Methodological Issues in Doing Research Involving the Near-Death Phenomenon

To the Editor:

It is always interesting to read research conducted in other countries involving individuals in the medical field. Laura Cunico's research report, "Knowledge and Attitudes of Hospital Nurses in Italy Related to Near-Death Experiences," published in the Fall 2001 issue of the Journal, was both exceptional and extremely frustrating. It was exceptional in that she was able to get 476 nurses to complete a complex questionnaire. I know from personal experience how difficult it is to get health professionals to participate in any study, particularly those
involving the completion of a lengthy questionnaire. For this accomplish-ment alone she would be commended.

But now to my extreme frustration: Cunico’s research was a replication study utilizing a questionnaire created by Nina Thornburg in 1988 to study the knowledge and attitudes of registered nurses in intensive or cardiac care units concerning the near-death experience (NDE). Cunico’s study was intended to compare and contrast the results of a crosscultural study of nurses with those collected by Thornburg and other researchers on hospice nurses, clergy, physicians, and psychologists. The opportunity to read an empirical crosscultural study involving a large number of nurses intrigued me.

But when I discovered how she analyzed her data I was initially puzzled, then surprised, which quickly evolved into dismay. It was very apparent that Cunico had used statistical tools designed for data that have interval properties, while her data were clearly ordinal. With interval data you know the distance between specific numbers or values. For example, you know that six feet is twice the length of three feet. It can be measured and verified and is consistent no matter when or where the measurement is taken. Ordinal data only tell you that one unit is longer, taller, larger than a second and, at best, demonstrate that a rank order is present.

If you had two individuals, one who agreed with a question and the second who strongly agreed (where the alternatives were strongly agree, agree, don’t know, disagree, and strongly disagree), all you would know for sure is that their responses differed. To assign a value of, say, 4 to strongly agree and a 3 to agree implies that a known and measured difference exists. But the actual distance between the two responses is unknown. The distance could be very narrow – almost identical – or extremely wide.

In addition, to take a set of items – as was also done by Thornburg – and give each question or statement a value of one and then adding up those items the respondent checked, implies that each item makes an equal contribution to a respondent’s score. However, it could be that one or two items in the list of 29 (the number of items included in the scale that measured the attitude that the respondent has toward NDEs), are more important than all the other 27 combined. Until you evaluate the relative contribution of each and every item included in the scale, you have no idea what any item included in the scale means.

As Cunico had based her study on that of Thornburg, I decided to look at Thornburg’s study to see what she had done. I discovered that Thornburg had two major objectives. The first was to create an
instrument that could be used to study the attitudes and knowledge various groups had of the near-death experience. The second was to use the instrument to see how well it worked.

To do this she used a sample of 20 registered nurses and gave them her instrument, which consisted of 29 true/false questions and 54 Likert items. Likert items are statements that respondents are asked to indicate as to whether they strongly agree, agree, do not know, disagree, or strongly disagree. These items were then summarized into sets of scores that reflected individuals' beliefs or attitudes. As to the items that comprised her questionnaire, Thornburg noted:

The literature was reviewed in order to construct specific items for each scale in the questionnaire. The items constructed were reviewed for content validity by experts in the disciplines of nursing, sociology, and psychology. Items deemed ambitious or inappropriate were reworked or discarded. (p. 224)

Thornburg did not state how the various items were deemed ambiguous or inappropriate, other than that experts in three disciplines said so. For subsequent researchers who desire to assess the reliability and validity of the research instrument, specifics as to which items were selected and how they were evaluated should be specified. Thornburg did report that the internal consistency of the scales were evaluated using Cronbach's coefficient alpha and found to range from .81 to .84 for the three scales: knowledge of NDEs, attitudes toward near-death phenomena, and attitudes toward patient care.

While I could question the utility of some of the questions included (especially the two scales relating to near-death), the major problem with the statements utilized in Thornburg's study was the inappropriate statistical tools used to evaluate the responses. It was clear to me why Cunico used the statistical tools she did. They were the statistical tools used in the original study and allowed comparisons of her data with those collected by Thornburg. But as both Cunico and Thornburg used interval statistical tools to analyze their ordinal data, serious problems in interpretation exist.

Thornburg did attempt to evaluate the items included in her scale. She reported doing a factor analysis in which items that accounted for the majority of the variance in the respondents' knowledge of near-death experiences were identified. But to do the factor analysis she attempted, the data must be interval, and these data clearly were not. While Thornburg's efforts to create a questionnaire were innovative for the time, the statistical tools used to evaluate the data she collected were inappropriate and hence negated any conclusions drawn. When
Thornburg stated that the "25 items of the attitude toward patient care scale were factored together, and yielded three factors accounting for 53.4 percent of the variance" (p. 235), she believed that her scale accounted for all but 46.6 percent of the variance existing for what constituted attitude toward patient care. If this were in fact true, it could make a significant contribution to the care of patients.

Thornburg's study was exploratory in nature, and she and her questionnaire reflected this fact. With all the research that has been done since 1988, I expected that subsequent researchers would have made some effort to refine and improve the original instrument. Sadly, that has not been the case. For example, in 1989, Barbara Walker and Robert Russell gave Thornburg's questionnaire to 326 psychologists, from whom 117 usable questionnaires were collected. They accepted the questionnaire's validity as a given. While they did not do a factor analysis, they computed correlations and showed their levels of statistical significance, which implied that there was a strong relationship that could not have happened by chance. But the level of data they collected was not amenable to the statistics they used, thus negating any conclusions reported. One additional problem with their study is the fact that they did not state how they determined that those who participated in the study were representative of either the 326 individuals surveyed or of psychologists in general.

This study was followed by one conducted by Lori Bechtel, Alex Chen, Richard Pierce, and Walker in 1992. This time they used the clergy as their research population. From a sample of 2,722 clergy they received 320 usable questionnaires; but as Michael Perry noted in his subsequent letter to the editor, "those who returned the questionnaire may be systematically different form those who did not" (1992, p. 129). Nevertheless, Thornburg's questionnaire was used, with a few coding modifications, and assumed to be valid. As in the earlier studies, parametric statistical tools, such as means and correlation coefficients, were used to evaluate nonparametric data.

In 1994, Linda Moore used the questionnaire to assess the knowledge and attitudes of physicians. Her sample consisted of 143 staff physicians in the Baptist Memorial Hospital system. While mean scores, which are interval data, were computed from ordinal responses, the majority of the data consisted of percentages showing how the physicians in their sample responded to specific questions or statements (that is, what percent responded true or false or strongly agree, agree, and so forth). Percentages are acceptable tools as long as the researcher does not attempt to draw inferences from them. Large numbers of
percentages can be somewhat informative but conceptually weak. The reader is forced to try to make sense of a long list of percentages. As earlier researchers did not present similar data, it would be difficult to make meaningful comparisons between this study and others.

Kay Ketzenberger and Gina Keim in 2001 used Thornburg’s questionnaire to assess the knowledge and attitude of college students. Their sample size was 50 and, I suspect, being a college professor myself, that this sample was probably drawn from one of their classes. These researchers also accepted without question the validity and reliability of the questionnaire and they also inappropriately used interval statistical tools, such as means, standard deviations, variances, and correlation coefficients, to evaluate their data.

Now to Cunico’s study of Italian nurses’ knowledge and attitudes relating to near-death experiences: Cunico was well aware of the studies done to date utilizing Thornburg’s questionnaire, as all were cited in her article. This long tradition (more than 13 years at the time) associated with the questionnaire’s use would suggest that it was a well developed, accepted, and valid instrument and therefore an appropriate and desirable tool to use in a replication study. I can therefore see why Cunico felt it justified and appropriate to use. I can also see why she utilized the statistical tools she did: they were the ones used by the creator of the questionnaire and by subsequent researchers.

Following in the footsteps of all those who preceded her, she failed to recognize the implications and limitations of the data she collected. But she went even further than her predecessors and computed regression coefficients. It is very tempting to want to know how much of the existing variance in the dependent variable, knowledge of the near-death experience, can be explained by specific variables. But given the nature of the data, such conclusions cannot be made, as the most basic assumptions underlying the statistical tools utilized were not met. It is sad that all the work that went into this and the earlier studies were negated by the inappropriate ways they were analyzed. Any conclusions reported in these studies were not justified and implications for future research are negated. While Cunico’s report was filled with statistics, they were meaningless.

What this and earlier authors need to do now is to rework their data using appropriate nonparametric statistical tools appropriate to the type and level of data collected. What can happen when researchers are not careful in the selections of their statistical tools occurred in a national study done in Canada. The researchers reported that the impact of premarital cohabitation on those who subsequently married
resulted in a reduced divorce rate. Because this finding contradicted all research done to date, various researchers asked to see the research instrument, how the data were analyzed, and the raw data. They discovered that the researchers had used inappropriate statistical tools. When the data were reanalyzed using appropriate statistical tools, the results were completely reversed.

Given that 17 years have now passed since the questionnaire was created, it seems strange that no one has made any attempt to validate or update it. Has nothing changed in our understanding of the near-death experience since 1988? I am not arguing that replication studies are irrelevant. Replication studies are very useful in testing the utility of a questionnaire across radically different groups of people, especially across cultures. What I am arguing is that the type of data collected specifies the statistical tools that can be used. Before anyone else utilizes Thornburg's questionnaire, it needs to be refined to reflect the current state of near-death experience research and knowledge. It should also be designed either to collect data that are amenable to the use of parametric statistical analysis, or researchers should use available nonparametric tools.

References


Harold A. Widdison Ph.D.
Professor of Medical Sociology
Department of Sociology
Box 15300
Northern Arizona University
Flagstaff, AZ 86011
E-mail: harold.widdison@nau.edu