
Clinical Trials

Part 6

ISSUES IN DATA ANALYSIS

- General comments about data analysis
- Baseline comparability of randomized groups
- Non-compliance with protocol and intention-to-treat analysis
- Competing events

ISSUES IN DATA ANALYSIS

- Covariate adjustment
- Multiplicity problem in clinical trials
- Some special analyses – survival analysis, longitudinal analysis
- Some final recommendations

GENERAL COMMENTS

- The better a clinical trial is designed and conducted, the easier and simpler will be the data analysis
- Too much emphasis on p-values; don't forget to look at the data (descriptive statistics –frequency distributions, means, standard deviations, Box plots, etc.)

GENERAL COMMENTS

- Data analyses should be specified in the protocol
 1. Baseline comparability of groups
 2. Analyses for all primary and secondary hypotheses
 3. How protocol non-compliance will be handled – missing data, drop-outs, drop-ins, non-compliers

GENERAL COMMENTS

- Data analyses should be specified in the protocol
 4. Covariate adjustment
 5. Subgroup analyses
 6. Interim statistical monitoring
 7. Analysis of exclusions from randomization -- generalizability

BASELINE DESCRIPTION AND COMPARABILITY OF GROUPS

- One of the first analyses performed
- Select 5-10 important baseline descriptors
 1. Demographics – age, gender, race
 2. Some severity of illness measures (e.g., EF in CHF, HbA1C or FBS in Diabetes, etc.)
 3. Selected comorbidities (CAD, CHF, COPD, Diabetes, HTN, PVD, CVD, etc.)
 4. Clinical or laboratory variables
 5. Variables that are prognostically important to primary outcome

BASELINE DESCRIPTION AND COMPARABILITY OF GROUPS

- Does the randomization look reasonable (i.e., Are the groups balanced on the baseline variables? Expect 5% to be statistically significant spuriously)

INTENTION-TO-TREAT (ITT) PRINCIPLE

- The standard or primary method of analysis for RCT
- All randomized patients are analyzed in the treatment group they were randomized to
- Includes patients who:
 - Do not get assigned treatment
 - Receive wrong treatment
 - Die before treatment given
 - Do not adhere to study protocol
 - Drop out of study

WHY ITT?

- Preserves the randomization and balance of groups
- p-values are valid no matter what
- Evaluates the policy of assigning patients to one group or another
- Provides conservative estimate of treatment effect, what you might expect if you apply the treatment to a large population (Effectiveness trial)
- If trial is under control, no issue

ARGUMENTS AGAINST ITT

- Does not provide a true test of treatment efficacy (effect of treatment in those who follow the study protocol)
- Alternatives often suggested
 - "Per protocol" or "compliers" analysis – excludes patients who did not follow protocol
 - "As treated" analysis – analyzes patients according to the treatment they received
- Can lead to bias because excluded patients may differ from analyzed patients, or groups might no longer be balanced

COMPETING EVENTS

- Can reduce power and bias the trial by reducing the number of patients available for follow-up (e.g., deaths in a trial where the major outcome is something other than death) (e.g., cost-effectiveness of HBHC)

COMPETING EVENTS

- In a trial where cause-specific mortality is primary outcome, deaths from other causes can reduce power (e.g., in a trial of antiarrhythmic drug reducing sudden cardiac death, death from cancer would reduce power)

COMPETING EVENTS

- In a trial of treatment for a cause-specific mortality, total total mortality should also be considered, since the treatment might be harmful in other respects. (e.g., some lipid lowering agents have been shown to reduce coronary heart disease mortality but increase mortality from other causes)

COVARIATE ADJUSTMENT

- A covariate is an extraneous factor (independent variable) which is not the primary independent variable of interest but may affect outcome
- In any study, one should collect data on important covariates (e.g., demographic factors – age, gender, race, SES; disease severity; other factors of prognostic importance to outcome) and report them

COVARIATE ADJUSTMENT

- The goal of a clinical trial is to produce balance on all known and unknown factors, through randomization, except for the intervention under study
- Covariate adjustment is essential in observational, non-randomized studies, but is less essential in randomized clinical trials

COVARIATE ADJUSTMENT

- Covariate adjustment might be done in an RCT because:
 - 1) In spite of randomization, some baseline covariates might still be imbalanced
 - 2) Covariate adjustment can improve precision (precision can also be worsened if there is large measurement error in the covariate)

COVARIATE ADJUSTMENT

- Covariate adjustment should be done with care because adjusting for covariates that are moderately imbalanced between treatment groups or moderately predictive of the outcome variable or judicious selection of certain covariates could change the results of the RCT

RECOMMENDATION FOR COVARIATE ADJUSTMENT

- Randomization should be done carefully and well. This is one of the most critical elements of the RCT
- The primary analysis of the RCT should be the unadjusted analysis. Secondary covariate adjusted analyses could be done using multiple regression for continuous outcome variable, logistic regression for dichotomous outcome variable, Cox regression for time to event outcome variable

RECOMMENDATION FOR COVARIATE ADJUSTMENT

- Any covariate adjustment should be prespecified in protocol
- In the planning of the RCT, select a few covariates known from previous studies to be important predictors of outcome and stratify and adjust for them in the analysis

MULTIPLICITY IN CLINICAL TRIALS

- Many types of multiplicity
 - Comparison of multiple treatment groups
 - Multiple outcomes
 - Repeated measures over time
 - Multiple looks at outcome data as data are accumulating
 - Analyzing results of trial in various subgroups
 - Use of different statistical methods on same data

MULTIPLICITY IN CLINICAL TRIALS

- General problem: In 100 independent tests at $\alpha = 0.05$ level, on average 5 will be statistically significant spuriously. Which ones are real, and which are spurious?
- Independence/dependence of tests generally not known

MULTIPLE TREATMENT GROUPS

- Encourage one experimental intervention and one control
- Experimental intervention – the most important one that will have greatest chance of changing clinical practice
- More than 2 treatment arms will require larger sample size and inevitably a multicenter study

MULTIPLE TREATMENT GROUPS

- Effect on sample size is more than just adding a treatment arm; also have to reduce α by using $\alpha = 0.05/k$, where $k = \#$ comparisons

EXAMPLE

- CSP #411
 - CABG/PCI vs. No Coronary Intervention Prior to Patients Undergoing Elective Vascular Surgery; physician chooses CABG or PCI
 - N = 673 for alpha = 0.05, Power = 0.90, 85% vs. 75% Survival
 - If you wanted to randomize patients to CABG vs. PCI vs. No Coronary Intervention and compare CABG vs. no intervention and PCI vs. no intervention with alpha = 0.05/2 = 0.025, Power = 0.90, 85% vs. 75% survival, then N = 1191 (more than 673 + 336 = 1009).
 - If you also wanted to compare CABG vs. PCI, this would make sample size even larger since you are comparing 2 active treatments and would expect a survival difference less than 10 percentage points

MULTIPLE OUTCOMES

- Some possible outcomes: Mortality, morbidity, symptoms physiologic measures, lab tests, QOL, adverse effects, health care utilization, costs
- Recommend you identify one most important primary outcome on which to base sample size, others secondary

MULTIPLE OUTCOMES

- One combined primary outcome is a possibility –
 - CSP #465 “Glycemic Control and Complications in Type II Diabetes:” (CV events – death, MI, PTCA/CABG, CVA, development of CHF, Amputation for Ischemic Gangrene)
 - Secondary outcomes – angina, TIA, claudication, retinopathy, nephropathy, microalbuminuria, neuropathy, QOL, costs, adverse events (hypoglycemia)

MULTIPLE OUTCOMES

- Components of the one combined primary outcome should not have wide variability in importance (e.g., do not include CHF hospitalization or angina with mortality or major morbidity)

MULTIPLE OUTCOMES

- If multiple primary outcomes are used, need to adjust alpha level or do global statistical test
 - Adjusting alpha – less powerful, but permits individual testing of each outcome (e.g., Bonferroni, α/k , $k = \#$ outcome variables)
 - Global test methods – more powerful but lack interpretability (e.g., Hotelling's T or MANOVA)

MULTIPLE OUTCOMES

- In CSP #304 (Cochlear Implants) we used factor analysis/principal components analysis to develop one primary outcome variable from a battery of 20+ audiologic tests (Henderson, Controlled Clinical Trials, 11:199-214, 1990)

REPEATED MEASURES OVER TIME

- Common in studies with outcomes such as physiologic measures (BP, pulmonary function tests, etc.), lab measures (e.g., FBS, HbA1c, cholesterol), or QOL measures
- Do not compare treatment groups with a test at each time point at 0.05 level – this will produce a multiplicity problem
- There are statistical techniques of longitudinal data analysis that permit keeping experiment wide alpha level at 0.05

MULTIPLE LOOKS AT OUTCOME DATA AS DATA ARE ACCUMULATING

- In trial where primary outcome is mortality or major morbidity, this is ethically mandatory
- Rules for monitoring outcomes will be defined by Data and Safety Monitoring Board
- As number of looks at alpha = 0.05 increases, overall true alpha error increases beyond 0.05
- Various interim statistical monitoring schemes are available to keep overall alpha level at 0.05 (Friedman, et al. Chapter 15)

SUBGROUP ANALYSES

- The overall results of clinical trial give average treatment effects across a large group of patients
- A natural question is: Are there particular subgroups where the treatment works better or worse than average?

SUBGROUP ANALYSES

- It is best to specify subgroups of interest as secondary hypotheses in the protocol, prompted by what is suggested by the literature. Otherwise, subgroup analysis will be viewed as a “fishing expedition”.
- Both alpha and beta errors can be increased by doing subgroup analyses

SUBGROUP ANALYSES

- Beta error (probability of not finding treatment effect in sample when one exists in population) is increased because sample size is smaller for subgroups than the entire sample
- Alpha error (probability of finding treatment effect in sample that does not exist in population) is increased because you are doing multiple testing

SUBGROUP ANALYSES

Prob. (finding at least one treatment effect in G Subgroups) =

$$1 - (1 - \alpha)^G$$

For alpha = 0.05 G = 5, Prob. = 23%
 G = 10, Prob. = 40%

SUBGROUP ANALYSIS

- Contexts in which subgroups arise (strongest to weakest)
 - Specific secondary hypothesis in protocol motivated by the literature – strongest
 - Stratification factors, e.g., age, gender, stage of disease
 - Subgroups identified by other studies
 - Subgroup findings emerging during trial
 - Post hoc analysis (“data dredging” or “fishing”) – weakest – many are possible; if you do enough, some will be statistically significant spuriously; hypothesis generating only; need confirmation by other studies

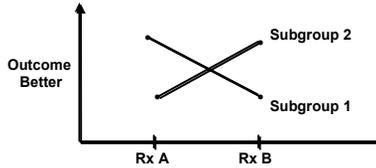
SUBGROUP ANALYSIS

- Believeability is strengthened by:
 - Similar results found in several trials
 - Internal consistency within trial for a number of subgroups and sites
 - Biological explanation Unless the main overall result is significant, investigators should be particularly conservative in evaluating significant subgroup findings

SUBGROUP ANALYSIS

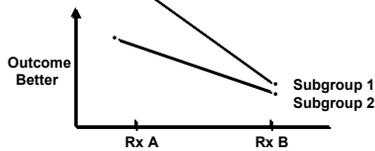
- It is sometimes helpful to distinguish between quantitative (noncrossover) and qualitative (crossover) subgroup treatment effects.

- Qualitative (crossover) effects – Treatment A is better than Treatment B in Subgroup 1, while Treatment B is better than Treatment A in Subgroup 2. This is relatively rare and of greater clinical importance



SUBGROUP ANALYSIS

- Quantitative (noncrossover) effects – Treatment A is superior to Treatment B in all subgroups, but magnitude of treatment effects differ; these are common in clinical trials and do not necessarily imply that treatment effects are truly different in subgroups



EXAMPLES: CSP #127

		Overall (n=683)	Caucasian (n=295)	African-American (n=388)
SBP	Prop.	-10.4	-13.2	-8.2
	HCTZ	-18.1	-15.3	-20.3
		p<.001	N.S.	P<.001
DBP	Prop.	-10.8	-12.6	-9.5
	HCTZ	-12.0	-10.9	-13.0
		p=0.03	p=0.02	p<.001

EXAMPLES: CSP #320

- Amiodarone vs. Placebo in Patients with Congestive Heart Failure & Ventricular Arrhythmia

USE OF DIFFERENT STATISTICAL METHODS

- Different statistical methods applied to the same data will not yield identical p-values
- e.g., In analysis of mortality in CSP #411, we could use:
 - Chi-square test (with or without Yates correction) or Fisher's Exact Test on % of deaths
 - Kaplan-Meier estimates of survival and log-rank test or Wilcoxon test
 - Assuming exponential distribution, a parametric test comparing the hazard rates of the groups
 - Cox regression, taking into account covariates
 - Life table analysis

USE OF DIFFERENT STATISTICAL METHODS

- Different statistical methods are based on different assumptions about the data
- Analyses should be chosen on basis of the most reasonable assumptions about the data, and should be pre-specified in protocol
- Do not analyze the data in different ways until you get a p-value less than 0.05

USE OF DIFFERENT STATISTICAL METHODS

- Example in Friedman, et al. (p.309) of use of different cutpoints

Cutpoint	Rx A		Rx B		
	Pre-HR	Post-HR	Pre-HR	Post-HR	
	73.96	73.2	74.40	73.96	
<input type="checkbox"/> 7		8%	0%		P = 0.15
<input type="checkbox"/> 5		24%	0%		P = 0.009
<input type="checkbox"/> 3		32%	28%		P = 0.76

SURVIVAL ANALYSIS (CH. 14)

- Standard method of analyzing time-to-event data
- Commonly used in studies where primary outcome variable is mortality and/or major morbidity
- Takes into account staggered entry of patients into the trial and patients who are alive or without major morbidity at end of trial (censored observations)

SURVIVAL ANALYSIS (CH. 14)

- Types of analyses
 - Life table analysis if survival time known in intervals
 - Kaplan-Meier survival curve if exact survival time is known
 - Log-rank test to compare survival curves
 - Cox proportional hazards model if you want to use covariates also

LONGITUDINAL DATA ANALYSIS

- Useful for outcome measures repeated over time in follow-up (e.g., blood pressure, depression scale, QOL measures)
- For both categorical and continuous variables
- Measure patients at fixed time points in follow-up (e.g., every 3 or 6 months or annually)
- Missing data is the biggest problem, due to missed visits, lost to follow-up, deaths, etc.

MISSING DATA IN LONGITUDINAL ANALYSIS

- Three Classes Distinguished
- MCAR – Missing completely at random
Missingness (nonresponse mechanism) is independent of outcome of interest; can drop subjects from analysis
- MAR – Missing at random; missingness depends only on observed and not missing outcome; can analyze data by likelihood methods without modelling non-response mechanism

MISSING DATA IN LONGITUDINAL ANALYSIS

- MNAR – Missing not at random; non-response mechanism depends on both the observed and missing outcomes; requires joint modelling of both observed and missing data mechanisms

MISSING DATA IN LONGITUDINAL ANALYSIS

- Avoid if possible
- Collect data on reasons for missingness
- Inadequate methods
 - Analysis of complete cases if data are not MCAR
 - (Watch out for software!)
 - LOCF – Last observation carried forward
- Great new reference:
 - Fairclough, Diane. Design and Analysis of
 - Quality of Life Studies in Clinical Trials.
 - Chapman & Hall/CRC, 2002.

FINAL RECOMMENDATIONS

- Involve a biostatistician early in the design stages of your clinical trial; he/she can help in all aspects of protocol development
- Keep design of trial simple and clean so that analysis will be straight forward; if you need MNAR techniques, you are in trouble!
- Consider estimation techniques as well as p-values

FINAL RECOMMENDATIONS

- Use descriptive statistics generously – get to know your data
- Prespecify all techniques in protocol
- ITT is king and watch out for multiplicity
