General Principles: FDA Introduction

The Science of Small Clinical Trials
Great Room, White Oak Campus
FDA

Lisa M. LaVange, PhD
Director, Office of Biostatistics
Office of Translational Sciences, CDER, FDA
November 27, 2012
Motivation
Rare Diseases and Small Trials

- Issues
  - Affected populations <200,000; may be many fewer
  - Patients are a scarce resource
    ➨ Careful attention required to design an efficient study able to provide a clear signal for decision making
  - Number of clinics qualified to treat a very rare condition may be small, and screening processes may be complex
    ➨ Making it all the more important to work with each clinic for successful recruitment and study conduct
Rare Diseases and Small Trials

- **Opportunities**
  - Patient registries of some rare diseases are available to help in study planning
  - Clinical networks may be established, with infrastructure built to help in study conduct
  - Patient advocacy may be well-organized and able to be tapped for assistance in recruitment
  - Patients are often very motivated to contribute to research efforts, particularly if unmet medical need
Rare Diseases and Small Trials

- What are the statistical aspects of small trials that have the potential to improve efficiency?
  - Study design
  - Study conduct
  - Data analysis
Study Design
The Basics

- Tendency in small trials to view fundamentals of quality trial design as impractical or unfeasible
- Before choosing a single-arm, non-randomized, open-label design, consider what is being sacrificed with respect to:
  - Choice of controls
  - Randomization
  - Masking
Control group

- **Treatment effect:**
  - The effect of any treatment for a given patient is the difference between what happened to the patient as a result of giving him the treatment and what would have happened had treatment been denied
  - Presents practical difficulties: We cannot observe what would have happened
  - 3 possibilities:
    - Historical controls
    - Within-patient change from baseline
    - Concurrent controls

*S. Senn, Statistical Issues in Drug Development, 2007*
Control group

Problems in measuring what might have happened:

1. **Historical controls** are different patients using different treatments at different times and in different places
   - Caution: Placebo arms in clinical trials are often not indicative of the natural course of the disease, making use of patient registry data or natural history studies difficult as a comparator

2. **Within-patient change** (comparing each patient’s outcome with their baseline value) does not account for changes unrelated to treatment, e.g., deterioration in chronic diseases or resolution of acute episodes
   - Caution: Regression to the mean can occur
Regression to the Mean

- We expect the means from a random sample of stable patients to be similar at two time points.
- For a distribution to remain stable at two time points:
  - The worst patients tend to improve
  - The best patients tend to deteriorate
- Discovered by Galton (1822-1911)
  - ‘Revert to mediocrity’
- How is this relevant to clinical trials? Consider an example:
  - Patients selected for a clinical trial of an antihypertensive agent
  - Inclusion criteria is SPP>160 and DPB>120
  - Expected value of next BP measurement is lower than baseline
  - Without a concurrent control group, improvement is misinterpreted as a treatment effect
  - Within a control group, improvement is called placebo response
Percent of patients by Systolic Blood Pressure as a function of their Baseline Systolic Blood Pressure: *Regression to the Mean*
Control group

- Gold Standard = Concurrent Controls
  - An experimental treatment is studied at the same time as an alternative control treatment
- Eliminates bias due to differences in patients studied during different time periods
- Eliminates bias due to ‘Regression to the Mean’
  - Regression to the mean is equally likely to occur in all treatment groups
  - Treatment effect is the difference between groups, above and beyond that attributed to regression to the mean
- Allows consistent reporting of prognostic factors
Placebo as control

• Dose response studies often benefit from inclusion of a placebo arm
• May not be ethical, and not always required when ethical, e.g., when placebo response is known a priori to be zero or nearly zero (e.g., spontaneous cures of serious infections not likely to occur)
• Interpretation of dose response may differ based on inclusion of placebo – see example on following slides
Evidence of a dose response?

![Bar graph showing efficacy response across different dose levels: Low, Medium, High.](image)
Evidence of a dose response?
Randomization

- Allows for unbiased treatment assignment
  - Without this, more diseased patients may be assigned to the test drug, if the drug seems promising, or to the control, if the test drug seems risky
- Allows for inferential statistics to be employed
  - Tests of hypotheses about treatment efficacy assume randomization
- Ensures statistical validity of the conclusions of the study
- Ensures comparability of treatment groups, even for factors unable to be measured (if baseline differences exist, they are due to chance)
- Problem: Patients may not want to risk randomization to placebo
Randomization

- **Strategies:**
  - Use standard of care in lieu of placebo
  - Allow early rescue treatment
  - 2:1 or other allocation skewed toward experimental treatment
    - Has the advantage of providing more exposure data, but will also have some associated loss of power
  - Consider randomized withdrawal designs
  - Allow placebo patients to switch to experimental treatment for a safety follow-up phase
    - Caution: efficacy data collected after the point of switching is difficult to interpret, especially if subjective endpoint
Randomization

- Stratification can enhance inference, ensure balance
- Stratification by site commonly used
  - Captures differences in physician practice, patient populations, geographic regions, and other unmeasured factors
  - Simplifies drug handling in some settings
- Caution: avoid over-stratification, e.g., by site, age, and disease severity
  - Low enrolling sites will result in imbalances—this is a particular problem for small studies
- If unable to stratify by site, central randomization via web portals or IVR allows stratification by other factors, balanced across site
- Note -- minimization algorithms are not random
Masking

- Gold standard = double-masked study
- Allows for unbiased evaluation of patient’s response
- Always strive for this standard, unless unethical or impossible
- Alternatives:
  - Physician and patient are aware of treatment assignment, but all other clinical personnel, sponsor’s personnel, and CRO’s personnel are masked
  - In particular, clinical outcome assessor is masked
  - Operate on a ‘need to know’ basis
  - Caution required to guard against unwilling unmasking, e.g., through side effects, lab parameter levels, etc.
Study Conduct
Study conduct

- Quality control is a critical element in ensuring the validity of a study’s findings
  - Even more important in small studies, where achieving a high level of accuracy is essential

- Three areas of focus:
  - Data quality
  - Training
  - Monitoring
Data quality

- Increase in data quality $\rightarrow$ increase in precision of estimates $\rightarrow$ study more likely to provide a clear signal about the test treatment and less likely to waste resources

- Data quality efforts should start early and include:
  - Case Report Form (CRF) design using proven tools
  - Careful consideration of mode of data collection, esp. for questionnaire administration, symptom diaries, etc.
  - Edit specifications, both real-time and periodic checks
  - Data transfer plans (and QA) for central labs and reading centers

- Overarching goal is to reduce measurement error in key variables

- Interim review of masked data for QC oversight
  - Evaluation of measurement error and variability (in time to modify analysis plans, if needed)
Case Study

- Phase 2 pulmonary trial in rare disease population
- Two key design decisions based on QC considerations
  - Use of a central reading center for pulmonary function testing
    - Standardization and calibration of spirometry equipment
    - QC oversight provided by reading center; re-trained clinic staff as needed
  - Average of two screening tests used as baseline value in ANCOVA to reduce variance, increase power
- Result of both procedures was to reduce the hypothesized standard deviation of the primary outcome by approx. 1/3

*Deterding, et al., 2007*
Training and monitoring

- Central training of clinic personnel on data acquisition, data entry, and response to data queries/edit resolution
  - Certification of clinic personnel and re-training, as needed

- Clinical monitoring
  - Site visits still most common, but use of electronic data capture tools makes central monitoring useful
  - Statistical plans for site visits can help focus resources
    - Stratified random sampling of sites to monitor at periodic intervals
    - Possible stratification factors include # patients enrolled, rate of early discontinuation and other protocol violations, and region/country
    - Incorporate ‘for-cause’ monitoring to allow re-training at problem sites
Open-label studies

- Open-label studies present a particular challenge
  - Operate on a ‘need to know’ basis
  - Important that sponsors and CROs working for sponsors remain masked
  - Randomization schedules should be protected and not shared, just as with a double-masked study
    - Possible to influence composition of treatment arms if know what the next assignment will be
  - Cumulative event rates or summary statistics should be the purview of Data Monitoring Committees and not the sponsor, clinical staff, or others
    - May be tempting for clinical monitors to view summary data and infer results, if EDC is used
    - Results could influence clinic practice for remainder of study
Clinic networks

- Establishing a research network of clinical centers treating a rare disease can help ensure efficient and high quality study design and conduct

- Case study: Cystic Fibrosis Foundation Therapeutics Development Network (CFFTDN)
  - Network of clinics with access to CF patient population and trained personnel
  - Infrastructure in place, e.g., EDC systems
  - Coordinating Center at Seattle Children’s Research Institute experienced in study design and conduct, analysis and reporting

- End result
  - Shorter study start-up and close-out times
  - Invaluable advice provided to industry sponsors
Data Analysis
Data analysis

- The small sample sizes expected in rare disease studies make it essential to optimize data analysis strategies

- Strategies:
  - Planning for and minimizing attrition
  - Controlling for subject to subject variability
  - Controlling Type 1 error
  - Adaptive designs
Attrition and its impact

- Design elements to minimize attrition include
  - Training clinical staff on importance of retention and its impact on the study’s power
  - Planning for continued follow-up, even if subjects discontinue treatment
    - Should be reflected in the informed consent document
    - Use of data collected post-discontinuation described in statistical analysis plan
  - Consider allowing all patients to receive experimental drug for a limited period after randomized study period concludes
    - Motivates patients to remain in the study
    - Extreme case is a full cross-over study with the intent to analyze the first period as primary
Missing data

- Too much missing data is difficult to overcome, even with the best analysis methods
  - > 10% attrition causes suspicion
  - > 20% attrition can invalidate the study
- If data are reasonably complete, several analysis strategies available to manage missing values
  - Single value imputations popular in the past, but increasingly criticized, e.g., last observation carried forward (LOCF)
  - Repeated measurements modeling also popular, but assume randomly missing values
  - Time-to-event analyses incorporate censored values, so less problematic
Missing data

- Final analysis plan should include the planned strategy for handling missing data
- Sensitivity analyses should be included and pre-specified, insofar as possible
  - Purpose is to assess the sensitivity of the study results to the method used for handling missing data
  - A key component of the statistical regulatory review is to determine whether methods and assumptions are valid and study findings can be supported, in light of any missing data that may occur
- National Academy of Sciences report on missing data issued in 2011 at request of FDA
  - Working group in Office of Biostatistics formed to assess review practices in light of this report
Covariate adjustment

- Minimizing variability of study outcomes increases the ability of detecting a clear signal
- Statistical approaches to provide control for various patient level characteristics during data analysis include:
  - Stratified chi-squared tests for categorical outcomes (e.g., extended Mantel-Haenszel statistics)
  - Stratified log rank tests or Cox proportional hazards models for time-to-event outcomes
  - Analysis of covariance models, parametric or non-parametric, and possibly with repeated measurements of outcomes
Covariate adjustment

- Factors to consider for control include
  - Stratification factors used in randomization
  - Covariables known a prior to be strongly correlated with response or outcome

- Other considerations
  - Factors should be assessed prior to randomization
  - Factors should be pre-specified in the protocol or final analysis plan

- Clinical center is a common factor accounted for in analysis
  - May adequately capture variation in geographic region, patient populations, and medical practice
  - Note that with valid randomization, baseline imbalances are due to chance; unadjusted analysis is valid, but usually sub-optimal
  - Purpose of controlling for baseline factors is to increase power of hypothesis tests and precision of estimates
Multiplicity, multiplicity, multiplicity

- Multiple assessments ➔ multiple opportunities for findings due to chance
- Example: 5 tests at \( \alpha = 0.05 \); overall Type I error = 0.226
- Arises from:
  - Multiple comparisons, e.g., pairwise comparisons among multiple treatment arms or doses
  - Multiple primary endpoints, e.g., clinical signs and symptoms
  - Multiple looks at the data, e.g., interim analyses
Multiplicity, multiplicity, multiplicity

- Various methods of adjustment
  - All methods impact significance testing
  - Some impact power
- Common procedures
  - Bonferroni, Holm, and Hochberg for multiple endpoints or doses
  - O’Brien-Fleming, Lan-DeMets, and Pocock for multiple tests (interim and final analyses)
- Protocol should clearly state method of adjustment
  - How the criteria for success of the study is affected
- If two or more endpoints are required to establish efficacy (co-primary endpoints), inflated Type II error may be of concern
- Draft guidance document due for release in 2013
Adaptive Designs

- Studies can be designed and conducted with adaptive features that are guided by the examination of accumulating data at interim time points
- Objectives:
  - Conduct more efficient studies, e.g., with fewer patients or of shorter duration
  - Increase the likelihood of success
  - Improve understanding of the treatment effect
- Prospective planning and careful implementation needed to avoid inducing bias, inflating Type I error, or producing results difficult to interpret
- Draft guidance document issued 2010; final guidance reflecting public comments in progress
Adaptive Designs

- Common elements for adaptation include: eligibility criteria, randomization procedures, doses and dose regimens, sample sizes, endpoints, analysis methods
- To date, more traditional approaches, e.g., group-sequential methods, used in confirmatory trials, while more novel strategies, e.g., Bayesian methods, used in exploratory trials
- Trend for increasing acceptance of novel approaches at the confirmatory stage in some disease areas, e.g., oncology (I-SPY), anti-bacterials, etc.
- Later in this conference, both frequentist and Bayesian approaches to adaptive designs will be discussed
- The pay-off for small trials and rare diseases can be significant!
References
References

Back-up slides