Analysis Issues in Outcomes Research

Sandra L. Ferketich, PhD, RN, FAAN
Joyce A. Verran, PhD, RN, FAAN

Traditional statistical methods have been applied to the analysis of outcomes data. In general, these approaches are appropriate for examining the effect of causal factors on client outcomes. However, there are a set of special issues that are pertinent to the analysis of outcomes data. The purpose of this paper is to describe some of these key issues in the application of traditional statistical techniques to outcomes analysis.

The first issue is the need to consider the complex nature of the factors influencing outcomes. Multivariate techniques are necessary to appropriately analyze the existing relationships. Relationships may involve interaction effects as well as mediator and moderator variables. Furthermore, complex models may be compromised by collinearity. A second area of discussion is analysis of data from clinical trials. Combined with this is a review of the problems that arise when random assignment is not possible. Although, clinical trials are frequently used in outcomes research, random assignment of individuals to the control and experimental groups may be difficult. In this case, there is a need to implement designs and analysis techniques that preserve the validity of causal results. One such technique is the regression discontinuity design which allows assignment of the experimental group on the basis of a cut-off score on a pertinent variable. Other procedures are available to examine groups for initial differences.

A third issue in the analysis of research data relates to the examination of rare events as health outcomes. Although such outcomes occur infrequently, they may be highly pertinent due to the high social and financial costs and/or demands on the health care system. To show a causal effect of intervention, the investigator must demonstrate a phenomenal effect of the treatment or have a large sample size. Some analysis techniques under the broad heading of survival analysis and research designs are suggested to manage the analysis of rare events.

A fourth issue that affects research with group level outcomes is the need to analyze data at a group rather than individual level even though the data may have been collected from individuals. Group level analysis requires assurance that individual data is appropriately aggregated to reflect the group opinion. Other problems with group level analysis include reduced variability and reduced sample sizes. Several strategies can be applied to reduce the possibility of incurring ecological or individualistic fallacies. Furthermore, contextual variables which may affect all members of the group may affect the individual's outcomes. Contextual regression is also discussed as an approach to managing the analysis of this type of data.

A fifth issue that plagues the researcher is incomplete data. Missing data may more strongly affect the outcomes of multivariate procedures. If data on one

variable are missing, the entire case could be lost to analysis. Some imputation strategies for dealing with this difficulty are discussed.

A number of procedures are available to analyze data from outcomes research. The major concern with such analysis is not that clinical investigators need to develop new procedures, but they should be cognizant of the advantages and limitations of the various techniques they select to use. More educational programs need to focus on outcomes research and the analysis of that data. In addition, professional support is needed for students and programs that choose to emphasize outcomes research.

Introduction

The intent of this paper is to review issues and problems in analyzing outcomes data. In point of fact, traditional statistical methods can be applied to analysis of outcomes data and the issues of analysis are the same as those in any research study. However, as we reviewed the outcomes research literature, there seemed to be a few areas that have particular importance for the outcomes researcher. Although our general topic areas are not exhaustive, they encompass some interesting analysis problems. The five areas we selected are analysis of: 1) complex models, 2) data from clinical trials, 3) rare events/outcomes, 4) aggregated data, and, 5) data sets with missing observations. Each area is reviewed and the pertinent issues are elucidated. Although we attempted to keep them separate, design issues are sometimes integrated with analysis issues. In some cases they are so integrally related to the problem that design could not be separated from analysis.

Complex Models

Much of outcomes research has focused on a limited number of outcomes and causal factors (Hegyvary, 1991). An example is Kovner's (1989) study to examine the relationship between nurse patient agreement on the importance of surgical outcomes to the actual outcomes of patient satisfaction and length of hospital stay. Kovner hypothesized that the greater the nurse patient agreement, the higher the patient satisfaction scores and the lower the length of stay. Although this research contains two outcome measures, its focus is on one causal factor. In the analysis, Kovner did include selected demographic characteristics as control variables. However, the model of primary interest contained one causal factor. Researchers and clinicians are now suggesting that health care is multivariate in nature and may not be adequately captured in a univariate model. Once this notion gains greater recognition, researchers will be forced to analyze data with multivariate techniques. Not only may the treatment have a specific effect but attribute variables and contextual variables may affect the outcome. Furthermore, each of those variables may be related to or interact with other variables in the set.

Collinearity

Collinearity, is not a new problem to behavioral science researchers. It has received attention for over twenty years (Gordon, 1968). When considering complex variable sets, it is almost impossible to select a group of variables that is expected to affect the outcome but are orthogonal to one another. In fact, if two or more causal factors are related to the outcome, the likelihood that they are related to one another or are collinear is relatively high. For example, a number of investigators have attempted to determine the effect of hospital teaching status and volume of procedures on the outcome of mortality (See for example, Flood, Scott & Ewy, 1984a, 1984b). Although there have been conflicting results, in general, hospitals with medical school affiliations and those that do more surgical procedures have lower than expected mortality rates when these rates are corrected

for severity of illness. It is also likely that hospitals affiliated with a medical school do a greater volume of surgical procedures.

Many researchers believe that collinearity is not a problem, since most statistical software programs will estimate parameters if there is not an exact linear relationship. Although this is true, the stability of the parameter estimation is in question as the value of the matrix determinant approaches zero. Furthermore, theoretically interesting variables may never enter the equation because they have a relationship to one or more variables already in the model or those that enter may appear unimportant due to reduced magnitude of effect.

Although indirect assessment of collinearity can be obtained by a visual examination of the correlation matrix of independent variables, this method falls far short of what is now readily available to researchers. In point of fact, the use of such an approach will almost always fail to find the linear composites that predict an outcome variable. In health outcomes research, if the trend is indeed to approach the study of outcome variables with very complex models, the shortcomings of detection become even more severe due to the number of potential variables in the set. Historically, other indicators of collinearity were also used. Evidence of collinearity was often claimed when hypothesized signs of beta weights were incorrect or important explanatory variables had low t-statistics. Obviously, neither of these conditions is a necessary or sufficient indicator of collinearity (Belsley, Kuh & Welsch, 1980). The presence of such indicators may in fact mean that theory is being challenged by the particular data set under examination. In summary, then the investigator cannot use such approaches to accurately assess the presence, magnitude and location of the collinearity. Fortunately, there are several newer diagnostic procedures available to the analyst.

Schroeder (1990) did a fine review of techniques readily available in popular statistical software. Her discussion of the use of the variance-inflation factor is especially helpful. One additional technique bears mention since it is now available on both SAS and SPSS output and provides extensive information to the researcher. Belsley, Kuh, and Welsch (1980) propose a method that takes into account two values to aid in determining the magnitude and location of the linear dependencies. They combine data on the condition index and the variance-decomposition proportion for each independent variable in the equation. When both of these values are of a sufficient magnitude, collinearity may be affecting parameter estimation (Belsley, Kuh, & Welsch, 1980, p. 112-113). Since these diagnostics are obtained after the regression has been performed, it is clear that the matrix could be inverted and the determinant was non-zero. The fact that the matrix was inverted does not, however, mean that the parameter estimation was stable.

Belsley, Kuh, and Welsch (1980) propose cut-off areas for both the condition index and the variance-decomposition proportion but acknowledge that these values may be somewhat sample specific. What is exciting about the use of these diagnostics is that the location of the high variance-decomposition proportions informs the researcher of the exact variables involved. This is beyond our present approaches which provide insight into the magnitude of the collinearity. Pedhazur (1982) suggests that if collinearity is present, additional regressions, that is regressing each independent variable in turn on all others, can be used to locate the exact problem. The method proposed by Belsley, Kuh, and Welsch (1980) is more efficient. We can gain extremely valuable information about the stability of the estimation of model parameters by using their method.

Once collinearity has been detected, the researcher must handle the issue of what action should be taken. There are a number of options that can be selected. One obvious approach is to drop a variable from the model. Given that all the variables are theoretically significant, an often used approach is to create a composite score of the highly related variables. This is especially appropriate when the investigator hypothesized that the shared variance is attributable to a single outside influence or latent variable. Using a factor transformation matrix to weight the individual variables permits the influence of each to be incorporated in the total score. However, the separate effect of each variable on the outcome becomes less apparent.

Another approach to dealing with the issues of collinearity is to use ridge regression. Ridge regression introduces a bias to the estimation procedure that is less severe than that which results from the bias due to collinearity. Encompassed in this method is both the process for selecting the constant that will be added to the main diagonal of the correlation matrix and the graphic estimation, ridge trace, of the effects of collinearity (Hoerl & Kennard, 1988; Schroeder, 1990).

Instead of using either of these approaches, the investigator could move directly to the specification and analysis of a latent variable model. One of the most attractive aspects of the latent variable approach is its intuitively appealing assumption that there is no one perfect measure of a construct. The construct may be measured more accurately by a set of measures that share variance.

There are additional reasons for using latent variable models. Latent variable model analysis is purported to manage measurement problems, dependencies among independent variables, and competing model testing. The advantages of the latent variable model approach have been discussed in detail in numerous publications such as Bentler (1980); James, Muliak and Brett (1982); and Loehlin (1987). The two steps in the latent variable model process, referred to as the measurement model and the causal model portions, are proposed as ways to manage the problems noted above. In point of fact, researchers who add a measurement aspect to their design, namely a factor analytic portion, as well as the analysis of a causal model, are using both steps. What is compelling about latent variable analysis software is the ease of obtaining output and certain statistical estimates. However, such estimates can be obtained through other statistics as suggested by Ferketich and Verran (1990).

If a model is properly specified, results from path analytic and latent variable modeling approaches should be similar. An example of the similarity of findings when different methods are carefully used is found with the reanalysis of the Weisman and Nathanson (1985) data. Weisman and Nathanson (1985) reported on a study of the effect of an aggregate variable of job satisfaction of nursing staff on client outcomes in 77 family planning clinics. There were two anomalous findings that Hays and White (1987) thought might have been due to the method used. They reanalyzed the data using the LISREL software. Their reanalysis strengthened the findings of the original study. We believe such examples support the need for carefully specifying and testing models as in the Weisman and Nathanson study (1987).

Although there is promise in using latent variable analysis, there are considerations that are often given inadequate attention when investigators propose using them. Latent variable modeling cannot be used just because there is a large data set available and the manifest variable model testing did not provide the desired results. Important conceptual preconditions must exist in order to use the method effectively. These preconditions relate to: 1) the design of the study, 2) the sample size, 3) the method of data collection, 4) and the vast

realm of issues pertaining to the manifest variables and their measurement (Bentler, 1990). For example, the issue of sample size constantly appears when investigators propose the use of a latent variable model. In the usual regression procedures, the number of parameters that are to be estimated is fairly easy to determine and a count of these parameters can be used as the basis for estimation of sample size. Knapp and Campbell-Heider (1989) provide a good discussion of the various estimates of sample size for regression and regression based procedures. Certainly the matter of sample size with latent variable modeling is not as easy to solve. The sample size must be sufficient for not only the beta weights that will be estimated but for the variance-covariance structure of the model. Large scale health outcomes studies should not have as much difficulty as smaller studies or small clinical trials. In this case the analysis of complex models may be beyond the data.

Since latent variable models manage large numbers of related manifest variables, they hold much promise for outcome researchers. However, Bergner (1990) states that it is unlikely that the full potential of latent variable modeling will be achieved until four conditions are met. First there needs to be a clear non-technical presentation of the method that can be understood by methodologically sophisticated but perhaps mathematically unsophisticated audiences. Second, software needs to be available that will provide diagnostic displays to allow examination of the form of relationships not just the goodness-of-fit of the model. Only in this way will the researcher know if a functional misspecification is creating the lack of fit. In regression procedures, residual displays are readily available that allow the researcher to make such assessments. Third, the software needs to be designed so that ways are available to test robustness of the model while recognizing the problems of misspecification, unreliability of measurements, and inappropriate functional forms. And, fourth, more complete elucidation of each stage of the analysis is needed so that the researcher is more thoroughly informed of what is happening during the analysis procedure itself.

We have been reviewing models whose complexity is defined by numbers of indicators and possible problems due to interdependence among independent variables. Complexity can also occur in models with relatively few variables but for which interaction, mediator, and/or moderator effects may be present.

Interaction terms, Moderators, Mediators

Interaction can be defined as the joint effects of two variables on an outcome. As Pedhazur (1982) states "...a given combination of treatments...may be particularly effective because they enhance the effects of each other, or particularly ineffective because they operate at cross purposes,..." (p. 350). Interaction effects may be apparent over the full range of the variables in question or they may be more effective in a given region. For our purposes in discussing special problems in health outcomes research, interaction in its broadest definition includes two special subcases. These are mediating and moderating effects. The latter is particularly compelling when considering variables in outcomes research since organizational, educational, and attribute variables may be considered moderators in certain models.

Moderators and mediators are third variables (as opposed to the dependent and independent variable) which alter the effects of the cause on the outcome. The moderator variable serves to partition the focal variable into subgroups over which it has maximal effectiveness in regard to the outcome variable. The mediator variable represents the means by which the focal variable affects the outcome variable (Baron & Kenny, 1986). In a visual representation of the moderator model, the moderator is placed in the same stage of the model as the independent variable and the interaction term. Thus

the main effects for the focal (independent) variable, main effects for the moderator and interaction effects are all in the same stage. However, for the mediator model, the mediator is placed between the independent and dependent variable.

"Mediator-oriented" research tends to focus on the mechanism by which the effect on the outcome variable is altered by the mediator rather than by the independent variable itself. Moderator variables, on the other hand, are often considered when there are unexpected weak or inconsistent relations from one study to another or from one population to another (Baron & Kenny, 1986). It is possible to propose very complex models in which both moderator and mediator effects are hypothesized. Baron and Kenny (1986) provide some specific approaches to analyzing complex models with mixed mediator and moderator effects. They propose that a model be developed and a series of hypotheses be tested based on the model to examine the postulated relations. Once the hypotheses have been set, the analysis is relatively straight forward based on regression analysis in which causal paths are expected to exist, disappear, or remain stable. Peters and Champoux (1979) provide another set of testing procedures for complex moderator models. Their work is based in organizational research.

An example of outcomes research which investigated disordinal interaction is provided by Szymanski and Parker (1989). These researchers used the Attitude-Treatment Interaction (ATI), also called the Trait-Treatment Interaction, approach to examine for specific areas of interaction across the variable of educational level. They found positive client outcomes were dependent upon educational level of rehabilitation counselors. However, there was an interaction of years of experience with education. Findings indicated the range of years of agency service within which master's prepared counselors would achieve better client outcomes than Bachelor's prepared counselors. The authors note that the complex nature of the relationship of level of rehabilitation counselor experience and education to client outcomes resulted in disordinal interaction of regression lines. They further suggest that the ATI design was quite helpful in accounting for such interactions. The ATI approach is discussed both by Borich (1986) and Pedhazur (1982).

Each of the techniques to investigate complex interaction models presupposes that the investigator has posed sets of hypotheses prior to the institution of the study or at a minimum a priori to the analysis. When theory is lacking regarding the effect of a moderator, post hoc analysis could be used to explore the data. Although a moderator might not have been hypothesized and data collected specifically on that moderator, the effects of the "third" variable can be assessed. If no moderator exists which is affecting the power of the model, then the effect or explanatory power of the model should be equal across all levels of the independent variables. In regression analysis an unequal effect can be examined post hoc through an assessment of the unexplained variance (residuals). The residuals can be plotted against various independent variables. The plot should show an equal band of residuals on both sides of the zero line. If there are areas of unequal spread, the conclusion is that the model is not operating the same way across different levels of the independent variable under scrutiny. One possible explanation for such unequal explanatory power is that a moderator variable is affecting the action of the independent variable on the outcome variable. Ferketich and Verran (1984) describe this process of plotting residuals.

Given the wide ranging possibilities for analyzing the effects of mediators and moderators, it is conceivable that arguments can occur over which phenomena is representative of reality. While such arguments are necessary to clarify the theoretical intention of the

researcher, the analysis is relatively straight forward. We hope that the theoretical question drives activity and not the analysis of special relations.

Clinical Trials

The clinical trial provides a powerful design to study the outcomes associated with various interventions. A critical component of the clinical trial is the use of a control group with which the responses of the treatment group or groups are compared. It is important to note that the term "control group" does not necessarily mean than randomization to treatment or control was employed. There are a number of ways to obtain controls. The two most common methods beyond randomization are the use of historical data and the use of intact groups.

Randomized Clinical Trials

The randomized clinical trial is the "gold standard" against which all other types of clinical experiment or quasi-experiment are compared (Murdaugh, 1990). Randomization insures that bias is eliminated from the assignment of treatments, that groups are balanced in covariates whether or not these variables are known and guarantees the validity of statistical tests of significance (Byar et al., 1976). In regard to statistical tests, randomization allows use of probability distributions and, hence, significance tests, for the difference in outcomes between treatment and control groups.

There are a number of ways to randomize subjects to groups and some of these techniques affect the results of analysis. For example, Ellenberg (1984) describes the modified-consent design, which allows randomization prior to obtaining informed consent. This approach eliminates the uncomfortable situation of having to obtain consent without being able to explain the treatment to the patient. The modified-consent design presents analysis problems because there will be patients who refuse to accept the randomly assigned treatment and may then be assigned to a group receiving standard therapy. If these groups are compared, using the usual analysis procedures applicable to randomized treatment assignment, the treatment effect may be diluted. If willingness to follow a particular treatment influences outcomes, elimination of subjects who refused treatment from the analysis may bias results. Therefore, all patients must be included in the analysis whether or not they initially refused the experimental treatment. The dilution of treatment effect is handled by increasing the sample size by an inflation factor calculated with the formula: $(1 - r_1 - r_2)^{-2}$, where r_1 and r_2 indicate the refusal rates for each treatment.

Even with randomization, some statistical adjustment may be necessary to correct for chance pretreatment differences in groups on possible confounding variables. It is for this purpose that Fisher first introduced the Analysis of Covariance (Kaplan & Berry, 1990). Analysis of Covariance is also useful as a method for reducing error variance thus increasing the precision of the examination of differences between groups. Of course, if some unmeasured or unknown variable influences outcomes, these factors should exert the same influence in both the treatment and control groups.

Non-randomized Clinical Trials

Some questions cannot be answered with randomized clinical trials either because randomization is difficult, not feasible or considered unethical. Difficulties with randomization include those situations where intact groups must be used for the research. Randomization would not be feasible in the situation where an insufficient number of patients or groups are available for study. Finally, randomization may be considered unethical when treatment is desperately needed by individuals.

When comparison groups are chosen by means other than randomization, the researcher must make the assumption that the groups are identical on all important variables except for the treatment or that some type of statistical correction will be made for pretreatment differences. Analysis of Covariance and multiple regression techniques are often used to make these corrections. However, they must be used with caution as these techniques cannot serve as a substitute for randomization. The objective of these statistical adjustments is to reduce bias, however, they can do nothing to solve bias due to unequal effects that unmeasured or unknown variables will have on the outcome. Nor can these techniques correct for the bias that may result due to measurement error in the covariate (Kaplan & Berry, 1990). This measurement error will tend to attenuate the adjustment for group pre-treatment differences.

Some authors (e.g. Lord, 1960; Campbell & Boruch, 1975) have proposed strategies to resolve at least the problem of measurement error in the covariate. Most of these strategies use the estimated true score of the covariate rather than the observed score with measurement error. This true score is estimated with a correction for attenuation formula using a variety of reliability coefficients.

Mishel (1990) offers a set of recommendations to reduce bias associated with non-randomized clinical trials. These include: 1) planning for multiple measures of covariates to increase reliability, 2) using multiple covariates to increase analytical power, 3) selecting covariates on the basis of previous research and theory to reduce the effect of unmeasured variables, 4) considering a time series or repeated measures design to examine discontinuities over time in the same group rather than focusing on between group comparisons, and, 5) calculating regressed gain scores for separate groups rather than across groups.

Regression Discontinuity Design

The analysis strategies for non-randomized clinical trials mentioned above are appropriate when randomization is not used because it is difficult or not feasible. They do not necessarily solve the problem when ethical considerations negate randomization. One design-analysis strategy that may be useful for these situations is the Regression Discontinuity (RD) Design. According to Trochim (1984), the RD design is appropriate when ethical questions are involved in the randomization of needy individuals to no treatment or restrictive treatment groups. Although there are several modifications of the RD design, in its simplest form it is a pretest-posttest control group approach that assigns subjects to intervention or control groups based upon a cutoff score on a variable pertinent to the investigation such as the pretest score on a patient outcome or symptomatology. The expectation of the design is that, if there is a treatment effect, the regression line for the treatment group of the post-test measure on the pre-test measure will show a discontinuity when compared to a modeled regression line had there been no treatment. For example, assume patients with high scores on a highly reliable and valid measure of pain were placed in a treatment group utilizing guided imagery while those with low pain scores were assigned to a control group. Further, assume that the cutoff score on the pain measure was 50. An assumption of the design is that there would be a strong positive correlation between pre-test and post-test scores. If no treatment were given to any patients, a regression line could be calculated along the full range of this relationship. However, a treatment will be given to those subjects with scores higher than the cut-off of 50 points. If the treatment is effective, the regression line of pre-test and post-test scores for those individuals receiving guided imagery will be offset, or lower, from the regression line which is modeled from the subjects who were in the control group.

There are several assumptions that must be met to implement an RD design. First, it must be possible to quantitatively measure a pertinent cutoff criterion and then it must be followed without exception. Second, the pre-post distribution must be able to be described by a polynomial function. Third, there must be sufficient individuals within the control group to model the hypothetical regression line had there been no treatment. Fourth, both groups of subjects must come from a continuous pre-test distribution with the division between groups determined by the cutoff criterion. Finally, the treatment must be given uniformly to all recipients in the experimental group.

The analysis of an RD study is somewhat complex and involves the determination of the appropriate polynomial function that fits the results of the study. The first step in analysis is transformation of the pre-test score so that the intercept of the regression is equal to the cut-off score. Next, the relationship is examined visually to determine whether there is any discernible discontinuity at the cutoff. It is possible that a discontinuity may be masked by variability in the data, therefore, even if a treatment effect is not visually apparent, the analysis moves to the determination of the degree of polynomial that may fit the data. This is usually done subjectively by visualizing the number of bends in the plot of the data. Higher order terms are created along with interactions between the exponential terms of the transformed pre-test score and a group membership variable. The final regression will contain the first order transformed pre-test scores, the group membership variable, the higher order polynomial terms of the transformed pre-test scores and the interactions between these terms and the grouping variable with post-test scores as the dependent variable. The model is subsequently refined by dropping non-significant polynomial terms and the appropriate interactions. The regression coefficient associated with the group membership variable is the estimate of the main effect of the program.

Although the RD design is a viable alternative when ethical considerations do not allow randomization, it is not the solution to all situations in which a randomized clinical trial is inappropriate. As Williams (1990) points out, the assumptions of the design and analysis are also its greatest limitations. The technique requires larger sample sizes than does the randomized clinical trial. In fact, Trochim (1990) indicates that as many as two and one-half times as many participants many be required as in a randomized experiment. The second major limitation is that there must be clear and distinct measures to determine a cutoff criterion. It is possible to combine several measures to form this cut-off, but they still need to be quantitatively modeled.

Even with its limitations, the RD design is a valuable addition to the techniques available to the researcher interested in examining patient outcomes. It provides a unique way to avoid the ethical dilemma of withholding treatment from those who are most in need. In addition, as Luft (1990) notes, “The RD approach also relies on the careful wedding of analysis and design, which is likely to lead to improved research” (p. 142-143).

Rare Events

Some health outcomes occur rarely. Rare events are defined as those that occur less than one time per 100 years or where the probability of occurrence in the sample is equivalent to the probability of occurrence of the event and competing events in the parent sample. The first definition can best be explained from analysis of the probability of an extremely rare event, such as a nuclear meltdown. The second definition can best be explained by events that occur with greater frequency but there are many competing events occurring concur-

rently. For example, the occurrence of cancer from exposure to a chemical used in automobile tire production (Rowe, 1986). Even though health outcomes events of interest to nursing may not be as rare or as imbedded with other events, we believe there are some useful parallels that deserve attention.

In health outcomes research two variations on the general theme of rare events are 1) a given outcome has a low probability of occurrence in any population and 2) a population is so small that the total number of cases, even of a common occurrence, is small. In each of these cases it is assumed that the occurrence of the event is costly to society. In the first case for example, the probability of the birth of a very low birth weight infant is relatively small within the overall population of births. Yet the occurrence of such an event places an extraordinary demand on resources and may create an inordinate social and financial burden on the family and society. Thus, the cost per event is high even though the probability of such an event occurring is relatively low. Such events often become the focus of outcomes research. The second class of rare events may be illustrated by the situation of teenage pregnancy. Even though teenage pregnancies have become more frequent, in small populations, such as that in a rural community, the proportion of teen pregnancies will be quite low simply because the population is so small. When health care research is aimed at decreasing the probability of the event, the researcher is faced with critical issues regarding sample size. Either the investigator will need to demonstrate a phenomenal effect on a smaller population or the sample size will need to be very large. Most often with rare events, the researcher must focus on the design of the study in order to use the most parsimonious sample necessary to answer the question.

The first approaches discussed for analysis of rare events are drawn from the analysis of events such as nuclear meltdown (Rowe, 1986). The first strategy is to model estimates to study the outcomes of similar systems. Logical conclusions are drawn about the occurrence in the system of interest. The validity of such generalizations is, of course, dependent on the confidence one has in the similarity of the systems. A second strategy is based on system structuring. The researcher analyzes the failure of component parts of the system. If a failure occurs at a critical step in a system, such as in the sequence of care events, this particular failure might be studied. Subsequent conclusions drawn could be then made to the entire system. Of course, there are problems with the degree of confidence one can have in such analyses since there may be any number of single or combined places that may be critical in the sequence of care. We recognize that this may be a different way of thinking about the delivery of health care but may hold some interesting possibilities for rare events analysis.

Retrospective Studies

There is another set of approaches that can be used to study rare events. The broad heading is that of retrospective studies which include case control methods. In the case of retrospective studies all phenomena of interest occur prior to the research. When it is difficult to predict the occurrence of an event or the event occurs infrequently, retrospective studies are very appropriate. The retrospective cohort design, also termed the historical cohort design, allows the researcher to enter subjects by past exposure to the event of interest. For example, consider a study of falls among post surgical patients. The antecedent event might be exposure to a particular ambulating procedure used by the nursing staff. The analysis would then focus on the difference in the number of falls between the groups who experienced the antecedent event and those who did not. In the case-control method the subjects are entered according to the presence or absence of the outcome variable.

In this case the interest is on the client who has experienced a fall. Since the incidence of falls is infrequent, if a random sample of patients were selected, the number of patients necessary for the analysis might be quite high. Therefore, even with all of the known problems about supporting causal links with such designs, they offer effective approaches to the analysis of rare events without the difficulties of accruing large samples.

Contingency Tables

Once the design for the study of the rare event is determined, the statistical approaches are fairly straight forward. The analysis in part depends on the outcome of interest and the number of variables considered. For example, the researcher is frequently interested in a dichotomous outcome, that is, either the event occurred or it did not occur. The examination of independence of cells in a 2 x 2 contingency table would be appropriate if only a limited number of discrete choices were available in the treatment and outcome variables. Contingency table analysis can be expanded for a three dimensional contingency table and log-linear models can be used for multidimensional contingency tables. Helpful sources for analysis of multidimensional categorical data are by Cox (1970); Bishop, Feinberg, and Holland (1975) and Feinberg (1980). Additionally, for dichotomous events, the binomial distribution can be used to test for nonrandom occurrence. For rare events, the poisson distribution can be used to approximate the binomial distribution. Again, however, the sample required for sufficient occurrence of the rare event may be large.

Survival Analysis

In clinical studies, the outcome of interest may not occur during the study period. Survival analysis encompasses a set of techniques to analyze survival data. Survival data represents time to the occurrence of a specific event, such as, length of time before death of clients under varying treatment conditions. For example, nursing might be interested in care that delays the onset of severe disability in people with Alzheimer's disease. Thus, one collects data on the time to the occurrence of a specific level of mental disability. The key distinguishing factor for survival analysis is the presence of censored data which is most often right censored. Data are said to be censored when the value of the random variable is unobserved for some of the sample. Censored data is distinct from missing data in that the unobserved event conveys information about the phenomena under study (McCool, 1982). For example, if people receiving a particular nursing protocol do not display a particular level of disability, then much information is conveyed relative to those people who do display the particular level. Right censored data relate to data from a subject who is removed from study prior to the occurrence of the event of interest. For example, a longitudinal study might end prior to the display of the behavior of subjects. Left censoring, on the other hand, occurs when death or some other failure occurs prior to the designated time.

There are two assumptions that are necessary to proceed with analysis of survival data. The first assumption is that failure mechanisms are independent across individuals over the given time interval. The second assumption is that the conditional probability of failing in time t, given events up to time t, is the same as the conditional probability of failing up to the given time t. These assumptions are stated more rigorously in sources such as Cox and Oakes (1984). A derivative of these assumptions is that any censoring mechanisms that withdraws subjects who have a particularly high or low risk of failure is prohibited. Regression-discontinuity designs are one method to manage the issues that may arise by keeping subjects in or not allowing subjects to enter the study. If survival analysis is used, there are a number of ways that the data can be modeled. The analysis,

however, is aimed at drawing inferences about the distribution of survival times. The appropriateness of the selected model can be assessed by plotting procedures (Andersen & Vaeth, 1988).

Aggregating Data

Much outcomes research deals with dependent variables that were measured and, hence, analyzed at the group level. For example, mortality and morbidity are reported as rates of occurrence for a group of individuals. In addition, in multivariate research, some causal factors may only be measurable at the group level. Examples of these factors include structural components of the organizations within which care is delivered (e.g. degree of decentralization, type of institutional ownership). When groups are being compared, the focal unit of interest is the group, and the unit of analysis should also be the group, even though the independent variables may be measured at the individual level. This dilemma of different units of analysis within the same model leads to the use of aggregated data as a representation of a group phenomena that was measured at the individual level. For example, in the previously mentioned study, Weissman and Nathanson (1985) examined the impact of a series of organizational attributes on aggregated job satisfaction and the outcomes of aggregated client satisfaction and client compliance. In addition, they examined the relationship of job satisfaction to the two outcome measures. The data were collected in 77 family practice clinics and these clinics became the sample for the analysis even though data were collected from 344 registered nurses and approximately 2,204 clients.

There are many issues involved in selecting the appropriate instruments to be used in aggregating data, but we will concentrate on analysis issues and assume that accurate and consistent instruments were used to obtain the data. Analysis issues include, 1) assuring that individual data is acceptable to aggregate to the group level, 2) dealing with radically reduced sample sizes and variance, and, 3) use of specific analysis techniques.

Data that have been aggregated to form a measure of the group itself, rather than as an average of a group of individuals should be clearly representative of the group. Further, the group should be sufficiently homogeneous that one measure reflects the group position.

Representation can be assumed if a sufficient number of group members have been measured. A response rate of 50% on a measure of autonomy would not result in an aggregated score that accurately reflects the opinion of the group. Without assurance that a sufficient proportion of the group has responded to a survey or questionnaire, the analyst should not aggregate data and represent it as an accurate measure of group behavior or opinion.

Data should be used in an aggregate form only when the variability of scores among group members is low enough that the mean reflects the general opinion. One way to test for this lack of variability is to use a form of intraclass correlation to examine whether the within group variability is less than the between group variability (Bartko, 1976; James, 1982). An intraclass correlation of .60 or higher indicates sufficient within group agreement that the aggregated measure is an appropriate representation of the group (Glick, 1985).

Since Robinson (1950) warned about the dangers of using aggregate data to study individuals, researchers have recognized that correlations with aggregate data many differ from the correlations computed with the same data at the individual level. In general, correlations of aggregate variables based on homogeneous groups are higher than their individual-level counterparts. An ecological fallacy may result when these correlations are used to infer from the group to the individual. An individualistic fallacy is also possible when

data from individuals is incorrectly used to infer to the group. One of the reasons that correlations alter with aggregation is due to the changes in the variance and covariance structure of the variables. There are usually major decreases in variability when data are aggregated. Since the majority of statistical procedures depend on a reasonable amount of variance in the phenomena of interest, severely reduced variability will bias the results of statistical analysis.

A further bias to statistical analysis is the reduction in sample size that must occur when data are aggregated. As noted in the Weissman and Nathanson (1985) study, data from 344 nurses and 2,204 patients reduced to a sample size of 77 clinics. This sample is still sufficient for the path analytic model used in the study, however, other researchers have found that using the group as the unit of analysis will severely reduce analysis choices. For example, Verran, Murdaugh, Gerber and Milton (1988) are examining the effect of an innovative practice model on patient outcomes including satisfaction, costs, and the morbidity measures of infections etc. Staff satisfaction, as measured at the individual level and aggregated to the group, is predicted to influence retention factors and patient outcomes. The latter two sets of variables were measured at the group level. Data collection on the individual level has included sample sizes of 600 and more nurses, while at the group or nursing unit level the sample is 20. Although path analytic techniques or structural equation modeling would be appropriate for this research, only bivariate correlations are reasonable with this small sample because of the bias that would result in parameters should more sophisticated multivariate techniques be used.

Rousseau (1985) recommends a series of strategies to avoid making ecological or individualistic fallacies when using aggregated data. First, variability should be assured by increasing the sample size as much as possible. While this approach may be feasible when data are being abstracted from large preexisting data bases, it may be economically unfeasible when primary data are being collected. In that situation, only those analytic techniques that are appropriate for small sample sizes should be used, but attempts should be made to collect data from groups that are dissimilar so that variability will be increased. Second, whenever possible, global measures of variables should be used since they avoid ambiguity because they are directly linked to the focal level of interest. Third, the level of analysis should be consistent with the level of the focal unit. In general, the focal unit generally corresponds to the level of measurement of the dependent variables. In the event that different units of measurement must be included in the model, the data should be maintained at the lowest measurement level possible. This requires the building of data files with both individual and aggregate level data included. Finally, the psychometric properties of instruments should be examined at the level of analysis to assure level specific reliability and validity.

Some specific analytic techniques have been recommended to assist with the examination of group level data. Langbein and Lichtman (1978) deal with analysis and grouping techniques to avoid the ecological fallacy. Many of the approaches proposed depend on how data are formed into groups.

It is also possible to use measures other than the mean of aggregated data to represent the group. For example, when there is a high degree of within group variability, the variance may be of interest rather than a measure of central tendency. It is possible that the lack of agreement within a group may have greater impact on outcomes than the magnitude of the group score on a measure.

Contextual regression (Holzemer, Jennings, Chambers & Paul, 1989) provides a systematic method for looking at the effect of contextual variables mea-

sured at the group level, on outcome variables measured at the individual level. With contextual regression, analysis is done at the individual level, but group variables are included in the regression equation. The purpose in contextual regression is to determine the presence of a contextual effect which requires that two regression equations be examined. Both equations are designed with hierarchical inclusion and stepwise entry of variables within each hierarchy. A contextual effect would be considered present if group variables account for a significant variation in the dependent variable while controlling for individual characteristics. An individual effect would be present if individual variables account for a significant variation in the dependent variable while controlling for group characteristics.

The analytic techniques mentioned have individual level data for the dependent variable. Therefore, the researcher can analyze individual scores with the inclusion of grouped variables. When the dependent variable is at the group level, usually the researcher has no choice but to analyze at that level. This requires aggregation of data with attention to the following: 1) careful delineation of the measurement and analysis levels, 2) insuring variability by including sufficient numbers of dissimilar groups, 3) choosing instruments in which items have a group referent; 4) testing the psychometric properties of instruments at the focal level, 5) testing aggregated scores for within group variability as opposed to between group variability, and 6) choosing analytic techniques appropriate for the group level sample size. In summary, although analysis at the group level with measurement at the individual level presents some unique problems, it is possible to successfully manage this type of data with attention to the items indicated above.

Incomplete Data

In many studies, missing data presents unique problems for the analysis. Missing data can occur from a variety of sources, not all of which can be prevented. Different statistical procedures present different problems when there are missing data. For example, many researchers are familiar with the management of missing data in an analysis of variance model. For example, in some cases, pooled variance estimates can be used to simplify computations. When multivariate techniques are used, the absence of a single data point may lead the investigator to discard the entire case from the sample. There are many other reasons for missing data. One of the most damaging and least well handled in analysis is when there is an unknown systematic bias operating. For example, nonresponses in survey studies present particular problems. Although some data can be collected on nonrespondants, it may not be possible to gather the type of data that will provide sufficient information to estimate the amount of bias in the analysis.

There are several procedures for handling missing data. As was mentioned above, the investigator can discard the case from the study. For some studies, especially those with small samples, discarding subjects can be hazardous. There are alternate approaches which can be used.

Little and Rubin (1988) summarize several approaches that are used under the broad classification of imputation-based procedures. The class of procedures uses various schema for filling in missing data. The data are then analyzed as if they were complete. “Hot deck” imputation is a process whereby recorded units in the sample are substituted for the missing data. “Mean imputation” is the use of means from sets of recorded values. The subjects mean on a subscale might be used to substitute for a missing value in the subscale. The third major method in the set of procedures is regres-

sion imputation where the missing value is predicted based on the known value of the data for a given unit in the sample.

Weighting procedures can also be used to substitute values for missing data. The probability of selection of any sample subset is determined. The design weights are inversely proportional to that probability. If the design weights for a particular subset of the sample is constant, subclass means can be substituted for missing data.

Lastly, the incomplete data can be explicitly modeled. The advantages to this approach are that appropriate models for the underlying process can be selected and the variance estimates are better because they can take into account the missing data (Little & Rubin, 1988). Such modeling processes require advanced preparation and a statistical consultant is in order.

Summary and Conclusions

Nurse scientists are increasingly interested in the study of outcomes of nursing care. The analysis of outcomes data, as we indicated in the introduction to this paper, does not present unique problems. Some issues, however, require greater attention because of the nature of the outcomes research questions. For example, complex theory is more likely to model the nature of the health outcomes not dependent on a single causative agent. However, the analysis of complex models frequently is made more difficult by the presence of collinearity and/or interactions among the independent variables. Some researchers will be interested in analyzing outcomes that involve censored data. Although there is significant work in the field of survival analysis and analysis of rare events, little is reported in the nursing literature. Additionally, application of analysis techniques from the field of organizational research is particularly appropriate when the focus of study is on larger units of analysis than the individual. The investigator needs the specialized knowledge about aggregating data and analysis of contextual variables when the focus of research is at the group level, such as nursing units, families, or communities. We have suggested study of regression discontinuity designs as a way to handle some of the issues involved with clinical trials. However, as we stated previously, these developments in analysis are not new but are relatively unreported in nursing research.

We did not discuss some other types of outcomes analysis such as the use of secondary data from large data sets. There are special issues such as access to the data set, compatible computer coding, transfer from one system to another, and confidentiality of the data when such transfer occurs (Agency for Health Care Policy and Research, 1991). These, however, are not strictly analysis issues. There also may be error from a variety of sources. Examples of these errors are 1) the reliability of data that were collected for another purpose, 2) the number of variables may be inadequate for model specification, and 3) the measurement of the variables may not be consistent with the theoretical framework of the secondary study. As with any other research, these specification and measurement errors result in biased parameter estimates.

Another example of the use of secondary data is meta-analytic research which combines the results from a number of smaller studies. The issues involved with meta-analysis have been well documented (see for example, Glass, McGaw & Smith, 1981; Smith & Naftel; 1984, Moody, 1990; and Cordray, 1990). Meta-analysis has the potential to improve the strength of causal inferences of interventions on outcomes. It enhances statistical power and improves generalizabilty by combining results of studies with small sizes and coding for aspects of scientific quality. However, the techniques, the strengths and limitations of meta-analysis are no different with outcomes data than with other data.

Some analytic procedures, such as the modeling of missing data, require further development. We are not proposing that most nurse scientists take part in that statistical process. What is needed, however, are nurse researchers who have a broad knowledge base of analytic procedures that are and can be used in research involving nurse sensitive patient outcomes. Without such knowledge, the researcher is highly dependent on others who may not understand the theoretical base of the research. We believe that it is critical that nurse researchers be able to lead their research team and be full participants in the process of selecting the analytic model consonant with the research question and the substantive theoretical base. Thus, some degree of sophistication about analysis of issues involved with outcomes data is necessary. Nurse scientists need to be able to explain and support their choice of the selected analysis method and integrate the results with theory. To do this, specialized education, combining the theories of outcomes analysis with research practice, is needed. More educational programs will need to focus on outcomes research and the analysis of that data. Additionally, professional support will be necessary for those students and programs choosing to emphasize outcomes research.

References

Agency for Health Care Policy and Research. (1991). Report to Congress: The feasibility of linking research-related data bases to federal and non-federal medical administrative data bases. (AHCPR Publication No. 91-0003). Rockville, Maryland: U. S. Department of Health and Human Services.

Andersen, P. K. & Vaeth, M. (1988). Survival Analysis. In S. Kotz and N. L. Johnson (Chief Eds) Encyclopedia of Statistical Sciences: Vol. IX, (pp. 119-129). New York: John Wiley & Sons.

Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51(6), 1173-1182.

Bartko, J.J. (1976). On various intraclass correlation reliability coefficients, Psychological Bulletin, 83(5), 762-763.

Belsley, D. A., Kuh, E. & Welsch, R. E. (1980). Regression diagnostics: identifying influential data and sources of collinearity. New York: John Wiley & Sons.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. Annual Review of Psychology, 31, 419-456.

Bentler, P. M. (1990). Latent variable structural models for separating specific from general effects. In Sechrest, L., Perrin, E., & Bunker, J., (eds) Research methodology: strengthening causal interpretation of nonexperimental data. Washington, D.C.: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, pp. 61-80.

Bergner, M. (1990). Latent variable structural equation modeling in health services research. In Sechrest, L., Perrin, E., & Bunker, J., (eds) Research methodology: strengthening causal interpretation of nonexperimental data. Washington, D.C.: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, pp. 81-83.

Bishop, Y. M. M., Feinberg, S. E. & Holland, P. (1975). Discrete Multivariate Analysis: Theory and Practice. Cambridge, MA: MIT press.

Borich, G. D. (1986). Trait-treatment interactions in the classroom: Researching the effects of instruction treatments on different types of learners. Schweizerische Zeitschrift fur Kaufmannisches Bilduneswesen, 5, 128-139.

Byar, D.P., Simon, R.M., Friedewald, W.T., Schlesselman, J.J., DeMets, D.L., Ellenberg, J.H., Gail, M.H., & Ware, J.H. (1976). Randomized clinical trials. The New England Journal of Medicine, 295(2), 74-80.

Campbell, D.T. & Boruch, R.F. (1975). Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations tend to underestimate effects. In C.A. Bennett, & A.A. Lumsdaine, (Eds.) Evaluation and Experience: Some Critical Issues in Assessing Social Programs. New York: Academic Press.

Cordray, D. S. (1990). Strengthening causal interpretations of nonexperimental data: The role of meta-analysis. In Sechrest, L., Perrin, E., & Bunker, J., (eds) Research methodology: strengthening causal interpretation of nonexperimental data. Washington, D.C.: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, pp. 151-172.

Cox, D. R. (1970). Analysis of Binary Data. London: Methuen.

Cox, D. R. & Oakes, D. (1984). Analysis of survival data. New York: Chapman and Hall.

Ellenberg, S.S. (1984). Randomization designs in comparative clinical trials. The New England Journal of Medicine, 310(21), 1404-1408.

Feinberg, S. E. (1980). The analysis of cross-classified categorical data. Cambridge, MA: MIT Press.

Ferketich, S. L. & Verran, J. A. (1984). Residual analysis for causal model assumptions. Western Journal of Nursing Research, 8, 41-60.

Ferketich, S. L. & Verran, J. A. (1990). Causal modeling. In Moody, L. E. (ed.) Advancing nursing science through research: Vol. II (111-144). Newbury Park, CA: Sage Publications.

Flood, A., Scott, W. R., & Ewy, W. (1984a). Does practice make perfect? Part I: The relation between hospital volume and outcomes for selected diagnostic categories. Medical Care, 22(2), 98-114.

Flood, A., Scott, W. R., & Ewy, W. (1984b). Does practice make perfect? Part II: The relation between hospital volume and outcomes and other hospital characteristics. Medical Care, 22(2), 115-124.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage Publications.

Glick, W. (1985). Conceptualizing and measuring organizational and psychological climate: Pitfalls in multilevel research. Academy of Management Review, 10(3), 601-616.

Gordon, R. A. (1968). Issues in Multiple Regression. American Journal of Sociology, 73, 592-616.

Hays, R. D. & White, K. (1987). Professional satisfaction and client outcomes: a reanalysis. Medical Care, 25(3), 259-262.

Hegyvary, S. (1991). Issues in outcomes research. Journal of Quality Assurance, 5(2), 1-6.

Hoerl, A.E. & Kennard, R.W. (1988). Ridge regression. In S. Kotz & N. L. Johnson, (Chief Eds.), Enclyclopedia of statistical sciences: Vol VIII, (pp. 129-136). New York: John Wiley & Sons.

Holzemer, W., Jennings, B.M., Chambers, D. & Paul, S. (1989). Contextual regression analysis, Nursing Research, 38(2), 124-125.

James, L. (1982). Aggregation bias in estimates of perceptual agreement. Journal of Applied Psychology, 67(2), 219-229.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). Causal analysis: Assumptions, models, and data. Beverly Hills: Sage.

Kaplan, R.M. & Berry, C.C. (1990). Adjusting for confounding variables. In L. Sechrest, E. Perrin, & J. Bunker (Eds.). Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data (pp. 105-114). Washington, D.C.: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research.

Knapp, T. R. & Campbell-Heider, N. (1989). Numbers of observations and variables in multivariate analysis. Western Journal of Nursing Research, 11(5), 634-641.

Kovner, C. T. (1989). Nurse-patient agreement and outcomes after surgery. Western Journal of Nursing Research, 11(1), 7-19.

Langbein, L. & Lichtman, A. (1978). Ecological Inference. Sage: Beverly Hills.

Little, R.J. & Rubin, D.B. (1983). Incomplete data. In S. Kotz & N. L. Johnson, (Chief Eds.), Enclyclopedia of statistical sciences: Vol IV, (pp. 46-53). New York: John Wiley & Sons.

Loehlin, J. C. (1987). Latent Variable Models. Hillsdale, NJ: Lawrence Erlbaum.

Lord, F.M. (1960). Large-sample covariance analysis when the control variable is fallible. Journal of the American Statistical Association, 68, 304-305.

Luft, H.S. (1990). The applicability of the regression-discontinuity design in health services research. In L. Sechrest, E. Perrin, & J. Bunker, (Eds.). Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data (pp. 141-143). Washington, D.C.: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research.

McCool, J. I. (1982). Censored Data. In S. Kotz & N. L. Johnson, (Chief Eds.), Enclyclopedia of statistical sciences: Vol I, (pp. 389-396). New York: John Wiley & Sons.

Mishel, M.H. (1990). Confounding variables. In L. Sechrest, E. Perrin, & J. Bunker, (Eds.). Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data (pp. 115-117). Washington,

D.C.: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research.

Moody, L.E. (1990). Meta-analysis: Qualitative and quantitative methods. In Moody, L. E. (ed.) Advancing nursing science through research (70-110). Newbury Park, CA: Sage Publications.

Murdaugh, C.M. (1990). Maximizing the independent variable. Paper presented at Western Society for Research in Nursing Meeting. Denver, CO.

Pedhazur, E. J. (1982). Multiple regression in behavioral research. New York: Holt, Rinehart and Winston.

Peters, W. S. & Champoux, J. E. (1979). The role and analysis of moderator variables in organizational research in R. T. Mowday and R. M. Steers. Research in organizations (pp. 239-253). Santa Monica: Goodyear.

Robinson, W. (1950). Ecological correlations and the behavior of individuals. American Sociological Review, 15, 351-387.

Rousseau, D. (1985). Issues of level in organization research; multilevel and cross level perspectives. Research in Organization Behavior, 7, 1-37.

Rowe, W. D. (1986). Rare-event risk analysis, In S. Kotz & N. L. Johnson (Chief Eds.). Encyclopedia of statistical sciences: Vol. VII (pp 617-623). New York: John Wiley & Sons.

)Schroeder, M. A., (1990). Diagnosing and dealing with multicollinearity. Western Journal of Nursing Research, 11(1), 175-184.

Smith, M.C. & Naftel, D.C. (1984). Meta-analysis: A perspective for research synthesis. Image, 16, 9-13.

Szymanski, E. M. & Parker, R. M. (1989). Relationship of rehabilitation client outcome to level of rehabilitation counselor education. Journal of Rehabilitation. 55(4), 32-36.

Trochim, W.M.K. (1984). Research design for program evaluation, Beverly Hills, CA: Sage Publications.

Trochim, W.M.K. (1990). The regression-discontinuity design. In L. Sechrest, E. Perrin, & J. Bunker, (Eds.). Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data (pp. 119-139). Washington, D.C.: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research.

Verran, J.A., Murdaugh, C.L., Gerber, R.M. & Milton, D.A. (1988). Differentiated Group Professional Practice in Nursing. Grant No. 1 U01 NR02153, funded by National Center for Nursing Research and Division of Nursing.

Weissman, C.S. & Nathanson, C.A. (1985). Professional satisfaction and client outcomes. Medical Care, 23(10), 1179-1192.

Williams, S.V. (1990). Regression-discontinuity design in health evaluation. In L. Sechrest, E. Perrin, & J. Bunker, (Eds.). Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data (pp. 145-149). Washington, D.C.: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research.