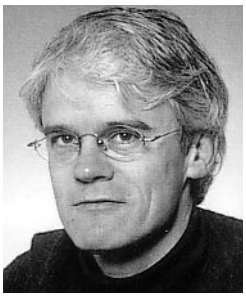
Editorial

Bias in Its Coat of Many Colors



Science is about approaching and understanding the truth, and this constant quest is hindered by error that is inherent in observation. Error comes in two varieties: random and systematic. Random error is caused by the play of chance that affects both the observer and the observed: most observable phenomena fluctuate around a certain mean value, and observers fluctuate around a mean aptitude to detect that phenomenon. Discrete events carry a chance to occur at a certain moment, but the exact time is hard to predict. Random error is expressed in the familiar statistical parameters such as a confidence interval around a mean. Increasing the number of observations combats this type of error: as this number increases, the chance increases that the mean of the observations is close to the true mean. Systematic error is usually called bias. Bias implies a flaw in the study design or measurement process that cannot be counteracted with increased numbers of observations. In clinical studies bias is usually caused by a flawed comparison between groups or states. It is the bane for the investigator, but a boon for epidemiologist-editorialists[1].

Clinical studies usually come in two varieties: the study of causes of disease, and the study of disease course over time, with or without treatment. In many instances observational studies are the most important source of information for the clinician. In the first step, one or a series of patients with features in common are described: they were exposed to the same putative risk factor or were treated with the same drug and had a notable outcome. In such cases or series the comparison is implicit: the author compares the results with results expected in "the population," however defined. For example, Semmelweis in 19th century Vienna knew what proportion of women on his ward died from sepsis after childbirth, and contrasted this with the mortality after introducing antiseptic procedures. It is clear that implicit comparisons carry great risks in terms of the potential for bias. However, very large effects can overcome bias.

The next step is to form an explicit control group to compare against. Thus, predefined groups of patients can be compared in terms of exposure to a putative causal or therapeutic agent and in terms of outcome (presence or absence of disease, improvement or healing). In Semmelweis' case, one of the reasons for his tenacity in trying to resolve the mystery was that another ward in the same hospital had much lower rates of sepsis and death after childbirth. This ward provided the same care, and finding the difference between the wards (the former trained midwives whereas Semmelweis' ward trained medical students who performed autopsies and transmitted bacteria to the mothers) in the end provided the clue to the solution[2]. When the groups are formed based on the outcome, the design is a case-control study; when the groups are formed based on the exposure before the outcome has occurred, the design is a cohort study. Finally, a special form of the cohort study is the randomized trial, only usable for treatment comparisons.

All these steps are taken to decrease the chance of bias. Each step increases the complexity and costs (time and money) of the study, so choosing the level of bias control involves a trade-off.

In studies with control groups that are not randomized, the most important forms of bias are confounding, recall, and detection bias[3]. Confounding occurs when, unbeknownst to the investigator, distorting factors are present that are related both to the exposure and the outcome[4]. For example, a study on a possible relationship between alcohol consumption and lung cancer would be useless if one was not aware of (and "controlled for") the smoking habits of the patients studied. Recall bias occurs when the exposure is remembered (and thus detected) better by those suffering from the putatively related outcome than patients not having this outcome. For example, patients with gastrointestinal bleeds may remember exposure to nonsteroidal antiinflammatory drugs better than asymptomatic patients. Detection bias occurs when the ability to detect the outcome is affected by the exposure. For example, women with cancer of the

See Bias in uncontrolled therapeutic trials in rheumatology due to selection of populations with extreme characteristics, *page 1881*

uterus may be detected earlier when exposed to estrogen, because the tumor will be more likely to "expose" itself with an abnormal bleeding pattern.

In randomized studies, bias can be caused by improper blinding and inadvertent influence on the randomization process (so that the balance in known and unknown prognostic factors between the groups is lost) or in the unblinded assessment of outcome. It is important to note that where in principle bias can work in all directions, i.e., inflate an already present association, reduce or even reverse such an association, or detect a non-existing association) the pressure to publish will usually work in the direction of newsworthy findings (publication bias).

In this issue, Ioannidis, *et al* take us back to biases present in uncontrolled series[5]. They specifically focus on the before–after comparison in uncontrolled treatment trials. In brief, they point out that patients in treatment trials are selected for being in a bad state: otherwise, they would not be eligible for the new treatment being described. The situation is not different when the trial is controlled. The key point, however, is that in the uncontrolled situation, all subsequent improvement is assumed to be caused by the treatment. In truth, any such improvement is the result of regression to the mean, natural course of the disease, placebo effect, true treatment effect, and random error. In controlled trials the nontreatment factors are assumed to work on both groups, so that the between-group comparison focuses on the true treatment effect. There is no simple remedy for this. Some of the regression effects can be countered by introducing a qualifying period, in which patients have to stay in their adverse state. In such a period patients with "chance," or short peaks of, high disease activity will drop out and not be studied further. Only when one is reasonably sure that patients will not spontaneously improve (i.e., regress to the mean) are they allowed into the study.

In sum, reports of case series and individual observations remain important as an "early warning system" and inform on potential treatment breakthroughs. Rarely is the described effect so large that the observation is conclusive in and of itself. In all other cases, such observations must always be interpreted with caution and followed up by controlled observations.

MAARTEN BOERS, MSc, MD, PhD,
Professor of Clinical Epidemiology,
Department of Clinical Epidemiology and Biostatistics,
Vrije Universiteit Medical Center,
PO Box 7057,
1007 MB Amsterdam, The Netherlands

REFERENCES

1. Sackett DL. Bias in analytic research. J Chronic Dis 1979;32:51-63.
2. Schmidt F. Ignaz Philipp Semmelweis (1818-1865): The saviour of mothers. 1998: http://www.geocities.com/Athens/Ithaca/1202/Semmelw.htm
3. MacMahon S, Collins R. Reliable assessment of the effects of treatment on mortality and major morbidity, II: observational studies. Lancet 2001;357:455-62.
4. Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. Lancet 2001;357:373-80.
5. Ioannidis JPA, Karassa FB. Bias in uncontrolled therapeutic trials in rheumatology due to selection of populations with extreme characteristics. J Rheumatol 2001;28:1881-7.