

Perspectives in Practice

Publishing Nutrition Research: A Review of Sampling, Sample Size, Statistical Analysis, and Other Key Elements of Manuscript Preparation, Part 2

CAROL J. BOUSHEY, PhD, MPH, RD; JEFFREY HARRIS, DrPH, RD; BARBARA BRUEMMER, PhD, RD; SUJATA L. ARCHER, PhD, RD

ABSTRACT

Members of the Board of Editors recognize the importance of providing a resource for researchers to insure quality and accuracy of reporting in the *Journal*. This second monograph of a periodic series focuses on study sample selection, sample size, and common statistical procedures using parametric methods, and the presentation of statistical methods and results. Attention to sample selection and sample size is critical to avoid study bias. When outcome variables adhere to a normal distribution, then parametric procedures can be used for statistical inference. Documentation that clearly outlines the steps used in the research process will advance the science of evidence-based practice in nutrition and dietetics. Real examples from problem sets and published literature are provided, as well as reference to books and online resources.

J Am Diet Assoc. 2008;108:679-688.

This is the second in a series of articles developed to interpret the guidelines for authors submitting manuscripts to the *Journal of the American Dietetic Association* (1) and provide relevant examples and interpretation for how to proceed with manuscript preparation

C. J. Boushey is an associate professor and director, Coordinated Program in Dietetics, Purdue University, West Lafayette, IN. J. Harris is an associate professor and Didactic Program director, Department of Health, West Chester University, West Chester, PA. B. Bruemmer is a senior lecturer, Department of Epidemiology, and director, Didactic Program in Dietetics, University of Washington Graduate Program in Nutritional Sciences, Seattle. S. L. Archer is a research assistant professor, Northwestern University, Feinberg School of Medicine, Department of Preventive Medicine, Chicago, IL.

Address correspondence to: Carol J. Boushey, PhD, MPH, RD, Coordinated Program in Dietetics, Purdue University, 700 W State St, West Lafayette, IN 47907.

E-mail: boushey@purdue.edu

Manuscript accepted: June 11, 2007.

Copyright © 2008 by the American Dietetic Association.

0002-8223/08/10804-0010\$34.00/0

doi: 10.1016/j.jada.2008.01.002

to advance the field of nutrition and its practical applications. Our purpose here is to review study sample selection, sample size, and common statistical procedures using parametric methods, and the presentation of statistical results. The first article in this series (2) focused on study design and the development of testable research hypotheses, the beginning of all good-quality research. A future article will highlight nonparametric statistics. Another in this series will address appropriate measurement tools and methods of analysis, such as sensitivity, specificity, validity, reliability, and relative validity. In addition, issues of judgment, such as making appropriate inferences based on the study design and results, a priori hypothesis testing, post hoc analyses, and extrapolation will be examined in the future articles, as will common epidemiologic methods, including the appropriate use and reporting of odds ratios, relative risk, confidence intervals, and statistical significance as well as the concepts of chance, confounding, and interaction. The goal of the series is to serve as review for some readers and provide new information for others to advance the field of nutrition and its practical applications.

SELECTING A STUDY SAMPLE

The first article in this series (2) clarified that the hypothesis specifies the population being studied. It is rare that an investigator would have the opportunity, time, or resources to measure an entire population of individuals. Consider the hypothesis, introduced previously (2), "There is no statistically significant difference at the $P < 0.05$ level in plasma clotting times among Asian-American men between ages 45 and 60 years taking 3 g/day n-3 fatty acids as combined docosahexaenoic acid and eicosapentaenoic acid in capsule form or a placebo for 6 consecutive weeks." Despite the narrow age range and the focus on a specific group of men, the population represented by this hypothesis still represents at least 840,000 individuals who live in all 50 states (3). Thus, it would be almost impossible to test this hypothesis on the entire population of interest. Thus, researchers will instead recruit a sample of the population to make conclusions about the entire population even if only one location in the United States is used for subject recruitment. The examples outlined below highlight the importance of using sampling methods with minimal bias that results can be assumed to apply to the population beyond the sample.

The difficulty with selecting a sample is ensuring that the sample is representative of the entire target population. The most straightforward approach to achieve this goal is to employ a simple random sample—individuals are selected in such a way that every individual has an equal chance of being selected. Methods of achieving this are available in several resources (4,5).

For the hypothesis above, each treatment group, in a completely randomized experimental design (ie, subjects randomized to receive n-3 fatty acids or subjects randomized to receive placebo), would be selected as part of a simple random sample. If a researcher in San Francisco had a list from the Department of Motor Vehicles of men between ages 45 and 60 years identifying themselves as Asian and residing in the San Francisco Bay area, this list could be used as the basis of selecting a simple random sample. However, those individuals without a driver's license would not be on the list and would not be included in the process of choosing the sample. This is an example of undercoverage. A complete list of any population is rarely available so the researcher needs to be aware of any possible bias (something that can lead to conclusions that are systematically different from the truth) created by those individuals not on the list. The Behavior Risk Factor Surveillance System of the Centers for Disease Control and Prevention conducts interviews based on probability samples of telephone numbers (6). On average, using telephone lists will miss about 6% of the population and usually those individuals without phones are more likely to be low income or homeless and probably differ from the rest of the population.

A more serious source of bias is nonresponse that can occur when those individuals selected through a simple random sample refuse to participate either due to transportation, lack of time, or disinterest. In addition, the recruiter may be unsuccessful at contacting an individual even after several tries. A nonresponse rate >25% would cause concern about potential bias (5). The American Association for Public Opinion Research has developed definitions for response rates, cooperation rates, refusal (nonresponse) rates, and contact rates that are useful in describing overall quality of the final data (7). The response rate would be a fundamental piece of information to include in a manuscript. In addition, when making conclusions about the results of a study, the response rates would be taken into consideration.

When a comprehensive list of the target population is not available, a sampling approach commonly used is to focus recruitment to specific geographic or community areas to capture a large proportion of the target group through posted flyers, newspaper ads, or media ads. This is referred to as a convenience sample. In this case, the subjects choose themselves rather than being randomly selected. These individuals can present some type of systematic bias. This is particularly true for opinion polls where those individuals who have strong opinions either for or against the topic are more likely to participate. Thus, a research study about taste properties of flavored milk would be more likely to attract those individuals that like and drink milk. Nonetheless, the convenience sample is often the only viable method available for most clinical studies. Researchers can minimize the potential bias by constructing a systematic, purposive methodology

of recruitment and outlining these steps in the report. For example, a concerted effort to select consecutively every accessible person that meets the study criteria will help minimize the volunteerism effect.

Another aspect of selecting a study sample is establishing selection criteria as applicable to the research question. These criteria would be specific inclusion criteria or exclusion criteria. In the example above, the research hypothesis dictated the inclusion of the demographic characteristics of Asian origins, men, and a specific age group. Other inclusion criteria to consider would be clinical characteristics, such as generally healthy with no diagnosis of cardiovascular disease. If the research is limited to a specific location then the inclusion criteria may include patients seeking treatment at a specific medical center and even further specification of time such as January to June of a particular year. The exclusion criteria address threats to the subjects or to the quality of the data. For example, recruiting individuals outside of the San Francisco Bay area would be a threat to data because the budget may not support recruitment efforts and transportation for subjects beyond the target area. Thus, excluding individuals outside of the catchment area will help preserve sample size, response rate, and retention. Whereas if an individual is receiving treatment for diagnosed cardiovascular disease, the addition of 3 g/day n-3 fatty acids may interfere with treatment. Thus to avoid putting a subject at risk for side effects, current treatment for cardiovascular disease may be an exclusion criterion.

A researcher is responsible for describing the validity of the sample as appropriate for answering the research question. This would include reference to the sample design, methods of recruitment, the rate of nonresponse, inclusions and exclusions, and the final sample size that is large enough to meet the study needs. These factors need to be considered to make conclusions about how much the sample can be generalized to the population. Lohse and colleagues (8) conducted an extensive survey among women aged 18 years and older participating in the Special Supplemental Nutrition Program for Women, Infants, and Children. The study sample used was a convenience sample; however, the authors clearly outline the eligibility criteria and the purposive sampling plan employed to minimize bias and best represent the Special Supplemental Nutrition Program for Women, Infants, and Children population. When biological factors are examined in observational and experimental studies, generalizing the results to a wider population is more acceptable than descriptive studies that enumerate the distribution of a factor in a sample population (5). For example, the strength of fruits and vegetables as a risk modifier for certain cancers tends to be more consistent among diverse populations than the prevalence of individuals consuming five or more fruits and vegetables per day. Therefore, the decision to generalize results from a single sample to a wider population requires consideration of many issues including sampling design, participation rates, and biological processes.

IMPORTANCE OF ESTIMATING SAMPLE SIZE

One of the common mistakes in research is failure to estimate an appropriate sample size before embarking on

Values needed to estimate sample size	Example
Choose the main endpoint of interest and the method by which it will be measured.	Difference in average calcium intakes from food between two groups using a two-sample <i>t</i> test.
Specify the size of the difference between the experimental groups that is meaningful to detect.	Meaningful difference would be 225 mg, an amount equivalent to about 6 oz milk.
Estimate the expected variability (ie, the estimated standard deviation).	Based on pilot data the standard deviation for a group that likes milk is 540 mg and the group that doesn't like milk is 430 mg.

Figure 1. Assumptions needed about the conditions of a study to complete sample size calculations.

a research project. If the sample size is too small even the best study cannot detect an important effect and this can contribute to further confusion surrounding a topic. The process of estimating sample size can be technically complex and a statistician can assist with this process. Researchers will find useful a reference written specifically for nutrition and dietetics that outlines seven steps to estimating sample size (9), as well as a reference directed to clinical studies (10,11). In addition, there are sample size calculators available online. One example of such a site is one that was created with support from the National Institutes of Health General Clinical Research Center Program (http://hedwig.mgh.harvard.edu/sample_size/size.html). Other programs can be found by using an online search engine to find “sample size calculators.”

Access to these calculators is a wonderful convenience; however, it does not preclude a researcher from completing the steps to determine the data needed for the calculations. To prepare for a visit to a statistician or a Web site calculator, a researcher needs to complete at a minimum the steps outlined in Figure 1. The researcher is the most qualified individual to state the outcome to be measured, how the outcome will be measured, what will be a meaningful result, and how much variation exists in the selected measure. A statistician can assist with the selection of an appropriate statistical test, as well as provide guidance with regard to choosing an appropriate level of error and power for the study. A statistician cannot provide advice unless the researcher has either completed a pilot study or extracted from the published literature information from other studies measuring the same or similar outcomes.

DATA ANALYSIS

Descriptive Statistics

Before data analysis, all data need to be checked even if data entry involved verification methods (dual entry), scanning, computer entry, or Web-based entry. Useful first steps are to run frequency analyses of every variable

Table. Sample characteristics of women participating in rural health screening program (N=289)^a

Characteristic	Mean ± SD ^b	
Age (y)	48 ± 12	
Weight (lb)	160.6 ± 37.3	
Height (in)	64.1 ± 2.4	
Body mass index	27.8 ± 6.0	
Percent body fat	33.4 ± 8.0	
Waist (in)	33.3 ± 5.5	
	<i>n</i>	%
Education		
High school or less	46	17.6
Some college	67	25.7
College graduate	148	56.7
Annual household income		
<\$30,000	81	31.6
\$30,000-\$49,999	83	32.4
≥\$50,000	92	35.9
Marital status		
Married	197	69.6
Not married	86	30.4
Supplement users		
Any supplement	215	74.4
Supplement containing folic acid	156	54.0

^aFrequencies may vary due to missing values.
^bSD = standard deviation.

Figure 2. Example of a table that outlines the characteristics of a sample population. Often this is the first table in a manuscript that describes the study sample (eg, age, sex, and body mass index). In this example, quantitative variables are summarized as means and standard deviations. Because age was collected as a whole year, the mean is reported as a whole year. Categorical variables are summarized as frequencies and proportions. Data from reference 13.

and then review the results to ensure that the output matches the expected values for each variable. Errors, implausible values, and outlier values need to be checked and any errors need to be corrected. A height of 93 in among 5-year-olds is outside of the expected range of 39 to 47 in and most likely represents an error that needs to be corrected before data analysis can begin. If a more realistic value is not available and the value of 93 in is considered biologically implausible, then the value is best changed to missing. On the other hand, if the value is checked and it is within the realm of reality (eg, 49 in) then it would be considered an outlier. Because an outlier represents an observational value, any analyses should include the outlier value. Under these circumstances, analysis methods to consider using include separation into groups, such as quartiles, and nonparametric analytical methods (a topic of a future article in this series). As an alternative, results can be presented with and without outliers. Once the data have been checked and edited, the first step of analysis is to simply look at the data using descriptive statistics. The purpose of this step

is to become familiar with the data and to create a description of the study population (12).

For the quantitative variables, this might be plots or frequency histograms. Use of stemplots to examine the shape of a distribution and to detect outliers is thoroughly discussed by Moore and McCabe (12). There are no simple rules for dealing with outliers in data unless, of course, the outlier represents an error; in which case it may be removed. Otherwise the researchers need to com-

municate clearly any decisions regarding the handling of an outlier. Other summary statistics would be means, medians, standard deviations, and range of values. For categorical variables, frequency distributions would be completed.

The data may be more meaningful by creating groups and calculating the proportion of subjects in each group (eg, overweight and not overweight). The results of this step often become the first table in a manuscript that outlines the characteristics of the study sample (eg, sex, age, and body mass index). See Figure 2 for an example of a characteristics table adapted from Boushey and colleagues (13). Some sample populations can be more complex than the example provided here and as a result the characteristics table needs to be stratified by sex, intervention groups, ethnicity (14), or study site (8). Examination of the data is a fundamental piece of information to include in a manuscript. For example, Larson and colleagues (15) included in the statistical analyses section, "To examine the characteristics of adolescents . . . , descriptive statistics were calculated on the adolescents who responded to all of the measures used in the analysis."

Inferential Statistics

Upon completing a study a researcher wants to ultimately discover the outcome of the hypothesis (eg, Did the intervention make a difference? Are vitamin levels in location A different from vitamin levels in location B?) An observed effect that is so large that it would rarely occur by chance is called statistically significant. To make a conclusion about a result being statistically significant, the appropriate statistical test needs to be completed.

Before data collection, a researcher would have planned the inferential statistics to be used for data analysis (16,17). There are flowcharts available that assist with the decision-making process for which statistics to

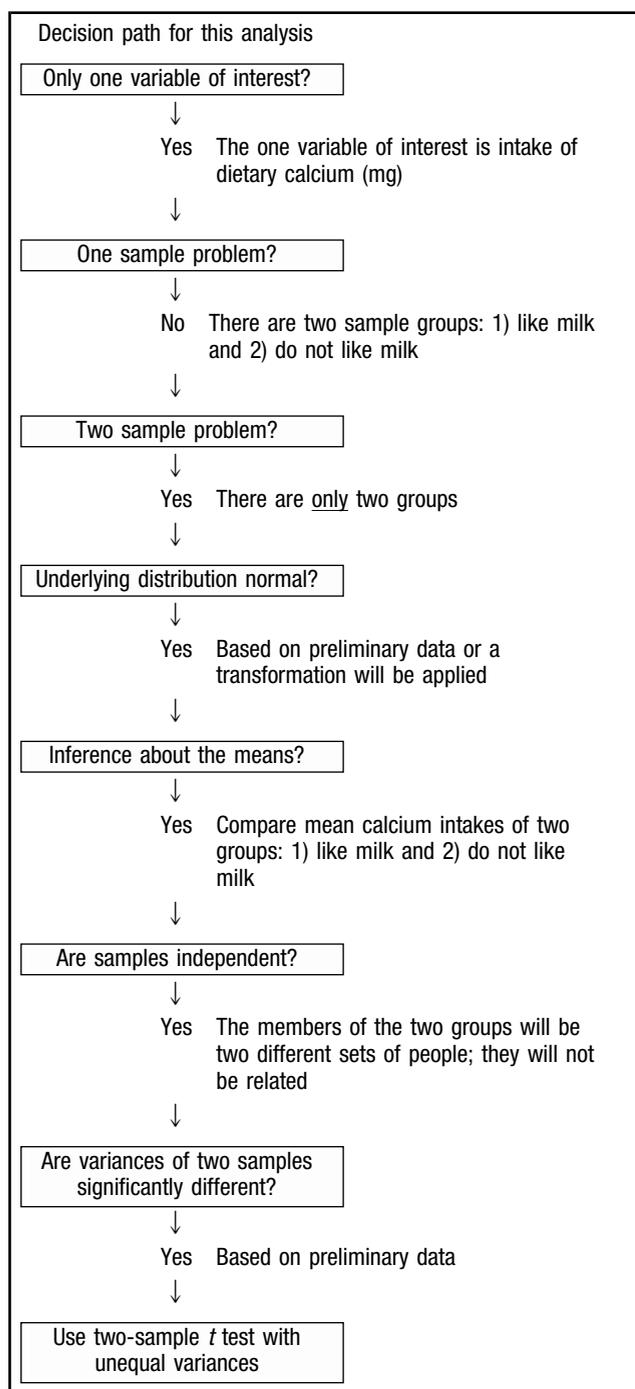
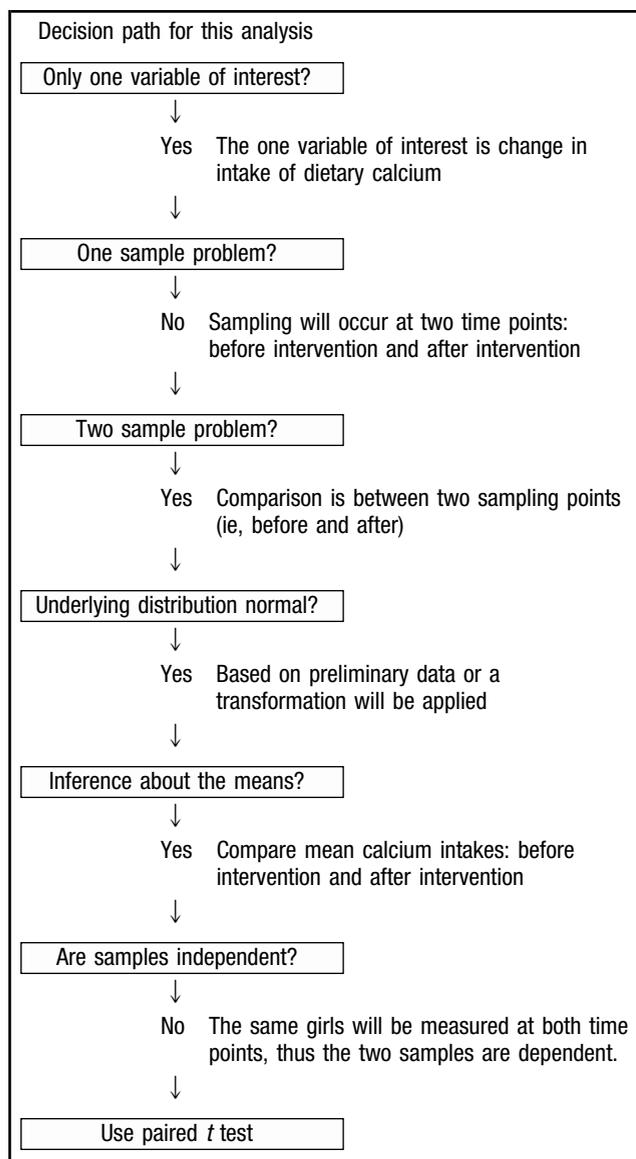


Figure 3. Selecting an inferential statistic by using a flowchart decision tree (see references 18 and 19). Assume a researcher is interested in finding if the average dietary calcium intake among girls aged 10 to 12 years that like milk is significantly different (defined as a $P < 0.05$) from girls who do not like milk. The primary outcome will be the estimate of dietary calcium intake (in milligrams) collected from girls using a previously tested food frequency questionnaire for estimating calcium intake. The exposure variable of interest is preference for milk that will be determined from a questionnaire developed to measure preference for milk. A total of 250 girls complete the questionnaires. Based on the decision tree, the researcher confirms a normal distribution (see Figure 3 for distribution plot) and confirms that the variances of the two samples are significantly different by using an F test (reference 30, p 318). Thus, the two-sample t test (also called an independent t test) is used to assess whether the average calcium intakes of the two groups are significantly different. The investigator finds that the P value for the two-tailed t test for equality of means is $P < 0.001$. Therefore, the investigator concludes that in this sample, the mean calcium intakes of girls who like milk (mean $1,106 \pm 548$ mg) is significantly higher than the mean calcium intake of girls who do not like milk (mean 748 ± 440 mg). It would be of interest to explore if providing flavored milk would change the calcium intakes of those girls who do not like milk.

use for analysis (18). The decision trees in these flowcharts are instructive because they highlight the importance of determining in advance the hypothesis, study design, and types of variables to be measured (these concepts were covered in the first article in this series [2]). There are flowcharts for methods of statistical inferences available in several resources (18,19) and an interactive online flowchart is available from the Institute for Social Research at The University of Michigan: (<http://www.microsirris.com/Statistical%20Decision%20Tree/>). Two examples of following a decision tree are outlined in Figures 3 and 4. A review of several concepts, such as normal distributions and independent or dependent samples will assist with appreciating the importance of making appropriate choices for statistical inference.



Density Curves and Normal Distributions

The most well-recognized measure of central tendency is the arithmetic mean or average. For the following five serum low-density lipoprotein (LDL) cholesterol level measurements: 74 mg/dL, 94 mg/dL, 113 mg/dL, 121 mg/dL, and 135 mg/dL, the mean is 107 mg/dL. The mean represents the value of the five values added together and divided by the number of observations. When summarizing data it is appropriate to report the mean when the data are normally distributed. Statistical tests that compare means, such as two-sample *t* tests and analysis of variance (ANOVA), are called parametric tests and assume that the data from the samples being compared are normally distributed. Adherence to a normal distribution can be evaluated by creating a density curve or plotting the data as a histogram and overlaying a normal curve as shown in Figure 5A for estimated dietary calcium intake among sixth-grade girls. Computer programs (Figure 6) enable researchers to easily examine whether data adhere to parameters of a normal distribution vs tedious plotting by hand. An alternative to the histogram is a normal probability plot that plots a variable's cumulative proportions against the cumulative proportions of a normal distribution. If the selected variable adheres to the normal distribution, the points cluster around a straight line as shown in Figure 5B for the same data shown in Figure 5A.

When the data are not normally distributed with skewed distributions or extreme values, the mean is not a very good measure of central tendency. For example, in the LDL cholesterol level example given above if the last listed value, 135 mg/dL, is changed to 180 mg/dL, the mean changes dramatically to 116 mg/dL. In this case, the median is a better descriptor of the data. The median is the data value that splits the data array in half. Half of the data values are below the median and half above it.

Figure 4. Selecting an inferential statistic by using a flowchart decision tree (see references 18 and 19). Assume a researcher is interested in finding if the average dietary calcium intake among sixth-grade girls who consume school lunch can increase significantly (defined as $P < 0.05$) if flavored milk is offered. The primary outcome will be the change in dietary calcium intake (in milligrams) from before the addition of flavored milk to after the addition of flavored milk. A total of 148 girls in sixth grade were offered flavored milk. Another 128 girls in the sixth grade were not offered flavored milk with their school lunches. Based on the decision tree, the researcher used a paired *t* test to determine whether the mean difference in calcium intakes of the girls increased significantly compared to zero, or no difference. The changes in each group of girls, those exposed to flavored milk and those girls not exposed to flavored milk, were analyzed separately. Using a computer program, the mean difference for calcium intake before and after the intervention was found to be 129 mg calcium/d for those girls exposed to flavored milk. The investigator notes that the *P* value for the two-tailed *t* test was $P = 0.001$. The mean difference for calcium intake before and after the same time period for those girls not exposed to flavored milk was found to be 6 mg calcium/d. The investigator noted that the *P* value for the two-tailed *t* test was $P = 0.914$. Therefore, the researcher concluded that the introduction of flavored milk in the school cafeteria was associated with a statistically significant positive change in calcium intake among sixth-grade girls.

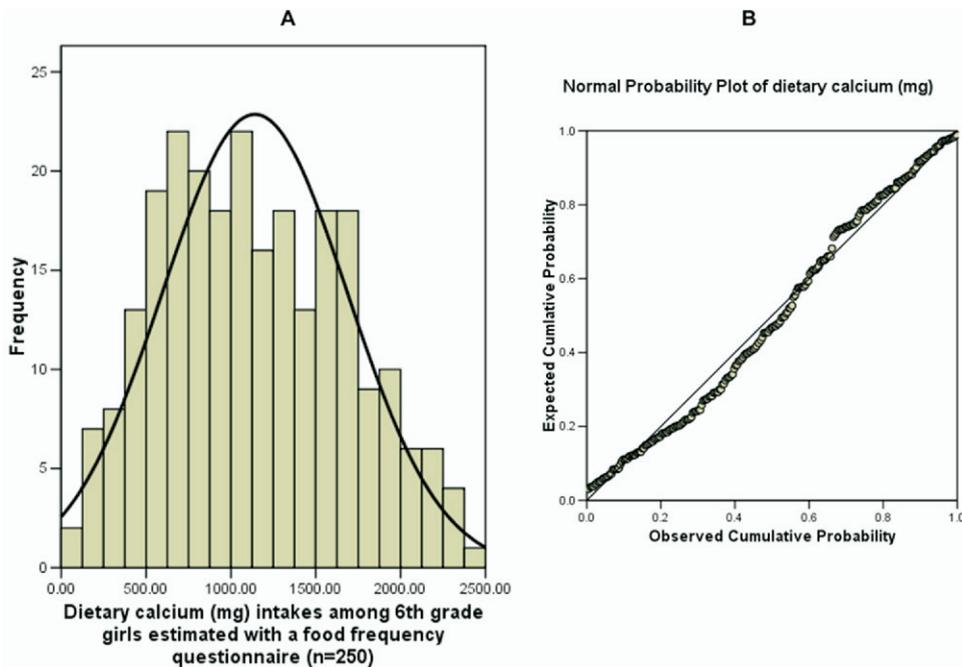


Figure 5. Adherence to a normal distribution can be evaluated by creating a density curve or plotting the data as a histogram and overlaying a normal curve as shown in A for estimated dietary calcium (mg) among sixth-grade girls. Alternatively, for a normal probability plot, if the selected variable adheres to the normal distribution, the points cluster around a straight line as shown in (B) for the same data as shown in (A).

SPSS	SPSS Inc 233 S. Wacker Dr 11th Fl Chicago, IL 60606 312/651-3000	www.spss.com
SAS	SAS Institute, Inc 100 SAS Campus Dr Cary, NC 27513-2414 919/677-8000	www.sas.com
Minitab	Minitab Inc Quality Plaza 1829 Pine Hall Rd State College, PA 16801-3008 800/448-3555	www.minitab.com
STATA	StataCorp, LP 4905 Lakeway Dr College Station, TX 77845 800/782-8272	www.stata.com
GraphPad Software	GraphPad Software, Inc 11452 El Camino Real, #215 San Diego, CA 92130 858/259-5770	www.graphpad.com

Figure 6. Names and contact information for commonly used statistical software packages.

For the LDL cholesterol examples above, the median is 113 mg/dL. Notice that it does not change with the addition of an extreme value. When data are not normally distributed it is important to report medians not means, unless the data can be transformed to normality successfully by using log or trigonometric transformations. By applying a function, such as the logarithm, data can be transformed to a more normal distribution. Dietary and nutrient data are often skewed as shown in [Figure 7](#) (A1

and B1). For a researcher to proceed with parametric statistical inference, it is necessary to transform the data by applying a function such as the logarithm or a cubed root. There are systematic principles that describe how transformations perform and can speed the process of applying an appropriate transformation (20,21). For example, right-skewed data as shown in [Figure 7](#) (A1 and B1) can usually be transformed to a more normal distribution with the use of the natural logarithm as shown in [Figure 7](#) (A2 and B2). It is important to note that the mean of the transformed data must be reported in this case and not the mean of the raw, untransformed data. This transformed mean is often not very meaningful for a reader, thus reporting of the median can be very useful.

If the data are not normally distributed and cannot be successfully transformed then parametric tests cannot be used. Nonparametric tests must be used, such as the Mann-Whitney test or the Kruskal-Wallis test that use proportions or rankings as the measures for comparison. As an alternative, the data can be divided into groups to create a categorical field. For categorical data, the χ^2 test (independent samples) or the McNemar test (dependent samples) can be used. Assumptions and uses of nonparametric tests and tests using the binomial distribution will be a topic of a future article in this series.

A common mistake researchers make is not testing their data for normality. This can sometimes result in authors reporting means and standard deviations and using parametric tests for inferential statistics when the data are not normally distributed. Thus, the results are not valid. The analytical step of inspecting the data for adherence to a normal distribution is a fundamental

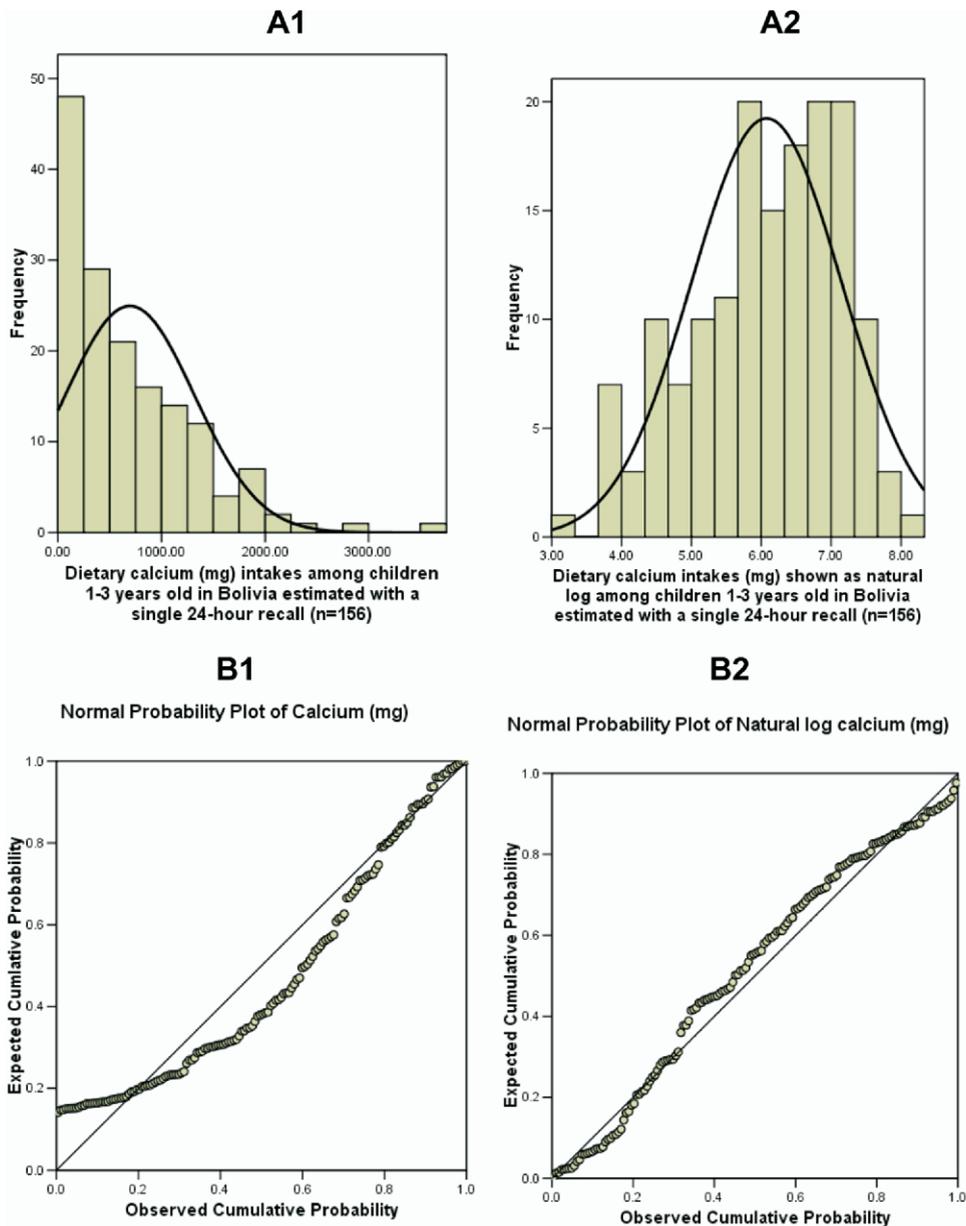


Figure 7. Nutrient intake data often do not adhere to normal distribution assumptions. As shown in (A1), the calcium intake data are skewed to the right and the data points do not cluster around the straight line on the normal probability plot as shown in (B1). Data can be transformed by applying a function such as the logarithm that can make a distribution more normal as shown in (A2) and (B2) for the same data.

piece of information to include. For example, authors can write, “Normal probability plots were used to assess the need for transformations. No variable required a transformation.” Or, if a variable needed transformation, specify the variable and transformation function used.

It is important to note that standard deviations also lose their relevancy if the data are not normally distributed. Standard deviations use all the data, including extreme values and means in the calculation. It is not appropriate to report standard deviations for data that do not meet the assumptions of a normal distribution.

Comparing Means of Independent and Dependent Samples

In the first example of using the flowchart decision tree (Figure 3), the exposure of interest was a categorical variable that created two distinct groups: girls who like milk and girls who dislike milk. The outcome of interest was dietary calcium intake, which is a quantitative variable. Thus, before applying a statistical test it was essential to determine if the calcium data met the assumptions of a normal distribution. For this particular case, the recommended statistical inference test was the two-sample *t* test or independent samples *t* test (same test, differ-

Arrangement of data for a study with a dichotomous exposure and a quantitative outcome variable among independent samples (Data for three observations shown. See Figure 3)			Arrangement of data for a study using before and after quantitative measures among the same people (ie, dependent samples) (Data for three observations shown. See Figure 4)		
ID	Exposure Variable	Outcome Variable	ID	Outcome Variable Calculated as Difference by Computer Program	
	Dislike milk=0 Like milk=1	Calcium intake (mg)		Calcium intake (mg) before intervention	Calcium intake (mg) post-intervention
1	1	835	1	303	509
2	0	531	2	779	922
3	1	1,193	3	530	484

Figure 8. Data tip: Example of data preparation for use with computer programs when the study design involves independent samples vs dependent samples.

ent name). This test is used when analysis involves a dichotomous categorical variable and a quantitative variable. This is a common test that is used in dietetics and nutrition research. For example, the two-sample *t* test would be used to compare means of glycated hemoglobin (quantitative variable) between two groups (dichotomous categorical variable) of individuals with type 1 diabetes; one group using a traditional insulin injection regimen and one group using a tightly controlled, multiple injection strategy. Examples of articles that used the two-sample *t* test can be found in work by Lohse and colleagues (8) and Gaetke and colleagues (22).

Sometimes the two-sample *t* test is confused with the paired *t* test. These are not the same. A paired *t* test is used when comparing a quantitative variable with related samples. Examples of this include measurements of the same patients before and after an intervention or pretest and posttest scores of students participating in a nutrition education session. See Figure 4 for the decisions that occur in the flowchart decision tree when working with dependent samples. Another example of using the paired *t* test is provided by Klohe-Lehman and colleagues (14) when comparing pre- and postintervention scores for nutrition knowledge.

Due to the availability of computer programs designed to calculate a wide variety of statistical tests (Figure 6), much of the burden in completing descriptive and inferential statistics has been reduced. Yet these programs do not have built-in systems to check if the statistical test used is appropriate. A researcher is still responsible for ensuring that the statistical tests used are appropriate based on the research design, sampling frame, the data distributions, and the outcomes. More importantly, the interpretation of any statistical test can only be made by a researcher; this is not provided by the computer program. Any data preparation to transform variables or recode a quantitative variable to a categorical variable is still the task of an investigator. For example, the data preparation will differ when study design involves independent samples vs dependent samples as shown in Figure 8. Authors need to include information about any computer program used, including program name, version, version release date, company name, company location.

It is important to appreciate that certain assumptions must be met when conducting *t* tests (or any test, for that matter). For the two-sample *t* test, the data from each sample must be normally distributed or be mathematically transformed as such. For the paired *t* test, the difference in the before and after measures must be normally distributed. These tests are robust when it comes to minor deviations from normality; however, nonparametric versions must be used when the violation of this assumption is more substantial (23).

Comparing Two Quantitative Variables

Correlation is the measure to use when looking for a potential relationship between two quantitative variables. A common correlation of interest in research is the relationship between two similar measures to see if one can be substituted for another; for example, dietary calcium intake estimated from a food frequency questionnaire vs 24-hour food recalls (24). A correlation is often calculated when conducting cross-sectional research because relationships between variables are analyzed at one isolated point in time. Correlations do not distinguish between the explanatory variable and the response variable; rather, it quantifies the relationship.

Correlations measure the direction and magnitude of a relationship between two quantitative variables by calculating the correlation coefficient. The Pearson correlation coefficient is usually written as *r* and has an absolute value and a sign. The absolute value represents the magnitude of the association between variables. The sign of *r* indicates the direction of the relationship. Correlation coefficients can range from -1.0 to 1.0. The larger the absolute value of the coefficient the stronger the relationship. The interpretation of the coefficient is dependent on the discipline and the variability that exists in the items being measured. A correlation of 0.9 may be very low under the condition of verifying a physical law using high-quality instruments, but may be regarded as very high in the social sciences where few factors can actually be controlled. Further, the strength of a correlation varies with sample size. At large sample sizes, smaller correlations can be statistically significant. Thus, set cutpoints

for interpretation of coefficients are in some ways arbitrary and should be used with discretion.

If the sign is negative it means as one variable increases the other tends to decrease and this can be described as a negative or inverse correlation. If the sign is positive there is a direct or positive relationship, meaning that as one variable increases the other variable also increases. A coefficient of -1.0 or 1.0 means there is a perfect correlation between the variables, either inverse or direct (25,26).

As with other parametric tests, assumptions must be met to conduct the test. The values for each variable being compared must be normally distributed or transformed as such. When describing the results of the correlation between two variables, provide the means and standard deviations of both variables, as well as the r value and the direction of the relationship as either positive or negative (25).

Methods for Comparing More Than Two Groups

One-way ANOVA is used when there is a situation in which a relationship is being examined between a quantitative dependent variable and an (independent) categorical variable. Typically, if the categorical variable has two groups, a two-sample t test is used as discussed above. This test of statistical inference is used to test the hypothesis that several means are equal. For example, a one-way ANOVA would be used to compare mean hemoglobin A1c (HbA1c) values between persons with type 2 diabetes receiving an exercise program, a low-glycemic-index diet, metformin, or a placebo. The ANOVA result only determines if there is a statistically significant difference somewhere between the means, but not which groups are different from one another. If the test statistic result (F statistic) for an ANOVA is significant (eg, $P < 0.05$) then the investigator can proceed with completing post hoc tests, which can distinguish which specific groups have statistically significantly different means. Common post hoc tests include the Bonferroni, Scheffe, and Tukey tests. There are principles to apply when selecting an appropriate post-hoc test to determine which specific groups are statistically significantly different (27). Consult Cheney (28) for a step-by-step sequence of analysis using ANOVA accompanied by instructive visuals.

There are other ANOVA statistics that are used when the research question is more complex and requires more than a single variable. These techniques are called multifactor or multiway ANOVA, repeated measures ANOVA, and multivariate ANOVA (MANOVA). All types of ANOVA assume that the samples compared are normally distributed and variances between samples are equal. If there is too much deviation from these assumptions non-parametric versions of these tests must be used. Some of these ANOVA methods are not necessarily covered in an introductory statistics class. Thus, researchers are encouraged to consult a statistician if their research questions are similar to the scenarios outlined below. For those individuals familiar with these tests that merely desire a refresher, there are comprehensive textbooks (29,30) available and there are Web sites that provide information (eg, www.stat-help.com, www.statsoft.com/textbook/

stanman.html#index, and www.socialresearchmethods.net/kb/).

Multifactor ANOVA is used when a relationship is being examined between a continuous variable and more than one categorical variable. For example, continuing with the diabetes groups above, examining differences in HbA1c levels between treatment groups as well as African Americans and non-Hispanic whites may also be important. With the multifactor ANOVA, the independent effects of the treatment groups and the race/ethnicity status as well as their joint effects can be examined. The independent effects are called main effects and the joint effects are referred to as interactions. If it is found that the interactions are statistically significant, then the relevance of the main effects is mute. For example, if metformin is shown to significantly lower HbA1c for the African-American group, but not the non-Hispanic white group, then the general question, "Is metformin a more effective treatment?" becomes irrelevant. The results would indicate that the effectiveness is dependent on whether the subject was in the African-American group or the non-Hispanic white group. Similar to one-way ANOVA, post hoc tests need to be used to find where specific differences between groups exist.

Repeated measures ANOVA is used to compare changes in a continuous variable over time or changes in a group of subjects when different treatments are used. If a researcher wanted to look at the effect of different diets on systolic blood pressure, one approach might be to recruit 50 people who would consume four diets in a randomized order. Each diet would be consumed for 2 weeks followed by a 2 weeks washout period. The four diets might be a typical American diet, the Dietary Approaches to Stop Hypertension diet, the Mediterranean diet, and a high animal protein/low carbohydrate diet. Systolic blood pressure would be measured for each person at the beginning of each diet period, specified times within each diet period, and at the end of each diet period. Repeated-measures ANOVA could be used to look at the differential effect of the diets on systolic blood pressure. Because all subjects would receive all diets, there would be a need to account for the variation within each subject. Repeated-measures ANOVA accounts for this within- and between-subject variations. Just as with one-way ANOVA, post hoc tests are needed to determine where the differences between diets exist. The effectiveness of interventions are often assessed with repeated measures ANOVA (14,22).

MANOVA is used when a relationship between one or more categorical variables and more than one continuous variable are being examined. Using one of the examples given above, the research question now expands to examining the relationship between different treatment modalities (independent variables) for type 2 diabetes (ie, exercise program, a low-glycemic-index diet, metformin, or a placebo) and both HbA1c and LDL cholesterol levels (dependent variables). This multivariate statistical test allows for a result with one statistical test rather than completing multiple tests. The changes in HbA1c and serum LDL cholesterol levels could be tested separately; however doing multiple tests on the same sample can increase the chances of committing a Type 1 error. Thus the completion of the single MANOVA decreases the risk of a Type 1 error. With MANOVA, post-hoc tests will need

to be conducted to ferret out the specific categorical differences. A description of the variables used and the inferential statistics is important. Rousset and colleagues (31) used MANOVA to assess the effectiveness of nutrition messages and sex to change the consumption of six protein-rich foods.

CONCLUSIONS

All research begins with a research question that is the precursor to a testable hypothesis. Implementation of study design, sample size calculations, sampling methods, and inferential statistics provides the basis to assess if the hypothesis is true or not. The opportunities to develop meaningful research hypotheses occur frequently in the field of dietetics. Food and nutrition professionals are encouraged to pursue research methods to assist with building an evidence base for practice. The final step in research is preparing a report that may enter the peer-review process as a manuscript. To make the publishing process less intimidating, prospective authors can follow the information outlined in this series of articles that is devoted to publishing nutrition research.

Data presented in this article come from actual research problems and have been modified to illustrate the ideas presented.

References

- American Dietetic Association. *Journal of the American Dietetic Association* guidelines for authors. *J Am Diet Assoc.* 2006;106:140-147.
- Boushey C, Harris J, Bruemmer V, Archer SL, Van Horn L. Publishing nutrition research: A review of study design, statistical analyses, and other key elements of manuscript preparation, Part 1. *J Am Diet Assoc.* 2006;106:89-96.
- Race and Hispanic or Latino origin by age and sex for the United States: 2000 (PHC-T-8). US Census Bureau, Population Division, Racial Statistics Branch Web site. <http://www.census.gov/population/www/cen2000/phc-t08.html>. Accessed January 19, 2006.
- Moore DS, McCabe GP. Producing data. In: Moore DS, McCabe GP, eds. *Introduction to the Practice of Statistics*. New York, NY: WH Freeman and Co; 2006:191-250.
- Hulley SB, Newman TB, Cummings SR. Choosing the study subjects: Specification, sampling, and recruitment. In: Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB, eds. *Designing Clinical Research: An Epidemiological Approach*. Philadelphia, PA: Lippincott Williams & Wilkins; 2001:25-35.
- Behavioral Risk Factor Surveillance System operational and user's guide. Centers for Disease Control and Prevention Web site. <http://www.cdc.gov/brfss/pdf/userguide.pdf>. Accessed January 19, 2006.
- Standard definitions: Final dispositions of case codes and outcomes rates for surveys. The American Association for Public Opinion Research Web site. http://www.aapor.org/pdfs/standarddefs_4.pdf. Accessed January 30, 2007.
- Lohse B, Stotts JL, Priebe JR. Survey of herbal use by Kansas and Wisconsin WIC participants reveals moderate, appropriate use and identifies herbal education needs. *J Am Diet Assoc.* 2006;106:227-237.
- Cheney CL, Boushey CJ. Estimating sample size. In: Mosen ER, ed. *Research: Successful Approaches*. Chicago, IL: American Dietetic Association; 2003:389-398.
- Browner WS, Newman TB, Hearst N, Hulley SB. Getting ready to estimate sample size: Hypotheses and underlying principles. In: Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB, eds. *Designing Clinical Research: An Epidemiological Approach*. Philadelphia, PA: Lippincott Williams & Wilkins; 2001:51-63.
- Browner WS, Newman TB, Cummings SR, Hulley SB. Estimating sample size and power: The nitty-gritty. In: Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB, eds. *Designing Clinical Research: An Epidemiological Approach*. Philadelphia, PA: Lippincott Williams & Wilkins; 2001:65-91.
- Moore DS, McCabe GP. Looking at data—Distributions. In: Moore DS, McCabe GP, eds. *Introduction to the Practice of Statistics*. New York, NY: WH Freeman and Co; 2006:3-99.
- Boushey CJ, Edmonds JW, Welshimer KJ. Estimates of the effects of folic-acid fortification and folic-acid bioavailability for women. *Nutrition.* 2001;17:873-879.
- Klohe-Lehman DM, Freeland-Graves J, Anderson ER, McDowell T, Clarke KK, Hanss-Nuss H, Cai G, Puri D, Milani TJ. Nutrition knowledge is associated with greater weight loss in obese and overweight low-income mothers. *J Am Diet Assoc.* 2006;106:65-75.
- Larson NI, Story M, Eisenberg ME, Neumark-Sztainer D. Food preparation and purchasing roles among adolescents: Associations with sociodemographic characteristics and diet quality. *J Am Diet Assoc.* 2006;106:211-218.
- Cheney CL, Mosen ER. Statistical analysis, data presentation, conclusions, and applications. In: Mosen ER, ed. *Research: Successful Approaches*. Chicago, IL: American Dietetic Association; 2003:29-35.
- Glantz SA. Front cover. In: Glantz SA, ed. *Primer of Biostatistics*. New York, NY: McGraw-Hill; 1997:0.
- Rosner B. Flowchart: Methods of statistical inference. *Fundamentals of Biostatistics*. Belmont, CA: Duxbury/Thomson Brooks/Cole; 2006: 849-854.
- Appendix C. In Dawson B, Trapp RG. *Basic and Clinical Biostatistics*. New York, NY: McGraw-Hill; 2001:381-385.
- Moore DS, McCabe GP. Looking at data—Relationships. In: Moore DS, McCabe GP, eds. *Introduction to the Practice of Statistics*. New York, NY: WH Freeman and Co; 2006:143-145.
- Hassard TH. *Understanding Biostatistics*. St Louis, MO: Mosby Year Book; 1991:275.
- Gaetke LM, Stuart MA, Truszczynska H. A single nutrition counseling session with a registered dietitian improves short-term clinical outcomes for rural Kentucky patients with chronic diseases. *J Am Diet Assoc.* 2006;106:109-112.
- Moore DS, McCabe GP. Inference for distributions. In: Moore DS, McCabe GP, eds. *Introduction to the Practice of Statistics*. New York, NY: WH Freeman and Co; 2006:462-463.
- Jensen JK, Gustafson D, Boushey CJ, Auld G, Bock MA, Bruhn CM, Gabel K, Misner S, Novotny R, Peck L, Read M. Development of a food frequency questionnaire to measure calcium intake of Asian, Hispanic, and white youth. *J Am Diet Assoc.* 2004;104:762-769.
- Moore DS, McCabe GP. Looking at data—Relationships. In: Moore DS, McCabe GP, eds. *Introduction to the Practice of Statistics*. New York, NY: WH Freeman and Co; 2006:123-127.
- Rosner B. Continuous probability distributions. *Fundamentals of Biostatistics*. Belmont, CA: Duxbury/Thomson Brooks/Cole; 2006:142-145.
- Rosner B. Multisample inference. *Fundamentals of Biostatistics*. Belmont, CA: Duxbury/Thomson Brooks/Cole; 2006:557-591.
- Cheney CL. Statistical applications. In: Mosen ER, ed. *Research: Successful Approaches*. Chicago, IL: American Dietetic Association; 2003:399-416.
- Moore DS, McCabe GP. *Introduction to the Practice of Statistics*. New York, NY: WH Freeman and Co; 2006.
- Rosner B. *Fundamentals of Biostatistics*. Belmont, CA: Duxbury/Thomson Brooks/Cole; 2006.
- Rousset S, Croit-Volet S, Boirie Y. Change in protein intake in elderly French people living at home after a nutritional information program targeting protein consumption. *J Am Diet Assoc.* 2006;106:253-261.