NURSE EDUCATOR
Volume 26, Number 3, pp 136–141

Mastering Research Critique and Statistical Interpretation

Guidelines and Golden Rules

Janet Hoey Robinson, PhD, RN

Mastery of statistical analysis research critique is an important skill for professional nurses. **A Guideline for Statistical Analysis** ***and*** **Golden Rules for Statistical Analysis Adequacy** ***are presented and applied to classroom use. Students who learn how to critique research and statistics usage effectively are satisfied consumers and report using knowledge in other clinical courses. Effective strategies to teach statistical analysis critique are discussed.***

Critiquing quantitative research findings is an essential skill for baccalaureate-prepared nurses. Most Bachelor of Science in Nursing (BSN) curricula require introductory statistics and nursing research courses. The nursing research course must build on prior statistical understanding, while faculty who teach research often do not cover statistical analysis, since students have already had a separate course. This leaves the integration of statistical research understanding to students who are often not prepared for this level of synthesis.

Nursing students often undervalue statistical knowledge and frequently report fear of failure before a statistics class. Building on this prior analytic knowledge base is difficult when there is continued group perception of inadequate statistical understanding. Undergraduate students consistently report concern that their statistics course work has little relevance to the nursing program. Nursing educators need to use students' perceptions about statistics courses in planning curricula.

The content, process, and sequence of statistics and research courses are important curricular discussions. The increased emphasis on interdisciplinary learning and integration of understanding across disciplines prompts questions about the most effective ways to teach statistics. Stranahan[1] reported a study on student outcomes (final grades) in statistics and nursing research courses when sequence was varied within the same institution. Findings suggested students who took statistics concurrently with nursing research did better than the students who took statistics prior to the research course or the students who took statistics after the research course. All students had the same faculty in both courses, controlling for differences in teaching methods as a variable. These outcomes suggest that the assumption and common practice of statistics being a prerequisite for research be questioned. It also points to an advantage for immediate integration of statistical knowledge with research content. Application of statistical concepts in nursing research increases student understanding.

In an attempt to integrate nursing and statistical understanding, nursing and mathematics faculty at Youngstown State University developed an interdisciplinary 5-hour course in which abstract mathematical concepts and skills were immediately applied to the discipline of nursing.[2] Faculty observed greater increase in student retention and understanding of abstract concepts. Student and faculty dissatisfaction with the required statistics course at the University of Victoria also prompted a multidisciplinary effort.[3] An experiential quantitative and qualitative analysis course was developed and implemented for undergraduate nursing, social work and child care majors that could be taken on campus or via distance education. This analysis course was in addition to a nursing inquiry and research course.

These interdisciplinary efforts documented in the literature point to continued efforts to create effective teaching strategies that increase understanding and evaluation of statistical skills in critiquing healthcare research. Though more creative integration of statistics and nursing research courses is possible, most baccalaureate programs require a discrete statistics course because of requisites for admission to Master of Science in Nursing (MSN) programs. Therefore, faculty need to ensure that statistical knowledge is integrated with other critiquing skills in their research courses.

For statistical analysis and utilization, critical thinking and mathematical aptitudes are essential for success, as are effective teaching strategies that foster understanding of complex concepts and processes. A paucity of nursing educational literature is available to assist faculty with course preparation to teach statistical analysis skills. Reported strategies to facilitate greater understanding of statistical analysis in the classroom include creation of group games[4,5] and the use of computer technologies.[2–4,6] Beitz[7] advocates a metacognitive approach in helping students learn and apply statistical analysis. Her cogent analysis of multiple parametric and nonparametric statistics, with definitions and application, includes a useful table for

Janet Hoey Robinson, PhD, RN, Professor, Associate Dean of the Graduate Program, jrobinson@mco.edu, School of Nursing, Medical College

students and professional nurses. Beitz and Wolf[4] describe a variety of creative strategies useful in promoting optimal learning and teaching critical concepts. Their pragmatic approaches attempt to influence student attitudes by decreasing their anxiety, gaining their attention, and increasing active participation.

The purpose of this article is to make the statistical analysis part of the critique process for quantitative research studies more concrete. Presented is a guideline developed, edited, and used for several years by the author to assist students with critiquing research articles. The guideline format lists the variables, postulated relationships between variables, sample characteristics, and statistics. Students are also taught how to apply the *Golden Rules for Statistical Analysis Adequacy* to data listed on the guideline. After describing the teaching strategy of using a guideline and rules, a research study is critiqued illustrating the application of both processes. Classroom implications are discussed and evaluated.

Teaching Strategy

Directives for the *Guidelines for Statistical Analysis* (Figure 1) include:

1. List all variables in the study under review;
2. Categorize each variable (independent, dependent, intervening, predictor, outcome, correlational);
3. Find operational definitions of each variable;
4. Identify level of measurement for each variable;
5. List descriptive statistics for each variable;
6. List reliability measures for each variable (from the literature and for this study);
7. List kinds of validity described;
8. Report sample size, power analysis, and descriptive statistics of the sample;
9. Draw the relationships between variables; and
10. Identify the descriptive and inferential statistics (parametric and nonparametric) reported in the study with the level of significance.

Students learn theoretical concepts (e.g., type of variable or level of measurement) from research texts and classes and then apply concepts to research literature. Students also develop skills in identifying key items from quantitative research reports. The guideline is part of the total critique format for the course; it gives students the "big picture." The guideline has been used and positively evaluated in nursing research courses by several different faculty.

Each week in class, students work with their instructor to review published research articles selected by the instructor. Students fill in elements on the guideline each week as they gain new knowledge. During the semester, five different critiques are done step by step from beginning to end in classroom exercises. Choosing a variety of articles enhances learning, but the choice of articles must be congruent with the level of learner. In addition to the five critiques done in class, students complete two critiques outside of class as part of a small group research utilization project. During final sessions of the course, the entire class critiques an additional research article as a capstone experience to summarize learning.

Students are also required to pass a critique examination by filling out the sample critique guideline on an article they have not read. This culminating critique experience is an open-book, open-notes exam. Students are always very nervous about this kind of exam because it requires very different critical-reasoning skills than the typical NCLEX format nursing tests. However, results of exams provide evidence of beginning mastery of critiquing skills that have been carefully nurtured throughout the semester. All semester, learning expectations are geared toward writing a mastery content exam. Nothing succeeds like success. Positive results are followed by great student enthusiasm. Students believe in their own abilities, rate the course and teaching methodology as very beneficial (contrary to the usual nursing research course evaluation), and volunteer to do research projects independently in other clinical courses.

After using the guideline for several years, I found myself repeatedly reminding students of salient points. Beginning consumers of research find a plethora of information that is overwhelming, so students need cautionary check-in points. An analysis of my critiques of student work on the guide-

	Variable	Categorize variable	Operational definition	Level of measurement	Descriptive statistics	Reliability	Validity
1							
2							
3							
4							
5							
6							

Sample size, power analysis, and descriptive statistics

Draw out the relationships between variables

Kind of statistics reported in study (parametric and nonparametric)

1. Descriptive statistics (not reported above)

2. Inferential statistics

Figure 1. *Guideline for Statistical Analysis.*

lines resulted in the *Golden Rules of Statistical Analysis Adequacy* (Figure 2). These rules are built on the integration of nursing science and statistical understanding to help undergraduate and graduate students make judgments about research findings. The rules are developed for student use in evaluating their own interpretations. The statistical part of the critique must be done in concert with other critique guidelines that are clearly outlined in classic nursing research texts.[8,9]

Application Exercise

To illustrate this critique process guideline, directives and rules are applied to a published research study on grief responses, coping processes, and social support.[10] A completed data guideline sample is presented in Figure 3. Six variables (social support, social network, income/education, spiritual beliefs, coping process, and grief response) are listed in the first column and categorized in the second. It is interesting to note that variables are categorized differently in different hypotheses and research questions. This becomes more evident in the center box of the guideline, where students draw out the postulated relationships. In the example critique, coping process is a dependent variable in the first hypothesis (H1), an independent variable in the second hypothesis (H2), and an intervening variable in the research question (RQ).

Operational definitions of variables are abbreviated and listed in the third column of the guideline. The operational definitions in this study include: the Norbeck Social Support Questionnaire[11] (NSSQ); the Robinson Bereavement Questionnaire[10] (RBQ); the Jalowiec Coping Scale (JCS), which is reported on by A. Jalowiec (unpublished data, 1989); the Grief Experience Inventory[12] (GEI); and the Role Function Measure[10] (RFM). The level of measurement is listed in column four. Students are taught that this information is essential in assessing the appropriateness of the statistics used in the study. Column five lists the descriptive statistics reported on study variables.

Reliability measures in the literature, as well as reliability measures in the study, are listed in column six. Students are taught that reliability measures are sample-specific. For example, a measure of general health that is piloted and tested on college-age students may or may not fit a sample of elderly participants. Similarly, reliable measures that are designed for the prominent United States culture may not be reliable for minority groups. The last column of the guideline lists types of validity reported in the study.

The middle section of the guideline asks for the sample size, power analysis, and descriptive statistics of the sample in order to assess whether statistical findings can be generalized to populations. Power analyses were not and still are not always incorporated into reported research studies. If not reported, students are asked to determine if there are at least ten subjects per study variable. Students are also asked to carefully observe subscale analysis. When subscales are treated as variables, sample size adequacy needs to be reassessed. Larger samples may be needed to maintain the power required for drawing conclusions about subscale analysis.

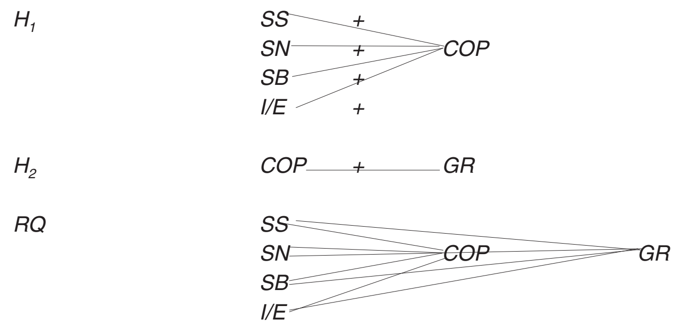
The final and most important step is to list the statistics used and the level of significance reported. During this step, students are referred to textbooks that define and/or describe var-

1. **Don't try to interpret a measure of centrality without a measure of variability.** Samples with equal means may have vastly different ranges and standard deviations indicating more differences than similarities between groups.
2. **Do check the level of measurement on all variables.** Make sure interval or continuous measures are used for all parametric statistics.
3. **Don't mix correlation and causation conclusions.** The existence of a relationship does not imply causation. Causality is established by design, not by analysis. Lazarsfeld[13] lists three criteria for causality: a) temporality; b) association; and c) cannot be explained by another variable. Remember findings support hypotheses and do not prove them and that the magnitude of the correlation always needs assessment.
4. **Do check the definition/description of all statistics used and their application criteria.** Statistics used have to be appropriate and need to be congruent with hypothesis(es) and research question(s) as well as the research design.
5. **Don't try to interpret an inferential statistic reported that does not list its level of significance.** The result could have occurred by chance it it is not assessed for its statistical significance. Remember, statistical significance is very different from clinical significance.
6. **Do check assumptions for statistical tests used and make sure they are met.** Problems with homoscedasticity and lack of normal distribution need to be addressed.[14]
7. **Do check reliability and validity reports in the literature on all instruments.** Use caution in interpreting new instruments without reported psychometric data and on instruments with coefficients below 0.70. Remember, a reliable tool is not necessarily a valid one.
8. **Don't assume instruments with reported reliability measures (internal consistency) are reliable for any sample.** Reliability measures are sample specific. A scale with an established reliability for college-aged students may not be adequate for the frail, elderly. Use interpretive caution for all instruments with reliability coefficients below 0.70.
9. **Do check adequacy of sample size.** Power analysis should be reported. If not, make sure there are at least 10 subjects per variable. When sub-scale analyses are reported, sample size adequacy needs to be reassessed. Type II errors are more likely to occur with inadequate size samples.
10. **Don't forget to assess the risk of Type II error.** The validity of nonsignificant findings needs to be questioned when the power is below 0.80.

Figure 2. *Golden Rules for Statistical Analysis Adequacy.*

Variable	Categorize Variable	Operational Definition	Level of Measurement	Descriptive Statistics	Reliability	Validity
Social support [SS]	Independent	NSSQ[11]	Interval		Internal consistency & test-retest in literature	Construct concurrent & predictive
Social network [SN]	Independent	NSSQ[11]	Interval		Internal consistency & test-retest in literature	Construct concurrent & predictive
spiritual beliefs [SB]	Independent	RBQ[10]	Interval		Internal consistency 0.58 for sample	Content, new instrument
Income/ education [I/E]	Independent	RBQ[10]	Interval			
Coping process [COP]	Dependent H1 independent H2 intervening RQ	JCS	Interval	Range 66-352	Internal consistency 0.88–0.94 range in literature 0.88 in sample;	Reports valid instrument in literature
Grief response [GR]	Dependent	GEI[12] and RFM[10]	Interval	GEI Range 16-107	GEI Cronbachs 0.88–0.94 range in literature; sample= 0.93 RFM = 0.70	Convergent & discriminant in literature; RFM - content only

Draw out the relationships between variables



Sample size, power analysis and descriptive statistics

- *N* = 65
- power analysis calculated for multiple regression with six variables, medium size effect, alpha set .05 and power .80
- age: mean - 64.8 years; *SD* 10.7; range 45–84
- formal education: mean - 11.6 years *SD* 2.3; range 7–16
- years married: mean – 38.8 years; *SD* 14.9; range 1–65
- race: 92% Euro-American, 8% African American
- religion: 91% Protestant; 9% Catholic
- income: 29% below $10,000; 38% between $10,000 and $20,000; and 32% above $20,000
- time since spousal death: mean 18.74; *SD* 3.7; range 13–24 months

Kind of statistics reported in study

1. Hypothesis 1 was tested with Pearson's r statistical tests. Table 1 reported a correlation matrix of all variables (SS, SN, SB, I/E, COP, GR); asterisks indicated which correlations were significant. Correlations supported a positive relationship between SS and COP and SN and COP.
2. Hypothesis 2 was tested with Pearson's r statistical tests and reported in text (COP to GEI -0.29, *p* = .01) and (COP to RFM 0.36, *p* = .01). Correlations supported a negative relationship between COP and GR.
3. Path analysis with SS, I/E and SB as independent variables, COP as intervening and GI as the dependent variable. Social network was dropped from the model because of strong positive correlation with SS (multicollinearity). The explained variance was 18%.

Figure 3. Critique of Robinson's[10] research article.

ious statistics with their application considerations. Both Polit and Hungler[9] and Beitz[7] have published tables that are excellent references to use in making the evaluation of a given statistic. Students need to interpret results and also decide if the appropriate statistic was used based on the study hypothesis(es) and/or research question(s). Students can make assessments of strengths and limitations of the study once materials are put on the data guideline.

In reviewing Figure 3, which is typical of critiques done in class by students, the following conclusions can be drawn. The six variables all have operational definitions, are interval level, and are categorized differently according to specific hypotheses and research question. Descriptive statistics were listed for two variables and only included "ranges;" another measure of centrality would have been helpful. There was literature support for reliability for five of the variables, and study reliability coefficients for the coping process and grief response were sufficient (> 0.70). The spiritual beliefs measure was not a reliable measure in the sample and was noted as a limitation. Some type of validity was reported for five study variables. The sample size of 65 was adequate for the six variables (power analysis). Demographic characteristics of sample were listed with appropriate descriptive statistics (measures of centrality and variability appropriate for level of measurement).

The statistics used were Pearson r correlations and path analysis. A Pearson product-moment correlation is used to test if a relationship exists between two interval level variables. This was an appropriate statistic for this study with multiple interval level variables. The article used a correlation matrix to display data; significant correlations were starred in the article, lending support (weak to moderate) in the directions predicted in the hypotheses. Path analysis is a multivariate statistic used to study relationships among variables. Path analysis is based on simple regression techniques and calculates the percent of explained variance in an interval level dependent variable with multiple independent and intervening interval level variables. Path analysis is an appropriate statistic to use for the research question that asks about the explained variance of independent and intervening variables on a dependent variable.

Classroom Implications and Evaluation

While statistics is a science that takes many years to master, students can learn to apply statistical concepts as beginning research consumers. Students need to be taught how to read a research report and how to differentiate main points from minutia. The *Guideline for Statistical Analysis* needs to be introduced early in the research course and needs to be applied in every class period. Students like the chart format and will easily hunt for data that are missing. Caution needs to be exercised, as students will try to fill in every box on the guideline, even when there are no data in the research report. Students get excited when they can identify the limitations of studies before they get to that section of the report.

The data guideline is conducive to small and large group classroom work. When used with small groups, competition among groups for answers turns into a game. When used with a large group, students are asked to vote as to whether the item entry is accurate or not. This forces students to make decisions about an answer, makes class more fun, and encourages critical-reasoning skills. We laugh a lot as we try to make judgments. While I have only used the guideline in a traditional classroom, a colleague is preparing to pilot it with a virtual classroom next semester. The guideline will add desired structure to the distance education format, as it is designed for interactive learning that would lend itself to an online presentation.

The *Golden Rules for Statistical Analysis Adequacy* are currently being piloted with students. These rules are check-off points and summarize a great deal of material. A frequent complaint of students in research classes is the grayness of interpretation; students want to see things as more black and white. While some may argue that it is important to keep things gray, I believe education needs to address the level of the learner. The intent is to start at a concrete level and proceed to the abstract in more advanced courses. While I know there is rationale for exceptions to these rules for certain circumstances, my intent is to give students a general set of statistical "commandments" for interpretation.

These "do and don't" rules are used in class where there is in an interactive environment to make judgments. For example, I will say, "The researcher says that there is a relationship between these two variables. What do you say?" At least one third of the class will quickly note that there isn't a reported level of significance, which means the correlation could have been due to chance. We gently chastise the rest of the class for not noting the omission, and then the whole class recites the rule that applies to this interpretation. Another time I might say, "Look at the 'anxiety level variable' and tell me what you know about it." Most of the class will quickly find its operational definition; then, they compete to find all the reported psychometric data on it. Students will also apply the rules to their independent critique projects.

Students are instructed to keep their Golden Rules open for each class period. The more students can use and recite from them, the more students will learn the concepts and principles behind the rules. Using the guideline and rules has helped me transform the research classroom into a more dynamic exchange between faculty and students instead of a "sage on the stage" atmosphere. I now lecture only about 10% of the time. Course and teaching evaluations by basic students are much more positive than they used to be. RN/BSN students value interactive learning in the curriculum and respond very positively to this teaching strategy. Students now look to themselves, their colleagues, and the text for answers instead of to the instructor. The short-term gain for students from this strategy is greater self-confidence and self-reliance. The long-term gain for the profession is yet to be realized; however, students who become avid, enthusiastic research consumers early in their career can only improve their professional practice and the care that patients receive.

Acknowledgment

I acknowledge the late Dr. Thomas J. Duggan for facilitating my appreciation and understanding of statistics through five semesters of course work and consultation on my dissertation. Tom role modeled excellence in teaching, mentored my professional growth, and gifted me with his passion for statistics. This manuscript is my gift to him. He read it shortly before his death.

References

1. Stranahan SD. Strategies for teaching nursing research: sequence of research and statistics courses and student outcomes. *West J Nurs Res.* 1995;17(6):695–699.
2. Schuster P, Ritchey N. Teaching introductory statistics to baccalaureate nursing students. *Nurse Educator.* 1998;23(5):34,41.
3. Shields LE, Brunt JH., Milliken PJ. Teaching quantitative and qualitative analysis to undergraduate students. *J Nurs Educ.* 1999;38(3):136–138.
4. Beitz JM, Wolf ZR. Creative strategies for teaching statistical concepts in nursing education. *Nurse Educator.* 1997;22(1):30–34.
5. Berry J, Kenny C. Using a game to teach research and statistics. *Nurse Researcher.* 1994;1(4):69–77.
6. Ihlenfeld JT. Teaching computerized data analysis in the classroom. *Comput Nurs.* 1993;11(6):269–270.
7. Beitz JM. Helping students learn and apply statistical analysis: a metacognitive approach. *Nurse Educator.* 1998;23(1):49–51.
8. Burns N, Grove SK. *Understanding Nursing Research.* 2nd ed. Philadelphia: WB Saunders; 1999.
9. Polit, DF, Hungler, BP. *Essentials of Nursing Research: Methods. Appraisal, and Utilization.* 4th ed. Philadelphia: Lippincott; 1997.
10. Robinson JH. Grief responses, coping processes and social support of widows: research with Roy's model. *Nursing Science Quarterly.* 1995;8(4):158–164.
11. Norbeck JS, Lindsey AM, Carrieri VL. The development of an instrument to measure social support. *Nurs Res.* 1981;(30):264–269.
12. Sanders CM, Mauger PA, Strong PN. *A Manual for the Grief Experience Inventory.* Palo Alto: Consulting Psychologist Press: 1985.
13. Lazarsfeld P. Foreword. In Hyman H., ed. *Survey Design and Analysis.* New York; The Free Press: 1955.
14. Munro BH. *Statistical Methods for HealthCare Research.* 3rd ed. Philadelphia: Lippincott; 1997.

News Notes & Tips

Continued from page 135

The students had gathered most of the objective information related to the client: unresponsive, NG tube with drainage, Foley catheter with urine in bag, etc.

The students were intrigued to see how many things they had missed, including using the sister as a source of data. While every student made his/her assessment, the sister kept trying to communicate that the client had been sick with "HIV" for a long time and was taking some medicine that started with an "A." The students discovered that they were so focused on the objective part of the assessment that they did not form a holistic picture with subjective information from the family member. Figure 1 shows types of questions used to stimulate student discussion of the assessments.

This One Minute Assessment activity benefited both the student and faculty. Student feedback was overwhelming positive. The most common comment was that the exercise required the students to "think on their feet" and prioritize as they progressed. The faculty found the activity to be an effective and time-efficient exercise in assessment and critical thinking.

Incorporating Critical Thinking Exercises and Clinical Case Scenarios Through Time in the Classroom

Edith A. West, MSN, RN, CS, Assistant Professor of Nursing, ewest@gw.lssu.edu. Lake Superior State University, Sault St. Marie, Michigan.

The dilemma that exists between what healthcare managers expect their baccalaureate prepared graduate nurses to do and what these graduate nurses are actually prepared to do is not a new one, but it is growing worse. Having only recently entered educational practice, I knew from recent experience as a staff nurse that lengthy orientation programs and hospitals filled with experienced RNs with whom a "new" nurse could readily consult and be mentored no longer existed. Yet, the care of more acutely ill patients demands that beginning nurses come to institutions ready and able to critically think, prioritize, organize, delegate to unlicensed personnel, and take appropriate actions. So I asked myself, what needs to be done? What can I do to better prepare my students to enter, stay in, and facilitate positive changes now?

Realizing the case studies I had been using did not fully reflect the varied components of a patient care situation, I began searching for more realistic and comprehensive case studies for use in the classroom. I found the workbook *Strategies, Techniques and Approaches to Thinking: Case Studies in Clinical Nursing* by DeCastillo.[1] I began incorporating the critical thought exercises and case scenarios through time into my junior level adult medical/surgical nursing course this past year with some surprising results. DeCastillo's critical thinking exercises and case scenarios include progression through time on a given patient and changes in the patient's condition that warrant assessment, analysis, intervention, and evaluation by the student. The scenarios include kardex information, medication, intake and output and laboratory records, and/or nurses notes. The student has to review more than one source of information and interpret, analyze, and evaluate the information to make appropriate inferences. They also have to explain the rationale for the nursing diagnoses and interventions formulated from this information and prioritize care. The scenarios enable students to see how a patient's condition can

Continues on page 152