Introduction to Clinical Trials

Lecture 3: Randomization & Data Management

Giorgos Bakoyannis, PhD
Associate Professor
Department of Biostatistics and Health Data Science
Indiana University Indianapolis

Notes by Prof. Constantin T. Yiannoutsos, City University of New York

Section 1

Randomization

Why randomization is used in clinical trials

Assigning treatment in a random fashion in clinical trials is conceptually problematic as it contravenes the usual manner by which treatment is assigned by clinicians to patients. However, randomization addresses a number of potential problems:

- *Biological variability versus treatment effect*
  The biological component of the variability of treatment response is much larger than the treatment-mediated component of the treatment response. This results in (beneficial or adverse) responses that, without randomization might be incorrectly attributed to the treatment.

Why randomization is used in clinical trials (continued)

- *Confounding by indication*
  In many situations, the disease drives the kind of treatment that will be given to the patient leading to individuals with certain characteristics (e.g., more severely ill patients) receiving a particular treatment with higher probability. This confounds the treatment effect with prognostic factors.

  For example, if a more severely ill patient is more likely to receive treatment, receiving treatment is associated with disease severity (a "reverse causality" effect) thus erroneously appearing that treatment causes the severity of disease
- *Patient heterogeneity*
  Without randomization, researchers are unable to ensure that the subject groups will be homogeneous.

Randomization controls the influence of unknown factors

Randomization reduces selection bias, the bias that occurs when treatment assignment is based (consciously or not) on patient prognostic factors.

Randomization prevents investigators, for example, from assigning patients with better or worse prognosis to treatments that they feel are superior.

The main effect of randomization however is that it prevents confounding (i.e., change) of the treatment outcome by an effect that the investigator may have ignored or failed to measure (such as having assigned more women than men to a cardiovascular treatment; gender, in that case, is a confounder).

Control of confounders (cont')

But even if confounders are measured and even if they are adjusted for statistically, this does not necessarily solve the problem as statistical modeling may be unable to properly reflect the relationships between effect or ensure that its underlying assumptions are met.

Thus, having a randomized study obviates these problems and provides studies with credibility.

How do we randomize to ensure balance

Once randomization has been decided upon, randomly assigning treatment to subject groups is far from an easy exercise.

Consider the probability of having equal-sized groups in a two-arm study where subjects are randomized by the flip of a coin. If N (even) subjects are randomized, half to treatment A if the coin comes up "heads" and half to B of the coin comes up "tails" the probability that the two treatment groups will be equal (by considering the binomial distribution) is

$$P\left(X = \frac{N}{2}\right) = \left(\begin{array}{c} N \\ N/2 \end{array}\right) \frac{1}{2^N}$$

If $N = 100$ for example, $P\left[X = 50\right] = \left(\frac{100!}{(50!)(50!)}\right) \frac{1}{2^{100}} \approx 0.08$ In other words, there is only 8% probability that we will get two balanced groups of 50 patients if we randomize in this manner.

Blocking

Randomization in blocks improves balance.

A "block" contains a series of treatment assignments in a pre-specified proportion. The block size must be a multiple of the number of the treatments considered.

A number of blocks are generated, with treatment assignment being randomly permuted.

Illustration of randomization with permuted blocks

If blocks of size four are used, there are

$$\frac{4!}{2!2!} = 6$$

(permutations of the *multiset* $\{A, A, B, B\}$) possible ways to permute two group assignments as follows (Table 13.1 in the textbook):

Table: Blocked randomization of two treatments using blocks of size four

Within-block	Permutation number					
assignment number	1	2	3	4	5	6
1	A	A	A	B	B	B
2	A	B	B	B	A	A
3	B	A	B	A	B	A
4	B	B	A	A	A	B

Blocking and stratification of known prognostic factors

Blocking can accommodate stratification of known prognostic factors. Suppose that there are two prognostic factors, age (young/old) and gender (Males/Females) i.e., four strata. The blocks are (Table 13.2 in the textbook):

Table: Blocked randomization of two treatments using blocks of size four

	Strata			
Block	Males		Females	
Number	Young	Old	Young	Old
1	A	B	B	A
1	B	B	A	B
1	B	A	B	A
1	A	A	A	B
2	A	B	A	B
2	A	A	B	B
2	B	B	A	A
2	B	A	B	A
⋮	⋮	⋮	⋮	⋮

Patients are assigned to the next available treatment code in their stratum.

Adaptive allocation

So far we have considered randomization schemes where treatment allocation has constant probability throughout the study.

Adaptive allocation is an alternative approach to treatment allocation where the probability of receiving a treatment changes through the trial and is based on intervening events such as the current balance composition in the groups or even the observed treatment response.

Urn schemes

Consider randomizing subjects to two treatments by picking a ball from an urn. If a red ball is chosen then the subject will be randomized to treatment A and, if a white ball is chosen, to treatment B.

Suppose that we start with two balls, one of each color, and a red ball is chosen. Then, that red ball *plus an additional white ball* is returned to the urn. On the next try, the probability of getting a white ball is 2/3.

Eventually of course, adding balls to the "urn" has less and less effect and the randomization results in a simple 50/50 randomization. However, urn schemes are very good in ensuring tight balance of randomization, which may not be ensured with blocking schemes if the sample sizes are very small and/or the number of strata is very large.

Play the winner schemes

To minimize the number of patients receiving inferior treatments, several investigators have recommended schemes where the probability of treatment allocation depends on treatment efficacy.

Consider the same urn scheme described above. We start with two balls (one red and one white say) and assign a patient to each treatment based on the color of the drawn ball. Now, if the treatment is a failure, a ball of the opposite color is added to the urn along with the previously drawn ball. If the treatment is a success, a ball of the same color is added to the urn along with the previously drawn ball. Thus, the next patient has a higher probability of receiving the superior treatment (based on the last patient's response).

Play the winner schemes (continued)

The problem of implementing this design is that each response must be assessed *before* the next patient is randomized. A modification is to have a multi-stage procedure where the allocation ratio changes after review of responses at the end of each stage.

Another problem is the fact that sample size caclulations become difficult.

These designs are easier to "sell" to patients and Institutional Review Boards.

Randomization and blinding

In most trials, treatment assignment (randomization) lists are generated beforehand. It is easy to see that a study investigator having access to these lists may undermine the objectivity of treatment assignment.

To demonstrably show that this has not occurred, the randomization process must be kept separate from the clinical investigators. This is done by designating an institution or an off-site location as the trial coordinating center. Treatment assignments are generated there and are communicated to the investigators, *one at a time*, through the telephone or secure computer access.

Unequal treatment allocation

Usually studies employ equal treatment allocation. This is done because, equal allocation is the most efficient (in terms of variability, power and sample size) design. However, there are many situations where unequal treatment allocation is needed:

- Maximize the number of subjects assigned to the new treatment
- Treatments may differ in cost.
  It is mentioned in the textbook (Section 11.7.2) that a treatment allocation ratio $r : 1$ where $r = \sqrt{C}$ (the square root of the relative cost C) the total cost of study is minimized. For example, if one treatment is 10 times more expensive an allocation ratio of 3:1 will minimize cost.
- Variability in treatment response is unequal between the treatments. In this case, an allocation ratio $r = \frac{n_1}{n_2} = \frac{1/\sigma_1}{1/\sigma_2}$ equal to the ratio of the *inverse* of the two standard deviations will maximize the power of the statistical test.

Section 2

Data management in clinical trials

Management of data in clinical trials

- Data Management Plan
- Data Definition, Forms, and Database Design
- Patient Registration
- Quality Assurance and Control
- Good Clinical Practice
- Follow-up and Closeout Phase

Data Management Plan

The data management plan typically includes:

- Description of data management system
- Timeline for development
- Provisions for reporting
- Patient registration and randomization
- Patient confidentiality and security systems
- Quality control and assurance procedures
- Budget estimate and justification

Data definition, forms, and database design

Data definition, form and database design:

- Is critical to the success of a study
- These activities are interrelated
- They shouuld be best done in the order listed
- The entire trial team should participate in this process
- Trials should not be activated before the data collection forms are finalized

Data definition

Types of data to be collected

- Identification: uniquely identify each study patient (study no., patient ID, site/institution)
- Research: directly address study objectives (baseline characteristics, side effects, end points)
- Administrative: aid in the management of the trial (sample tracking, study personnel)
- Regulatory: show compliance with regulations (IRB approval, patient consent)
- Reference Center: demonstrate quality control

Data definition (continued)

- Should be done during protocol development phase
- Should involve all members of the trial team
- Limit data collection to the essential items needed to address study objectives and conduct the trial
- Review analysis plan to ensure all necessary data are included

Design of case report forms (CRF)

- Case Report Forms (CRFs) are printed or electronic documents that are used to collect the required trial data.
- Measurement and recording of data are the most critical steps in the overall data management process.
- Activating a trial before CRFs are in place will likely result in inconsistent and incomplete data.
- Pilot testing of CRFs can identify problems before the trial begins and gives the end-users a chance to have meaningful input into the design

CRF design

- When designing CRFs consider the following questions:
  - Who will be completing the forms?
  - When will the data be available?
  - Where will the data be collected?
- Understanding the timing, location, and personnel involved in collection will help determine which data items belong on which forms.

CRF design (continued)

- Well-designed forms should employ clear, precise, unequivocal questions and coded responses.
- When appropriate, responses should be mutually exclusive and include units of measurement.
- Use of open-ended questions should be avoided.
- Consider the overall organization of data both cross-sectional (one-time collection) and longitudinal (multiple observations collected).
- Cross-sectional data such as gender and DOB should not be re-collected on follow-up forms.
- Information that is collected repeatedly over time should be formulated in the same manner at each time point.

Database design (continued)

- Closely linked to the definition of data
- A common database structure is one that has a separate table for each CRF.
- In order to link these tables, we need to define the key fields that uniquely identify a record (i.e. patient identifiers and visit date).
- Select appropriate data attributes (type, length, and format)
- Choose meaningful field names
- Prepare for missing data

Patient registration

- Registration is the process of entering a patient on a trial.
- In trials where there are multiple institutions participating and one central coordinating center, each institution normally submits patient information to the coordinating center which in turn provides patient identifiers and details about the assigned treatment.

Patient registration (cont.)

The registration process typically includes:

- Verification that the institution is currently a participant in good standing.
- Verification that the investigator is authorized.
- Verification that all regulatory requirements have been met including approval from an institutional Ethics Committee or Review Board and proper patient informed consent.
- Verification that the patient is eligible.
- Collection of demographic data for monitoring purposes or assignment of treatment.

Patient registration (cont.)

- Direct treatment assignment
  - All patients or a specific subset are assigned to a particular treatment.
  - More complex designs may involve direct assignment of some patients and randomization of others based on certain patient characteristics.
- Randomization of treatment assignment
  - Patients are assigned at random to one of the treatment arms of the trial.
  - Requires safeguards to ensure integrity of the randomization process and minimize the possibility of treatment selection.

Quality assurance

Quality Assurance is the set of processes, procedures, and activities that are initiated prior to data collection to ensure that the expected level of quality will be reached.

- Data Dictionary
- Form Design
- Electronic error prevention and control
- Missing data
- Training and informing of clinical staff

Electronic error prevention and checking

- Electronic interface should mirror the paper form as much as possible.
- Use of drop-down menus ease selection and restrict entry to only valid responses.
- Entry of required fields such as patient ID and date should precede entry of additional data.
- System should prevent entry of duplicate patient ID's or duplicate observations on the same patient for the same visit.
- Utilization of range restrictions on numeric fields will prevent entry of erroneous data. Include logical checks to conditionally restrict entry.

Managing missing data

- *In the clinic*
  If possible, review of CRFs or source documents prior to patient departure will minimize missing data.
  - This procedure should be part of the clinical staff training and oversight.
- *Point of data entry*
  Entry of CRFs that are missing key variables such as patient ID or visit date should not be permitted. Alert data entry clerk about other missing variables so that forms can be tagged for review.
- *Post-data entry*
  Missing data queries should be generated by the Coordinating Center and sent to the participating sites for review.

Training and informing of clinical staff

- Regular training in accurate completion of encounter forms is vital to ensuring data quality.
- Review of CRFs by an independent individual ensures proper completion and improves data quality.
- Inclusion of the signature of the person completing the forms (and/or the site PI) encourages and ensures accountability.

Quality control

- Quality Control (QC) is the process of monitoring and maintaining the reliability, accuracy, and completeness of the data during the conduct of the trial.
- The QC process should provide a consistent and objective review of data and assessment of endpoints.
- It is recommended that the review of data include those involved in the collection of trial data and someone independent of this group.

Ensuring data quality

Point of assessment

- *Patient Registration*
  Confirm eligibility criteria are met
- *Collection*
  Review completeness of form before patient leaves the clinic
- *Entry*
  Range restrictions, logical checks, double-entry
- *Post-entry*
  Performance of clean-up queries
- *Statistical analysis*
  Assessment of study endpoints

Good Clinical Practice (GCP)

- Standards by which clinical trials are designed, conducted and reported.
- Protect the rights of patients entered on trials.
- Ensure confidence in the integrity of the data collected and the results published.
- As a result of several tragedies such as the use of thalidomide in the early 1960s, the World Health Organization initiated discussions to regulate clinical research involving humans.
- This led to the Helsinki Declaration, an international agreement signed in 1964 to address ethical issues in the conduct of clinical trials.
- US FDA expanded these requirements and introduced the Code of Good Clinical Practice in 1978.

GCP Responsibilities for participating sites

- Ethics Committee approval of trial protocol
- Recruitment of patients to the trial
- Obtaining patient consent and informing them of the possible risks of the treatment
- Collection and recording of data required by the protocol
- Reporting of adverse events
- Ensuring protocol compliance
- Ordering, storage, and administration of study drug

Patient informed consent

As part of the consent process, patients should receive the following:

- Description of research plan
- Information about possible risks and benefits
- Summary of any procedures or requirements that go beyond standard practice
- Information about the alternative treatment options
- Name and phone number of the responsible physician to contact if there are problems or questions

Patient informed consent

As part of the consent process, patients should receive the following:

- Patients should also receive:
- Assurance that confidentiality will be maintained but that certain people will have access to trial data and may also have access to patient's medical records
- Information that the patient has the right to refuse participation without affecting access to future medical care or the relationship with their clinician
- Assurance that the patient has the right to withdraw from the trial at any time

GCP Responsibilities for coordinating center

- Confirming Ethics Committee approval at each site
- Confirming informed consent of patients
- Source verification of submitted data – typically requires visit to site to review source documents
- Screening researchers to ensure that they are appropriately qualified
- Quality Control and computerization of CRFs
- Ongoing monitoring of the safety of the treatment
- Analysis of the trial and reporting of results

Implementing GCP

Maintaining Source Documentation

- GCP requires that data submitted on CRFs can be verified by review of source materials at the participating sites
- Source materials include patient's medical or clinical record, x-rays, scans, pathology slides, trial chart, copies of CRFs submitted, and data queries from the Coordinating Center
- US guidelines require that original records be retained for a minimum of 7 years from the approval of treatment by the FDA

Completing CRF

These procedures should be followed in order to comply with GCP:

- Read and follow all instructions carefully
- Write legibly using permanent ink
- Submit original form and retain copy at local site
- Write answers inside the space provided
- Answer all questions
- Use the units of measurement requested on the form

Completing CRF (continued)

Additional procedures to comply with GCP:

- When a data correction is needed, cross thru the erroneous information with a single line and write in the correct response above or to the side. Initial and date all corrections.
- Check forms for completeness and correctness before submission to the Coordinating Center
- Ensure that identifier information is correct
- Submit forms according to the schedule defined in the protocol

Monitoring visit/audit

- The Coordinating Center will organize visits to each participating site to conduct periodic data reviews.
- Data from submitted CRFs will be systematically compared with data in the patient's medical record (source documents)
- Audit may involve all patients, a random sample, or a specific subgroup (such as all patients who have died)

Follow-up and closeout phase

Preparing data for analysis

- Freezing data
  - Errors can be corrected but no new data is added
- Progress reports for monitoring accrual, eligibility, safety and attrition
- Formal interim reports
  - For Data Safety and Monitoring Board(DSMB)
  - IRB Continuing Review

Follow-up and closeout phase

- Participating sites are asked
  - Submit all overdue CRFs and responses to queries
  - Return study materials such as drugs or devices
  - Prepare trial records for archiving
- Coordinating Center is responsible for
  - Completion of all required analysis
  - Submission of trial materials (CRFs, regulatory documents, randomization details, copy of database, unused drug) to sponsor.
  - Preparation of a copy of the database, related software, analysis files, and supporting documents for archiving.

References

1. McFadden, Eleanor (1998). Management of Data in Clinical Trials. New York: Wiley.
2. US Department of Health and Human Services, Food and Drug Administration (April 1996). Guidance for Industry E6 Good Clinical Practice: Consolidated Guidance. http://www.fda.gov/cder/guidance/index.htm