Debunking the myth:
Random block size does not decrease selection bias

Wenle Zhao

Medical University of South Carolina, Charleston, SC, USA

Randomization Working Group, October 28, 2025

Contents

Magic Square
(Lo Shu, ~2000 BC)

16	2	3	13
5	11	10	8
9	7	6	12
4	14	15	1

- 1 to n^2 integers
- One number one cell
- Row sum = C
- Column sum = C
- Diagonal sum = C

Latin Square
(Euler, ~1782)

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

- N letters
- One letter one cell
- Each row has all letters
- Each column has all letters

Randomized Block Design
(Ronald A. Fisher, ~1935)

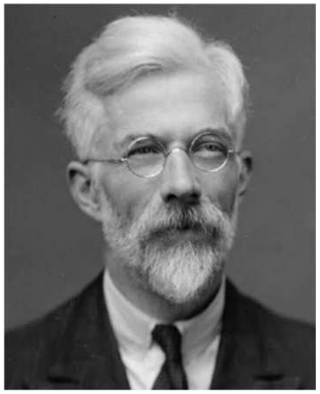
A	A	B	B
A	B	A	B
A	B	B	A
B	A	A	B
B	A	B	A
B	B	A	A

- M arms
- Each block is balanced
- Permutate sequences

Sudoku
(Howard Garns, 1979)

			9		2		8	7
8		1						
9			1	8		4		
4					7	8	9	1
		9			8		5	
	8	5			9			2
7				9	4	6		
	9						7	
	4	6		2				

- 9 rows 9 columns
- 9 3x3 sub squares
- Each row, column, and sub square contains each of the 9 integers once.



Fisher described a design for potatoes in 1931 at Rothamsted: **9 blocks of 9 plots each**, testing 3 levels of nitrogen × 3 levels of potash in the field.

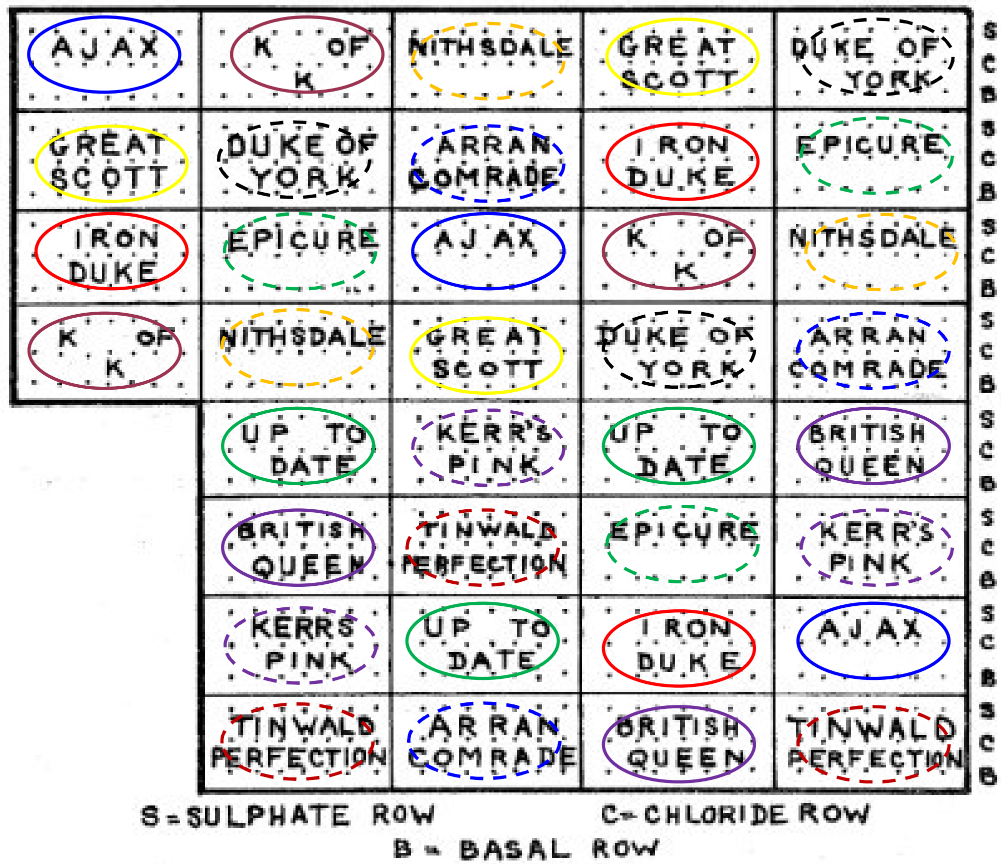
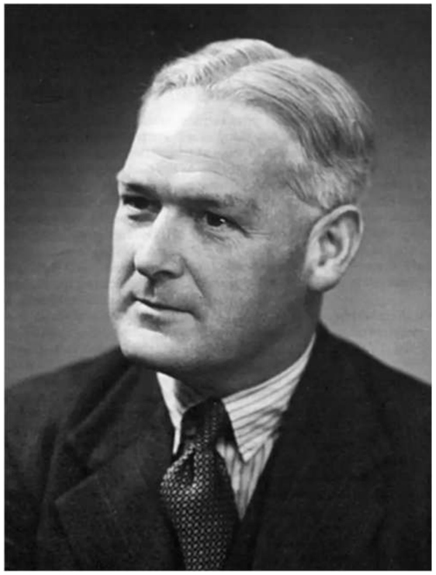


Diagram 1. Plan of experiment. Farmyard manure series.



The 1948 streptomycin trial for pulmonary tuberculosis, run by the UK Medical Research Council, is widely recognized as the first modern randomized controlled clinical trial (RCT).

They used a random numbers table prepared by statistician Austin Bradford Hill.

BRITISH MEDICAL JOURNAL

LONDON SATURDAY OCTOBER 30 1948

STREPTOMYCIN TREATMENT OF PULMONARY TUBERCULOSIS

A MEDICAL RESEARCH COUNCIL INVESTIGATION

The following gives the short-term results of a controlled investigation into the effects of streptomycin on one type of pulmonary tuberculosis. The inquiry was planned and directed by the Streptomycin in Tuberculosis Trials Committee, composed of the following members: Dr. Geoffrey Marshall (chairman), Professor J. W. S. Blacklock, Professor C. Cameron, Professor N. B. Capon, Dr. R. Cruickshank, Professor J. H. Gaddum, Dr. F. R. G. Heaf, Professor A. Bradford Hill, Dr. L. E. Houghton, Dr. J. Clifford Hoyle, Professor H. Raistrick, Dr. J. G. Scadding, Professor W. H. Tytler, Professor G. S. Wilson, and Dr. P. D'Arcy Hart (secretary). The centres at which the work was carried out and the specialists in charge of patients and pathological work were as follows:

Brompton Hospital, London.—Clinician: Dr. J. W. Crofton, Streptomycin Registrar (working under the direction of the honorary staff of Brompton Hospital); Pathologists: Dr. J. W. Clegg, Dr. D. A. Mitchison.

Colindale Hospital (L.C.C.), London.—Clinicians: Dr. J. V. Hurford, Dr. B. J. Douglas Smith, Dr. W. E. Snell; Pathologists (Central Public Health Laboratory): Dr. G. B. Forbes, Dr. H. D. Holt.

Harefield Hospital (M.C.C.), Harefield, Middlesex.—Clinicians: Dr. R. H. Brent, Dr. L. E. Houghton; Pathologist: Dr. E. Nassau.

Bangour Hospital, Bangour, West Lothian.—Clinician: Dr. I. D. Ross; Pathologist: Dr. Isabella Purdie.

Killingbeck Hospital and Sanatorium, Leeds.—Clinicians: Dr. W. Santon Gilmour, Dr. A. M. Reevie; Pathologist: Professor J. W. McLeod.

Northern Hospital (L.C.C.), Winchmore Hill, London.—Clinicians: Dr. F. A. Nash, Dr. R. Shoulman; Pathologists: Dr. J. M. Alston, Dr. A. Mohun.

Sully Hospital, Sully, Glam.—Clinicians: Dr. D. M. E. Thomas, Dr. L. R. West; Pathologist: Professor W. H. Tytler.

The clinicians of the centres met periodically as a working subcommittee under the chairmanship of Dr. Geoffrey Marshall; so also did the pathologists under the chairmanship of Dr. R. Cruickshank. Dr. Marc Daniels, of the Council's scientific staff, was responsible for the clinical co-ordination of the trials, and he also prepared the report for the Committee, with assistance from Dr. D. A. Mitchison on the analysis of laboratory results. For the purpose of final analysis the radiological findings were assessed by a panel composed of Dr. L. G. Blair, Dr. Peter Kerley, and Dr. Geoffrey S. Todd.

Introduction

When a special committee of the Medical Research Council undertook in September, 1946, to plan clinical trials of streptomycin in tuberculosis the main problem faced was that of investigating the effect of the drug in pulmonary tuberculosis. This antibiotic had been discovered two years previously by Waksman (Schatz, Bugie, and Waksman, 1944); in the intervening period its power of inhibiting tubercle bacilli *in vitro*, and the results of treatment in experimental tuberculous infection in guinea-pigs, had been reported; these results were strikingly better than those with any previous chemotherapeutic agent in tuberculosis. Preliminary results of trials in clinical tuberculosis had been published (Hinshaw and Feldman, 1945; Hinshaw, Feldman, and Pfuetze, 1946; Keefer *et al.*, 1946); the clinical results in pulmonary tuberculosis were encouraging but inconclusive.

The natural course of pulmonary tuberculosis is in fact so variable and unpredictable that evidence of improvement or cure following the use of a new drug in a few cases cannot be accepted as proof of the effect of that drug. The history of chemotherapeutic trials in tuberculosis is filled with errors due to empirical evaluation of drugs (Hart, 1946); the exaggerated claims made for gold treatment, persisting over 15 years, provide a spectacular example. It had become obvious that, in future, conclusions regarding the clinical effect of a new chemotherapeutic agent in tuberculosis could be considered valid only if based on adequately controlled clinical trials (Hinshaw an Feldman, 1944). The one controlled trial of gold treatmen (and the only report of an adequately controlled trial i tuberculosis we have been able to find in the literature reported negative therapeutic results (Amberson, McMahon and Pinner, 1931). In 1946 no controlled trial of streptomyci in pulmonary tuberculosis had been undertaken in th U.S.A. The Committee of the Medical Research Counci decided then that a part of the small supply of streptomyci allocated to it for research purposes would be best employe in a rigorously planned investigation with concurren controls.

The many difficulties of planning and conducting a tria of this nature are important enough to warrant a ful description here of the methods of the investigation.

Plan and Conduct of the Trial

Type of Case

A first prerequisite was that all patients in the trial shoul have a similar type of disease. To avoid having to mak allowances for the effect of forms of therapy other tha bed-rest, the type of disease was to be one not suitable fo other forms of therapy. The estimated chances of spon taneous regression must be small. On the other hand, th type of lesion should be such as to offer some prospect o action by an effective chemotherapeutic agent; for thi reason old-standing disease, and disease with thick-walle

cavities, should be excluded. Finally the age group must be reasonably limited, since the total number of patients in the trial could not be large.

Such closely defined features were considered indispensable, for it was realized that no two patients have an identical form of the disease, and it was desired to eliminate as many of the obvious variations as possible. For these several reasons the type of case to be investigated was defined as follows: acute progressive bilateral pulmonary tuberculosis of presumably recent origin, bacteriologically proved, unsuitable for collapse therapy, age group 15 to 25 (later extended to 30).

The selection of this type of disease constituted full justification for having a parallel series of patients treated only by bed-rest, since up to the present this would be considered the only suitable form of treatment for such cases. Additional justification lay in the fact that all the streptomycin available in the country was in any case being used, the rest of the supply being taken up for two rapidly fatal forms of the disease, miliary and meningeal tuberculosis.

Recruitment and Admission of Cases

Co-operation in the trial was obtained in the first place from Brompton Hospital (drawing on London County Council cases), Colindale Hospital (London County Council), and Harefield County Hospital (Middlesex County Council). The L.C.C. and the M.C.C. gave full co-operation, permitting recruitment of suitable cases from the areas served by them, covering a population of nearly six million persons. Accordingly letters were sent, through the tuberculosis departments of these authorities, to tuberculosis officers and to medical superintendents of general hospitals outlining the proposed trial and asking that particulars and *x*-ray films of possibly suitable patients be sent to the co-ordinator of the trials for consideration. Visits were paid to the tuberculosis clinics and hospitals to show by representative *x*-ray films the type of case sought and to explain in detail the nature of the controlled trial. When cases were submitted the clinical particulars and *x*-ray films were taken to the Committee's selection panel for consideration. When a patient had been accepted as suitable, request was made through the local authority for admission to one of the streptomycin centres; in spite of long waiting-lists these patients were given complete priority, and the majority were admitted within a week of approval.

The first patients to be accepted were admitted to the centres in January, 1947. At first the impression was that cases of the type defined are seen often. In fact, such cases are not common. As it became evident after three months that enough cases could not be found in the London and Middlesex areas, other authorities were approached. The Welsh National Memorial Association, the Department of Health for Scotland, and the Leeds Tuberculosis Service made available centres at Sully, Bangour, and Killingbeck, and cases were recruited to those centres from the respective areas. In addition, another centre was opened in the London area, at the Northern Hospital (L.C.C.).

By September, 1947, 109 patients had been accepted, and no more were admitted to this trial. Two patients had died within the preliminary observation week; these are excluded from the analysis. Of the remaining 107 patients 55 had been allocated to the streptomycin group and 52 to the control group.

The Control Scheme

Determination of whether a patient would be treated by streptomycin and bed-rest (S case) or by bed-rest alone (C case) was made by reference to a statistical series based on random sampling numbers drawn up for each sex at each centre by Professor Bradford Hill; the details of the series were unknown to any of the investigators or to the co-ordinator and were contained in a set of sealed envelopes, each bearing on the outside only the name of the hospital and a number. After acceptance of a patient by the panel, and before admission to the streptomycin centre, the appropriate numbered envelope was opened at the central office; the card inside told if the patient was to be an S or a C case, and this information was then given to the medical officer of the centre. Patients were not told before admission that they were to get special treatment. C patients did not know throughout their stay in hospital that they were control patients in a special study; they were in fact treated as they would have been in the past, the sole difference being that they had been admitted to the centre more rapidly than was normal. Usually they were not in the same wards as S patients, but the same regime was maintained

It was important for the success of the trial that the details of the control scheme should remain confidential. It is a matter of great credit to the many doctors concerned that this information was not made public throughout the 15 months of the trial, and the Committee is much indebted to them for their co-operation.

By definition, cases accepted for the trial were unsuitable for collapse therapy; clinicians were therefore asked to adopt collapse therapy only if the course of the disease so changed that some collapse measure became indispensable and urgent. In the S cases collapse therapy was in fact never applied during the four treatment months. It was given to five of the 52 C cases during that period.

Observation and Treatment Period

Each patient was to remain in bed at the centre for at least six months, and the results were to be assessed on the clinical status at the end of that period. In addition to the usual hospital records, clinical observations were entered on standard record forms designed particularly for this trial; these forms provided for details of history, criteria of acceptance, examination on admission, monthly routine re-examinations with assessment of progress since last examination, observation of toxic reactions, temperature and treatment records, and finally a pathological record form. Instructions on required frequency of examinations were given.

Clinicians and pathologists' meetings were held during the trials to discuss the work as it proceeded. The co-ordinator visited centres and was constantly in touch with the clinicians concerned to discuss the progress of the trial and the problems arising. The working subcommittee of pathologists established the technical laboratory procedures, discussed the findings at intervals, and arranged for independent checking of sensitivity tests of tubercle bacilli and streptomycin levels in the blood.

Analysis of Results

The general trend of results during the course of the trial was followed through the monthly reports from the centres. The analysis of results up to six months after the patient's admission is presented here; it is based on information from the standard record forms completed for each patient and on the *x*-ray films which have been made available by the hospitals concerned.

The films have been viewed by two radiologists and a clinician, each reading the films independently and not knowing if the films were of C or S cases. One of the radiologists had been attached to a centre taking part in the trial; the other two specialists had not been connected with the trial in any way. There was fair agreement among the three; at a final session they met to review and discuss

The Control Scheme

Determination of whether a patient would be treated by streptomycin and bed-rest (S case) or by bed-rest alone (C case) was made by reference to a statistical series based on random sampling numbers drawn up for each sex at each centre by Professor Bradford Hill; the details of the series were unknown to any of the investigators or to the co-ordinator and were contained in a set of sealed envelopes, each bearing on the outside only the name of the hospital and a number. After acceptance of a patient by the panel, and before admission to the streptomycin centre, the appropriate numbered envelope was opened at the central office; the card inside told if the patient was to be an S or a C case, and this information was then given to the medical officer of the centre. Patients were not told before admission that they were to get special treatment. C patients did not know throughout their stay in hospital that they were control patients in a special study; they were in fact treated as they would have been in the past, the sole difference being that they had been admitted to the centre more rapidly than was normal. Usually they were not in the same wards as S patients, but the same regime was maintained

- Randomization sequences are created for each sex-center stratum.
- 7 hospitals, 2 sex category, a total 14 strata.
- Final distribution: 55 in treatment arm and 52 in control arm.
- No detail on how the random numbers were drawn.
- Under simple randomization, the chance to have this imbalance is 0.3.

THE CLINICAL TRIAL

A. BRADFORD HILL Ph.D. D.Sc.

Professor of Medical Statistics
London School of Hygiene and Tropical Medicine

Honorary Director, Statistical Research Unit
Medical Research Council

" Therapeutics ", said Professor Pickering in his Presidential Address to the Section of Experimental Medicine and Therapeutics of the Royal Society of Medicine, " is the branch of medicine that, by its very nature, should be experimental. For if we take a patient afflicted with a malady, and we alter his conditions of life, either by dieting him, or by putting him to bed, or by administering to him a drug, or by performing on him an operation, we are performing an experiment. And if we are scientifically minded we should record the results. Before concluding that the change for better or for worse in the patient is due to the specific treatment employed, we must ascertain whether the result can be repeated a significant number of times in similar patients, whether the result was merely due to the natural history of the disease or in other words to the lapse of time, or whether it was due to some other factor which was necessarily associated with the therapeutic measure in question. And if, as a result of these procedures, we learn that the therapeutic measure employed produces a significant, though not very pronounced, improvement, we would experiment with the method, altering dosage or other detail to see if it can be improved. This would seem the procedure to be expected of men with six years of scientific training behind them. But it has not been followed. Had it been done we should have gained a fairly precise knowledge of the place of individual methods of therapy in disease, and our efficiency as doctors would have been enormously enhanced " (Pickering, 1949).

It would be difficult to put the case for the clinical trial more cogently or more clearly. It is the gradual development of this attitude of mind coupled with the concurrent introduction of one antibiotic, one modern drug, after another, that has led in the past few years to the highly organized and efficiently controlled therapeutic trial of new remedies. For instance, as Marc Daniels (1950) wrote of one field, " it is now becoming generally accepted that scientific appraisement of new drugs in tuberculosis is a fundamental necessity, but it is hard to realize that this particular progress has been made almost entirely within the past five years ". Its absence in the past led, he suggests, to the many years of inconclusive work on gold therapy, while Pickering stresses the much earlier and positively dangerous methods of therapeutics, such as blood-letting, purging and starvation, of which the dangers could not have failed to be exposed by comparative observations, impartially made. In more recent years much work has, of course, been done on the efficacy of methods of treatment, e.g. the use of artificial pneumothorax in pulmonary tuberculosis. But many of these studies, as was pointed out in the review of fifty years of medicine published by the *British Medical Journal* (1950), suffer from the handicap that no comparative observations were made, sometimes for ethical reasons, but too often because their importance was not appreciated.

1. Imperfect Contrasts

As a result of this situation very many second-best, or even much inferior, " controls " have been put forward. Thus the following ways and means have been used from time to time, and are still used:

i. The treatment of patients with a particular disease is unplanned but naturally varies according to the decision of the physicians in charge. To some patients a specific drug is given, to others it is not. The progress and prognosis of these patients are then compared. But in making this comparison in relation to the treatment the fundamental assumption is made—and must be made—that the two groups are equivalent in all respects relevant to their progress, except for the difference in treatment. It is, however, almost invariably impossible to believe that this is so. Drugs are not ordered by doctors at random, but in relation to a patient's condition when he first comes under observation and also to the subsequent progress of his disease. The two groups are therefore not remotely comparable and more often than not the group given the specific drug is heavily weighted by the more severely ill. No conclusion as to its efficacy can possibly be drawn.

ii. The same objections must be made to the contrasting of volunteers for a treatment with those who do not volunteer, or between those who accept and those who refuse. There can be no knowledge that such groups are comparable; and the onus lies wholly, it may justly be maintained, upon the experimenter to prove that they are comparable, before his results can be accepted. Particularly, perhaps, with a surgical operation the patients who accept may be very different from those who refuse.

iii. The contrast of one physician, or one hospital, using a particular form of treatment, with another physician, or hospital, not adopting that treatment, or adopting it to a lesser degree, is fraught with much the same difficulty—apart from the practicability of being able to find such a situation (with, it must be noted, the same forms of ancillary treatment). It must be proved that the patients are alike in relevant group characteristics, i.e., age, sex, social class, severity of illness, before they can be fairly compared and their relative progress, or lack of progress, interpreted. That proof is clearly hard to come by.

iv. The ' historical ' control relies upon a contrast of past records of the pre-drug days with those of the present treated patients. Of the former group 10 per cent, say, died and 90 per cent recovered while for the present group the ratios are 5 and 95 per cent. If everything else remained constant

- Formally introduced stratified permuted block randomization to clinical trials by Bradford Hill in 1951.

Guidance for Industry

E9 Statistical Principles for Clinical Trials

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
September 1998
ICH

for

incipl

rials

U.S. Department of Health and
Food and Drug Admini
Center for Drug Evaluation and I
Center for Biologics Evaluation an
September 1998
ICH

randomization, to administer the assignment of randomized treatment. In addition, clinical assessments should be made by medical staff who are not involved in treating the subjects and who remain blind to treatment. In single-blind or open-label trials every effort should be made to minimize the various known sources of bias and primary variables should be as objective as possible. The reasons for the degree of blinding adopted, as well as steps taken to minimize bias by other means, should be explained in the protocol. For example, the sponsor should have adequate standard operating procedures to ensure that access to the treatment code is appropriately restricted during the process of cleaning the database prior to its release for analysis.

Breaking the blind (for a single subject) should be considered only when knowledge of the treatment assignment is deemed essential by the subject's physician for the subject's care. Any intentional or unintentional breaking of the blind should be reported and explained at the end of the trial, irrespective of the reason for its occurrence. The procedure and timing for revealing the treatment assignments should be documented.

In this document, the blind review (see Glossary) of data refers to the checking of data during the period of time between trial completion (the last observation on the last subject) and the breaking of the blind.

2. *Randomization (2.3.2)*

Randomization introduces a deliberate element of chance into the assignment of treatments to subjects in a clinical trial. During subsequent analysis of the trial data, it provides a sound statistical basis for the quantitative evaluation of the evidence relating to treatment effects. It also tends to produce treatment groups in which the distributions of prognostic factors, known and unknown, are similar. In combination with blinding, randomization helps to avoid possible bias in the selection and allocation of subjects arising from the predictability of treatment assignments.

The randomization schedule of a clinical trial documents the random allocation of treatments to subjects. In the simplest situation it is a sequential list of treatments (or treatment sequences in a crossover trial) or corresponding codes by subject number. The logistics of some trials, such as those with a screening phase, may make matters more complicated, but the unique preplanned assignment of treatment, or treatment sequence, to subject should be clear. Different trial designs will necessitate different procedures for generating randomization schedules. The randomization schedule should be reproducible (if the need arises).

Although unrestricted randomization is an acceptable approach, some advantages can generally be gained by randomizing subjects in blocks. This helps to increase the comparability of the treatment groups, particularly when subject characteristics may change over time, as a result, for example, of changes in recruitment policy. It also provides a better guarantee that the treatment groups will be of nearly equal

size. In crossover trials, it provides the means of obtaining balanced designs with their greater efficiency and easier interpretation. Care should be taken to choose block lengths that are sufficiently short to limit possible imbalance, but that are long enough to avoid predictability towards the end of the sequence in a block. Investigators and other relevant staff should generally be blind to the block length; the use of two or more block lengths, randomly selected for each block, can achieve the same purpose. (Theoretically, in a double-blind trial predictability does not matter, but the pharmacological effects of drugs may provide the opportunity for intelligent guesswork.)

In multicenter trials (see Glossary), the randomization procedures should be organized centrally. It is advisable to have a separate random scheme for each center, i.e., to stratify by center or to allocate several whole blocks to each center. More generally, stratification by important prognostic factors measured at baseline (e.g., severity of disease, age, sex) may sometimes be valuable in order to promote balanced allocation within strata; this has greater potential benefit in small trials. The use of more than two or three stratification factors is rarely necessary, is less successful at achieving balance, and is logistically troublesome. The use of a dynamic allocation procedure (see below) may help to achieve balance across a number of stratification factors simultaneously, provided the rest of the trial procedures can be adjusted to accommodate an approach of this type. Factors on which randomization has been stratified should be accounted for later in the analysis.

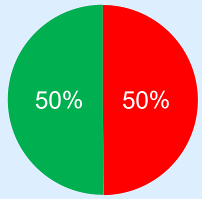
The next subject to be randomized into a trial should always receive the treatment corresponding to the next free number in the appropriate randomization schedule (in the respective stratum, if randomization is stratified). The appropriate number and associated treatment for the next subject should only be allocated when entry of that subject to the randomized part of the trial has been confirmed. Details of the randomization that facilitate predictability (e.g., block length) should not be contained in the trial protocol. The randomization schedule itself should be filed securely by the sponsor or an independent party in a manner that ensures that blindness is properly maintained throughout the trial. Access to the randomization schedule during the trial should take into account the possibility that, in an emergency, the blind may have to be broken for any subject. The procedure to be followed, the necessary documentation, and the subsequent treatment and assessment of the subject should all be described in the protocol.

Dynamic allocation is an alternative procedure in which the allocation of treatment to a subject is influenced by the current balance of allocated treatments and, in a stratified trial, by the stratum to which the subject belongs and the balance within that stratum. Deterministic dynamic allocation procedures should be avoided and an appropriate element of randomization should be incorporated for each treatment allocation. Every effort should be made to retain the double-blind status of the trial. For example, knowledge of the treatment code may be restricted to a central trial office from where the dynamic allocation is controlled, generally through telephone

Permuted Block and Selection Bias

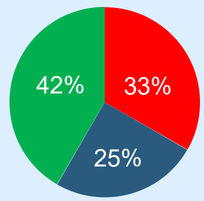
b = 2

AB   BA



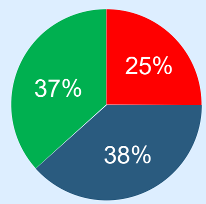
b = 4

AABB   ABAB   ABBA
BAAB   BABA   BBAA



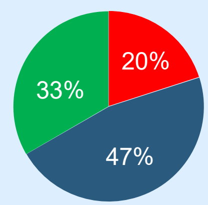
b = 6

AAABBB   AABABB   AABBAB   AABBBA
ABAABB   ABABAB   ABABBA   ABBBAA
BAAABB   BAABAB   BAABBA   BABBAA
BBAAAB   BBAABA   BBABAA   BBBAAA



b = 8

AAAABBBB   AAABABBB   AAABBABB   AAABBBAB
AAABBBBA   AABAABBB   AABABABB   AABABBAB
……
BBBAAABA   BBBAABAA   BBBABAAA   BBBBAAAA



Number of permutation blocks $=\binom{b}{b/2} = \frac{b!}{(b/2)!^2}$

Proportion of deterministic assignments $= \frac{1}{1+(b/2)}$

Completely random

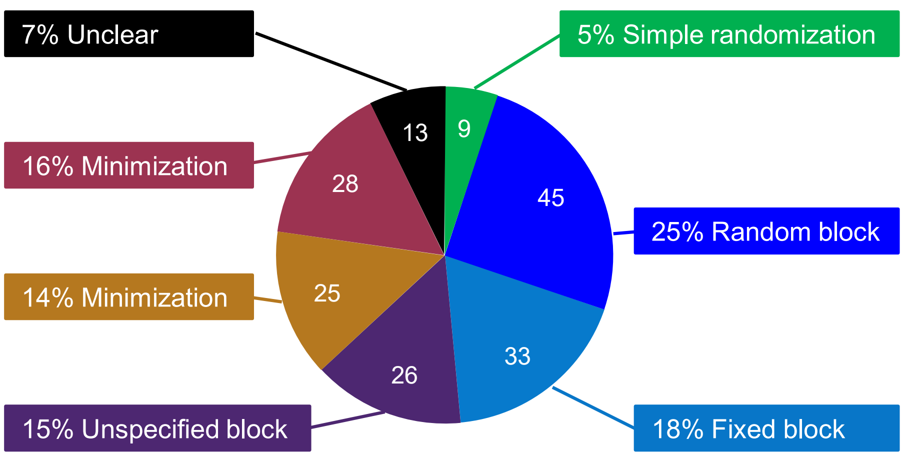
Biased coin

Deterministic

Random Block Design - Widely Used

ICH E9 (1998 guideline)

size. In crossover trials, it provides the means of obtaining balanced designs with their greater efficiency and easier interpretation. Care should be taken to choose block lengths that are sufficiently short to limit possible imbalance, but that are long enough to avoid predictability towards the end of the sequence in a block. Investigators and other relevant staff should generally be blind to the block length; the use of two or more block lengths, randomly selected for each block, can achieve the same purpose. (Theoretically, in a double-blind trial predictability does not matter, but the pharmacological effects of drugs may provide the opportunity for intelligent guesswork.)



ELSEVIER
Journal of Clinical Epidemiology 141 (2022) 90–98
Journal of Clinical Epidemiology

REVIEW

A review found small variable blocking schemes may not protect against selection bias in randomized controlled trials

Laura Clark*, Lauren Burke, Rachel Margaret Carr, Elizabeth Coleman, Gareth Roberts, David J. Torgerson

York Trials Unit, Department of Health Sciences, University of York, York, YO10 5DD, United Kingdom

Accepted 7 September 2021; Available online 11 September 2021

Abstract

Objective: Blocking is associated with prediction of the allocation sequence and subversion. This paper explores if blocking strategies lead to an increase in baseline age heterogeneity (a marker for potential subversion) and, whether the use of blocking is changing over time.

Study Design and Settings: The British Medical Journal, Journal of the American Medical Association, The Lancet and the New England Journal of Medicine were hand searched to identify open RCTs published in January between 2001 and 2020. To explore heterogeneity of baseline age meta-analyses were performed on trials implementing blocking, minimization, and simple randomization.

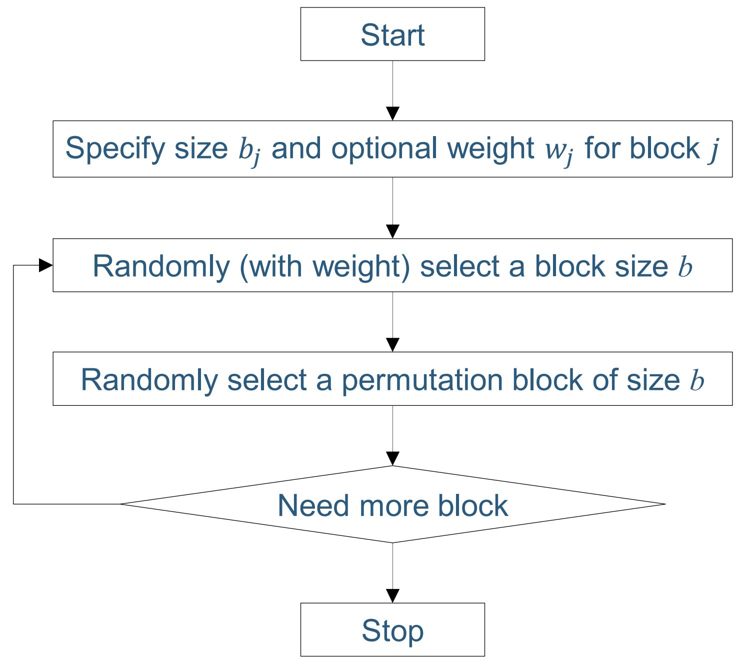
Results: One hundred seventy-nine open RCTs were identified: nine (5.0%) undertook simple randomization, 104 (58.1%) blocking, 25 (13.9%) minimization, and one (0.6%) both. Baseline age heterogeneity of 24% (P= 0.02) was observed in all trials implementing blocking, 62% (P = 0.001) in trials implementing a fixed block of four, 40% (P = 0.07) implementing variable blocks including a 2 and 0% for both simple randomization and minimization. Small block sizes are implemented in modern trials.

Conclusion: Variable block sizes including two are associated with subversion and should not be implemented. If center only stratification is necessary, it should be used alongside larger blocking schemes. Authors should consider alternative methods to restrict randomization.

Random Block Design - Definition

ICH E9 (1998 guideline)

size. In crossover trials, it provides the means of obtaining balanced designs with their greater efficiency and easier interpretation. Care should be taken to choose block lengths that are sufficiently short to limit possible imbalance, but that are long enough to avoid predictability towards the end of the sequence in a block. Investigators and other relevant staff should generally be blind to the block length; the use of two or more block lengths, randomly selected for each block, can achieve the same purpose. (Theoretically, in a double-blind trial predictability does not matter, but the pharmacological effects of drugs may provide the opportunity for intelligent guesswork.)



Example:

$$b_1 = 2,\ b_2 = 4,\ b_3 = 6.$$

$$w_1 = w_2 = w_3 = \frac{1}{3}$$

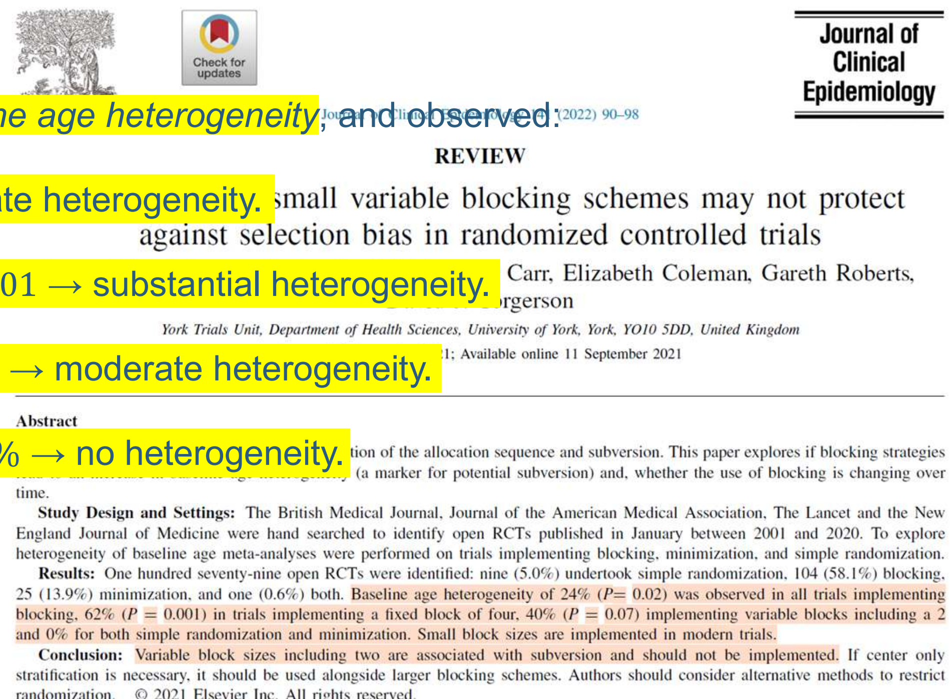
Random Block Design Example

Prediction Method	Prediction	Prediction Rate (PR)	Correct	Correct Guess Probability (CG)	Selection Bias Risk Score (SBRS)
Assume b=2	18	50%	8	44%	-0.0056
Assume b=4	14	39%	9	64%	0.1111
Assume b=6	12	33%	10	83%	0.2222
Assume b=8	4	11%	4	100%	0.1111
Completely Random		100%		50%	0

$$SBRS = PR \cdot (2 \times CG - 1)$$

Block size	No.	Tx.	b=2	b=4	b=6	b=8
b=6	1	A	-	-	-	-
	2	B	B	-	-	-
	3	B	-	-	-	-
	4	B	A	A	-	-
	5	A	-	-	A	-
	6	A	B	-	A	-
b=2	7	A	-	B	-	-
	8	B	B	B	-	B
b=2	9	B	-	-	-	-
	10	A	A	-	-	-
b=4	11	A	-	-	-	-
	12	A	B	B	B	-
	13	B	-	-	-	-
	14	B	A	-	-	-
b=4	15	B	-	A	-	-
	16	A	A	A	A	A
	17	B	-	-	A	-
	18	A	A	-	A	-
b=6	19	A	-	-	-	-
	20	B	B	B	-	-
	21	A	-	-	-	-
	22	A	B	-	-	-
	23	B	-	B	B	-
	24	B	A	B	B	B
b=2	25	A	-	-	-	-
	26	B	B	-	-	-
b=6	27	B	-	-	-	-
	28	B	A	A	-	-
	29	A	-	-	A	-
	30	A	B	-	A	-
	31	A	-	B	-	-
	32	B	B	B	-	B
b=4	33	B	-	-	-	-
	34	B	A	-	-	-
	35	A	-	A	A	-
	36	A	B	A	A	-

Random Block Design - Selection Bias



Journal of Clinical Epidemiology

Journal of Clinical Epidemiology 141 (2022) 90–98

REVIEW

small variable blocking schemes may not protect against selection bias in randomized controlled trials

Carr, Elizabeth Coleman, Gareth Roberts, ...rgerson

York Trials Unit, Department of Health Sciences, University of York, York, YO10 5DD, United Kingdom

...1; Available online 11 September 2021

Abstract

...tion of the allocation sequence and subversion. This paper explores if blocking strategies ... (a marker for potential subversion) and, whether the use of blocking is changing over time.

Study Design and Settings: The British Medical Journal, Journal of the American Medical Association, The Lancet and the New England Journal of Medicine were hand searched to identify open RCTs published in January between 2001 and 2020. To explore heterogeneity of baseline age meta-analyses were performed on trials implementing blocking, minimization, and simple randomization.

Results: One hundred seventy-nine open RCTs were identified: nine (5.0%) undertook simple randomization, 104 (58.1%) blocking, 25 (13.9%) minimization, and one (0.6%) both. Baseline age heterogeneity of 24% ($P = 0.02$) was observed in all trials implementing blocking, 62% ($P = 0.001$) in trials implementing a fixed block of four, 40% ($P = 0.07$) implementing variable blocks including a 2 and 0% for both simple randomization and minimization. Small block sizes are implemented in modern trials.

Conclusion: Variable block sizes including two are associated with subversion and should not be implemented. If center only stratification is necessary, it should be used alongside larger blocking schemes. Authors should consider alternative methods to restrict randomization.

Clark et al. performed meta-analysis for the imbalance in *baseline age heterogeneity*, and observed:

- All trials with blocking: $I^2 = 24\%,\ p = 0.02$ → moderate heterogeneity.
- All trials with fixed block size b = 4: $I^2 = 62\%,\ p = 0.001$ → substantial heterogeneity.
- All trials used varying block sizes: $I^2 = 40\%,\ p = 0.07$ → moderate heterogeneity.
- All trials simple randomization or minimization: $I^2 = 0\%$ → no heterogeneity.

Interpretation:

1. Poor baseline age balance possibly due to allocation predictability in blocking randomization.
2. Prediction without certainty may occur in trials with random block design.

Convergent Prediction

……, the *convergent* prediction, which predicts that treatment which has hitherto occurred less often, ……

David Blackwell, J. L. Hodges Jr. Design for the Control of Selection Bias, *Ann. Math. Statist*. 28(2): 449-460 (June 1957).

For equal allocations, always predict the next assignment be the smallest arm.

DESIGN FOR THE CONTROL OF SELECTION BIAS[1]

By David Blackwell and J. L. Hodges, Jr.

University of California, Berkeley

0. Summary. Suppose an experimenter E wishes to compare the effectiveness of two treatments, A and B, on a somewhat vaguely defined population. As individuals arrive, E decides whether they are in the population, and if he decides that they are, he administers A or B and notes the result, until nA's and nB's have been administered. Plainly, if E is aware, before deciding whether an individual is in the population, which treatment is to be administered next, he may, not necessarily deliberately, introduce a bias into the experiment. This bias we call *selection bias*. We propose to investigate the extent to which a statistician S, by determining the order in which treatments are administered, and not revealing to E which treatment comes next until after the individual who is to receive it has been selected, can control this selection bias.

Thus a design d is a distribution over the set T of the $\binom{2n}{n}$ sequences of length $2n$ containing nA's and nB's. We shall measure the bias of a design by the maximum expected number of correct guesses which an experimenter can achieve, knowing d, attempting to guess the successive elements of a sequence $t \varepsilon T$ selected by d, and being told after each guess whether or not it is correct. The distribution of the number G of correct guesses depends both on d and on the prediction method p used by the experimenter. We shall consider particularly two designs, the *truncated binomial*, in which the successive treatments are selected independently with probability $\frac{1}{2}$ each until n treatments of one kind have occurred, and the *sampling* design, in which all $\binom{2n}{n}$ sequences are equally likely. We shall consider particularly two prediction methods, the *convergent* prediction, which predicts that treatment which has hitherto occurred less often, and the *divergent* prediction, which predicts that treatment which has hitherto occurred more often, except that after n treatments of one kind have been administered, the divergent prediction agrees with the convergent predictions that the other treatment will follow; when both treatments have occurred equally often, either method predicts A or B by tossing a fair coin, independently for each case of equality.

We find that among all designs, the truncated binomial minimizes the maximum expected number of correct guesses. For this design, the expected number of correct guesses is independent of the prediction method, and is

$$n + n\binom{2n}{n}\Big/2^{2n} \sim n + (n/\pi)^{1/2}$$

Received March 26, 1956; revised November 2, 1956.

[1] This investigation was supported (in part) by research grant from The National Institutes of Health, Public Health Service.

Convergent Prediction

Block size	No.	Tx.	b=2	b=4	b=6	b=8	n_B	n_A	Convergent
b=6	1	A	-	-	-	-	0	0	-
	2	B	B	-	-	-	1	0	B
	3	B	-	-	-	-	1	1	-
	4	B	A	A	-	-	1	2	A
	5	A	-	-	A	-	1	3	A
	6	A	B	-	A	-	2	3	A
b=2	7	A	-	B	-	-	3	3	-
	8	B	B	B	-	B	4	3	B
b=2	9	B	-	-	-	-	4	4	-
	10	A	A	-	-	-	4	5	A
b=4	11	A	-	-	-	-	5	5	-
	12	A	B	B	B	-	6	5	B
	13	B	-	-	-	-	7	5	B
	14	B	A	-	-	-	7	6	B
b=4	15	B	-	A	-	-	7	7	-
	16	A	A	A	A	A	7	8	A
	17	B	-	-	A	-	8	8	-
	18	A	A	-	A	-	8	9	A
b=6	19	A	-	-	-	-	9	9	-
	20	B	B	B	-	-	10	9	B
	21	A	-	-	-	-	10	10	-
	22	A	B	-	-	-	11	10	B
	23	B	-	B	B	-	12	10	B
	24	B	A	B	B	B	12	11	B
b=2	25	A	-	-	-	-	12	12	-
	26	B	B	-	-	-	13	12	B
b=6	27	B	-	-	-	-	13	13	-
	28	B	A	A	-	-	13	14	A
	29	A	-	-	A	-	13	15	A
	30	A	B	-	A	-	14	15	A
	31	A	-	B	-	-	15	15	-
	32	B	B	B	-	B	16	15	B
b=4	33	B	-	-	-	-	16	16	-
	34	B	A	-	-	-	16	17	A
	35	A	-	A	A	-	16	18	A
	36	A	B	A	A	-	17	18	A

Prediction Method	Prediction	Prediction Rate (PR)	Correct	Correct Guess Probability (CG)	Selection Bias Risk Score (SBRS)
Assume b=2	18	50%	8	44%	-0.0056
Assume b=4	14	39%	9	64%	0.1111
Assume b=6	12	33%	10	83%	0.2222
Assume b=8	4	11%	4	100%	0.1111
Convergent Prediction	23	64%	18	78%	0.3611

$$SBRS = PR \cdot (2 \times CG - 1)$$

Convergent Prediction for Fixed Blocks

For 2-arm equal allocation trials:

$P_A = \Pr(T = A)$, conditional allocation probability

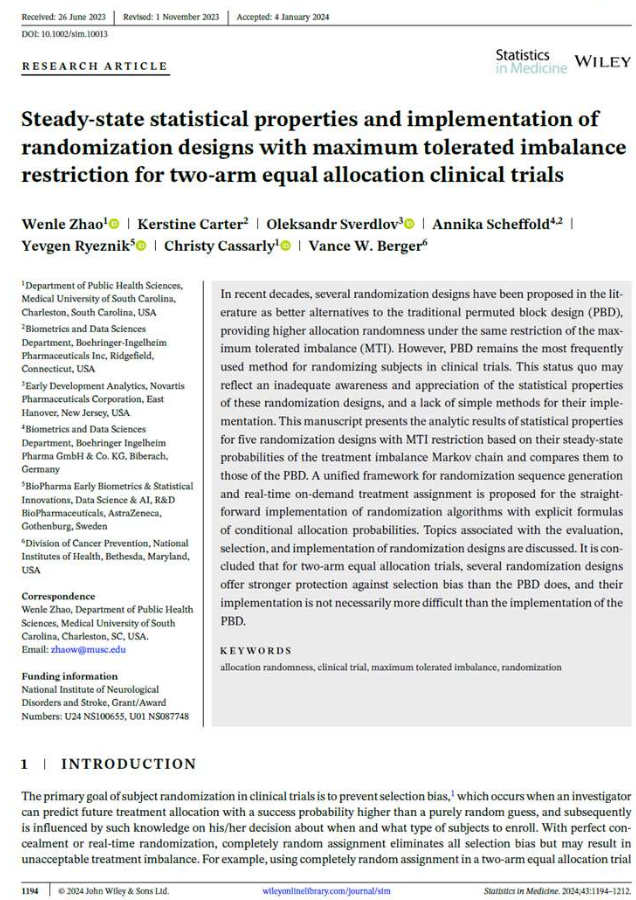
$D = N_A - N_B$, treatment imbalance

$DA = E(P_A = 0) + E(P_A = 1)$, deterministic assignments

$CR = E(P_A = 0.5)$, complete random assignments

$CG = 0.5 + E(|P_A - 0.5|)$, correct guess probability

$SD = \sqrt{E(D^2)}$, treatment imbalance standard deviation


Received: 26 June 2023 | Revised: 1 November 2023 | Accepted: 4 January 2024
DOI: 10.1002/sim.10013

RESEARCH ARTICLE

Statistics in Medicine WILEY

Steady-state statistical properties and implementation of randomization designs with maximum tolerated imbalance restriction for two-arm equal allocation clinical trials

Wenle Zhao[1] | Kerstine Carter[2] | Oleksandr Sverdlov[3] | Annika Scheffold[4,2] | Yevgen Ryeznik[5] | Christy Cassarly[1] | Vance W. Berger[6]

[1]Department of Public Health Sciences, Medical University of South Carolina, Charleston, South Carolina, USA
[2]Biometrics and Data Sciences Department, Boehringer-Ingelheim Pharmaceuticals Inc, Ridgefield, Connecticut, USA
[3]Early Development Analytics, Novartis Pharmaceuticals Corporation, East Hanover, New Jersey, USA
[4]Biometrics and Data Sciences Department, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany
[5]BioPharma Early Biometrics & Statistical Innovations, Data Science & AI, R&D BioPharmaceuticals, AstraZeneca, Gothenburg, Sweden
[6]Division of Cancer Prevention, National Institutes of Health, Bethesda, Maryland, USA

Correspondence
Wenle Zhao, Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA.
Email: zhaow@musc.edu

Funding information
National Institute of Neurological Disorders and Stroke, Grant/Award Numbers: U24 NS100655, U01 NS087748

In recent decades, several randomization designs have been proposed in the literature as better alternatives to the traditional permuted block design (PBD), providing higher allocation randomness under the same restriction of the maximum tolerated imbalance (MTI). However, PBD remains the most frequently used method for randomizing subjects in clinical trials. This status quo may reflect an inadequate awareness and appreciation of the statistical properties of these randomization designs, and a lack of simple methods for their implementation. This manuscript presents the analytic results of statistical properties for five randomization designs with MTI restriction based on their steady-state probabilities of the treatment imbalance Markov chain and compares them to those of the PBD. A unified framework for randomization sequence generation and real-time on-demand treatment assignment is proposed for the straightforward implementation of randomization algorithms with explicit formulas of conditional allocation probabilities. Topics associated with the evaluation, selection, and implementation of randomization designs are discussed. It is concluded that for two-arm equal allocation trials, several randomization designs offer stronger protection against selection bias than the PBD does, and their implementation is not necessarily more difficult than the implementation of the PBD.

KEYWORDS
allocation randomness, clinical trial, maximum tolerated imbalance, randomization

1 | INTRODUCTION

The primary goal of subject randomization in clinical trials is to prevent selection bias,[1] which occurs when an investigator can predict future treatment allocation with a success probability higher than a purely random guess, and subsequently is influenced by such knowledge on his/her decision about when and what type of subjects to enroll. With perfect concealment or real-time randomization, completely random assignment eliminates all selection bias but may result in unacceptable treatment imbalance. For example, using completely random assignment in a two-arm equal allocation trial

 |

Convergent Prediction for Fixed Blocks

Block Size	Complete Random Assignments	Deterministic Assignments (DA)	Correct Guess Probability (CG)	Selection Bias Risk Score (SBRS)	Treatment Imbalance (SD)
Fixed b=2	0.5	0.5	0.75	0.5	0.707
Fixed b=4	0.4167	0.3333	0.7083	0.4166	0.9325
Fixed b=6	0.3667	0.25	0.6833	0.3666	1.0847
Fixed b=8	0.3321	0.2	0.6661	0.3322	1.2258
Fixed b=10	0.3063	0.1667	0.6532	0.3064	1.3543

Convergent Prediction for Random Blocks

Predictability for random block design is the weighted average of predictabilities of member fixed blocks.

Definition:

block size b_j and weight w_j $(j = 1,2,\cdots,m)$.

CG_j: correct guess probability for block size b_j.

$CG = \frac{\sum_{j=1}^{m} b_j w_j CG_j}{\sum_{j=1}^{m} b_j w_j}$, overall correct guess probability.

$SBRS = 2CG - 1$, overall selection bias risk score.

Convergent Prediction for Random Blocks

Block Size	Complete Random Assignments	Deterministic Assignments (DA)	Correct Guess Probability (CG)	Selection Bias Risk Score (SBRS)	Treatment Imbalance (SD)
Fixed b=2	0.5	0.5	0.75	0.5	0.707
Fixed b=4	0.4167	0.3333	0.7083	0.4166	0.9325
Fixed b=6	0.3667	0.25	0.6833	0.3666	1.0847
Fixed b=8	0.3321	0.2	0.6661	0.3322	1.2258
Fixed b=10	0.3063	0.1667	0.6532	0.3064	1.3543
Varying b=2,4 equal weight	0.4447	0.3889	0.7222	0.4444	0.8639
Varying b=2,4,6 equal weight	0.4056	0.3194	0.7028	0.4055	0.9805
Varying b=2,4,6,8 equal weight	0.3762	0.2717	0.6881	0.3762	1.0853
Varying b=4,6 equal weight	0.3867	0.2833	0.6933	0.3866	1.0265
Varying b=4,6,8 equal weight	0.3624	0.2653	0.6812	0.3624	1.1195
Varying b=4,6,8,10 equal weight	0.3424	0.2179	0.6712	0.3424	1.2086

Random Block Design

Conclusion:

Random block design (varying block design) may reduce the motivation of allocation prediction for some investigators, it does not reduce the risk of selection bias, if convergent prediction is used.

there is no theoretical support that the RBD is less predictable than the PBD.

- Shao and Rosenberger

Properties of the Random Block Design for Clinical Trials

Hui Shao and William F. Rosenberger

Abstract To avoid deterministic treatment allocations in the permuted block design (PBD), many clinical trialists prefer randomizing the block sizes as well. While such a procedure is rarely formalized, it is generally assumed that the design will be less predictable. In this paper, we formalize the random block design by assuming a discrete uniform distribution for block size. The aim of this study is to provide a statistical understanding of the RBD, by investigating its distributional properties, including the degree of predictability and variability of treatment imbalance.

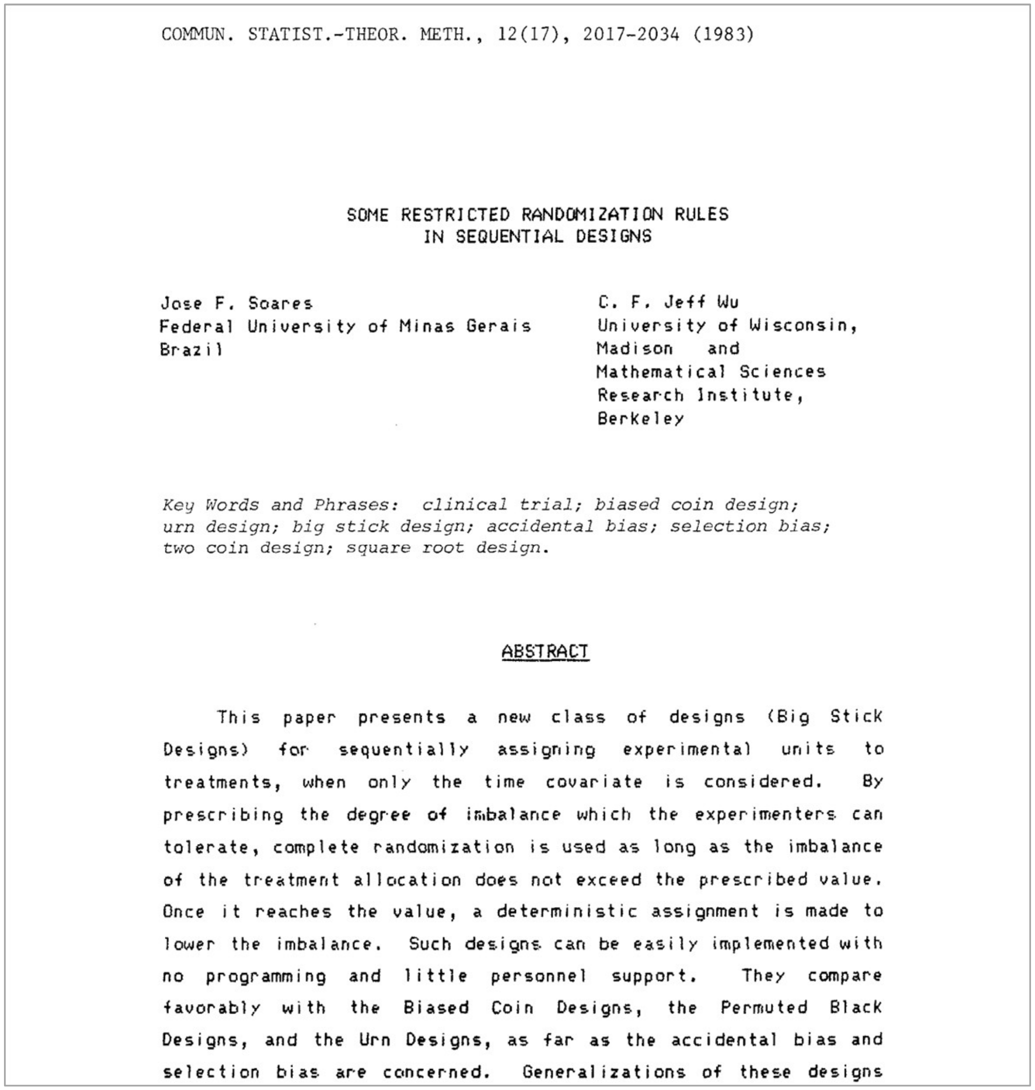
1 Introduction

The permuted block design (PBD) is the most popular randomization procedure. In a PBD procedure, first a number of blocks with equal even block size is established, then the treatment assignments are randomized within each block. The main advantage of using the PBD is that it assures balance or approximately balance for each treatment group throughout the course of the trial, especially when the sample size is small. A drawback of this design is that one or more treatment assignments in each block are deterministic and predictable when the trial is not blinded. Therefore, the random block design (RBD), in which the block size is randomly selected from a sequence of even integers, is proposed to reduce the predictability of future assignments and achieve balanced treatment allocation. However, this procedure is rarely carefully defined when it is employed, [5, 7] and there is no theoretical support that the RBD is less predictable than the PBD. We formalize the procedure by selecting block sizes according to a discrete uniform distribution on the even integers, and quantify the predictability of the RBD in two-arm clinical trials.

H. Shao (✉) • W.F. Rosenberger
Department of Statistics, George Mason University, 4400 University Drive, MS 4A7, Fairfax, VA 22030, USA
e-mail: hshao2@gmu.edu; wrosenbe@gmu.edu

J. Kunert et al. (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, DOI 10.1007/978-3-319-31266-8_26

Better Alternatives – Big Stick Design


COMMUN. STATIST.-THEOR. METH., 12(17), 2017-2034 (1983)

SOME RESTRICTED RANDOMIZATION RULES
IN SEQUENTIAL DESIGNS

Jose F. Soares
Federal University of Minas Gerais
Brazil

C. F. Jeff Wu
University of Wisconsin,
Madison and
Mathematical Sciences
Research Institute,
Berkeley

Key Words and Phrases: clinical trial; biased coin design; urn design; big stick design; accidental bias; selection bias; two coin design; square root design.

ABSTRACT

This paper presents a new class of designs (Big Stick Designs) for sequentially assigning experimental units to treatments, when only the time covariate is considered. By prescribing the degree of imbalance which the experimenters can tolerate, complete randomization is used as long as the imbalance of the treatment allocation does not exceed the prescribed value. Once it reaches the value, a deterministic assignment is made to lower the imbalance. Such designs can be easily implemented with no programming and little personnel support. They compare favorably with the Biased Coin Designs, the Permuted Black Designs, and the Urn Designs, as far as the accidental bias and selection bias are concerned. Generalizations of these designs

Better Alternatives – Big Stick Design

Soares JF, Wu CF, Some restricted randomization rules in sequential designs. *Commun. Stat*. 1983; 12:2017-2034.

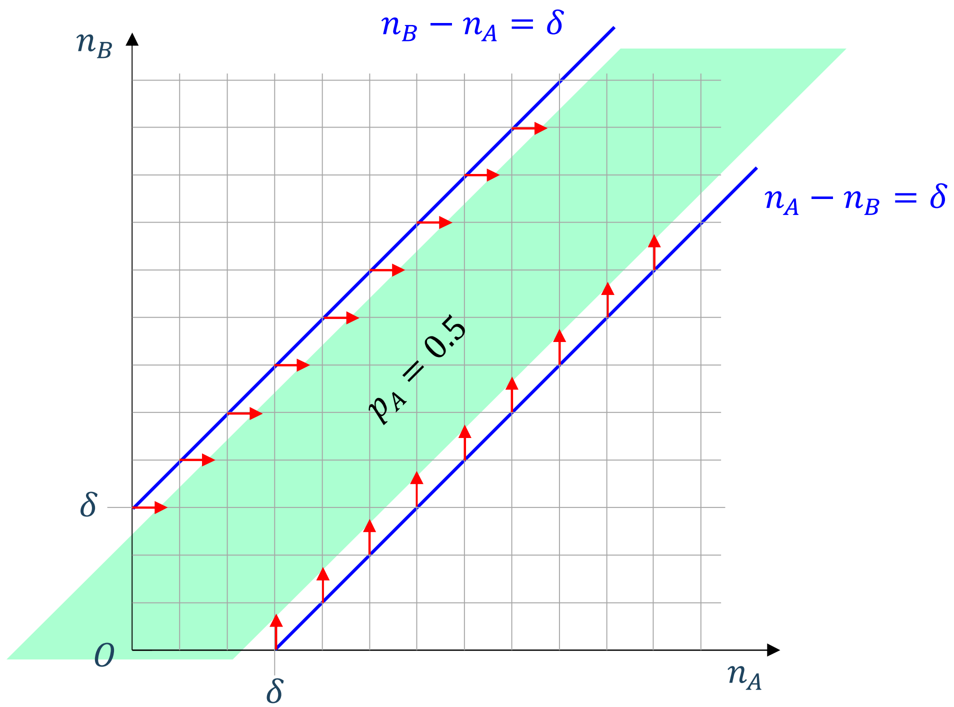
d : treatment imbalance,

$\delta = \frac{b}{2}$: maximal tolerated imbalance.

$$p_A\,(BSD) = 0.5 - 0.5 \times sign(d) \times int\left(\frac{|d|}{\delta}\right)$$

Using complete random assignments by default,

and deterministic assignment to reduce imbalance when it reaches the boundary.



Better Alternatives – Block Urn Design

Contemporary Clinical Trials 32 (2011) 953–961

Contents lists available at SciVerse ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

ELSEVIER

Block urn design — A new randomization algorithm for sequential trials with two or more treatments and balanced or unbalanced allocation

Wenle Zhao *, Yanqiu Weng

Department of Medicine, Medical University of South Carolina, Charleston, SC, USA

ARTICLE INFO

Article history:
Received 4 May 2011
Received in revised form 21 July 2011
Accepted 9 August 2011
Available online 22 August 2011

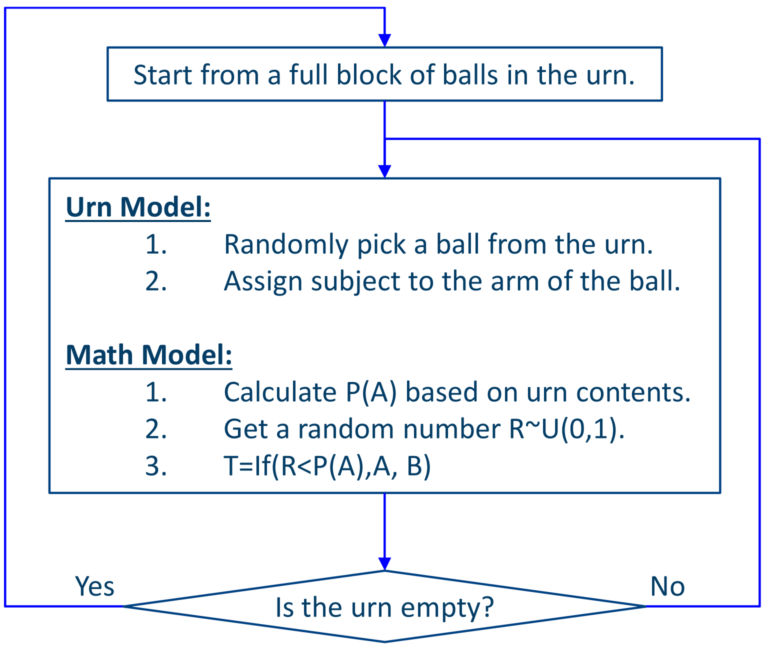
Keywords:
Randomization
Block urn design
Sequential clinical trial
Treatment imbalance
Allocation randomness
Deterministic assignment
Correct guess

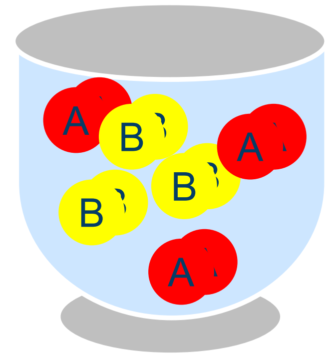
ABSTRACT

Permuted block design is the most popular randomization method used in clinical trials, especially for trials with more than two treatments and unbalanced allocation, because of its consistent imbalance control and simplicity in implementation. However, the risk of selection biases caused by high proportion of deterministic assignments is a cause of concern. Efron's biased coin design and Wei's urn design provide better allocation randomness without deterministic assignments, but they do not consistently control treatment imbalances. Alternative randomization designs with improved performances have been proposed over the past few decades, including Soares and Wu's big stick design, which has high allocation randomness, but is limited to two-treatment balanced allocation scenarios only, and Berger's maximal procedure design which has a high allocation randomness and a potential for more general trial scenarios, but lacks the explicit function for the conditional allocation probability and is more complex to implement than most other designs. The block urn design proposed in this paper combines the advantages of existing randomization designs while overcoming their limitations. Statistical properties of the new algorithm are assessed and compared to currently available designs via analytical and computer simulation approaches. The results suggest that the block urn design simultaneously provides consistent imbalance control and high allocation randomness. It can be easily implemented for sequential clinical trials with two or more treatments and balanced or unbalanced allocation.

The Urn Model for Permuted Block Design

Rosenberger WF, Lachin, JM. *Randomization in Clinical Trials Theory and Practice*. Wiley: New York, 2002.





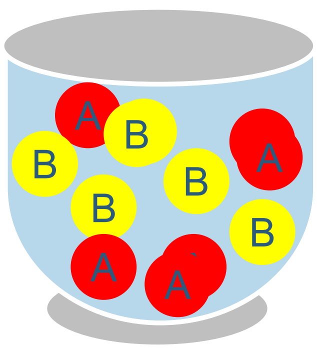
R=0.3205	P(A) = 0.50	A
R=0.6743	P(A) = 0.40	B
R=0.8971	P(A) = 0.50	B
R=0.9296	P(A) = 0.67	B
R=0.2408	P(A) = 1.00	A
R=0.0523	P(A) = 1.00	A

Return all balls of a full block when the urn is empty.

Better Alternatives – Block Urn Design

Zhao W, Weng Y. Block urn design - a new randomization algorithm for sequential trials with two or more treatments and balanced or unbalanced allocation. *Contemp Clin Trials*. 2011 Nov;32(6):953-61. doi: 10.1016/j.cct.2011.08.004.

$$p_A\,(BUD) = 0.5 - 0.5 \times \frac{d}{2\delta - |d|} = \pi r^2$$

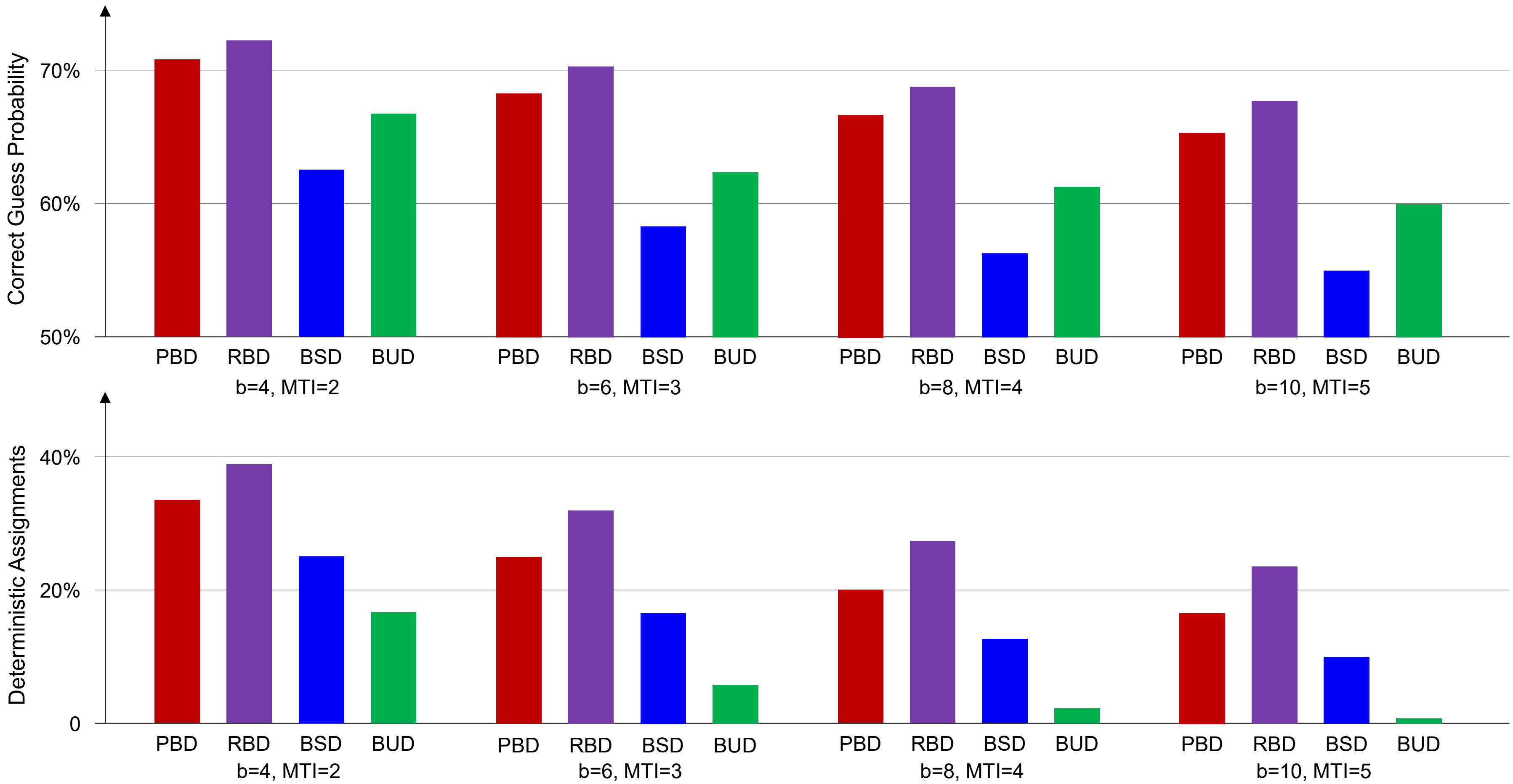


$p_A = 0.50$	R = 0.3205	A
$p_A = 0.40$	R = 0.6743	B
$p_A = 0.50$	R = 0.8971	B
$p_A = 0.60$	R = 0.9296	B
$p_A = 0.75$	R = 0.2408	A
$p_A = 0.60$	R = 0.0523	A

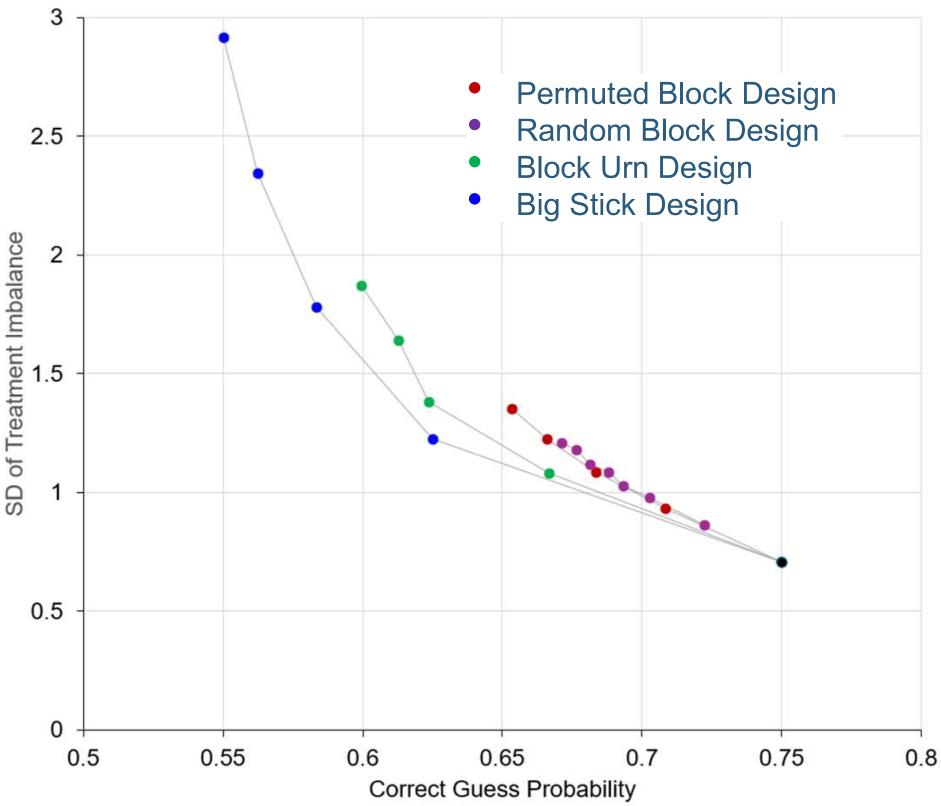
Permuted Block Randomization: Return a whole block of balls when the urn is empty.

Block Urn Design:
Return a set of balanced balls when it is available.

Better Alternatives: Big Stick and Block Urn



Trade-off Performance Comparison



Conclusion:

The myth about the advantages of random blocking is not true.

Actions Required

1. Call for an update of the ICH Guideline (E9)

2. Encourage IRT industry to offer better alternatives to blocking

medidata

ALMAC

SUVODA

PPD

ICON

IQVIA

perceptive

Guidance for Industry

E9 Statistical Principles for Clinical Trials

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
September 1998
ICH

Questions?