



Evaluating the Literature

Published January 1, 2006

As new studies are published, the evidence base for understanding the risks and benefits of treatment options changes. A basic understanding of the types of studies and the meaning of the analyses helps healthcare providers evaluate the evidence and implications for clinical practice.

Types of studies. The two major types of studies are *experimental* and *observational*. In experimental studies, the interventions and conditions are strictly defined and controlled. In observational studies, investigators observe outcomes in relation to variables of interest, but they do not assign participants to the study exposure. The most common types of studies are listed here, ordered by the strength of evidence they provide.

Experimental studies. The two types of experimental studies are randomized controlled trials and crossover trials:

- *Randomized controlled trials* are the gold standard of scientific inquiry. In these studies, a group of subjects with similar characteristics is identified. Each subject is then randomly assigned to an intervention or a control group. In this way, participants have an equal and unbiased (ie, random) chance of being assigned to each treatment under study.

These types of trials are well suited to situations in which exposure to treatment is modifiable, a legitimate uncertainty exists regarding the benefit and/or harm of treatment, and outcomes are reasonably common. However, the selection criteria may limit the generalizability of the results.

Depending on the intervention, both participants and investigators may be blinded (ie, not informed) as to which treatment a participant is receiving. Blinding controls for potential placebo effects and the effects of a participant's expectation of benefit. Randomized controlled trials assess the efficacy of the treatment in a controlled setting, which may not reflect its actual effectiveness in a real-world, clinical-practice setting. Often, the trials use a highly defined patient population, so it may not be accurate to extrapolate the results to other patient populations. The Women's Health Initiative is an example of a randomized controlled trial.

- *Crossover trials* allow subjects to serve as their own controls. Participants are randomly assigned to one treatment arm and later switched to the other treatment arm. The crossover study methodology has often been used in trials to assess efficacy of medications, such as in the treatment of vasomotor symptoms.

Observational studies. Types of observational studies include cohort studies, case-control studies, case reports, and case series:

- *Cohort studies* begin with a defined group of subjects (eg, individuals of a certain age or who work in a certain industry) called the cohort. This cohort is then followed over time for a variety of outcomes. Most commonly, data are collected in a similar manner on all participating subjects at the beginning of the study (baseline) and at set intervals during follow-up.

The studies provide a clearer temporal sequence of exposures and/or outcomes, are well suited for rare exposures, and can study multiple exposures and/or outcomes. However, they can be time consuming and expensive, have the potential for bias, and may lose participants during follow-up.

Cohort studies are usually prospective, but they may be retrospective (ie, all relevant exposures and events will already have occurred when the study is initiated). Evidence from prospective cohort studies is considered stronger because data on exposures are collected before the outcomes occur. The Nurses' Health Study and the Framingham Study are examples of large, prospective, cohort studies.

- *Case-control studies* most commonly begin with an outcome of interest (eg, myocardial infarction, breast cancer) and then compare the characteristics of individuals with the outcome (cases) and without the outcome (controls). Data are analyzed using a snapshot approach, determining at a single point in time the differences that may account for the outcome.

Matching subjects for specific characteristics and defining strict eligibility criteria lessens, but cannot eliminate, the possibility that the results are caused by bias. For example, women who use estrogen therapy are known to smoke less and be generally healthier, and this biases any observation of estrogen therapy and health outcomes.

Despite these limitations, case-control studies have many advantages. Because they begin with an outcome of interest, they can be performed efficiently and at less cost than cohort studies. They are important in situations in which it would be unethical to assign individuals to an exposure (eg, asbestos) or when an outcome is relatively rare so that the number of identified cases in any given cohort would be too small to analyze (eg, birth defects).

- In *case reports* and *case series*, the experience of a single patient or series of patients is described. Such reports are useful in bringing new diseases or phenomena to the attention of the clinical and scientific community and for generating new hypotheses. However, without further study, case reports can be considered only suggestive.

Analyses. The results of epidemiologic studies and clinical trials are frequently presented as a relative risk (RR). Other risk-related nomenclature follows:

- The RR tells the estimated magnitude of the change in risk related to the presence versus the absence of a factor of interest.

An RR *less than* 1.0 is associated with lower risk. For example, an RR of 0.50 means that there is a 50% reduction in risk among those with the factor versus those without it. An RR of 0.3 means a 70% reduction in risk.

An RR *greater than* 1.0 means that the factor increases risk. For example, an RR of 1.2 means that there is a 20% increase in risk in the group with the factor versus those without it. An RR of 2.0 means a doubling of risk.

- The *P value* is the probability of obtaining the observed RR (or a more extreme value) by chance.
- The *confidence interval* (CI), usually cited with the RR, indicates the range within which the true magnitude of the measured effect lies. A 95% CI gives the range of values that have a 95% probability of containing the true RR. When a 95% CI does not contain the number 1.0 (eg, 0.40-0.80 or 1.12-1.37), the measured RR is significant by at least $P < 0.05$. A wide CI reflects a wide variation in the data.
- The *odds ratio* is an estimate used in case-control studies that approximates the RR.

The RR is the rate of disease in a group exposed to a potential risk factor, divided by the rate of disease in the unexposed group. For example, if the annual rate of myocardial infarction in women who smoke is 220 per 100,000 and the annual rate in women who do not smoke is 110 per 100,000, the RR associated with smoking would be determined as follows:

$$\text{RR} = \frac{220}{100,000/\text{yr}} \div \frac{110}{100,000/\text{yr}} = 2.00$$

This means that compared with unexposed women, the rate of myocardial infarction for smoking women is twice that of nonsmoking women.

The impact of RR depends on incidence. This can be quantified by *attributable risk* (AR; or risk difference). This is the difference in the incidence rates in the exposed and unexposed groups (the groups with and without the risk factor being studied). The AR

serves to quantify the effect of exposure and thus gives a measure of its public health impact. For example, in the calculation presented earlier, the AR would be as follows:

$$AR = \frac{220}{100,000/\text{yr}} - \frac{110}{100,000/\text{yr}} = \frac{110}{100,000/\text{yr}}$$

This means that for every 100,000 women who smoke, there would be 110 additional cases of myocardial infarctions per year. Depending on the baseline rates of disease, the AR can vary greatly, given the same RR. For example, if the baseline rate of a disease is 6 per 100,000 per year and smoking doubled the risk to 12 per 100,000 per year, the RR would be 2.0; however, the AR would be only 6 per 100,000 per year.

A *meta-analysis* is an analytic technique used to pool the results from many smaller studies. Pooling studies has the effect of increasing the sample size, thereby gaining statistical power. Thus, a meta-analysis may pool the results of clinical trials that are too small to have statistical significance in themselves but that may show significance when pooled. Specific criteria (eg, eligibility criteria of subjects, data completeness) are established to determine which studies will be included in the analysis. Observational studies may also be pooled in a meta-analysis. It must be remembered that any biases present in the contributing studies will be present in the meta-analysis.

Bibliography

Koepsell TD, Weiss NS. *Epidemiologic Methods, Studying the Occurrence of Illness*. New York, NY: Oxford University Press; 2003.

Last JM, Spasoff RS, Harris SS, Thuriaux MC. *A Dictionary of Epidemiology*. 4th ed. New York, NY: Oxford University Press; 2000.

LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997;337:536-542.

McKinlay SM. Issues in design, measurement, and analysis for menopause research. *Exp Gerontol* 1994;29:479-493.

Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of reporting of meta-analyses. *Lancet* 1999;354:1896-1900.

Source: *Menopause Practice: A Clinician's Guide*. Cleveland, OH: The North American Menopause Society; 2004.