ASA Biopharmaceutical Section
Regulatory – Industry Workshop

ICON
A Symbol of Excellence

Multi-stage adaptive designs: Simulation considerations

Sergei Leonov (ICON Innovation Center)
Joint work with
Greg Cicconetti (AbbVie), Alan Hartford (AbbVie),
Yili Pritchett (MedImmune), Michael Smith (Astellas)

September 27, 2017

Outline

- Designs with sample size re-estimation
- Designs with potential for early stopping for efficacy/futility
- Confirmatory multi-stage designs
- Maintenance of type I error

Background

DIA Adaptive Design Scientific Working Group activities

- Simulation practices for adaptive trial designs

Motivating sample size re-estimation (SSR)

Problem: at study design stage, there is limited information on estimates of the treatment effect and nuisance parameters

Possible solutions

1. Run a pilot study to gain information on study design parameters, after which a more reliable sample size calculation can be made
2. Start the trial with a sample size based on 'best guesses' for the study parameters: plan an interim analysis (IA) of the accumulated data, to update estimates of these parameters; perform sample size re-estimation
   - Blinded SSR (**bSSR**): use pooled blinded data to update nuisance parameters
   - Unblinded SSR (**ubSSR**): observed estimates of treatment effect and nuisance parameters are used

Motivating designs with early stopping

Problem: accumulating evidence suggests little chance to meet the planned study objectives (or study objectives have already been met)

Possible solutions

1. Continue to end despite any ethical or economic considerations
2. Design study with pre-planned IAs of unblinded data to assess the likelihood that the trial will meet its objectives and allow for early stopping
   - Group sequential designs*:* allow for single or multiple IAs performed sequentially, with an option of stopping for futility or success
   - Adaptive designs: allow for more general design changes; can include SSR, and the ability to stop the trial based on efficacy and futility
     - This extends to designs with a *prospectively planned opportunity for modification* of one or more specified aspects of the study design and hypotheses based on analysis of interim data

Protocol planning considerations

- **Feasibility**: whether enrollment rate and time for endpoint readout allow for trial modifications
- **Logistics**: Determine number and information rates for IAs
- **SSR**
  - Identify parameters to be re-estimated and SSR methodology
  - Specify maximum sample size
  - Identify the method of final data analysis with adjustment for potentially inflated Type I error rate due to sample size increase
- **Studies with stopping rules**
  - Identify the assessments considered at IAs: stop for efficacy and/or futility along with corresponding criteria for taking actions
  - Understand the effect of stopping on the final analysis and appropriate adjustment for these effects

The role of simulation

- Means to study the implications of prospectively defined trial modifications in a very controlled setting
- Key features
  - Patient generation (endpoint, enrollment, dropout)
  - Protocol modification rules to be followed
  - Which metrics to record
- Some basic questions
  - How do trial modifications impact study end results?
  - Are there '*optimal*' timings for interim analyses?
  - What's the likelihood of stopping early?
  - How often can we expect SSR to call for an increase in study size?
  - Robustness: what happens when simulation assumptions are violated?

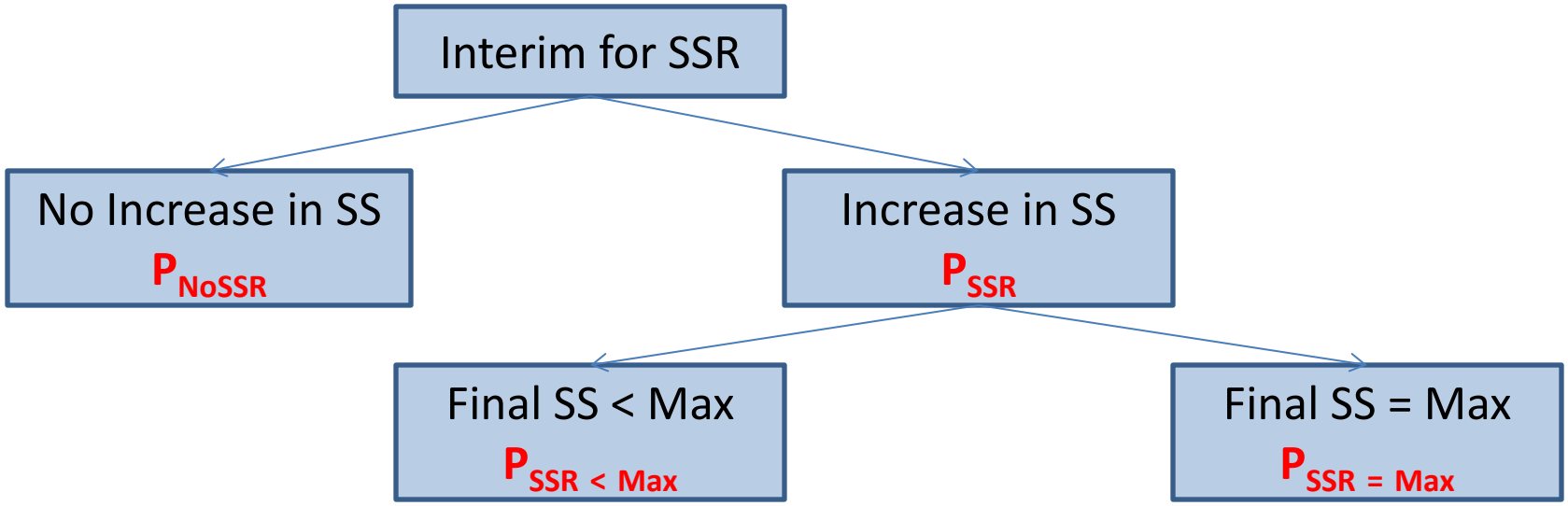
Decision rules

- How do trial modifications impact study end results?
  - Metrics: power, type I error, bias, coverage of the confidence intervals
- Are there '*optimal*' timings for IAs? Do trial characteristics limit choices?
  - Enrollment vs. data available for decision-making
  - Time lag between IA trigger and taking an action on a DMC decision
  - Metric: proportion of studies stopped early for efficacy/futility
    - Single IA: what is the impact of choosing an earlier or later time?
    - Multiple IAs: what added value does each IA bring?
  - Metric: proportion of studies with correct interim calls for efficacy/futility (This is assessable if the simulation 'plays out' stopped trials as if stopping rules were not applied)

Decision rules (cont.)

- How often can we expect SSR to call for an increase in study size?
  - Metrics: Proportion of trials where SSR is actioned, final sample size
- What happens if enrollment and/or dropout modeling is incorrect?
  - Enrollment/dropout impacts the totality of data available at an interim
    - Slower enrollment and higher dropout rates
  - Run simulations under faster/slower enrollment assumptions and greater/smaller dropout rates (check robustness)
- Additional considerations for time to event trials
  - SSR: recalibration of the number of events collected
  - Impact of trial modifications on trial duration
  - Increasing the number of events: do we increase number of subjects to mitigate extension to trial duration?
    - Impact on median observation time

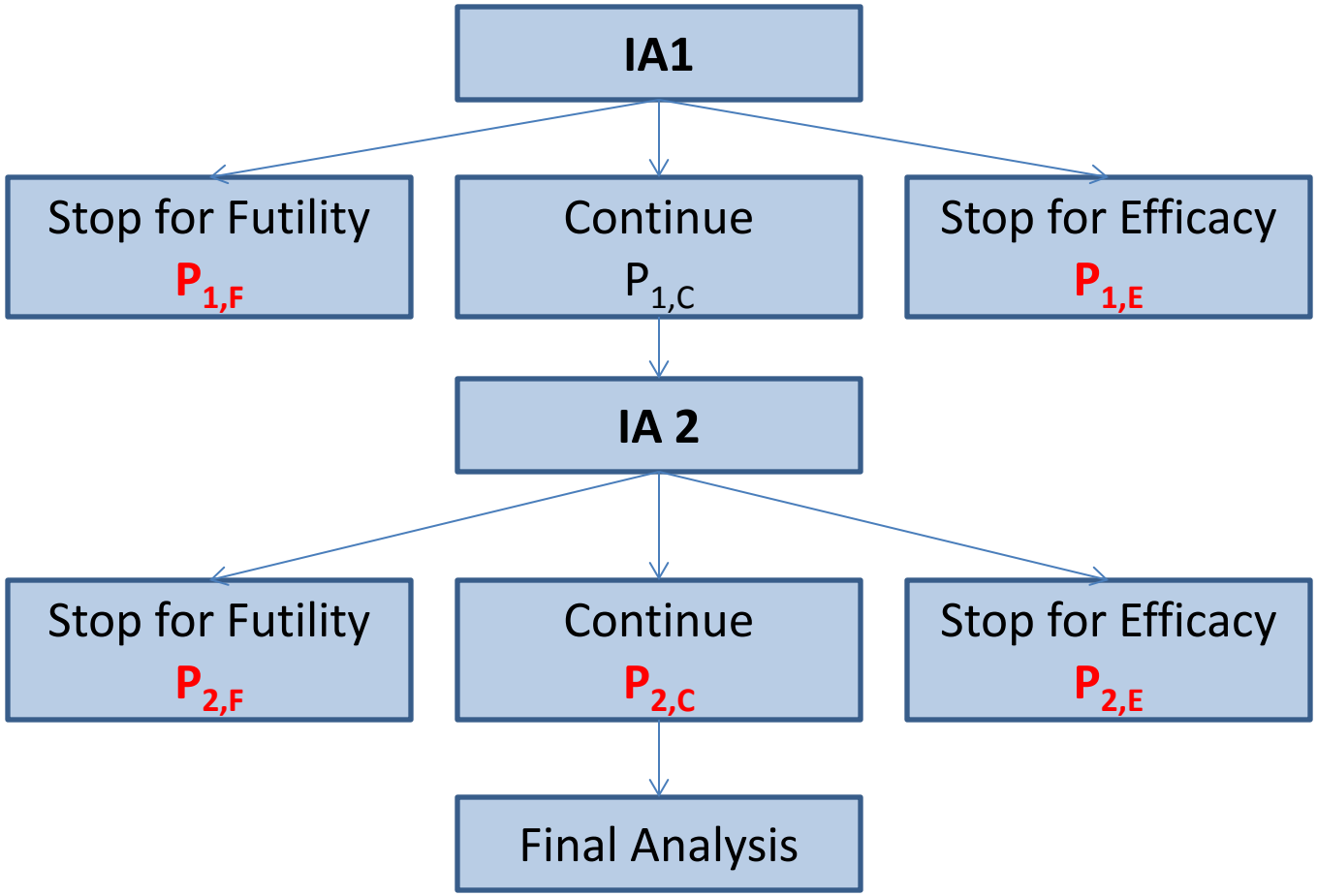
Flow diagram: SSR with single IA



Possible outcomes

- Increase in the sample size (SS): whether this increase is at the maximum sample size limit, or it is below the imposed limit
- No increase
- **Red probabilities provide a full accounting of possible trial outcomes**

Flow diagram: stopping rule, two IAs



Possible outcomes

- Stop for some pre-defined ‘efficacy’ rule, pre-defined ‘futility’ rule or not to stop

Summary table: SSR and stopping for efficacy/futility

IA 1		IA 2	
Stopping Rule	**SSR**	**Stopping Rule**	**Probability**
Stop for futility	NA	NA	$P_{1,F}$
Stop for efficacy	NA	NA	$P_{1,E}$
Continue			$P_{1,C}$
Continue	No change in sample size	ALL	P_{1CN}
Continue	No change in sample size	Stop for futility	$P_{1CN,F2}$
Continue	No change in sample size	Stop for efficacy	$P_{1CN,E2}$
Continue	No change in sample size	Continue	$P_{1CN,C2}$
Continue	Increase in sample size	ALL	P_{1CI}
Continue	Increase in sample size	Stop for futility	$P_{1CI,F2}$
Continue	Increase in sample size	Stop for efficacy	$P_{1CI,E2}$
Continue	Increase in sample size	Continue	$P_{1CI,C2}$

Red probabilities provide a full accounting of possible trial outcomes

- Assessment for stopping (for efficacy or futility), then SSR at the first IA
- Assessment for stopping (for efficacy or futility) at the second IA

Simulation report template: design with SSR & stopping rules

1. **General section:** description of study objectives and key elements of the design (patient population, duration, treatment groups, study endpoints, etc.)
2. **Designs to be considered** (traditional fixed design vs. proposed adaptive design)

A. Statistical methods for SSR

- bSSR vs. ubSSR
- Analytic derivations (if appropriate) of re-estimated sample size
- Method to control Type I error rate and parameter inferences for final data analysis when sample size is increased after SSR
- Original/maximum sample size and corresponding range for unknown parameters

B. Statistical methods for stopping rules

- Type of stopping rules (futility or efficacy) and clinical interpretation, shape of stopping region, relevant parameters such as conditional power
- Methodology for calculating stopping rules, with a reference to software used
- Description of calculation of IA results
- Controlling family-wise type error rate (e.g., α-spending function)

C. Statistical methods for estimation

- Point estimate/ confidence intervals

D. Diagram(s) illustrating the full set of possible outcomes from each adaptation point

Simulation report template: design with SSR & stopping rules

3. Simulation plan

A. Simulation objectives

1. Show that the adaptive design adequately meet requirements for essential operating characteristics (control of false positive rate, point estimates, confidence regions, power)
2. Compare with alternative designs (e.g., fixed design without SSR/early stopping)
3. Perform sensitivity analysis with respect to timing of IAs, enrollment rate, drop-out rate, maximum sample size, etc.

B. Algorithm for data generation and procedures involved in the simulation process

C. Design settings: number and timing of IAs, stopping boundaries, SSR rule, etc.

D. Number of simulation runs, random seed

E. Software used in the simulation studies (EAST, ADDPLAN, gsDesign, etc.)

- Provide code for the proprietary (own) software

Simulation report template: design with SSR & stopping rules

4. Simulation results

1. Operating characteristics (according to Table of simulation metrics)
2. Probabilities of reaching each of adaptation outcomes according to the simulations
3. Thoughtfully designed data tables and graphical presentations are encouraged

5. Summary and recommendation

1. Select design parameters based on simulation
2. Compare the adaptive design to the fixed design

- **References**
- **Appendix: calculation details**
  - Can be used to capture detailed mathematical formulas used in study design and simulations

Confirmatory multi-stage designs

- Included: seamless phase II/III and enrichment designs with adaptive choice of sub-populations
- Control of type I error rate: combination tests and the conditional error rate principle have been proposed
- Need to account for treatment/population selection
  - Take the k best, unconditionally
  - Take the k best, subject to some criteria
- Sample size and allocation ratio following treatment [population] selection
  - Final sample size for controls and each selected arm: fixed or flexible
  - The latter case: total sample size is fixed but the sample size per arm depends on the number of arms to be continued to the final analysis
- Simulations are typically used to investigate/optimize designs' properties due to the complex nature of such multi-stage designs

Maintenance of type I error rate

- **Best case**: analytic demonstration of Type I error control
- **Good case**: analytic demonstration of asymptotic Type I error control
- General agreement among industry, academic and regulatory bodies:
  - **Type I error rate control cannot be established via simulation**
- FDA Guidance ***Use of Bayesian Statistics in Medical Devices Clinical Trials*** (2010): **adequate characterization of operating characteristics, including type I error, may require extensive simulation**
  - Ways to reign in type I error rate
    - Changes to aspects of study design (e.g., the number/timing of interims)
    - Altering study futility/success criteria
    - Increasing of study size
  - Note: changes to study design require a new round of simulations

Maintenance of type I error rate (cont.)

- FDA Guidance ***Adaptive Design Clinical Trials for Drugs and Biologics*** (2010): use of simulation to show control of type I error is '**controversial and not fully understood**'
- Lin et al. (2016): role of simulations in submissions to CBER
  - Simulations have helped evaluate type I error rate in circumstances when control of type I error is only guaranteed asymptotically
  - Need for extensive use of simulation when trial designs are complex; if study end results differ from the range of assumptions used in trial simulations, study interpretation could be compromised
- **Prescription Drug User Fee Act (PDUFA) VI Commitment Letter** (Aug 2017)
  - Starting in FY 2018, FDA will conduct a pilot program for highly innovative trial designs for which analytically derived properties (e.g., Type I error) may not be feasible, and **simulations are necessary to determine trial operating characteristics**
  - By end of 2020: FDA will develop or revise, as appropriate, relevant MAPPs, SOPPs and/or review templates and training to incorporate **guidelines on evaluating complex clinical trial designs that rely on computer simulations to determine operating characteristics**

Maintenance of type I error rate: simulation size

Let p be type I error associated with the significance test

- Typical simulation iteration
  - Sample observations
  - Run a statistical test, record 1 [0] if [not] significant → $U_1, \ldots, U_n$
  - Variance of the mean of n i.i.d. Bernoulli r.v.: Var = $p(1-p)/n$
- We want st. deviation to be bounded by *small c*

$$St.\ dev = \sqrt{\frac{p(1-p)}{n}} < c \quad \Rightarrow \quad n > \frac{p(1-p)}{c^2}$$

- $p \sim 0.05$, $c \sim 0.001$ → $n > 50{,}000$
- Can we do better than 50,000?
  - Mukhopadhyay , Cicconetti (2004): 2-stage sequential sampling methodology for simulation size determination

Concluding remarks

- **PDUFA VI Commitment Letter** (2017): increased role of simulations in evaluating complex trial designs
- **21st Century Cures Act** (2016): incorporating complex adaptive and other novel trial designs into proposed clinical protocols
- **FDA Guidance on Adaptive Design for Medical Devices** (2016): computer simulations can play a crucial role in adaptive designs
- **FDA Guidance on Adaptive Designs for Drugs and Biologics** (2010): detailed documentation on computer simulations required in the study protocol

Our WG: an attempt to create a framework for conducting simulations and developing a simulation report for adaptive designs

Selected references

- Bauer P, Bretz F, Dragalin V, König F, Wassmer G (2016). Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. *Statist. Med*., **35**(3), 325–347.
- Bauer P, Köhne K (1994). Evaluation of experiments with adaptive interim analyses. *Biometrics*, **50**(4), 1029–1041.
- Bretz F, Koenig F, Brannath W, Glimm E, Posch M (2009). Adaptive designs for confirmatory clinical trials. *Statist. Med*., **28**(8), 1181-1217.
- Cui L, Hung HMJ and Wang SJ (1999). Modification of sample size in group sequential trials. *Biometrics*, **55**(3), 853–857.
- Gaydos B, Anderson KM, Berry D, Burnham N, Chuang-Stein C, Dudinak J, Fardipour P, Gallo P, Givens S, Lewis R, Maca J, Pinheiro J, Pritchett Y, Krams M (2009). Good practices for adaptive clinical trials in pharmaceutical product development. *Drug Inf. J.*, **43** (5), 539-556.
- Gould AL, Shih WJ (1998). Modifying the design of ongoing trials without unblinding. *Statist. Med*., **17**(1), 89-100.
- Jennison C, Turnbull BW (2000). *Group Sequential Methods. Applications in Clinical Trials*. Chapman & Hall/CRC, Boca Raton, FL.
- Lehmacher W, Wassmer G (1999). Adaptive sample size calculations in group sequential trials. *Biometrics*, **55**(4), 1286–1290.

Selected references (cont.)

- Lin M, Lee S, Zhen B, Scott J, Horne A, Solomon G, Russek-Cohen E (2016). CBER's experience with adaptive design clinical trials. *Ther. Innov. Regul. Sci.,* **50**(2), 195-203.
- Mukhopadhyay N, Cicconetti G (2004). How many simulations should one run? In: *Applied Sequential Methodologies: Real World Examples with Data Analysis.* (Mukhopadhyay N, Datta S, Chattopadhyay S, eds.), pp. 261-292. Marcel Dekker, NY.
- Müller HH, Schäfer H (2001). Adaptive group sequential designs for clinical trials: Combining the advantages of adaptive and of classical group sequential approaches. *Biometrics*, **57**(3), 886–891.
- Pritchett YL, Menon S, Marchenko O, Antonijevic Z, Miller E, Sanchez-Kam M, Morgan-Bouniol CC, Nguyen H, Prucka WR (2015). Sample size re-estimation designs in confirmatory clinical trials—current state, statistical considerations, and practical guidance. *Statist. Biopharm. Res.,* **7**(4), 309-21.
- Proschan MA, Hunsberger SA (1995). Designed extension of studies based on conditional power. *Biometrics*, **51**(4), 1315–1324.
- Proschan, MA, Lan KKG, Wittes JT (2006). *Statistical Monitoring of Clinical Trials. A Unified Approach*. Springer, NY.
- Wassmer G, Brannath W (2016). *Group Sequential and Confirmatory Adaptive Designs in Clinical Trials*. Springer Series in Biopharmaceutical Statistics, Switzerland.

Back up: simulation size

- Simulation size from a sequential estimation perspective
  - Mukhopadhyay , Cicconetti (2004): 2-stage sequential sampling methodology for simulation size determination.
  - Motivating scenario: Type I error estimation when observations are Tukey random variables
    - Mixtures of two normal random variables with common mean → symmetric, mound shaped distributions with tails that are lighter/heavier than normal
  - Sample size determination as a bounded-risk problem: 1st and 2nd order asymptotic properties of a two-stage simulation size determination framework

Questions:

Sergei.Leonov@iconplc.com