Perspectives in Practice

Publishing Nutrition Research: A Review of Study Design, Statistical Analyses, and Other Key Elements of Manuscript Preparation, Part 1

CAROL BOUSHEY, PhD, MPH, RD; JEFFREY HARRIS, DrPh, MPH, RD, CNS; BARBARA BRUEMMER, PhD, RD; SUJATA L. ARCHER, PhD, RD; LINDA VAN HORN, PhD, RD

ABSTRACT

To enhance the *Journal's* position as the premier source for peer-reviewed research in the science of food, nutrition, and dietetics, members of the Board of Editors recognize the importance of providing a resource for researchers to ensure quality and accuracy of reporting in the *Journal*. This first monograph of a periodic four-part series focuses on the study hypothesis, study design, and collaboration with a statistician. The basics of study design start with a clear hypothesis or research question and a definitive outcome measure. Throughout the development of a research project, the questions of what is to be discovered and why the research is being conducted need to be addressed. Decisions about parameters to measure and study design most appropriate to test a hypothesis create the foundation for future conclusions. Collaboration with a statistician can aid in the research development process. Documentation that coherently communicates the research process will advance the science of evidence-based practice in nutrition and dietetics. Real examples from published literature are provided, as well as references to books and online resources.
J Am Diet Assoc. 2006;106:89-96.

C. Boushey is an associate professor and director, Coordinated Program in Dietetics, Purdue University, West Lafayette, IN. J. Harris is associate professor and director, Didactic Program in Dietetics, Department of Health, West Chester University, West Chester, PA. B. Bruemmer is senior lecturer, Department of Epidemiology and director, Didactic Program in Dietetics, and technical advisor, Center for Public Health Nutrition, University of Washington, Graduate Program in Nutritional Sciences, Seattle. S. L. Archer is a research assistant professor, Department of Preventive Medicine, Northwestern University, Feinberg School of Medicine, Chicago, IL. L. Van Horn is editor-in-chief, Journal of the American Dietetic Association, and professor of preventive medicine, Northwestern University, Feinberg School of Medicine, Chicago, IL.

Address correspondence to: Carol Boushey, PhD, MPH, RD, Purdue University, 700 West State St, West Lafayette, IN 47907. E-mail: boushey@purdue.edu

0002-8223/06/10601-0016$32.00/0
doi: 10.1016/j.jada.2005.11.007

Research plays a vital role in the practice and advancement of dietetics. The importance of research is reflected in the standards of professional practice, which include "Each dietetics professional effectively applies, participates in, or generates research to enhance practice" (1). In the training of future dietetics professionals, integration of research is emphasized in four knowledge, skills, or competency areas (2). Paramount with documenting the effectiveness of dietetics services is the ability of dietetics professionals to conduct well-designed studies and use research in daily practice. The House of Delegates, in 2001, recommended that "the ADA strengthen members' ability to understand and interpret research and increase practice-based research conducted in all practice settings using a variety of organizational units" (3). The *Journal of the American Dietetic Association* is the primary vehicle for communicating outcomes of dietetic and nutrition research. The overarching goal is to be the premier peer-reviewed journal in the field of food, nutrition, and dietetics and to support the mission of the American Dietetic Association by publishing well-designed, data-driven research. With this in mind, a working group with statistical expertise within the Board of Editors of the *Journal* was convened to promote statistically sound nutrition research design and analyses among the dietetics professions and to provide guidance to better document valid evidence related to diet in the prevention and treatment of disease.

The goals of this working group are to (a) assure the quality, accuracy, and validity of statistical methods and documentation in *Journal* research articles; (b) encourage involvement, at the earliest possible stages, of appropriate biostatistical expertise to establish a solid science base for the research itself and support statistical analyses needed to accurately interpret and convey results; and (c) develop a framework to guide, direct, and educate members, readers, and students on appropriate use of statistical methods using workshops, print, and Web-based information. This monograph represents an essential part of this process and is the introduction to a series of four monographs that will address and interpret the author guidelines (4) and provide relevant examples and interpretation for how to proceed with manuscript prep-

aration. The series also introduces the *Journal* as a potential source for statistical training and expertise targeting nutrition research reporting and providing a special focus on clinical nutrition issues.

The purpose of this monograph is to lay the foundation for the rest of the series by reviewing and defining common terms and basic resources available for use. The focus is on study design and development of testable research hypotheses, the beginning of all good-quality research. The second monograph will address statistical procedures using nonparametric and parametric variables and presentation of statistical results. The third monograph will explore common epidemiological methods, including appropriate use and reporting of odds ratios, relative risk, confidence intervals, statistical significance and the concepts of chance, confounding, and interaction. The final monograph 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. The *Journal* is pleased to provide readers with these peer-reviewed monographs, which will hopefully serve as review for some readers and new information for others to advance the field of nutrition and its practical applications.

DEVELOPING RESEARCH QUESTIONS AND HYPOTHESES

The Research Question

In the initial stages of designing research, it is important to clearly define what to study. There are three initial steps in deciding how to proceed forward with a research idea. The first is to state a research question, including concepts of interest. Second, brainstorm about the primary and confounding variables and appropriate study subjects. Finally, state a specific, measurable hypothesis from which research designs and methods can be defined. The research process starts with a research question or study hypothesis, specifying concepts of interest. A well-crafted research question includes the treatment or exposure concept, the outcome, the identity of the subjects, and over what time period the study will be conducted. For instance, a research question might be, "Does intake of n-3 fatty acids over 6 weeks among Asian-American men 45 to 60 years of age affect blood clotting?" See Figure 1 for additional features of a good research question. A common mistake made by researchers at this point is to start planning the research design and appropriate methods for the study without carefully defining the outcome variables to be measured and the treatment or exposure variables. Because it is imperative to define the expected outcomes of the research, the researcher needs to determine which parameters (variables) will represent the desired outcomes (5-7). In addition to specifying outcome variables to be measured, it is important to carefully and specifically define the treatment or exposure. Using the research question previously posed, how will the subjects ingest n-3 fatty acids and how much will they consume? This stage of the research process is the time to start seeking opportunities to collaborate with a biostatistician (see Sidebar).

Element	Essential characteristics
Research question (objective of study)	• Feasibility • Interesting • Novel or innovative • Ethical • Relevant or worth doing
Hypothesis (based on research question and basis for testing statistical significance of findings)	• Measurable • Specifies population being studied • Identifies time frame • Indicates type of relationship being examined • Includes variables being studied • Defines level of statistical significance (ie, the α level)

Figure 1. Features of a good research question and hypothesis. (Adapted from Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. *Designing Clinical Research: An Epidemiological Approach.* 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2001 and *National Institutes of Health, Grant Writing Tips Sheets* available at: http://deainfo.nci.nih.gov/extra/extdocs/gntapp.htm).

Types of Variables

There are two general types of variables, quantitative and categorical (8,9). Quantitative variables are also referred to as continuous variables, as they are measured on a scale in which a value could be placed between any two numbers. Serum cholesterol, blood urea nitrogen, systolic blood pressure, and waist circumference are examples of variables measured on a continuous scale.

Categorical variables consist of distinct categories either with an inherent order (ordinal) or with no defined order (nominal). Examples of ordinal variables would be pain severity, tastiness of a meal, or job satisfaction measured on a Likert scale (see Glossary) (10). Examples of nominal variables would be ethnicity, marital status, or geographic area.

As one defines variables to measure the concepts specified in the research question, the researcher needs to identify if the variables are categorical or quantitative, and determine how the concepts will be measured. For the research question about n-3 fatty acids and blood clotting, measurable parameters would need to be defined for these concepts. To measure treatment with n-3 fatty acids or exposure to n-3 fatty acids, some choices are consumption of measured portions of fatty fish or n-3 fatty acids administered in supplemental form as fish-oil capsules (11). Treatment applications must be well defined and validated. Dosages and amounts are specified and validation measures must be used to make sure all subjects are getting equal treatment application. If capsules are being used, researchers can count those that remain in the subjects' bottles at the end of the study to make sure subjects truly complied with the treatment. If food is being consumed, a dietary assessment method is adopted to validate the consumption of n-3 fatty acids among the subjects. Cell-membrane composition of n-3 fatty acids could also be used to determine if subjects

were consuming the required capsules or food. To measure the outcome variable, clotting time derived from a plasma sample can be used (12). Considering our example, clotting time would be measured before and after exposure to examine treatment effects. The key is to specify variables that are measurable and represent the concepts of the research question.

At this stage of the research-planning process, another important step is to identify potential confounding variables (13). Early involvement of a biostatistician can ease some of the complexity involved with this step. Potential confounding parameters can either be controlled through the research design or through analysis, if the variables are sufficiently measured during the research process. A confounder is an extraneous factor that is related to both the outcome of interest and the exposure of interest. Confounding is an apparent association (see Glossary) between an outcome and an exposure caused by a third factor not taken into consideration. Common confounding variables are age, sex, tobacco use, or history of tobacco use. For example, if it was known that individuals prescribed a specific dose of aspirin by their doctors were also more likely to be conscientiously consuming n-3 fatty acids, then the treatment or exposure of interest (n-3 fatty acids) would be associated with aspirin use. Furthermore, aspirin is known to influence blood clotting (14); thus, the outcome of interest is directly affected by aspirin. This sets the stage for classic confounding. In this case, the confounding can be addressed through study design by excluding those individuals who regularly take aspirin, or aspirin ingestion could be carefully measured and accounted for in the statistical analysis. Awareness of confounding variables will influence the conclusions from the research. The researcher should create a list of all potential confounders in preparation for obtaining assistance from a biostatistician (9). The biostatistician can then help design ways to control the influence of confounders.

Writing Measurable Hypotheses

Finally, a measurable hypothesis must be developed to answer the question, "What is really being studied?" The hypothesis includes the population being studied, the time frame, the α level for defining statistical significance, the type of relationship being examined, and the variables being studied (7). The α level is the chance a researcher is willing to take that there is not a difference or relationship between variables even though a statistical test might indicate such. If the α level is set at $P \leq .05$, it means that the researcher is willing to take a $\leq 5\%$ chance, which is the highest chance most statisticians are willing to take. Typically, the hypothesis is stated in the null, which means that it is stated that there is no relationship or difference between variables. An appropriate hypothesis might be, "There is no statistically significant difference at the $P < .05$ level in plasma clotting times among Asian-American men between 45 and 60 years of age taking 3 grams per day of n-3 fatty acids as combined docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA) in capsule form or a placebo for six consecutive weeks." This statement includes the treatment (exposure) (DHA/EPA), the outcome of interest (plasma clotting times), the population (age, ethnic group, sex), the comparison that will be tested, the time frame, and the α level to reject the null hypothesis (see Figure 1). When these elements are included, researchers are much more likely to pick a valid research design, select appropriate methods for assessing the variables of interest, and be able to more specifically interpret the data derived from the study. The hypothesis also forms the basis for the sample size needed to observe the expected differences in outcomes between treatment and nontreatment groups with a reasonable degree of probability, or power (15,16). It should be noted that hypotheses can be stated in the null, ie, there is no difference or relationship, or as an alternative hypothesis, ie, there is a difference or relationship. Manuscripts and published articles void of a measurable hypothesis are usually unable to provide a cogent answer to a research question or have minimal basis for valuable interpretation.

STUDY DESIGN

In research, the study design represents the critical set of blueprints that is the foundation from which the study's conclusions will be drawn. A common mistake is to gloss over study design and opt for using a convenient practice situation and place an emphasis on analysis. However, Campbell and Machin (17) note that "a badly designed study can never be retrieved, whereas a poorly analyzed one can usually be reanalyzed." Study design dictates analysis of the data versus the other way around.

Paramount with documenting the effectiveness of dietetics services is the ability of dietetics professionals to conduct well-designed studies and use research in daily practice.

Study designs are often grouped as observational studies, such as cross-sectional, cohort; or case-control and experimental studies, such as clinical trials (see Figure 2). These two groups are distinguished by the intervention that accompanies an experimental study. In this case, something is intentionally manipulated by the investigators among the subjects, eg, providing one group supplemental DHA/EPA and one group placebo. If there is no intervention and the subjects' usual routines and behaviors are observed and measured, then it's an observational study.

Observational Study Designs

The cross-sectional study design is most commonly used and includes surveys, laboratory experiments, and studies to describe the prevalence of disease and/or exposure in a specified population at one point in time. The National Health and Nutrition Examination Survey is cross-sectional as it takes a "snapshot" of the population, eg, 1999 to 2000 (18). This study design has its limitations when examining the association of diet and its role in the onset of disease. Any disease with a long latency period, such as cancer or osteoporosis, would not work for this design or with diseases that may alter exposure. Both the exposure and the outcome are estimated at the same time; thus, the method cannot provide insight on causal

Study design	Distinguishing features	Example
Observational study designs		
Prospective study[a]	A group free of the disease or outcome of interest followed over time	Examine a cohort of men periodically over several years, observe the incidence of stroke in high consumers of fatty fish rich in n-3 fatty acids and low consumers of these fish.
Cross-sectional study	A group examined at one point in time	Examine group of men once, observing the prevalence of a history of cerebrovascular disease in high consumers of fish and low consumers of fish.
Case-control study	Two groups, based on the outcome	Examine a group of men with history of cerebrovascular disease (the "cases") and compare them with a group of healthy men (the "controls") assessing history of fatty fish consumption.
Experimental study designs		
Cross-over randomized trial	Two groups created by a random process, one group starts placebo arm and the other in a treatment arm, washout period, treatment arms switch	Randomly assign men to receive n-3 fatty acids as supplements or identical placebo, follow both treatment groups for 6 weeks, discontinue treatment for 6 weeks, reinitiate treatments for 6 weeks with alternate treatment, observe plasma clotting time.
Randomized blinded trial (short term)	Two groups created by a random process, one group in a placebo arm and the other in a treatment arm	Randomly assign men to receive n-3 fatty acids as supplements or identical placebo, follow both treatment groups for 6 weeks to observe plasma clotting time.
Randomized blinded trial (long term)	Two groups created by a random process, one group in a placebo arm and the other in a treatment arm	Randomly assign men to receive n-3 fatty acids as supplements or identical placebo, follow both treatment groups for several years to observe the incidence of stroke.

Figure 2. Examples of research study designs. [a]Also referred to as a cohort or longitudinal study. (Adapted from Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. *Designing Clinical Research: An Epidemiological Approach.* 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2001 and Monsen ER. *Research Successful Approaches.* 2nd ed. Chicago, IL: American Dietetic Association; 2003:1-482).

relationships. This is referred to as antecedent-consequent bias. Therefore, cross-sectional design cannot be used to establish cause and effect. For example, individuals diagnosed with hypertension may reduce their consumption of salt and salty foods, thus leading to the false conclusion that low sodium intake is associated with hypertension. In this case, results cannot distinguish if diet was a result of the disease or if the diet preceded the disease. Nonetheless, this is a valuable study design that can address many other research questions and sequential cross-sectional studies are especially valuable for documenting changes in behaviors in populations over time, such as consumption of fruits and vegetables (19) and prevalence rates of elevated body mass index (20).

Prospective studies are not affected by the antecedent-consequent bias inherent in cross-sectional studies as the prospective (also known as cohort or longitudinal) study design focuses on the potential causal factor (exposure) and follows individuals forward in time to identify health outcomes (21). Eligible subjects are free of the disease of interest at time of enrollment. In this case, exposure to a diet high in salt or salty foods can actually be measured prior to onset of any disease. In the case of prospective studies, subjects are followed over time and, ideally, exposures are measured several times rather than once at the beginning. This is especially important if the subjects are followed over years. The CARDIA Study (now in its 20th year) (22) and the Framingham Heart Study started in 1948 (23) are examples of prospective observational-type studies.

A prospective study can be conducted retrospectively (also called a historical cohort study) (21) and opportunities for this are fairly common in dietetic settings. With this design, the investigators are still interested in finding subjects with and without exposure and following them forward in time, but the entire endeavor involves historical data. An example would be selecting a sample of patients admitted to a hospital during a specified period of time with the same admitting diagnosis. Then the medical records would be used to identify those patients receiving nutrition therapy (treatment) and comparing outcomes of interest (eg, length of stay, readmission) to those patients not receiving nutrition therapy. If an institution has very complete, standardized record-keeping, then data collected months and even years earlier can be collected from clients' or patients' records. The main drawback to this approach is the possibility of inconsistency in assessment of factors of interest. Researchers need to be aware of legal limitations secondary to the Health Insurance Portability and Accountability Act of 1996, which limits access to patients' former records without seeking appropriate permissions (24).

Another retrospective study design is the case-control study (21,25,26). Strict adherence to this study design starts with careful selection of individuals with the disease, ailment, or problem of interest. Then, disease-free controls are

selected from the same population pool as the cases. Exposure information is collected retrospectively either by asking all subjects to recall past events or securing permission to gather information from medical or employee records. Case-control studies are prone to recall bias (see Glossary), as cases are more motivated to recall events of the past compared with controls not affected by a disease. Often authors will erroneously label their study as a case-control study when in reality it is a cross-sectional study that simply designated subjects as "cases" and "controls." The North Carolina Colon Cancer Study is a case-control study examining risks for colon cancer in African-American and non-Hispanic white adults (27).

Experimental Study Designs

Randomized trials are often considered the "gold standard" of the prospective study design. In the case of randomized trials, subjects generally have some disease, ailment, or problem that will be the focus of the experimental treatment. Furthermore, subjects are randomly assigned to treatment groups, thus allowing comparisons between groups as to any differences in the outcome of interest as a result of the treatment. Additional quality steps in a randomized trial include use of blinding (or masking), where the subject does not know whether he or she is receiving treatment, and double-blinding, where neither the subject nor the investigators know which group is receiving treatment. Some procedures cannot be blinded to the subjects, such as receiving five servings of fruits and vegetables daily, compared to five servings of fruit or vegetable juice, compared to receiving nothing. For these situations, an attempt is made to blind the investigators, especially those doing the statistical analyses of the data. Examples of randomized trials include the Women's Health Initiative Dietary Modification Clinical Trial (28) and the Diabetes Control and Complications Trial (29).

An alternative to the randomized clinical trial is the crossover design, which allows for a comparison of within- and between-groups (7,30). Half of the subject volunteers are randomly assigned to start with the control period and then switch to the experimental treatment and the other half do the opposite sequence. This allows for comparisons between-groups and within-groups. With this design, confounding is held to a minimum because each subject acts as his or her own control. The problem with using this study design is the potential for any residual influences from the intervention on the outcome during the period after the treatment has stopped. This is referred to as *carryover effects* and introducing a "washout period" can alleviate this problem. For example, after consuming additional n-3 fatty acids for 6 weeks, study subjects may need a washout period of longer than a month to allow plasma levels of n-3 fatty acids to return to pretreatment levels.

Every author needs to document and justify the study design used. This allows readers to immediately be aware of study limitations and potential biases. In addition, it prepares the reader for the type of analysis to expect and the type of inferences that can be made for the conclusions. If, when planning a study, a researcher is unable to articulate the study design, then it probably indicates that more planning is needed. The term *intervention* does not communicate enough about study design other than to let the reader know that a type of treatment, food, education, or medicine was used. However, the study design is not evident. An intervention or treatment can be used in a randomized trial, a school randomized trial, a crossover study, or a quasiexperimental design study (treatment allocation is not random). A concise outline of study design approaches can be found in readily available books (15,21,26,31) and online (32).

COMMUNICATING RESEARCH ACTIVITIES IN A MANUSCRIPT

Journals in the life sciences include guidelines for authors that stress inclusion of a precise objective or a specific hypothesis, or both, if applicable. Generally, this advice is followed by directions to elaborate information about the study design. The approach to communicate this important part of the research process will vary depending upon the research question and the methods used. The guidelines for authors of the *Journal of the American Dietetic Association* can act as a resource to authors in preparing an informative, readable manuscript (4). Some examples from published articles are highlighted here.

Randomized trials are often considered the "gold standard" of the prospective study design.

When preparing a manuscript for the *Journal*, the research objective or hypothesis is generally placed in two locations. An abbreviated version is placed in the abstract and the detailed version is placed at the end of the introduction, after the review of the literature. In a Research Brief authored by Holcomb and colleagues (33), the research objective appeared as the last sentence of the introduction, ". . . to examine data from the Third National Health and Nutrition Examination Survey (NHANES III) for the association of physical activity with body fatness and abdominal obesity in women aged 20 to 55 years." This statement includes all of the necessary elements of a research objective: population being studied (women between 20 and 55 years), the time frame (the dates for data collection of the national surveys called NHANES are well documented and available), type of relationship being examined (an association), and variables (physical activity, body fatness, abdominal obesity). There is no reference to an α level for defining statistical significance as this is often included in the statistical analyses section of a manuscript rather than in the research objective.

The article by Holcomb and colleagues (33) represents a unique situation as the study design is inherent in the source of the data, ie, NHANES. Because a survey is a type of cross-sectional study, specifying the cross-sectional study design can be considered optional, especially in a Research Brief. This is in contrast to a cross-sectional study reported by Baruffi and colleagues (34). In the research objective for this study, it begins, "This article presents a cross-sectional analysis of data from the 1997 to 1998 Hawaii Federal Supplemental Nutrition Program for Women, Infants, and Children (WIC) with the aim of comparing the prevalence of overweight among young children of different ethnic backgrounds and describing the age pattern of overweight in early childhood." Since

the WIC follows eligible clients over time, it would be feasible to initiate a prospective study. However, the authors clearly indicate immediately that in their study the data from WIC were analyzed as a cross-sectional study. An analysis of data from The Bogalusa Heart Study provides an example of a prospective study (35). The abbreviated research objective found in the abstract reads, ". . . to assess changes in food group consumption patterns from childhood to young adulthood." The study design is identified in the design section of the abstract as "longitudinal," another label for prospective study design.

Attention to detail about the research objective or hypothesis, in addition to the study design, will contribute positively to a reader's ability to comprehend the scope of the study.

A research objective is often used with a cross-sectional study as these observational studies are often descriptive in nature and describe distributions of diseases and health-related characteristics in the sample population. A hypothesis is the standard with a randomized trial study design. Jambazian and colleagues (36) finished the literature review in their article with, "We hypothesized that the addition of a single food high in α-tocopherol has the ability to raise lipid-adjusted plasma and RBC α-tocopherol levels." In the abstract, the study design was included with the hypothesis, "in a randomized, crossover feeding trial." The study design is further elaborated in the Methods section. This is necessary as the technique used for randomization needs to be elaborated, as does the length of the study and the steps for "crossover." An article by Karmally and colleagues (37) uses key parts needed in a hypothesis in the abstract, eg, "This randomized controlled trial of cholesterol lowering [type of relationship being studied] by an oat bran cereal containing beta glucan vs a corn cereal without soluble fiber [variables being studied] in Hispanic Americans [population being studied] was conducted for 11 weeks [time frame]." The statement does not specify direction or level of confidence with the results (α values, eg, P values), however, this can be noted in the Statistical Methods section. Including the information about the hypothesis and study design in the title of the article and detail in the abstract are useful to readers and useful for data collection methods used in meta-analysis.

A manuscript needs to clearly outline why and how a study was done. Attention to detail about the research objective or hypothesis, in addition to the study design, will contribute positively to a reader's ability to comprehend the scope of the study. Authors are encouraged to seek out resources highlighted in this series of articles and the Authors' Guidelines when preparing manuscripts (4). These guides will insure that dietitians and other health professionals will continue to have a resource through the *Journal* for evidence-based practice for treatment and prevention of disease and promotion of health and well-being.

References

1. The American Dietetic Association Web site. Available at: http://www.eatright.org/Member/83_9468.cfm. Accessed July 25, 2005.
2. Commission on Accreditation for Dietetics Education. *Accreditation Handbook.* Chicago, IL: American Dietetic Association; 2002. Catalog number 0430.
3. The American Dietetic Association Web site. Available at: http://www.eatright.org/Print/index_11557.cfm. Accessed on July 18, 2005.
4. American Dietetic Association. *Journal of the American Dietetic Association* guidelines for authors. *J Am Diet Assoc.* 2006;106:140-147.
5. Rothman KJ. Measuring disease occurrence and causal effects. In: Rothman KJ, ed. *Epidemiology: An Introduction*. New York, NY: Oxford University Press; 2002:24-56.
6. Moore DS, McCabe GP. Producing data. In: Moore DS, McCabe GP, eds. *Introduction to the Practice of Statistics*. 5th ed. New York, NY: W.H. Freeman and Company; 2006:191-250.
7. Monsen ER. *Research Successful Approaches.* 2nd ed. Chicago, IL: American Dietetic Association; 2003: 1-482.
8. Moore DS, McCabe GP. Looking at data—Distributions. In: Moore DS, McCabe GP, eds. *Introduction to the Practice of Statistics.* 5th ed. New York, NY: W.H. Freeman and Company; 2006:3-39.
9. Cheney CL. Statistical applications. In: Monsen ER, ed. *Research Successful Approaches*. 2nd ed. Chicago, IL: American Dietetic Association; 2003:399-416.
10. Pratt PE, Kwon J, Rew ML. Perceived job importance and job performance satisfaction of selected clinical nutrition management responsibilities. *J Am Diet Assoc.* 2005;105:1128-1133.
11. Institute of Medicine National Academy of Sciences. *Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids (Macronutrients).* Washington, DC: National Academy Press; 2002.
12. Knapp HR. Dietary fatty acids in human thrombosis and hemostasis. *Am J Clin Nutr.* 1997;65:S1687-S1698.
13. Newman TB, Browner WS, Hulley SB. Enhancing causal inference in observational studies. In: Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB, eds. *Designing Clinical Research: An Epidemiologic Approach*. 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2001:125-141.
14. Kadakia RA, Ferguson JJ. Optimal antithromobotic treatment for percutaneous coronary intervention. *Minerva Cardioangiol*. 2005;53:15-42.
15. Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. *Designing Clinical Research: An Epidemiological Approach.* 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2001.
16. Cheney CL, Boushey CJ. Estimating sample size. In: Monsen ER, ed. *Research Successful Approaches.* 2nd ed. Chicago, IL: American Dietetic Association; 2003: 389-398.
17. Campbell MJ, Machin D. *Medical Statistics: A Common Sense Approach.* 3rd ed. West Sussex, UK: John Wiley & Son Ltd; 1999.

18. National Center for Health Statistics, National Health and Nutrition Examination Survey information page. Centers for Disease Control and Prevention Web site. Available at: http://www.cdc.gov/nchs/nhanes.htm. Accessed July 15, 2005.
19. Kaminoto LA, Easton AN, Maurice E, Husten CG, Macera CA. Surveillance for five health risks among older adults—United States, 1993-1997. *MMWR CDC Surveill Summ*. 1999;48:89-130.
20. Ogden CL, Flegal KM, Carrol MD, Johnson CL. Prevalence and trends in overweight among US children and adolescents, 1999-2000. *JAMA*. 2002;288:1728-1732.
21. Rothman KJ. Types of epidemiologic study. In: Rothman KJ, ed. *Epidemiology: An Introduction*. New York, NY: Oxford University Press; 2002:57-93.
22. McDonald A, Van Horn L, Slattery M, Hilner J, Bragg C, Caan B, Jacobs D, Liu K, Hubert H, Gernhofer N, Betz E, Havlik D. The CARDIA dietary history: Development, implementation, and evaluation. *J Am Diet Assoc*. 1991;91:1104-1112.
23. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med.* 1975;4:518-525.
24. Erlen JA. HIPAA—Implications for research. *Orthop Nurs*. 2005;24:139-142.
25. Newman TB, Browner WS, Cummings SR, Hulley SB. Designing an observational study: Cross-sectional and case-control studies. In: Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB, eds. *Designing Clinical Research.* 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2001:107-124.
26. Boushey CJ. Nutritional epidemiology: Dietary assessment methods. In: Weaver CM, Heaney RP, eds. *Calcium in Human Health.* 1st ed. Totawa, NJ: Humana Press; 2006:39-64.
27. Satia-Abouta J, Galanko JA, Martin CF, Potter JD, Ammerman A, Sandler RS. Associations of micronutrients with colon cancer risk in African Americans and whites: Results from the North Carolina Colon Cancer Study. *Cancer Epidemiol Biomarkers Prev*. 2003;12:747-754.
28. Ritenbaugh C, Patterson RE, Chlebowski RT, Caan B, Fels-Tinker L, Howard B, Ockene J. The Women's Health Initiative Dietary Modification trial: Overview and baseline characteristics of participants. *Ann Epidemiol.* 2003;13(suppl 9):S87-S97.
29. Schmidt LE, Cox MS, Buzzard IM, Cleary PA. Reproducibility of a comprehensive diet history in the Diabetes Control and Complications Trial. The DCCT Research Group. *J Am Diet Assoc*. 1994;94:1392-1397.
30. Grady D, Cummings SR, Hulley SB. Designing an experiment: Clinical trials II. In: Hulley SB, Cummings SR, Browner D, Grady W, Hearst N, Newman TB, eds. *Designing Clinical Research*. 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2001:157-174.
31. Iverson C. *American Medical Association Manual of Style.* 9th ed. Baltimore: Williams & Wilkins; 1998.
32. Swinscow TDV. Revised by MJ Campbell. *Statistics at Square One*. 9th ed. BMJ Publishing Group; 1997. Available at: http://bmj.bmjjournals.com/collections/statsbk/index.shtml. Accessed on July 15, 2005.
33. Holcomb CA, Heim DL, Loughin TM. Physical activity minimizes the association of body fatness with abdominal obesity in white, premenopausal women: Results from the Third National Health and Nutrition Examination Survey. *J Am Diet Assoc*. 2004;104:1859-1862.
34. Baruffi G, Hardy CJ, Waslien CI, Uyehara SJ, Krupitsky D. Ethnic differences in the prevalence of overweight among young children in Hawaii. *J Am Diet Assoc*. 2004;104:1701-1707.
35. Demory-Luce D, Morales M, Nicklas T, Baranowski T, Zakeri I, Berenson G. Changes in food group consumption patterns from childhood to young adulthood: The Bogalusa Heart Study. *J Am Diet Assoc.* 2004;104:1684-1691.
36. Jambazian PR, Haddad E, Rajaram S, Tanzman J, Sebate J. Almonds in the diet simultaneously improve plasma α-tocopherol concentrations and reduce plasma lipids. *J Am Diet Assoc*. 2005;105:449-454.
37. Karmally W, Montez M, Palmas W, Martinez W, Branstetter A, Ramakrishnan R, Holleran SF, Haffner SM, Ginsberg HN. Cholesterol-lowering benefits of oat-containing cereal in Hispanic Americans. *J Am Diet Assoc*. 2005;105:967-970.

GLOSSARY

For a helpful glossary of research and statistical terms, refer to the Centers for Disease Control and Prevention website at http://www.cdc.gov/reproductivehealth/EpiGlossary/glossary.htm. Most of the following terms come from this resource.

Association: Statistical relationship between two or more events, characteristics, or other variables.

Exposed (group): A group whose members have been exposed to a supposed cause of disease or health state of interest, or possess a characteristic that is a determinant of the health outcome of interest.

Incidence rate: A measure of the frequency with which an event, such as a new case of illness, occurs in a population over a period of time. The denominator is the population at risk; the numerator is the number of new cases occurring during a given time period.

Likert scale: A Likert scale is used to measure the extent to which a person agrees or disagrees with a statement and is used to measure attitudes, preferences, and subjective reactions. Traditionally a 5-point scale is used; however, scales of less or more than 5 are also used.

Example: Milk goes great with cookies.
1=Strongly disagree
2=Disagree
3=Neither agree or disagree
4=Agree
5=Strongly agree

Prevalence: The number or proportion of cases or events or conditions in a given population.

Recall bias: A systematic error resulting from individuals in one group being more likely than individuals in a different group to remember past events.

CONVERSATION WITH A BIOSTATISTICIAN

George McCabe, PhD, Professor of Statistics at Purdue University, West Lafayette, IN, recently served as a member of the Institute of Medicine Committee on the Use of Dietary Reference Intakes in Nutrition Labeling. Dr McCabe directed the statistical consulting service on the Purdue University campus for over 30 years, thus gaining an appreciation for working with scientists in pursuing their research goals. Carol J. Boushey, PhD, MPH, RD, Associate Professor in the Department of Foods and Nutrition at Purdue University, reports here her conversation with Dr. McCabe.

CJB: When planning a research project, when in the process should a researcher include a statistician?

GM: When the researcher or research group are formulating ideas . . . in other words, very early in the process.

CJB: What is the most important piece of information that a researcher should bring to a meeting with a statistician?

GM: A description of what the investigator hopes to find out . . . a description of the expected outcomes.

CJB: As an experienced statistical consultant, what is most often missing when you meet with a researcher for the first time?

GM: This depends upon the stage of the research process. Early in the process—the lack of pilot study data. Later in the process—a lack of precise details about how the study was done, such as the randomization method used or lack of randomization. Other details often missing are a clear description of the experimental units or how groups were assigned to intervention and case status.

CJB: Is there a list of things to bring to a meeting with a statistician?

GM: Reprints of previous studies similar to the one a researcher is starting, pilot data, and ball-park estimates of costs to make reasonable sample size estimates within budget.

CJB: Speaking of sample size, is it appropriate to ask a statistician to calculate sample size?

GM: Yes, that is an appropriate request to make. However, sample size is just one of the technical details that can be checked-off. Statisticians really want to be involved in the thinking-through of a research project, including the selection of treatments and the development of the protocols. The statistician's contributions to the research process can extend far beyond the calculation of sample size.