Precision Medicine: Lecture 09
SMART Design & Analysis

Lina Montoya, Michael R. Kosorok,
Nikki L. B. Freeman and Owen E. Leete

Department of Biostatistics
Gillings School of Global Public Health
University of North Carolina at Chapel Hill

Fall, 2021

Acknowledgments & Resources

Lecture drawn from:

- Kidwell, Ch. 2 of Kosorok & Moodie book: "DTRs and SMARTs: Definitions, designs, and applications"
- Petersen, et. al. Ch. 10 of Kosorok & Moodie book: "Evaluation of longitudinal dynamic regimes with and without marginal structural working models"
- Nahum-Shani, et. al. (2012), Experimental Design and Primary Data Analysis Methods for Comparing Adaptive Interventions. *Psychol Methods*.
- Advanced Topics in Causal Inference - class by Petersen & Montoya
- "Causal Inference for Multiple Time-Point (Longitudinal) Exposures" - workshop by Balzer & Montoya

Outline

Outline

Recall...

- In precision medicine, we often want to know: who should get which treatment?
  - And at what time?
  - And at what dose?
  - And in what sequence?
  - Etc.
- Further: Is giving treatment in an individualized way better than giving it in a non-individualized way (e.g., giving everyone the same treatment)?
- These are fundamentally causal questions

Recall...

- In precision medicine, we often want to know: who should get which treatment?
  - And at what time?
  - And at what dose?
  - And in what sequence?
  - Etc.
- Further: Is giving treatment in an individualized way better than giving it in a non-individualized way (e.g., giving everyone the same treatment)?
- These are fundamentally <u>causal</u> questions
- **What kinds of trials – by design – easily allow for answering these causal questions?**
  - "easily allow for" = by design, satisfy causal assumptions

Sequential Multiple Assignment Randomized Trial (SMART)

What are SMARTs?

- Multistage trial in which administration of treatment (either random or not) is a function of pre-specified variable(s) measured during the trial

Sequential Multiple Assignment Randomized Trial (SMART)

What are SMARTs?

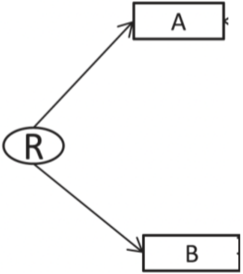
- Multistage trial in which administration of treatment (either random or not) is a function of pre-specified variable(s) measured during the trial
- Said another way: Multistage trial in which participants follow specific dynamic treatment regimes

Sequential Multiple Assignment Randomized Trial (SMART)

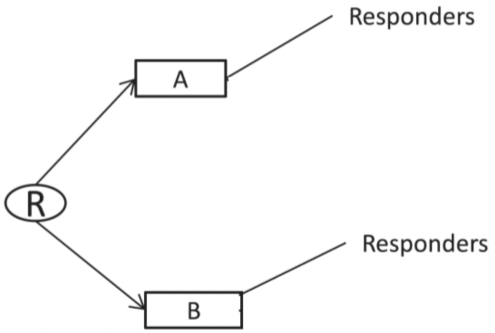
What are SMARTs?

- Multistage trial in which administration of treatment (either random or not) is a function of pre-specified variable(s) measured during the trial
- Said another way: Multistage trial in which participants follow specific dynamic treatment regimes
  - These "specific regimes" are a SMART's **embedded regimes**

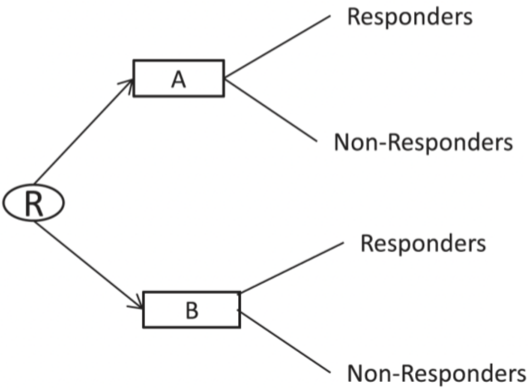
Example # 1



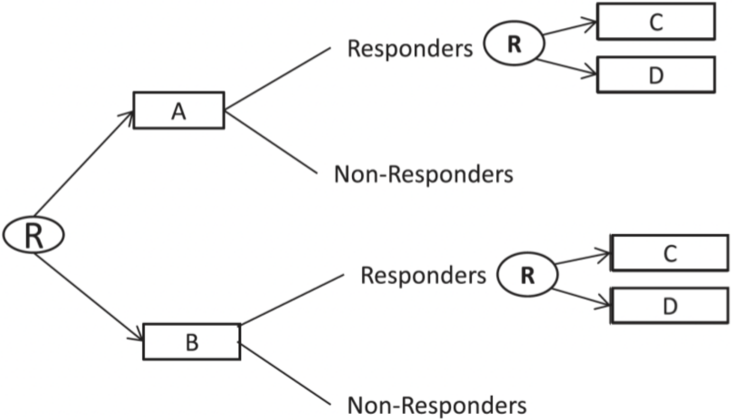
Example # 1



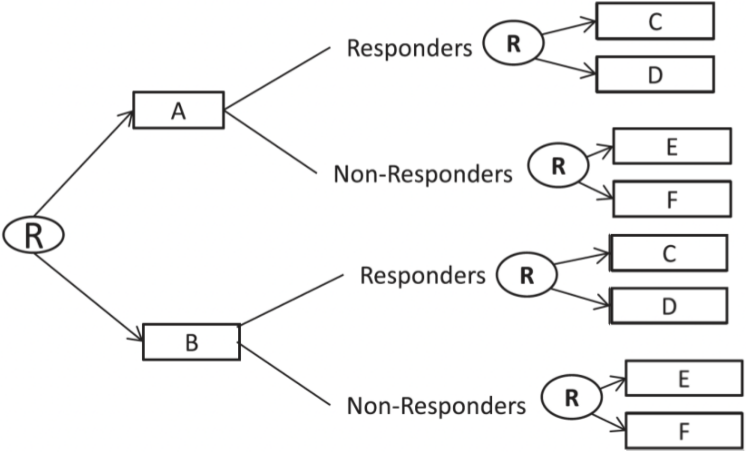
Example # 1



Example # 1


R
A
B
Responders
R
C
D
Non-Responders
Responders
R
C
D
Non-Responders

Example # 1


R
A
B
Responders
Non-Responders
Responders
Non-Responders
R
R
R
R
C
D
E
F
C
D
E
F

Example # 1

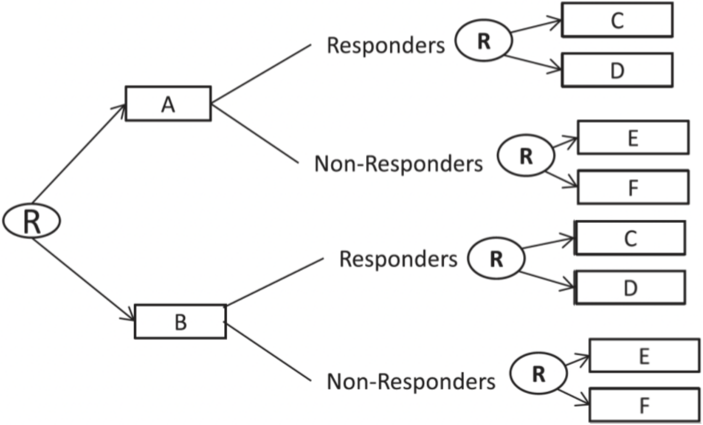
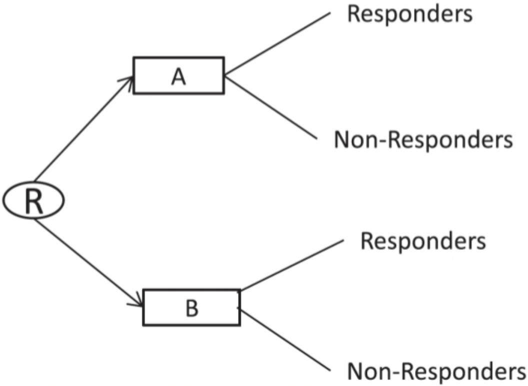
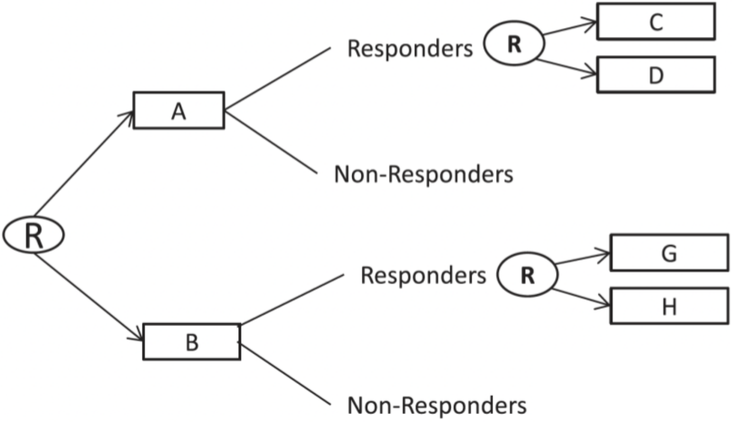


Figure 1: SMART in which administration of second treatment depends on an **intermediate covariate**.

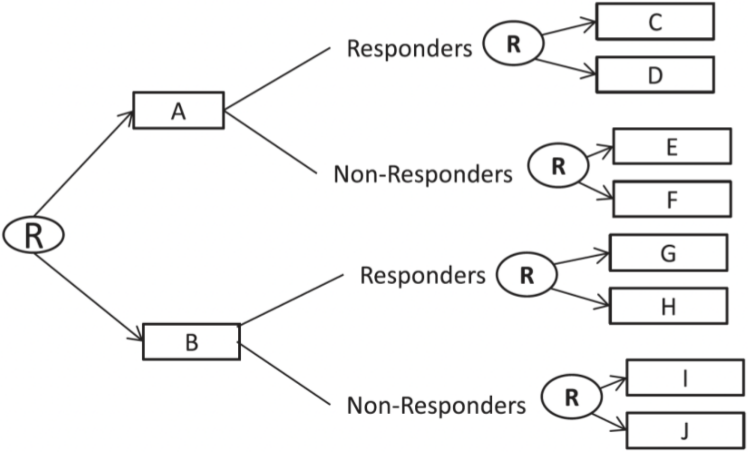
Example # 2



Example # 2


R
A
B
Responders
Non-Responders
R
C
D
Responders
Non-Responders
R
G
H

Example # 2


R
A
B
Responders
Non-Responders
Responders
Non-Responders
R
R
R
R
C
D
E
F
G
H
I
J

Example # 2

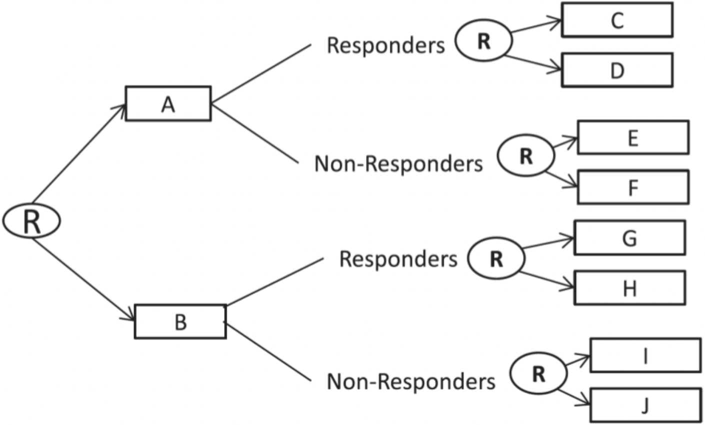


Figure 2: SMART in which administration of second treatment depends on ?

Example # 2

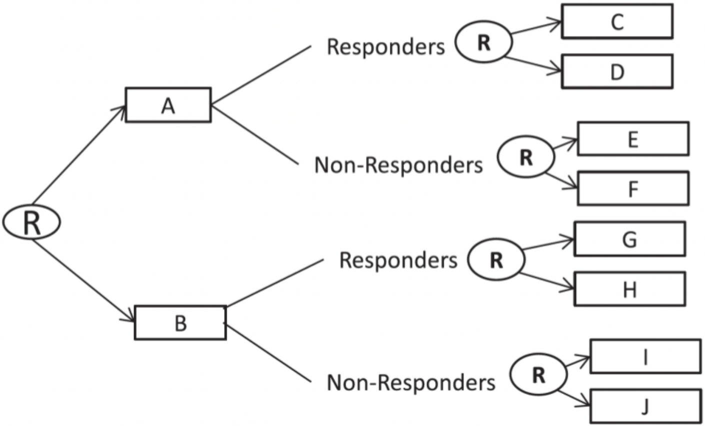
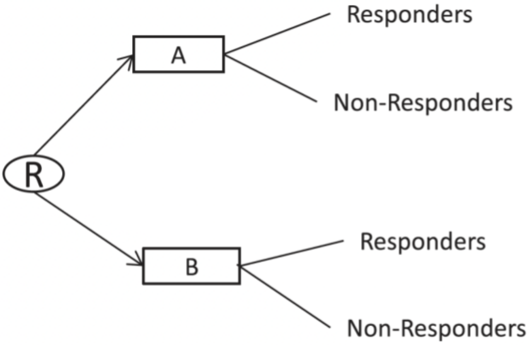
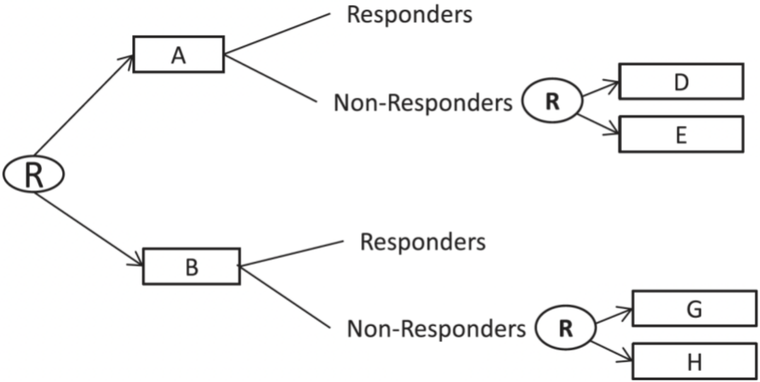


Figure 3: SMART in which administration of second treatment depends on **an intermediate covariate and prior treatment**.

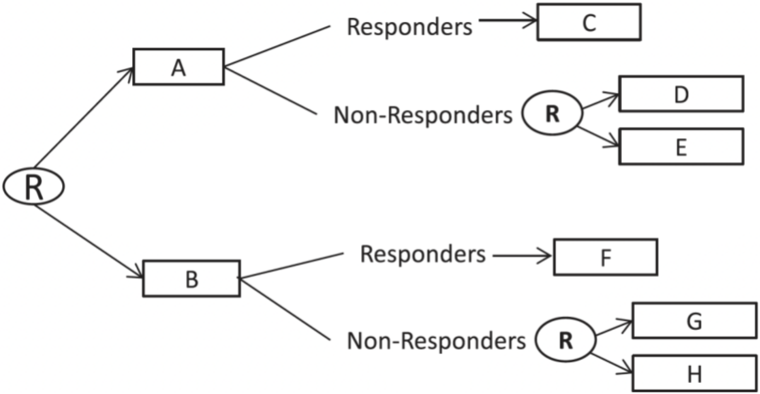
Example # 3



Example # 3


R
A
Responders
Non-Responders
R
D
E
B
Responders
Non-Responders
R
G
H

Example # 3



R
A
B
Responders
Non-Responders
R
C
D
E
Responders
Non-Responders
R
F
G
H

Example # 3

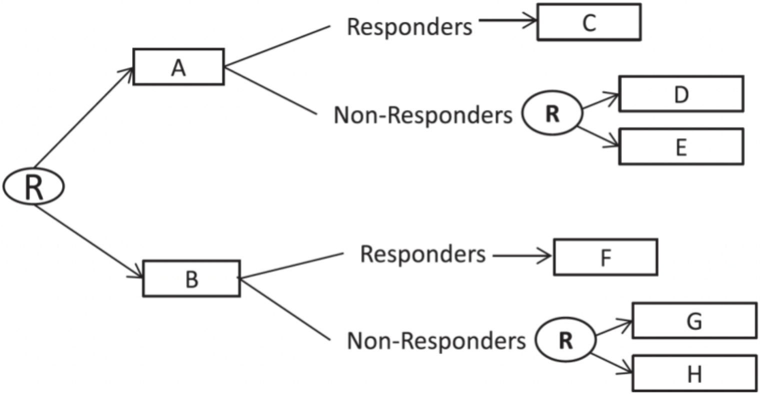


Figure 4: SMART in which administration of second treatment depends on ?

Example # 3

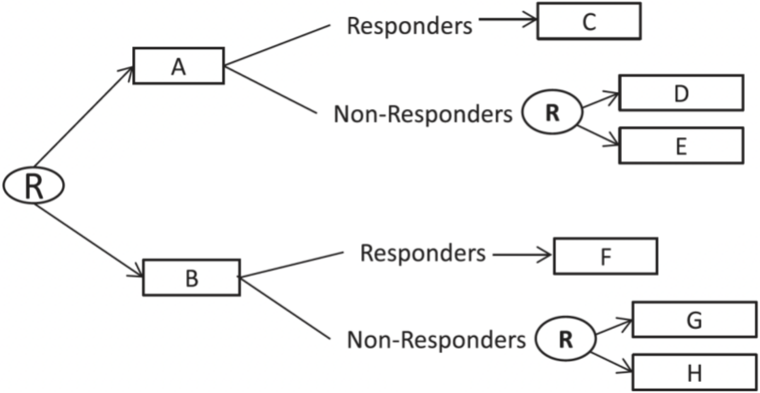
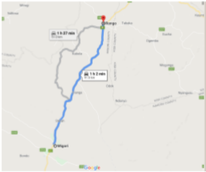


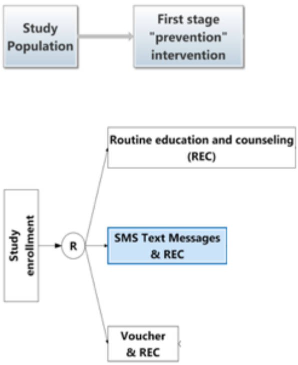
Figure 5: SMART in which administration of second treatment depends **an intermediate covariate and prior treatment...and randomness for non-responders, no randomness for responders**.

Running example - ADAPT-R

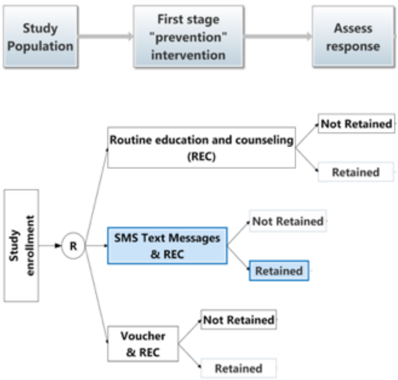
- Adaptive Strategies for Preventing and Treating Lapses of Retention in HIV Care (ADAPT-R; PIs: Petersen & Geng)
- Antiretroviral Therapy (ART) potent and effective, but 20-40% of HIV-infected patients in sub-Saharan Africa lost to follow-up within two years after enrolled in care (e.g., Geng, et. al. 2015)
- Many *diverse* barriers to HIV care retention
  - Structural barriers (e.g. transport to clinic)
  - Psychosocial barriers (e.g., stigma)
  - Other (e.g., long clinic waiting times)
- **No "one-size-fits-all" incentive to stay in care**
  - **Want to know which strategy works best for whom**



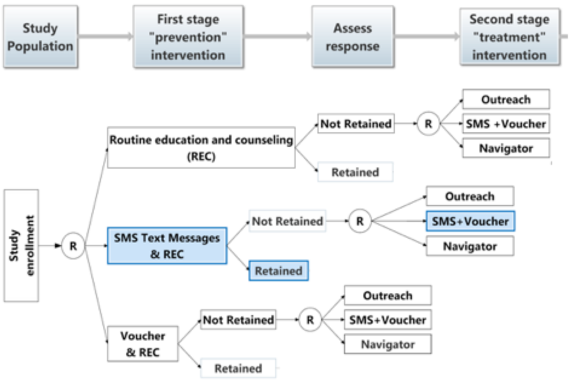

ADAPT-R Design



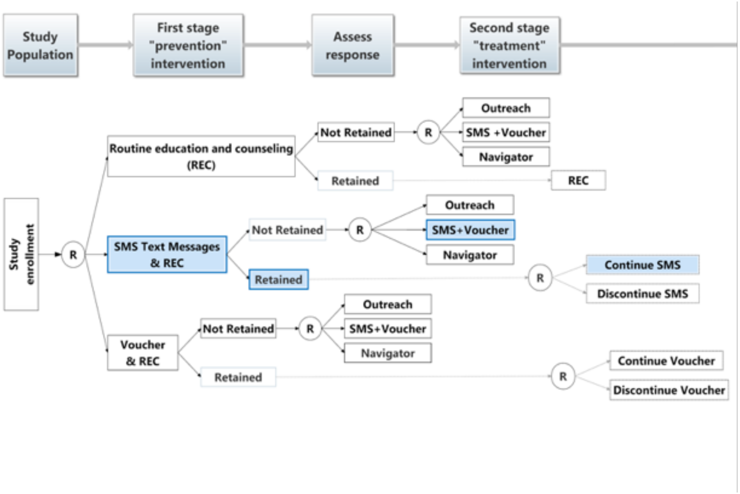
ADAPT-R Design



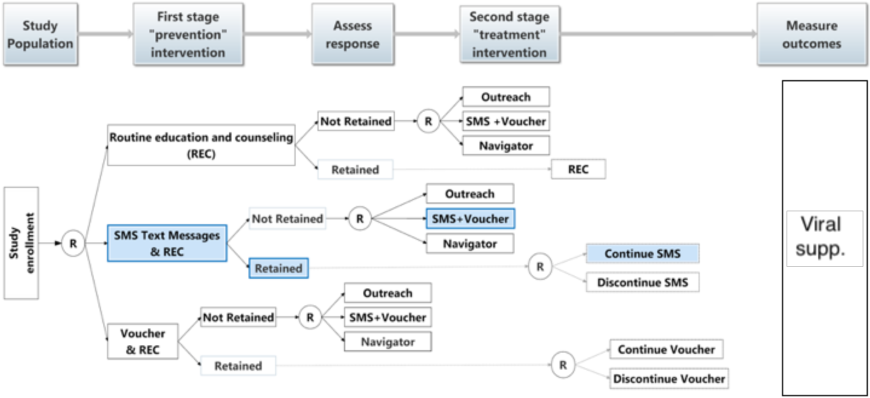
ADAPT-R Design



ADAPT-R Design



ADAPT-R Design



Outline

Outline

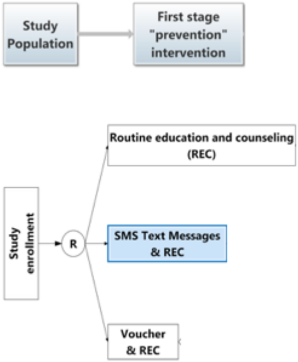
What data come out of a SMART?

Longitudinal data!

In general...

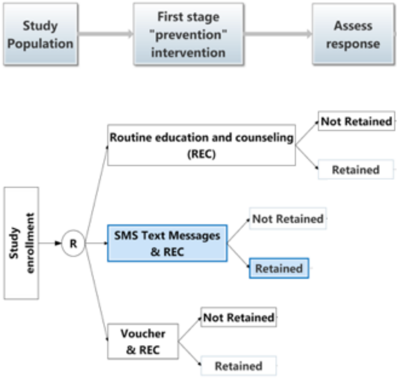
- For a time t, the endogenous variables are
  - Interventions $A(t) \in \mathcal{A}_t$, which could include right-censoring
  - Covariates $X(t)$, which could include baseline covariates and time-varying covariates
  - Outcome Y
- Overbars denote a variable's past history, e.g., $\bar{A}(t) = (A(1), \ldots, A(t))$ and $\bar{X}(t) = (X(1), \ldots, X(t))$
  - $A(0) = X(0) = \emptyset$

ADAPT-R Data



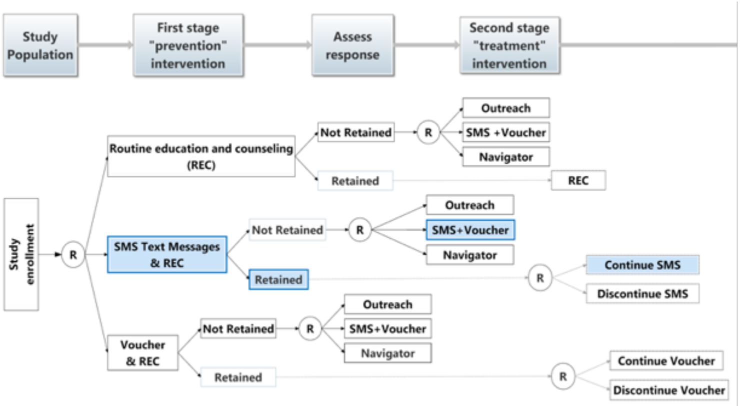
Data collected so far: baseline covariates $X(1)$, Stage 1 intervention $A(1)$

ADAPT-R Data



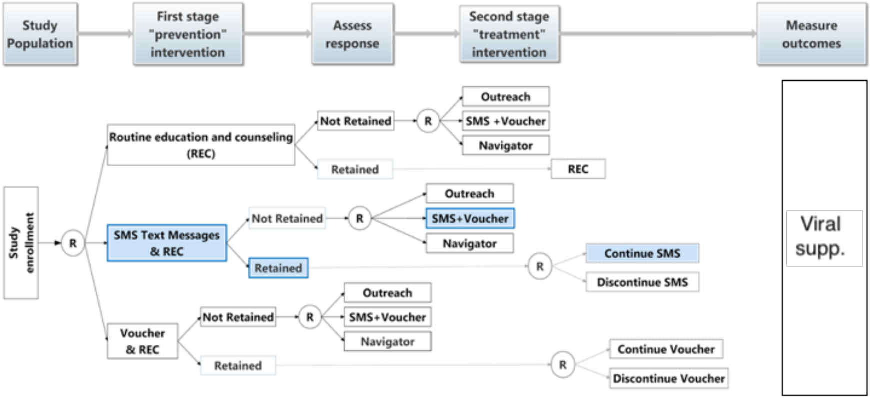
<u>Data collected so far:</u> baseline covariates $X(1)$, Stage 1 intervention $A(1)$, time-varying covariates (incl. first year lapse in care) $X(2)$

ADAPT-R Data



<u>Data collected so far:</u> baseline covariates $X(1)$, Stage 1 intervention $A(1)$, time-varying covariates (incl. first year lapse in care) $X(2)$, Stage 2 intervention $A(2)$

ADAPT-R Data



<u>Data collected so far:</u> baseline covariates $X(1)$, Stage 1 intervention $A(1)$, time-varying covariates (incl. first year lapse in care) $X(2)$, Stage 2 intervention $A(2)$, outcome Y

ADAPT-R Data

Note: $t = 1$ is time of first randomization; $t = 2$ is time of first retention lapse or end of first year

- **X(1)** = baseline covariates (e.g., sex, age, WHO stage, CD4 count, clinic site, etc.)
- **A(1)** = Stage 1 prevention intervention
  - standard of care, SMS text messaging, transportation voucher
- **X(2)** $= (S(2), L(2))$ = time-varying variables
  - $L(2)$ = indicator of lapse in care ($\geq$ 14 days late to a clinic visit) within the first year after enrollment
  - $S(2)$ = all other measured variables between $A(1)$ and $A(2)$, including intermediate outcomes and death
- **A(2)** = Stage 2 "treatment" intervention
  - outreach, SMS and voucher, peer navigator if $L(2) = 1$
  - discontinue, continue if $L(2) = 0$ and $A(1) \in \{$SMS, voucher$\}$
- **Y** = viral suppression at two-years

Outline

What do we know about how the variables relate to each other?

If there were no exclusion restrictions/no assumptions on the functional form...

For observation time $t = 1, \ldots, K$:

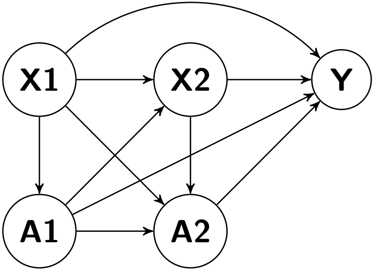
$$
\begin{aligned}
X(t) &= f_{X(t)}(U_{X(t)}, \bar{X}(t-1), \bar{A}(t-1)) \\
A(t) &= f_{A(t)}(U_{A(t)}, \bar{X}(t), \bar{A}(t-1)) \\
Y &= f_Y(U_Y, \bar{X}(K), \bar{A}(K))
\end{aligned}
$$

where exogenous variables are denoted $U = (U_{X(t)}, U_{A(t)}, U_Y)$

What do we know about how the variables relate to each other? 2 time-points

If there were no unmeasured confounders/no exclusion restrictions/assumptions on the functional form...

$$
\begin{aligned}
X(1) &= f_{X(1)}(U_{X(1)}) \\
A(1) &= f_{A(1)}(U_{A(1)}, X(1)) \\
X(2) &= f_{X(2)}(U_{X(2)}, X(1), A(1)) \\
A(2) &= f_{A(2)}(U_{A(2)}, A(1), \bar{X}(2)) \\
Y &= f_Y(U_Y, \bar{X}(2), \bar{A}(2))
\end{aligned}
$$



What do we know about how the variables relate to each other?

For observation time $t = 1, \ldots, K$:

$$\begin{aligned} X(t) &= f_{X(t)}(U_{X(t)}, \bar{X}(t-1), \bar{A}(t-1)) \\ A(t) &= f_{A(t)}(U_{A(t)}, \bar{X}(t), \bar{A}(t-1)) \\ Y &= f_Y(U_Y, \bar{X}(K), \bar{A}(K)) \end{aligned}$$

Is this the correct causal model for SMART data? Should there be any assumptions on dependencies or functional form?

What do we know about how the variables relate to each other?

- Because data were generated from a SMART, treatment assignment at time t only depends on a subset of endogenous variables that occur before $A(t)$ and/or random assignment

What do we know about how the variables relate to each other?

- Because data were generated from a SMART, treatment assignment at time t only depends on a subset of endogenous variables that occur before $A(t)$ and/or random assignment
- Let $\bar{Z}(t) \subseteq (\bar{A}(t-1), \bar{X}(t))$ denote this subset of endogenous tailoring variables used for the SMART's treatment assignment at time t

What do we know about how the variables relate to each other?

- Because data were generated from a SMART, treatment assignment at time t only depends on a subset of endogenous variables that occur before $A(t)$ and/or random assignment
- Let $\bar{Z}(t) \subseteq (\bar{A}(t-1), \bar{X}(t))$ denote this subset of endogenous tailoring variables used for the SMART's treatment assignment at time t
- Thus, we modify the causal model through $A(t)$. A SMART's causal model is:

$$
\begin{aligned}
X(t) &= f_{X(t)}(U_{X(t)}, \bar{X}(t-1), \bar{A}(t-1)) \\
A(t) &= f_{A(t)}(U_{A(t)}, \bar{Z}(t)) \\
Y &= f_Y(U_Y, \bar{X}(K), \bar{A}(K))
\end{aligned}
$$

for observation time $t = 1, \ldots, K$

ADAPT-R Causal Model

- Causal model for ADAPT-R? How do these variables relate to each other?

ADAPT-R Causal Model

- Causal model for ADAPT-R? How do these variables relate to each other?
- What is $\bar{Z}(2)$?

ADAPT-R Causal Model

$$
\begin{aligned}
X(1) &= f_{X(1)}(U_{X(1)}) \\
A(1) &= f_{A(1)}(U_{A(1)}, Z(1)) \\
X(2) &= f_{X(2)}(U_{X(2)}, X(1), A(1)) \\
A(2) &= f_{A(2)}(U_{A(2)}, \bar{Z}(2)) \\
Y &= f_Y(U_Y, \bar{X}(2), \bar{A}(2))
\end{aligned}
$$

where $\bar{Z}(2) = (A(1), L(2))$, i.e., first stage treatment and whether there was a first-year lapse in care.

ADAPT-R Causal Model

$$
\begin{aligned}
X(1) &= f_{X(1)}(U_{X(1)}) \\
A(1) &= f_{A(1)}(U_{A(1)}, Z(1)) \\
X(2) &= f_{X(2)}(U_{X(2)}, X(1), A(1)) \\
A(2) &= f_{A(2)}(U_{A(2)}, \bar{Z}(2)) \\
Y &= f_Y(U_Y, \bar{X}(2), \bar{A}(2))
\end{aligned}
$$

where $\bar{Z}(2) = (A(1), L(2))$, i.e., first stage treatment and whether there was a first-year lapse in care. Specifically...

- $A(1) \sim Multinom(n = 1, k = 3, p_{SMS} = p_{Vouch} = p_{SOC} = \frac{1}{3})$
- $A(2)$ is
  - $\sim Multinom(n = 1, k = 3, p_{SMS+V} = p_{Nav} = p_{SOC} = \frac{1}{3})$ if $L(2) = 1$
  - $\sim Bernoulli(p = \frac{1}{2})$ if $L(2) = 0$ and $A(1) \in \{\text{SMS, Vouch}\}$
  - SOC if $L(2) = 0$ and $A(1) = \text{SOC}$
- $(U_{A(1)}, U_{A(2)})$ are independent of each other and all other Us

Outline

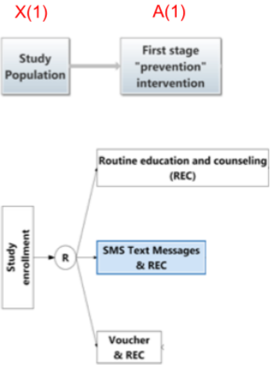
Outline

Review: Decision Rules

- A decision rule d_t is a function that takes as input information accrued on a participant up to time t and outputs a treatment decision for $A(t)$
  - $d_t : (\bar{X}(t), \bar{A}(t-1)) \rightarrow \mathcal{A}_t$
  - $\mathcal{D}_t$ is the set of decision rules at time t

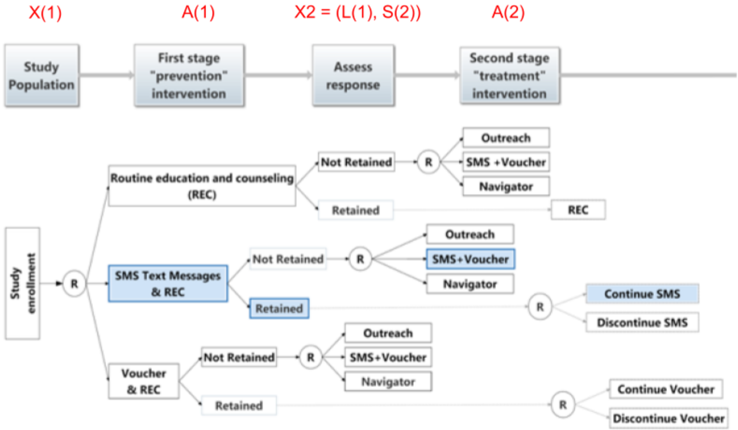
(see PM Lecture 7, slides 9-10)

First Decision Rule - ADAPT-R



First decision rule, $d_1(X(1))$, assigns either SMS, Voucher, or SOC to $A(1)$ based on $X(1)$

Second Decision Rule - ADAPT-R



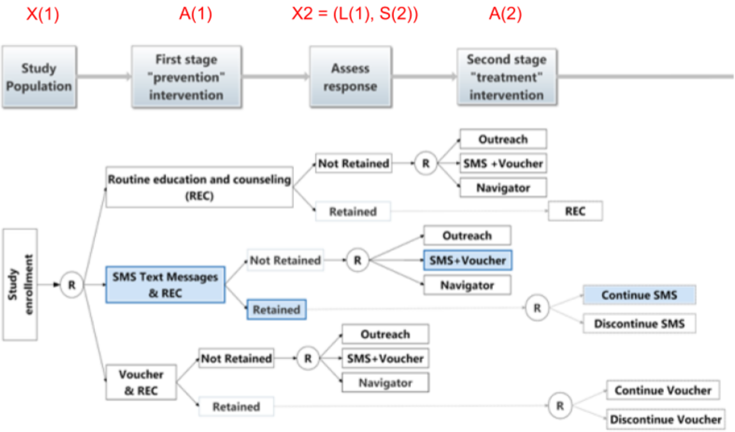
Second decision rule, $d_2(\bar{X}(2), A(1))$, assigns either SMS+Voucher, Nav., SOC, Continue or Discontinue to $A(2)$ based on $X(1)$, $A(1)$, and/or $X(2)$

Review: Dynamic Treatment Regimes

- Let $d = (d_1, d_2, ..., d_K)$ be a dynamic treatment regime (i.e., a sequence of rules)
  - $\mathcal{D}$ is the set of all dynamic treatment regimes

(see PM Lecture 7, slides 9-10)

Dynamic Treatment Regimes - ADAPT-R



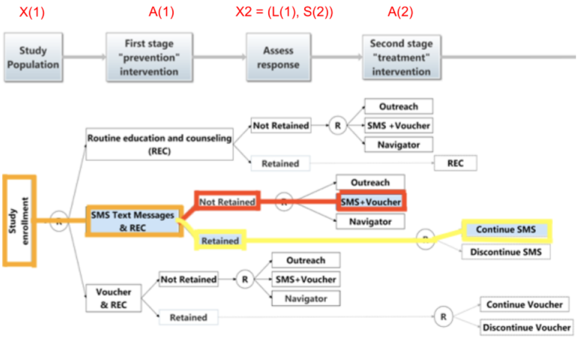
Dynamic regime is $d = (d_1(X(1)), d_2(\bar{X}(2), A(1)))$

Embedded Dynamic Treatment Regimes

A SMART's embedded dynamic treatment regime (AKA embedded regime) is a dynamic treatment regime that will only be a function $\bar{Z}(t)$ (i.e., a SMART's pre-defined tailoring variables)

- i.e., for $t = 1, ..., K$, $d = \big(d_1(Z(1)), ..., d_K(\bar{Z}(K))\big)$
- Let $\mathcal{D}^S \subset \mathcal{D}$ denote the D dynamic treatment regimes embedded within the SMART trial

Embedded Regimes - ADAPT-R



- Embedded regime $d \in \mathcal{D}^S$ is $d = (d_1(Z(1)), d_2(\bar{Z}(2)))$
  - Recall $\bar{Z}(2) = (L(2), A(1))$
  - So $Z(1) = \emptyset$, $Z(2) = (L(2), A(1))$
- Embedded regime example: give everyone SMS; then, give SMS+Voucher if lapse or give Continue SMS if no lapse

ADAPT-R: 15 Embedded Regimes

Embedded Regime #	Stage 1	Stage 2 if Fail	Stage 2 if Succeed
1	SOC	SOC	Continue
2	SMS	SOC	Continue
3	Voucher	SOC	Continue
4	SOC	SMS + Voucher	Continue
5	SMS	SMS + Voucher	Continue
6	Voucher	SMS + Voucher	Continue
7	SOC	Navigator	Continue
8	SMS	Navigator	Continue
9	Voucher	Navigator	Continue
10	SMS	SOC	Discontinue
11	Voucher	SOC	Discontinue
12	SMS	SMS + Voucher	Discontinue
13	Voucher	SMS + Voucher	Discontinue
14	SMS	Navigator	Discontinue
15	Voucher	Navigator	Discontinue

Counterfactual Outcomes

A counterfactual outcome under an embedded regime $d \in \mathcal{D}^S$ is derived under an intervention on the SCM:
For $t = 1, ..., K$:

$$\begin{aligned} X(t) &= f_{X(t)}(U_{X(t)}, \bar{X}(t-1)\bar{A}(t-1)) \\ A(t) &= d_t(\bar{Z}(t)) \\ Y_d &= f_Y(U_Y, \bar{X}(K), \bar{A}(K)) \end{aligned}$$

Y_d is an individual's outcome if, possibly contrary to fact, the individual had been assigned treatment according to the embedded regime d

Causal Question

What are the expected outcomes had all participants received each of the SMART's embedded regimes?

Causal Question/Parameter

What are the expected outcomes had all participants received each of the SMART's embedded regimes?

So, for one $d \in \mathcal{D}^S$:

$$\Psi_d^F(P_{U,X}) = \mathbb{E}_{P_{U,X}}[Y_d]$$

Causal Question/Parameter

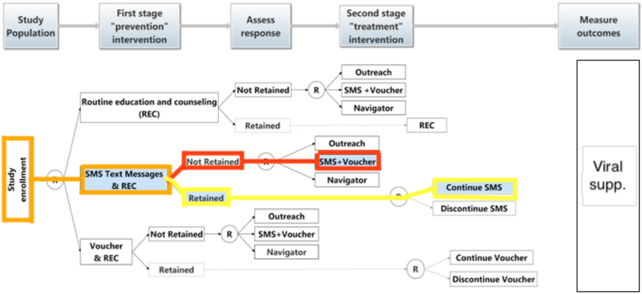
What are the expected outcomes had all participants received each of the SMART's embedded regimes?

So, for one $d \in \mathcal{D}^S$:

$$\Psi^F_d(P_{U,X}) = \mathbb{E}_{P_{U,X}}[Y_d]$$

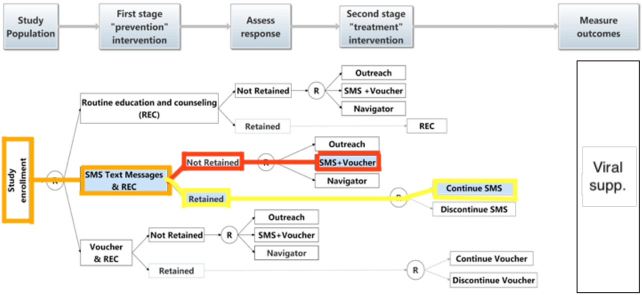
and the vector of the counterfactual values of the D embedded regimes is denoted $\Psi^F(P_{U,X}) = (\Psi^F_{d^{(1)}}(P_{U,X}), ..., \Psi^F_{d^{(D)}}(P_{U,X}))$.

ADAPT-R: Causal Questions/Parameters



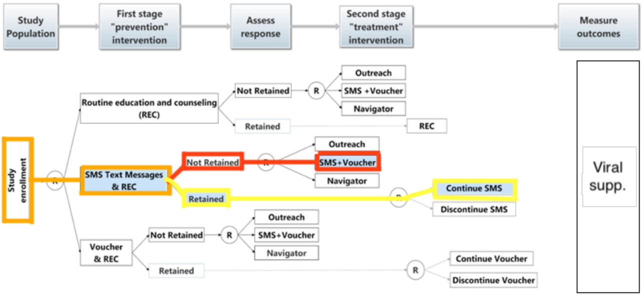
- Prob. 2-yr. viral supp. had all been assigned a specific *embedded regime*?

ADAPT-R: Causal Questions/Parameters



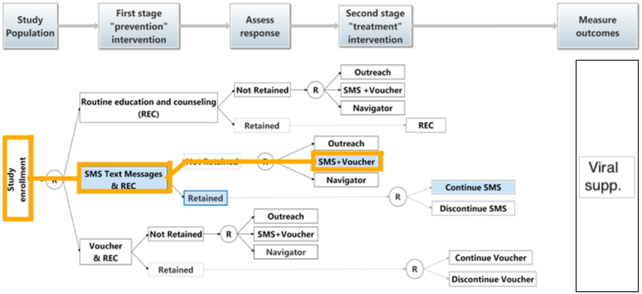
- Prob. 2-yr. viral supp. had all been assigned a specific *embedded regime*?
- More specifically: prob. of 2-yr. viral supp. had everyone been assigned the same prevention intervention, then each person assigned a treatment intervention based on whether or not he/she/they had a lapse in care?

ADAPT-R: Causal Questions/Parameters



- Prob. 2-yr. viral supp. had all been assigned a specific *embedded regime*?
- More specifically: prob. of 2-yr. viral supp. had everyone been assigned the same prevention intervention, then each person assigned a treatment intervention based on whether or not he/she/they had a lapse in care?
  - E.g.: prob. 2-yr. viral supp. had everyone been assigned SMS, then assigned SMS+Voucher if lapse or continue SMS if no lapse
- Causal parameter: $\mathbb{E}_{P_{U,X}}[Y_{A(1)=d_1(Z(1)),A(2)=d_2(\bar{Z}(2))}]$

ADAPT-R: Causal Questions/Parameters



Contrast this with causal question: prob. 2-yr. viral supp. had everyone been assigned SMS, then failed, then assigned SMS+Voucher?

$\mathbb{E}_{P_{U,X}}[Y_{A(1)=a(1),L(2)=1,A(2)=a(2)}]$

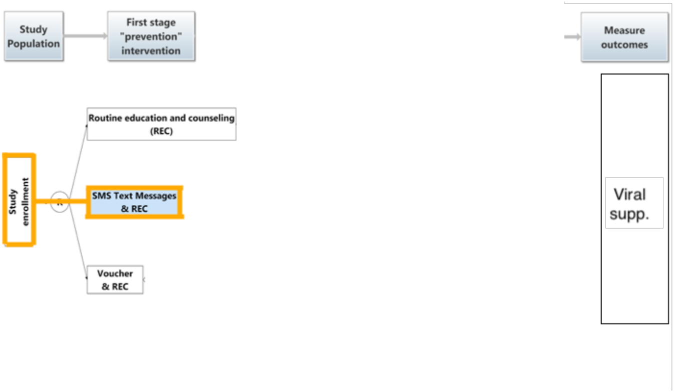
- Here, not evaluating a dynamic regime
- Requires assumptions beyond those satisfied with SMART data because intervening on $L(2)$ (which was not randomized)

Other Causal Parameters/Questions

1. Point treatment static regime: What is the expected outcome had everyone received $A(t) = a(t)$, i.e., $\mathbb{E}[Y_{a(t)}]$?

Other Causal Parameters/Questions

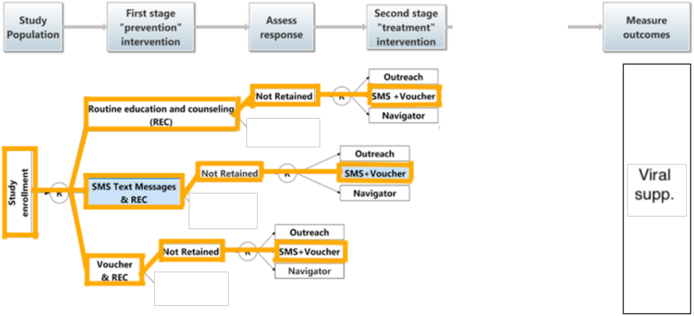
For example...



- Probability of 2-year viral suppression had everyone received the same prevention intervention (e.g., SMS)?
- Causal parameter: $\mathbb{E}[Y_{A(1)=a(1)}]$

Causal Questions

For example...



- Probability of 2-year viral suppression had those who had a lapse in care received the same treatment intervention (e.g., SMS + Voucher)?
- Causal parameter: $\mathbb{E}_{P_{U,X}}[Y_{A(2)=a(2)}|L(2)=1]$

Causal Questions

- In fact, in ADAPT-R, there are 4 simple RCTs pieced together => straightforward answers to causal questions
  - Among all, randomize to prevention intervention
  - Among failures, randomize to intensive intervention
  - Among SMS successes, randomize to discontinue vs. continue
  - Among Voucher successes, randomize to discontinue vs. continue

Causal Questions

1. Point treatment static regime: What is the expected outcome had everyone received $A(t) = a(t)$, i.e., $\mathbb{E}[Y_{a(t)}]$?

    - See, e.g., PM lecture on estimating point-treatment static regimes (Lecture 2)

Causal Questions

1. Point treatment static regime: What is the expected outcome had everyone received $A(t) = a(t)$, i.e., $\mathbb{E}[Y_{a(t)}]$?
    - See, e.g., PM lecture on estimating point-treatment static regimes (Lecture 2)
2. Point treatment optimal dynamic treatment rule: What is the optimal way to assign $A(t)$, i.e., $d_t^*(\bar{X}(t), \bar{A}(t-1)) = \text{argmax}_{d_t \in \mathcal{D}_t} \mathbb{E}[Y_{d_t}]$?
    - See, e.g., PM lecture on single-stage Q-learning and OWL (Lecture 4, 5, & 6)

Causal Questions

1. Point treatment static regime: What is the expected outcome had everyone received $A(t) = a(t)$, i.e., $\mathbb{E}[Y_{a(t)}]$?
    - See, e.g., PM lecture on estimating point-treatment static regimes (Lecture 2)
2. Point treatment optimal dynamic treatment rule: What is the optimal way to assign $A(t)$, i.e., $d_t^*(\bar{X}(t), \bar{A}(t-1)) = \text{argmax}_{d_t \in \mathcal{D}_t} \mathbb{E}[Y_{d_t}]$?
    - See, e.g., PM lecture on single-stage Q-learning and OWL (Lecture 4, 5, & 6)
3. Longitudinal optimal dynamic treatment regime: What is the optimal way to assign $\bar{A}(K)$ in sequence, i.e., $d^* = \text{argmax}_{d \in \mathcal{D}} \mathbb{E}[Y_d]$?
    - See, e.g., PM lecture on multi-stage Q- and A-learning (Lecture 7 & 8)

Outline

Outline

Observed Data & Statistical Model

- Assume the observed data $O_i \equiv (\bar{X}(K)_i, \bar{A}(K)_i, Y_i) \sim P_0 \in \mathcal{M}$, $i = 1, \ldots, n$ were generated by sampling n independent and identically distributed (i.i.d.) times from a data-generating system contained in the causal model
  - Assume ADAPT-R observed data $O_i = (\bar{X}(2)_i, \bar{A}(2)_i, Y_i) \sim P_0 \in \mathcal{M}$, $i = 1, \ldots, 1,809$ were generated by sampling $1,809$ i.i.d. times from a data-generating system contained in the causal model
- P_0 is the observed data distribution, an element of $\mathcal{M}$, the statistical model

Outline

Identification

Two conditions are necessary for identification. What is identification?

For $t = 1, \ldots, K$ and $d \in \mathcal{D}^S$

1. Sequential randomization assumption

$$Y_{d_t} \perp A(t) | \bar{X}(t), \bar{A}(t-1) = \bar{d}_{t-1}(\bar{Z}(t-1))$$

2. Sequential positivity assumption

$$g_0(A(t) = d_t(\bar{Z}(t)) | \bar{X}(t), \bar{A}(t-1) = \bar{d}_{t-1}(\bar{Z}(t-1))) > 0 - a.e.$$

where $g_{A(t),0} = P_0(A(t)|\bar{X}(t), \bar{A}(t-1))$

Identification

Two conditions are necessary for identification; that is, for determining that the causal parameter (a function of the counterfactual distribution, $P_{U,X}$) can be written as statistical parameters (a function of the observed data distribution P_0).

For $t = 1, \ldots, K$ and $d \in \mathcal{D}^S$

1. Sequential randomization assumption

$$Y_{d_t} \perp A(t)|\bar{X}(t), \bar{A}(t-1) = \bar{d}_{t-1}(\bar{Z}(t-1))$$

2. Sequential positivity assumption

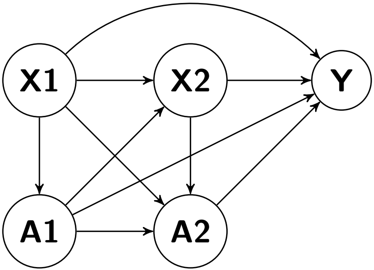
$$g_0(A(t) = d_t(\bar{Z}(t))|\bar{X}(t), \bar{A}(t-1) = \bar{d}_{t-1}(\bar{Z}(t-1))) > 0 - a.e.$$

where $g_{A(t),0} = P_0(A(t)|\bar{X}(t), \bar{A}(t-1))$

Why sequential assumptions? A detour...

Suppose this is our SCM (i.e., no unmeasured confounders, no assumptions on dependence or functional form)

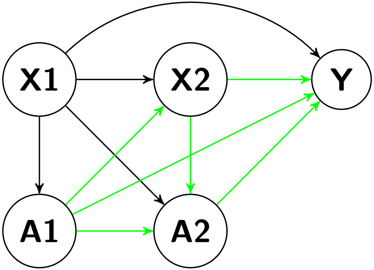
$$
\begin{aligned}
X(1) &= f_{X(1)}(U_{X(1)}) \\
A(1) &= f_{A(1)}(U_{A(1)}, X(1)) \\
X(2) &= f_{X(2)}(U_{X(2)}, X(1), A(1)) \\
A(2) &= f_{A(2)}(U_{A(2)}, A(1), \bar{X}(2)) \\
Y &= f_Y(U_Y, \bar{X}(2), \bar{A}(2))
\end{aligned}
$$



Why sequential assumptions? A detour...

And suppose our target causal parameter is $\mathbb{E}_{P_{U,X}}[Y_{a(1),a(2)}]$

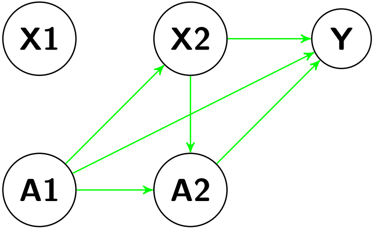
$$\begin{aligned} X(1) &= f_{X(1)}(U_{X(1)}) \\ A(1) &= a(1) \\ X(2) &= f_{X(2)}(U_{X(2)}, X(1), A(1)) \\ A(2) &= a(2) \\ Y &= f_Y(U_Y, \bar{X}(2), \bar{A}(2)) \end{aligned}$$



Why sequential assumptions? A detour...

As in point-treatment case: is conditioning on baseline covariates $X(1)$ enough to identify the effects of both $A(1) = a(1)$, $A(2) = a(2)$ on Y?

$Y_{a(1),a(2)} \perp A(1), A(2) | X(1)$?

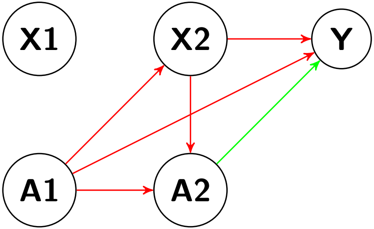


Why sequential assumptions? A detour...

Conditioning on baseline covariates $X(1)$ is **not** enough to identify the effects of both $A(1) = a(1)$, $A(2) = a(2)$ on Y

$Y_{a(1)} \perp A(1)|X(1)$, but
$Y_{a(2)} \not\perp A(2)|X(1)$

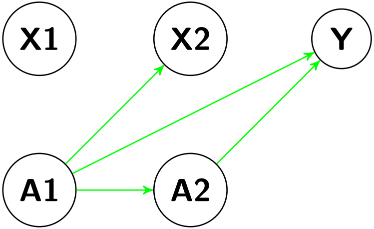
(intuition: $X(2)$ is a confounder for effect of $A(2)$ and Y)



$$\mathbb{E}_{P_{U,X}}[Y_{a(1),a(2)}]$$
$$\neq \sum_{X(1)} \mathbb{E}_0 \left[Y \mid X(1) = x(1), \bar{A}(2) = \bar{a}(2) \right] P_0 \left(X(1) = x(1) \right)$$

Why sequential assumptions? A detour...

Is conditioning on $X(1)$ and $X(2)$ enough to identify the effects of both $A(1) = a(1)$, $A(2) = a(2)$ on Y?



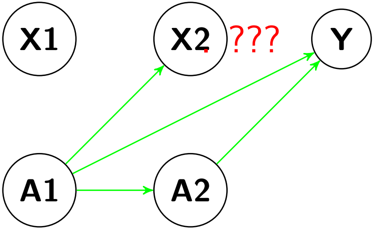
$$Y_{a(1),a(2)} \perp A(1), A(2) | X(1), X(2) \ ?$$

Why sequential assumptions? A detour...

Conditioning on $X(1)$ and $X(2)$ is **not** enough to identify the effects of both $A(1) = a(1)$, $A(2) = a(2)$ on Y

(intuition: losing the effect of $A(1)$ on Y through $X(2)$)

$Y_{a(1)} \not\perp\!\!\!\perp A(1), A(2) | X(1), X(2)$,
and
$Y_{a(2)} \not\perp\!\!\!\perp A(1), A(2) | X(1), X(2)$

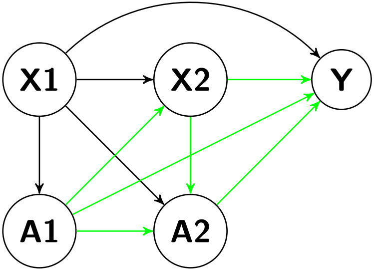


$$\mathbb{E}_{P_{U,X}}[Y_{a(1),a(2)}] \neq \sum_{X(1),X(2)} \mathbb{E}_0 \left[Y \mid \bar{X}(2) = \bar{x}(2), \bar{A}(2) = \bar{a}(2) \right] P_0 \left(\bar{X}(2) = \bar{x}(2) \right)$$

Why sequential assumptions? A detour...

Dilemma of time-dependent confounding:

- Need to standardize with respect to the distribution of both $X(1)$ and $X(2)$ for the effect of both $A1 = a(1)$ and $A2 = a(2)$...
- ...but $X(2)$ are simultaneously *confounders* and *mediators* for that effect



Key Insight

- We don't need to adjust for everything all at once!
- Consider the problem of identifiability *sequentially*
  - For each $A(t)$ in sequence, ask if its effect on Y can be identified by conditioning on some subset of the observed past

Identifiability for the Effects of Multiple Interventions

Sequential Randomization Assumption:

- For 2 timepoints:

$$Y_{a(1),a(2)} \perp A(1)|X(1)$$
$$Y_{a(1),a(2)} \perp A(2)|X(1), A(1), X(2)$$

- For K timepoints:

$$Y_{\bar{a}} \perp A(t)|\bar{X}(t), \bar{A}(t-1) = \bar{a}(t-1), t = 1, \ldots K$$

Identifiability for the Effects of Multiple Interventions

Sequential Randomization Assumption:

- For K timepoints:

$$Y_{\bar{a}} \perp A(t)|\bar{X}(t), \bar{A}(t-1) = \bar{a}(t-1), t = 1, \dots K$$

- When does this hold?
  - If at each time point, A(t) is randomized within strata of (some subset of) covariates and treatment observed up until then
- Then, at each time point the effect of A(t) on future nodes is identified
- (analagous to randomization assumption for a single intervention, which RCTs easily satisfy)

Identifiability for the Effects of Multiple Interventions

Sequential Randomization Assumption:

- For K timepoints:

$$Y_{\bar{a}} \perp A(t) | \bar{X}(t), \bar{A}(t-1) = \bar{a}(t-1), t = 1, \dots K$$

- When does this hold?
  - If at each time point, A(t) is randomized within strata of (some subset of) covariates and treatment observed up until then
- Then, at each time point the effect of A(t) on future nodes is identified
- (analagous to randomization assumption for a single intervention, which RCTs easily satisfy)

Sound familiar? What is ideal experiment to satisfy the sequential randomization assumption?

Identifiability for the Effects of Multiple Interventions

Sequential Randomization Assumption:

- For K timepoints:

$$Y_{\bar{a}} \perp A(t)|\bar{X}(t), \bar{A}(t-1) = \bar{a}(t-1), t = 1, \ldots K$$

- When does this hold?
  - If at each time point, A(t) is randomized within strata of (some subset of) covariates and treatment observed up until then
- Then, at each time point the effect of A(t) on future nodes is identified
- (analagous to randomization assumption for a single intervention, which RCTs easily satisfy)

The ideal experiment is a SMART!

Identification

Two conditions are necessary for identification:

For $t = 1, \ldots, K$ and $d \in \mathcal{D}^S$

1. Sequential randomization assumption ✓

$$Y_{d_t} \perp A(t) | \bar{X}(t), \bar{A}(t-1) = \bar{d}_{t-1}(\bar{Z}(t-1))$$

2. Sequential positivity assumption

$$g_0(A(t) = d_t(\bar{Z}(t)) | \bar{X}(t), \bar{A}(t-1) = \bar{d}_{t-1}(\bar{Z}(t-1))) > 0 - a.e.$$

where $g_{A(t),0} = P_0(A(t) | \bar{X}(t), \bar{A}(t-1))$

Sequential Positivity Assumption

Analogous to point treatment case, need some positive probability of following regime of interest at each time point, regardless of covariate history

For $t = 1, \ldots, K$ and $d \in \mathcal{D}^S$

$$g_0(A(t) = d_t(\bar{Z}(t)) | \bar{X}(t), \bar{A}(t-1) = \bar{d}_{t-1}(\bar{Z}(t-1))) > 0 - a.e.$$

Does this hold? Why?

Sequential Positivity Assumption

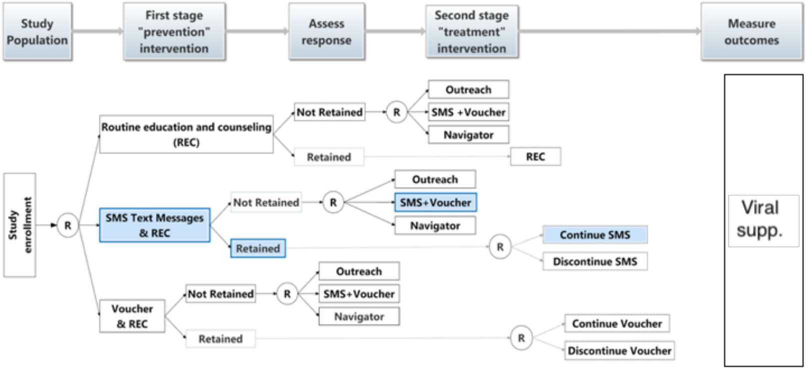
Analogous to point treatment case, need some positive probability of following regime of interest at each time point, regardless of covariate history

For $t = 1, \dots, K$ and $d \in \mathcal{D}^S$

$$g_0(A(t) = d_t(\bar{Z}(t))|\bar{X}(t), \bar{A}(t-1) = \bar{d}_{t-1}(\bar{Z}(t-1))) > 0 - a.e.$$

By design, at each time point, we know the probability $A(t)$ follows the embedded rule $d_t(\bar{Z}(t))$...

Sequential Positivity Assumption



$$g_0(A(1) = d_1(Z(1))|X(1)) = ?$$

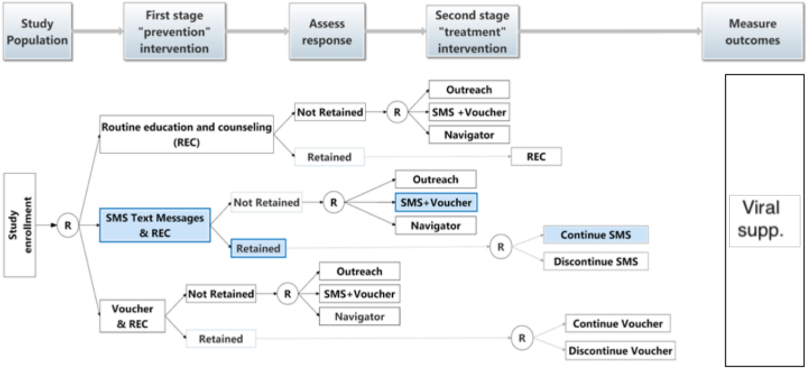
$$g_0(A(2) = d_2(\bar{Z}(2))|L(2) = \text{lapse}, S(2), A(1), X(1)) = ?$$

$$g_0(A(2) = d_2(\bar{Z}(2))|L(2) = \text{no lapse}, S(2), A(1) = \text{SMS}, X(1)) = ?$$

$$g_0(A(2) = d_2(\bar{Z}(2))|L(2) = \text{no lapse}, S(2), A(1) = \text{Voucher}, X(1)) = ?$$

$$g_0(A(2) = d_2(\bar{Z}(2))|L(2) = \text{no lapse}, S(2), A(1) = \text{SOC}, X(1)) = ?$$

Sequential Positivity Assumption



$$g_0(A(1) = d_1(Z(1))|X(1)) = 1/3$$

$$g_0(A(2) = d_2(\bar{Z}(2))|L(2) = 1, S(2), A(1), X(1)) = 1/3$$

$$g_0(A(2) = d_2(\bar{Z}(2))|L(2) = 0, S(2), A(1) = \mathrm{SMS}, X(1)) = 1/2$$

$$g_0(A(2) = d_2(\bar{Z}(2))|L(2) = 0, S(2), A(1) = \mathrm{Voucher}, X(1)) = 1/2$$

$$g_0(A(2) = d_2(\bar{Z}(2))|L(2) = 0, S(2), A(1) = \mathrm{SOC}, X(1)) = 1$$

Identification

Two conditions are necessary for identification:

For $t = 1, \ldots, K$ and $d \in \mathcal{D}^S$

1. Sequential randomization assumption ✓

$$Y_{d_t} \perp A(t) | \bar{X}(t), \bar{A}(t-1) = \bar{d}_{t-1}(\bar{Z}(t-1))$$

2. Sequential positivity assumption ✓

$$g_0(A(t) = d_t(\bar{Z}(t)) | \bar{X}(t), \bar{A}(t-1) = \bar{d}_{t-1}(\bar{Z}(t-1))) > 0 - a.e.$$

Outline

Identification Result

Statistical parameter corresponding to causal parameter $\Psi^F_d(P_{U,X})$ for one embedded regime $d = \bar{d}_K(\bar{Z}(K))$ is the g-computation formula:

$$\Psi_d(P_0) = \sum_{x_1,\ldots,x_K} \mathbb{E}_0\left[Y|\bar{x}(K), \bar{A}(K) = \bar{d}_K(\bar{Z}(K))\right]$$
$$\times \prod_{t=1}^{K} P_0\left(x(t)|\bar{x}(t-1), \bar{A}(t-1) = \bar{d}_{t-1}(\bar{Z}(t-1))\right)$$

where...

- the summation here generalizes to an integral for continuous $X(t)$
- $\bar{d}_t : (\bar{X}(t), \bar{A}(t-1)) \rightarrow \mathcal{A}_1 \times \ldots \times \mathcal{A}_t$ denotes the regime sequence until time t
  - e.g., $\bar{d}_t(\bar{Z}(t)) = (d_1(Z(1)), d_2(\bar{Z}(2)), ..., d_t(\bar{Z}(t))$ is a SMART's embedded regime until time t

Identification Result

Statistical parameter corresponding to causal parameter $\Psi^F_d(P_{U,X})$ for one of ADAPT's embedded regime:

$$\Psi_d(P_0) = \sum_{x_1,x_2} \mathbb{E}_0 \left[Y|\bar{x}(2), A(1) = d_1(Z(1)), A(2) = d_2(A(1), L(2))\right]$$
$$\times P_0 \left(x(2)|x(1), A(1) = d_1(Z(1))\right)$$
$$\times P_0 \left(x(1)\right)$$

The vector of all embedded regime values is identified as $\Psi(P_0) = (\Psi_{d^{(1)}}(P_0), ..., \Psi_{d^{(D)}}(P_0))$; in ADAPT-R, $D = 15$.

Outline

Outline

Outline

Outline

Outline