Statistical Methods for Dynamic Treatment Regimens and Sequential Multiple Assignment Randomized Trial

A Webinar from ASA Biopharmaceutical Section

Abdus S. Wahed (Biostatistics) and Yu Cheng (Statistics)
University of Pittsburgh

Learning objectives : to understand

- Basic concept of dynamic treatment regimes (DTR)
- Various methods for inference related to DTR
- Sequential multiple assignment randomized trials (SMART)
- Inverse-probability-weighting
- G-Computation
- Design issues, guidelines and power

Outline: how the objectives will be achieved

- Define DTR and related framework
- Discuss assumptions related to inference
- Motivate the methods of inference
- Analyze a SMART trial
- Discuss design guidelines

Dynamic Treatment Regimes

4

A general introduction

Motivation

- Drug Development & Approval Process
  - Compare A (new) vs. B (standard)
    - A is better (less or equivalent toxicity, better short-term efficacy) than B
  - Compare C (new) vs. B (standard)
    - C is better (less or equivalent toxicity, better short-term efficacy) than B
  - (May be) Compare A vs. C

Motivation

- Drug Use
  - Patient diagnosed with the disease
    - Physician has to make a decision – put patient on A or C
    - Observe response to the prescribed treatment over time
    - Stop treatment, continue treatment with/without dose modification, switch treatment

Motivation

- Decision process
  - Not all treatment works for all patients, e.g.
    - Treatment A works for patients if the disease is diagnosed early;
    - Treatment C is shown to be less effective in patients with history of diabetes
  - Covariate Treatment Interaction

Motivation

- Decision process
  - How long you should keep a patient on a treatment before you stop, modify, or switch?
    - Depends on
      - Adverse events
      - Intermediate response markers
      - Long-term effects
      - Options to switch to

Motivation

- Decision process
  - How to make a decision to stop, modify, or switch? When?
  - If a decision to modify treatment is taken, what the modification should be?
  - If a decision to switch the treatment is taken, what treatment should be switched to?

Definition

- Dynamic Treatment Regime
  - A set of specific rules to make decisions at each decision point of the therapy
  - Also known as adaptive treatment regime, adaptive treatment strategy
  - Example: “If the patient is a Caucasian female, age 50 or over, have normal HGB levels, (bla bla bla ..), start the patient on therapy A, observe for 4 weeks (?), if it seems to be working (?), continue A, if not, if PC > 130000 switch to B, if PC<130000, switch to C, observe for another 4 weeks(?)…… ”

Goal

- Dynamic Treatment Regime
  - Find the best treatment regime to best manage a disease, or
  - Compare several dynamic treatment regimes

Problem

- Curse of dimensionality
  - Theoretically, infinitely many treatment regimes are possible

"If the patient is a Caucacian female, age 50 or over, have normal HGB levels, (bla bla bla ..), start the patient on therapy A, observed for 4 weeks (?), if it seems to be working (?), continue A, if not, if PC > 130000 switch to B, if PC<130000, switch to C, observe for another 4 weeks(?)…… "

Solution and Issues

- Screen Candidate Regimes from Observational Data
  - How?
- Run a clinical trial
  - How?
- Combine clinical trials?

Important note:

Adaptive/Dynamic Treatment Regimes

$\neq$

Adaptive Design

Adaptive Treatment Regime vs. Adaptive Design

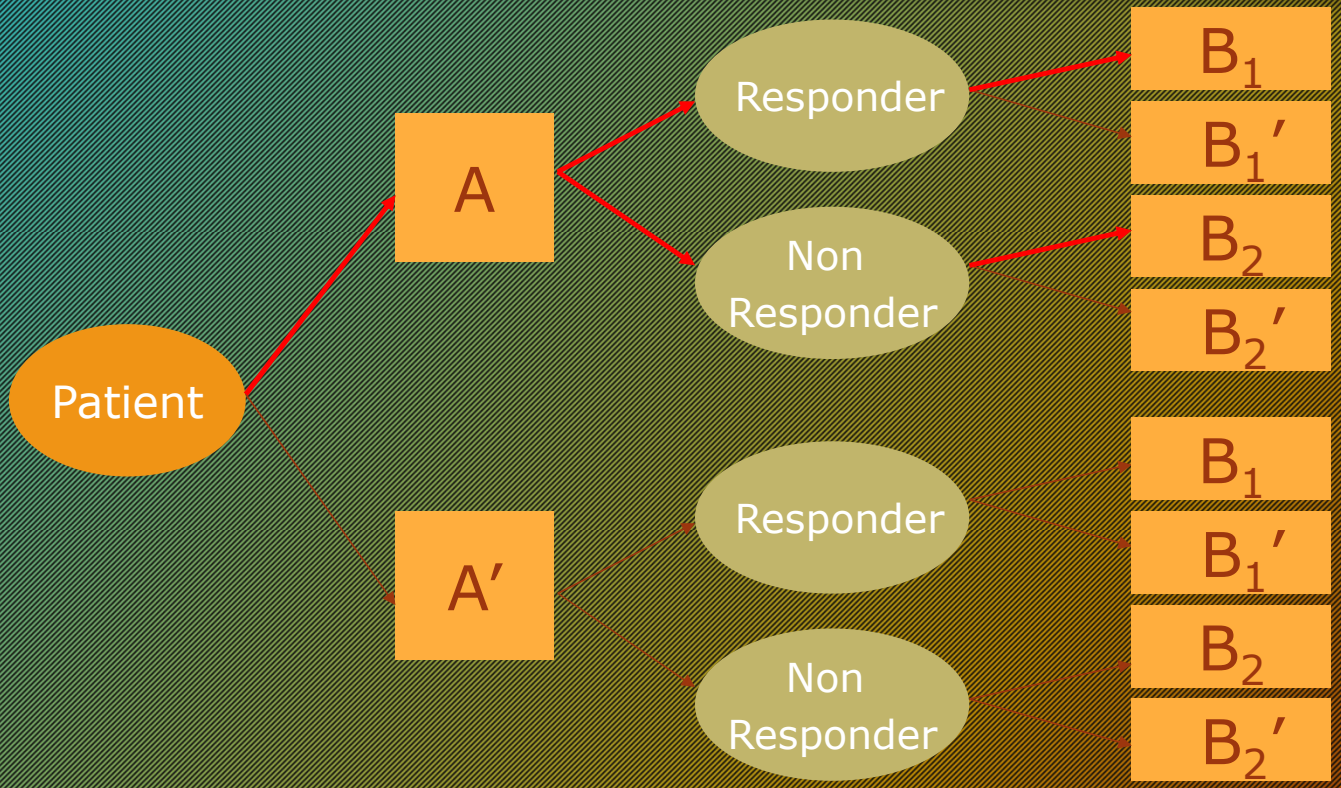
- Adaptive Treatment Regimes

  "...adaptive as used here refers to a time-varying therapy for managing a chronic illness" (Murphy,2005)

- Adaptive Design

  "...such as designs in which treatment allocation probabilities for the present patients depend on the responses of past patients" (Murphy,2005)

Example


Patient
A
A′
Responder
Non Responder
Responder
Non Responder
B1
B1′
B2
B2′
B1
B1′
B2
B2′

8 Possible Regimes

(1) Trt with A followed by B_1 if response, else B_2 (AB_1B_2)

(2) Trt with A followed by B_1 if response, else B_2' (AB_1B_2')

(3) Trt with A followed by B_1' if response, else B_2 (AB_1'B_2)

(4) Trt with A followed by B_1' if response, else B_2' (AB_1'B_2')

(5) Trt with A' followed by B_1 if response, else B_2 (A'B_1B_2)

(6) Trt with A' followed by B_1 if response, else B_2' (A'B_1B_2')

(7) Trt with A' followed by B_1' if response, else B_2 (A'B_1'B_2)

(8) Trt with A' followed by B_1' if response, else B_2' (A'B_1'B_2')

What is the objective of constructing Dynamic Treatment Regimes?

A treatment naïve patient walks through the door…

What treatment regime should the patient be assigned to?

Let T(XYZ) (larger is better) be the outcome of an individual, had he/she received treatment according to regime XYZ

If we knew that

1. $E[T(AB_1B_2)] = 15$
2. $E[T(AB_1B_2')] = 14$
3. $E[T(AB_1'B_2)] = 18$
4. $E[T(AB_1'B_2')] = 17$
5. $E[T(A'B_1B_2)] = 20$
6. $E[T(A'B_1B_2')] = 19$
7. $E[T(A'B_1'B_2)] = 13$
8. $E[T(A'B_1'B_2')] = 12$

Best Regime for the patient

In Reality…

Problems:

E[T(.)]'s are not known

How can one accurately and efficiently estimate E[T(.)]?

How to estimate the expected outcome under different Regimes?

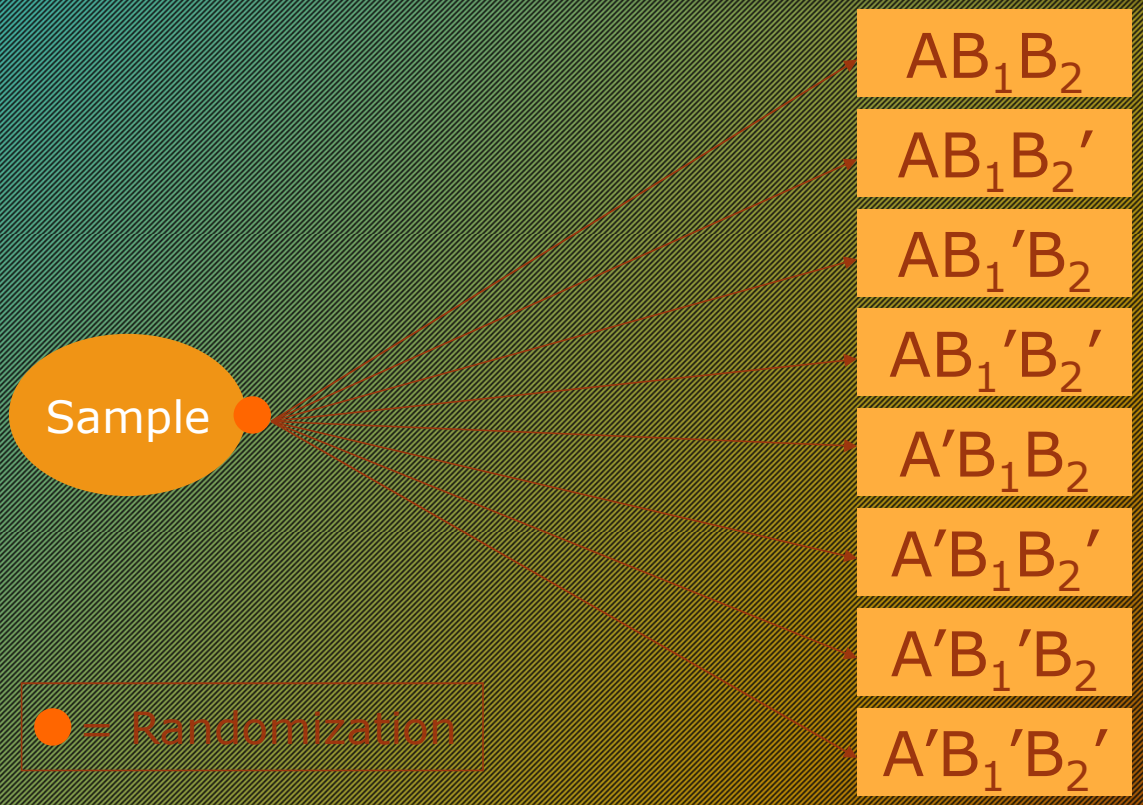
Three study designs:

1. A clinical trial with 8 arms

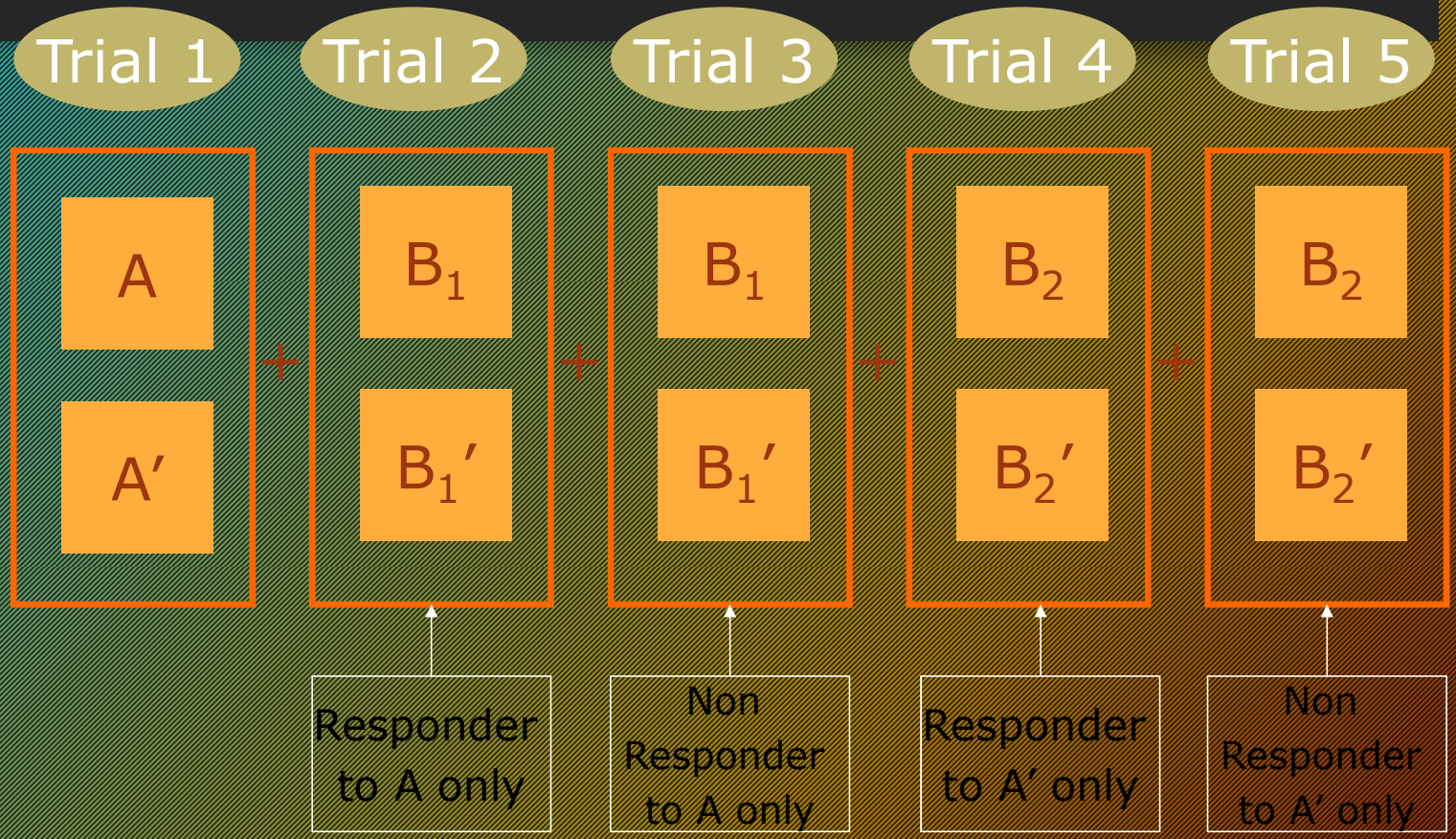
2. Combine 5 separate trials

3. SMART (Sequential Multiple Assignment Randomized Trials)

Design 1: A clinical trial with 8 Treatment Arms



Design 2: Combining Existing Trials



Design 3: SMART

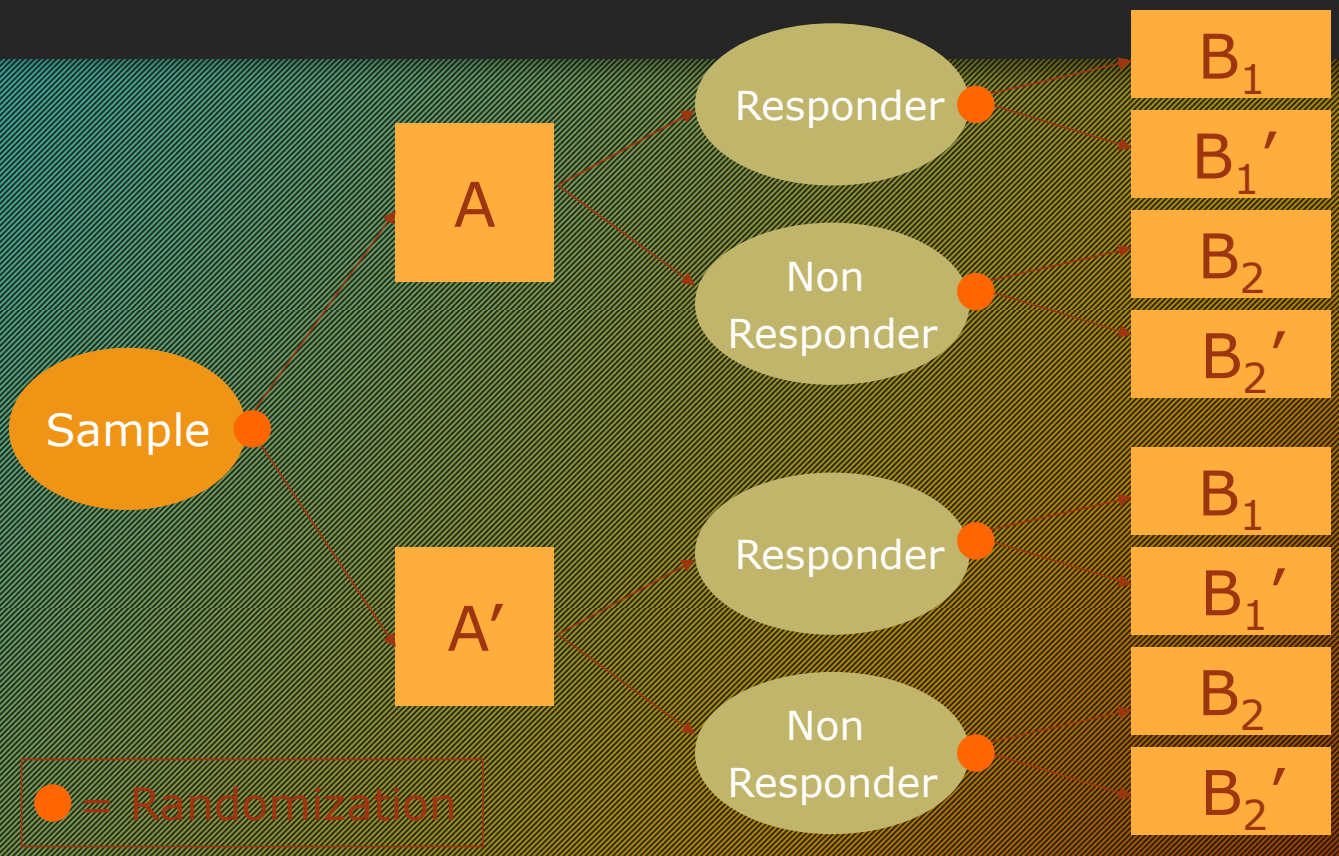
Sequential Multiple Assignment Randomized Trials (SMART)[Murphy, 2005; Lavori, 2001]

The SMART designs were adapted to:

- Cancer (Thall et al. 2000, Matthay et al., 2009)
- Alzheimer's Disease (Schneider et al. 2001)
- Depression [STAR*D] (Rush 2004)
- ADHD
- Drug Abuse
- Alcohol

https://methodology.psu.edu/ra/adap-inter/projects

Design 3: SMART



Comparison of 3 Study Designs

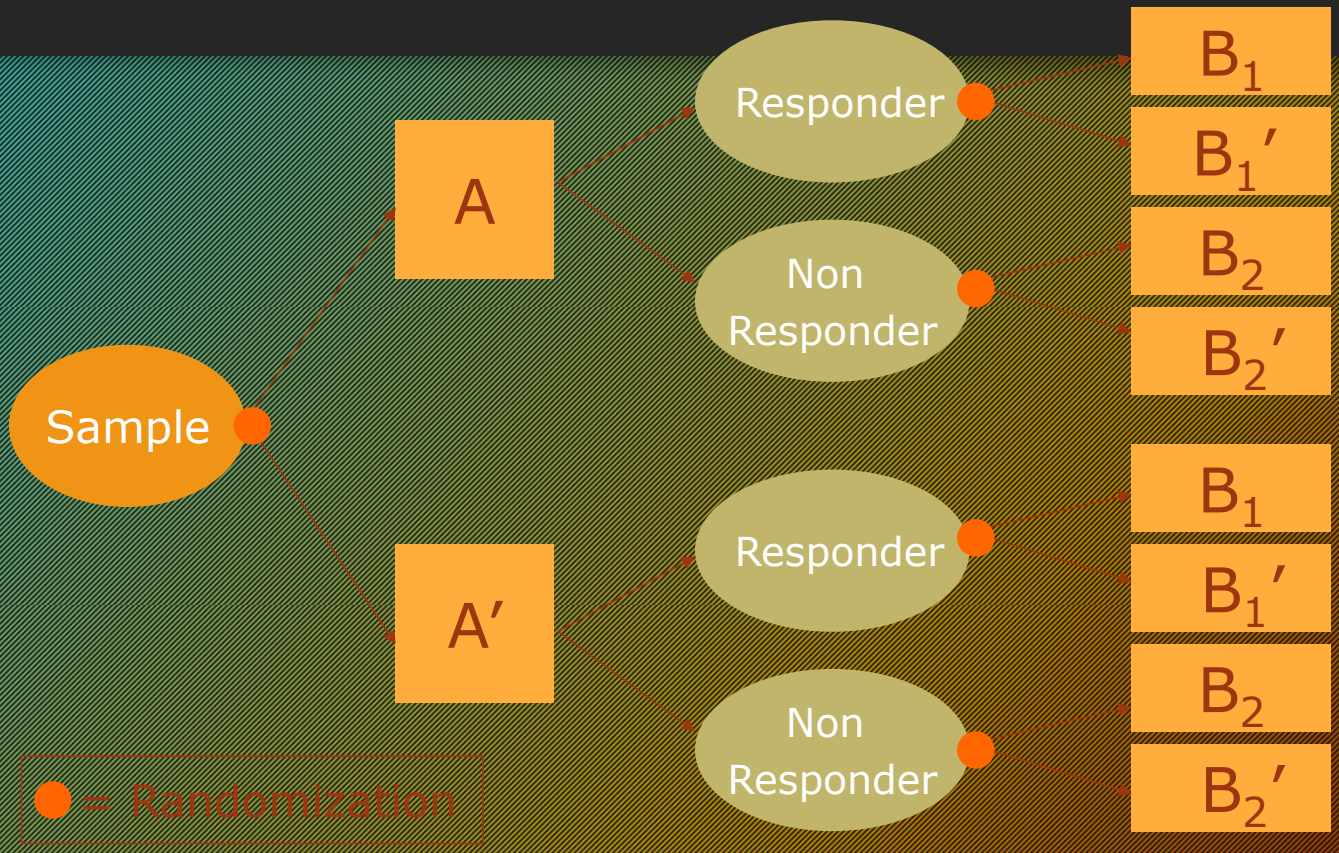
Question:	A Trial with 8 Trts	Combined Trial	SMART
1. Does it serve the purpose of finding the best strategy?	Yes	Maybe	Yes
2. Is it feasible?	No	No	Yes
3.Can we use standard statistical methods to analyze data?	Yes	Maybe	No

SMART Designs

27

Procedure, Assumptions, and Inference

Sequentially Randomized Designs



Causal Effect of a Treatment Regimen

- $T_i(XYZ)$ represents the outcome of the i-th individual treated under this regimen XYZ
- Unless individual i follows one of the following treatment paths, $T_i(XYZ)$ will not be observed:
  - Received X, responded, and then received Y
  - Received X, did not respond, and then received Z
- 8 such variables can be thought of for each individual

Counterfactual Variables

- All eight variables $\{T_i(AB_1B_2), T_i(AB_1B_2'), T_i(AB_1'B_2), T_i(AB_1'B_2'), T_i(A'B_1B_2), T_i(A'B_1B_2'), T_i(A'B_1'B_2), T_i(A'B_1'B_2')\} \equiv \mathcal{T}_i^*$, cannot be observed for the same individual i

- These are referred to as counterfactuals (contrary to what happens in reality)

- Causal inference focuses on the distribution of these counterfactual variables

The Estimands

$$\mu(XYZ) - E[T(XYZ)]$$

$$X \in \{A, A'\}$$
$$Y \in \{B_1, B_1'\}$$
$$Z \in \{B_2, B_2'\}$$

Observed data in SMART

$$\{I_i(A),\ R_i,\ R_iZ_{1i},\ (1-R_i)Z_{2i},\ T_i\},\ i=1,2,\ldots,n$$

$I(A)$ = Indicator of treatment A

R = Response indicator (1/0)

Z_1 = Treatment B_1 indicator (1/0)

Z_2 = Treatment B_2 indicator (1/0)

T = Observed outcome

Relationship between observed data and counterfactuals

Consistency Assumption (CA)

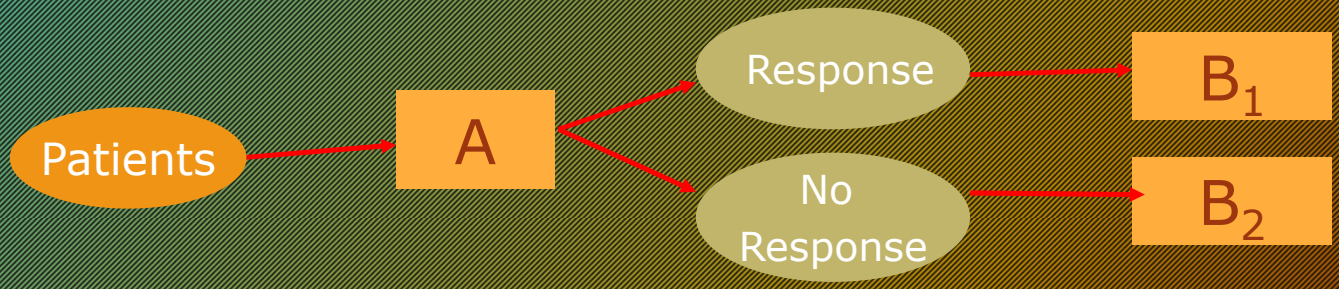
Counterfactual outcome = observed outcome, given the observed treatment history is consistent with the counterfactual;

for example,

$$T_i(AB_1B_2) = I_i(A)[R_iZ_{1i} + (1 - R_i)Z_{2i}]T_i$$

Estimation

- Let us focus on the policy AB_1B_2
  - What would we do if everyone in the sample were treated according to the regimen AB_1B_2 ?



Estimation

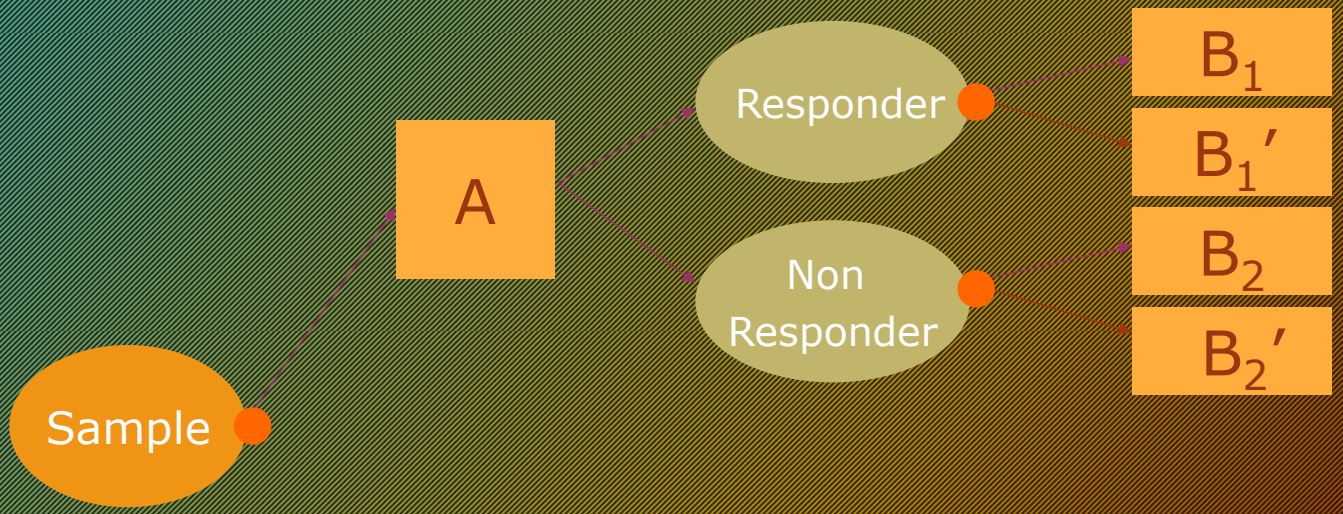
Answer: One-Sample Problem

$$\widehat{\mu}\,(AB_1B_2) = \frac{\sum T_i}{n} = \frac{\sum T_i(AB_1B_2)}{n}$$ (By CA)

Applies to the 8-arm randomization trial

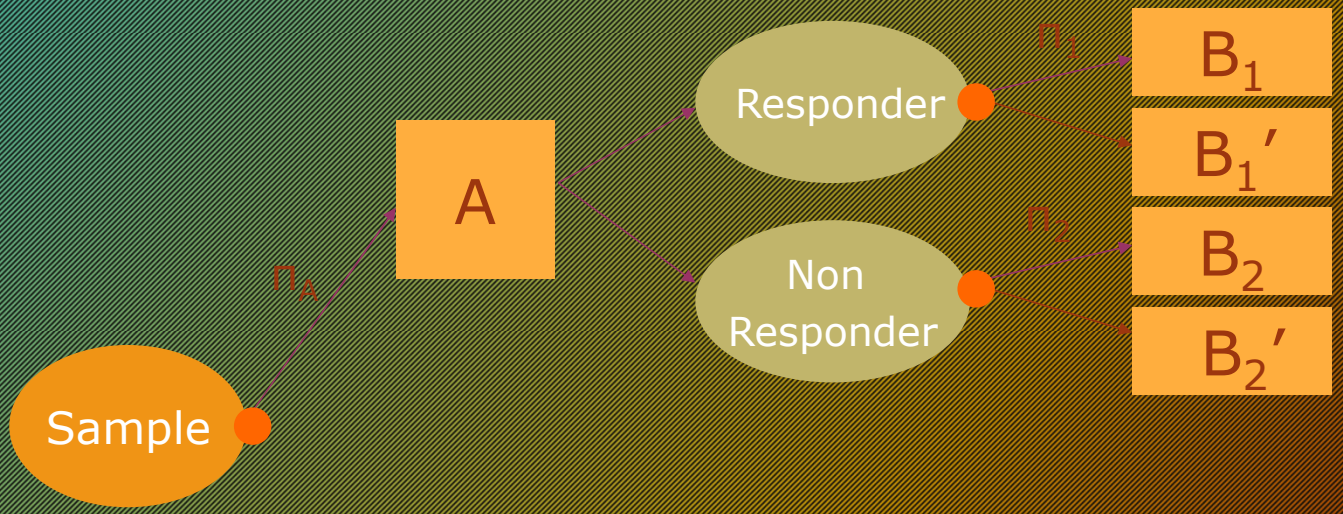
Estimation

- But in SMART, we have <u>not</u> treated everyone with AB_1B_2



Estimation

- Let $C(AB_1B_2)$ be the set of patients who were treated according to the policy AB_1B_2



Naïve estimator

$C(AB_1B_2) = \{i: I_i(A)[R_iZ_{1i} + (1-R_i)Z_{2i}]=1\}$

One would define

$$\hat{\mu}(AB_1B_2) = \sum I_i(A)\,[R_iZ_{1i} + (1-R_i)Z_{2i}]T_i/n'$$

Where

$$n' = \sum I_i(A)[R_iZ_{1i} + (1-R_i)Z_{2i}].$$

Naïve estimator

This estimator would be biased as it ignores the second randomization, except when

the probability of responders being randomized to B_1 is the same as the probability of non-responders being randomized to B_2.

Unbiased Estimation

Sequential Randomization Assumption (SRA)

Treatment assignment at each stage does not depend on counterfactual, given the treatment and covariate history prior to randomization;

For example,

$$P[U_i(A)=1 \mid V_i, \mathcal{H}_1^*] = P[U_i(A)=1 \mid V_i]$$

$$P[Z_{ji}=1 \mid R_i, V_i, \mathcal{H}_1^*] = P[Z_{ji}=1 \mid R_i, V_i], \; j = 1, 2$$

Unbiased estimator

Sequential Randomization Assumption (SRA)

SRA is conveniently satisfied in sequentially randomized trials, since patients are randomized at each stage with **known** probabilities;

Let,

$$P[I_i(A)=1 \mid V_i, \mathcal{H}_i^*] = P[I_i(A)=1 \mid V_i] = \pi_A$$

$$P[Z_{ji}=1 \mid R_i, V_i, \mathcal{H}_i^*] = P[Z_{ji}=1 \mid R_i, V_i] = \pi_j,\ j = 1, 2$$

Inverse-Probability-Weighting

Motivation

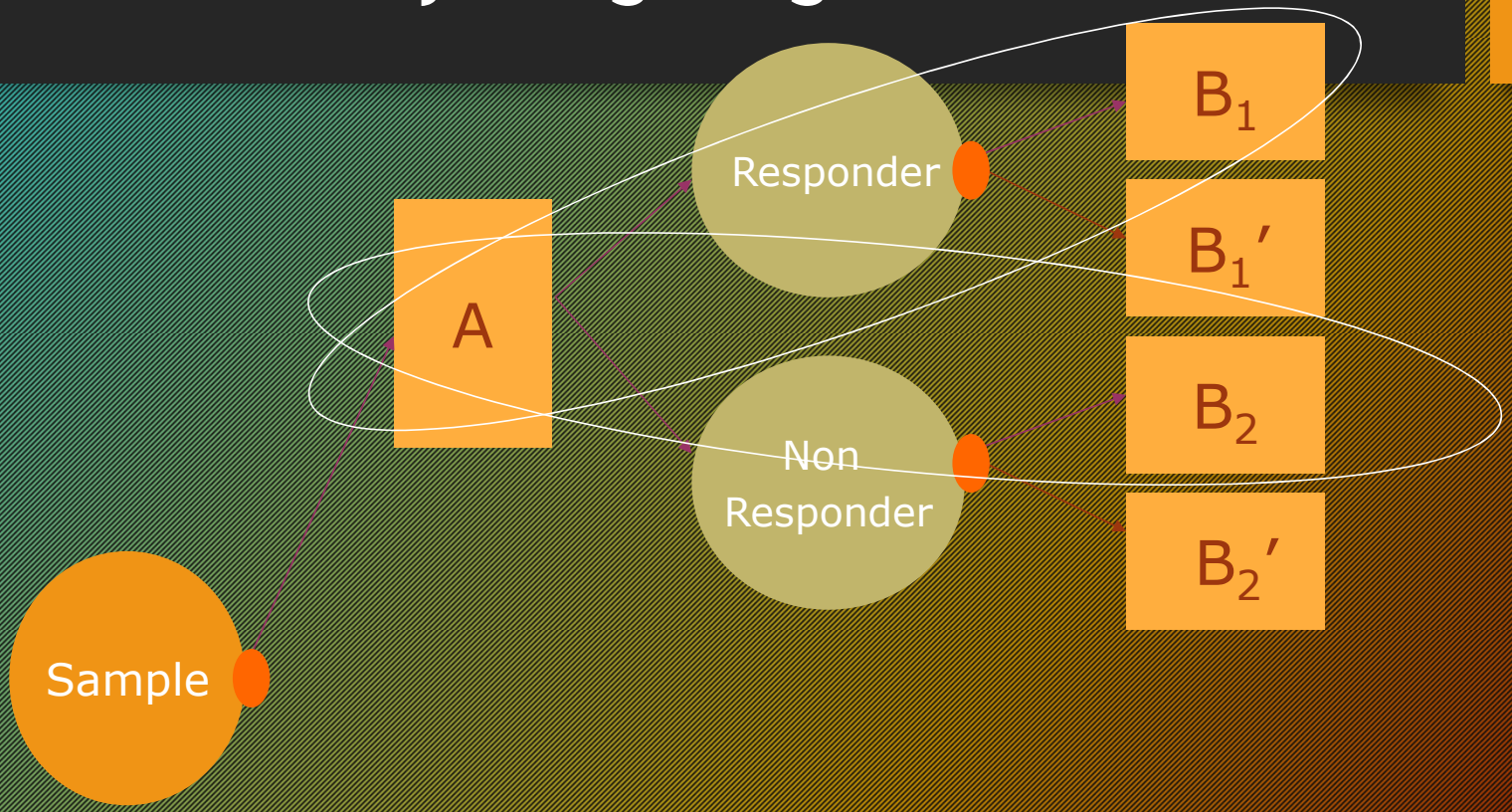
- There are two types of patients in the set $C(AB_1B_2)$ who were treated according to the regimen AB_1B_2

  Responders who received B_1

  and

  Non-responders who received B_2

Inverse Probability Weighting



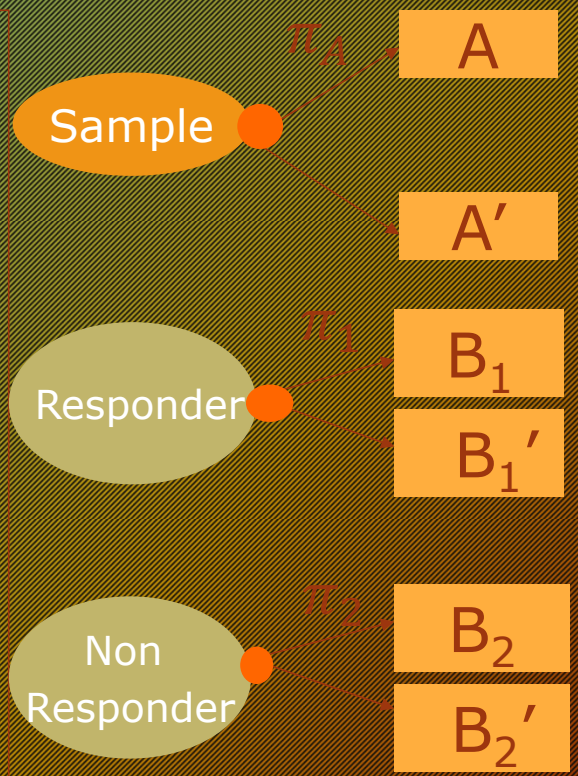
Inverse Probability Weighting

Suppose $\pi_A = \pi_1 = \pi_2 = 1/2$

A patient who received A' was equally eligible to receive A

A responder who received B_1' was equally eligible to receive B_1

A non-responder who received B_2' was equally eligible to receive B_2



Inverse Probability Weighting

- Thus a patient who received A in $C(AB_1B_2)$ is representative of another patient who received A'
- A responder who received B_1 in $C(AB_1B_2)$ is representative of another patient who received B_1' and
- A non-responder who received B_2 in $C(AB_1B_2)$ is representative of another patient who received B_2'

Inverse Probability Weighting

We define weights as follows

Every patient who received A in $C(AB_1B_2)$ receives a weight of 2 $[=1/(1/2)=1/\pi_A]$

Also,

A responder who received B_1 in $C(AB_1B_2)$ receives a weight of 2 $[=1/(1/2)=1/\pi_1]$

A non-responder who received B_2 in $C(AB_1B_2)$ receives a weight of 2 $[=1/(1/2)=1/\pi_2]$

While everyone else receives a weight of zero (since they received treatment inconsistent with regimen AB_1B_2.

Inverse Probability Weighting

Note that, under CA and SRA

$$E\left\{\frac{I_i(A)}{\pi_A}\left[R_i\frac{Z_{1i}}{\pi_1}+(1-R_i)\frac{Z_{2i}}{\pi_2}\right]T_i\right\}$$

$$=E\left\{\frac{I_i(A)}{\pi_A}\left[R_i\frac{Z_{1i}}{\pi_1}+(1-R_i)\frac{Z_{2i}}{\pi_2}\right]T_i(AB_1B_2)\right\}$$

$$=E\left(E\left\{\frac{I_i(A)}{\pi_A}\left[R_i\frac{Z_{1i}}{\pi_1}+(1-R_i)\frac{Z_{2i}}{\pi_2}\right]T_i(AB_1B_2)\right\}\middle| I_i(A),R_i,T_i(AB_1B_2)\right)$$

$$=E\left\{\frac{I_i(A)}{\pi_A}T_i(AB_1B_2)E\left[R_i\frac{Z_{1i}}{\pi_1}+(1-R_i)\frac{Z_{2i}}{\pi_2}\middle| I_i(A),R_i,T_i(AB_1B_2)\right]\right\}$$

$$=E\left\{\frac{I_i(A)}{\pi_A}T_i(AB_1B_2)\right\}=E\left\{T_i(AB_1B_2)E\left[\frac{I_i(A)}{\pi_A}\middle| T_i(AB_1B_2)\right]\right\}$$

$$=E\{T_i(AB_1B_2)\}=\mu(AB_1B_2)$$

Inverse Probability Weighting

Therefore,

$$\hat{\mu}^{IPW}(AB_1B_2) = \frac{1}{n}\sum_{i=1}^{n} \frac{I_i(A)}{\pi_A}\left[R_i \frac{Z_{1i}}{\pi_1} + (1-R_i)\frac{Z_{2i}}{\pi_2} \right] T_i$$

is an unbiased estimator of $\mu(AB_1B_2)$

A variant of IPW Estimator

$$\hat{\mu}_*^{IPW}(AB_1B_2) = \frac{\sum_{i=1}^{n} \frac{I_i(A)}{\pi_A}\left[R_i \frac{Z_{1i}}{\pi_1} + (1-R_i)\frac{Z_{2i}}{\pi_2}\right]T_i}{\sum_{i=1}^{n} \frac{I_i(A)}{\pi_A}\left[R_i \frac{Z_{1i}}{\pi_1} + (1-R_i)\frac{Z_{2i}}{\pi_2}\right]}$$

Compare treatment regimens

- Wald test of contrasts of regime means are possible, but requires covariance between estimators (which may not be independent of each other)

Covariance between two estimators

Consider two estimators

$$\hat{\mu}^{IPW}(AB_1B_2) = \frac{1}{n}\sum_{i=1}^{n} \frac{I_i(A)}{\pi_A}\left[R_i \frac{Z_{1i}}{\pi_1} + (1-R_i)\frac{Z_{2i}}{\pi_2}\right]T_i$$

$$\hat{\mu}^{IPW}(AB_1B_2') = \frac{1}{n}\sum_{i=1}^{n} \frac{I_i(A)}{\pi_A}\left[R_i \frac{Z_{1i}}{\pi_1} + (1-R_i)\frac{(1-Z_{2i})}{(1-\pi_2)}\right]T_i$$

The two estimators use information from the common set of patients (responders who received B_1. We expect the two estimators to be correlated.

Covariance between two estimators

$$Cov\left\{\hat{\mu}^{IPW}(AB_1B_2),\hat{\mu}^{IPW}(AB_1B_2')\right\}=Var\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{I_i(A)}{\pi_A}\left[R_i\frac{Z_{1i}}{\pi_1}\right]T_i\right\}$$

$$\operatorname{cov}\left\{\hat{\mu}^{IPW}(AB_1B_2),\hat{\mu}^{IPW}(AB_1B_2')\right\}=\frac{1}{n(n-1)}\sum_{i=1}^{n}(T_i^R-\bar{T}^R)^2$$

$$T_i^R=\frac{I_i(A)}{\pi_A}\left[R_i\frac{Z_{1i}}{\pi_1}\right]T_i$$

$$\bar{T}^R=\frac{\sum_{i=1}^{n}T_i^R}{n}$$

G-Estimation (Robins, 1990; Murphy, 2003)

- Weighted average of outcomes from the two stages, without modeling the probability of treatment

$$\hat{\mu}^{GCOMP}(AB_1B_2) = \frac{\sum_{i=1}^{n} I_i(A)Ri}{\sum_{i=1}^{n} I_i(A)} \times \frac{\sum_{i=1}^{n} I_i(A)RiZ_{1i}Y_i}{\sum_{i=1}^{n} I_i(A)RiZ_{1i}} + \frac{\sum_{i=1}^{n} I_i(A)(1-Ri)}{\sum_{i=1}^{n} I_i(A)} \times \frac{\sum_{i=1}^{n} I_i(A)(1-Ri)Z_{1i}Y_i}{\sum_{i=1}^{n} I_i(A)(1-Ri)Z_{1i}}$$

Other topics

- Augmented IPW Estimator[Lunceford et al., 2002; Wahed and Tsiatis, 2004]
- How would the analysis change if the outcome is survival and censoring is present? Log-Rank Test, IPCW estimators [Wahed and Tsiatis, 2004, 2006; Feng and Wahed 2008, 2009; Wahed, 2010; Wahed and Thall, 2013; Kidwell and Wahed, 2013]
- How to account for those patients who do not provide consent to the second stage treatment? (Open question)
- Competing Risks?[Yavuz et al., 2016]
- Adjustment for covariates [Tang and Wahed, 2011, 2013]
- Missing data[Open question]

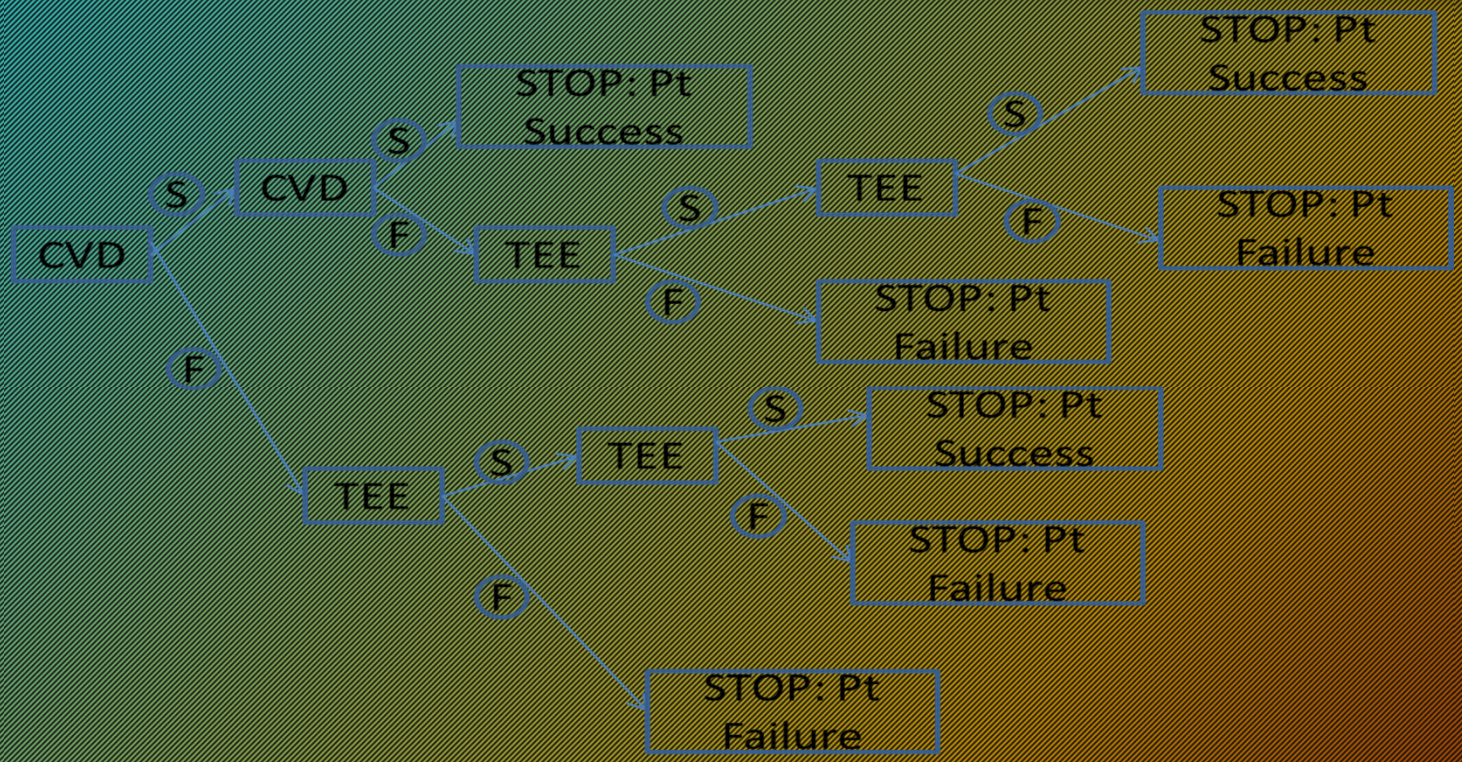
SMART Data Analysis

Example from Bembom and Van der Laan (2007), JNCI

The Prostate Cancer Trial:

- Overall Success: two *consecutive* successful responses
- Overall Failure: two *cumulative* unsuccessful responses
- Stopping Rules: Stop trial when either an overall success or an overall failure has occurred.

The Prostate Cancer Trial:



The Prostate Cancer Trial:

- 3 ways to get overall Patient Success
  - SS, FSS, SFSS (2 consecutive successes)

- 4 ways to get overall patient failure
  - FF, FSF, SFF, SFSF. (2 cumulative failures)

Goal of the Analysis:

- We will use
  - G computation Algorithm
  - Inverse Probability of Treatment Weighting.

G-comp of the Cancer data:

- 12 different dynamic regimes, ab, where a, b € {CVD, KA/VE, TEC, TEE}, a ≠ b.
- For example, the regime CVD-KA/VE dictates to “start with CVD and follow with KA/VE if two consecutive successes were not achieved with CVD.”

G-comp of the Cancer data:

First-Line Therapy				Salvage Therapy				Estimated Overall Success rate
Regimen	No.	S	P	Regimen	No.	S	P	P(95% CI)
CVD	26	4	0.15	KA/VE	10	5	0.50	0.58 (0.28, 0.86)
				TEC	6	1	0.17	0.29 (0.06, 0.63)
				TEE	6	0	0.00	0.15 (0.04, 0.31)
KA/VE	28	7	0.25	CVD	7	0	0.00	0.25 (0.10, 0.42)
				TEC	8	0	0.00	0.25 (0.10, 0.42)
				TEE	6	0	0.00	0.25 (0.10, 0.42)

G-comp of the Cancer data:

- Explaining the estimate for CVD-TEC regime
- Need only the distribution of overall success given treatment and covariate history.
  - 15% experienced overall success on CVD
  - Of the 85% who failed and where randomized to TEC, 17% experienced overall success.

  G-computation estimate of overall success =

  $0.15 + (0.85)(0.17) = 0.29$

G-comp of the Cancer data:

First-Line Therapy				Salvage Therapy				Estimated Overall Success rate
Regimen	No.	S	P	Regimen	No.	S	P	P(95% CI)
TEC	30	14	0.47	CVD	5	1	0.20	0.57(0.33, 0.85)
				KA/VE	4	0	0.00	0.47 (0.28, 0.65)
				TEE	7	0	0.00	0.47 (0.28, 0.65)
TEE	24	10	0.42	CVD	4	1	0.25	0.56 (0.28, 1.00)
				KA/VE	4	0	0.00	0.42 (0.22, 0.61)
				TEC	6	1	0.17	0.51 (0.28, 0.78)

IPTW of the Cancer Data:

- P(receive any 1st line trt)=1/4=0.25
  - IPTW weight = 1/(1/4) =4
- P(receive any salvage trt) = 1/3=0.33
  - IPTW weight =1/(1/3) = 3
- We can improve upon the IPTW performance by using empirical proportions rather than the known randomization probabilities.

IPTW of the Cancer Data:

Rule Overall	No	S1	S2	F	S1	S2	F	Estimated success rate
d(CVD,KA/VE)	14	4	5	5	2.2	5.9	5.9	0.58 (0.28, 0.86)

- P(receive 1st line CVD trt)= 26/108 [IPTW weight = 108/26]
- P(receive KA/VE as salvage after CVD)=10/22 [IPTW Weight = 22/10]

For 4pts with overall 1st line success, weight them by 108/26

For remaining 5 with overall success on salvage trt, upweight them by an additional 22/10

Then normalize to the original observed sample size

IPTW of the Cancer Data:

- 4*(108/26)= 16.61 [1st line success]
- Normalize to observed sample size of 14
  - 16.61*(14/108)=2.2
- 5* (108/26)*(22/10) = 45.69 [Salvage success]
- Normalize to observed sample size of 14
  - 45.69 * (14/108) =5.9
- Then estimated # of total failures= 14- (2.2+5.9) = 5.9
- Est overall success rate = (2.2+5.9)/14= 0.58

IPTW vs. G-computation

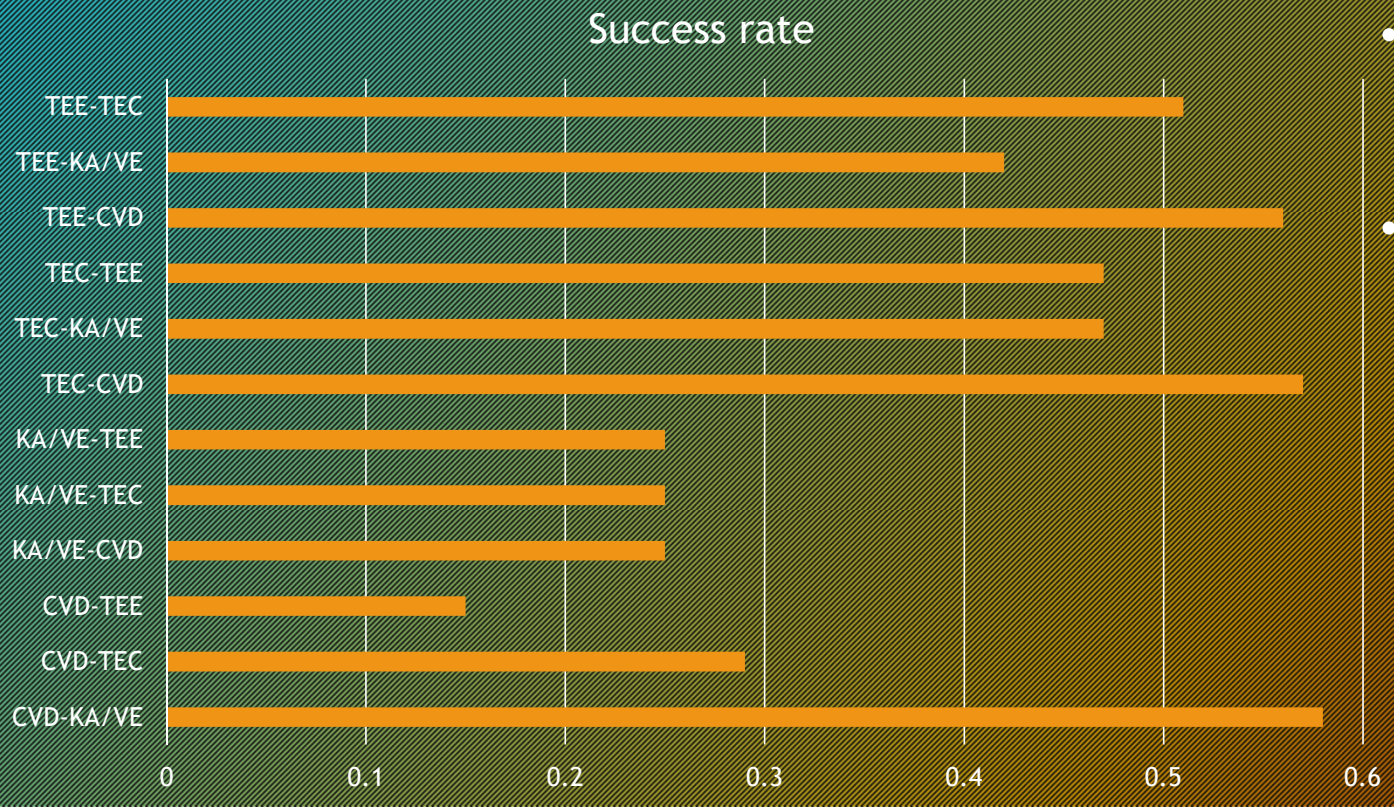
- Identical estimates if IPTW uses the empirical estimates rather than randomization estimates and G-comp does not rely on simplifying assumptions
- IPTW unlike G-comp is guaranteed to provide valid estimates in the absence of any additional assumptions.

Conclusion: Which regime is best?

First Line Therapy	Estimated Overall Success rate		
Regimen	No.	S	P
CVD	26	4	0.15
KA/VE	28	7	0.25
TEC	30	14	0.47
TEE	24	10	0.42

- TEC and TEE are good first line therapies.
- CVD is the worst first line therapy.
- Practical implication: consider TEC or TEE instead of CVD as the first line therapy for treatment naïve patients.

Conclusion: Which regime is best?



- Salvage Success rates are in general low except for KA/VE after CVD.
- The choice of salvage therapies after failure with a 1st line therapy other than CVD is unimportant

Design Issues

Guidelines, Sample size, and power

Design principals – limiting number of DTRs

- Wallace and Moodie (2014) suggested limiting the number of DTRs because of the dangers of creating high dimensional problems and impracticable sample sizes
- At each critical decision point, restrict the class of interventions based on ethical, feasible, or scientific considerations
- Use a low dimension summary instead of all intermediate outcomes to restrict class of next treatments

Design principals – clear adaption rules

- Use a well justified intermediate outcome and tailoring variables
  - How do you define responders and non-responders?
  - Can others use this definition?
  - If not available, a non-restricted SMART may be considered
- Specify when to assess response status
  - Too soon: may not see the initial treatment effects yet
  - Too late: condition may deteriorate so much that you can not rescue it

Design principals – primary and secondary hypotheses

- Choose a primary hypothesis
  - Scientifically important
  - Aids in developing DTR
  - Power to address this hypothesis
- Choose secondary hypotheses
  - Further develop DTRs
  - Trial may not be powered to address these hypotheses
- Collect intermediate outcomes that might be useful in ascertaining for whom each intervention works best (individualization/tailoring variables)
  - Information that may later enter into DTR

SMART in Psychiatry

- Recently SMART has drawn great attention in psychiatry
- A SMART design is uniquely suited to address questions about when to deliver which intervention to treat patients and achieve optimal long-term outcomes
- We discuss two SMART applications in Psychiatric research
  - to illustrate the above design principals
  - to discuss some practical issues

Perinatal weight intervention

- To date, perinatal interventions have not produced lasting improvements in weight or health at one year postpartum
  - Interventions to minimize excessive gestational weight gain (GWG) have had limited impact
  - Efforts to prevent postpartum weight retention have been only modestly successful
- Lifestyle interventions designed to address the full perinatal period can maximize maternal health in the first postpartum year

Rationale for SMART

- There are key questions regarding
  - the optimal timing of interventions for women who vary in pre-pregnancy weight status and GWG, and
  - how best to address their differing needs during pregnancy and the first postpartum year
- A SMART is designed to identify optimal intervention sequences
- It is uniquely suited to address questions about when to deliver intervention during the perinatal period

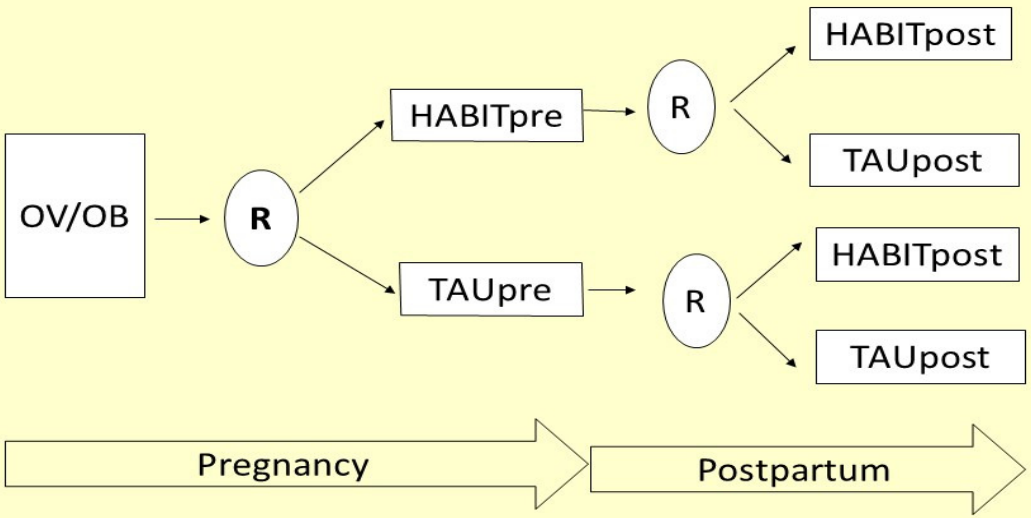
SMART weight

- Women overweight or obese before pregnancy remain at health risk in the postpartum period, regardless of GWG
- Pregnant women (N=300), stratified by prenatal weight category (OV/OB), are enrolled at entry into prenatal care and randomly assigned to
  - Health and Behaviors in Transition [HABITpre] or
  - Treatment As Usual [TAUpre]
- At delivery, all women again are randomized to
  - HABITpost or
  - TAUpost
- The outcomes are weight, cardiometabolic health, depressive symptoms and stress, that are measured at baseline, delivery, 6- and 12-months postpartum

Non-restricted SMART

- We considered to use whether women met the IOM GWG guideline as a tailoring variable
- However, there was no strong rationale for this being a tailoring variable
  - No sufficient data to suggest different interventions for those who have and have not met the GWG goal
- Thus, we have proposed a non-restricted SMART

Non-restricted SMART weight design



Specific aim 1

- To determine the combination of prenatal and postpartum lifestyle interventions that improves weight and secondary outcomes at 12-months postpartum
  - H1: HABITpre → HABITpost will lead to better outcomes than HABITpre → TAUpost or TAUpre → HABITpost
  - H2: TAUpre → HABITpost will lead to better outcomes than HABITpre → TAUpost or TAUpre → TAUpost

Specific aim 2

- To evaluate the impact of combinations of interventions by GWG on maternal weight and health outcomes
  - H3: Among women who gain excessive GWG, those who are assigned to HABITpost will have improved outcomes compared to those in TAUpost
  - H4: Among women who receive HABITpost, women who meet GWG goals will have improved outcomes compared to those who exceed GWG goals

Advantages of this SMART design

- This non-restricted SMART enables us
  - to test if GWG can be used as a tailoring variable
  - to examine the impact of pre-pregnancy weight status on the optimal strategies of interventions
  - e.g., if the strategy of HABITpre followed by HABITpost, regardless of GWG, optimizes outcomes for obese women, and if TAUpre then HABITpost is helpful for overweight women only when GWG is excessive
- It will provide data to develop DTRs which can be tested in a more definitive SMART study

Sample size calculation

Row	PREGNANCY	GWG	POSTPARTUM
1	HABITpre	Meet	HABITpost
2	HABITpre	Exceed	HABITpost
3	HABITpre	Meet	TAUpost
4	HABITpre	Exceed	TAUpost
5	TAUpre	Meet	HABITpost
6	TAUpre	Exceed	HABITpost
7	TAUpre	Meet	TAUpost
8	TAUpre	Exceed	TAUpost

- Power calculations were performed for primary aims:
  - H1: Rows 1, 2 vs. 3, 4, 5, 6
  - H2: Rows 5, 6 vs. 3, 4, 7, 8
  - H3: Rows 2, 6 vs. 4, 8
  - H4: Rows 1, 5 vs. 2, 6

Creating randomization lists

- Three hundred participants will be randomized with equal probability to one of the two initial interventions, TAUpre and HABITpre, stratified by their initial weight status (OV/OB)

- Participants will be further randomized with equal probability to one of the two postpartum weight interventions, TAUpost and HABITpost, no matter whether they have met the IOM guideline for GWG

Four randomization lists

- Without attrition, a quarter of subjects are expected to follow one of the four treatment sequences:
  - HABITpre → TAUpost
  - HABITpre → HABITpost
  - TAUpre → TAUpost
  - TAUpre → HABITpost
- We could create a randomization list of four treatment sequences, stratified by initial weight status

Pros and cons

- The advantage of this approach is that now the randomization is just as straightforward as a typical stratified RCT
- However, attrition is common in practice
- Attrition may affect a particular treatment sequence disproportionately

Hypothetical example

- Assume 30 assigned to TAUpre→TAUpost will drop out, and only 10 will drop out from any of the other three treatment sequences
- The timing of dropout is important
  - If all dropouts occur after the second stage randomization, this disproportional attrition cannot be avoided
  - It is also informative
- Now let us assume all dropouts occur during pregnancy
  - It is unlikely that 30 will drop out from TAUpre → TAUpost, and 10 will drop out from TAUpre → HABITpost during pregnancy

Two separate randomization lists

- Under this extreme situation, it is advantageous to first randomize subjects to TAUpre and HABITpre, and then to further randomize remaining subjects to TAUpost and HABITpost

- We will have 55 subjects going through TAUpre → TAUpost using two randomization lists
  - That is, 110 remained are further assigned to TAUpost compared to
  - 45 going with a randomization list of four treatment sequences (i.e., 75 are assigned to this sequence and 45 remain after attrition

Comparisons of the two approaches

- For a non-restricted SMART with equal probability assignment at both stages, a simple randomization list of all treatment sequences is attractive for its simplicity
- A 2-stage rand. strategy is more robust against extreme cases
- Performing all randomizations upfront is disadvantageous if there are important first-stage intermediate outcomes that might be strongly predictive of second-stage primary outcomes (Nahum-Shani et al., 2013)

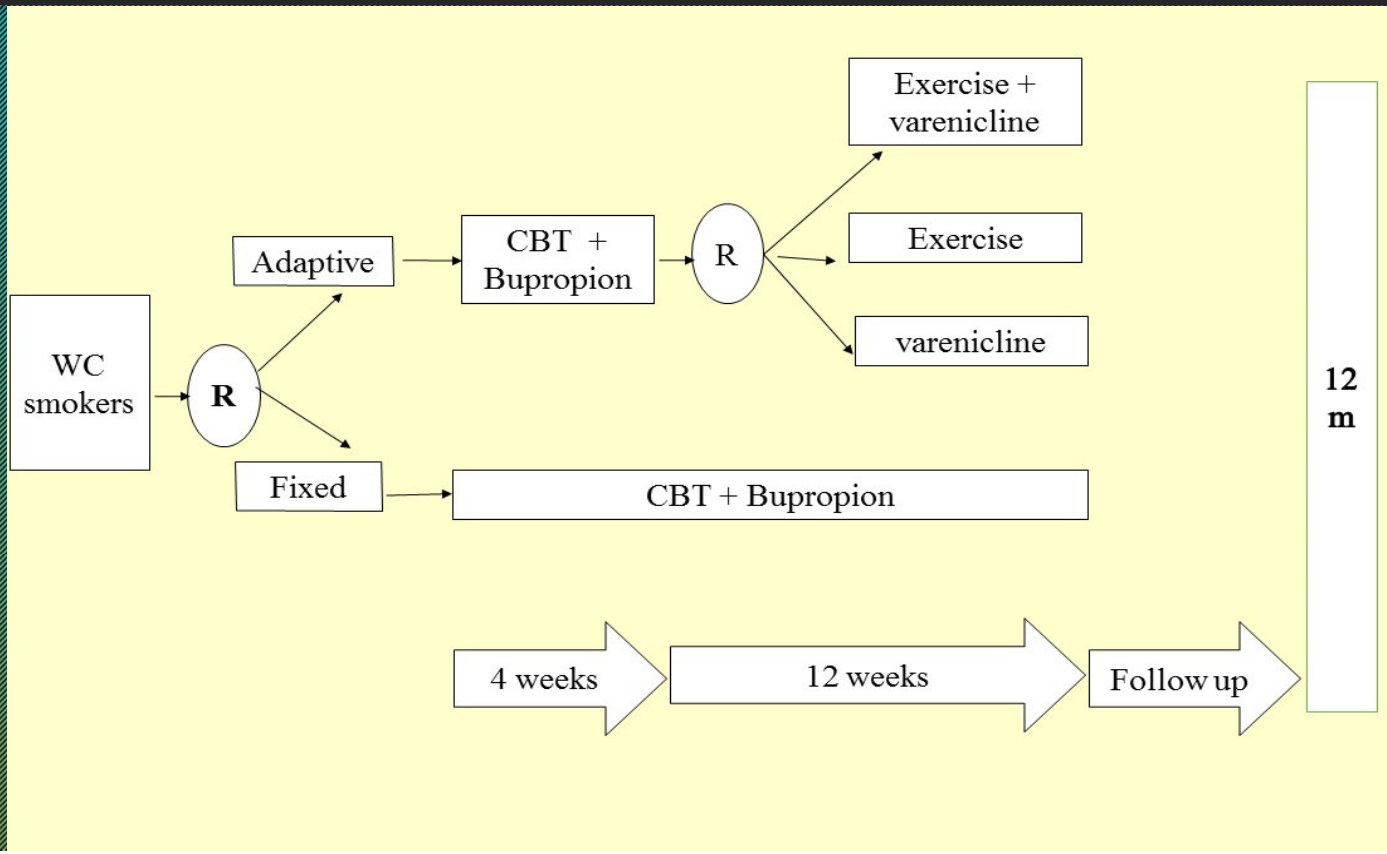
Comparisons of the two approaches

- Another implication of the two randomization approaches is how to perform the intent-to-treat analysis
- If a subject is assigned to a treatment sequence (e.g. TAUpre → TAUpost) upfront, she will be included in the analysis for that treatment sequence regardless of her completion status
- In the 2-stage randomization, if a subject is assigned to TAUpre and drops out of study during pregnancy, the treatment of this subject is consistent with both TAUpre → TAUpost and TAUpre → HABITpost

A smoking cessation SMART study

- To identify how to further improve quit rates at one-year in weight-concerned smokers being given an effective weight-concerns CBT plus bupropion treatment
- Very dependent smokers trying to quit are more successful with varenicline if it is used in combination with bupropion
- Engaging in a moderate exercise regimen may independently enhance ability to quit
- We propose a non-restrictive SMART to test the efficacy of adding varenicline, exercise, or both to augment initial quit rates and improve long-term maintenance of abstinence

SMART smoking



Alternative study designs

- One may consider a 2X2 factorial design

		Varenicline	
		No	Yes
Exercise	No	Standard	Standard + V
	Yes	Standard + Ex	Standard + Ex + V

- Proposed 2-stage randomization is similar to this simpler factorial design
- The multiple stage randomization can lead to the development of empirically-based adaptive interventions for long-term abstinence

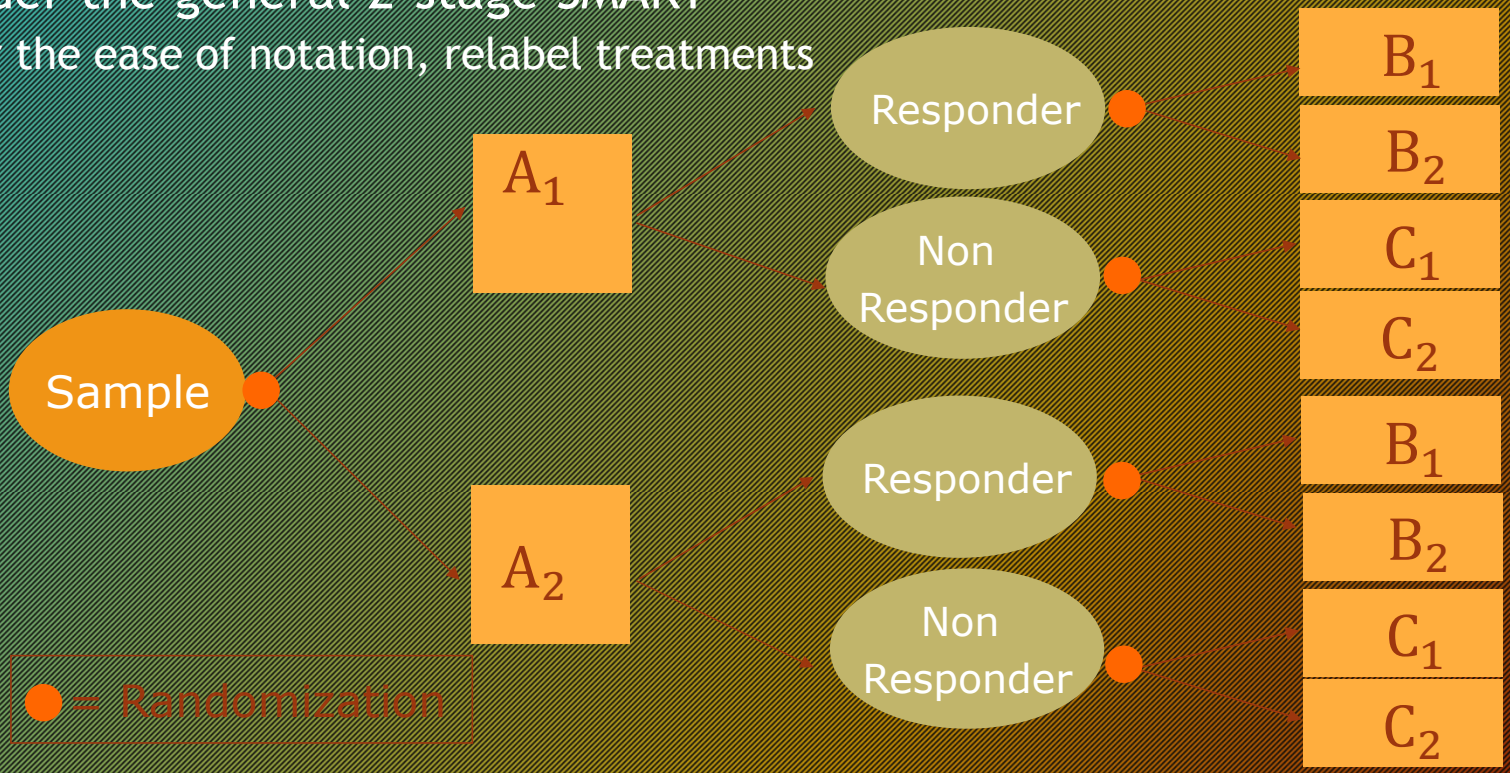
Sample Size and Power

Literature

- **Chapter 5 of *Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data* edited by** Michael R. Kosorok, Erica E. M. Moodie (Continuous and binary outcomes)
- Dawson and Lavori (2010). Continuous outcome
- Feng and Wahed(2008, 2009), Li and Murphy (2011). Survival outcome, limited
- Murphy(2005) . Continuous outcome, limited
- Oetting et al. (2008). Cont. outcome, limited
- Ogbagaber, Karp, and Wahed (2015). Continuous Outcome

Sample sizes and power calculation

- Consider the general 2-stage SMART
  - For the ease of notation, relabel treatments



Continuous outcome

- The mean under DTR $A_jB_kC_l$, $j,k,l = 1,2$, is
$$\mu_{jkl} = E\{Y(A_jB_kC_l)\}$$
- Conditioning on R_j, μ_{jkl} can be expressed as
$$\mu_{jkl} = \pi_j\mu_{A_jB_k} + (1-\pi_j)\mu_{A_jC_l},$$

Where $\mu_{A_jB_k} = E\{Y(A_jB_k)\}, \mu_{A_jC_l} = E\{Y(A_jC_l)\}$ and π_j is the probability of response to induction treatment A_j

Estimation

- The observed data
  $\{X_{ji}(A_j), R_i, R_i Z_{ki}(B_k), (1-R_i)Z'_{li}(C_l), Y_i, j, k, l = 1,2\}, i = 1, \ldots, n$
  - R_j is the observed response status
  - $X_{ji}(A_j)$, $Z_{ki}(B_k)$ and $Z'_{li}(C_l)$are the indicator functions for treatments, i.e., $X_{ji}(A_j) = 1$ if the i-th patient was assigned to A_j, and 0, otherwise
- A consistent estimator for the strategy mean μ_{jkl}

$$\hat{\mu}_{jkl}^{IPW} = \frac{\sum_i^n W_{jkli}\, Y_i}{\sum_i^n W_{jkli}}$$

Where $W_{jkli} = X_{ji}\{\frac{R_i Z_{ki}}{P_k} + \frac{(1-R_i)\, Z'_{li}}{Q_l}\}$

Variance

- $\sqrt{n}\left(\hat{\mu}_{jkl}^{IPW} - \mu_{jkl}\right) = \frac{1}{\sqrt{n}}\sum_i \psi_{jkli} + o_p(1)$ is asymptotically distributed $N(0, \sigma_{jkl}^2)$, where $\psi_{jkli} = k_j W_{jkli}\left(Y_i - \mu_{jkl}\right)$ is the influence func
  - k_j is the limit of $n/\sum X_{ji}$
- $\sigma_{jkl}^2 =$

$$K_j\left[\frac{\pi_j}{P_k}\left\{\sigma_{A_jB_k}^2 + \left(1-\pi_j\right)^2\left(\mu_{A_jB_k} - \mu_{A_jC_l}\right)^2\right\} + \frac{1-\pi_j}{Q_l}\left\{\sigma_{A_jC_l}^2 + \pi_j^2\left(\mu_{A_jB_k} - \mu_{A_jC_l}\right)^2\right\}\right]$$

- Covarinaces between estimators can be calculated as functions of sub-population means and variances and sample sizes.

Overall sample size

- Suppose we want to test
- $H_0: \mu_{111} = \mu_{112} = \mu_{121} = \mu_{122} = \mu_{211} = \mu_{212} = \mu_{221} = \mu_{222}$
- $H_0: C\,\mu = 0$ where C is a contrast matrix.
- Under H_0,

$$n\,\hat{\mu}^T C^T \left[C\,\hat{\Sigma}\,C^T\right]^{-1} C\,\hat{\mu}$$

follows a central chi-square distribution with 7 d.f.

Overall sample size cont.

- $\hat{\mu}^T = [\hat{\mu}_{111}, \hat{\mu}_{112}, \hat{\mu}_{121}, \hat{\mu}_{122}, \hat{\mu}_{211}, \hat{\mu}_{212}, \hat{\mu}_{221}, \hat{\mu}_{222}]^T$
- $C =$

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Overall sample size cont.

- The variance term $\hat{\Sigma} = \begin{bmatrix} \hat{\Sigma}_1 & 0 \\ 0 & \hat{\Sigma}_2 \end{bmatrix}$, where

$$\hat{\Sigma}_1 = \begin{bmatrix} \hat{\sigma}^2_{111} & \hat{\sigma}_{111,112} & \hat{\sigma}_{111,121} & \hat{\sigma}_{111,122} \\ \hat{\sigma}_{111,112} & \hat{\sigma}^2_{112} & \hat{\sigma}_{112,121} & \hat{\sigma}_{112,122} \\ \hat{\sigma}_{111,121} & \hat{\sigma}_{121,112} & \hat{\sigma}^2_{121} & \hat{\sigma}_{121,122} \\ \hat{\sigma}_{111,122} & \hat{\sigma}_{122,112} & \hat{\sigma}_{122,121} & \hat{\sigma}^2_{122} \end{bmatrix}$$

$$\hat{\Sigma}_2 = \begin{bmatrix} \hat{\sigma}^2_{211} & \hat{\sigma}_{211,212} & \hat{\sigma}_{211,221} & \hat{\sigma}_{211,222} \\ \hat{\sigma}_{211,212} & \hat{\sigma}^2_{212} & \hat{\sigma}_{212,221} & \hat{\sigma}_{212,222} \\ \hat{\sigma}_{211,221} & \hat{\sigma}_{221,212} & \hat{\sigma}^2_{221} & \hat{\sigma}_{221,222} \\ \hat{\sigma}_{211,222} & \hat{\sigma}_{222,212} & \hat{\sigma}_{222,221} & \hat{\sigma}^2_{222} \end{bmatrix}$$

Overall sample size cont.

- Under the alternative hypothesis, it follows a non-central chi-squared distribution with 7 d.f. and a non-centrality parameter λ

$$\lambda = n\,\mu^T C^T [\,C\,\Sigma\,C^T]^{-1} C\,\mu$$

- The sample size formula is given by

$$n = \frac{\lambda}{\mu^T C^T [C\,\Sigma\,C^T]^{-1} C\,\mu}$$

Pairwise comparison

- The sample size required to detect a difference between each pairwise comparison
- Bonferroni correction, $\frac{\alpha}{g}$, g is the total number of pairwise comparisons

$$n = \frac{\left[\sigma^2_{jkl} + \sigma^2_{j'k'l'} - 2\,\sigma_{jkl,j'k'l'}\right]\left[Z_{\left\{1-\frac{\alpha}{2g}\right\}} + Z_{\{1-\beta\}}\right]\text{^}2}{\left[\mu_{jkl} - \mu_{j'k'l'}\right]^2} \quad j,k,l = 1,2$$

Simulation study

- Design Parameters:
-
- Subgroup means: $\mu_{A_jB_1} = \mu_{A_jC_2} = 15, \mu_{A_jC_1} = 20, \mu_{A_jB_2} = 22$; subgroup variances: $\sigma^2_{A_jB_k} = 6^2, \sigma^2_{A_jC_l} = 8^2$, for $j, k, l = 1,2$.

- $H_0: \mu_{111} = \mu_{112} = \mu_{121} = \mu_{122} = \mu_{211} = \mu_{212} = \mu_{221} = \mu_{222}$

- Alternatives values: $\mu_{111} = 17.5, \mu_{112} = 15, \mu_{121} = 21, \mu_{122} = 18.5, \mu_{211} = 17.5, \mu_{212} = 15, \mu_{221} = 21, \mu_{222} = 18.5$

- $Q_1 = 0.5$

Simulation study

Scenario	π_1	π_2	P_1	Power	Sample	EP
1	0.5	0.5	0.5	0.8	70	0.84
	0.5	0.5	0.7	0.8	79	0.85
	0.5	0.5	0.5	0.9	89	0.92
	0.5	0.5	0.8	0.9	120	0.92
2	0.2	0.5	0.5	0.8	83	0.82
	0.2	0.5	0.7	0.8	92	0.83
	0.2	0.5	0.5	0.9	106	0.90
	0.2	0.5	0.8	0.9	134	0.92