AACN Advanced Critical Care
Volume 19, Number 2, pp.211–222

Interpretation and Use of Statistics in Nursing Research

Karen K. Giuliano, PhD, RN, FAAN
Michelle Polanowicz, MSN, RN

ABSTRACT

A working understanding of the major fundamentals of statistical analysis is required to incorporate the findings of empirical research into nursing practice. The primary focus of this article is to describe common statistical terms, present some common statistical tests, and explain the interpretation of results from inferential statistics in nursing research. An overview of major concepts in statistics, including the distinction between parametric and nonparametric statistics, different types of data, and the interpretation of statistical significance, is reviewed. Examples of some of the most common statistical techniques used in nursing research, such as the Student independent *t* test, analysis of variance, and regression, are also discussed. Nursing knowledge based on empirical research plays a fundamental role in the development of evidence-based nursing practice. The ability to interpret and use quantitative findings from nursing research is an essential skill for advanced practice nurses to ensure provision of the best care possible for our patients.
Keywords: nonparametric statistics, parametric, statistical significance, statistical tests

Over the past decade, the use of evidence-based medical practice has increased dramatically and become the standard for healthcare decision making. Its popularity has given rise to a proliferation of evidence-based articles, conferences, and tools related to healthcare delivery. French[1] reported that a frequency analysis of the key word *evidence-based* in the healthcare literature yielded citations for almost 6000 articles, with the majority published since 1995. The evidence-based movement has elicited strong interest among healthcare professionals as one of the key elements for optimal clinical decision making and provision of quality healthcare. It is likely that some aspects of healthcare reimbursement will be tied to the use of the best available evidence within the next decade as suggested by the current debates about pay for performance.[2]

Scholarly inquiry to develop and apply evidence-based practice in nursing can take variety of forms, and there is much evidence to support the premise that nursing knowledge develops from multiple perspectives and lenses.[3] In nursing, there has been an explosion in knowledge during the past 20 years, thus providing the discipline with diverse and multifaceted theoretical frameworks and practice paradigms. Although there is widespread agreement that the development of nursing knowledge should not support an epistemology in which empirical knowledge is the pinnacle or "criterion standard," the contribution of empirical inquiry to the science of nursing must be recognized. The philosophy of Aristotle and his

Karen K. Giuliano is Principal Scientist, Philips Medical Systems, 3000 Minuteman Rd, MS 500, Andover, MA 01810 (Karen.giuliano@philips.com); and Associate Research Professor, Boston College, William F. Connell School of Nursing, Chestnut Hill, Massachusetts.

Michelle Polanowicz is Senior Product Manager, Philips Medical Systems, Andover, Massachusetts.

impact on the process of empirical scientific inquiry continue to be a major force in the building of evidence-based practice paradigms.[4] Hence, traditional empirical approaches will continue to exert an important influence on the development of knowledge through research. As a result, it is imperative for nurses to strive for continual improvement in their understanding of quantitative research approaches including research design, data analysis, and interpretation of results. Such understanding is an absolute requirement for the clinical application of research evidence, a more complete understanding of individual patient responses to the provision of healthcare, and the conduct of new research. The purpose of this article is to provide an overview of some of the most common statistical terms and analytic methods found in nursing research and describe how statistical results can inform the process of clinical care provided by nurses.

Descriptive and Inferential Statistics

Statistics uses well-defined mathematical procedures to collect, analyze, interpret, and present data. The major purpose of these quantitative procedures is to summarize and reduce data into parts that are small enough to interpret within the context of a predetermined theoretical background. Two major forms of statistics are available: descriptive and inferential.[5–8]

Rather than collecting data from an entire population, researchers usually study a carefully selected subset of the population, which is known as the study sample. Descriptive statistics organize and describe the characteristics of the data in a particular study sample. The main purpose of descriptive statistics is to provide as much detail as possible about the characteristics of the study sample, which helps determine whether it is appropriate to apply research findings from that study sample to populations similar to the one that has been included in the sample. Descriptive statistics are commonly used in many aspects of daily life and are primarily used to measure central tendency and variance, which are described below. For example, descriptive statistics are used when the average body temperature, weight, or age is reported for a group of patients or study subjects.[5–8]

Descriptive statistics are used to describe a study sample, whereas inferential statistics use information or data collected about the study sample to make inferences about a larger population. Inferential statistics allow researchers and clinicians to make predictions about a specific population on the basis of information obtained from a sample that is representative of that population. The primary focus of this article is on the use and interpretation of inferential statistics in nursing research because knowledge based on results of inferential statistical analysis plays a critical role in the development of evidence-based nursing practice.[5–8]

Types of Data

A fundamental tenet of the interpretation of data involves understanding the type of data that are to be analyzed. Data can be categorized into 2 types: categorical or continuous. Categorical data are based on counts that simply put variables into categories that have no meaningful numerical or quantitative relationship. A common example of a categorical variable is gender. A statistical analysis software program such as the Statistical Package for the Social Sciences assigns values such as 1 for male and 2 for female to categorize gender for a study sample. The numbers assigned to the 2 gender categories allow the statistical software program to calculate the frequency of males and females in a given sample. The number assigned to each gender category is arbitrary and has no actual numerical significance.[5–8]

In contrast, continuous data provide information that can be measured on a continuum or a scale. Continuous data are based on a measurement scale that can be used mathematically to calculate additional values. These data have a potentially infinite number of values along a measurement continuum.[5–8] Common examples of continuous healthcare data include height, weight, cholesterol level, waist circumference, and temperature.

Measures of Central Tendency

Statistics are based on the idea of a normal distribution of data or what is more commonly referred to as a *bell curve*. Because most data are clustered around the center of the distribution, measures of central tendency are a fundamental component of statistics. The 3 main measures of central tendency are mean, median, and mode. Each measure provides a slightly different view of the distribution of the data for a

sample.[5–8] The mean is the most common measure of central tendency. It is the sum of all the values in a group, divided by the number of values in that group. The result is referred to as the mean or, more commonly, the average. The median defines the midpoint in data set, which is the point at which half the data fall above and half fall below the midway mark. The mode is the most general measure of central tendency, but it still plays an important role in understanding the characteristics of the data. The mode is the value that occurs most frequently in the sample. It can be easily ascertained visually by arranging the data set in order and reviewing it if the data set is small or by using a statistical analysis package for a large data set.[5–8]

The other concept fundamental to the use of descriptive data is variance, which measures the dispersion of values collected from a sample around the measure of central tendency. It is not possible to fully interpret measures of central tendency without knowing the variance or degree to which scores vary from the measure of central tendency. The most commonly used measure of variance in descriptive statistics is the standard deviation, which describes how widely spread the values in a data set are from the mean.[5–8] For example, let us look at 3 study groups, each with 3 men in the sample with a mean group weight of 160 lb. In study group 1, each of the 3 men weighs exactly 160 lb, which means there is no variance and the SD would be 0. In study group 2, the 3 men weigh 150, 160, and 170 lb, respectively, again with a mean group weight of 160 lb. This constitutes a small amount of variance, which would be reflected by a small standard deviation value. In group 3, 1 man weighs 100 lb, another weighs 120 lb, and the third weighs 260 lb. Although the mean weight for group 3 is still 160 lb, the weights among the 3 individuals vary much more than the first 2 groups, which means that the variance would be greater with a large standard deviation. Because the variance is so different in the 3 study groups, it is unlikely that research findings on weight dosing for a new medication could reliably be applied to all 3 groups with the same results, although the group means are identical.

Data from a study sample are often presented graphically as a standard normal distribution. Figure 1 displays a standard normal distribution curve with a mean of 0 and an SD of 1. One standard deviation above and below the midpoint includes 68% of the values in a sample that are normally distributed and about 95% of the values fall within ± 2 SDs of the mean.[5–8]

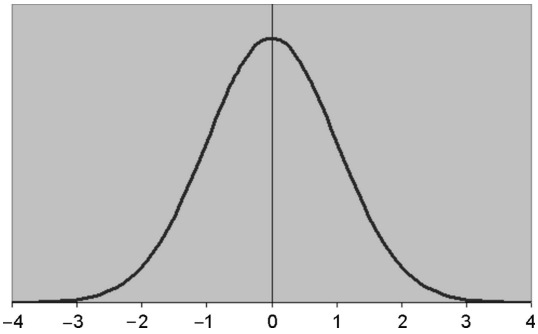

Figure 1: Standard normal distribution.

Hypothesis Testing and Significance Levels

Medical researchers and clinicians ask many different types of questions about relationships between interventions (such as a medication) and patient outcomes. An essential component of the research process is forming and evaluating hypotheses, which are empirically testable statements about a relationship between 2 or more variables. Hypothesis testing allows researchers to determine whether the variance between 2 or more study groups can be explained either by chance or by the intervention, such as administration of a medication to lower systolic blood pressure (SBP). Every clinical study requires a statement of a null hypothesis and an alternate hypothesis. The null hypothesis asserts that there is no relationship either between the variables being studied or between or among groups. The alternate hypothesis states that there is a true relationship between the variables of interest that is not attributable to chance. When asking any research question, the null hypothesis is assumed to be true unless another hypothesis is supported by the research findings.[5–8]

The significance level is another important concept in statistical analysis because it affects the likelihood that null hypotheses will be accepted or rejected. No research study is perfect and all research findings are subject to potential error due to either study design or chance. Thus, levels of statistical significance must be established before the conduct of any research or statistical analysis in terms of a probability or a *P* value. The significance level indicates the probability (ie, the chance) of rejecting the null hypothesis when it is actually

true. Although significance levels can be set at any value, the most common *P* values in nursing research are .05 (5%) or .01 (1%). This means that there is either a 5% or a 1% chance of rejecting the null hypothesis when, in fact, it is true. Rejecting the null hypothesis when it is true is known as a type I error or a false-positive result, and it occurs when an observed difference between study groups is actually due to chance. A type I error is also known as an *alpha error* denoted by the Greek letter α. The lower the significance level or *P* value, the less likely it is that a researcher will make a type I or false-positive error.[5–8]

Type II errors, or β errors, can also occur when a researcher is evaluating the statistical significance of study results. A type II error occurs when researchers fail to reject the null hypothesis and when the alternative hypothesis is actually true. This is referred to as a false-negative error and means that the researchers attributed the study results to chance, rather than a true difference between the study groups.[5–8]

Sensitivity and Specificity

Screening tests are often used to determine whether a patient does or does not have a disease. Two common ways to evaluate the accuracy of a screening test are sensitivity and specificity. *Sensitivity* refers to the likelihood that an instrument, measurement, or medical test will correctly identify those individuals who have a particular attribute, such as a disease. This is often referred to as a *true positive*. Conversely, the term *specificity* refers to the probability that an instrument, measurement, or test will correctly identify the absence of a disease in patients who do not have the disease, which is referred to as a *true negative*. Tests or measures with a high level of specificity minimize the number of times a healthy patient is diagnosed with a disease or a condition that he or she does not have.[5–8]

Table 1 shows the results when the Surviving Sepsis Campaign practice recommendations were applied to a group of critically ill patients at risk for sepsis.[9] According to Sepsis Campaign recommendations, a total of 232 individuals were predicted not to have sepsis compared with an observed number of 157 individuals without sepsis and 191 who had sepsis. With respect to prediction of sepsis, the guidelines estimated that 471 individuals would have the condition. However, there were 280 observed or confirmed cases of sepsis and 75 cases of predicted nonsepsis according to the Sepsis Campaign recommendations. Analysis of these data reveals that the recommendations were more accurate for the prediction of sepsis in individuals who actually had the condition, and 78.9% of the cases were correctly classified by the recommendations, which is the sensitivity of the test. The formula for calculation of sensitivity is number of true positives divided by the number of true positives plus false negatives (true positives/[true positives + false positives]). In the sepsis study example, the sensitivity is calculated as 280/(280 + 75), which equals 78.9%. In contrast, the guidelines correctly predicted only 45.1% of the individuals who did not have sepsis, which is a measure of the specificity of the recommendations for prediction of sepsis. Specificity is calculated by dividing the total number of true negatives by the total number of false positives plus the true negatives. In this example, specificity is calculated as 157/(157 + 191), which equals 45.1%, representing a low specificity value.[9]

Positive predictive value (PPV) provides a measure of the predictive value of an instrument or test and is based on the proportion of individuals tested who are truly positive. The

Table 1: Classification Table for Sepsis[a]

	Predicted		
Observed	**Nonsepsis**	**Sepsis**	**% Correct**
Nonsepsis	157	191	45.1
Sepsis	75	280	78.9
Overall %			62.2

[a]Used with permission from Giuliano KK. Physiologic monitoring for critically ill patients: testing a predictive model for the early detection of sepsis. *Am J Crit Care.* 2007;16:122–130.[10]

formula to calculate PPV is the total of true positives divided by true positives plus false positives (true positives/[true positive + false positives]). In the sepsis study example, the PPV is calculated as 280 observed cases of sepsis (true positives)/(191 [total number of true positives] + 280 [total number of false positives]), which equals 59.4%. Finally, the negative predictive value (NPV) can be calculated by totaling the number of true negatives and dividing this value by the total of true negatives plus false negatives. In the sepsis study, the NPV is 67.7% on the basis of the calculation of 157 true negatives/(157 [true negatives] + 75 [false negatives]).[9]

The weakness of the measure of sepsis used in this study is that whereas most patients with sepsis would be accurately identified as having sepsis (true positives), a fairly high number of patients without sepsis who were equally ill would also be screened positive for sepsis (false positives). This means that a fairly high number of patients without sepsis would be predicted to indeed have sepsis. In other words, although the measure of sepsis model performed reasonably well for screening in patients who had sepsis, it did a poor job of identifying those patients who did not have sepsis.[9]

Confidence Intervals

A confidence interval (CI) provides a range of values associated with the probability that a variable will fall within that range. The larger the CI, the less precise is the measurement of that variable. Confidence intervals of 95% and 99% are most common in medical research. For example, if we wanted to know the average SBP among a sample of 100 women treated with a medication to lower SBP and the CI was set to 95%, we could expect that 95% of the women in the sample would have an SBP that fell between the upper and lower limits of the range. If we had calculated a 99% CI, we would be more precise in estimating the average SBP among women receiving the treatment to lower blood pressure. However, a CI does not mean that 95% or 99% of each woman in the group has an SBP that equals the mean SBP for the entire sample. Rather, the CI suggests that 95% or 99% of the women will have SBP that falls between the lower and upper limits of the range of blood pressures. This concept is important for understanding the correct interpretation of research findings.[5–8]

Common Statistical Techniques Used in Nursing Research

Many research questions are asked to determine whether a difference exists between 2 groups on a variable or a group of variables and whether that difference is due to chance. Before deciding on the most appropriate statistical test to be used, a few questions must be answered about the individual variables and data set to be evaluated. First, the variable must be identified as either categorical or continuous. If the variable is a categorical one, the most frequently used statistical technique is the χ^2 test, which evaluates statistically significant differences between frequencies or proportions for 2 or more groups. The χ^2 test compares the actual frequency with the expected frequency of each variable measured in each group.[5–8] For example, in the study of patients with sepsis, 2 descriptive variables were survival to hospital discharge (yes or no) and gender (male or female). A χ^2 analysis was used to determine whether either of these variables was significantly different between the septic and nonseptic patient groups.[9]

Using a significance value of .05, the results in Table 2 show that there is a significant difference in survival to hospital discharge ($P = .008$) in patients with sepsis compared with those who are not affected by sepsis. However, the χ^2 test for differences in frequency of sepsis attributable to gender was not significant ($P = .455$), and we would conclude that the probability of having sepsis was not associated with gender.[10]

Assessment of group differences for continuous variables requires the researcher to first determine whether the variables are normally distributed (eg, conform to the standard normal distribution or bell curve) or whether the distribution of the variables is skewed, which means the data are not normally distributed. Statistical tests for skewness include the Fisher measure of skewness. Using this test, any values that exceed ± 1.96 SDs from the mean would be considered significantly skewed, because in a normal distribution, 95% of values for a variable fall between ± 1.96 SDs of the mean.[5–8] In the example data set for the patients with sepsis, all variables were significantly skewed as shown in Table 3.[10]

Because the variables were skewed and not normally distributed, it is necessary to calculate a nonparametric statistical test to determine whether group differences were statistically

Table 2: Chi-Square Test for Dichotomous Variables Among Demographic and Study Variables for Septic and Nonseptic Patient Groups[a]

	Sepsis Frequency, %	Nonsepsis Frequency, %	Total Frequency, %	*P*
Survival[b]				
Survivor	202 (55.6)	235 (64.7)	437 (60.2)	.008
Nonsurvivor	161 (44.4)	128 (35.3)	289 (39.8)	
Total	363	363	726	
Gender				
Male	191 (52.69)	189 (51.9)	380 (52.3)	.455
Female	172 (47.4)	175 (48.1)	347 (47.7)	NS
Total	363	364	727	

Abbreviation: NS, not significant.
[a]Used with permission from Giuliano KK. Physiologic monitoring for critically ill patients: testing a predictive model for the early detection of sepsis. *Am J Crit Care*. 2007;16:122–130.[10]
[b]Survival to hospital discharge.

significant. Nonparametric statistical tests allow interpretation of skewed data collected from a sample, which often occurs when data do not have a strong numerical relationship. Nonparametric statistical models rely on distribution-free statistical methods and make fewer assumptions than parametric statistics. Examples of nonparametric statistical tests include the χ^2 test, the Cochran *Q* test, the Fisher exact test, and the Mann-Whitney *U* test or Wilcoxon rank sum test. In contrast, parametric statistical tests are based on the assumption that the study variables are continuous, normally distributed, and have equal variance (referred to as homogeneity of variance). One of the major differences between a parametric statistical test and a nonparametric statistical test is that nonparametric tests are not used to estimate population parameters, because they do not have normally distributed data. Examples of parametric statistical tests include the *t* test, 1-way analysis of variance (ANOVA), repeated-measures ANOVA, Pearson correlation, simple linear and nonlinear regressions, and multivariate linear and nonlinear regressions.[5–8] Table 4 shows the data and significance levels for the Mann-Whitney *U* tests completed for each of the variables in the sepsis study.[9] The mean and the standard deviation are shown only as a point of reference but were not used in the analyses.

Table 3: Skewness for Study Variables (N = 727)[a]

Variable	Skewness[b]
Age	–7.38
SAPS II	6.26
24-h urine output	21.50
Hospital LOS	40.91
ICU LOS	54.10
24-h highest HR, beats/min	2.76
24-h highest RR, breaths/min	12.70
24-h lowest SBP	–4.04
24-h lowest MAP, mm Hg	–3.20
24-h lowest temperature, °C	–10.45
24-h highest temperature, °C	6.97

Abbreviations: HR, heart rate; LOS; length of stay; MAP, mean arterial pressure; RR, respiratory rate; SAPS II, Simplified Acute Physiology Score; SBP, systolic blood pressure.
[a]Used with permission from Giuliano KK. *Physiologic Monitoring for Critically Ill Patients: Testing a Predictive Model for the Early Detection of Sepsis* [dissertation]. Chestnut Hill, MA: Boston College; 2005.[9]
[b]Values ± 1.96 are significantly skewed at the .05 level.

Upon review of all of the variables measured in the sepsis study, it was determined that the lowest respiratory rate (RR) was, in fact, normally distributed. Therefore, a parametric *t* test shown in Table 5 was completed to determine whether lowest RRs differed significantly between the groups of patients

Table 4: Mann-Whitney *U* Test of Demographic and Study Variables for Septic and Nonseptic Patient Groups[a]

Variable	Total (*n* = 727)	Septic (*n* = 363)	Nonseptic (*n* = 364)	*P*
Patient age, y				
Mean (SD)	65.13	63.43	66.82	.001
Range	18–90	18–90	18–90	
Mean rank		339.02	388.91	
SAPS II score				
Mean (SD)	52.38 (19.18)	52.61 (19.33)	52.16 (19.04)	.755
Range	20–113	20–113	20–110	
Mean rank		354 48	349.69	
Urine output,[b] mL				
Mean (SD)	1572.84 (1457.68)	1422.11 (1341.12)	1721.05 (1551.44)	.003
Range	0–12 450	0–8275	0–12 450	
Mean rank		333.51	379.11	
Hospital LOS, d				
Mean (SD)	21.63 (26.48)	25.45 (30.58)	17.82 (20.98)	<.001
Range	1–232	1–232	1–232	
Mean rank		401.94	326.17	
ICU LOS, d				
Mean (SD)	7.15 (9.59)	9.04 (11.78)	5.26 (6.18)	<.001
Range	0.4–126	0.4–126	0.4–37.8	
Mean rank		411.36	316.77	
Highest HR,[b] beats/min				
Mean (SD)	117.91 (25.52)	120.62 (23.87)	115.21 (26.83)	.002
Range	59–231	70–231	59–204	
Mean rank		405.08	355.92	
Highest RR,[b] breaths/min				
Mean (SD)	27.79 (9.07)	28.77 (8.49)	26.80 (9.53)	<.001
Range	8–80	8–62	10–80	
Mean rank		394.84	332.16	
Lowest SBP,[b] mm Hg				
Mean (SD)	83.22 (24.99)	78.62 (22.20)	87.81 (26.74)	<.001
Range	0–180	0–180	0–163	
Mean rank		320.28	407.60	

(*continues*)

Table 4: Mann-Whitney *U* Test of Demographic and Study Variables for Septic and Nonseptic Patient Groups[a] (*Continued*)

Variable	Total (*n* = 727)	Septic (*n* = 363)	Nonseptic (*n* = 364)	*P*
Lowest MAP,[b] mm Hg				
Mean (SD)	56.74 (17.08)	52.21 (15.15)	60.30 (18.17)	<.001
Range	0–120	0–110	0–120	
Mean rank		317.62	404.74	
Lowest temperature,[b] °C				
Mean (SD)	36.21 (0.93)	36.31 (0.96)	36.11 (0.88)	.001
Range	31.1–39.3	31.1–38.9	31.9–39.3	
Mean rank		389.50	338.57	
Highest temperature,[b] °C				
Mean (SD)	37.85 (1.03)	38.05 (1.08)	37.64 (0.93)	<.001
Range	33.3–42.2	34.8–42.2	35.3–41.9	
Mean rank		405.89	322.23	

Abbreviations: HR, heart rate; LOS, length of stay; MAP, mean arterial pressure; RR, respiratory rate; SAPS II, Simplified Acute Physiology Score; SBP, systolic blood pressure.

[a]Used with permission from Giuliano KK. Physiologic monitoring for critically ill patients: testing a predictive model for the early detection of sepsis. *Am J Crit Care.* 2007;16:122–130.[10]

[b]During the first 24 hours of ICU admission.

with and without sepsis.[5] The results indicate that the mean lowest RR for patients who did not have sepsis was significantly lower at 13.25 than the mean lowest RR of 13.98 observed in patients with sepsis ($P < .000$).[9]

Analysis of data from the sepsis study required multiple comparisons between variables. Consequently, it was necessary to calculate the Bonferroni correction test for all continuous variables. This test is done to reduce the chance of a type I error, which would assume that a true difference exists between groups when that difference is actually due to chance. Applying the Bonferroni correction to the Mann-Whitney *U* tests completed for the sepsis study, a $P \leq .004$ was required for a difference to be considered statistically significant and not due to chance. Upon review of Table 4, the only variable that was not significantly different between the groups was the Simplified Acute Physiology Score (SAPS II), which provides a measure of illness acuity. The mean SAPS II score was 52.61 for patients who developed sepsis compared with 52.16 in

Table 5: *t*-Test of Demographic and Study Variables for Septic and Nonseptic Patient Groups[a]

	Lowest RR,[b] breaths/min		
	Mean (SD)	Range	*P*
Total (*n* = 760)	13.98 (4.88)	0–30	.000
Septic (*n* = 363)	14.72 (4.99)	0–30	.000
Nonseptic (*n* = 364)	13.25 (4.65)	0–30	.000

Abbreviation: RR, respiratory rate.

[a]Used with permission from Giuliano KK. Physiologic monitoring for critically ill patients: testing a predictive model for the early detection of sepsis. *Am J Crit Care.* 2007;16:122–130.[10]

[b]During the first 24 hours of ICU admission.

patients not affected by sepsis. The *P* value for this difference was .755 and did not meet the Bonferroni correction value of 0.004 or less. This can be interpreted as an indication of no significant difference in illness acuity between the septic and nonseptic patient groups, which was expected because the two groups were matched for acuity.

Analysis of Variance

An important concept to understand when interpreting results from any research study is the difference between independent and dependent variables. Independent variables are deliberately manipulated as part of a study design (such as medication vs placebo, or dietary intervention vs regular diet) to achieve a desired outcome such as lower blood pressure and are also referred to as predictor or explanatory variables. Dependent variables are those that change in response to exposure to the different independent variable groups and may also be called response variables or outcome variables. For example, one could ask the following research question: "Does treatment with an angiotensin-converting enzyme (ACE) inhibitor or placebo for hypertension cause changes in measures of blood pressure (where the dependent variable is blood pressure values of patients enrolled in the study)?" Notice that the independent variable (treatment with either an ACE inhibitor or placebo medication) is a categorical variable and the dependent variable of blood pressure is a continuous variable.

Analysis of variance is a statistical method that allows simultaneous comparisons between 2 or more groups (independent variables) on a dependent variable to determine whether any observed differences in the dependent variable are statistically significant or due to chance. If they are due to chance, the researcher will accept the null hypothesis. However, if the ANOVA results reveal statistically significant differences between groups, the researcher can assume that the alternate hypothesis has been supported. Several different ANOVAs are available, but the most commonly used types of ANOVA in nursing research include 1-way ANOVA, 2-way ANOVA, and repeated-measures ANOVA. A 1-way ANOVA is used to test for differences on 1 independent variable with 2 or more levels. A 2-way ANOVA includes more than 1 independent variable. A repeated-measures ANOVA is performed when multiple measurements of the dependent variable are collected from each of the subjects in the study samples. In the example of a study comparing the effect of an ACE inhibitor with placebo on blood pressure, multiple measurements of blood pressure over time would require analysis with repeated-measures ANOVA. Another more complicated form of ANOVA is a multivariate ANOVA, also known as MANOVA. A MANOVA is conducted when there is more than 1 dependent variable, such as diastolic and systolic blood pressure levels.[5–8]

Repeated-Measures ANOVA

The following describes a statistical analysis with repeated-measures ANOVA for a study that was conducted to compare the effect of varying degrees of backrest elevation on cardiac output measurements in critically ill patients. The research question for this study was "what is the effect of backrest elevation and time on CO measurement in critically ill patients using the CCO method?"[11(p244)] This study included 2 independent variables: backrest elevation at angles of 0,° 30°, and 45° (3 levels) and duration of elevation for 0, 5, and 10 minutes (3 levels) after each change in backrest elevation. The dependent variables were cardiac output assessed by 4 measures including continuous cardiac index, stroke volume (SV), heart rate (HR), and mean arterial pressure (MAP). Each dependent variable was a continuous measure. Analysis of study results involved a single-group, repeated-measures ANOVA with time (0, 5, and 10 minutes) and backrest elevation (0°, 30°, and 45°). A repeated-measures ANOVA was appropriate for this study because a total of 9 measurements of cardiac index were obtained for each patient at each time point and level of backrest elevation including 3 backrest elevations for 3 different time intervals.[11]

The overall results of the repeated-measures ANOVA are displayed in Table 6. No significant differences were reported in any of the 4 dependent measures of cardiac index across the 9 repeated measurements for either time or degree of backrest elevation. In addition, the results of the statistical analysis revealed no interaction between the 2 independent variables of elevation and time.[11] An ANOVA is a particularly informative statistical test because it evaluates the independent effects of the predictor or independent variables and also takes into account any significant

Table 6: Results of Within-Subjects Repeated-Measures ANOVA[a]

	CCI, L/min			SV, mL			HR, beats/min			MAP, mm Hg		
Source	***df***	**F**	***P***	***df***	**F**	***P***	***df***	**F**	***P***	***df***	**F**	***P***
Time	2	0.71	.472	2	0.103	.843	2	1.6	.212	2	0.779	.413
Backrest elevation	2	1.51	.234	2	0.928	.383	2	0.043	.916	2	3.47	.051
Interaction	4	0.367	.715	4	0.470	.614	4	1.23	.289	4	1.45	.246

Abbreviations: CCI, continuous cardiac index; df, degrees of freedom; HR, heart rate; MAP, mean arterial pressure; SV, stroke volume.
[a]Used with permission from Giuliano KK, Scott SS, Brown V, Olson M. Backrest angle and cardiac output measurement in critically ill patients. *Nurs Res.* 2003:52(4):242–248.[11]

interactions between these variables. *Interaction* is a statistical analysis term that means that 1 variable is affected by the values of 1 or more other variables. In this study, an interaction would be present if 1 or more of the 4 measures of cardiac index was affected by a specific combination of the degree of backrest elevation and amount of time spent at the different degrees of elevation. Table 6 indicates that there are no statistically significant independent effects of backrest elevation or duration of time elevated. Nor is the interaction of backrest elevation with duration of time elevated associated with statistically significant changes in any of the 4 measures of cardiac index. In short, the null hypothesis must be accepted and it be assumed that backrest position, duration of time elevated, and the interaction of these 2 variables had no significant effect on the cardiac index in the critically ill patients of this study.[11]

In this research, the variables of HR, SV, and MAP were included because they are physiologically related to cardiac output and cardiac index, which is the cardiac output adjusted for the body size. Additional analyses were done in order to confirm that the lack of significant differences in cardiac index values was due to a real absence of change, not just the occurrence of physiologic compensation resulting from changes in either HR or SV. Because the values of SV and HR did not change across the 9 repeated measurements, the results of this research can be generalized with greater confidence to the clinical setting.[11]

Regression

Regression is a statistical technique that provides an evaluation of the relationship between a dependent variable with specific independent variables. Regression analysis uses the degree of relationship or correlation between the dependent and independent variables to develop a regression equation that can be used for prediction. It is assumed that the dependent variable is representative of a standard normal distribution. Regression is particularly useful as a statistical method to explain interrelationships among a set of variables, which is why it is so widely used in nursing research. For example, regression analysis might be performed to evaluate results from a study that examined the effects of 3 increasing dosing regimens (continuous variable) of 2 different types of medications or a placebo (categorical variable) intended to lower blood pressure or a placebo on patients' blood pressure levels (continuous variable) following the administration of 3 different medications at 3 doses. The independent variables would be the different doses of the 2 blood pressure medications or a placebo, whereas the dependent variable includes blood pressure readings.

Several different types of regression are commonly used in nursing research including linear regression, multivariate regression, and logistic regression. To develop the regression equation, linear regression assesses the degree of correlation between 2 variables whereas multivariate regression evaluates correlations among several variables. One of the limitations of both linear and multiple regressions is that they must use *dependent* variables that are continuous measures.

Binary or binomial logistic regression uses only 2 dependent variable groups based on categorical data that are mutually exclusive. It has become increasingly common in nursing research.[12] An assumption of logistic regression

Table 7: Significance and Odds Ratios for Final Logistic Regression Model[a]

Variable	β	SE	Wald	*P*	Odds ratio	CI (95%)
RR, breaths/min	0.253	0.200	1.596	.207	1.288	0.870–1.906
MAP, mm Hg	1.354	0.227	35.577	<.001	3.874	2.483–6.046
Temperature, °C	0.754	0.164	21.190	<.001	2.126	1.542–2.930
HR, beats/min	0.184	0.235	0.616	.433	1.202	0.759–1.904

Abbreviations: CI, confidence interval; HR, heart rate; MAP, mean arterial pressure; RR, respiratory rate.
[a]Used with permission from Giuliano KK. Physiologic monitoring for critically ill patients: testing a predictive model for the early detection of sepsis. *Am J Crit Care*. 2007;16:122–130.[10]

is that the model correctly contains all of the relevant predictors and no irrelevant predictors (such as eye color, which is unlikely to have anything to do with a clinical outcome such as septic or not septic). In clinical practice, however, this assumption is rarely met because it is virtually impossible to identify and control all irrelevant predictors. Fortunately, logistic regression is a robust statistical method that is able to withstand violations such as inclusion of some irrelevant predictors.[12]

Binomial logistic regression was used to analyze data from the sepsis research study, which was intended to answer the research question: "Can a combination of the physiologic parameters of heart rate, mean arterial blood pressure, body temperature, and RR measured during the first 24 hours of the critical care admission distinguish between critically ill adult patients who develop sepsis and those who are not diagnosed with sepsis?"[10(p123)]

In this study, the independent variables were recoded as dichotomous variables using the cutoff values recommended by the current clinical practice standards of the Surviving Sepsis Campaign.[13] These dichotomous variables were then entered into a binomial logistic regression model to predict whether or not patients developed sepsis. The final logistic regression model used the 4 dependent variables of HR, MAP, body temperature, and RR and their values correspond with results from the binomial logistic regression analysis presented in Table 7.

Two of the 4 independent variables were significantly and independently associated with being septic: MAP ($P < .001$) and temperature ($P < .001$). Patients who have a MAP of 69 mm Hg or less and a temperature of 38°C or higher during the first 24 hours of intensive care unit (ICU) admission were more likely to develop sepsis than those who did not have blood pressure or temperatures that met these cutoff values. The 2 other variables in the model, RR and HR, were not significant predictors of sepsis diagnosis with *P* values of .207 and .433, respectively.[10]

An indication of the strength of the association between the independent variable and the dependent variable is the odds ratio (OR). The results of the logistic regression provide evidence that patients with a temperature of 38°C or higher during the first 24 hours of ICU admission were approximately twice as likely to be septic than patients with a temperature below 38°C. The other significant predictor, MAP, had an OR of 3.874, which can be interpreted to mean that a MAP of 69 mm Hg or less increased the odds of developing sepsis almost 4-fold. Because a low MAP is one of the most salient physiologic features of sepsis, a high OR for this variable was expected. As these data have shown, septic patients tend to have both clinically and statistically significantly lower blood pressures than equally sick patients with normal MAP. The remaining 2 predictor variables, RR and HR, were not statistically significant, meaning that they were not significant predictors for differentiating between septic and nonseptic patients.[10]

Conclusions

Although both qualitative and quantitative methods are important research techniques to advance nursing knowledge, this article reviewed some of the key concepts related to the use of quantitative research in nursing. The conduct and interpretation of findings from quantitative nursing research represents a thoughtful and scholarly strategy that can yield answers and knowledge to inform the practice of nursing care and promote application of evidence-based research to ultimately

enhance the care of patients. Common quantitative approaches are useful for the study of nursing research issues that have made and will continue to make important contributions to the science of nursing. Although it is not completely unbiased, the quantitative approach provides a level of objectivity that increases our confidence in the conclusions we draw with regard to scientific facts and existing scientific principles. To use these types of data most effectively, a sound understanding of major research principles and statistical methods is required.

References

1. French P. What is the evidence on evidence-based practice? *J Adv Nurs*. 2002;37(3):250–257.
2. Gilmore AS, Zhao Y, Kang N, et al. Patient outcomes and evidence-based medicine in a preferred provider organization setting: a six-year evaluation of a physician pay-for-performance program. *Health Serv Res*. 2007;42(6, pt 1):2140–2149.
3. Giuliano KK, Tyer-Viola L, Palan-Lopez R. Unity of knowledge as a necessity for the development and advancement of nursing knowledge. *Nurs Sci Q*. 2005;18(3):243–248.
4. Giuliano K. Can we bridge the gap between knowledge and clinical practice? *Nurs Philos*. 2003;4(1):44–52.
5. Munro B. *Statistical Methods for Healthcare Research*. New York: Lippincott Williams & Wilkins; 2004.
6. Fowler J, Jarvis P, Chevannes M. *Practical Statistics for Nursing and Health Care*. New York: John Wiley & Sons Inc; 2002.
7. Research Methods Knowledge Database. http://www.socialresearchmethods.net/kb/index.php. Accessed January 17, 2008.
8. Columbia Quantitative Methods in Social Sciences. http://www.columbia.edu/ccnmtl/projects/qmss/index.html. Accessed January 17, 2008.
9. Giuliano KK. *Physiologic Monitoring for Critically Ill Patients: Testing a Predictive Model for the Early Detection of Sepsis* [dissertation]. Chestnut Hill, MA: Boston College; 2005.
10. Giuliano KK. Physiologic monitoring for critically ill patients: testing a predictive model for the early detection of sepsis. *Am J Crit Care*. 2007;16:122–130.
11. Giuliano KK, Scott SS, Brown V, Olson M. Backrest angle and cardiac output measurement in critically ill patients. *Nurs Res*. 2003;52(4):242–248.
12. Grimm L, Yarnold P. *Reading and Understanding Multivariate Statistics*. Washington, DC: American Psychological Association; 1996.
13. Dellinger R, Carlet J, Masur H, et al. Surviving Sepsis Campaign guidelines for the management of severe sepsis and septic shock. *Crit Care Med*. 2004;32(3):858–873.