This article will focus on the processes that require attention before data collection is initiated. The processes include the design of data collection tools or instruments and estimation of the number of subjects required to produce reliable and generalizable results to be used by other clinicians. The importance of designing or evaluating data collection tools during the initial stages of the development of research protocols cannot be underestimated.

As with the previous two articles in the series, this article is a continuation of the examination of the data analysis for an intervention study for subjects diagnosed with breast cancer in a support group (Coward, 2003). Coward’s study will be reviewed using the first two steps of the six-step process presented in Table 1. Issues related to identification of limitations also will be discussed, because limitations should be identified and noted throughout the research process—not just at the end. The primary objective of Coward’s study was to pilot a support group intervention that promotes self-transcendence perspectives in women diagnosed with breast cancer. The second objective was to assess whether changes in well-being would occur over time between patients participating in support groups and those not participating.

Ideally, all researchers should meet with a statistician at least three times (sometimes more often) during the study process. The first meeting is to perform a sample-size estimation, often referred to as a power analysis. The second meeting is to design a data collection tool or evaluate the strengths and limitations of using an established tool. The third meeting and subsequent meetings take place throughout the process of data summarization. Researchers should keep in mind that data analysis is a process—not a one-time analysis of data. It requires ongoing discussions between researchers and statisticians to ensure clarity and understanding of the questions being asked.

**Estimate the Sample Size**

The pilot study enrolled 41 subjects, all diagnosed with breast cancer. The study design incorporated two groups: the experimental group (support group participants, n = 22) and the control group (did not participate in support group activities, n = 17). The author did not report a power analysis or sample-size estimation, which is common. Many trials, especially pilot studies, do not have sufficient information available to calculate estimated sample sizes required when designing research projects. In fact, one of the key roles of pilot studies is to obtain preliminary information to justify the need (and expense) for larger studies.

**Assess the Appropriateness of the Design of the Data Collection Tool**

Data collection tools serve two very important roles. First, they force researchers to identify all (or almost all) of the data elements required to address primary and secondary objectives. Second, researchers and statisticians can determine how data should be collected...
for optimal statistical summarization. “Salvaging” poorly collected data sets often is difficult for statisticians (Fink, 2003a). One common example is when researchers collect the ages of subjects and then realize that birth dates should have been collected instead. Birth dates allow for the creation of additional variables. By collecting birth dates, researchers then can use the data (in the form of dates) to calculate age at diagnosis, age at completion of treatment, etc. For this reason, statisticians often recommend that data be collected in the most raw and potentially most usable form. Consultation during the development of instruments can help prevent the collection of unusable or limited data sets.

After meeting with a statistician, researchers should pilot test their data collection tools or survey instruments. Data collection tools are forms used to collect information relative to study participants (e.g., patient characteristics, laboratory values, drug doses, survival status). Survey instruments are forms used to collect information about patients’ quality of life, cognitive well-being, symptom distress, etc. Survey instruments can be presented to subjects via four formats: self-administration, interview, structured record review (data collection tool), or structured observation (Fink, 2003b). Researchers often are advised to use their tools to collect data elements for 5–10 subjects, then meet with a statistician, summarize the results together, and make appropriate changes. Teams often refine the tools, spending less time retrieving only 5–10 patients or charts rather than 200 or more.

Survey instruments typically are structured so that multiple questions are used to represent and glean information about the same objective. The primary objective also is referred to as an index or domain. For example, David Cell a provided validated instruments specific to various types of cancer such as breast cancer (Brady et al., 1997). One of the Functional Assessment of Chronic Illness Therapy surveys was developed to address quality of life in patients diagnosed with breast cancer (FACT-B) (Brady et al.). The survey has five domains (physical well-being, social and family well-being, emotional well-being, functional well-being, and additional concerns). Patients answer six items or questions using a Likert scale that represents and provides information about patients’ physical well-being in response to breast cancer. Because the scale has been validated with other patients with breast cancer, researchers can see how their subjects compare to others with the same diagnosis.

Many researchers elect to use instruments that were used previously by other researchers. An advantage of the strategy is that information about reliability and validity of the instruments may be established already. Instruments often have been revised many times and are designed to facilitate easy data collection. Often, a norm exists to which findings can be compared. For example, the FACT-B has been normed with women with stage III and IV breast cancer (Brady et al., 1997). If a researcher chooses to use the FACT-B, using it with a similar population may be most appropriate. If a researcher tries to use the instrument with women with in situ or early-stage breast cancer, the findings may not be valid.

Disadvantages to using existing instruments are that the instruments may not have been used with the populations to be studied or may not address a particular domain or construct. In most cases, researchers must obtain permission to use instruments, and sometimes a cost is involved. Authors of surveys also should provide instructions for scoring survey items.

Another option is to create specific instruments to be used in studies. Benefits and limitations exist when choosing preexisting, validated instruments. Researchers must keep in mind that the validation process of a survey is specific to the type of subjects it was initially designed to analyze. A survey also is validated and assessed for reliability specific to the primary and secondary objectives. Therefore, researchers should revalidate tools for the specific objectives and subject populations of their current studies. Creating new instruments can be time consuming. When instruments are not used with large numbers of subjects, researchers may have difficulty determining the reliability and validity of the instruments.

Reliability addresses whether information collected is repeatable or can be replicated. Five types of reliability measurement exist: test-retest, intraobserver, interobserver, alternate form, and internal consistency (see Table 2). Validity also has five types: face, content, criterion: concurrent, criterion: predictive, and construct (Litwin, 2003) (see Table 3). Validity addresses whether an instrument measures the construct or question. Basically, validity answers the question, “Am I measuring what I think?” Reliability answers the question, “Is the information gained repeatable?” If a survey instrument is proven to be valid, then it also is reliable. However, an instrument may be proven reliable but not necessarily valid.

Coward (2003) used a demographic data collection form that was modified (based on previous use) to better assess participants’ current treatment status, stage of disease, and developmental stage at diagnosis. She added additional items to identify factors preventing women from participating in support groups. The participants in the breast cancer support group study were given eight surveys to complete at three time points (see Table 4). The survey tools were assessed for reliability and validity.
Always note whether surveys have been validated for reliability to ensure that the participants understood the questions (in the way the researchers intended the questions to be understood) and that the questions addressed the stated primary and secondary objectives. One of the statistical methods used to determine internal consistency is Cronbach’s coefficient alpha. The statistic is a measurement of the strength of the internal consistency (or homogeneity) of a set of survey questions. For example, it is an assessment that measures the extent to which items included on a questionnaire focus on a particular domain (e.g., patient satisfaction, well-being).

Table 4 gives a list of the eight tools used by Coward (2003) as indicators of whether the support group intervention was successful. The table also provides the authors and creators of the study tools and instruments. Mentioning authors and creators of instruments is important and professionally considerate. Researchers should obtain permission to use instruments for their own studies. When obtaining permission, researchers also might want to take the opportunity to question authors of tools regarding specific issues they encountered.

### Table 2. Types of Reliability Measures

<table>
<thead>
<tr>
<th>TYPE OF RELIABILITY</th>
<th>PURPOSE</th>
<th>EXAMPLE</th>
<th>ISSUES OF CONCERN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternate form</td>
<td>The responses to the questions are reworded or reordered but maintain functional equivalency or continue to ask and answer the same question.</td>
<td>“Yes or no” versus “no or yes”; reword the response as follows: one to two times per day versus 12–24 hours per day.</td>
<td>Provides a mechanism of compensating for the practice effect, the idea that a rater or respondent recalls or becomes familiar with questions on a survey after repeated use (Litwin, 2003)</td>
</tr>
<tr>
<td>Internal consistency</td>
<td>To assess whether a batch or group of questions represents the same concept (e.g., emotional well-being)</td>
<td>Six questions are used to assess emotional well-being by the Functional Assessment of Chronic Illness Therapy—Breast Cancer survey.</td>
<td>Beware of using a collection of questions that do not “fit well” together.</td>
</tr>
<tr>
<td>Interobserver</td>
<td>To assess whether two or more raters agree</td>
<td>Four different nurses provide an oral assessment scale (OAS) for the same patient.</td>
<td>Observers will have to practice assessment so ratings are consistent.</td>
</tr>
<tr>
<td>Intraobserver</td>
<td>To assess whether an individual rates the same issue in a consistent manner</td>
<td>A nurse provides an OAS for the same patient (e.g., one in the morning and another in the evening).</td>
<td>Beware of the practice effect.</td>
</tr>
<tr>
<td>Test-retest</td>
<td>To measure how stable a person’s response is by giving the respondent the same survey at two different times</td>
<td>Postoperative patients measure pain using a scale two days and seven days after surgery.</td>
<td>Beware of variables that change over a short period of time.</td>
</tr>
</tbody>
</table>

### Table 3. Types of Validity Measures

<table>
<thead>
<tr>
<th>TYPE OF VALIDITY</th>
<th>PURPOSE</th>
<th>EXAMPLE</th>
<th>ISSUES OF CONCERN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct</td>
<td>To measure how meaningful a survey instrument is, usually after many years of experience</td>
<td>Results from the instrument have been reported in numerous research projects.</td>
<td>Not easily quantifiable</td>
</tr>
<tr>
<td>Content</td>
<td>To obtain an opinion from a trained individual</td>
<td>Obtain subjective opinions from social workers, oncologists, and nurses about a survey that assesses quality of life after cancer treatment.</td>
<td>Provides a subjective opinion, not an objective measurement, of the appropriateness of a survey or question</td>
</tr>
<tr>
<td>Criterion: concurrent</td>
<td>To compare a newly developed survey with a gold standard (previously validated survey)</td>
<td>Compare the results obtained from a survey developed by a researcher’s own institution with a previously validated survey measuring the same indexes.</td>
<td>A gold standard must be available to use for comparison.</td>
</tr>
<tr>
<td>Criterion: predictive</td>
<td>To find the predictable usefulness of a score, determined from a survey, with some associated patient outcome</td>
<td>Use oral assessment scale scores for prediction of pain medication doses.</td>
<td>Not recommended for longitudinal clinical studies; the time interval between survey and outcome is too long.</td>
</tr>
<tr>
<td>Face</td>
<td>To obtain an opinion of the survey from an untrained individual</td>
<td>Obtain a subjective opinion of the survey from a roommate or spouse.</td>
<td>Not considered a true measurement or assessment of validity; usually performed at a pretest stage</td>
</tr>
</tbody>
</table>
Once researchers obtain permission to use tools, they must test whether the instruments can be used reliably to assess their particular study subjects. Coward (2003) took the steps to verify that the eight previously validated instruments also were reliable for use with the 41 subjects enrolled in her study who recently were diagnosed with breast cancer. Coward carried the analysis a step further by also validating whether the instruments were consistently reliable over the course of the study (pre- and postintervention). Cronbach’s alpha statistic was used as a measurement of reliability for seven of the eight instruments. Levels of 0.7 or higher generally are accepted as representing good reliability (Litwin, 2003). Most of the reliability coefficients (except one) were higher than 0.7, suggesting that the instruments could be used reliably for measuring whether the intervention was successful for patients newly diagnosed with breast cancer in terms of quality of life. Most of the reliability coefficients were relatively high (higher than 0.8) and remained consistent for the three time periods. However, the coefficients for one of the instruments (Affect Balance Scale) tended to exhibit a decrease in reliability over the course of the study. This may have been related to the effects of treatment.

Limitations of the Study

Throughout the process, researchers should consider that no study is perfect. Identification of limitations is an ongoing process. Often during construction or selection of survey tools, trade-offs are made to balance the usefulness of tools with the feasibility of studies. Limitations can be put to good use, leading to future considerations and studies. Frequently encountered limitations include low sample size, use of surrogate measures, and a heterogenic population of study participants (e.g., in terms of diagnosis and treatment regimens). Limitations should be identified throughout the study and corrected whenever possible.

Limitations always should be documented and reported in manuscripts summarizing study results. Awareness of limitations is important to readers for many reasons. Limitations provide not only a better understanding of why resulting outcomes happened but also valuable information to people who might want to pursue and conduct similar research projects. If follow-up studies were to be conducted by other researchers, the issues, if known, might be preventable.

Limitations also give a more accurate description and understanding of the findings, especially the unexpected issues that arise during the course of study. Finally, limitations answer questions of whether study results are generalizable to other subject populations at other cancer centers. One of the unexpected limitations noted by Coward (2003) related to a potential for bias. Many of the participants selected to be in the experimental group, suggesting that they were aware of the value of support groups during treatment for breast cancer.

Summary

Parts I through III of this series of articles gave an introduction to some of the issues requiring attention when performing clinical studies and summarizing the results statistically as well as clinically. Statistical methods have been described for the identification of statistically significant differences and associations between an experimental group and a control group or over the course of time after an intervention. The next article will address the methods used for the determination of the strength or magnitude of these identified statistically significant differences.

Author Contact: Dana Oliver, MT(ASCP), MPH, can be reached at oliverda@slu.edu, with copy to editor at CJONEditor@ons.org.

References


