UCDAVIS HEALTH

Design of Phase II Clinical Trials

CLINICAL AND TRANSLATIONAL SCIENCE CENTER

Susan Stewart, Ph.D.

Division of Biostatistics

The UC Davis CTSC receives support from the NIH National Center for Advancing Translational Sciences (award TR001860).

Topics

- Objectives
- Types
  - Multi-stage
  - Randomized
  - Platform
  - Crossover

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 ≤ 5%
- Alternative hypothesis: ORR ≥ 15%
- Tested at the 0.025 level, 1-sided

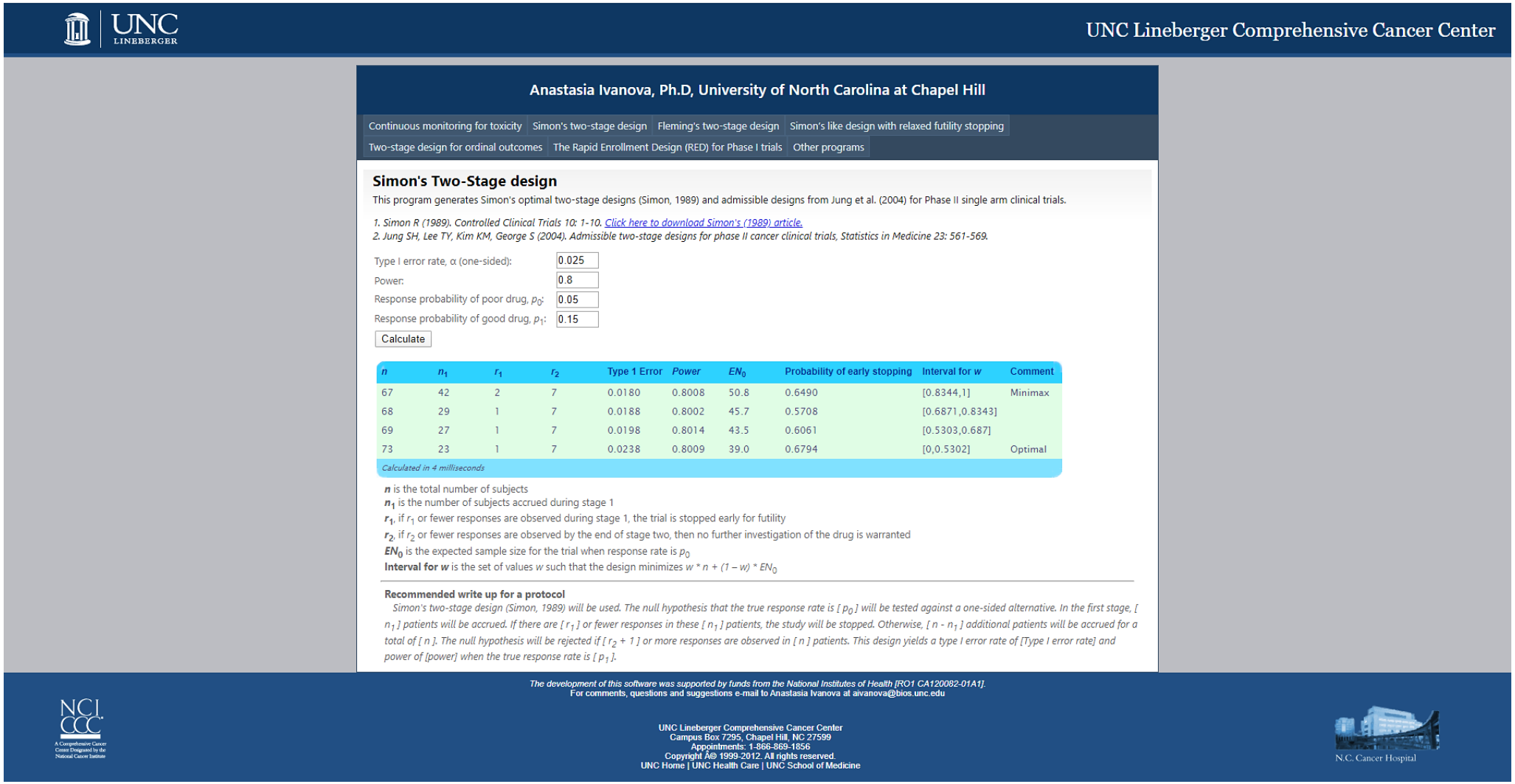
Tew et al. Cancer 2014; 120:335-43

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



UNC LINEBERGER — UNC Lineberger Comprehensive Cancer Center

Anastasia Ivanova, Ph.D, University of North Carolina at Chapel Hill

Continuous monitoring for toxicity | Simon's two-stage design | Fleming's two-stage design | Simon's like design with relaxed futility stopping | Two-stage design for ordinal outcomes | The Rapid Enrollment Design (RED) for Phase I trials | Other programs

Simon's Two-Stage design

This program generates Simon's optimal two-stage designs (Simon, 1989) and admissible designs from Jung et al. (2004) for Phase II single arm clinical trials.

1. Simon R (1989). Controlled Clinical Trials 10: 1-10. Click here to download Simon's (1989) article.
2. Jung SH, Lee TY, Kim KM, George S (2004). Admissible two-stage designs for phase II cancer clinical trials, Statistics in Medicine 23: 561-569.

Type I error rate, α (one-sided): 0.025
Power: 0.8
Response probability of poor drug, p_0: 0.05
Response probability of good drug, p_1: 0.15
Calculate

n	n_1	r_1	r_2	Type 1 Error	*Power*	EN_0	Probability of early stopping	Interval for w	Comment
67	42	2	7	0.0180	0.8008	50.8	0.6490	[0.8344,1]	Minimax
68	29	1	7	0.0188	0.8002	45.7	0.5708	[0.6871,0.8343]	
69	27	1	7	0.0198	0.8014	43.5	0.6061	[0.5303,0.687]	
73	23	1	7	0.0238	0.8009	39.0	0.6794	[0,0.5302]	Optimal

Calculated in 4 milliseconds

n is the total number of subjects
n_1 is the number of subjects accrued during stage 1
r_1, if r_1 or fewer responses are observed during stage 1, the trial is stopped early for futility
r_2, if r_2 or fewer responses are observed by the end of stage two, then no further investigation of the drug is warranted
EN_0 is the expected sample size for the trial when response rate is p_0
Interval for *w* is the set of values w such that the design minimizes $w * n + (1 - w) * EN_0$

Recommended write up for a protocol

Simon's two-stage design (Simon, 1989) will be used. The null hypothesis that the true response rate is [p_0] will be tested against a one-sided alternative. In the first stage, [n_1] patients will be accrued. If there are [r_1] or fewer responses in these [n_1] patients, the study will be stopped. Otherwise, [n - n_1] additional patients will be accrued for a total of [n]. The null hypothesis will be rejected if [r_2 + 1] or more responses are observed in [n] patients. This design yields a type I error rate of [Type I error rate] and power of [power] when the true response rate is [p_1].

The development of this software was supported by funds from the National Institutes of Health [RO1 CA120082-01A1].
For comments, questions and suggestions e-mail to Anastasia Ivanova at aivanova@bios.unc.edu

NCI CCC A Comprehensive Cancer Center Designated by the National Cancer Institute

UNC Lineberger Comprehensive Cancer Center
Campus Box 7295, Chapel Hill, NC 27599
Appointments: 1-866-869-1856

UNC Home | UNC Health Care | UNC School of Medicine

N.C. Cancer Hospital

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

Data Analysis

- Quit rates: binary
  - Compared each treatment group separately vs. placebo
  - Compared pooled dosage groups vs. placebo
  - Step-down procedure to account for multiple comparisons
  - Logistic regression
    - Independent variables: treatment and center
    - Computed odds ratios with 95% confidence intervals
- MNWS (withdrawal), mCEQ (cigarette evaluation): numeric
  - Analysis of covariance (ANCOVA)
    - Covariate: baseline level of outcome variable
    - Independent variables: treatment and center

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

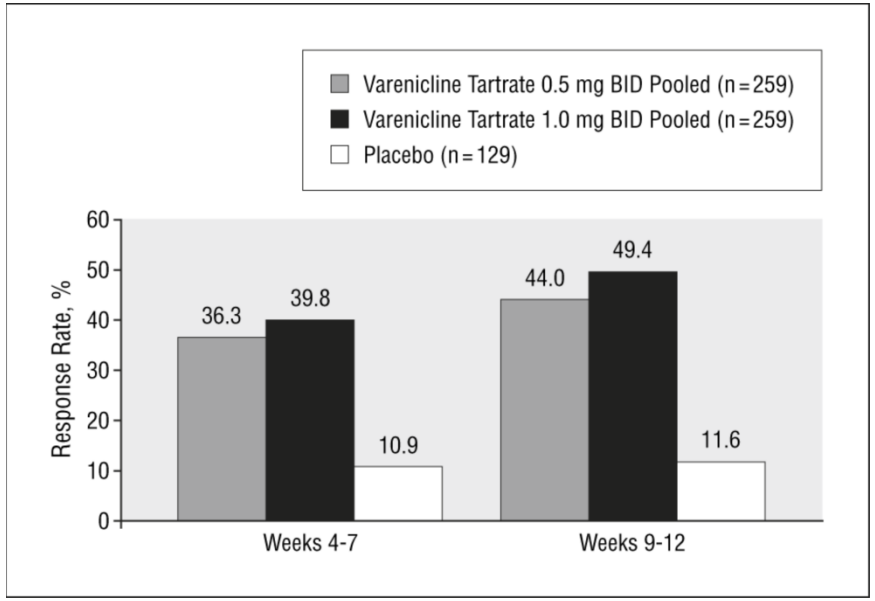


Figure Legend:

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.

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

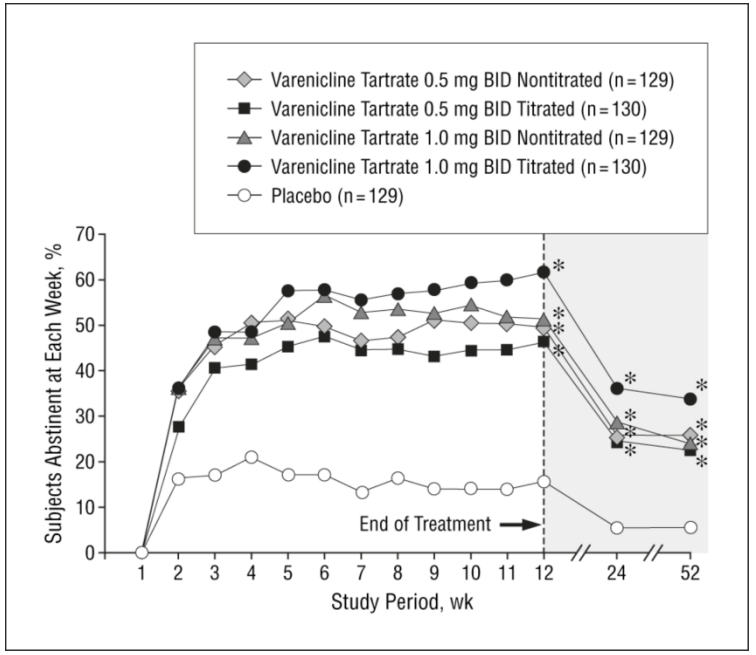


Figure Legend:

Carbon monoxide–confirmed weekly point prevalence abstinence rates. BID indicates twice daily. *P<.001 vs placebo.

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

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

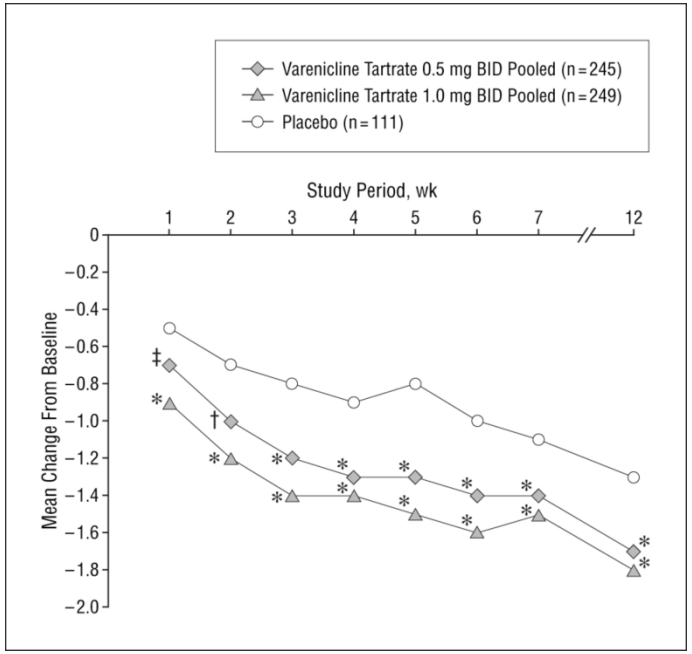


Figure Legend:

Mean changes in Minnesota Nicotine Withdrawal Scale “urge to smoke” scores from week 1 to week 12 for all subjects. BID indicates twice daily. In comparison with placebo, asterisk indicates P<.001; dagger, P<.01; and double dagger, P<.05.

Conclusion

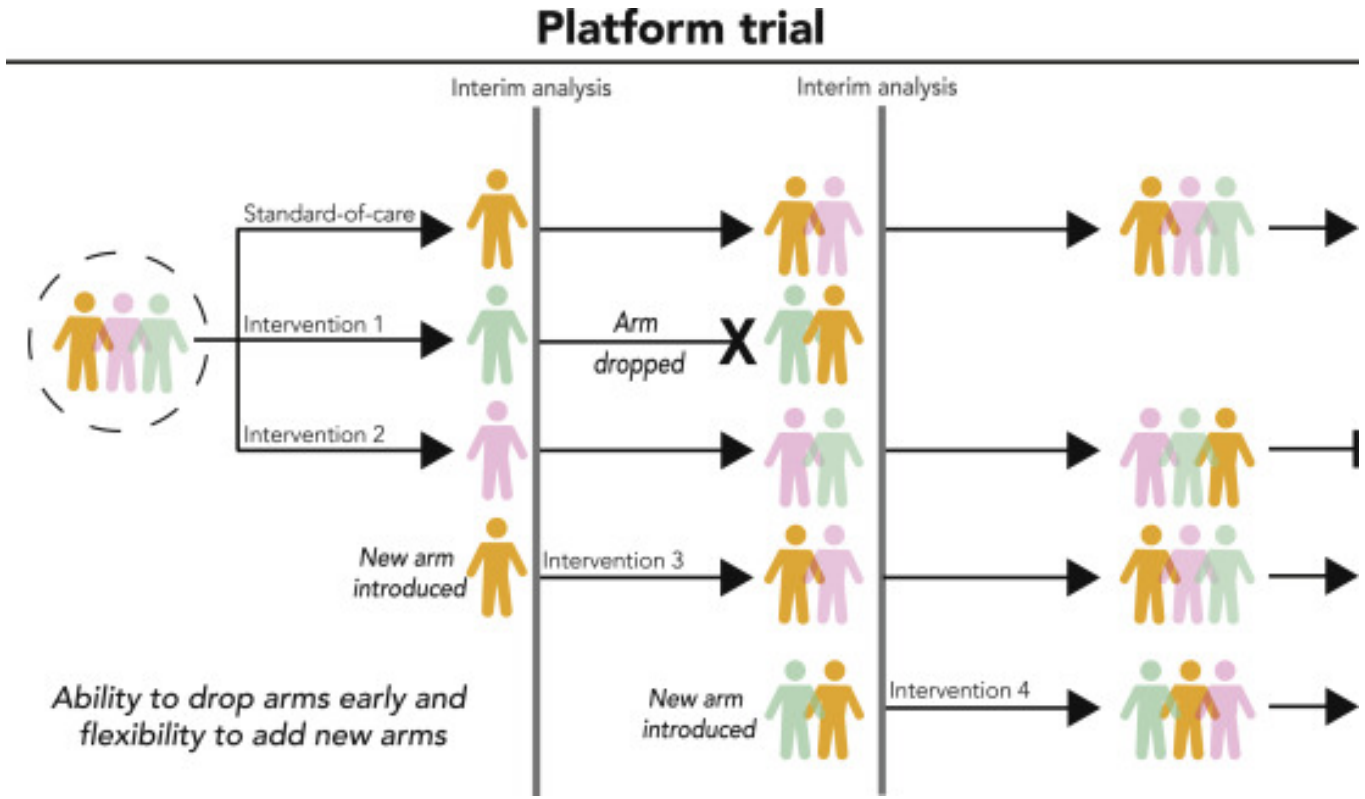
- Varenicline tartrate , 0.5 mg and 1.0 mg twice daily, is efficacious for smoking cessation.

Platform Trials

- Multiple treatments evaluated simultaneously
- Single master protocol
- Adaptive platform designs
  - Drop treatments for futility
  - Declare one or more treatments superior
  - Add new treatments
- Multi-arm, multi-stage
- More efficient than traditional RCT designs

Saville & Berry. Efficiencies of platform clinical trials: A vision of the future. Clin Trials. 2016 Jun;13(3):358-66.

Park et al. An overview of platform trials with a checklist for clinical readers. J Clin Epidemiol. 2020 Sep;125:1-8.



Park et al. An overview of platform trials with a checklist for clinical readers. J Clin Epidemiol. 2020 Sep;125:1-8.

Example: ACCORD Seamless Phase 2 Platform Study to Assess Multiple COVID-19 Treatments

- Objectives:
  - Stage 1 (screening stage): Evaluate safety and efficacy of candidate agents as add-on therapy to standard of care (SoC) in hospitalized patients
  - Stage 2 (expansion stage): Confirm efficacy of agents selected based on evidence from Stage 1
- Participants:
  - Hospitalized patients age ≥18 with Grade 3-5 COVID-19 in UK
- Main outcomes:
  - Time to sustained clinical improvement ≥2 points on WHO 9 point ordinal scale
  - Live discharge or fit for discharge (0-2 on WHO scale) by Day 29

Wilkinson et al. Trials (2020) 21:691

ACCORD trial (cont'd)

- Comparator and candidate interventions
  - Current SoC for COVID-19
  - Bemcentinib
    - Could reduce viral infection; blocks spike protein
  - MEDI3506
    - Anti-IL-33 monoclonal antibody; could treat respiratory failure
  - Acalabrutinib
    - BTK inhibitor; anti-viral and anti-inflammatory
  - Zilucoplan
    - Complement C5 inhibitor; may block severe inflammatory response
  - Nebulized heparin
    - Binds with spike protein
  - Others TBD

ACCORD trial (cont'd)

- Randomization
  - Stratified by baseline severity grade
  - Equal allocation to each experimental arm and contemporaneous SoC arm
  - May be changed to 2:1 in favor of experimental arms
- Sample size per agent
  - Stage 1: 60
  - Stage 2: 126
  - Total: up to 1800

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≤ 0·05, ** P≤ 0·01, *** P≤ 0·001.

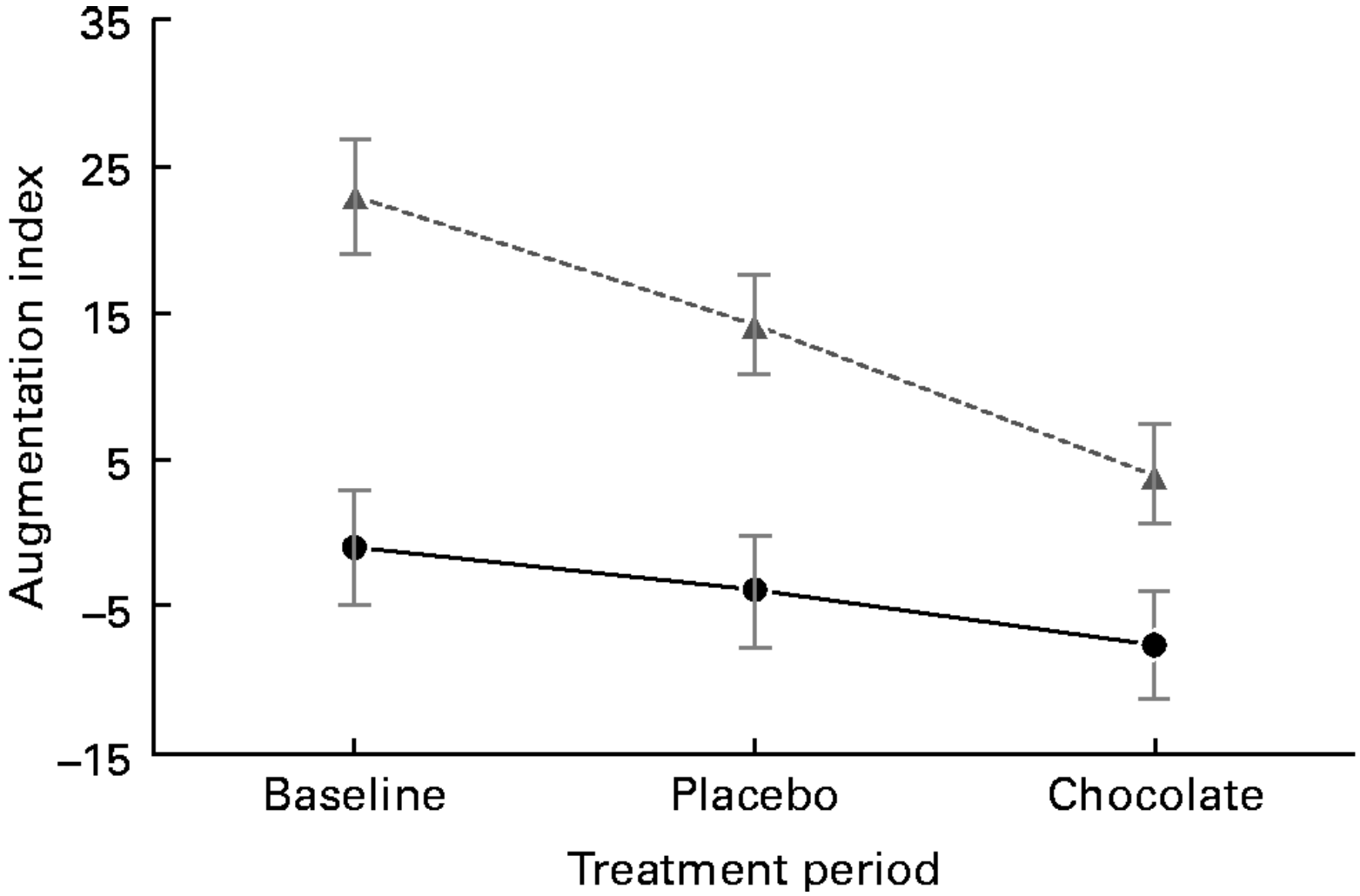


Fig. 1 Sex difference in vascular response to the cocoa+dark chocolate treatment. Women () exhibited significant reductions in the augmentation index, whereas men () did not (sex × treatment interaction, P= 0·01).

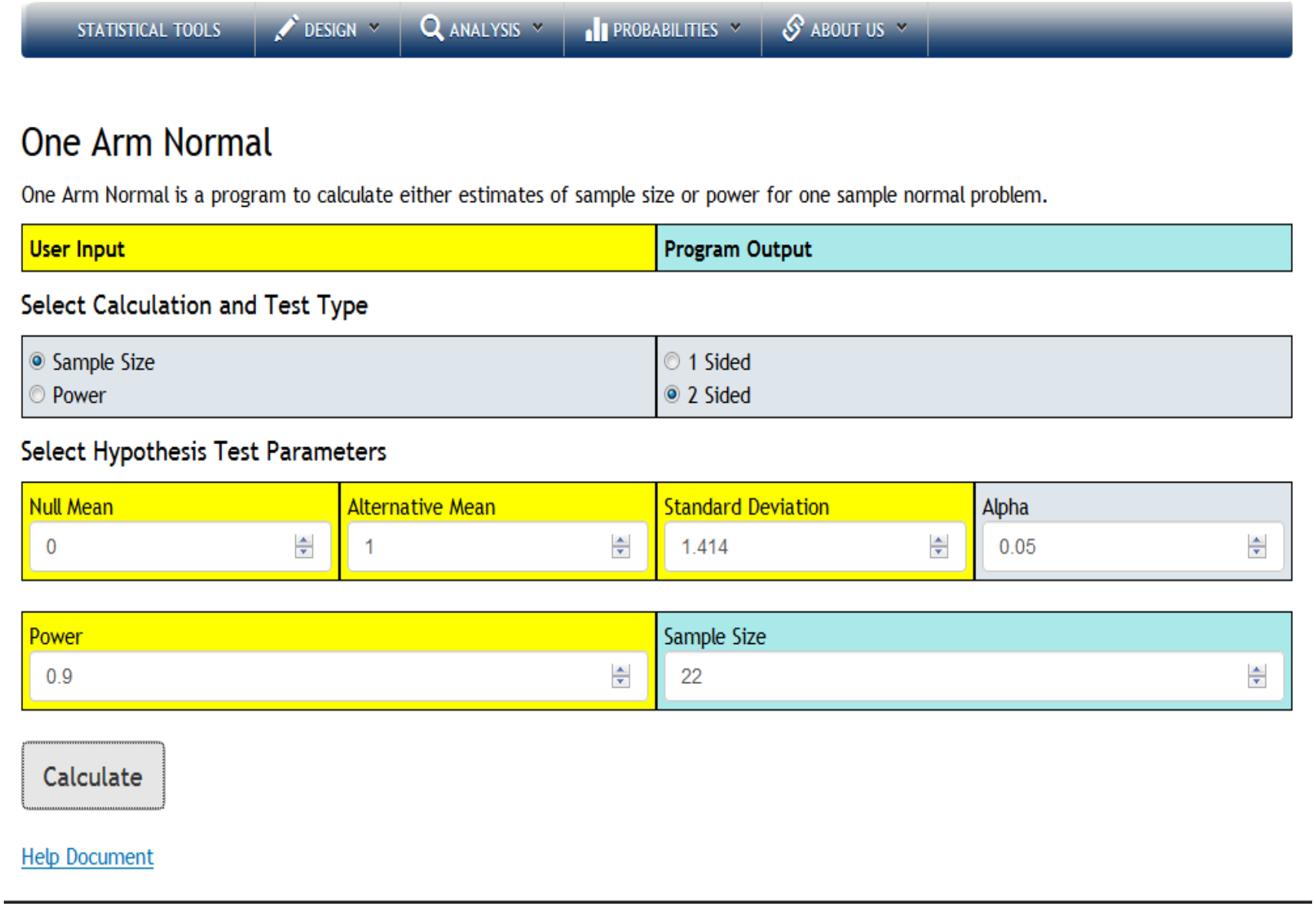
2-Period 2-Treatment Crossover Trial: Outcome by Sequence & Period

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



Crossover Trial vs. Parallel Group Sample Size

- For a given
  - difference in treatment mean responses μ_B-μ_A=δ
  - 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

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

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