

UNBIASED RESEARCH AND THE HUMAN SPIRIT: THE CHALLENGES OF RANDOMIZED CONTROLLED TRIALS

Kenneth F. Schulz, PhD, MBA

Abstract • Résumé

Research by Klein and associates provides useful information on the relation between episiotomy and outcomes such as perineal trauma, but the methodologic implications of their work are especially fascinating. Physicians who participated in their randomized controlled trial (RCT) were supposed to adhere to a policy of either liberal or restrictive use of episiotomy according to the study arm to which each patient was assigned. However, some used the procedure for approximately 90% of patients regardless of allocation. Klein and associates' post-hoc study (see pages 769 to 779 of this issue) sheds light on the relation between physician attitudes and the practice of episiotomy. The author contends that the noncompliance encountered by Klein and associates reflects the fact that randomized trials are anathema to the human spirit. He offers suggestions for making RCTs more meaningful and stresses that, although RCTs are indispensable to the advancement of medical knowledge, they necessitate assiduous attention to matters of design and implementation.

Une recherche de Klein et de ses collaborateurs fournit les renseignements utiles sur le lien entre l'épisiotomie et des résultats comme les traumatismes périnéaux, mais les répercussions méthodologiques de leurs travaux sont particulièrement fascinantes. Les médecins qui ont participé à leur étude contrôlée randomisée (ECR) devaient utiliser l'épisiotomie à volonté ou de façon limitée, selon le volet de l'étude auquel chaque patiente était affectée. Certains ont toutefois pratiqué l'intervention chez environ 90 % des patientes, sans tenir compte de l'affectation. L'étude ultérieure de Klein et de ses collaborateurs (voir pages 769 à 779 du présent numéro) explique le lien entre les attitudes des médecins et la pratique de l'épisiotomie. L'auteur affirme que l'inobservation constatée par Klein et ses collaborateurs démontre que les études randomisées sont un anathème pour l'être humain. Il présente des suggestions pour rendre les ECR plus significatives et insiste sur le fait que si elles sont indispensables au progrès des connaissances médicales, ces études obligent à porter une attention assidue aux questions de conception et de mise en oeuvre.

Klein and associates' conscientious and thoughtful report on the relation between physician beliefs and the use of episiotomy (see pages 769 to 779 in this issue) augments their earlier work on episiotomy and postpartum outcomes.^{1,2} Their results constitute a persuasive argument for more restricted use of episiotomy. Their studies also present fascinating methodologic implications.

The original study¹ was a randomized controlled trial (RCT). Unfortunately, its usefulness was compromised by poor physician compliance with the trial protocol. Depending upon the random allocation of patients to the "treatment" or "control" study arms, contributing physicians were to have followed a policy of either liberal or restrictive use of episiotomy. However, Klein and

associates detected that many did not alter their use of episiotomy as required and employed episiotomy approximately 90% of the time for participants in either trial arm. Not surprisingly, the culprits turned out to be those who viewed episiotomy favourably.

Advances in health care depend upon the knowledge gained from unbiased studies. Because RCTs provide the only hope of eliminating selection biases from investigations, they serve as the foundation for advancing medical science.

However, Klein and associates' work exemplifies another side of RCTs: they are anathema to the human spirit. Researchers need to realize that, given the opportunity, trial implementers will frequently subvert the intended aims of random assignment. Subversion can be

Dr. Schulz is in the Division of STD/HIV Prevention, National Center for Prevention Services, Centers for Disease Control and Prevention, Atlanta.

Reprint requests to: Dr. Kenneth F. Schulz, Division of STD/HIV Prevention, CDC Mailstop E02, 1600 Clifton Road NE, Atlanta GA 30333

prevented or deflected, however, with painstaking attention to design and implementation.

THE BEST-LAID PLANS

Klein and associates are to be commended for their fine work. First, they executed an intention-to-treat analysis of their RCT. They termed further nonrandomized comparisons "post-hoc cohort" analyses and the present study an "exploratory and hypothesis-generating exercise." They appear to have analysed and reported their results responsibly.

Their RCT was, however, seriously affected by physician noncompliance with the randomly assigned therapy. Physicians who viewed episiotomy favourably or very favourably performed the procedure with similar frequencies regardless of the trial arm to which the participants had been assigned. Obviously, trials in which so much therapeutic contamination occurs are of limited value. Hence, Klein and associates subjected their results to further research and analysis. Because the results of their post-hoc studies are perhaps more meaningful than those of the original trial, one might surmise that doing an RCT under such circumstances wastes time, energy and money. Why not just begin with a nonrandomized, observational study?

Part of the answer resides in the fact that the investigators did not expect to encounter such levels of noncompliance. This point aside, however, we should be circumspect about reflex reactions to nonrandomized studies. They have certain characteristics that lead to biases of some degree. Moreover, lack of empirical research into the quality standards for observational studies makes their conduct and interpretation fraught with difficulty. In contrast, RCTs are demonstrably less susceptible to bias than nonrandomized studies³⁻¹⁰ and apply quality standards that have been extensively evaluated.^{3,4,7-20} Granted, observational studies are useful and necessary when an RCT proves impossible. When RCTs are feasible, however, they should be preferred. The conditions under which Klein and associates conducted their trial strained the limits of feasibility.

NONCOMPLIANCE AND ALLOCATION CONCEALMENT

Selection bias could have been introduced into Klein and associates' trial through the manipulation of assignments. The physicians who viewed episiotomy more favourably decided more frequently not to randomly assign certain participants, who had been enrolled, to a study arm.¹² Did they have knowledge of the next assignment that led to the decision not to allocate participants who, from the physicians' point of view, would

have been allocated to the "wrong" group? Moreover, did knowledge of the next assignment lead physicians to direct some participants to a "desired" group? Who would doubt those possibilities, given the performance of some physicians once the participants had been allocated?

Whether selection bias of this kind is introduced hinges largely upon the adequacy of the random-allocation process. Random allocation in a trial should involve both generating an unpredictable assignment sequence and concealing that sequence until participants are actually allocated. Many medical researchers inaccurately regard the sequence-generating process as "random allocation" and overlook the allocation concealment process, which is perhaps the more important of the two.¹³ Allocation concealment should prevent anyone involved in the trial from having foreknowledge of a treatment assignment up to the point of allocation.¹⁴ Crucially, concealment prevents those who admit patients to a trial from knowing the upcoming assignments.

Breaches of allocation concealment probably happen more often than we suspect. Only about one quarter of trials report even the most minimal of allocation concealment procedures.^{14,16} Moreover, my colleagues and I have found that trials in which the allocation sequence had been inadequately concealed yielded larger estimates of treatment effects (odds ratios were exaggerated by 30% to 40% on average) compared with trials in which authors reported adequate allocation concealment.¹³ This result provides empiric evidence of bias in trials with inadequate allocation concealment. Furthermore, many residents and junior faculty will admit to deciphering, or witnessing someone else deciphering, an assignment scheme.²¹ Although most published RCTs probably provide reliable results, I believe allocation breaches to be more than a rare occurrence.

Klein and associates' description of allocation concealment¹ surpassed the concealment descriptions provided in most of the reports my colleagues and I have reviewed.^{14,16} Yet they used envelopes that arguably are, in general, susceptible to deciphering.²² "The envelopes were opaque, sequentially numbered, and contained instructions printed on opaque cards."¹ The opaque envelopes and opaque cards would have effectively prevented transillumination of the envelopes, a frequently cited problem.²¹ However, Klein and associates did not specify that the envelopes were sealed, and unsealed envelopes have led to deciphering in other trials.²¹

Even with adequate concealment, some deciphering of sequences can occur. Klein and associates' trial could not be blinded, and thus the assignments became known after allocation. Unblinded treatments, in particular, require an unpredictable assignment sequence. If a sequence is predictable (e.g., following an ABAB pattern or consisting of

short, fixed blocks), a sequence can be deduced from the order of the past assignments, and physicians can foresee all or some of the upcoming assignments.

Because Klein and associates did not report how they generated the assignment sequence we do not know whether it was unpredictable.¹ A random-number table or generator should have played a role. We may perhaps infer that some form of blocking was used, since "study envelopes were prepared separately by parity in each hospital."¹ If the blocks were of fixed size, and especially if they were small (six patients or fewer), the sequence may have been too predictable. In this case, the block size could have been deciphered and selection bias introduced, whatever the effectiveness of the allocation concealment.

Although these considerations may appear picayune, seemingly trivial issues have been implicated in the sabotaging of trials.²¹ The methodologic details of random assignment require diligent attention. I presume that the assignment envelopes in Klein and associates' trial were sealed. Given the other potential problems with allocation concealment and sequence generation, however, I do not know, and perhaps the investigators themselves do not know, whether their assignment sequence was partly deciphered. My guess is that any biases that entered their trial favoured episiotomy. If this is the case, it only strengthens their conclusions as to the lack of benefit of episiotomy.

NONCOMPLIANCE AND THE HUMAN SPIRIT

Klein and associates' problems with physician non-compliance and other examples¹³ of assignment manipulation in clinical trials are manifestations of a larger conflict: namely, the scientific community's need to obtain unbiased data from an inherently biased source — human beings. Unfortunately, RCTs are anathema to the human spirit. Whereas practitioners understand in theoretic terms the need for unbiased research, once they are engaged in a trial they may find it too difficult to maintain a dispassionate stance. They may "know" what treatment works better for patients and so want certain patients to benefit from that particular therapy; or they may want the results of a study to confirm what they already believe. Those who fail to comply with an assignment scheme do not necessarily have sinister motives: many subversions may reflect simple curiosity rather than scientific malevolence. Furthermore, practitioners involved in conducting a trial that does not have proper procedures for sequence generation and allocation concealment may find the challenge of deciphering the allocation scheme irresistible. Whatever the motivation, the effect is the same if the introduction of bias invalidates the comparisons made in the trial.

MINIMIZING THE EFFECTS OF PHYSICIAN NONCOMPLIANCE

How can more meaningful RCTs be conducted under conditions of potential physician noncompliance? Klein and associates suggest that random assignment of physicians rather than patients (presumably on the basis of their attitude toward episiotomy) might have been a viable solution. Such an approach has promise, although it would be subject to variability among physicians and would answer a slightly different question. Nevertheless, the additional variability would probably pale in comparison with the noncompliance that Klein and associates encountered. The results from such a trial would likely provide clearer and more meaningful results than those of the trial conducted.

Under the threat of physician noncompliance several actions could increase the likelihood of obtaining reliable results. First, investigators could conduct preliminary research on whether physician compliance might be a problem. The results of this preparatory work could prompt changes to the trial design, foster enhancements to trial implementation or scuttle the use of a randomized design. Second, investigators need to give participating physicians and other collaborators extensive training in the trial protocol. Although this is frequently done, investigators may not devote enough time and effort to protocol training. Third, investigators should educate those involved in implementation on the fundamentals of RCT methodology. People who implement trials sometimes undermine them without knowing it.²¹ Last, more effort should be devoted to monitoring compliance with the research protocol once allocation has occurred. Recalcitrant contributors could be instructed to adhere to the protocol or be excluded from further involvement (although their data would be used in the analysis up to that point, of course). This step would have minimized the impact of noncompliance on Klein and associates' results.

The type of monitoring I am espousing could be carried out when data come into a central trial office. None of these recommendations would entail expensive on-site monitoring. Rather, I advocate that, in general, trials be simplified in order to permit larger samples to be used. Nevertheless, some on-site auditing may be necessary.²³

With regard to random allocation, we must acknowledge the human factors that influence this important scientific process. Educating those who implement trials in the rationale for random allocation is an important first step, but a surprisingly difficult one. Importantly, the scientific community, particularly granting agencies and journals, should insist upon adequate sequence generation, adequate allocation concealment and the reporting

of both.^{11,14,16,21} No excuse should be accepted for a failure to meet these requirements.

If investigators use envelopes for allocation concealment, these should be sequentially numbered, opaque and sealed. Moreover, the investigators should ensure that the envelopes are opened sequentially, and only after they are inscribed with the participating patient's name and other details.²⁴ I also recommend using pressure-sensitive or carbon paper inside the envelope to copy the details onto the allocation record and thus create a valuable audit trail.

CONCLUSION

Are untoward, counterproductive human tendencies sufficient reason to abandon RCTs? Quite the contrary: these tendencies are what lie behind the need for random allocation. We need to expand our use of RCTs. Human inclinations simply make them a challenge to implement properly. To avert any distortion of results, researchers must erect methodologic barriers to prevent or deflect bias; this requires assiduous and excruciating attention to design, implementation and reporting.^{11-14,16}

References

1. Klein MC, Gauthier RC, Jorgensen SH et al: Does episiotomy prevent perineal trauma and pelvic floor relaxation? *Online J Curr Clin Trials* [serial on line] 1992; July 1 (doc 10)
2. Klein MC, Gauthier RJ, Robbins JM et al: Relationship of episiotomy to perineal trauma and morbidity, sexual dysfunction, and pelvic floor relaxation. *Am J Obstet Gynecol* 1994; 171: 591-598
3. Chalmers TC, Celano P, Sacks HS et al: Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983; 309: 1358-1361
4. Chalmers TC, Matta RJ, Smith H et al: Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *N Engl J Med* 1977; 297: 1091-1096
5. Greenland S: Can meta-analysis be salvaged? *Am J Epidemiol* 1994; 140: 783-787
6. Petitti DB: Of babies and bathwater. *Am J Epidemiol* 1994; 140: 779-782
7. Sacks H, Chalmers TC, Smith H: Randomized versus historical controls for clinical trials. *Am J Med* 1982; 72: 233-240
8. Miller JN, Colditz GA, Mosteller F: How study design affects outcomes in comparisons of therapy: II. Surgical. *Stat Med* 1989; 8: 455-466
9. Colditz GA, Miller JN, Mosteller F: How study design affects outcomes in comparisons of therapy: I. Medical. *Stat Med* 1989; 8: 441-454
10. Devine EC, Cook TD: A meta-analytic analysis of effects of psychoeducational interventions on length of post-surgical hospital stay. *Nurs Res* 1983; 32: 267-274
11. Standards of Reporting Trials Group: A proposal for structured reporting of randomized controlled trials. *JAMA* 1994; 272: 1926-1931
12. Altman DG: Randomization: essential for reducing bias. *BMJ* 1991; 302: 1481-1482
13. Schulz KF, Chalmers I, Hayes RJ et al: Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273: 408-412
14. Schulz KF, Chalmers I, Grimes DA et al: Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA* 1994; 272: 125-128
15. Chalmers TC, Levin H, Sacks HS et al: Meta-analysis of clinical trials as a scientific discipline. I: Control of bias and comparison with large co-operative trials. *Stat Med* 1987; 6: 315-325
16. Altman DG, Doré CJ: Randomization and baseline comparisons in clinical trials. *Lancet* 1990; 335: 149-153
17. Wortman PM, Yeaton WH: Synthesis of results in controlled trials of coronary artery bypass graft surgery. In Light RJ (ed): *Evaluation Studies Review Annual*, vol 8, Sage, Beverly Hills, Calif, 1983: 536-551
18. 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
19. DerSimonian R, Charette LJ, McPeck B et al: Reporting on methods in clinical trials. *N Engl J Med* 1982; 306: 1332-1337
20. Emerson JD, Burdick E, Hoaglin DC et al: An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Control Clin Trials* 1990; 11: 339-352
21. Schulz KF: Subverting randomization in controlled trials. *JAMA* (in press)
22. Pocock SJ: Statistical aspects of clinical trial design. *Statistician* 1982; 31: 1-18
23. Rennie D: Accountability, audit, and reverence for the publication process. *JAMA* 1993; 270: 495-496
24. Bulpitt CJ: *Randomized Controlled Clinical Trials*, Martinus Nijhoff, The Hague, The Netherlands, 1983