CHAPTER **21**

# CLINICAL RESEARCH: Show Me the Evidence

*D. Scot Malay, DPM, FACFAS*

It is the purpose of this paper to provide an overview of the basic considerations related to research design and presentation. Moreover, an emphasis will be put on those features of study design that contribute to the validity of a publication.

Let me start this review of research design with a quote from Galen, the Greek surgeon who preceded Vesalius by over 1200 years. Galen, circa 150 AD, concluded: "All who drink of this remedy recover in a short time, except those whom it does not help, who all die. Therefore, it is obvious that it fails in only incurable cases." This statement may be true, however it is just as likely to be inaccurate, and we, as readers, need additional information before we can agree or disagree with the conclusion. In all likelihood, Galen's conclusion was based on a case series in which a considerable amount of the surgeon's own bias probably led to systematic error which could have compromised his conclusion. This is seen, rather frequently, in much of the medical literature...even in some refereed journals.

Lind, on the other hand, being the surgeon to the Royal Navy in the late 1700s, employed more scientific rigor in his study of scurvy. He noted: "...I took twelve patients...(with) scurvy...There cases were as similar as I could have them...They lay together in one place and had one die common to all. Two of these were ordered each a quart of cyder a day. Two others took twenty-five drops of elixir of vitriol three times a day upon an empty stomach. Two others took two spoonfuls of vinegar three times a day...Two of the worst patients were put upon a course of seawater. Two others had each two oranges and one lemon given them every day. The two remaining patients took an electuary recovered by a hospital surgeon made of garlic, mustard, balsam of Peru and myrrh. The consequence was that the most sudden and visible good effects were perceived from the use of oranges and lemons; one of those who had taken them being at the end of six days fit for duty. The other was the best recovered of any in his condition and was appointed nurse to the rest of the sick."[1] Clearly Lind's attention to the case-control design, restrictions and inclusion criteria for the treatment groups, yields a result that is more likely to be valid (and has stood the test of time).

## CAUSALITY

There are several types of associations between a cause (intervention, treatment, disease) and an effect (outcome, morbidity, death or cure). These include:

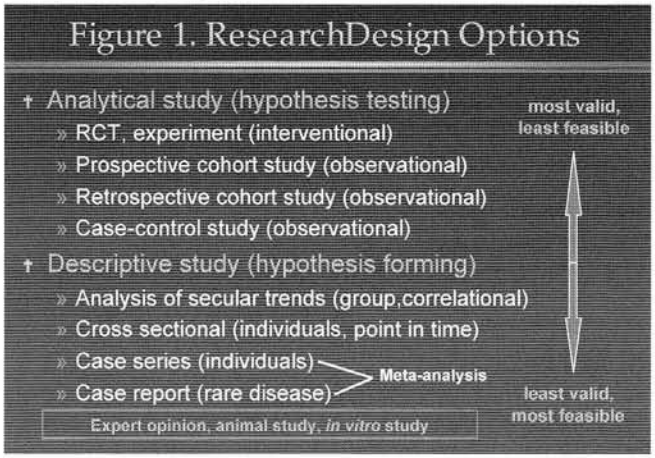
1. None (independent)
2. Art factual, spurious
   a. Chance
   b. Bias
1. Indirect
   a. Confounding
4. Direct (causal)

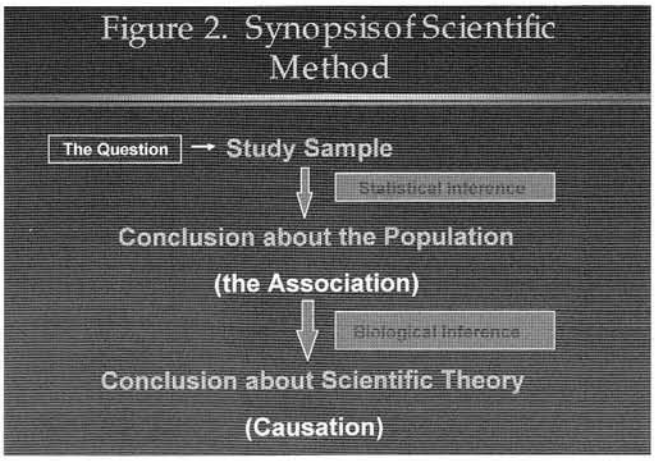
Support for a causal association entails:

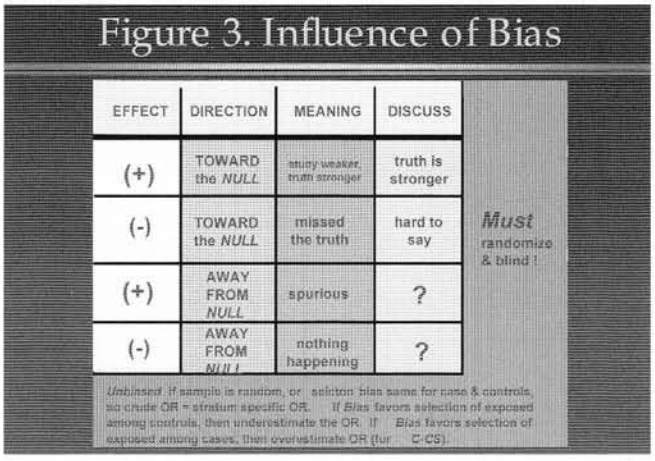
1. Coherence with existing information
2. Consistent, repeatable results when tested at different locations and times
3. Time sequence of events (the outcome follows the intervention chronologically)
4. Strength of the association
   a. Magnitude of the result (relative risk, odds ratio)
   b. Dose-response (more of an effect with more of the intervention)
   c. Study design (RCT, cohort, case-control, trend analysis, cross-sectional, case reports, expert opinion, animal study, in-vitro experiment)
5. Specificity (rare in biology) (e.g. only smoking causes lung cancer)

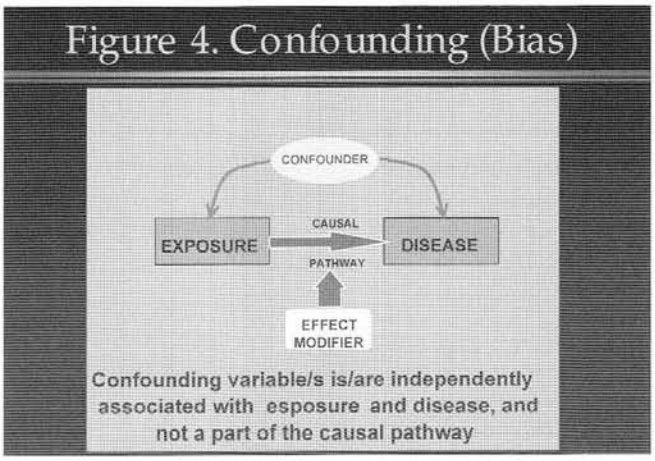
## SCIENTIFIC METHOD OF INVESTIGATION

Ideally, the fundamental technique employed in clinical research is the scientific method. The goal of the scientific method is to identify the causal relationship of an association between an intervention and its affects. Much of the appreciation for implementation of the scientific method in clinical research design stems from the field of epidemiology, specifically clinical epidemiology. The epidemiologist studies the distribution and determinants




of disease in populations. Whereas, the clinical epidemiologist extends these basic principles to the critical evaluation of diagnostic and therapeutic modalities used in clinical practice. There are several useful ways to design a clinical research study (Figure 1), each of which conveys different strengths and weaknesses regarding the ability to identify "the truth." And, it is important for the clinician to be aware of these basic research techniques, because the medical literature is often littered with meaningless information based on poorly designed studies. In other words, it is important for the clinician to have a good sense of what makes an author's conclusions legitimate. At the very least, the practicing clinician should focus his/her attention on the medical literature contained within refereed journals. These publications initiate the review process for the clinician, since their editorial board has the job of filtering articles for scientific merit prior to publication. This is not always the case with textbooks and symposia and, unfortunately, even refereed journals publish poor quality research on occasion. In a comparison of randomized clinical trials published in podiatric journals (JAPMA & JFAS) and a mainstream medical journal (JAMA), Turlik and colleagues[2] concluded that the podiatric medical journals were, based on a standardized evaluation scale,[3-6] less credible than the same type of study published in the mainstream medical journal. To assess the validity of a published clinical trial, Turlik and colleagues recommend that the reader ask the following questions of the article:

1. Was randomization explained?
2. Was concealment allocation explained?
3. Were the treatment groups comparable at baseline?
4. Were subjects accounted for at the end of the study?
5. Were outcomes assessors blinded?
6. Was the outcome instrument validated?
7. Was power analysis done prior to the start of the study?
8. Was statistical analysis compatible with type and distribution of the data?
9. Was the 95% confidence interval calculated for point estimates of the outcome?

The scientific method, therefore, underpins the research design (Figure 2). It can be summarized in the following way: 1. A research question is formulated (how does a specific treatment affect a specific disease). 2. A study is designed to answer the research question. 3. A sample population is selected and the study is performed on this finite group of subjects. 4. The data are analyzed statistically and an inference is made about the association between the treatment and the outcome. 5. A biological inference is made that enables us to make a conclusion about causation in the general population. In the end, knowledge is gained and, if attention was paid to minimizing bias, the findings will most likely be valid.

## BIAS

A critical appraisal of the medical literature requires an understanding of the many ways in which bias can affect the results and conclusions of a study. Bias in a study is a systematic error in the collection and/or interpretation of data (measurement). Bias threatens the validity of an inference to the general population, and bias can not be "fixed" with statistical methods. Bias has both a magnitude and a direction, either toward or away from the null hypothesis (Figure 3). The null hypothesis states that the outcome occurred by chance, rather than due to the intervention (the alternative hypothesis). Bias can take several common forms, including:

1. Selection (non-participant, drop out)
2. Informational
   a. Differential effort (Hawthorne effect)
   b. Misclassification (calculation, assay, or recall error)
3. Confounding variable (Figure 4)

Chance can also influence the magnitude and direction of an effect, however attention to proper power and statistical methods should resolve chance related bias. Confounding variable/s is/are independently associated with both the exposure (intervention) and disease (outcome), and are not a part of the causal pathway between the exposure and the disease. Methods used to decrease bias include:

1. Research design (quality of the study)
2. Restriction (inclusion and exclusion criteria)
3. Blinding of participants and examiners
4. Matching participants (caution in case-control studies)
5. Data analysis
6. Stratified analysis
7. Mathematical modeling

## BUILDING BLOCKS OF PATIENT ORIENTED RESEARCH

Every study is comprised of a variety of elements that, when combined, enable us to reduce the impact of bias and to, ideally, make valid statements about causality. It is the researchers' responsibility to avoid bias in the various elements of the study. The building blocks of unbiased patient oriented research include[2]:

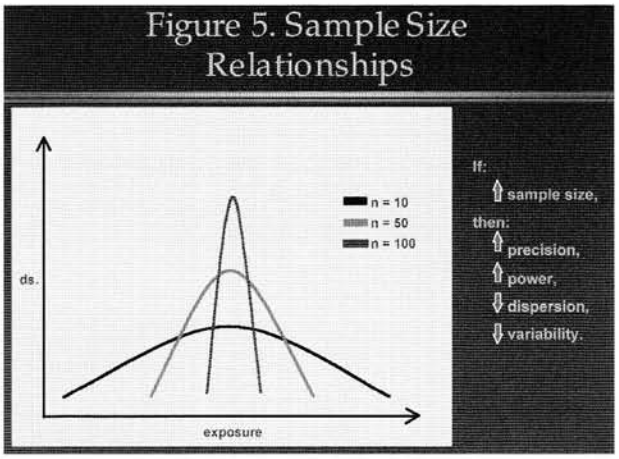
1. Randomization
2. Double blinding
3. Comparable treatment arms (avoidance of selection bias)
4. Validated health measures
5. Power analysis
6. Intent-to-treat
7. Compatible statistical analysis
8. 95% confidence interval

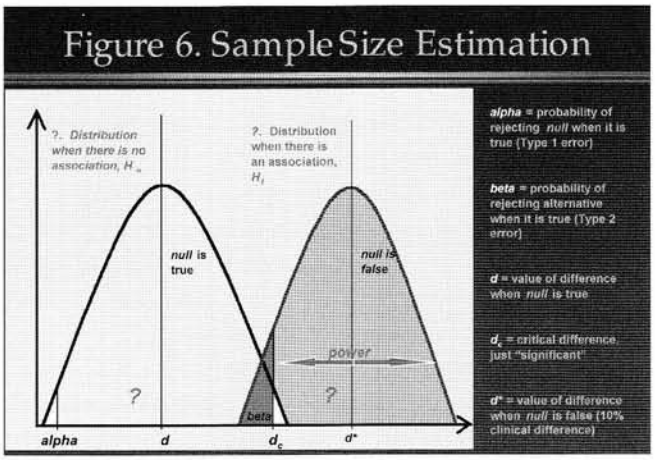
Randomization - Subject sampling is absolutely critical to the outcome of any clinical study. Use of the probability, or random, sample implies that all of the subjects have an equal chance of selection. In this sense, there is no bias in the selection process. A variety of methods are available for establishing a random sample, including simple (random number generator or table), systematic (every ith after a random start), stratified (all of a subset and random selection from larger set), and clustered (random subset) sampling techniques. Non-probability (non-random) samples always convey a certain degree of bias and certain subjects are therefore more likely to be selected. Non-random methods include purposeful (selection of extremes, compliant, non-compliant subjects), judgmental (defined by expertise), quota (age, race, geographical), and convenience or haphazard sampling. Most case series are examples of samples of convenience.

Double blinding -Both subjects and outcomes assessors need to be blinded, in order to reduce the risk of bias.

Comparable treatment arms - It is crucial that cases and controls be of similar make-up in as many ways as possible. A typical bias displayed in case-control studies, as well as case reports, is disparity between the degree of disease in the study groups (e.g. cases more ill than controls, or vice versa).

Validated health measure - Outcomes measures such as pain, ability to perform the outcomes of daily living, and quality of life require implementation of validated




instruments (e.g. McGill Pain Questionnaire,[7] Medical Outcomes Survey Short Form 36.[8] Hard measurements, such as radiographic angles, age, weight, range of motion, and many others, require the use of appropriate measuring techniques, however do not necessarily have any relationship to the quality of a subject's outcome. In other words, it may not matter that the hallux abductus angle significantly improved if the patient is unable to work after bunion surgery. Similarly, a subject may be much better off following bunion surgery, despite limited improvement in the hallux abductus angle.

Power analysis - Hypothesis testing requires that the sample being studied be of sufficient size in order to have enough statistical power to show whether or not a significant difference is identified (Figures 5 and 6). As a general rule, the study should be able to detect a 5% difference at 80% power. Data from small studies can be pooled by means of meta-analysis, in order to achieve statistical power that may enable a significant difference to be identified. Meta-analysis requires comparison of studies that are similar based on specified inclusion and exclusion criteria. The specific strategy used for inclusion should be defined, including the test of heterogeneity.

Intent-to-treat analysis - In order to minimize bias related to study drop-outs and/or cross-over treatments, the intent-to-treat analysis should be employed. In this way, subjects are analyzed based on the treatment group to which they were originally randomized. Furthermore, this enables the investigators to account for all of the subjects at the end of the study. Failure to employ the intent-to-treat analysis subverts randomization, and compromises interpretation of the outcome.

Compatible statistical analysis - Statistical analysis of the data enable us to make conclusions that go beyond the limits of our finite sample population. Biostatistics is a form of applied statistics useful in epidemiological research. Statistics are used to enable us to accurately describe our data and, more importantly, to make inferences about our results. An inference is a conclusion that is made without having complete information or absolute knowledge of the entire population (beyond the study sample). Inferential statistics are founded in probability theory, and a basic understanding of the available techniques will allow the researcher to determine the best method of data analysis for the specific study goals. Although a detailed discussion of biostatistics is beyond the scope of this presentation, in general, descriptive statistics entail the use of measures that define location (mean, median, mode) and measures of spread (range, quantiles, variance, standard deviation, and coefficient of variation). Such data are usually displayed in tables, graphs (bar, histogram, stem-and-leaf plots, box plots), and frequency distributions. Inferential statistics assume the use of a random variable, and allow us to determine relative risk, odds ratios, sensitivity and specificity, predictive value, prevalence and incidence, confidence intervals and other probabilities. The specific type of statistical analysis used will vary with the type and distribution of the data collected.

The analysis should also yield results that enable the reader to understand the association of an intervention or treatment on the outcome in question. Measures of frequency include prevalence, or the number of case occurring in a population at a point in time. Prevalence is often used in cross-sectional studies and serves as an indication of the burden of a disease on the population. The cumulative incidence indicates the number of new cases per susceptible individuals in the population, and the incidence density indicates the number of new cases in the population per total person-time and is often used in life table analysis. Measures of association (Figures 7, 8) include the relative risk, which specifically is the incidence in the exposed group divided by the incidence in the unexposed group of subjects. The attributable risk

is the incidence in the exposed less the incidence in the unexposed group of subjects.

Relative Risk = a/a+b / c/c+d = Ie / Io
Odds Ratio (RR for C-Cs) = ad / bc

Other useful inferential statistical measures include the confidence interval, t-distribution, and the Chi-square distribution, among many others. In and of itself, statistical validity does not convey causality.

95% confidence interval - Probabilities and point estimates should be depicted in conjunction with the 95% confidence interval. The length of the confidence interval gives some idea of the precision of the point estimate. A large interval implies less precision and more variance. A confidence interval that crosses 1 indicates an insignificant value. The confidence interval is governed by the sample size (Figure 5), the standard deviation from the mean, and degree of confidence desired (95% versus 99%). As the sample size increases, the confidence interval decreases. As the standard deviation decreases, the confidence interval also decreases. Finally, as the degree of confidence desired increases, the confidence interval increases. The confidence interval can be used as a description of the precision with which the parameters of a distribution can be estimated. Moreover, the confidence interval can be used to make clinical decisions on the basis of data.

## HIERARCHY OF STUDY DESIGN

In general, the hierarchy of clinical research design,[9,10] starting with the least valid but usually most feasible (less expensive) study format, is as follows:

1. Case report and case series
2. Cross sectional study
3. Analysis of secular trends
4. Case-control study
5. Retrospective cohort study
6. Prospective cohort study
7. Randomized controlled trial (RCT)

Analytical studies, which allow hypothesis testing, include the case-control, cohort, and RCT designs. Descriptive studies, which allow hypothesis formation, include case reports, cross sectional studies, and the group, correlational, analysis of secular trends. Even less valid are expert opinions, animal studies (as they relate to causality in humans), and in-vitro experiments.

Case report or series

Advantages

1. Describes person/s, place/s, and time/s
2. Ideal for rare disease, unusual event
3. Post-marketing adverse drug reactions (www.MEDWatch)
4. Hypothesis formulation

Disadvantages

1. Non-random
2. No control of bias or confounding

Cross-sectional & analysis of secular trends

Advantages

1. Person, place, point or period prevalence
2. Public health issues (trends)
3. Hypothesis formulation

Disadvantages

1. No concept of elapsed time or incidence rate
2. No understanding of exposure and disease
3. Provides point prevalence only
4. No control of bias/confounding

The analysis of secular trends can be sub-divided into cross-sectional (individual) and correlational (group) studies. The cross-sectional design is the quintessential public health tool, and has been used to evaluate screening tests (e.g. Pap smear, colonoscopy), disease burden at point/period in time, lifetime exposures (birth defects, blood type, race, sex, etc.), change over time (new treatment, technology, or environmental change), and events over time to identify a trend (correlation).

Case-control study (Figures 8, 9)

Advantages

1. Improved efficiency over cohort study (less time, less cost)
2. Odds Ratio (OR ~ RR, rare disease assumption)
3. Risk set sampling allows OR ~ RR

Disadvantages

1. Must sample from random cohort (selection bias)
2. Can not determine incidence (start with disease or exposure)

Cohort study (Figures 7, 9)

Advantages

1. Outcome unknown (can not bias)
2. Random sample prevents biased exposure
3. Can determine incidence rate and relative risk
4. Prospective or retrospective (Figure 10)
5. External validity (generalizeable)

Figure 7. Measure of Association in Cohort Study

Relative Risk = a/a+b / c/c+d = $I_e / I_o$

case-control study

| | Ds. | No Ds. | |
|---|---|---|---|
| Exp. | A | B | A+B |
| Un-exp. | C | D | C+D |
| | A+C | B+D | A+B+C+D |

cohort study

Figure 8. Measure of Association in Case-Control Study

Odds Ratio (*RR* for C-Cs) = ad / bc

| | Case | Control | |
|---|---|---|---|
| Exp. | A | B | A+B |
| Un-exp. | C | D | C+D |
| | A+C | B+D | A+B+C+D |

Figure 9. Case-Control & Cohort Studies

OR

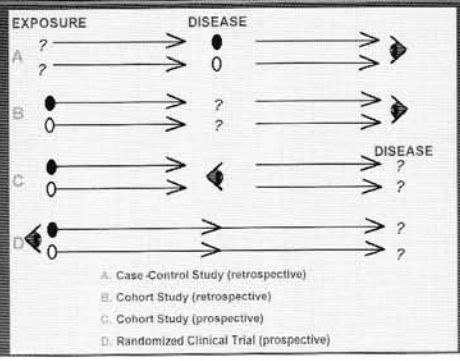
case-control study

| | Ds. | No Ds. | |
|---|---|---|---|
| Exp. | A | B | A+B |
| Un-exp. | C | D | C+D |
| | A+C | B+D | A+B+C+D |

cohort study

RR

Figure 10. Retrospective or Prospective?




Disadvantages

1. Time consuming and expensive
2. Can bias outcome if non-random exposure
3. Need large sample (drop outs, changes over time
4. Must standardize for specific features of the cohort
   a. Direct - apply the features of known standard to study cohort (assume study population has the same characteristics as the standard [i.e. US Census data] population)
   b. Indirect – rates from a standard population are applied to the rates in the exposed group to control for effects of potential confounders

Randomized controlled trial (RCT)

Advantages

1. Most generalizeable
2. Unbiased (randomized, blinded, monitored)
3. Criteria well-defined
4. Restrictions, matching, endpoint/s (ordinal, continuous), drop outs & missing data, arms
5. Statistically manageable and documented
6. Most definitive evidence for/against a treatment
7. Prospective (Figure 10)

Disadvantages

1. Expensive and difficult to perform (monitor/deadlines)
2. Internal validity may not generalize if too focused (selection bias)

In summary, a clinical research study should indicate in clear terms, the following points:

1. Purpose...What is the Question?
2. Research design and methodology
3. Statistical power of the study
4. Clear and concise graphic description/s of the data
5. Support for the abstract and conclusions
6. Stated shortcomings and limitations of the study, such as bias and confounding

There are several proper ways to perform clinical research, and there exists a hierarchy of research designs that differ in ability to identify the truth. Application of biostatistics enable investigators to make inferences related to the general population. Bias can threaten the validity of the results of the study. In podiatric medicine and surgery, we have to strive for high quality clinical research that produces valid results.

Suggested databases for clinical research guidelines:

Centres for Health Evidence
(http://www.cche.net/CHE/home.asp.)
MEDWatch
Cochrane Collaboration
TRIP
Info POEMs
National Guideline Clearing House
Health Care Guideline
Medline
National Center for Health Statistics
(http://www.cdc.gov)
FDA (http://www.fda.gov)
PS Power & Sample Size Calculations
www.mc.vanderbilt.edu/prevmed/ps.htm

## REFERENCES


1. Reprinted in: CP Stewart, D Guthrie, Eds. Edinburgh University Press, 1953.
2. Turlik MA, Kushner D, Stock D. Assessing the validity of published randomized controlled trials in podiatric medical journals. *J Am Podiatric Med Assoc* 2003;93:392.
3. Turlik MA, Kushner D. Levels of evidence of articles in podiatric medical journals. *J Am Podiatric Med Assoc* 2000;90:300-4.
4. Greenhalgh T, Donald A. Papers that report drug trials (RCTs). In Evidence Based Health Care Workbook for Individual & Group Learning, Vol.1, Greenhalgh T, Donald A., Eds. BMJ Books, London; 2000, p. 59.
5. Jadad A. Assessing the quality of RCTs: why, what, how, & by whom? In Randomized Controlled Trials, Vol.1, Jadad A, Ed. BMJ Books, London; 1998, p. 45.
6. Dixon RA, Monroe JF, Silcocks PB. A paper on clinical agreement. In The Evidence Based Medicine Workbook: Critical Appraisal for Clinical Problem Solving, Vol. 1, Dixon RA, Nonroe JF, Silcocks PB, Eds. Reed Educational and Professional Publishing, Oxford, 1997, p. 17.
7. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain* 1975;1:277-99.
8. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). Conceptual framework and item selection. *Med Care* 1992;30:473-83.
9. Hennekins CH, Buring JE. Epidemiology in Medicine. Baltimore: Lippincott Williams & Wilkins; 1987.
10. Guyatt G, Rennie D. User's Guide to the Medical Literature: Essentials of Evidence-Based Clinical Practice. Washington (DC): AMA; 2002.