The Quest for Unbiased Research: Randomized Clinical Trials and the CONSORT Reporting Guidelines

Kenneth F. Schulz, PhD, MBA

Editor's Note

The large and increasing number of randomized clinical trials in all of medicine has prompted the introduction of guidelines that are intended to improve the quality of the research and of the published papers derived from such investigations. These guidelines for reporting clinical trials, the Consolidated Standards of Reporting Trials (the CONSORT statement) were set forth by a group of journal editors, statisticians, biomedical investigators, and editors. The CONSORT statement has been published in several major medical journals and is likely to have a major impact on editorial policies as well as the design of such investigations. I have invited Dr Kenneth F. Schulz of the Centers for Disease Control and Prevention, a leading investigator in the evaluation of clinical trials and a participant in the design of the CONSORT guidelines, to prepare this brief summary for the *Annals*. The CONSORT statement is expected to have a substantial effect on the peer review of clinical trials, and it promises to elevate the quality of such investigations. Future submissions to the *Annals* will be evaluated in the context of these guidelines.

Robert A. Fishman, MD
Editor

Assessing treatments can be misleading unless investigators ensure unbiased comparison groups. Random allocation to those groups remains the only method that eliminates selection and confounding biases. Thus, randomized controlled trials (RCTs) serve as the foundation for advancing medical science. Deservedly, medical researchers consider them the “gold standard.”

The design and implementation of an RCT, however, takes more knowledge and effort than investigators seem to fathom. They often encounter unanticipated, and sometimes unnoticed, difficulties. In particular, trial implementers sometimes subvert the intended aims of randomization [1-3]. While proper attention to design issues prevents or deflects attempts at subversion, investigators often fail to attain that requisite level of attention.

The poor reporting of RCTs reflects an inadequate attention to methodology. An article published in 1990 detailed the poor reporting of RCTs in the premier general medical journals in the world [4]. That article spawned another analysis that yielded similar results for specialty journals [5]. Regrettably, the most recent analysis indicates persistent poor reporting [6]. These and other recent reports [7, 8] buttressed earlier analyses that found inadequate reporting [9-11]. Justifiably, the medical community frets that shoddy reporting reflects shoddy methods, and that with shoddy methods come biased results.

With the aim of directly addressing shoddy reporting, and with the hope of also dealing with shoddy methods, a group recently published reporting guidelines for RCTs, the Consolidated Standards of Reporting Trials (CONSORT) statement [12]. In this special report, I discuss the important elements of randomization, review the recent research on bias and allocation methods, refer to some instances in which investigators have provided anonymous accounts of subverting randomization, and address recent findings on bias and double-blinding and exclusions after entry. I then introduce and discuss the CONSORT guidelines.

**Important Elements of Randomization**

The success of randomization depends on two interrelated processes [1, 13]. The first process entails generating a sequence by which the participants in a trial are allocated to treatment and control groups. To ensure unpredictability of that allocation sequence, investigators should generate it by a random process. The sec-
ond process involves meticulous implementation of that allocation sequence through an assignment mechanism (allocation concealment process) that prevents foreknowledge of the treatment assignments to those implementing the allocation sequence [5]. Many medical researchers mistakenly regard the process of generating an allocation sequence as "randomization" and overlook the process of allocation concealment [1, 2]. Without adequate allocation concealment, however, even random, unpredictable assignment sequences can be subverted [13, 14]. For example, suppose that an investigator creates an adequate allocation sequence using a random number table. However, the investigator then affixes the list of that sequence to a bulletin board, thereby essentially providing no allocation concealment. Those responsible for admitting participants could ascertain the upcoming treatment allocations and then route participants with better prognoses to the experimental group and those with poorer prognoses to the control group, or vice versa. Bias could easily be introduced. Crucially, allocation concealment shields those who admit participants to a trial from knowing the upcoming assignments [1, 2]. The decision to accept or reject a participant must be made and informed consent obtained without knowledge of the treatment to be assigned [15].

Investigators must not confuse blinding with allocation concealment. Substantial differences between the two exist. In particular, allocation concealment focuses on preventing selection and confounding biases, safeguards the assignment sequence before and until allocation, and can always be successfully implemented [1, 3]. In contrast, blinding focuses on preventing ascertainment bias, safeguards the sequence after allocation, and cannot always be implemented [1, 3].

Recent Research on Bias and Allocation Methods

As with all studies, bias jeopardizes even RCTs if investigators improperly execute them. For example, in a study of 250 controlled trials from 33 metaanalyses, we found that alleged RCTs with inadequate allocation concealment yielded larger estimates of treatment effects than did trials in which the authors reported adequate concealment [13]. Odds ratios were exaggerated, on average, by 30 to 40%. Furthermore, two other recent analyses in different subject matter areas also revealed exaggerated treatment effects related to inadequate allocation concealment [16, 17]. These exaggerated estimates of treatment effects likely reflect methodological problems. They also reveal meaningful levels of bias. If a medical researcher aspires to find a decrease in mortality of 35% from a particular treatment, certainly potential biases of 30 to 40% would overwhelm estimates of the treatment effect. Indeed, the elimination of bias becomes especially crucial in trials designed to detect moderate treatment effects.

Randomization and the Human Spirit

In the previous section, I referred to studies that discovered empirical evidence of bias. In practice, do investigators actually subvert the intended purpose of randomization? Have they divulged the details of deciphering allocation sequences? While that has happened [18], documented accounts are rare. Nevertheless, when physicians responded to queries during epidemiological workshops, more than half related instances of deciphering an allocation concealment scheme [1]. This should not be interpreted as representing more than half of all the trials [1]; indeed, most published RCTs probably estimate treatment effects reliably. Allocation breaches, however, seem to be something more frequent than a rare occurrence.

The personal accounts of those decipherings run the gamut from simple to intricate operations [1]. One frequently mentioned approach, the posting of an allocation sequence on a bulletin board, takes little effort to decipher. Other examples of simple deciphering operations include opening unsealed assignment envelopes and holding translucent envelopes up to a regular lightbulb. Workshop participants also related more involved efforts. For example, a few took sequentially numbered, opaque, sealed envelopes to the "hot light" (an intense incandescent bulb) in the radiology department for deciphering of the assignment scheme. Another admitted to searching for code in the office files of the principal investigator. The stories frequently reflected ingenious efforts.

The stories also frequently reflected naive, dubious judgment [1–3]. Assignment manipulations by those in trials reveal larger underlying conflicts. The need for unbiased research struggles with the inherently biased sources of that information—human beings. While investigators understand the need for unbiased research in the abstract, they may have difficulty maintaining an impartial posture once they are engaged in a trial. They may want certain patients to benefit from one of the treatments that they believe to be better or may want the results of a study to show what they believe to be valid [2, 3]. Aspects of properly conducted RCTs, then annoy investigators, because trial procedures attempt to impede human inclinations. Unfortunately, RCTs are an anathema to the human spirit [1–3].

For those conducting a RCT with inadequate allocation concealment, the challenge of deciphering the allocation scheme may become too great a temptation to resist. Those who thwart assignment mechanisms do not necessarily have corrupt motives. Frequently, they are unaware of the scientific ramifications of their actions [1]. Whether their motives reflect innocent or
pernicious intents, however, their actions undermine the validity of a trial.

Given the human ingredients in RCTs, we must realize that trial implementers, if provided the opportunity, sometimes subvert randomization. Fortunately, trialists can usually prevent subversions with diligent attention to design and implementation issues [1-3]. Trial designers and implementers should adopt the methods and standards of proper randomization and understand the rationale behind them [1, 3].

Bias and Nonallocation Methodological Elements
Methodological elements in RCTs other than proper allocation also reduce bias. We found that trials that were not double-blinded yielded larger estimates of treatment effects compared with trials in which authors reported double-blinding (odds ratios exaggerated, on average, by 17%) [13]. While double-blinding appears to prevent bias, its effect appears weaker than allocation concealment. Another recent analysis also denoted the importance of double-blinding [17].

Absence of bias persists throughout a trial only if the analysis includes all properly randomized participants in the originally assigned groups. Thus, we had expected to find some evidence of bias in those trials that reported having excluded participants after randomization. However, those trials did not yield exaggerated estimates of treatment effects compared with trials in which the reports gave the impression of no exclusions [13]. Some of the trials with reports that gave the impression of no exclusions probably had exclusions [13].

Reporting and CONSORT
While investigators must execute RCTs properly to minimize bias, they also must communicate those efforts to the reader. Without proper reporting, readers cannot discern the trials with valid results from those with questionable results. The use of proper methodology should be transparent to editors, reviewers, and readers. They should not have to assume or guess the methods employed; they should be informed explicitly. Yet assessments of the reporting quality of published trials have consistently found major flaws [4-11].

Violations of randomization probably happen more frequently than suspected. Only 9% of reports in the specialist journals and 15% in the general journals reported both an adequate method of generating random sequences and an adequate method of allocation concealment [4, 5, 19]. Of the reports on studies considered double-blind, only 45% described similarity of the treatment and control regimens and only 26% provided information on the protection of the allocation schedule [20]. Investigators reported testing the efficacy of blinding in only 2 of 31 trials, and found substantial unblinding of assignments in both [20].

Perhaps the most dismal reporting stems from authors neglecting to address, or incompletely addressing, exclusions of participants after randomization. Two studies found that trials with no apparent reported exclusions may in fact have made exclusions during the trial but ignored them in the report [13, 20]. Moreover, two trials with documented exclusions have published reports indicating no exclusions [21]. Readers of RCT reports should warily regard the potentially misleading information provided. Published information on exclusions may frequently provide a misleading impression of trial quality [20].

To address the prevailing flaws in reporting, two groups independently met and separately published guidelines for the reporting of RCTs [22, 23]. A subsequent editorial suggested that the two groups met to generate unified guidelines [24]. This flurry of activity culminated in the CONSORT statement [12]. It includes a checklist of 21 items (Table) that pertain to a RCT report. Those items reflect the fundamental information necessary to evaluate accurately the internal and external validity of a trial. The CONSORT statement also includes a flow diagram (Fig) that depicts the progress of participants throughout a two-group parallel-design RCT, the type of trial most commonly reported. Modifications will need to be made in reports of trials with a larger number of groups or trials using a different design [12].

While this guideline avalanche may seem precipitous, it really is long overdue [3]. Instigated by the pioneering work of A. Bradford Hill, the British Medical Journal (BMJ) published the first randomized trial almost 50 years ago [25]. For all these years, calls for improved reporting published in medical and statistical journals have languished on library shelves.

The CONSORT statement, however, represents a change. A number of high quality and high profile journals are adopting the guidelines, including JAMA, Lancet, and BMJ. The journals encourage authors to follow the statement immediately, but from January 1, 1997, they will be required to do so. After almost 50 years, the age of adequate reporting of RCTs approaches.

With such a long legacy of poor reporting, can the research community hope that these guidelines will produce a positive impact? While surely many factors will affect events, at least three factors justify hope. First, the greater number of methodologically astute journal editors constitutes perhaps the most substantial change. The editors contributing to the CONSORT guidelines portray a symptom of that trend. They represent a large number of editors worldwide who share similar views. Second, empirical evidence indicates that many improperly implemented RCTs may indeed be
Table. Consolidation of Standards for Reporting Trials—CONSORT

<table>
<thead>
<tr>
<th>Heading</th>
<th>Subheading</th>
<th>Descriptor</th>
<th>Was It Reported?</th>
<th>On What Page No.?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Abstract</td>
<td>Identify the study as a randomized trial.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
<td>Use a structured format.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methods</td>
<td>Protocol</td>
<td>State prospectively defined hypothesis, clinical objectives, and planned subgroup or covariate analyses.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assignment</td>
<td>Describe Planned study population, together with inclusion/exclusion criteria.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Planned interventions and their timing.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Primary and secondary outcome measure(s) and the minimum important difference(s), and indicate how the target sample size was projected.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rationale and methods for statistical analyses, detailing main comparative analyses and whether they were completed on an intention-to-treat basis.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Masking (Blinding)</td>
<td>Prospectively defined stopping rules (if warranted).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td>Participant Flow</td>
<td>Identify this as a randomized trial.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>and Follow-up</td>
<td>State estimated effect of intervention on primary and secondary outcome measures, including a point estimate and measure of precision (confidence interval).</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>State results in absolute numbers when feasible (e.g., 10/20, not 50%).</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Present summary data and appropriate descriptive and inferential statistics in sufficient detail to permit alternative analyses and replication.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Describe prognostic variables by treatment group and any attempt to adjust for them.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Describe protocol deviations from the study as planned, together with the reasons.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comment</td>
<td></td>
<td>State specific interpretation of study findings, including sources of bias and imprecision (internal validity) and discussion of external validity, including appropriate quantitative measures when possible.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>State general interpretation of the data in light of the totality of the available evidence.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


biased. Journal editors and readers of scientific reports could understandably be less concerned with poor reporting if they believed that poor reporting did not indicate biased results. However, with evidence indicating an association between poor reporting and exaggerated estimates of treatment effects [13, 16, 17], they
will be less comfortable simply presuming insulation from bias. Third, journals adopting the CONSORT guidelines do not merely suggest cooperation. They mandate adherence to the guidelines as a precondition for publication.

References