# Design of Phase I and II Clinical Trials

Susan Stewart
Division of Biostatics

UCDAVIS HEALTH | Clinical and Translational Science Center

# Topics

- Phase I
  - Objectives
  - Types
    - Rule-based
    - Model-based
    - Model-assisted
- Phase II
  - Objectives
  - Types
    - Multi-stage
    - Randomized
    - Crossover

# Phase I clinical trials

- Phase I (NIH definition): Tests a new biomedical intervention in a small group of people (e.g. 20-80) for the first time to determine efficacy and evaluate safety (e.g., determine a safe dosage range and identify side effects).
  - Healthy volunteers or patients who have failed other treatments
  - Determine maximum tolerated dose (MTD)
  - Assess pharmacokinetics (PK, what the body does to a drug ) & pharmacodynamics (PD, what a drug does to the body)

# Cancer clinical trials design

- Most cancers are life-threatening
- Most anticancer drugs are cytotoxic with a narrow therapeutic window
  - Low doses: ineffective, but toxic
  - Higher doses: effective, but even more toxic
- Phase I and II studies designed to
  - Minimize the number of people exposed to toxic, ineffective treatments
  - Select efficacious treatments with an acceptable safety profile efficiently

# Phase I cancer trials

- Participants are usually patients who have exhausted their treatment options
- Goal is to determine the maximum tolerated dose (MTD)
- Doses above the MTD have unacceptable levels of dose-limiting toxicity (DLT)

# General considerations

- Starting dose may be based on animal studies
- Participants are usually
  - Critically ill
  - From a small pool of available patients
  - Heterogeneous, e.g., with different tumor types
- Phase I cancer trials
  - Are a screening process to find potentially effective drugs with an acceptable safety profile
  - Determine the MTD with a minimal number of patients in a minimal amount of time
  - Establish the MTD from below (due to the extreme cytotoxicity of the drugs)

# Types of Phase I designs

- Rule-based: Assign patients to dose levels according to pre-specified rules based on observations of events (e.g., DLT) from the clinical data
- Model-based: Assign patients to dose levels based on estimation of the target toxicity level using a model of the dose-toxicity relationship
- Model-assisted: Use a model for efficient decision-making, but specify dose escalation & de-escalation rules before trial starts

Le Tourneau et al. J Natl Cancer Inst 2009; 101:708-20
Zhou et al. Clin Cancer Res 2018; 24(18):4357-4364

# Rule-based designs

- No prior assumption of dose-toxicity curve
- Up-and-down designs
  - Escalate or de-escalate dose with diminishing fractions of preceding dose depending on presence or absence of severe toxicity in previous dose cohort
  - Simple up-and-down design converges to  dose with 50% probability of severe toxicity
  - Not used much because they risk exposing patients to unacceptable levels of toxicity

# Traditional 3+3 design

- Most common design for phase I cancer trials
- Only assumption: toxicity increases with dose
- Rules
  - Start with a cohort of 3 patients at a dose considered safe based on animal studies
  - If none experiences a DLT, treat another 3 patients at the next dose level
  - If one of the first 3 patients experiences a DLT, treat another 3 patients at the same dose level
  - Dose escalation continues until 2 of 3-6 patients experiences a DLT
- Recommended phase II dose is the dose below this toxic dose level

# Example: ASP3026 in patients with advanced solid tumors

- Oral anaplastic lymphoma kinase (ALK) inhibitor
- First-in-human
- Open-label
- Multi-center
- Dose escalation
- Dose expansion
- Objectives: Determine
  - Primary: Safety & tolerability (MTD)
  - Secondary: Pharmacokinetics & antitumor effects

Li et al. J Hematol Oncol 2016 9(1):23

# Dose Escalation & Expansion

- 28-day treatment cycles, continuous dosing
- First patient in each dose cohort evaluated for DLTs on cycle 1, day 4. If no DLTs, rest of cohort enrolled
- 8 dose levels: 25, 50, 75, 125, 200, 325, 525*, 800 mg
- Expansion of higher dose cohorts allowed to include crizotinib-refractory ALK-positive patients (target population)
- Dose expansion at MTD: crizotinib-refractory ALK-positive patients

*MTD

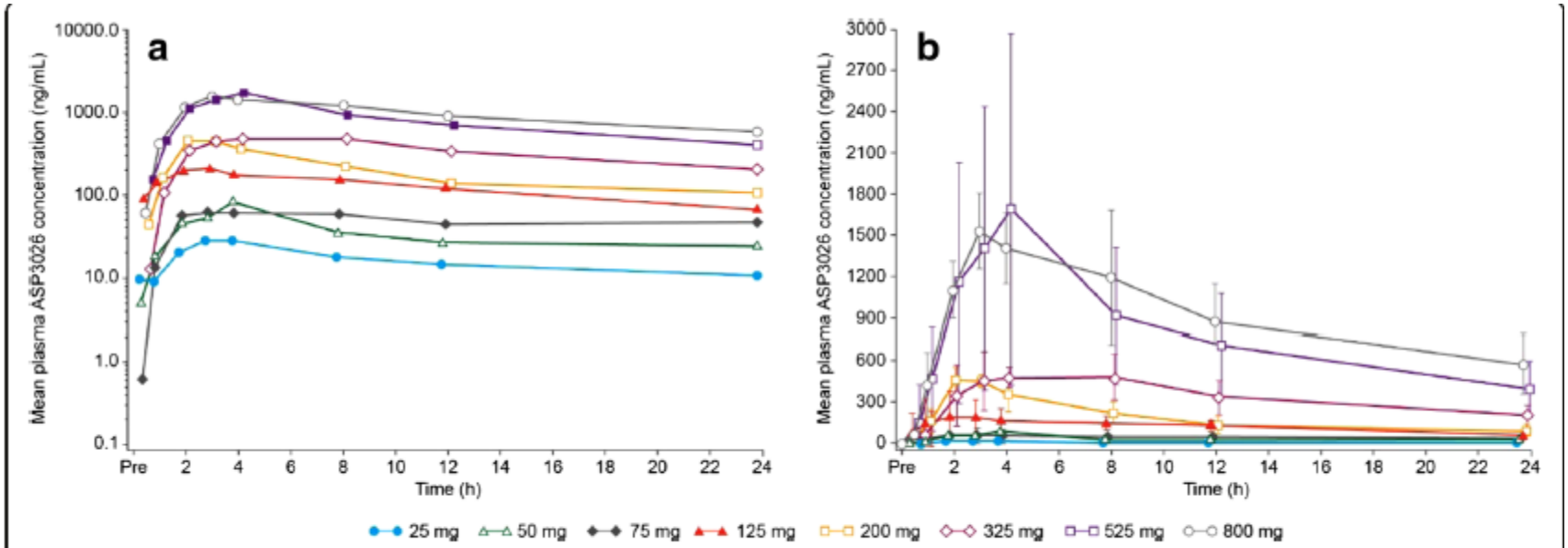



**Fig. 1** Mean plasma concentration of ASP3026, cycle 1, day 1. **a** Semi-log plot. **b** Linear plot. For patient numbers at each dose, refer to Tables 3 and 4

J Hematol Oncol 2016 9(1):23

# Traditional 3+3 design

- Advantages
  - Simple and safe
  - Provides info on inter-patient PK variability
- Disadvantages
  - Too many escalation steps
  - Few patients get therapeutic doses

# Accelerated titration design

- Cohorts of one new patient per dose level start at the lowest dose level
- Intra-patient dose escalation is allowed
- Reverts to 3+3 design if one DLT or 2 moderate toxicities are observed
- Advantage: more patients treated at therapeutic doses
- Disadvantage: intra-patient dose escalation may mask cumulative effects of treatments

# Pharmacologically guided dose escalation

- Assumes DLTs can be predicted by plasma drug concentrations based on animal data
- Not widely used

# Summary of rule-based designs

- Easy to implement, but
- Inefficient in establishing MTD
- Only use information from last dose
- Are widely used

# Model-based designs

- Continual reassessment method (CRM)
  - First Bayesian model-based method
  - Requires initial estimate of the slope of the dose-toxicity curve
  - This estimate is adjusted based on observed data
  - Estimated probability of DLT is updated for each patient

# Model-based designs

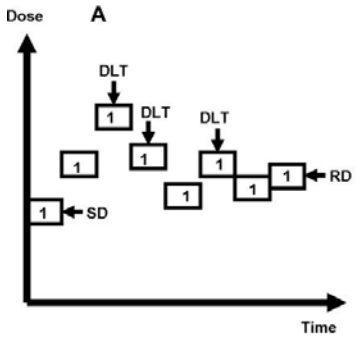
- Escalation with overdose control (EWOC)
  - Modification of CRM
  - Probability of exceeding MTD assessed after each patient
  - Stop dose escalation if probability of exceeding MTD gets too high
- Bayesian logistic regression (BLRM)
  - Modification of CRM
  - “Optimal” dose has highest posterior probability of being within the proper dosing interval, i.e., with probability of DLT within specified limits
  - Has overdose control similar to EWOC

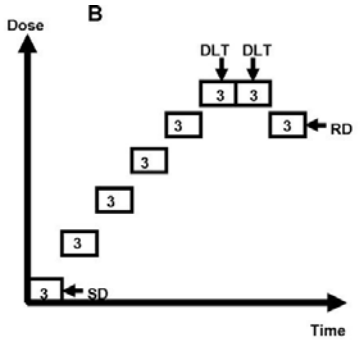
# Summary of model-based designs

- Use all data accumulated during trial
- Efficient, but
- Difficult to implement
- May fail to reach recommended phase II dose if initial estimate of dose-toxicity curve slope is wrong

# Graphical depiction of dose escalation methods for phase I cancer clinical trials.

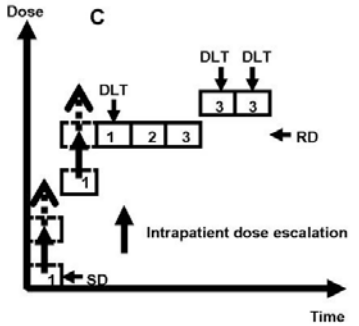
A: Simple up-and-down design

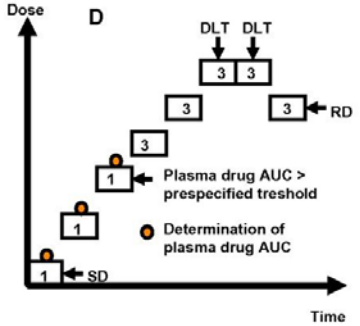



B: Traditional 3+3 design

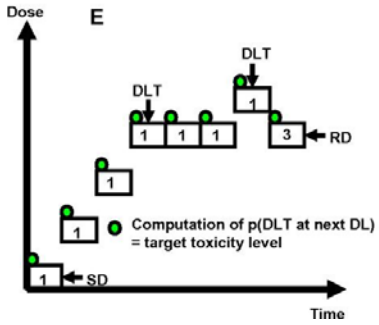
C: Accelerated titration design

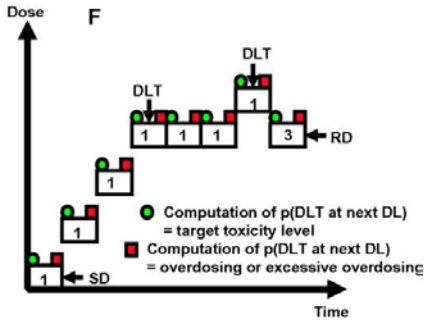



D: Pharmacologically guided dose escalation

E: Modified continual reassessment method (CRM)




F: Escalation with overdose control (EWOC)

**Le Tourneau C et al. JNCI J Natl Cancer Inst 2009;101:708-720**

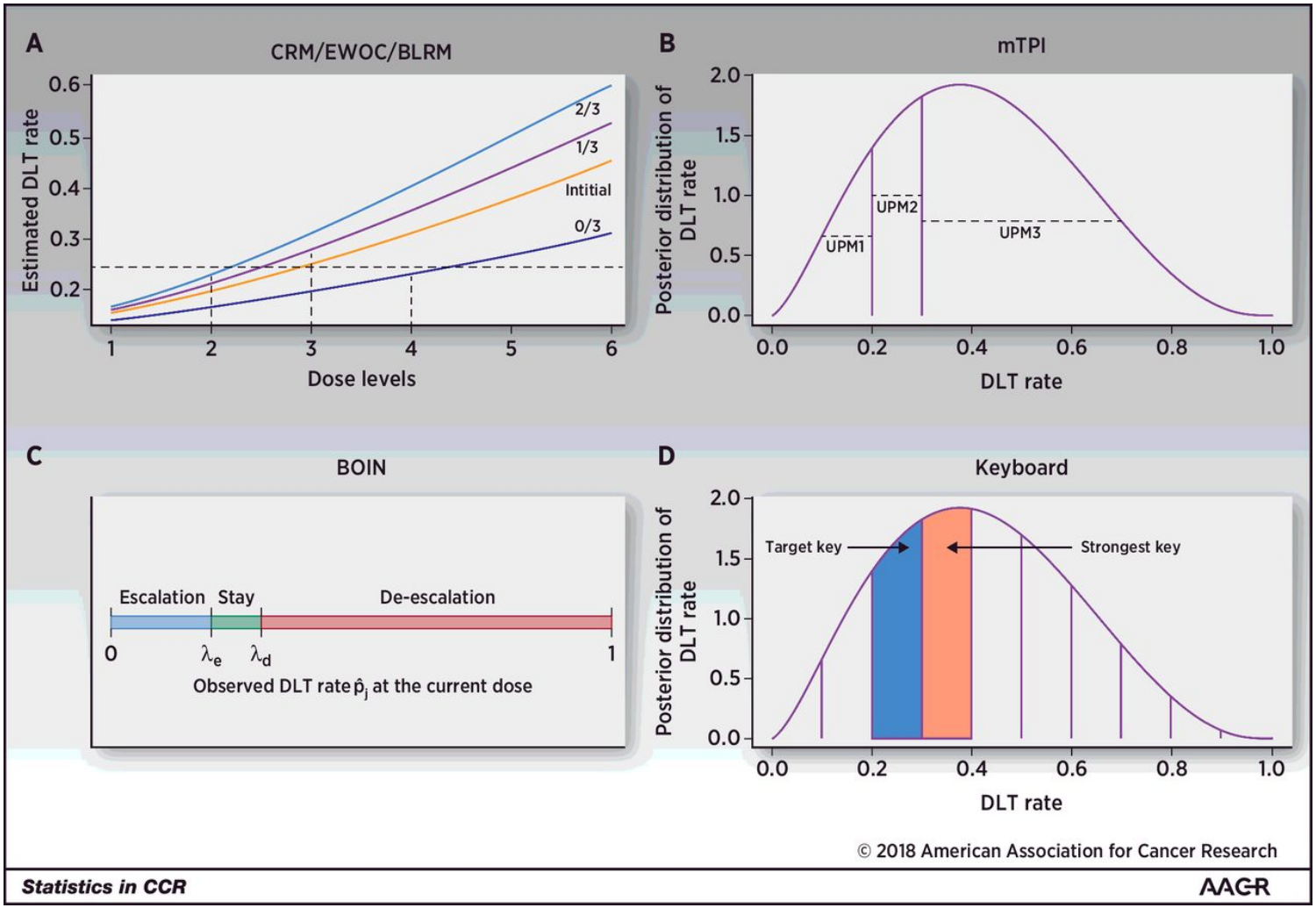


JNCI

# Model-assisted designs

- Modified toxicity probability interval (mTPI)
  - Specifies 3 intervals: proper dosing interval, underdosing interval, overdosing interval
  - Escalate, de-escalate or stay at same dose based on posterior distribution of the DLT rate in the intervals at current dose
- Keyboard design
  - Similar to mTPI, but has several intervals of equal length (keys)
- Bayesian optimal interval (BOIN)
  - Compare observed DLT rate with pre-determined dose escalation & de-escalation boundaries, which are derived from pre-specified toxicity probability thresholds

# Decision of dose escalation and de-escalation under the CRM/EWOC/BLRM, mTPI, BOIN, and keyboard designs.

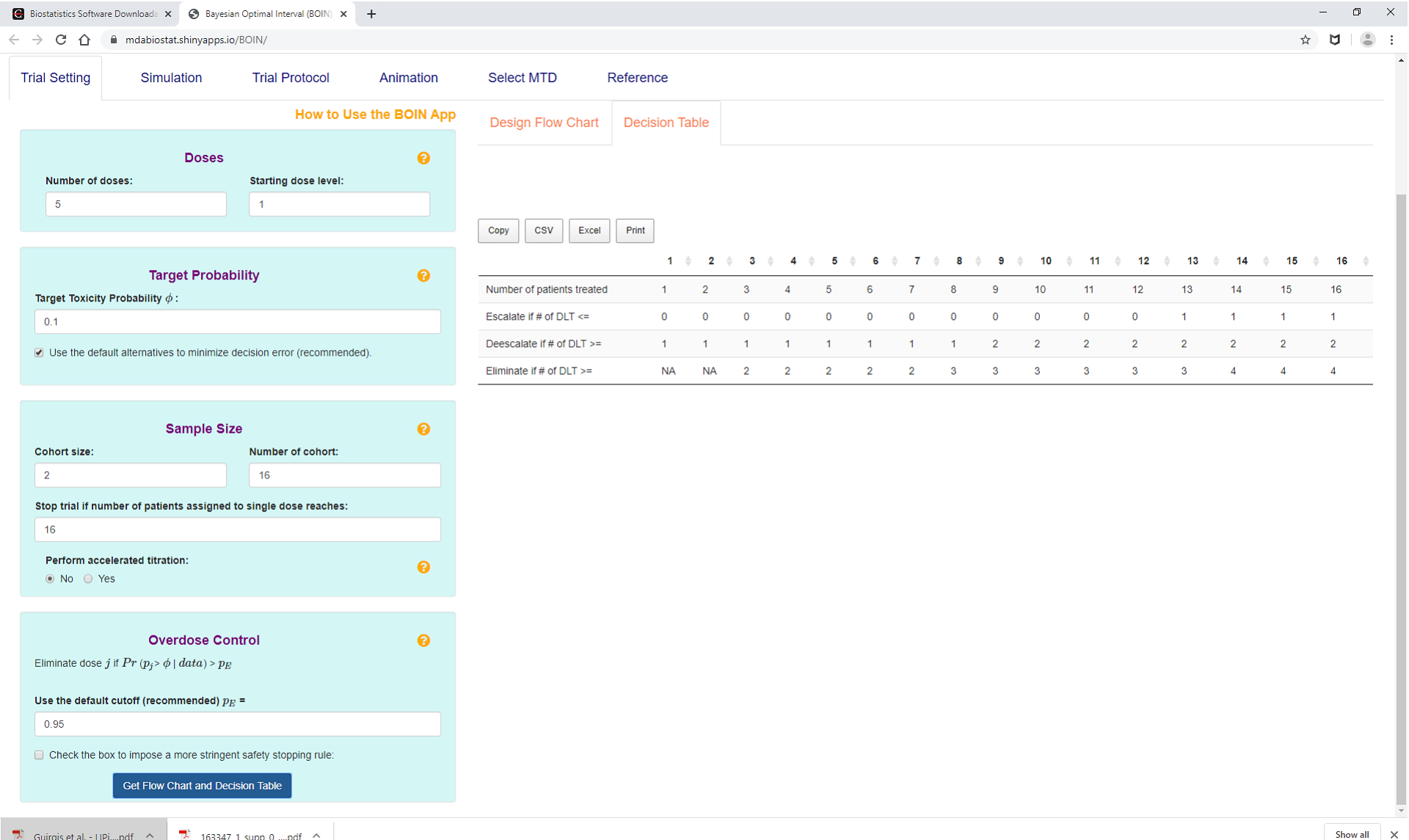



Heng Zhou et al. Clin Cancer Res 2018;24:4357-4364



Clinical Cancer Research
AACR American Association for Cancer Research

# Study designs: https://biostatistics.mdanderson.org/SoftwareDownload/SoftwareOnline




| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number of patients treated | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| Escalate if # of DLT <= | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| Deescalate if # of DLT >= | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| Eliminate if # of DLT >= | NA | NA | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 |

# Summary of model-assisted designs

- Combine superior performance of model-based designs with simplicity of rule-based designs
- BOIN and keyboard have similar performance and are easy to implement
- BOIN may be particularly appealing because it uses the observed DLT rate to determine dose escalation & de-escalation

Zhou et al. Clin Cancer Res 2018; 24(18):4357-4364

# Phase II clinical trials

- Phase II (NIH definition): Study the biomedical or behavioral intervention in a larger group of people (several hundred) to determine efficacy and further evaluate safety.
  - Is there any biological activity?
  - May or may not have concurrent controls
  - May be shorter term with different outcome and more exclusion criteria than phase III trials
  - Phase IIA-evaluate dosing; phase IIB -determine effectiveness

# Phase II: Multi-stage designs

- Purpose
  - Identify drugs that are promising for further testing in a Phase III trial
  - Preliminary efficacy assessment
  - Avoid exposing patients to sub-therapeutic dose levels
  - Terminate the study if the treatment is ineffective

# Single arm trials

- Optimal two-stage designs
  - Permit early stopping if there is a moderately long sequence of initial failures
  - Enroll $n_1$ patients in stage 1
  - If $\leq r_1$ responses, stop the trial
  - Otherwise, enroll $n_2$ more patients
  - Decide whether or not treatment is promising based on the $n_1+n_2$ patients

# Two-stage designs

- Null hypothesis: probability of response is unacceptably low
- Alternative hypothesis: probability of response is sufficiently high to warrant further study
- Simon's optimal two-stage design minimizes the expected sample size under the null hypothesis for the given error constraints
- Simon's minimax design minimizes the maximum sample size for the given error constraints

# Example: Intravenous aflibercept in patients with ovarian cancer

- Drug is a vascular endothelial growth factor (VEGF) inhibitor
- 2 dose levels tested (2 mg/kg and 4 mg/kg), based on previous phase 1 & 2 studies
- Patients with advanced platinum-resistant ovarian cancer
- Simon minimax 2-stage design
- Primary outcome: objective response rate (ORR)
- Null hypothesis: ORR $\leq$ 5%
- Alternative hypothesis: ORR $\geq$ 15%
- Tested at the 0.025 level, 1-sided

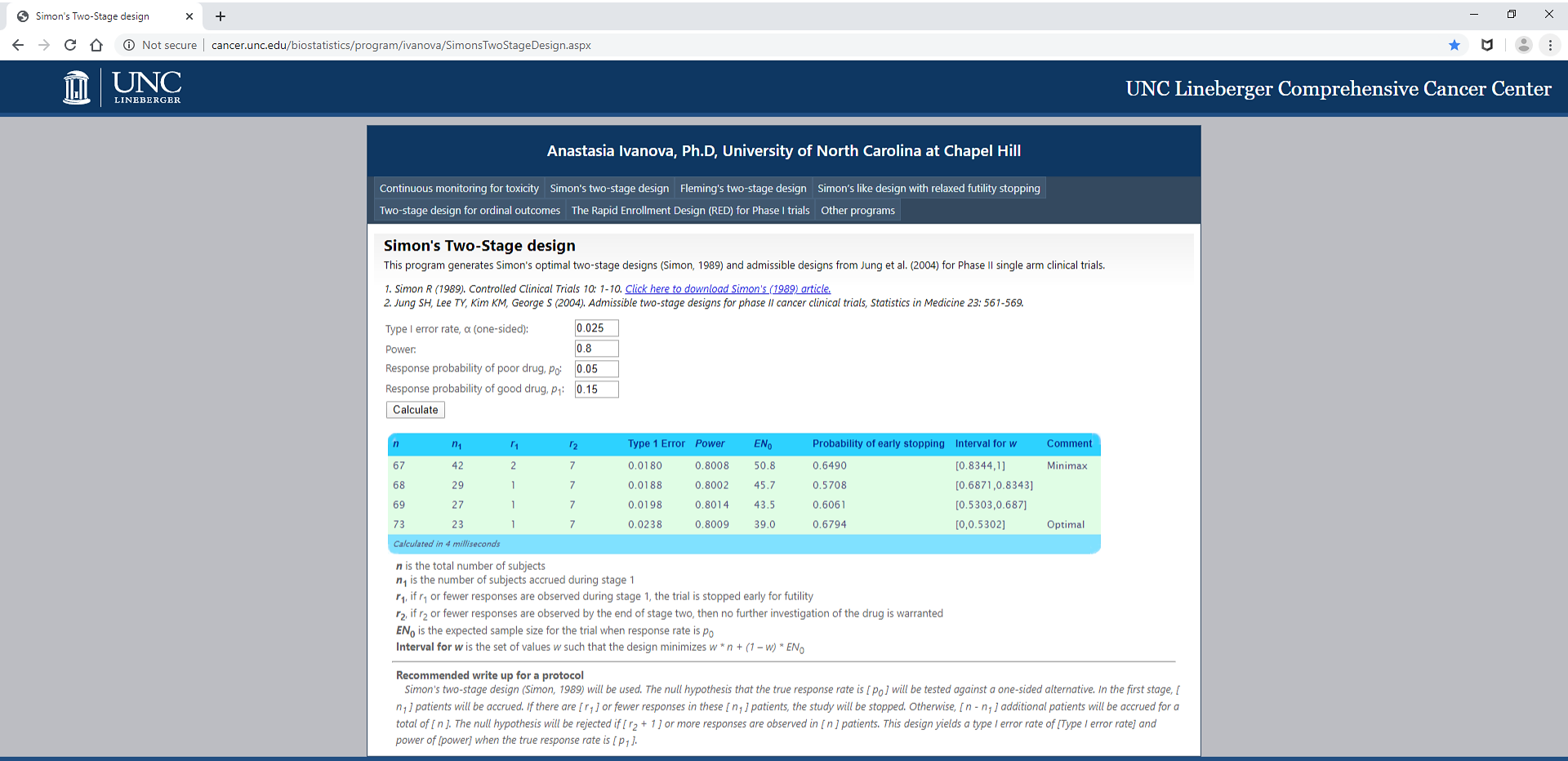
Tew et al. Cancer 2014; 120:335-43

# 2-stage design

- Plan: enroll 42 patients in each group in stage 1
- If at least 3 responders in stage 1in a group, go on to enroll 25 patients in stage 2
- Declare drug suitable for future study if at least 8 responders total (stages 1 & 2) in a group
- Allowed to enroll additional patients beyond the 2-stage design to reach a planned total sample size of 200

# Sample size calculation

http://cancer.unc.edu/biostatistics/program/ivanova/SimonsTwoStageDesign.aspx



# Multiple stage designs

- Can extend to 3 (or even 4 stages)
- May require at least one response at first stage to go on to the second stage
- Considerations for any multi-stage design
  - How long will it take to determine whether there are enough responses to proceed to the next stage?
  - Will we stop the study or keep on enrolling while waiting for the results from the previous stage?

# Randomized phase II designs

- May randomize patients to different drugs or dose levels of the same drug
- Can estimate differences between treatments
- Can pick the treatment with best response
- Randomization produces balanced groups

# Example: Phase II trial—Oncken (2006)

- Background: Evaluated 4 varenicline dose regimens for promoting smoking cessation.
- Methods: Multicenter, double-blind, placebo-controlled. Randomized healthy smokers aged 18-65 to varenicline tartrate or placebo twice daily for 12 weeks
  - 0.5 mg non-titrated (n=129); 0.5 mg titrated (n=130)
  - 1.0 mg non-titrated (n=129); 1.0 mg titrated (n=130)
  - placebo (n=129)

  with 40-week follow-up to assess long-term efficacy.

  Primary efficacy outcomes: carbon-monoxide confirmed 4-week continuous quit rates; continuous abstinence

Arch Intern Med. 2006 166(15):1571-7

# Results

- Weeks 9–12 continuous quit rates greater in 1.0 mg group and 0.5 mg group than placebo
- Weeks 9–52 abstinence rates greater in 1.0 mg group and 0.5 mg group than placebo
- Generally well tolerated
  - Nausea in 16%–42% of varenicline treated subjects
  - Less nausea with titrated dosing

From: **Efficacy and Safety of the Novel Selective Nicotinic Acetylcholine Receptor Partial Agonist, Varenicline, for Smoking Cessation**

Arch Intern Med. 2006;166(15):1571-1577. doi:10.1001/archinte.166.15.1571

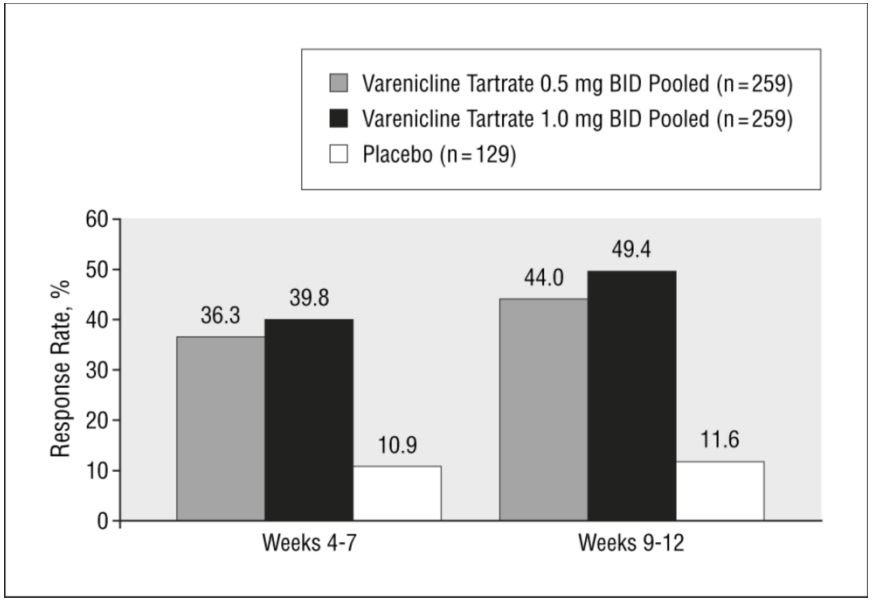



**Figure Legend:**

Continuous quit rates. P<.001 for each treatment group vs placebo. BID indicates twice daily. The odds ratios (ORs) and 95% confidence intervals (CIs) for the weeks 4 through 7 evaluation were 4.96 (95% CI, 2.66-9.22) for the 0.5-mg group and 5.86 (95% CI, 3.16-10.90) for the 1.0-mg group; for the weeks 9 through 12 evaluation, 6.32 (95% CI, 3.47-11.50) and 8.07 (95% CI, 4.42-14.70), respectively.

# Conclusion

- *Varenicline tartrate , 0.5 mg and 1.0 mg twice daily, is efficacious for smoking cessation.*

# Crossover Trial

- Definition (Chow & Liu): Modified randomized block design in which each block receives more than one treatment at different dosing periods.
- Simplest case: each participant is randomized to receive 2 treatments, A and B, in the order AB or BA.
- Between the 2 treatments, there is a washout period.

Design and Analysis of Clinical Trials (3rd Ed.) Chow & Liu, Wiley, 2014

# Crossover Trial

- Advantages
  - Each participant serves as his or her own control
  - Removes inter-patient variability from the comparison of treatments
  - Therefore, requires a smaller sample size than a parallel groups design
- Disadvantage
  - Have to worry about carryover between treatments
    - Carryover effects may not be equal
  - Vulnerable to dropouts

# Higher Order Crossover Designs

- Definition (Chow & Liu):
  - Number of periods > number of treatments
    - Two-sequence dual (extra period) design: ABB, BAA
    - Doubled (replicated) design: AABB, BBAA
  - Number of sequences > number of treatments
    - Balaam's design: AA, BB, AB, BA
  - Both
    - Four-sequence design: AABB, BBAA, ABBA, BAAB
- These designs allow estimation of carryover effects and intra-patient variability

# Crossover Trial

- Example: Randomized double blind trial of dark chocolate/cocoa snack vs. control snack in overweight people aged 40-64 (n=30)
- 2 periods, 4 weeks each, with 2-week washout period
- Outcomes: large & small blood vessel dilatation, peripheral blood flow, arterial stiffness
- Comparison: Active vs. control & baseline

West et al., British Journal of Nutrition 2014; 111:653-61

# Data Analysis

- Initial model
  - Fixed effects: treatment (baseline, active, control), period, treatment X period interaction
  - Random effect: participant
- Treatment X period was not statistically significant
- Some models included treatment X sex interaction
- Tukey's post-hoc tests for multiple comparisons

# Table 4: Results

| | Pre-treatment‡ | | Control§ | | Active§ | |
|---|---|---|---|---|---|---|
| | Mean | SE | Mean | SE | Mean | SE |
| Ultrasound measurements | | | | | | |
| Basal arterial diameter (mm)‖ | 4·20*** | 0·17 | 4·21*** | 0·17 | 4·47 | 0·17 |
| Peak arterial diameter (mm)‖ | 4·39*** | 0·18 | 4·42*** | 0·18 | 4·65 | 0·18 |
| FMD (% change) | 4·73 | 0·41 | 5·12 | 0·44 | 4·25 | 0·44 |
| Doppler-derived measures | | | | | | |
| Basal flow volume (ml/s)‖ | 166** | 18 | 176* | 18 | 214 | 18 |
| Peak flow volume (ml/s)¶ | 1059* | 76 | 1032* | 77 | 1153 | 77 |
| Reactive hyperaemia (% change)†† | 612* | 37 | 567 | 39 | 503 | 39 |
| EndoPAT variables | | | | | | |
| RHI | 2·26 | 0·14 | 2·19 | 0·12 | 2·20 | 0·11 |
| fRHI | 0·60 | 0·09 | 0·55 | 0·08 | 0·49 | 0·07 |
| AI‡‡ | 9·92** | 3·9 | 5·90** | 3·6 | −0·57 | 3·5 |
| AI at 75 bpm§§ | 2·75** | 3·9 | −2·72** | 3·6 | −8·53 | 3·5 |
| Anthropometrics | | | | | | |
| Weight (kg) | 80·9 | 2·3 | 80·7 | 2·3 | 81·3 | 2·3 |
| BMI (kg/m$^2$) | 27·4 | 0·5 | 27·5 | 0·5 | 27·7 | 0·5 |
| Waist circumference (cm) | 94·6 | 1·2 | 94·7 | 1·2 | 95·5 | 1·2 |
| Hip circumference (cm) | 106·8 | 0·9 | 106·9 | 0·9 | 106·9 | 0·9 |
| Waist:hip ratio | 0·89 | 0·01 | 0·89 | 0·01 | 0·89 | 0·01 |

Mean values were significantly different from those of the active group: * $P \leq 0 \cdot 05$, ** $P \leq 0 \cdot 01$, *** $P \leq 0 \cdot 001$.

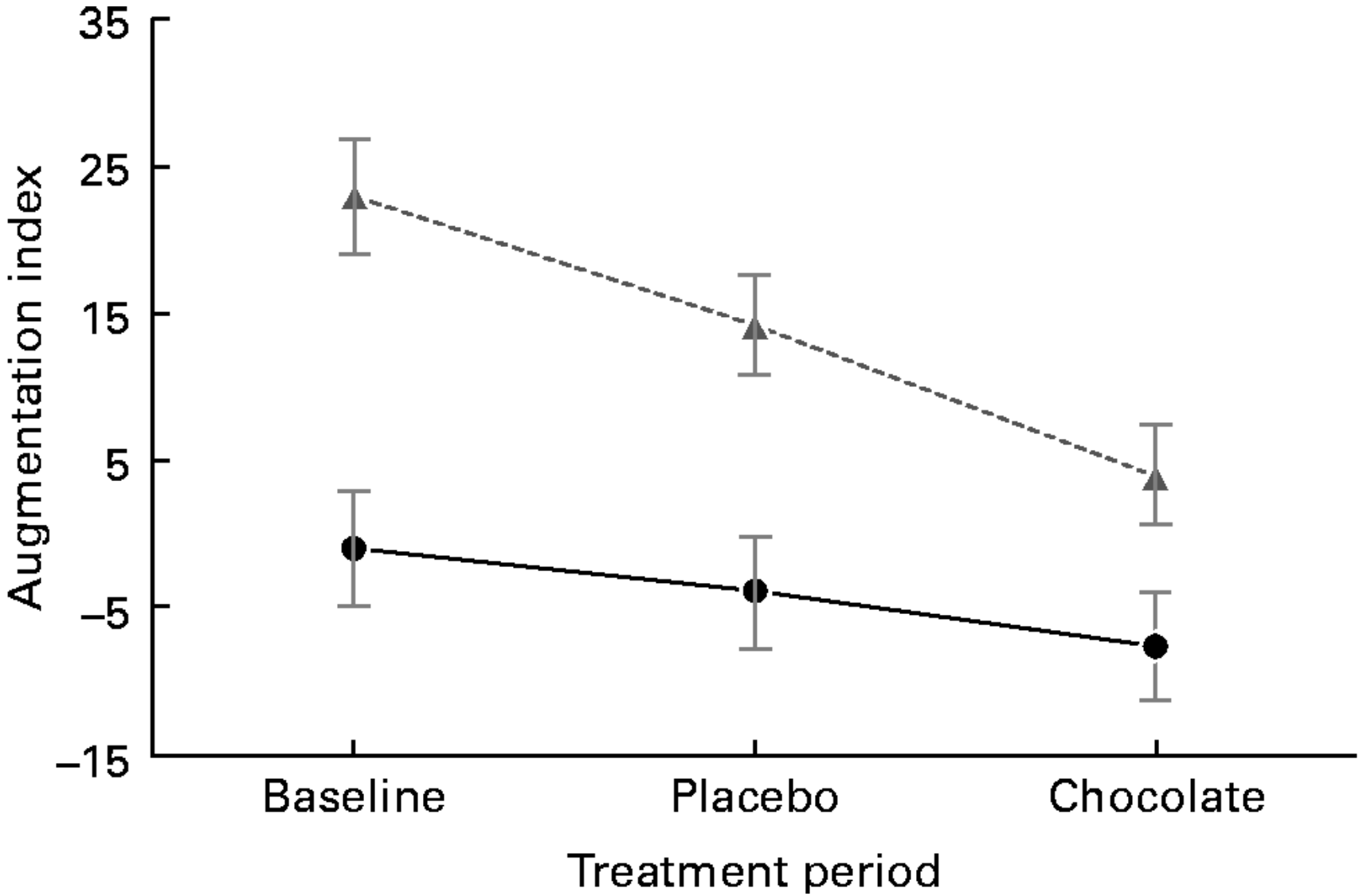



Fig. 1 Sex difference in vascular response to the cocoa+dark chocolate treatment. Women () exhibited significant reductions in the augmentation index, whereas men () did not (sex × treatment interaction, P= 0·01).

# 2-Period 2-Treatment Crossover Trial: Outcome by Sequence & Period

| Sequence | Period 1 | Period 2 |
|---|---|---|
| AB | $Y_A$ | $Y_B$ |
| BA | $Y_B$ | $Y_A$ |

# Simplifying Assumptions

- $H_0$: $\mu_B = \mu_A$; $H_a$: $\mu_B \neq \mu_A$
- Specify $\mu_B - \mu_A = \delta$
  (difference in treatment effects)
- No sequence or period effect: paired t-test comparing treatment B with treatment A over the entire sample
  - Specify SD$=\sqrt{2}$*(within-person SD)$=$SD$(Y_B - Y_A)$
  - Or specify SD$(Y_B)$, SD$(Y_A)$, and corr$(Y_A, Y_B)$

https://stattools.crab.org

STATISTICAL TOOLS | DESIGN | ANALYSIS | PROBABILITIES | ABOUT US

# One Arm Normal

One Arm Normal is a program to calculate either estimates of sample size or power for one sample normal problem.

| User Input | Program Output |
|---|---|

## Select Calculation and Test Type

| | |
|---|---|
| (•) Sample Size<br>( ) Power | ( ) 1 Sided<br>(•) 2 Sided |

## Select Hypothesis Test Parameters

| Null Mean | Alternative Mean | Standard Deviation | Alpha |
|---|---|---|---|
| 0 | 1 | 1.414 | 0.05 |

| Power | Sample Size |
|---|---|
| 0.9 | 22 |

Calculate

Within-person SD=1

Help Document

# Crossover Trial vs. Parallel Group Sample Size

- For a given
  - difference in treatment mean responses $\mu_B - \mu_A = \delta$
  - treatment response variance Var(Y)
    - (between-person plus within-person)
  - levels of type I & II error

$$\frac{\text{n crossover}}{\text{n parallel}} = 0.5*[1\text{-corr}(Y_B, Y_A)]$$

  - Even if there is no within-person correlation, the crossover trial requires half the sample size
  - The greater the correlation, the greater the reduction in sample size

# Selecting a Design

- Need to consider (Chow & Liu)
  - Number of treatments to be compared
  - Characteristics of the treatment
  - Study objectives
  - Availability of participants
  - Inter- and intra-person variability
  - Duration of the study
  - Dropout rates

# Considerations

- If intra-patient variability ≥inter-patient variability, parallel groups preferred to crossover
- If inter-patient variability is large and the number of treatments is small, consider a cross-over design
  - However, disease state must be stable