

MEDICAL WRITINGS

Notice: This Material May Be
Protected by Copyright Law
(Title 17 U.S. Code).

The Landscape and Lexicon of Blinding in Randomized Trials

Blinding in medical research possesses a rich history spanning a couple of centuries (1). Most researchers and readers grasp its meaning. Unfortunately, beyond that general understanding lies confusion. In addition to terms such as “single blind” and “double blind” meaning different things to different people, some steadfastly refuse to use the term “blinding” and insist instead on the term “masking.” Others confuse blinding with other methodologic precautions, such as concealment of allocation during the process of creating comparison groups. Still others consider that randomization is of little use unless accompanied by “double-blinding,” thus revealing that they have not understood that these separate aspects of methodology address separate sources of bias.

A recent survey addressed whether the process historically termed “blinding” should be termed “masking” (2). The survey revealed a lack of accord on that question and inconsistencies concerning other blinding terminology. Although many resources address the lexicon of blinding, including clinical trial textbooks (3-5), clinical trial dictionaries (6, 7), and a recently released epidemiology dictionary (8), these sources do not entirely clear the lexicographic fog. Indeed, a recent study found that investigators, textbooks, and published articles all varied greatly in their interpretations of single-, double-, and triple-blinding (9). In other words, terminologic tangles abound with blinding. We delve into the landscape and lexicon of blinding in randomized trials in the hope of untangling some of that terminology.

SYNOPSIS OF THE HISTORY OF BLINDING

Scientists sometimes portray blinding as a recent methodologic achievement, but researchers have used blinding for more than 200 years. Lavoisier and Franklin introduced blinding in the late 18th century to test therapeutic claims made for Mesmerism—a therapy founded on the notion that magnetism had healing properties (1, 10). Toward the middle of the 19th century, many homeopaths used blinding in their “provings” and in comparisons of homeopathy with mainstream medicine (11). By the late 19th century, psychological researchers began to use blinding for traditional questions, more to minimize bias than to expose

fraud (1). The beginning of the 20th century found some physiologists and pharmacologists, particularly in Germany, using blind assessment. That use became more frequent in Germany by the 1930s.

Researchers in Britain and the United States developed interest in blind assessment, but a different rationale motivated their interest. The rationale in Germany for blinding centered on the elimination of bias. In contrast, the interest in Britain and the United States initially centered on preventing attrition problems (1). Without blinding and a placebo intervention, recruiting and retaining participants for a no-intervention control group became daunting. “For Anglo-American clinical researchers, the initial adoption of a placebo sham in an experiment was an architectural device to create a viable and camouflaged concurrent no-treatment arm in a clinical trial” (1). Toward the end of the 1930s, British and U.S. researchers also began to acknowledge the benefits of blinding in avoiding bias.

The evolution of the randomized, controlled trial during the first half of the 20th century promoted greater use of blinding. Properly concealed random allocation to comparison groups abolished selection bias at entry to a trial, and clinical investigators began to appreciate fully the biases that could affect studies after participants had entered a trial. That realization transferred greater credibility to blinding arguments (1). We recommend Kaptchuk (1) for a historical account.

BACKGROUND

Blinding is intended to reduce bias in medical research. Although blinding is often associated in people’s minds with randomized, controlled trials, it can be used in a variety of study designs to reduce observer biases (12). For example, investigators can assess outcome measures blinded to exposure status in nonrandomized cohort studies or exposure status blinded to case or control status in case-control studies. Indeed, when investigators first used blinding in the 18th century, they assessed the effects of Mesmerism in nonrandomized experiments (1, 10). Having noted this, however, we focus on blinding in the context of randomized comparisons of interventions.

Blinding is widely recognized as reducing differen-

tial assessment of outcomes of interest (known as ascertainment bias, information bias, or observer bias), prompted by knowledge of the group assignment of individuals being observed (3–5). Blinding is less frequently recognized as also operationally improving compliance and retention of trial participants and reducing biased supplemental care or treatment (sometimes called co-intervention) (3, 5). We provide glimpses of the potential disadvantages of participants', investigators', and outcome assessors' knowing the intervention group to which the participants have been assigned. In many cases, the biases that result might well be subconscious, but they are biases nonetheless.

Possible Consequences of Participants' Knowing

Psychological effects could arise from participants' knowing that they have received a “promising” new treatment, a thoroughly tested standard treatment, an untested new treatment, or a “disappointing” standard treatment. In other words, how the treatment options are perceived may influence the way in which they are evaluated. Despite evidence suggesting that new treatments are as likely to be inferior as they are to be superior to standard treatments (13), we have the impression that participants generally assume that new treatments will be better than standard treatments. In any case, knowledge of the intervention received can affect the psychological or physical responses of the participants (3–5).

Furthermore, knowledge of the intervention could influence participants' cooperation. For example, if participants believe that they were assigned to what they perceive as an inferior intervention, they may not comply well with the regimen. Moreover, they may not adhere to follow-up procedures, leading to a potentially biased loss to follow-up.

Possible Consequences of Investigators' Knowing

We define investigators in an aggregate sense to include a broad trial team—for example, trial designers, participant enrollers, randomization executors, health care providers, intervention counselors, and routine-data collectors. Investigators particularly pertinent to blinding include health care providers (such as an attending physician or nurse) and intervention counselors (for example, someone delivering a behavioral prevention mes-

sage) who interact with the participants throughout the trial. The inclinations of investigators for or against the interventions can be directly transferred to participants by their attitudes (14). Their inclinations may also be manifested in, for example, differential use of ancillary interventions of supplemental care or treatment (co-interventions). Of note, the implementer could also encourage or discourage continuation in the trial on the basis of knowledge of the intervention group assignment.

Possible Consequences of Outcome Assessors' Knowing

When they know the intervention group assignment of the participants whom they are assessing, outcome assessors with inclinations for or against any of the interventions being compared may make biased assessments. For example, if they believe the new intervention is superior, then they could register more generous responses to that intervention. Obviously, more subjective outcomes present greater opportunities for bias. Pain scores assessed by participants are a good example of a subjective outcome. Even some outcomes considered objective can be fraught with subjectivity—for example, pelvic inflammatory disease and myocardial infarction.

In general, blinding becomes less important to reduce observer bias as the outcomes become less subjective. “Hard” outcomes leave little room for bias. For example, knowledge of the intervention would have little effect on measuring a “hard” outcome, such as death (but still could influence the attributed cause of death). Of importance, even when participants and investigators have not been blinded, blinding of outcome assessors is often possible and advisable (12).

PLACEBOS AND BLINDING

Blinding frequently leads to the use of placebos. Placebos may or may not have effects mediated through psychological mechanisms, but they are administered to participants in a trial because they are otherwise “inactive.” An “active placebo” is a placebo with properties that mimic the symptoms or side effects (for example, dry mouth, sweating) that might otherwise reveal the identity of the (pharmacologically) active test treatment. The effect, in practice, of using placebos is contentious (15), but the widespread view remains that placebos should be administered, whenever possible, to partici-

pants in control groups when assessing the effects of proposed new treatments for a condition for which no effective treatment already exists (3, 4). Placebos are generally used in trials of drugs, vaccines, and other medicinal interventions but can sometimes also be used in trials of procedures, such as ultrasonography, acupuncture, and, occasionally, surgery. For example, perhaps when placebo treatments are inappropriate, some closely analogous approach, such as placebo wound dressings, may suffice in a trial of laparoscopic versus open appendectomy (16).

When an effective standard treatment exists, it is frequently used in the control group for comparison against a new treatment. Thus, investigators might compare two active treatment groups without a placebo group. Even then, however, investigators frequently attempt to achieve blinding by using the “double-dummy” method—in essence, two placebos (12, 17). For example, in comparing two agents, one in a blue capsule and the other in a red capsule, the investigators would prepare blue placebo capsules and red placebo capsules. Then both treatment groups would receive a blue and a red capsule, one active and one inactive.

TERMINOLOGY TANGLES

Double Definitions of “Double-Blinding”

Blinding (masking) describes a situation in which knowledge of intervention assignments is hidden from participants, investigators, or outcome assessors in a trial. The term “double-blind” denotes a trial in which the participants, investigators, and assessors all remain unaware of the intervention assignments throughout the trial. Given that three groups are kept ignorant, the term “double-blind” is sometimes misleading. In medical research, however, the same individual often is both investigator and assessor, so in that instance, the terminology accurately refers to two categories.

The term “single-blind” denotes a trial in which one of the three categories of individuals remains unaware of the intervention assignments throughout the trial. Typically, however, it connotes a trial in which the participant, not the investigator, remains ignorant (4). Moreover, a single-blind trial may confusingly refer to the participant and implementer both knowing the intervention but the assessor remaining unaware of it.

The terms “nonblind” and “triple-blind” also deserve attention. “Nonblinded” (open label) denotes trials

in which all three categories of individuals know who has received which interventions throughout the trial. When investigators use the term “triple-blind,” they usually mean a double-blind trial that also maintains a blind data analysis (3). Some investigators, however, denote trials as triple-blind if investigators and assessors are distinct people and both, as well as participants, remain unaware of assignments. Investigators rarely use the term “quadruple-blind,” but some use it to denote blinding of participants, investigators, assessors, and data analysts (7).

Readers may note some fuzziness in our discussions of single-, double-, and triple-blinding. That fuzziness reflects true ambiguity. Indeed, universally accepted definitions have eluded the scientific community, and scanty reporting on blinding pervades the literature. Authors have frequently reported their study only as double-blind and not provided much further clarifying information (18–21). Of concern, a recent study (9) found that the term “double-blind” was interpreted differently among authors, readers, and “experts.” A survey of physicians and a review of recent textbooks and reports revealed numerous interpretations of the designation “double-blind.” For example, some thought it meant that the patients (participants) and clinicians (investigators) were blinded, whereas others thought that the patients and outcome assessors were blinded (9). In sum, investigators do not define “double-blinding” consistently, and to make matters worse, they frequently fail to report their definitions clearly in their articles.

Most important, we urge that authors explicitly state what steps were taken to keep whom blinded. If they choose to use terminology such as single-, double-, or triple-blinding in reporting randomized controlled trials, they should explicitly define those terms. When we use the term “double-blinding” in this article, we are referring to a situation in which steps have been taken to blind participants, investigators, and outcome assessors to group assignments.

“Blinding” or “Masking”

Some people use the term “masking” in place of “blinding.” Similarly, “double-blinded” becomes “double-masked.” For example, one prominent text uses blinding and double-blind (3), whereas another prominent text uses masked and double-masked (4). With that disparity, we are not surprised that researchers ques-

tion and debate the proper terminology. Remarkably, few researchers have investigated the effect of terminology on trial participants or the meaning of terminology to investigators themselves (9).

“Masking” may have originated relatively recently as a euphemism for “blinding” in trials involving participants who have impaired vision. In those trials, “masking” may indeed be more appropriate because it may be less upsetting and offensive. It also would be less confusing in trials in which blindness is an outcome. In trials not addressing impaired vision, however, the term “blinding” does not seem to be upsetting or offensive to trial participants (2).

Both the process and the terminology of blinding, furthermore, appear to be well understood by most participants and investigators. That understanding evolved over two centuries of use (1). Blinding became ingrained in the principles of medical research. That blinding terminology seems to be universally recognized represents a lofty achievement for any term, let alone a methodologic term. Discarding a widely understood and instantly identifiable term would seem unwise, particularly in trials not addressing impaired vision, unless evidence to the contrary emerges.

In addition, blinding terminology conveys a strong bias-prevention message. Apparently, “blinding” terminology emerged when Benjamin Franklin and colleagues actually blindfolded participants to shield them from knowledge in their evaluations of the therapeutic claims made for Mesmerism (10). The visual imagery of blindfolding, a total covering of the eyes, conveys stronger bias prevention than masking, in which eye holes may permit extensive viewing. Moreover, the International Conference on Harmonization (ICH) guidance primarily uses “blinding” terminology (22). (The ICH is an intensive tripartite collaboration between regulatory authorities in Europe, Japan, and the United States to develop common guidelines for the design, execution, and the reporting of clinical trials.) The long history, pervasive general understanding, strong visual imagery, and adoption by the ICH lead us to suggest that “blinding” should remain the predominant terminology. With global electronic access to articles, if authors use “masking,” we feel they risk the loss of comprehension in many parts of the world. Clearly, however, some investigators and editors prefer “masking” terminology (23).

“Blinding” and “Allocation Concealment”

The success of randomization depends on two interrelated processes (24–26). First, an unbiased allocation sequence must be generated, preferably by some random procedure. Second, strict implementation of that schedule must be secured through an assignment process that prevents foreknowledge of treatment assignment (24–26). That second strict implementation process, which we term “allocation concealment,” is often confused with blinding. Allocation concealment prevents those who admit patients to a trial from knowing the upcoming assignments.

To judge from reports of controlled trials (25, 27, 28), the crucial importance of the allocation concealment process has not been widely recognized until recently (24, 29). Any notion of its importance seems to have been usurped under the rubric of “double-blinding.” Moreover, one paper that gave rare recognition to the distinct process of “allocation concealment” termed it “randomization blinding,” perhaps yielding further confusion (30).

“Allocation concealment” refers to a distinct process and should not be confused with blinding according to the terms in which we have described the latter. Allocation concealment seeks to prevent selection bias, protects the allocation sequence *before and until* assignment, and can *always* be successfully implemented regardless of the study topic (24, 25). In contrast, blinding seeks to prevent ascertainment bias, protects the sequence *after* allocation, and cannot always be implemented—for example, in trials comparing surgical with medical treatments. Thus, allocation concealment up to the point of assignment of the intervention and blinding after that point address different sources of bias and differ in their practicability. In light of those considerations, we introduced the term “allocation concealment” (24, 25) to describe the process used to prevent foreknowledge of intervention allocations before assignment. We exclusively reserve the term “blinding” for measures taken to conceal group identity after assignment.

“Double-Blinding” and “Randomization”

“Double-blinding” is intertwined with randomization. Apparently, double-blinding emerged as a concept more easily understood than randomization. With that understanding came confused terminology. One egregious confusion particularly annoys us. Our impression

garnered over years of observation suggests that many investigators and readers delineate a randomized trial as high quality if it is “double-blind,” as if double-blinding is the sine qua non of a randomized controlled trial. For example, a widely used scoring system for the quality of randomized trials accounts for double-blinding but not directly for allocation concealment (31). A randomized trial, however, can be methodologically sound in terms of controlling selection bias (proper randomization) and not be double-blind or, conversely, double-blind and not methodologically sound in terms of controlling selection bias. Although double-blinding reflects good methods, it should not be used as a surrogate marker of overall trial quality. As we discuss below, adequate allocation concealment actually appears the more important indicator. Moreover, many trials cannot be double-blinded. Those trials must be judged on merit and not on an inapplicable standard based on double-blinding.

DOES BLINDING PREVENT BIAS?

Although we feel that investigators, readers, and editors sometimes overestimate the importance of “double-blinding,” we do not suggest that blinding is unimportant. Intuitively, blinding should reduce bias, particularly in estimating effects on some kinds of outcomes. More important, a systematic review of the empiric methodologic evidence supports that intuition (32).

Although not double-blinding appears to introduce bias, its average effect—exaggerating estimates by about 19% (32)—appears weaker than that of allocation concealment. Trials with inadequate or unclear allocation concealment have been shown to yield larger estimates of treatment effects compared with those that used adequate concealment (on average, 41% and 33%, respectively) (24), with commensurate results when the inadequate and unclear categories are lumped (29, 32). Double-blinding appears important in preventing bias but not as important as allocation concealment.

Blinding must succeed to reap its benefits. Investigators who use blinding can assess the success of the blinding by directly asking participants, investigators, or outcome assessors which intervention they think was administered. In principle, if blinding was successful, these individuals should not be able to do better than chance when guessing the intervention. In practice, adverse side effects may sometimes provide strong hints about the

intervention. Furthermore, individuals may be reluctant to expose their unblinding efforts, if undertaken, by providing accurate responses to the queries—in other words, if they have deciphered group assignments, they might not be willing to incriminate themselves as having done so. To be sure, testing the success of blinding involves some difficulties in interpretation. Nevertheless, if investigators attempt to judge the success of their blinding, they should provide the results of those attempts. At the least, investigators need to report any failure of the blinding procedure, such as use of nonidentical placebo or active preparations.

SUMMARY

“Blinding” terminology rests on a long history, evokes widespread understanding, creates strong impressions, and pervades the international harmonization guidelines. We prefer “blinding” over “masking” terminology. The term “double-blinding” cannot be a surrogate marker for the methodologic quality of a randomized controlled trial. Each trial must be judged on its own merit, with blinding being just one component. Another, more important component—allocation concealment—should not be confused with blinding. A trial can use an adequate allocation concealment mechanism but use a totally inadequate blinding method.

Trial reports would benefit from authors’ following the CONSORT guidelines (*Consolidated Standards of Reporting Trials*) (33). They should eschew using only the single-, double-, triple-blinding terminology in favor of explicitly stating who was blinded and, to the extent possible, if blinding was successful. If they use terminology such as “double-blind,” they should explicitly define it. Moreover, if the authors contend that trial participants, investigators, and outcome assessors were blinded, then the authors should describe, at a minimum, the mechanism (such as capsules, tablets) and the similarity of treatment characteristics (for example, appearance, taste, administration). Readers deserve information on those elements to judge the adequacy of blinding attempts. Moreover, better reporting might inspire better understanding.

Kenneth F. Schulz, PhD, MBA
Family Health International
Research Triangle Park, NC 27709

Iain Chalmers, DSc

United Kingdom Cochrane Centre
Oxford, United Kingdom OX2 7LG

Douglas G. Altman, DSc

ICRF Medical Statistics Group, Centre for Statistics in Medicine,
Institute of Health Sciences
Oxford, United Kingdom OX3 7LF

Acknowledgment: The authors thank Dr. Ted J. Kaptchuk for his review of earlier drafts of this manuscript.

Requests for Single Reprints: Kenneth F. Schulz, PhD, MBA, Family Health International, Box 13950, Research Triangle Park, NC 27709.

Current Author Addresses: Dr. Schulz: Family Health International, Box 13950, Research Triangle Park, NC 27709.

Dr. Chalmers: UK Cochrane Centre, NHS Research and Development Programme, Summertown Pavilion, Middle Way, Oxford, United Kingdom OX2 7LG.

Professor Altman: ICRF Medical Statistics Group and Centre for Statistics in Medicine, Institute of Health Sciences, Old Road, Headington, Oxford, United Kingdom OX3 7LF.

Ann Intern Med. 2002;136:254-259.

References

- Kaptchuk TJ. Intentional ignorance: a history of blind assessment and placebo controls in medicine. *Bull Hist Med.* 1998;72:389-433. [PMID: 9780448]
- Lang T. Masking or blinding? An unscientific survey of mostly medical journal editors on the great debate. *MedGenMed.* 2000;E25. [PMID: 0011104471]
- Pocock SJ. *Clinical Trials: A Practical Approach.* Chichester, UK: Wiley; 1983.
- Meinert CL, Tonascia S. *Clinical Trials: Design, Conduct, and Analysis.* New York: Oxford Univ Pr; 1986.
- Friedman LM, Furberg CD, DeMets DL. *Fundamentals of Clinical Trials.* 3rd ed. St. Louis: Mosby; 1996.
- Meinert CL. *Clinical Trials Dictionary.* Baltimore, MD: The Johns Hopkins Center for Clinical Trials; 1996.
- Day S. *Dictionary for Clinical Trials.* Chichester, UK: Wiley; 1999.
- International Epidemiological Association. *A Dictionary of Epidemiology.* Last JM, ed. 4th ed. Oxford: Oxford Univ Pr; 2001.
- Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA.* 2001;285:2000-3. [PMID: 11308438]
- Franklin B, Bailly JS, Lavoisier AL. Rapport des Commissaires chargés par le roi, de l'examen du magnetisme animal. A Nice: Chez Gabriel Floteron; 1785.
- Dean ME. A homeopathic origin for placebo controls: 'an invaluable gift of God'. *Altern Ther Health Med.* 2000;6:58-66. [PMID: 10710804]
- Day SJ, Altman DG. Statistics notes: blinding in clinical trials and other studies. *BMJ.* 2000;321:504. [PMID: 10948038]
- Chalmers I. What is the prior probability of a proposed new treatment being superior to established treatments? [Letter] *BMJ.* 1997;314:74-5. [PMID: 9001503]
- Wolf S. Effects of suggestion and conditioning on action of chemical agents in human subjects—pharmacology of placebos. *J Clin Invest.* 1950;29:100-9.
- Hróbjartsson A, Gøtzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *N Engl J Med.* 2001;344:1594-602. [PMID: 11372012]
- Lintula H, Kokki H, Vanamo K. Single-blind randomized clinical trial of laparoscopic versus open appendectomy in children. *Br J Surg.* 2001;88:510-4. [PMID: 11298617]
- Altman DG. *Practical Statistics for Medical Research.* London: Chapman and Hall; 1991.
- Mosteller F, Gilbert JP, McPeck B. Reporting standards and research strategies for controlled trials: agenda for the editor. *Control Clin Trials.* 1980;1:37-58.
- Gøtzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials.* 1989;10:31-56. [PMID: 2702836]
- DerSimonian R, Charette LJ, McPeck B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med.* 1982;306:1332-7. [PMID: 7070458]
- Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ.* 1996;312:742-4. [PMID: 8605459]
- International conference on harmonisation; guidance on statistical principles for clinical trials; availability—FDA. Notice. *Fed Regist.* 1998;63:49583-98. [PMID: 10185190]
- Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med.* 2001;134:663-94. [PMID: 11304107]
- Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA.* 1995;273:408-12. [PMID: 7823387]
- Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA.* 1994;272:125-8. [PMID: 8015122]
- Schulz KF. Subverting randomization in controlled trials. *JAMA.* 1995;274:1456-8. [PMID: 7474192]
- Altman DG, Doré CJ. Randomisation and baseline comparisons in clinical trials. *Lancet.* 1990;335:149-53. [PMID: 1967441]
- Adetugbo K, Williams H. How well are randomized controlled trials reported in the dermatology literature? *Arch Dermatol.* 2000;136:381-5. [PMID: 10724201]
- Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet.* 1998;352:609-13. [PMID: 9746022]
- Chalmers TC, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. *N Engl J Med.* 1983;309:1358-61. [PMID: 6633598]
- Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials.* 1996;17:1-12. [PMID: 8721797]
- Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ.* 2001;323:42-6. [PMID: 11440947]
- Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med.* 2001;134:657-62. [PMID: 11304106]