

Achieving Balance in Clinical Trials

An unbalanced view from EU regulators

Marc Buyse and Damian McEntegart

The CPMP's position discouraging dynamic allocation techniques is unfair.



Illustration Paul A. Belci

The two principles underpinning the design, execution, and analysis of comparative clinical trials are minimization of bias and maximization of the precision in estimating treatment effect. Randomization, in conjunction with blinding where possible, is a fundamental tool used in meeting the first of these objectives.¹ Simple (unrestricted) randomization prevents any conscious or unconscious selection bias by allocating subjects to treatment groups entirely at random. In the long run it achieves balance of all known and unknown prognostic factors, and therefore the results of randomized trials are always “valid” in the sense that they cannot be biased in any systematic way, for or against, regarding the treatments under comparison.²

In a single trial, however, chance may lead to imbalances with respect to important prognostic factors, which may in turn cause “accidental” bias.³ Clinical trials with substantial imbalances come under criticism, even when these imbalances are due to chance alone^{4,5} or to randomization methods that poorly control for balance.^{6,7} Imbalances in baseline subject characteristics are often blamed when trials fail to show the expected results.⁸

For these reasons, trialists have sought

to use methods that, in contrast to randomization, ensure balance (to some extent) with respect to prespecified subject characteristics. This policy has been endorsed by the International Conference on Harmonization (ICH) guideline on statistical principles in clinical trials: “Stratification by important prognostic factors measured at baseline (e.g., severity of disease, age, sex, etc.) may sometimes be valuable in order to promote balanced allocation within strata; this has greater potential benefit in small trials.”⁹

The two methods that are most commonly used to balance for prognostic factors are randomly permuted blocks within strata and minimization.^{10,11} These methods are best explained through a simple example. Suppose a randomized trial is carried out to compare two treatment groups, “A” and “B,” and balance is sought with respect to the disease stage (early versus advanced disease) and the subjects’ gender. Suppose, further, that the trial has already entered some subjects and the next subject to be randomized is a female with early disease. Permuted blocks within strata can be implemented as follows (truncated binomial design²):

- count the number of female subjects with early disease already allocated to A



and B: let these numbers be NA and NB, respectively

- let IA and IB be the integer part of the division of NA and NB by $b/2$ where b is the block size (e.g., if $b = 4$ and $NA = 9$, $IA = 4$)
- if $IA > IB$, allocate B; if $IA < IB$, allocate A; otherwise, allocate A or B at random. Minimization can be implemented as follows (for example):

- count the number of subjects with early disease (regardless of gender) already allocated to A and to B: let these numbers be EA and EB, respectively
- count the number of female subjects (regardless of disease stage) already allocated to A and to B: let these numbers be FA and FB, respectively
- if $EA + FA > EB + FB$, allocate B; if $EA + FA < EB + FB$, allocate A; otherwise, allocate A or B at random.

As can be seen from this simple but realistic example, the method of randomly permuted blocks forces balance within strata formed by the cross-classification of factor levels, and as a result its balancing properties break down as the number of strata becomes large relative to the size of the trial. The method of minimization considers the individual factor levels separately from each other and can therefore cope with more factors than permuted blocks within strata, making it a more effective method of achieving balance across a large number of factors. There is no other conceptual difference between the two methods, even though randomly permuted blocks within strata are sometimes characterized as being a “static” method (because lists of permuted blocks can be generated ahead of time if desired) while minimization is called a “dynamic” method (because it depends on the order in which subjects with different characteristics enter the trial).

Both methods have been used extensively by academic institutions, cooperative groups, and the pharmaceutical industry.¹²⁻¹⁴ Yet, in a surprising departure from usually thoughtful and balanced positions, a recent guideline of the European Committee for Proprietary Medicinal Products (CPMP) claims that dynamic methods such as minimization “remain highly controversial” and are “strongly discouraged,” though they express no similar concern for static methods such as

randomly permuted blocks within strata.¹⁵

Unscientific viewpoint

In our view, the CPMP’s position is unfair, unfounded, and unwise. It is unfair, because it ignores recent methodological literature that encourages wider use of minimization and does not cite any references supporting its own views.^{16,17} At least some of the CPMP’s concern relates to the appropriate analysis following the use of dynamic allocation: Can the conventional asymptotic tests be used, or should “rerandomization” tests be used to reflect the order in which patients with different characteristics entered the trial?

Although there is some theoretical justification for the latter, by no stretch of the imagination can this have any meaningful impact on the results of a trial. The limited empirical work available in this respect shows the appropriateness of conventional analyses even in the presence of

best of our knowledge, no study using minimization has ever reported the use of a rerandomization test in the medical literature. An article in a leading medical journal described minimization as the platinum standard for clinical trials without drawing any adverse response.²¹

The CPMP’s position is unfounded because it endorses static balancing methods and rejects dynamic methods, when in fact there is no essential difference between the two, as illustrated in the previous example. The use of conventional analyses that ignore subject ordering does not have an underlying theoretical justification for either method. Both methods induce constraints on the randomization, and the constraints of permuted blocks are often far more stringent than those of minimization. For instance, when a block size of four is used to allocate two treatments, one-third of all treatment allocations are deterministic. Deter-

Clinical investigators who plan trials for European registration will now be faced with two conflicting regulatory guidances.

time trends, a situation in which conventional analyses would be questioned the most.¹⁸

Another concern with minimization is that this method could conceivably create imbalances with respect to unknown prognostic factors. A recent article has shown this presumption to be without merit, the balance on unknown covariates following dynamic randomization being at least as good, on average, as that obtained with simple randomization.¹⁹

Neither of these concerns was seen as an impediment to the use of minimization following two recent reviews in highly respected, peer-reviewed journals.^{11,16} The CONSORT (Consolidated Standards of Reporting Trials) statement, which has been endorsed by leading medical journals, states that “Minimization is an acceptable alternative to random assignment” without any reservations.²⁰ To the

ministic treatment allocations are undesirable, at least in open-label trials, because foreknowledge of these treatment allocations may cause selection bias.²² Some implementations of minimization may also lead to deterministic treatment allocations, but most do not because they employ a random element as suggested by the CONSORT statement²⁰ and by the ICH Guideline.⁹

Avoiding accidental bias

The CPMP’s position is unwise, because it favors use of randomization methods that expose trialists and the medical community to the risk of accidental bias, when this risk could have been limited through use of balancing methods that are especially valuable when the trials are small or cannot easily be repeated. If there were any controversy over the use of minimization, it would be expected of an indepen-

dent agency to weigh all scientific arguments, for and against minimization, before castigating the use of a method that has long been adopted in the clinical community.¹¹⁻¹⁴

Clinical investigators who plan trials for European registration will now be faced with two conflicting regulatory guidelines.^{9,15} If they consider avoidance of accidental bias a goal worth pursuing, they will find the CPMP's latest guideline to be seriously out of balance.

References

1. D.G. Altman, "Randomisation," *British Medical Journal*, 302, 1481-1482 (1991).
2. W.F. Rosenberger and J.M. Lachin, *Randomization in Clinical Trials: Theory and Practice* (New York: Wiley, 2002).
3. M. Buyse, "Centralized Treatment Allocation in Comparative Clinical Trials," *Applied Clinical Trials*, 32-37 (August 2000).
4. D.G. Altman, "Randomisation in the SYMPHONY Trial [Letter]," *Lancet*, 356 1521-1522 (2000).
5. The SYMPHONY investigators, "Comparison of Sibrafiban with Aspirin for Prevention of Cardiovascular Events After Acute Coronary Syndromes: a Randomized Trial," *Lancet*, 355, 337-345 (2000).
6. K. Wheatley and R. Gray, "ICON4/AGO-OVAR-2.2: Questions About Trial Design, Cost-Effectiveness, and Clinical Benefit [Letter]," *Lancet*, 362, 1333 (2003).
7. N. Colombo *et al.* on behalf of ICON4 collaborators, "ICON4/AGO-OVAR-2.2: Questions about Trial Design, Cost-Effectiveness, and Clinical Benefit [Letter]," *Lancet*, 362, 1334 (2003).
8. B. Leyland-Jones on behalf of the BEST Investigators and Study Group, "Breast Cancer Trial with Erythropoietin Terminated Unexpectedly," *Lancet Oncology*, 4 (8) 459-60 (2003).
9. International Conference on Harmonisation, E-9 Document, "Guidance on Statistical Principles for Clinical Trials," Federal Register 63 (179) 49583-49598 (1998).
10. S.J. Pocock and R. Simon, "Sequential Treatment Assignment With Balancing for Prognostic Factors in the Controlled Clinical Trial," *Biometrics*, 31, 103-115 (1975).
11. C.J. Weir and K.R. Lees, "Comparison of Stratification and Adaptive Methods for Treatment Allocation in an Acute Stroke Clinical Trial," *Statistics in Medicine*, 22 705-726 (2003).
12. EuroCODE, "A New Approach to Collaborative Research in Clinical Oncology," *European Journal of Cancer and Clinical Oncology*, 25, 1905-1906 (1989).
13. ISIS-3: "A Randomised Comparison of Streptokinase vs Tissue Plasminogen Activator vs Anistreplase and of Aspirin Plus Heparin vs Aspirin Alone Among 41299 Cases of Suspected Acute Myocardial Infarction," *Lancet*, 339, 753-770 (1992).
14. E.A. Wickham, "A Paper that Changed my Practice: Minimisation," *British Medical Journal*, 309, 796 (1994).
15. Committee for Proprietary Medicinal Products. Points to Consider on Adjustment for Baseline Covariates. CPMP/EWP/2863/99, 2003. Available from <http://www.emea.eu.int/pdfs/human/ewp/286399en.pdf>.
16. N.W. Scott *et al.*, "The Method of Minimisation for Allocation to Clinical Trials: A Review," *Controlled Clinical Trials*, 23, 662-674 (2002).
17. D.J. McEntegart, "The Pursuit of Balance Using Stratified and Dynamic Randomization Techniques: an Overview," *Drug Information Journal*, 37 (3) 293-308 (2003).
18. L.A. Kalish and C.B. Begg, "The Impact of Treatment Allocation Procedures on Nominal Significance Levels and Bias," *Controlled Clinical Trials*, 8, 121-135 (1987).
19. M. Aickin, "Randomization, Balance and the Validity and Efficiency of Design-Adaptive Allocation Methods," *Journal of Statistical Planning and Inference*, 94, 97-119 (2001).
20. D.G. Altman *et al.*, CONSORT Group (Consolidated Standards of Reporting Trials). "The Revised CONSORT Statement for Reporting Randomised Trials: Explanation and Elaboration," *Annals of Internal Medicine*, 134, 663-694 (2001).
21. T. Treasure and K. MacRae, "Minimisation: the Platinum Standard for Trials?," *British Medical Journal*, 317, 362-363 (1998).
22. B. Efron, "Forcing a Sequential Experiment to be Balanced," *Biometrika*, 58, 403-417 (1971).

Marc Buyse, ScD, is director, *International Drug Development Institute (IDDI)*, Brussels, Belgium, +32 2 6468918, fax: +32 2 6468662, email: marc.buyse@iddi.com. **Damian McEntegart**, * MSc FIS, is statistics and medication forecasting manager with *ClinPhone Group Ltd.*, Lady Bay House, Meadow Grove, Nottingham, U.K., +44 115 955 7333, fax +44 115 955 7555, email: dmcenteg@clinphone.com.

*To whom correspondence should be addressed.