Implementing Mail Survey Questionnaires

by Shane P. Desselle

Today’s health care practitioner assumes greater responsibility for improving patients’ health outcomes while keeping health care costs at bay. Increasingly greater numbers of managed care pharmacists are being placed in managerial positions requiring that they be adept at making critical decisions. A crucial component of the decision-making process is the collection, analysis, and interpretation of relevant data. The mail survey questionnaire is one of the most powerful methodological tools for gathering data. Mail surveys can be utilized to gather extremely diverse types of data, such as estimates of patient compliance to drug therapy, member preferences for prescription drug benefits design, patients’ ease of access to health care practitioners, sources of employee dissatisfaction, patients’ quality-of-life, and patterns of health care utilization, just to name a few.

Survey methodology can become quite complex for researchers engaged in theoretical research, such as in developing instrumentation to quantify previously unmeasured constructs or determining the appropriateness of various psychosocial models. Utilizing surveys in more applied research or simply to gather data as part of the decision-making process does not have to be that difficult. All too often, however, the researcher or administrator does not follow simple tenets of proper survey construction and is left with an abundance of data which he or she cannot use or, even worse, which appears to answer a question in a certain way when, in fact, it provides evidence to the contrary. It is better to base a decision on a hunch or on a lack of data than to base it on data obtained from unreliable data collection procedures. Administrators, practitioners, and researchers who are novices to survey methodology sometimes submit manuscripts to refereed journals in hopes that they will be accepted as an original research contribution. Unfortunately for the authors, the use of weak survey methodology and design frequently does not pass the muster of peer review, and they are left highly disappointed.

This is the first of two articles whose purpose is to assist practitioners, administrators, and budding researchers in constructing, implementing, and analyzing the results of survey questionnaires. This article focuses primarily on implementing surveys and is divided into several sections, including deriving the purpose and objective of the project for which the survey will be employed, some brief comments on research design and sampling, selecting the proper scaling method, pilot testing, cleansing the scale, assessing the potential for non-response bias, and...
analyzing the results. The second article (to be published in the May/June 2002 JMCP) will focus on how to actually design a survey instrument, particularly regarding the construction of items comprising an attitudinal or behavioral scale.

### Purpose of the Project or Study

Even if a project is being undertaken to obtain data to answer a specific question and not necessarily for theoretical research purposes, it is best to proceed according to basic steps of scientific inquiry to obtain usable results. If these procedures are not followed, or if the project is relegated to “something we’ll just quickly piece together,” the researcher falls prey to making the same mistakes that the layman makes every day in observing phenomena, such as overgeneralization, selective observation, inaccurate observation, illogical reasoning, and ex post facto hypothesizing (aka, “See, I told you so.”). Scientific inquiry affords the investigator critical confidence in the results of the project, no matter what methodology is used to gather the data. The scientific process involves stating and shaping the problem, generating relevant and testable hypotheses, reasoning and deduction, observation, illogical reasoning, and ex post facto hypothesizing (aka, “See, I told you so.”). Scientific inquiry affords the investigator critical confidence in the results of the project, no matter what methodology is used to gather the data. The scientific process involves stating and shaping the problem, generating relevant and testable hypotheses, reasoning and deduction, observation, and analysis of the results.

As elementary as it may appear, many researchers get hung up or make misjudgments during the very first step—stating and shaping the problem. The novice researcher fails to gain a firm grasp on why the project is being conducted in the first place. An administrator may simply state, for example, that he wants to know if beneficiaries in the plan are satisfied. If the project’s purpose is not further clarified, the administrator will end up wasting time and money on a survey that will ultimately fail to address disenrollment patterns or identify specific strengths that could be used in subsequent educational and/or marketing efforts. The next logical step involves a careful examination of the research problem or question and generating hypotheses that can be tested with the use of survey data. An example of a testable hypothesis is as follows:

**Members enrolled in our open formulary plan perceive easier access to needed medications than do members enrolled in our closed formulary plan.**

It is beyond the scope of this article to go into further detail on deriving and testing hypotheses. A continuing education article published in JMCP in 1998 adequately addresses these issues.

### Research Design and Sampling Considerations

Prior to constructing the survey, the researcher must consider the design of the project along with an adequate sampling scheme of potential respondents. The design of the study will confer varying degrees of the confidence the investigator has in the results not being due to external confounders and his or her ability to extrapolate the results beyond those who responded to the survey. It will also determine whether the results can be used to explain, correlate, or predict the occurrence of one or more phenomena. The use of random sampling mitigates many of the factors that may jeopardize the validity of the results. Random sampling is not always feasible, however, due to a lack of time and resources or the inability to obtain contact information on the entire population of interest. Convenience sampling may be appropriate so long as the researcher does not attempt to extrapolate the results beyond the sampled population. On the other hand, sometimes the research question requires more sophisticated sampling strategies, particularly if responses are desired from certain subgroups that may be small in proportion to the overall population. A good discussion of research design and sampling strategies has been published in a prior issue of JMCP.

### Selecting a Scaling Method

Many projects employing the use of surveys are interested in getting at perceptions, attitudes, or behavioral intentions of persons within the sampled population. To do so, the investigator needs to develop some sort of scale that accurately measures respondents’ reactions to the stimulus or referent object of interest as dictated by the project’s goals and objectives. It is this component of the questionnaire that involves the application of scientific techniques or psychometrics to develop one or more scales to suit the research objectives. There are several types of scaling methods available to the researcher. Of these, the Likert scale is the most commonly utilized. It is a type of summated rating scale that is relatively easy to construct. Its responses are also easy to score. Although ordinal in nature, scores typically follow a pattern of normal distribution and lend themselves well to the use of robust statistical procedures for analysis. A Likert scale is one that involves a set of item stimuli that refer to the referent object. Respondents are asked to respond to each item stimulus on a numbered scale anchored at bipolar extremes. The scale often solicits from respondents their level of “agreement” to the item stimuli. Other terms may be utilized, however, such as “satisfaction,” “preference,” or “likelihood.” Scholars in psychometrics suggest that the scale employ a 5- or 7-point format; i.e., the scale ranges from “1” to either “5” or “7.” It is important that the range of responses be properly balanced across the scale. An example would be:

- 1 = Very dissatisfied
- 2 = Dissatisfied
- 3 = Neither satisfied nor dissatisfied
- 4 = Satisfied
- 5 = Very satisfied

A common mistake made by novice researchers is to skew the scale to make more than half of the response choices “positive” or “negative” which may confuse respondents and result in their supplying unreliable answers to the item stimuli.

To score respondents’ answers to the scale, it is often just a matter of summing the responses to each item stimulus for a total score. Subtotals may be obtained for various components or domains of the scale uncovered through a factor analysis procedure (to be discussed later).

Another type of scale is the semantic differential scale. Rather than responding to a set of item stimuli on one scale, the semantic differential solicits responses to a number of different subscales.
Each is scored on a 5- or 7-point basis as well. Each subscale is anchored by a set of bipolar descriptors of the referent object. The descriptors are typically one-word adjectives or short phrases. For example, a researcher may be interested in obtaining customers’ perceptions of the staff employed at a pharmacy by having them respond to the following subscales:

- **How would you rate the friendliness of the staff?**
  - Rude
  - Courteous

- **How would you rate their ability to provide you with the services you need?**
  - Incompetent
  - Competent

- **How would you rate their desire to help you solve problems that arise?**
  - Unhelpful
  - Helpful

- **How would you rate their appearance?**
  - Unprofessional
  - Professional

A scale ranging from 1 to 5 or 1 to 7 would be placed under each set of bipolar descriptors. Respondents would be asked to circle the number that corresponds to their perceptions. The semantic differential is also a summated rating scale whose scores are determined by summing responses to each subscale.

A limitation of any summated scale is the fact that each item or subscale is assumed to have equal weight or importance in comprising the respondent's attitude toward the referent object. For example, an employer may want to gain insight into its employees' level of satisfaction on the job. A typical Likert scale might solicit respondents' level of satisfaction with various attributes of the job, such as the salary, fringe benefits, support of supervisors, and opportunity for advancement. In summing the responses to each stimulus, it is assumed that they contribute equally to the total score when, in fact, some aspects of the job may be more important than others to the employee. The differences in the importance of each stimulus is accounted for in an interval-level scale. Unlike summated scales that collect ordinal data, an interval scale incorporates the actual distance or value between each stimulus along a continuum. The methodology known to produce the most precise measurements in determining these distances is the method of paired comparisons. In this method, each item stimulus is paired against every other item stimulus in a completely randomized fashion. A panel of judges selects the stimulus that makes the most important contribution toward the referent object, NOT the one on which they most agree. These judgments are recorded in a series of matrices from which calculations are made to determine the value of each item stimulus. Once these values are determined, the scale can then be administered to the population of interest. With this format, respondents are asked to select the three item statements that are most reflective of their own beliefs. The score assigned to them is the median value associated with those item statements.

The paired comparisons technique is difficult and burdensome for the panel of judges and for the researcher. Another method of interval-level scale development is the method of equal-appearing intervals. Despite a slight improvement in ease of administration, all interval-level scaling procedures are quite complex, and this is a significant drawback to their use. An excellent resource to learn more about interval-level scale development is Allen Edwards' *Techniques of Attitude Scale Construction.*

### Constructing the Survey Instrument

Constructing the items comprising the survey instrument is probably the component of survey research in which the most egregious errors are made. Inexperienced researchers assume that once you have selected a scaling procedure, constructing the survey is simply a matter of writing down a few statements and putting some numbers next to them. Constructing the items comprising a scale is the most scientifically rigorous and problematic component of survey research. Proper survey construction employs the use of psychometric techniques and a few simple, but commonly overlooked, “dos and don’ts.”

Accompanying any mail survey questionnaire is a properly crafted cover letter. The cover letter explains to potential respondents the purpose of the research project and assures them of the confidentiality and anonymity (should that be the case) of their responses. Aside from informing survey recipients of their rights and protections as a participant in the study, the cover letter is a tool used to maximize the number of persons who complete and return the mail survey.

The second article in this two-part series will be devoted expressly to constructing a cover letter and survey questionnaire, with a thorough review of how to properly elicit demographic data and more complex attitudinal and behavioral information from respondents.

### Pilot Testing

After all of the questions have been generated, the investigator can construct an initial rough draft of the survey questionnaire. There are excellent resources that address issues such as formatting, font size, the color of paper, the use of graphic illustrations, and providing survey respondents with adequate directions for completing the survey.

Researchers often take some things for granted, particularly the reading level and clarity of the survey. This is why it is a good idea to share a draft of the survey with colleagues and with lay persons who meet the sampling criteria. Obtaining expert advice from a colleague familiar with the subject area or from known survey methodologists is commonly referred to as securing the *face validity* of the survey.

The investigator should pilot-test the initial draft of the survey instrument. While securing the face validity of the survey is important, it may not be enough to ensure that the survey instrument is as good as it can be. It is important to pilot-test the instrument with a population that is similar to the one being sampled or
Cleansing the Scale

After the surveys have been mailed and responses have been received, the job of scale development still has not yet been completed. Prior to utilizing the survey data to test the hypotheses, the investigator must take steps to ensure the validity and reliability of the scale. A statistical procedure called factor analysis utilizes the covariance existing between responses to the item stimuli to group them together into “factors” or domains. This enables the investigator to determine if items load onto the domains previously hypothesized while constructing the scale. It may also reveal that certain items did not load onto any of the domains, thus calling to question whether these items should be retained. The investigator should follow this up with an item analysis to determine the scale’s overall consistency and the adequacy of each individual item in the scale. Item analysis (and factor analysis) procedures are available on statistical software packages such as SPSS. The scale’s reliability, or the consistency of respondents in reacting to the stimuli, is determined with the calculation of a Cronbach’s alpha. Although subject to debate, a Cronbach’s alpha of 0.70 is often considered acceptable. The item analysis calculates the overall Cronbach’s alpha, but also calculates Cronbach’s alphas following the deletion of each item stimulus. If the Cronbach’s alpha is improved with the deletion of an item, the investigator may consider deleting the item prior to analysis. The decision to retain or delete items in the scale should be derived from a compilation of evidence, including the factor analysis, item-to-total correlation, and change in Cronbach’s alpha.

Where reliability addresses whether a scale measures a phenomenon consistently across multiple iterations, validity addresses whether the scale measures what it is purported to measure. Aside from the relatively simple check for face validity discussed previously, the investigator should determine the construct validity of the scale, particularly in theoretical research endeavors. Construct validity can be compartmentalized into convergent validity and discriminant validity. Responses to a scale demonstrating convergent validity should be highly correlated with responses to a scale measuring a similar construct. Even better, responses to the scale should be correlated with data obtained from another source purportedly measuring the true attitude or behavior. For example, if a researcher developed a scale measuring traits that predispose a pharmacist to making fewer dispensing errors, the responses to that scale should be correlated with actual data revealing the number of dispensing errors made by the pharmacist. On the other hand, that same scale should exhibit discriminant validity, or be unrelated to other measures. For example, the scale should not be highly correlated with the pharmacists’ grade point averages while they were in pharmacy school.

Determining the Potential for Non-Response Bias

Prior to analyzing the results, the investigator should determine the survey’s response rate. This is accomplished by dividing the number of usable responses returned by the number of surveys delivered to their proper addresses. Surveys returned by mail as undeliverable should be subtracted from the total number of surveys sent out. The investigator must assess the potential for non-response bias; that is, whether those who did not respond to the survey are in some way different from those that did respond. This can and should be approached from several angles. For one, the responding population should be compared to the entire population from which sampling occurred. There should be no statistical difference across these two groups when compared by proportions of race/ethnicity, gender, age, income level, and other personal characteristics. It is also useful to conduct a wave analysis comparing early and late responders to the survey by these same characteristics as well as by their responses to the scale used in the survey. Wave analysis is based upon the premise that late responders may be likely to respond to the survey in a manner similar to that of non-respondents. If the information is available, and if it does not violate ethical principles established in the proposal approved by the institutional review board (IRB), a final method to determine non-response bias is to telephone persons who did not respond to the survey and ask them to respond by telephone. Concerns about possible non-response bias would be allayed if their responses would be statistically equal to those obtained from mail survey respondents.

Analyzing the Results

Results to summed scale are typically reported as means for each item along with means for each domain, if applicable, and a grand scale mean. Scholars have suggested this to be appropriate despite the fact that the data are ordinal and not ratio in nature. Therefore, the use of parametric statistics, such as student t tests, one-way analyses of variance (ANOVARs), multivariate ANOVAs, and multiple regression procedures may be employed to test the research hypotheses. The investigator should consult with an appropriate expert or other resource to identify the statistical procedure(s) most appropriate. A third in a series of continuing education articles published in JMCP on conducting research addresses the use of statistical tests. A more advanced discussion of higher order statistical procedures such as factor analysis is also available.

Conclusion

In an era of increased accountability among managed care organizations and their respective providers, survey questionnaires are a powerful tool that can be utilized to obtain any of a broad array of
data. Too often, surveys are constructed haphazardly without respect to proper scientific methods, the ramifications of which are unreliable and invalid results. In this situation, the investigator is better off not conducting the survey at all. This continuing education article should help to steer administrators and researchers in the right direction when conducting survey research. For novices, it is best that they contact a more experienced researcher to help them with their initial efforts.

REFERENCES
5. Likert R. A technique for the measurement of attitudes. Arch Psychol 1932, No. 140.
Upon completion of the continuing education module, pharmacists will be able to:
1. Identify situations for which surveys may be utilized to collect data.
2. Identify important research design and sampling considerations in survey research.
3. Design an appropriate cover letter to accompany a mail survey that maximizes its rate of return.
4. Select an appropriate scaling procedure.
5. Describe methods used to assess the validity and reliability of a summated ratings scale.
6. Describe procedures used to determine the potential for non-response bias in mail surveys.

**SELF-ASSESSMENT QUESTIONS**

1. Which of the following types of data may be useful to collect in mail survey research?
   a. Health care utilization patterns of consumers
   b. Medication compliance of patients
   c. Job satisfaction of employees
   d. All of the above

2. Which of the following is NOT part of the process of scientific inquiry?
   a. Generating hypotheses
   b. Selective observation
   c. Analysis of the results
   d. Stating the problem

3. The researcher’s ability to extrapolate the results of a mail survey beyond the respondent population hinges upon:
   a. The items used to comprise the survey
   b. The scaling procedure
   c. The study’s design and sampling procedures
   d. The adequacy of the pilot test

4. A type of scaling procedure in which item stimuli are weighted or valued according to their level of contribution to the referent object is called ____________.
   a. Interval-level
   b. Semantic differential
   c. Summated rating
   d. Likert

5. Which of the following is true?
   a. A Likert scale is a type of summated ratings scale.
   b. Summated ratings scales are so called because scoring them usually involves simply adding up the respondent’s numeric choices to each item.
   c. Summated ratings scales are usually easier to construct and implement than are interval-level scales.
   d. All of the above

6. When an initial draft of a survey questionnaire is administered to a group of 50 to 100 people prior to mailing it out, this is called ____________.
   a. Pilot testing
   b. Securing face validity
   c. Sampling
   d. Preliminary analysis

7. Which of the following is a commonly employed mechanism to gauge a survey’s internal consistency?
   a. Conducting a t test
   b. Conducting a simple linear multiple regression analysis
   c. Employing a factor analysis procedure
   d. Calculating a Cronbach’s alpha

8. ____________ can be utilized to identify underlying domains of a referent object and create relevant subscales. It is also used as a tool to discern the construct validity of a scale.
   a. Cronbach’s alpha
   b. Factor analysis
   c. Item-to-total correlation
   d. Wave analysis

9. A researcher determining whether the responses to a scale correlate with other methods purported to measure the same construct is assessing the ____________ of the scale.
   a. Convergent validity
   b. Discriminant validity
   c. Face validity
   d. Reliability

10. Which of the following is true?
    a. A wave analysis compares responders to non-responders of a survey.
    b. The presence of nonresponse bias jeopardizes the researcher’s ability to draw accurate conclusions from the data.
    c. Nonresponse bias is a more imminent threat when response rates to the survey are high.
    d. All of the above
### DEMOGRAPHIC INFORMATION (not for scoring)

11. In what type of setting do you work?  
(Leave blank if none of the responses below applies.) 
- a. HMO  
- b. PPO  
- c. Indemnity insurance  
- d. Pharmacy benefits management  
- e. Other

12. Did this program achieve its educational objectives?  
- a. Yes  
- b. No

13. How many minutes did it take you to complete this program, including the quiz? (Fill in on answer sheet.)

14. Did this program provide insights relevant or practical for you or your work?  
- a. Yes  
- b. No

15. Please rate the quality of this CE article.  
- a. Excellent  
- b. Good  
- c. Fair  
- d. Poor

### Instructions

This test affords 1 hour (0.10 CEU) of continuing pharmaceutical education in all states that recognize the American Council on Pharmaceutical Education. To receive credit, you must score at least 70% of your test answers correctly. To record an answer, darken the appropriate block below. Mail your completed answer sheet to: Academy of Managed Care Pharmacy, 100 N. Pitt Street, Suite 400, Alexandria, VA 22314. If you score 70% or more, a certificate of achievement will be mailed to you within eight weeks. If you fail to achieve 70% on your first try, you will be allowed only one retake. The ACPE Provider Number for this lesson is 233-000-02-001-H04. This offer of continuing education credit expires January 30, 2003.

### Participant Identification: Please type or print.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td>8.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
<td></td>
<td>9.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td></td>
<td>10.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Social Security #:</th>
<th>Work Phone #:</th>
</tr>
</thead>
</table>

**For Identification Purposes Only**

**Name:**

<table>
<thead>
<tr>
<th>LAST</th>
<th>FIRST</th>
<th>MIDDLE</th>
</tr>
</thead>
</table>

**Company:**

**Address:**

<table>
<thead>
<tr>
<th>STREET (with Apt. No.) or PO. Box</th>
<th>CITY</th>
<th>STATE</th>
<th>ZIP</th>
</tr>
</thead>
</table>

**State & Lic. No.:**

<table>
<thead>
<tr>
<th>STATE</th>
<th>LICENSE NO.</th>
</tr>
</thead>
</table>

**Member Type:**

- ☐ Active  
- ☐ Supporting Associate  
- ☐ Student  
- ☐ Nonmember

**Signature:**

*I verify by my signature above that I have completed this examination independently.*

**Note:** If you wish you can now take this CE Exam online. Go to www.amcp.org.