POPULATION MEDICINE NEWS

Epidemiology, Preventive Medicine, Public Health
Production Medicine, Computer Applications in Vet Med

September 30, 1991 4(23)

Critical Evaluation of the Scientific Literature

Confounded data

Confounding bias is the great nemesis of empirical research, and the ability to identify it in scientific papers is fundamental to critical evaluation.

Statistical error or "noise" has 2 broad deleterious impacts on our ability to accurately estimate the effect (B) of a treatment or management procedure (Fig 1). Imprecision, we have seen (Vol 4 No 22, Sept 16, 1991), reduces our ability to resolve true effects, but the problem can be avoided by providing adequate sample size. Readers of research reports can almost always evaluate the degree of imprecision and the accuracy of conclusions by, at most, a few simple computations (confidence intervals). Bias, the other deleterious impact of error, is much more insidious, and the reader must seek clues in the methods and results sections. The degree to which an estimate is free of bias is termed its validity.

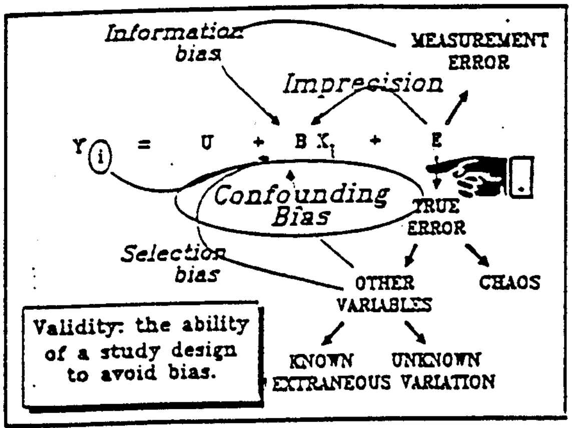

Figure 1. Confounding bias results from differences in extraneous variables between comparison groups.

Fig. 2 contrasts precision and validity. Where imprecision results in a random scatter of the "bullet holes" around the "bull's eye", bias results in the pattern being displaced away from the true effect. Increasing sample size will always improve an imprecise estimate but will only make a biased one worse since the pattern will "tighten" around the wrong center. Several sources of bias exist, but confounding bias is the most difficult for researchers to avoid and thus must receive substantial attention in critically reviewing scientific papers.

As an example of confounding, suppose a clinical trial was attempting to evaluate the effect of iron injections in newborn Holstein heifers on growth rate and morbidity. In the trial herd, like in most US herds, 40% of newborn calves had low passive immune levels as defined by < 5 G/dl total serum protein (TP). Suppose further that in allocating calves to treatment group, the control group ended up with 60% low TP and the treatment group with only 20%. Even if iron injections truly have no effect on morbidity or growth rate, we would not be surprised to see an advantage (especially in morbidity) among injected calves. Our ability to evaluate the effect of iron injections is said to be confounded by TP. Even though we have every reason to believe that TP is an important variable in the health of calves, it is considered an extraneous variable here because it is not the variable of immediate interest. Confounding is produced by extraneous variables that are unequally distributed among the treatment groups. It can either obscure a true effect, create the illusion of an effect where none exists, or give us an effect estimate (eg, for benefit/cost analysis) that is artificially high or low.

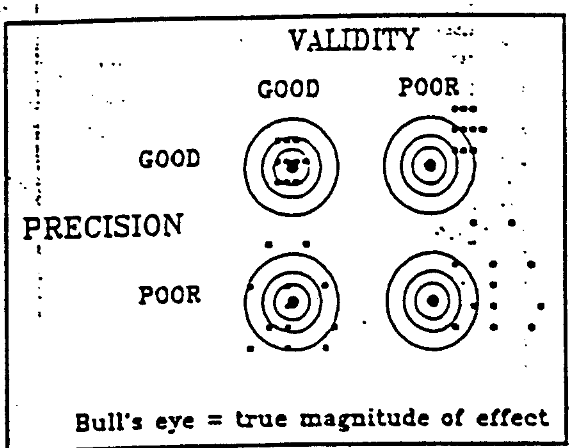

Figure 2. Imprecision results in scatter around the bull's eye. Bias results in a pattern centered away from the bull's eye.

4 ways to control confounding.

Several tools are used to control confounding bias. The following is a brief description of each. Case studies in upcoming issues will relate them to particular scientific articles.

Random allocation. In the iron injection trial, the strong aggregation of low TP in one group should lead us to conclude that effective random allocation was not used. The treatment and control group will not be perfectly identical in all ways even when random allocation is used correctly, but large differences are very unlikely. The beauty of effectively used random allocation is that it will usually distribute, at least roughly, a myriad of potential confounders, both known and unknown. Indeed, random allocation is the only way to control for unknown confounders, and we cannot be fully confident in the validity of any effect estimate where random allocation was not used.

At least 3 good methods of random allocation exist. The best is to use a random numbers table to assign animal identification numbers to treatment. Systematic allocation of alternating animals in some sequential order (say birth order or eartag number) is effective if it is faithfully followed. Another scheme that is sometimes used effectively is to place coded pieces of paper (poker chips, etc) into a paper bag (hat, urn, etc) with each code standing for a different treatment. As each animal becomes available, a marker is drawn for it and it is allocated to the appropriate group; this method could be called "random distribution".

Assessing the random allocation scheme is the single most important aspect of critical reading of articles reporting the results of clinical trials and experiments. Some writers make this simple: they spell out the means of random allocation (on what basis was a particular animal assigned), and they show us the distribution of suspected confounders (say TP) in the different treatment groups. Most do not, and we must look for clues:

Clue 1. If materials and methods states that random allocation was used but does not indicate how it was done, the most likely reason for the omission is that the scheme did not adhere to any conventional standard (ie. the ever-popular Haphazard Allocation Scheme was used instead).

Clue 2. If a random allocation scheme is mentioned in the materials and methods that would be expected to result in roughly equal group sizes but the group sizes reported are greatly different, it is likely that at least some animals found their way into groups without being randomly allocated.

Clue 3. If the descriptive statistics of potential confounders (or any variable even if not a suspected confounder) are greatly dissimilar between treatment groups, it is unlikely that an effective allocation scheme was followed (e.g., differences in TP distribution in the treatment groups of a neonatal calf trial).

Statisticians who have analyzed data from haphazardly "randomized" trials are extremely critical of the above omissions and discrepancies because they have seen the insidious effects of confounding and how common they are unless effective allocation is used. The wary reader will not trust the conclusions of studies in which such problems are observed.

Restriction. Two forms of restriction exist. In **complete restriction**, we limit our trial to animals that meet specified criteria. For example, the above mentioned iron injection trial in neonatal calves could be restricted to calves with TP > 5.0 g/dl. Complete restriction is sometimes used in an attempt to improve power of morbidity/mortality studies. In this use, the trial is limited to high risk individuals--say, calves with TP < 5 g/dl. A smaller sample size is required to detect an effect of specified magnitude if high risk individuals are used. Yet, this use of complete restriction inevitably raises questions of external validity, a concept to be addressed in a later issue.

In **partial restriction** we attempt to provide equal numbers of animals in the strata of potential confounding variables. We randomly allocate within the strata. For example, in the iron injection trial, we could determine a calf's TP status ($<$ 5 or 5+ G/dl) and randomly allocate within each group. This would ensure equal numbers of calves in the TP strata, thus avoiding confounding. In pair matching, the strata consist of only one individual per treatment closely matched on several extraneous variables. Though this approach has intuitive appeal, it is usually unnecessary in properly randomized and analyzed trials. We should not necessarily discriminate against studies that do not use partial restriction even when obvious confounders exist.

Physical control of extraneous variation. In certain types of scientific studies (clinical trials and observational studies), the animals under observation are not maintained in a carefully controlled research facility but are dispersed in different locations, fed different diets, housed somewhat differently, and generally exposed to wide differences of environment. If animals are randomly allocated across the different types of environments, failure to physically control the environment does not introduce confounding. If randomization is not possible (observational studies), differences in environment introduce some major difficulties in interpretation.

Statistical control of confounding. Avoiding confounding does not require that treatment groups be identical with respect to potential confounders as long as we know what the confounding variables are and measure them for each individual in the trial. A toolbox of statistical methods has been developed to segregate the effect of confounders from the effect of interest (e.g., the treatment effect). An important bonus of this segregation is that, by removing the effects of selected extraneous variables, we reduce the noise and thus increase the power of the study. In the iron injection trial the segregation of TP from error would occur by extending the model:

$$Y = BX + CT + E$$

where Y is, say, number of sick days for a given calf, X is the treatment group (iron injection=1 or no injection =0), B is the effect of iron injection vs no injection, T is the total protein of a given calf, C is change in Y for each increment change in TP (ie, the slope), and E is error. Compared to the simpler model in which the effect of TP is contained in E, the power of the model wit CT is greater since E has been reduced (assuming that C is non zero and that TP does have an effect on number of sick days).

Anyone who follows the scientific literature, particularly the Animal Sciences, encounters mathematical models far more complex than the simple example shown here. A full grasp requires a few graduate level statistics courses, but an intuitive understanding is needed to critically evaluate articles containing such models. Simply put, their purpose is to isolate the effects of selected confounders from error. This results in 2 benefits:
(1) the distortion of confounding bias is eliminated if the treatment groups were not identical with respect to the confounders, and our estimate of treatment effect is thereby "adjusted" as if the confounders did not exist; and
(2) by reducing the overall noise in the model we get more power out of a given sample size and can thus get a more precise estimate of treatment effect.
Returning to the target shooting metapho of Fig 2, statistical control of confounders helps to center our shot pattern around the bull's eye and to tighten up its distribution. What statistical models cannot do, however, is eliminate the confounding due to variables not in the model (ie, unknown extraneous variables). This point is key to interpreting observational studies, a subject to be taken up later.

Coming in the next issue: critical review of allocation in 3 clinical trials evaluating prostaglandin therapy for postpartum disease in dairy cows.

POPULATION MEDICINE NEWS is produced by the Washington State University Field Disease Investigation Unit, Pullman, WA 99164-6610 (509-335-0711) and is supported in part by the WSU Cooperative Extension Service. Viewpoints expressed are not necessarily those of the University or its Administration. This newsletter may be copied by anyone. Editors: Dale Hancock and Susan Waller.