligible persons who had not responded by the third mailing could be reached by telephone, we did not include them in table 1. If the field phase had ended after the first mailing, the estimate of the percent of sampled veterans who were aware that they were eligible for VA dental care would have been 69.4 percent. This estimate would have been 70.0 percent if the field phase had ended by the second mailing and 70.4 percent if ended after three mailings.

Trends are presented regarding the veteran's report of his current sources of dental care and medical care, the interval since his last dental visit, VA dental eligibility, and age. Although trends are evident regarding knowledge of eligibility, trends for the interval since the last dental visit and age are not statistically significant. Termination of the field phase after one mailing would have made little substantive difference regarding the estimates of the univariate distributions, as shown by comparison of the cumulative distributions across the separate mailings. These findings are consistent with negligible bias owing to nonresponse.

As discussed, another method of evaluating nonresponse is comparison of characteristics of nonresponders to responders, not just comparison of early, middle, and late responders as was done in table 1. One variable available for all persons in the sample was age, regardless of whether they actually returned a survey. The mean age of the 150 eligible persons who did not respond to any mailing or telephone followup and whose address was correct was 55.6 years (standard deviation 14.4). They were significantly younger than those who did respond (student's t-test, $t = 4.8$, $P < 0.001$).

Multivariate characteristics by mailing. We were interested in how univariate distributions would have varied if only early or middle responders had been surveyed, as well as how substantive conclusions regarding the multivariate distributions might have differed had the field phase been abbreviated. We duplicated the logistic regression modeling reported (*1*), except that we limited the analyses to first or second mailing respondents, or both. A tabular presentation of these results is available from the authors.

An abbreviation of the field phase would have had an effect on the magnitudes of all the parameter estimates in the logistic regression model. The percentage changes in the magnitude of the estimates, from the model that included only respondents to the first mailing to the model that included respondents to all three mailings, ranged from 4 to 213 percent. However, conclusions regarding the statistical significance of any single parameter estimate would have differed in only one instance. With inclusion of more respondents, the standard error of the parameter estimates decreased, and differences in the probabilities of statistical significance of the estimates were evident for four explanatory covariates in the model.

To evaluate further the effect of response time on multivariate modeling, we used four measures of model fit. We calculated a percent deviance reduction in the regression model by subtracting the deviance in the full multivariate model from the deviance in a model that contained only the intercept, and dividing this difference by the deviance in the intercept-only model. This is analogous to an R^2 value in linear regression. Other measures of model fit were sensitivity, specificity, and percent pairs that were concordant; the latter three measures of model fit were calculated from a 2 x 2 contingency table of observed versus predicted classifications. None of the four measures of model fit differed by more than 2 percent when respondents from the second or third mailings, or both, were included.

Missing data. As discussed, item nonresponse can be a source of bias. We compared the characteris-

tics of persons who answered all questions, and hence for whom there were no missing values in the logistic model, with those with at least one value missing in the explanatory covariates or the outcome of interest. Persons with at least one missing value did not differ significantly by education, marital status, or usual transportation pattern. However, persons who did have at least one missing value in the model were significantly older (mean age, 64.9 years) than those with no missing values in the covariates (mean age, 60.7 years, one-way ANOVA, $F = 13.3$, 1 df, $P < 0.001$).

We hypothesized that late responders would answer fewer of the 46 questions on the survey. Without regard to response time, item omission ranged from 0.6 percent (for a question that inquired about when the veteran served in the military) to 11.2 percent (for a question that inquired about the veteran's perception of VA dental care quality, regardless of whether the veteran had ever actually received any).

Age was not provided by 1.8 percent of respondents, as compared to 0.7 percent for marital status, 1.2 percent for race, and 0.9 percent for years of formal education. Four percent did not respond to the "interval since last dental visit" question, and 2.9 percent did not provide information on their number of remaining teeth. Although 8.7 percent did not answer a question on annual household income, only 5.0 percent did not respond to a question two items later that asked about their perception of whether their income met their expenses.

Fifty-five percent of those who did not answer the household income question also did not respond to the question about income meeting expenses, although 94 percent of those who did not answer the perceived income question also did not answer the household income question. Fifty-eight percent of respondents answered all the questions on the survey (see figure). Ninety percent of respondents answered at least 91 percent of the questions. Nine persons returned blank surveys.

Persons who responded to the first mailing answered a mean of 97 percent of the questions on the survey. Persons who responded to later mailings answered slightly fewer questions (93 percent of items were answered by responders to the second mailing, and 94 percent by responders to the third mailing). Because almost all respondents answered the questions regarding age and level of education, we calculated the mean percent of questions answered, by age and education. There were no statistically significant trends in missing data related to these two characteristics.

'Little attention has been given in the literature to distributions of missing data and their potential impact on the substantive conclusions of investigations.'

Validity of the response to the VA dental eligibility inquiry was compared by age group, education level, the mailing responded to, and eligibility class. The validity of the respondents' report for this item was very high (98.4 percent of pairs overall were concordant). Validity did not vary significantly by age group or education, but persons who responded only after three mailings did have a significantly lower validity (99.1 percent of responders to the first mailing had a valid response, compared to 99.6 percent for second mailing responders, and 95.2 percent for third mailing responders; $P < 0.01$, chi-square test). As expected, persons who were ex-POWs reported their eligibility status with the highest validity (99.1 percent), followed by persons who had a 100 percent service-connected (SC) disability (99.0 percent), followed by persons who had a SC oral condition or disability (95.3 percent).

Change in analysis plans. Our analytic goal was to model the probability of veterans' use of the VA as a source of dental care. However, a substantial percent of persons reported being unaware that the VA offered dental care to eligible veterans (24 percent of the sample), or being unaware of their eligibility for it (19 percent). Persons who were unaware could not be expected to report VA use, so we could not justify including them in the analysis. This reduced the sample size available for analysis from 1,009 (respondents who answered both questions) to 671 (respondents who were aware of VA dental care and eligibility). Although not originally planned for during our projection of the sample size required, sufficient statistical power remained.

An additional consideration was encountered during logistic modeling. The logistic regression model requires a proportional odds (parallel lines) assumption when the outcome variable is on an ordinal scale. This assumption is not required when the outcome variable is dichotomous. Under this assumption, the odds that a subject's response exceeds any given level depends on the variables in the model, but the ratio of the odds of exceeding

Table 2. Estimates of statistical power[1], by sample sizes in a survey of male veterans residing in Massachusetts and eligible for VA outpatient dental care, 1989

Sample size	*Sample*	*Percent power at α = 0.05*
1,255	Total number of survey questionnaires mailed	95
1,049	Number of questionnaires received or responses taken in telephone interview	91
855	Respondents for whom no responses were missing	85
750	Projection made before the field phase began of the size of the sample needed	80
565	Respondents for whom no responses were missing and who were aware of VA dental care eligibility	68
422	Respondents after 1 mailing for whom no responses were missing and who were aware of both VA dental care and their eligibility, if assuming equal numbers of VA (211) and non-VA users (211)	56
422	Respondents after 1 mailing for whom no responses were missing and who were aware of both VA dental care and their eligibility, using the unequal numbers of VA (134) and non-VA (288) users actually observed	51

[1] Power to detect a 10 percent difference in perception of VA dental care quality between VA users and non-VA users, accepting a type I error of α = 0.05.

any two different levels stays constant. The log odds corresponding to different splits of the response into high and low are then parallel linear functions of the explanatory variables.

In this study, the three outcomes of interest were ordinal in scale, current source of dental care, current source of medical care, and interval since last dental visit (table 1). The proportional odds assumption was violated when each of the three outcomes were analyzed in their original five-category scale. This meant that the number of categories of responses to a specific question had to be reduced by combining some categories or, in the extreme, dichotomizing them. The distribution of current source of dental care shown in table 1 suggested an obvious dichotomy (persons who reported the VA as their only or primary source of care were combined into one category, and persons whose current source was primarily or only non-VA sources were grouped). For the "interval since last dental visit" variable, the "less than 6 months" category was pooled with the "6 months to 1 year" category. The proportional odds assumption was no longer violated after these two categories were pooled.

Regarding the current source of medical care, however, neither theoretical basis nor the distribution of the categories suggested an obvious dichotomy or pooling of categories. Consequently, for this outcome, we proceeded with model development by comparing models in which dichotomous coding of the outcome variable differed. For example, we compared model fits when modeling category 1 versus a pooling of categories 4 and 5, when modeling category 1 versus a pooling of categories of 3, 4, and 5, and when modeling category 1 versus a pooling of categories of 2 through 5. The categories of responses to questions are shown in table 1. Other relevant combinations were also modelled.

The best model fit was obtained with a dichotomy that compared persons in category 1 with a pooling of persons in categories 4 and 5. However, this reduced the sample size to 568 (the sum of categories 1, 4, and 5), and the statistical power was substantially less than if we had been able to model the 977 respondents from all categories. We had not anticipated this when planning the project.

Table 2 shows how some of these considerations affected statistical power. During the grant proposal preparation phase, we projected a need for 750 respondents in order to have an 80 percent power to detect a 10 percent difference in perception of VA dental care quality between VA users and non-VA users, with an α level of 0.05. We projected conservatively that we would have a 60 percent response rate, so that our final sample size for the first mailing should be 1,250 persons (750 ÷ 0.60). With random sampling, the sample size actually was 1,255 persons. Although we had projected a deliberately conservative response rate, we did not take into account a decrease in sample size resulting from a large number of persons unaware of their eligibility (for current source of dental care), or sample size reduction due to dichotomization of the outcome variable (for current source of medical care as described in the previous paragraph). This caused a substantial reduction in power.

The power to detect a 10 percent difference in perceived quality would have been 91 percent if all 1,049 questionnaires returned had been complete.

However, only 855 respondents had no missing values in all of the explanatory covariates in the logistic regression model, and this reduced the power to 85 percent. Because a substantial number of veterans were not aware that the VA offers dental care to eligible veterans, and consequently were not included during logistic modeling, the power was further reduced to 68 percent. If the field phase had been stopped after only one mailing, statistical power would have been only 56 percent.

O'Brien (*30*) has described an approach to power analysis for log-linear models in which the contemplated analysis is performed on a table of hypothetical counts. This allows the power analysis to be as detailed as any contemplated analysis, and we used this method to calculate the power with the distribution of the outcome variable (current source of dental care) that actually occurred. The unequal distribution of 132 VA dental users and 288 non-VA dental users that actually occurred in the sample further contributed to the loss of power. Accounting for the actual distribution, and including only persons who responded to the first mailing, and who were both aware of VA dental care and their eligibility for it, reduced the power to 51 percent.

Discussion

Because the response rate for this survey was high, we were able to measure trends in response time. Although the high response rate in this study limited the potential for nonresponse bias, it did not eliminate it. We compared respondents to nonrespondents and measured trends in characteristics among responders. As is typically the case, information on nonrespondents was limited to sociodemographic information and in this survey was limited to age. The significantly younger age of nonresponders is consistent with substantive bias owing to nonresponse. However, no substantively important trends among early, middle, and late responders were observed.

Univariate estimates would have differed little if the field phase had been abbreviated, although some of the directions of the associations between characteristics and response time were actually the opposite of what we had hypothesized. For example, we had anticipated that persons who were aware of VA dental care would have been more likely to respond early, not late. Even more unanticipated was the trend observed with reporting the number of remaining teeth, in which those without teeth were more likely to be early responders, despite the expected trend observed with persons who reported a more recent dental visit being more likely to have responded early.

Although not reported previously, we did model statistically the mailing responded to with logistic regression and the date of return with linear regression, using as explanatory covariates age, education, interval since last dental visit, importance placed on dental care, source of dental care, and other variables hypothesized to be related to response time. These models accounted only for 1 percent of the variance in response time, suggesting that response time was largely random with respect to respondent characteristics, or that it could only be accounted for by variables other than those measured in this survey.

Some substantive differences were noted in multivariate modeling when only early or middle respondents were included. A third mailing required a significant effort, and attempting to reach veterans by telephone required an even greater effort. Because some differences were noted, we judged that the third mailing was worthwhile, but the effort exerted to obtain 16 additional respondents by telephone made no difference in univariate estimates, and because telephone respondents were only asked to cooperate using an abbreviated questionnaire, they could not be included during multivariate modeling.

Little attention has been given in the literature to distributions of missing data and their potential impact on the substantive conclusions of investigations. Later responders were more likely to be missing data, but trends with age and level of formal education were not significant. Missing data during multivariate modeling reduced the available sample size by approximately 15 percent. This type of reduction should be planned for when projecting the sample size needed for a project if imputation will not be done.

Another factor that substantially affected the size of the sample available for analysis (during multivariate modeling of current source of medical care) was the distribution of the outcome variable in relation to its explanatory covariates. To the extent that this can be anticipated, this factor should be planned for during sample size projection.

The strategy of this survey successfully achieved all objectives. A large amount of information was gathered from the population of interest at a relatively low cost, with a sample size that had sufficient statistical power to detect differences among groups, while preserving the anonymity of

the respondents. Bias from nonresponse was likely limited, given that no substantively important trends in early to later responders were observed.

However, the fact that nonresponders were older than responders suggests that such bias cannot be ruled out. We judged that analysis of these trends is justified for all similar mailed surveys because the analysis may indicate the likelihood of bias in the sample. Further, many mailed surveys have initial response rates that are much lower than that observed for our survey; consequently, the potential for substantively important bias would presumably be greater. Nonresponse, ineligibility of potential respondents, item nonresponse, and unanticipated problems with violation of logistic regression model assumptions combined to reduce statistical power to detect differences among groups or alter the analysis of these differences. All of these factors need to be planned for by investigators organizing similar surveys.

References

1. Gilbert, G. H., Branch, L. G., and Longmate, J.: Eligible veterans' choice between VA-covered and non-VA-covered dental care. J Public Health Dent. 52: 277-287 (1992).
2. Donald, M. N.: Implications of nonresponse for the interpretation of mail questionnaire data. Public Opinion Q 24: 99-114 (1960).
3. Warwick, D. P., and Lininger, C. A.: The sample survey: theory and practice. McGraw-Hill, Inc., New York, NY, 1975.
4. Locker, D., Wiggins, R., Sittampalam, Y., and Patrick, P.: Estimating the prevalence of disability in the community: the influence of sample design and response bias. J Epidemiol Community Health 35: 208-221 (1981).
5. Fowler, F. J., Jr.: Survey research methods. Sage Publications, Beverly Hills, CA, 1984.
6. Locker, D., and Grushka, M.: Response trends and nonresponse bias in a mail survey of oral and facial pain. J Public Health Dent 48: 20-25 (1988).
7. Barton, J., et al.: Characteristics of respondents and non-respondents to a mailed questionnaire. Am J Public Health 70: 823-825 (1980).
8. DeMaio, T. J.: Refusals: who, where, and why. Public Opinion Q 44: 223-233 (1980).
9. Stinchcombe, A. L., Jones, C., and Sheatsley, P.: Nonresponse bias for attitude questions. Public Opinion Q 45: 359-375 (1981).
10. Hochstim, J. R.: A critical comparison of three strategies of collecting data from households. J Am Stat Assoc 62: 976-989 (1967).
11. Kanuk, L., and Berenson, C.: Mail surveys and response rates: a literature review. J Marketing Res 12: 440-453 (1975).
12. Kaplan, S., and Cole, P.: Factors affecting response to postal questionnaires. Br J Prev Soc Med 24: 245-247 (1970).
13. Linsky, A. S.: Stimulating responses to mailed questionnaires: a review. Public Opinion Q 39: 82-101 (1975).
14. Herzog, A. R., and Dielman, L.: Age differences in response accuracy for factual survey questions. J Gerontol 40: 350-357 (1985).
15. Spry, V.M., et al.: Recruiting survey respondents to mailed surveys: controlled trials of incentives and prompts. Am J Epidemiol 130: 166-172 (1989).
16. Campbell, M. J., and Waters, W. E.: Does anonymity increase response rate in postal questionnaire surveys about sensitive subjects? A randomized trial. J Epidemiol Community Health 44: 75-76 (1990).
17. Rimm, E. B., et al.: Effectiveness of various mailing strategies among nonrespondents in a prospective cohort study. Am J Epidemiol 131: 1068-1071 (1990).
18. O'Toole, B. I., Battistutta, D., Long, A., and Crouch, K.: A comparison of costs and data quality of three health survey methods: mail, telephone and personal home interview. Am J Epidemiol 124: 317-328 (1986).
19. Hawkins, D. F.: Estimation of nonresponse bias. Sociol Meth Res 3: 461-488 (1975).
20. Colsher, P. L., and Wallace, R. B.: Data quality and age: health and psychobehavioral correlates of item nonresponse and inconsistent responses. J Gerontol 44: P45-P52 (1989).
21. Helöe, L. A.: Comparison of dental health data obtained from questionnaires, interviews and clinical examination. Scand J Dent Res 80: 495-499 (1972).
22. Hochstim, J. R., and Renne, K. S.: Reliability of response in a sociomedical study. Public Opinion Q 35: 69-79 (1971).
23. McFarland, S. G.: Effects of question order on survey responses. Public Opinion Q 45: 208-215 (1981).
24. Mathiowetz, N. A., and Groves, R. M.: The effects of respondent rules on health survey reports. Am J Public Health 75: 639-644 (1985).
25. Kelsey, J. L., O'Brien, L. A., Grisso, J. A., and Hoffman, S.: Issues in carrying out epidemiologic research in the elderly. Am J Epidemiol 130: 857-866 (1989).
26. Metzner, H. L., et al.: Comparison of surrogate and subject reports of dietary practices, smoking habits and weight among married couples in the Tecumseh diet methodology study. J Clin Epidemiol 42: 367-375 (1989).
27. Andrews, F. M., and Herzog, A. R.: The quality of survey data as related to age of respondent. J Am Stat Assoc 81: 403-410 (1986).
28. Rodgers, W. L., and Herzog, A. R.: Interviewing older adults: the accuracy of factual information. J Gerontol 42: 387-394 (1987).
29. SAS Institute Inc.: SAS/STAT user's guide, version 6. Ed 4. SAS Institute Inc., Cary, NC, 1989.
30. O'Brien, R. G.: Using the SAS system to perform power analysis for log-linear models. Proceedings of the 11th Annual SAS Users Group Conference, SAS Institute, Cary, NC, 1986.