Using Data from a SMART to Address Primary Aims

Module 3

General Objectives

- A taste of how data from a SMART can be analyzed to address various scientific questions
  - How to frame scientific questions
  - Experimental cells to be compared
  - Resources you can use for data analysis

Outline

- Illustrative Example: ADHD SMART Study (PI: Pelham)
- Learn how to analyze data from SMART to address two typical primary research questions

  (a): Main effect of first-stage options

  (b): Main effect of second-stage options/tactics
- Prepare for a third primary aim analysis by

  (c): Learning to estimate the mean outcome under each of the embedded AIs (separately) using an easy-to-use weighting approach.

d³lab

Note about SAS code

My slides include SAS Code

The goal is to provide the intuition for the data analysis
Not to make you experts on SAS

Outline

- **ADHD SMART study (PI: Pelham)**
- Learn how to analyze data from SMART to address two typical primary research questions

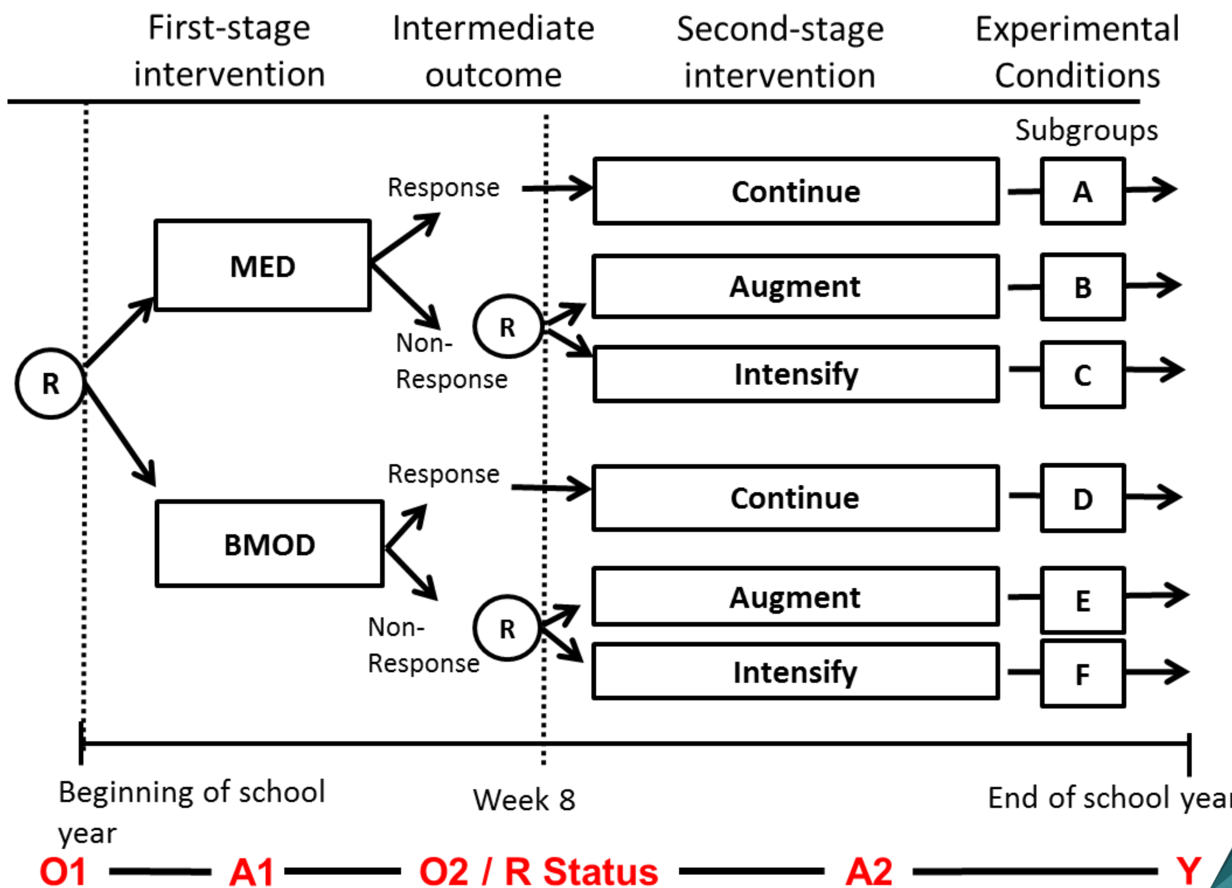
  (a): Main effect of first-stage options

  (b): Main effect of second-stage options/tactics
- Prepare for a third primary aim analysis by

  (c): Learning to estimate the mean outcome under each of the embedded AIs (separately) using an easy-to-use weighting approach.

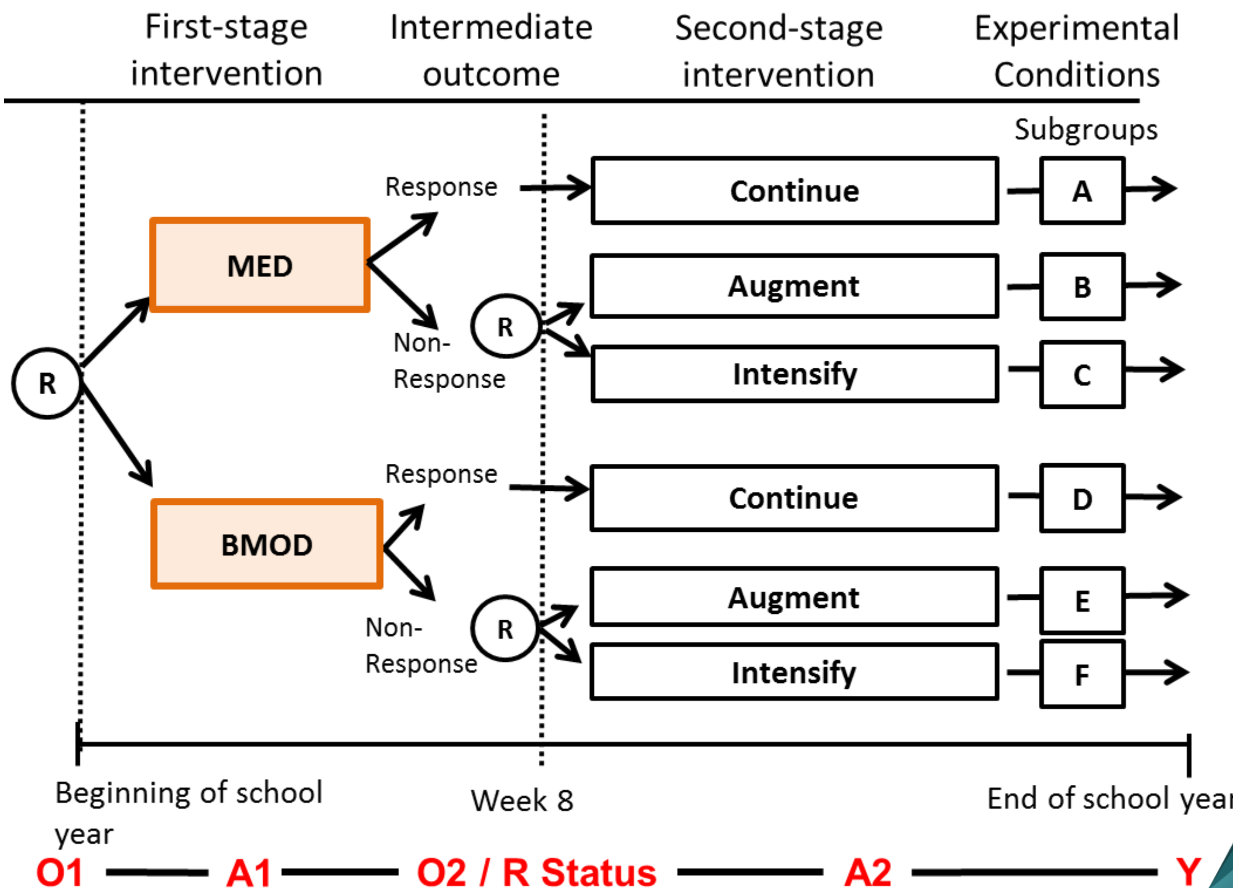
d³lab

ADHD Study (PI: Pelham)



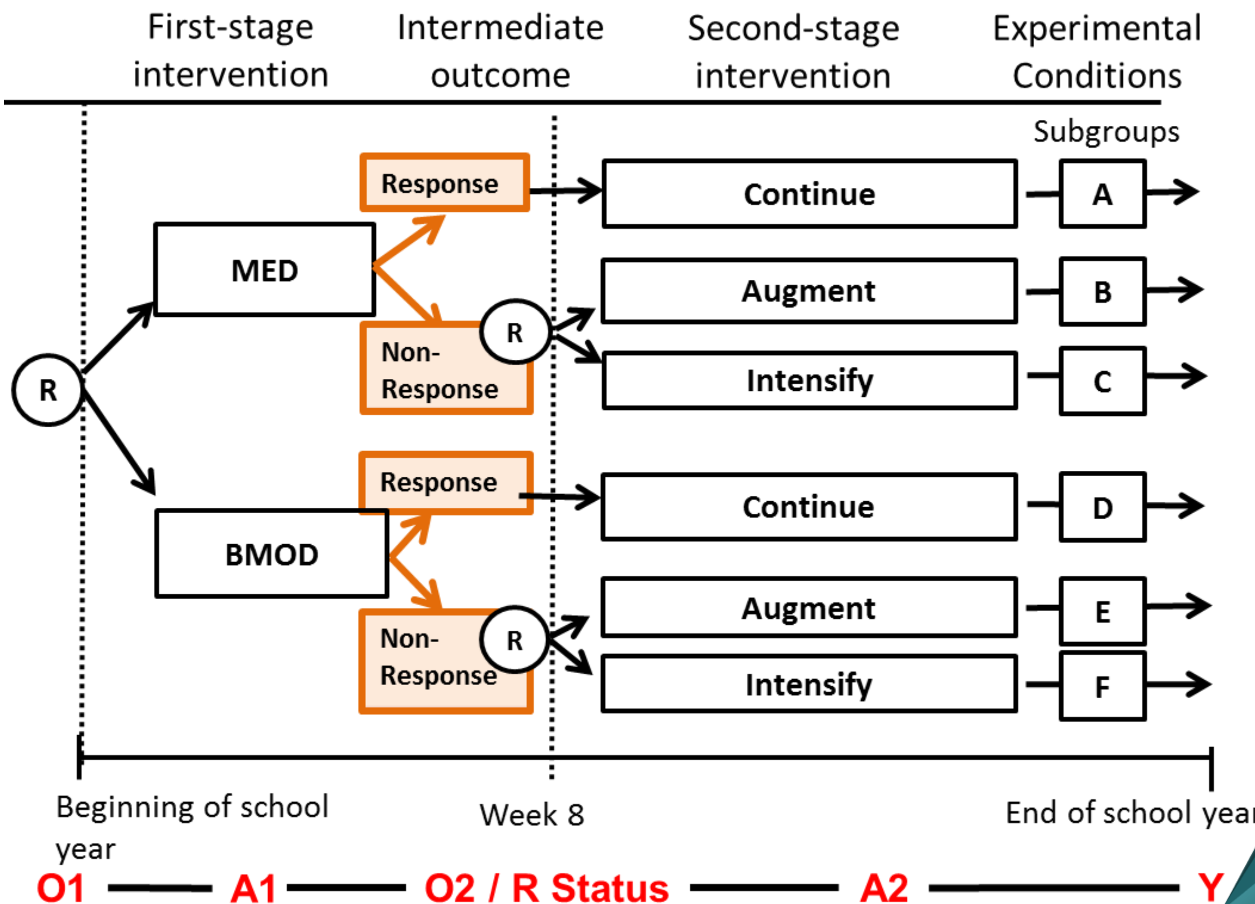
ADHD Study (PI: Pelham)

2 initial intervention options are being compared



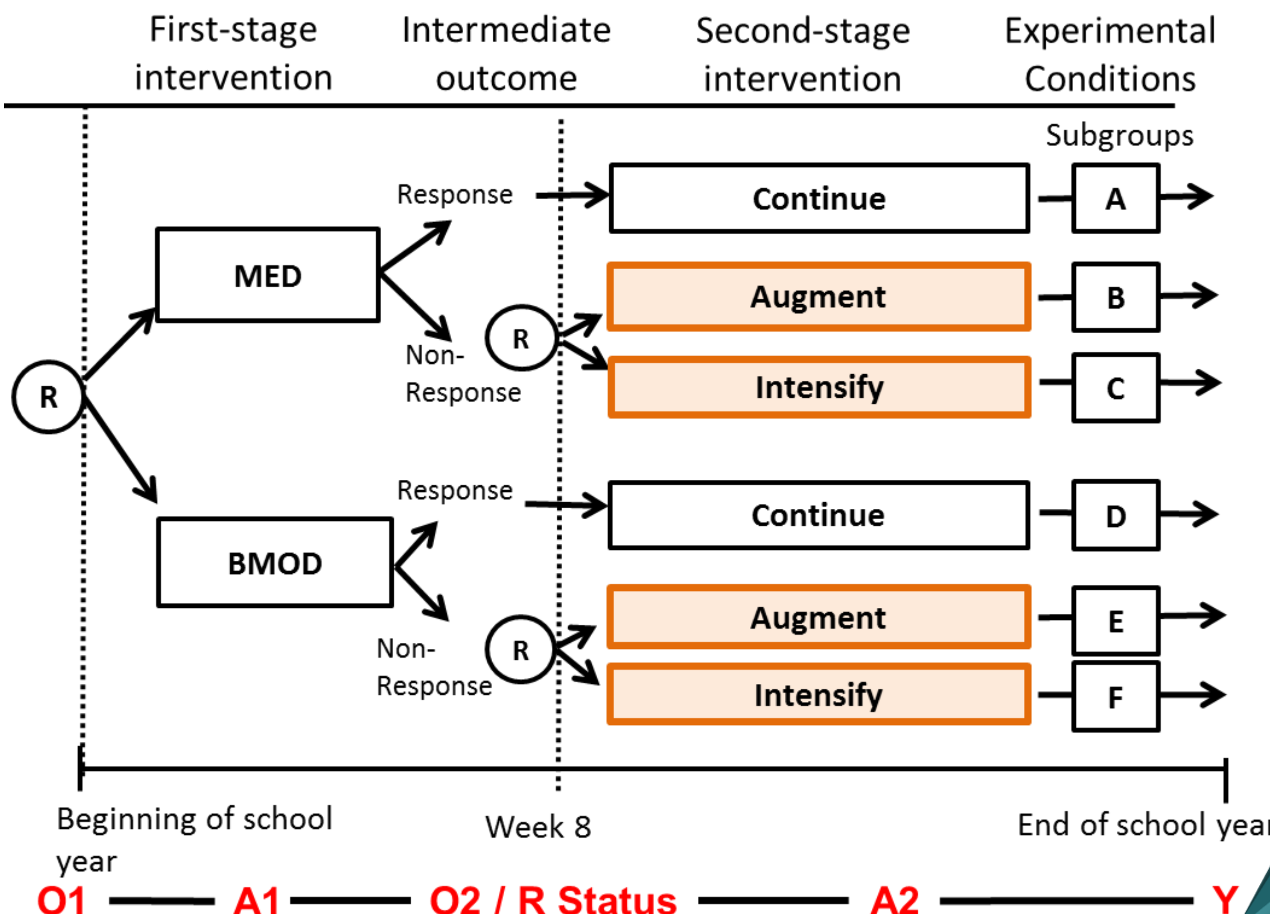
ADHD Study (PI: Pelham)

Embedded tailoring variable: Response status



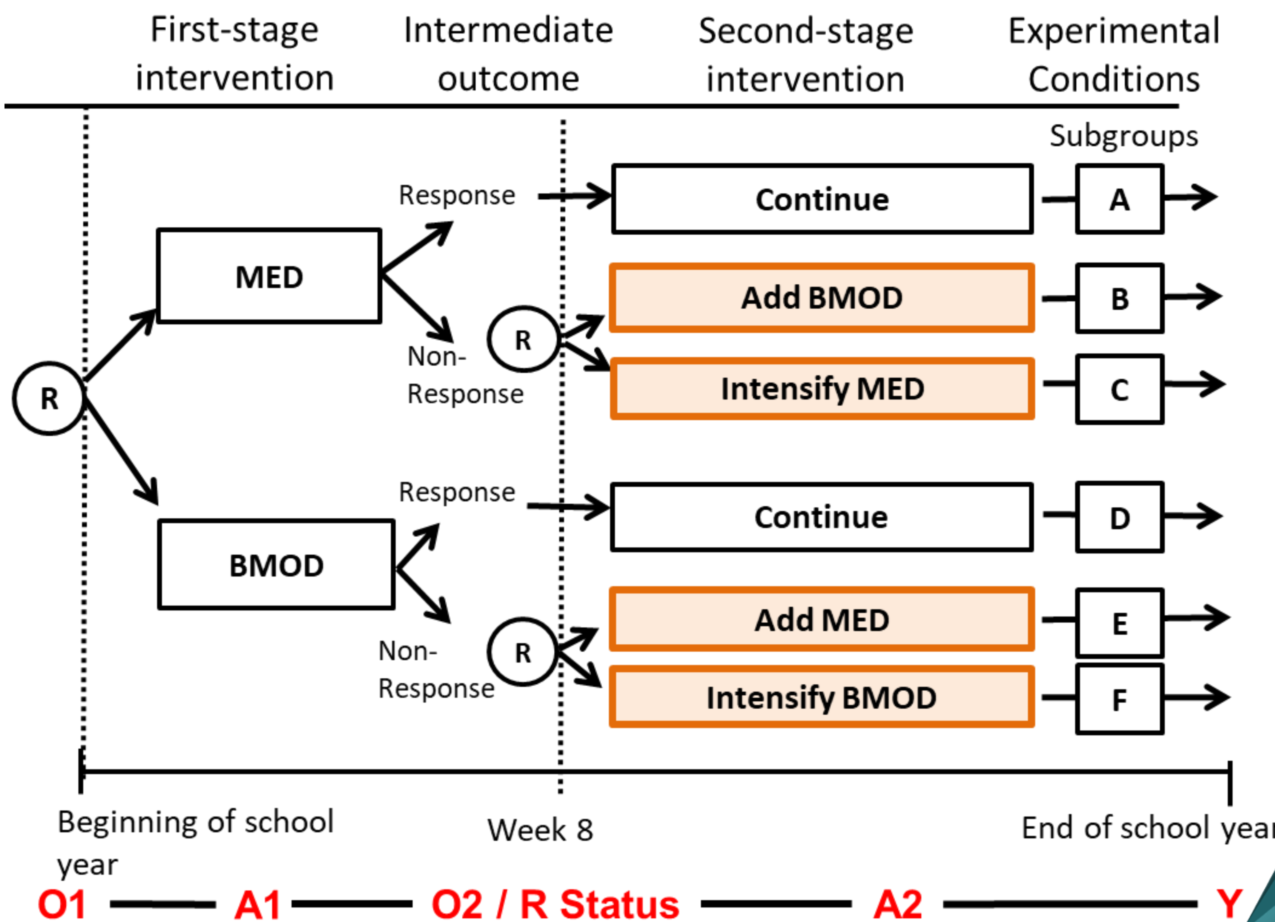
ADHD Study (PI: Pelham)

2 second stage options compared for non-responders



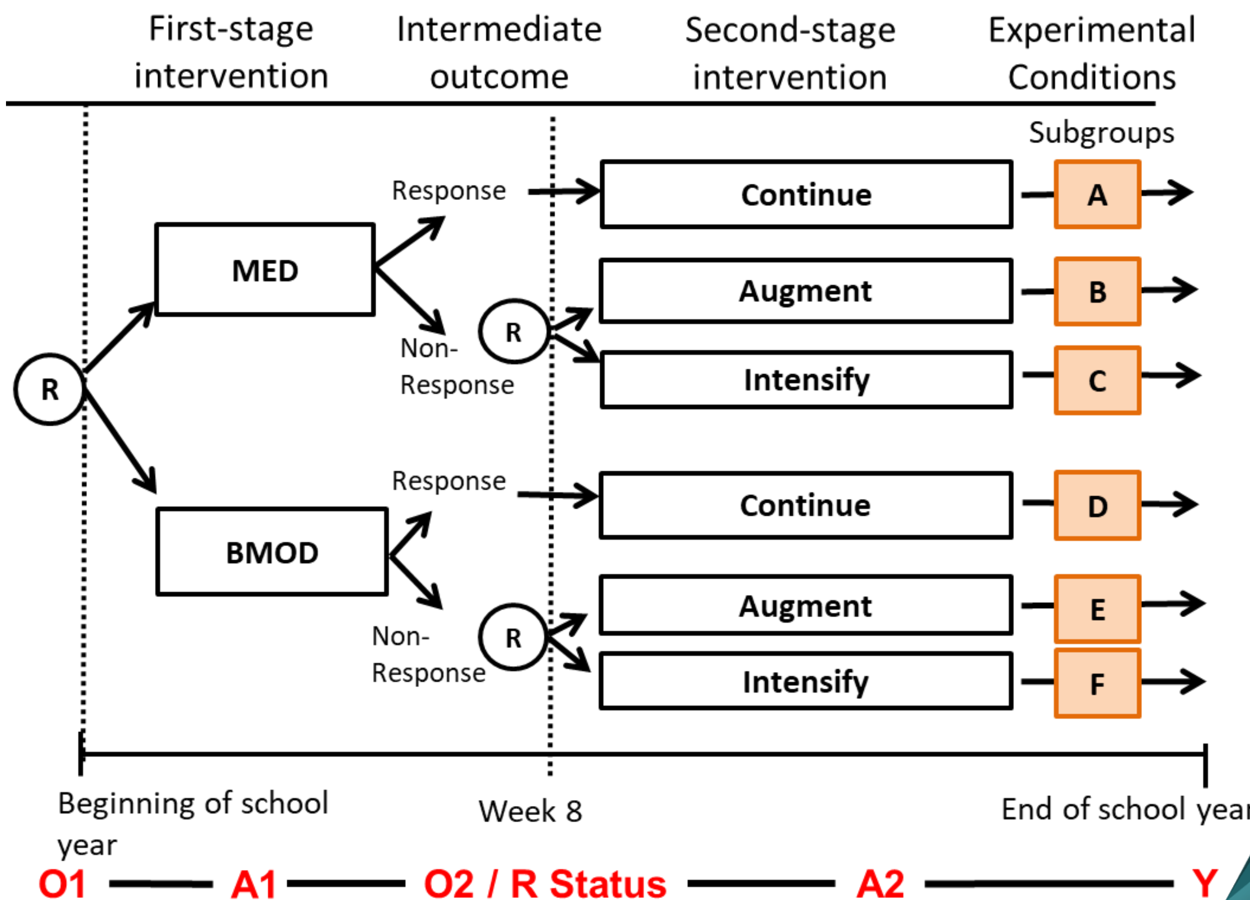
ADHD Study (PI: Pelham)

2 second stage options compared for non-responders



ADHD Study (PI: Pelham)

Total of 6 individual experimental conditions



ADHD Study (PI: Pelham)

4 embedded AIs: #1

At the beginning of school year

Stage 1 = {**MED**},

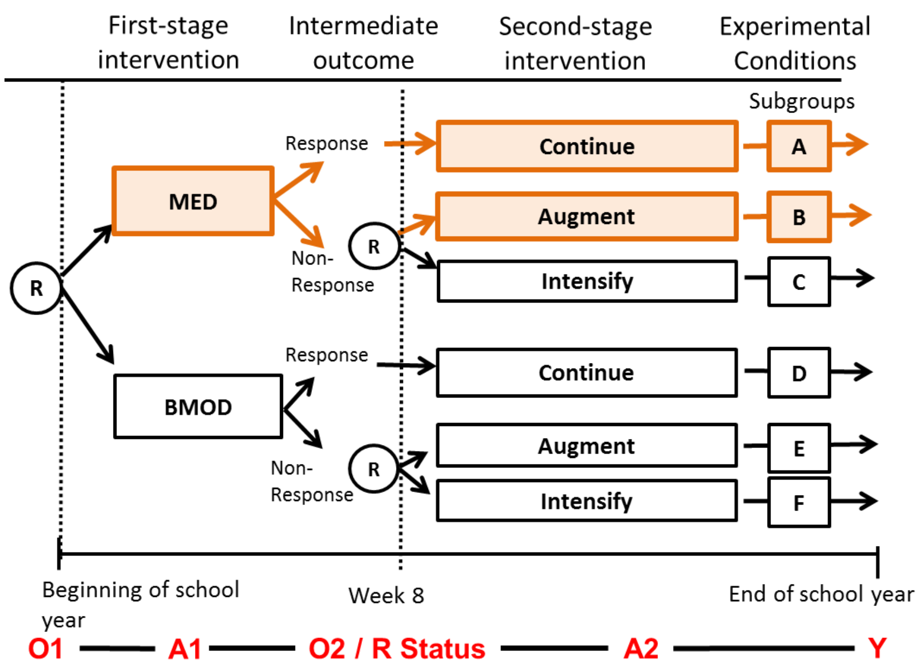
Then, every month, starting week 8

IF response status = {NR}

THEN Stage 2 = {**AUGMENT**}

ELSE IF response status = {R}

THEN **CONTINUE** Stage 1



Notice, AI is not randomized; it is a recommended decision rule

ADHD Study (PI: Pelham)

4 embedded AIs: #2

At the beginning of school year

Stage 1 = {**BMOD**},

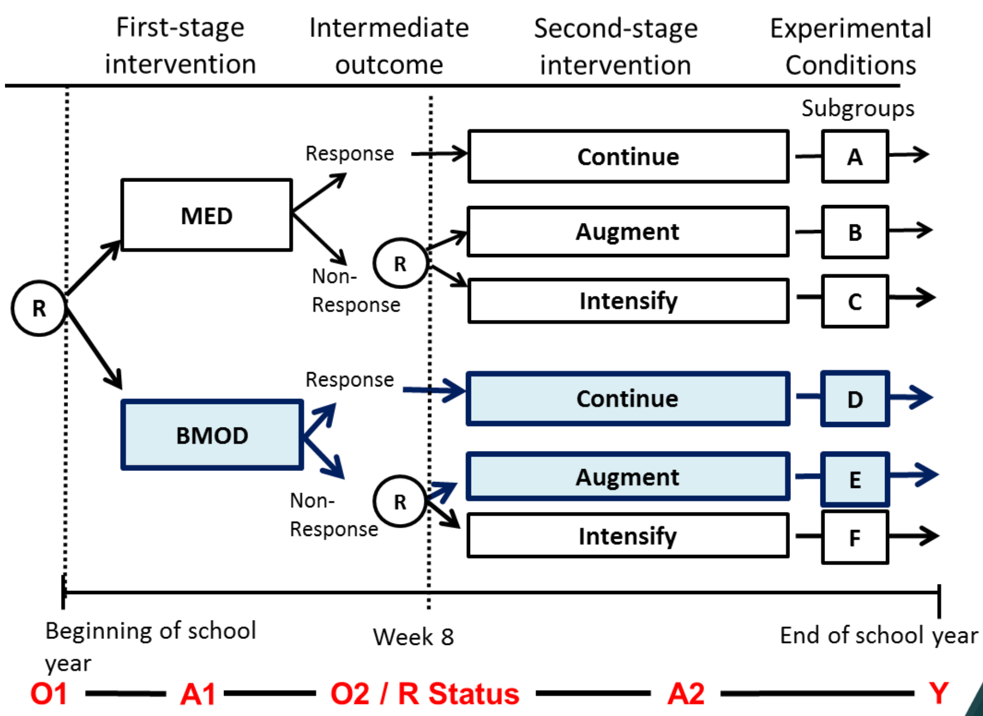
Then, every month, starting week 8

IF response status = {NR}

THEN Stage 2 = {**AUGMENT**}

ELSE IF response status = {R}

THEN **CONTINUE** Stage 1



ADHD Study (PI: Pelham)

4 embedded AIs: #3

At the beginning of school year

Stage 1 = {**MED**},

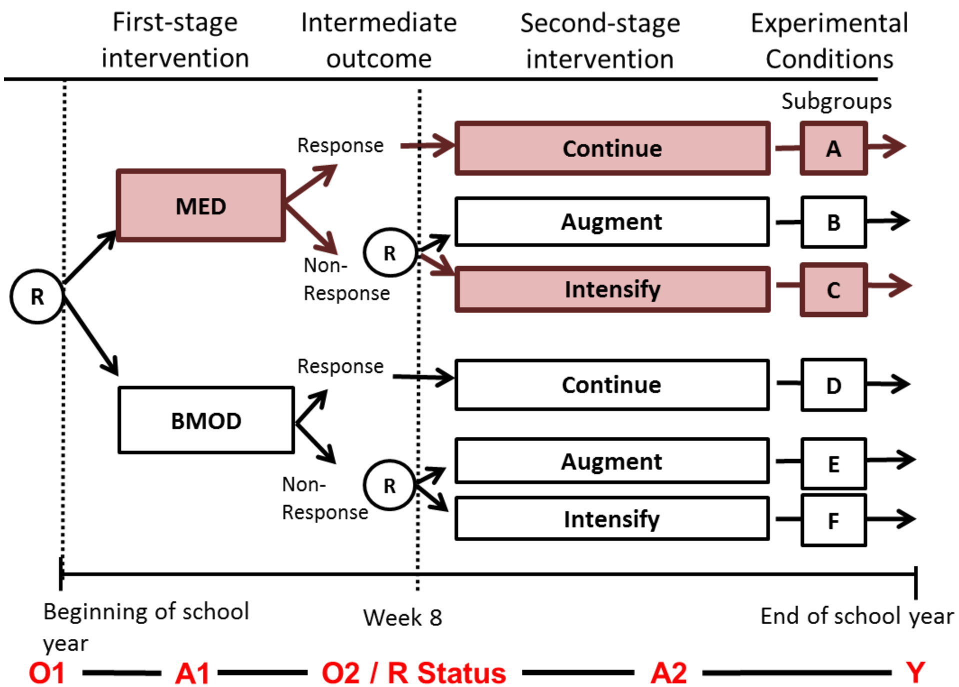
Then, every month, starting week 8

IF response status = {NR}

THEN Stage 2 = {**INTENSIFY**}

ELSE IF response status = {R}

THEN **CONTINUE** Stage 1



ADHD Study (PI: Pelham)

4 embedded AIs: #4

At the beginning of school year

Stage 1 = {**BMOD**},

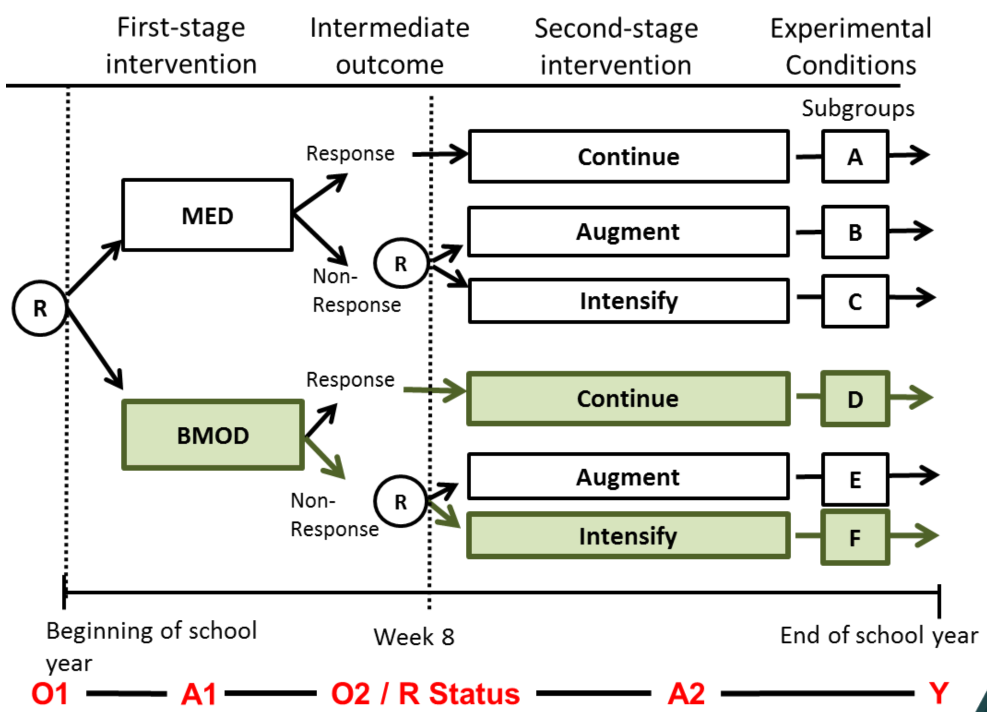
Then, every month, starting week 8

IF response status = {NR}

THEN Stage 2 = {**INTENSIFY**}

ELSE IF response status = {R}

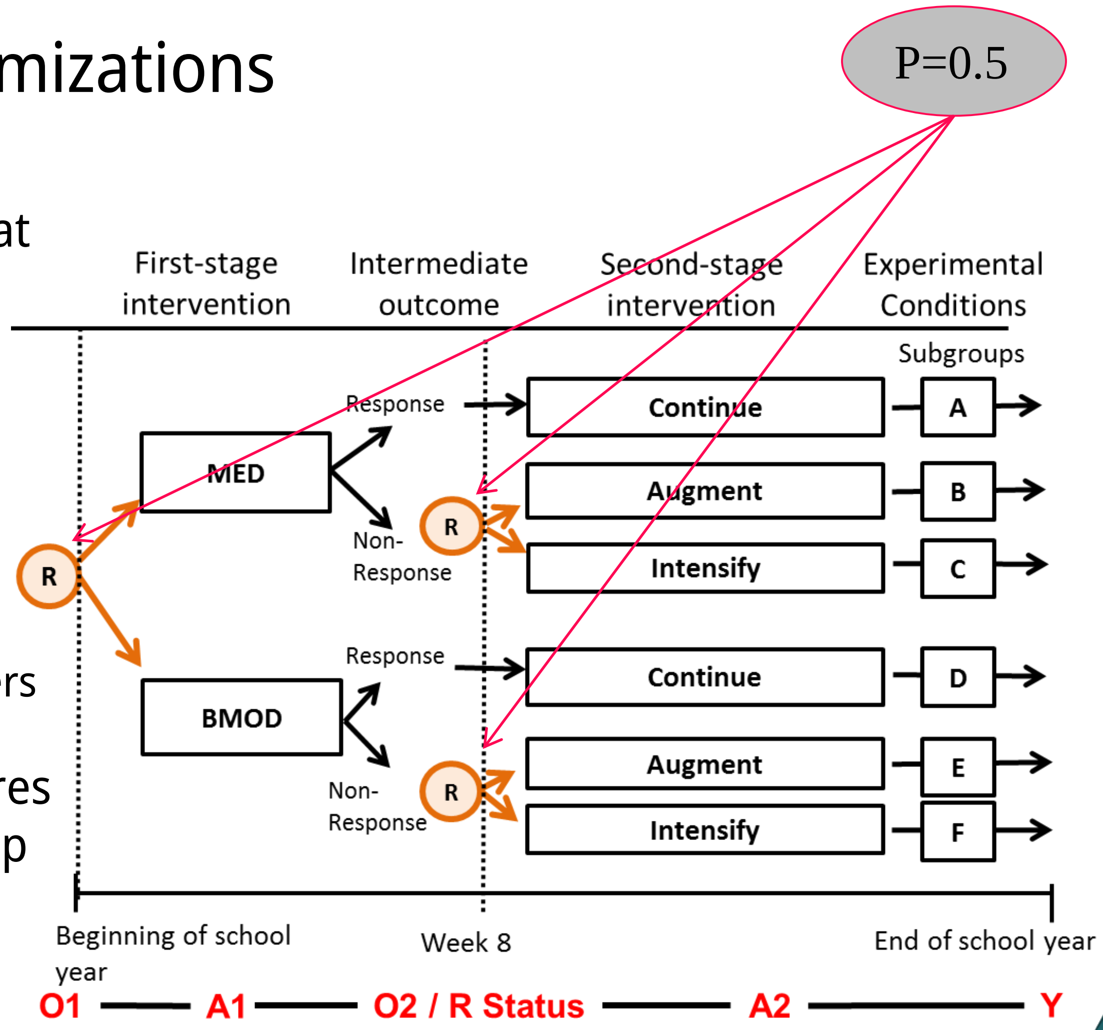
THEN **CONTINUE** Stage 1



ADHD Study (PI: Pelham)

Sequential Randomizations

- Ensures unbiased comparisons of options at each stage.
- No alternative explanations in comparison of
  - First-stage options
  - Second-stage options among non-responders
- Done in a way that ensures between treatment group balance.



d³lab

ADHD Study (PI: Pelham)

What the data looks like, Part I:

	ODD at baseline?	Baseline ADHD Score	Prior Med?	Race	Stage 1 Option
ID	O11	O12	O13	O14	A1
1	1 (YES)	1.18	0 (NO)	1 (White)	-1 (MED)
2	0 (NO)	-0.567	0	0 (Other)	-1
3	0	0.553	1 (YES)	0	1 (BMOD)
4	0	-0.013	0	0	1
5	0	-0.571	1	0	1
6	0	-0.684	1	0	1
7	0	1.169	0	0	-1

*****This data is simulated**

ADHD Study (PI: Pelham)

What the data looks like, Part II:

	Response/ Non-Response	Time until NR (months)	Adherence	Stage 2 Tactic	School Perfm
ID	**R**	**O21**	**O22**	**A2**	**Y**
1	1 (R)	.	0 (NO)	.	3
2	0 (NR)	6	0	1 (INTSFY)	4
3	0	1	1 (YES)	-1 (AUGMENT)	4
4	0	7	0	-1	4
5	0	5	1	1	2
6	0	3	1	-1	4
7	1	.	0	.	3

*****This data is simulated**

d³lab

Outline

- ADHD SMART study
- **Learn how to analyze data from SMART to address two typical primary research questions**

  (a): Main effect of first-stage options

  (b): Main effect of second-stage options/tactics
- Prepare for a third primary aim analysis by

  (c): Learning to estimate the mean outcome under each of the embedded AIs (separately) using an easy-to-use weighting approach.

Typical Primary Aim 1:
Main effect of Stage 1 Options

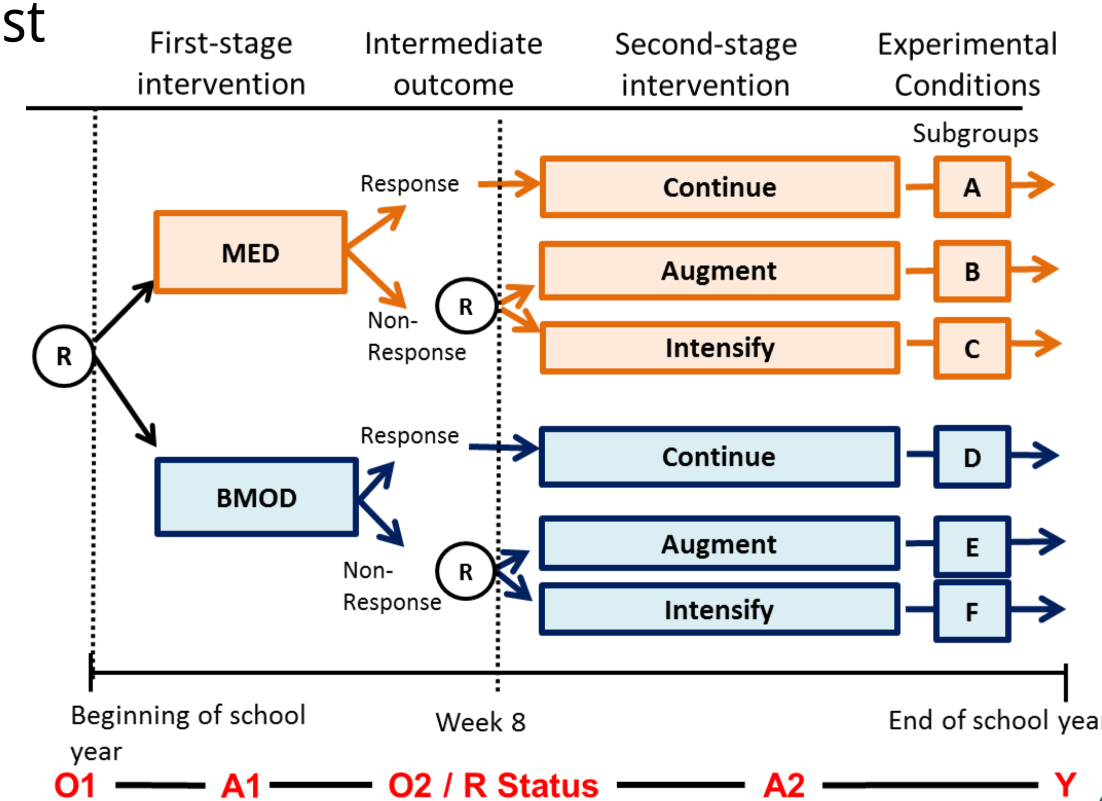
How to frame the question?

1. ***What is the best*** first-line treatment in terms of end of study school performance, controlling for future treatment by design?
2. ***What is the effect*** of starting with BMOD vs with MED in terms of end of study school performance?
3. ***Is it better on average*** to begin treatment with BMOD or with MED, in terms of end of study school performance?

Typical Primary Aim 1:
Main effect of Stage 1 Options

Simply a comparison of two groups:

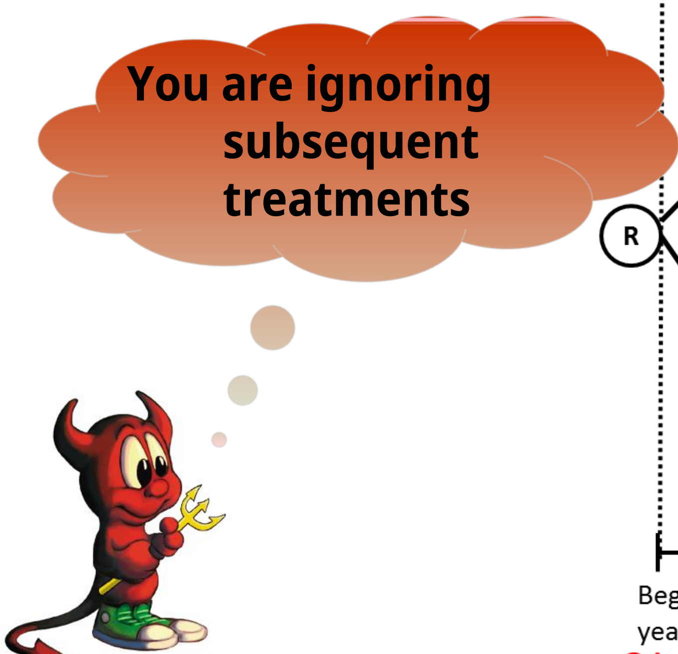
- A two-sample t-test

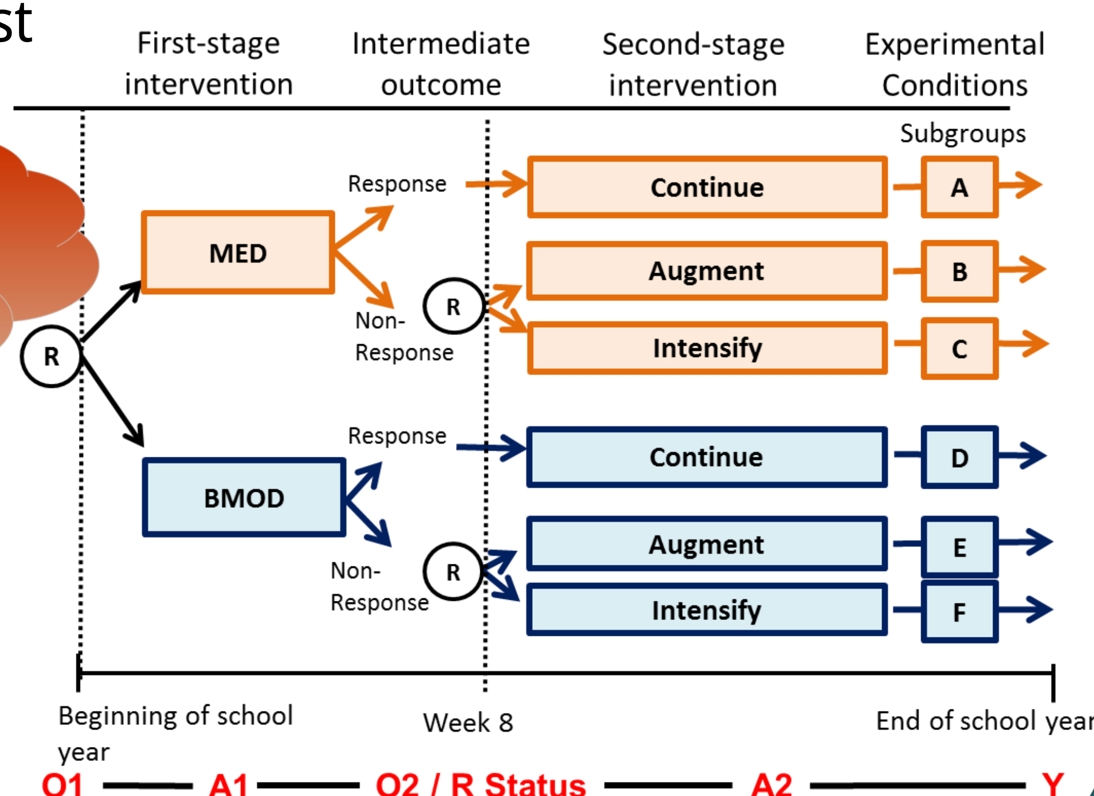


Typical Primary Aim 1:
Main effect of Stage 1 Options

Simply a comparison of two groups:

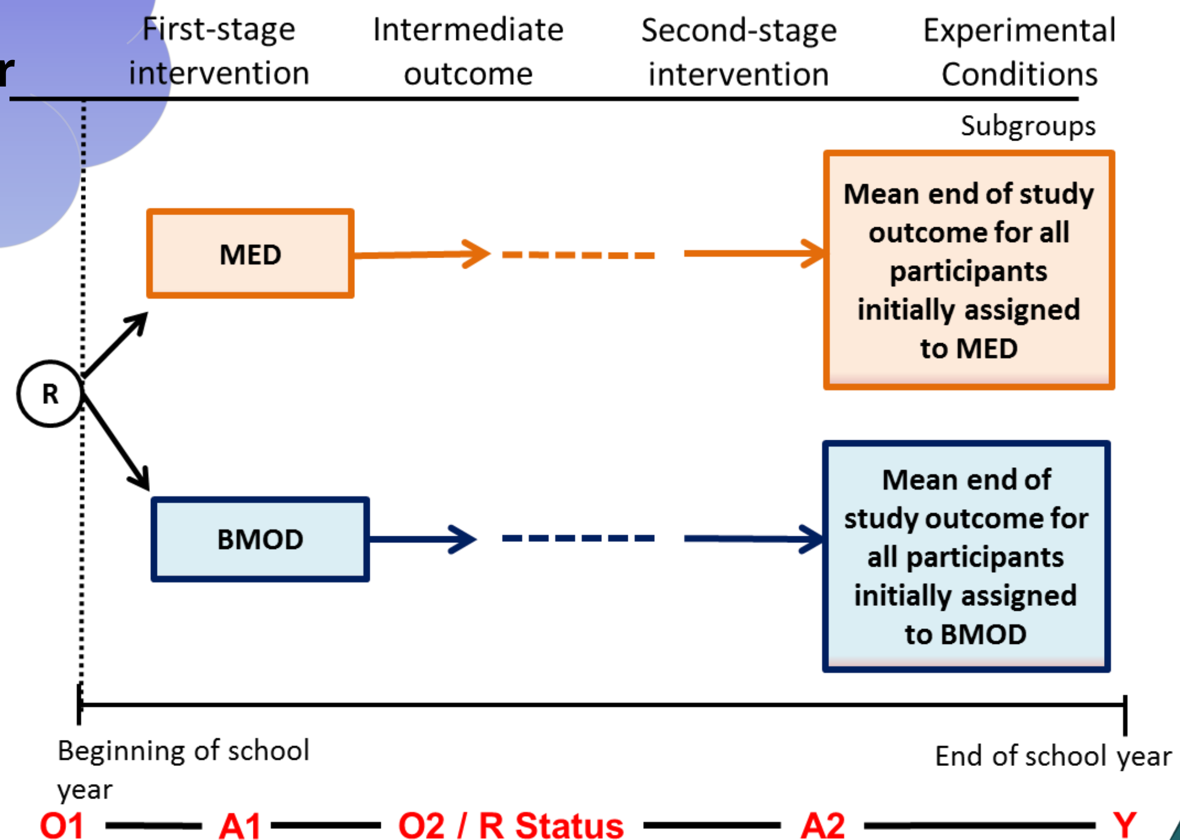
- A two-sample t-test





Typical Primary Aim 1:
Main effect of Stage 1 Options

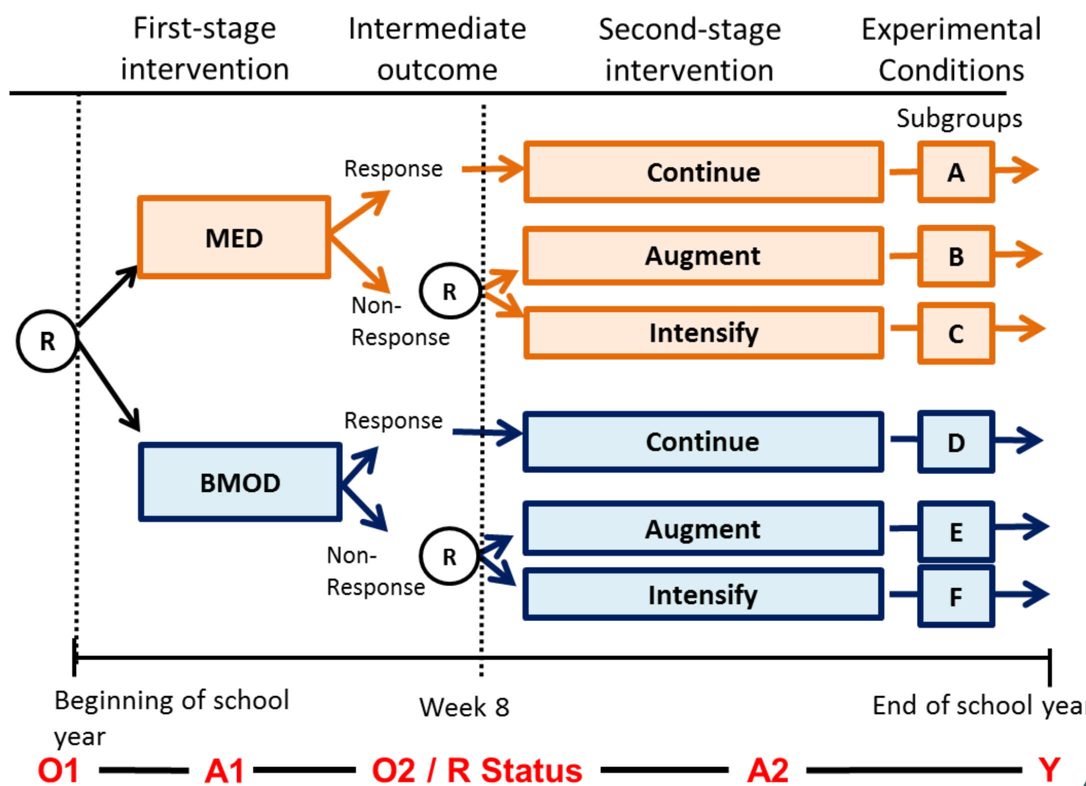
Think about an RCT, where things "happen" after treatment is offered...



Typical Primary Aim 1:
Main effect of Stage 1 Options

Not ignoring;

averaging over!



Before we show you SAS code: Review Coding Scheme

Recall $A_1 = 1$ => BMOD

$A_1 = -1$ => MED

The Regression and Contrast Coding Logic:

$$E[Y|A_1] = b_0 + b_1 A_1$$

c for centered

or you can fit

$$E[Y|A_1, \boldsymbol{O_1}] = b_0 + b_1 A_1 + b_2 O_{11c} + b_3 O_{12c} + b_4 O_{13c} + b_5 O_{14c}$$

Overall Mean Y under BMOD = $b_0 + b_1 \times 1$

Overall Mean Y under MED = $b_0 + b_1 \times (-1)$

Between groups diff = $(b_0 + b_1) - (b_0 - b_1) = 2b_1$

d3lab

Logic for SAS Code

$$E[Y|A_1, \boldsymbol{O_1}] = b_0 + b_1 A_1 + b_2 O_{11c} + b_3 O_{12c} + b_4 O_{13c} + b_5 O_{14c}$$

```
proc genmod data = dat1;
  model Y = A1    O11c O12c O13c O14c;
  estimate 'Mean Y under BMOD' intercept 1 A1  1;
  estimate 'Mean Y under MED'  intercept 1 A1 -1;
  estimate 'Between groups difference'     A1  2;
run;
```

- GENMOD fits generalized linear models-- an extension of traditional linear models
- MODEL statement specifies the outcome, and the independent variables
- ESTIMATE statement enables to estimate linear functions of the parameters

Logic for SAS Code

In ESTIMATE statements, If I leave a coefficient blank, it means I set it to zero.

```
proc genmod data = dat1;
  model Y = A1    O11c O12c O13c O14c;
  estimate 'Mean Y under BMOD' intercept 1 A1  1 O11c 0;
  estimate 'Mean Y under MED'  intercept 1 A1 -1;
  estimate 'Between groups difference'     A1  2;
run;
```

Logic for SAS Code

```
proc genmod data = dat1;
  model Y = A1    O11c O12c O13c O14c;
  estimate 'Mean Y under BMOD' intercept 1 A1  1;
  estimate 'Mean Y under MED'  intercept 1 A1 -1;
  estimate 'Between groups difference'     A1  2;
run;
```

The Regression Logic:

$$E[Y|A_1, \boldsymbol{O_1}] = b_0 + b_1 A_1 + b_2 O_{11c} + b_3 O_{12c} + b_4 O_{13c} + b_5 O_{14c}$$

Overall Mean Y under BMOD $= b_0 + b_1 \times 1$

Overall Mean Y under MED $= b_0 + b_1 \times (-1)$

Between groups diff $= (b_0 + b_1) - (b_0 - b_1) = 2b_1$

Logic for SAS Code

```
proc genmod data = dat1;
  model Y = A1    O11c O12c O13c O14c;
  estimate 'Mean Y under BMOD' intercept 1 A1  1;
  estimate 'Mean Y under MED'  intercept 1 A1 -1;
  estimate 'Between groups difference'     A1  2;
run;
```

<u>The Regression Logic</u>:

$$E[Y|A_1, \boldsymbol{O_1}] = b_0 + b_1 A_1 + b_2 O_{11c} + b_3 O_{12c} + b_4 O_{13c} + b_5 O_{14c}$$

Overall Mean Y under BMOD = $b_0 + b_1 \times 1$

Overall Mean Y under MED = $b_0 + b_1 \times (-1)$

Between groups diff = $(b_0 + b_1) - (b_0 - b_1) = 2b_1$

Logic for SAS Code

```
proc genmod data = dat1;
  model Y = A1    O11c O12c O13c O14c;
  estimate 'Mean Y under BMOD' intercept 1 A1  1;
  estimate 'Mean Y under MED'  intercept 1 A1 -1;
  estimate 'Between groups difference'     A1  2;
run;
```

The Regression Logic:

$$E[Y|A_1, \boldsymbol{O_1}] = b_0 + b_1 A_1 + b_2 O_{11c} + b_3 O_{12c} + b_4 O_{13c} + b_5 O_{14c}$$

Overall Mean Y under BMOD $= b_0 + b_1 \times 1$

Overall Mean Y under MED $= b_0 + b_1 \times (-1)$

Between groups diff $= (b_0 + b_1) - (b_0 - b_1) = 2b_1$

Aim 1 Results

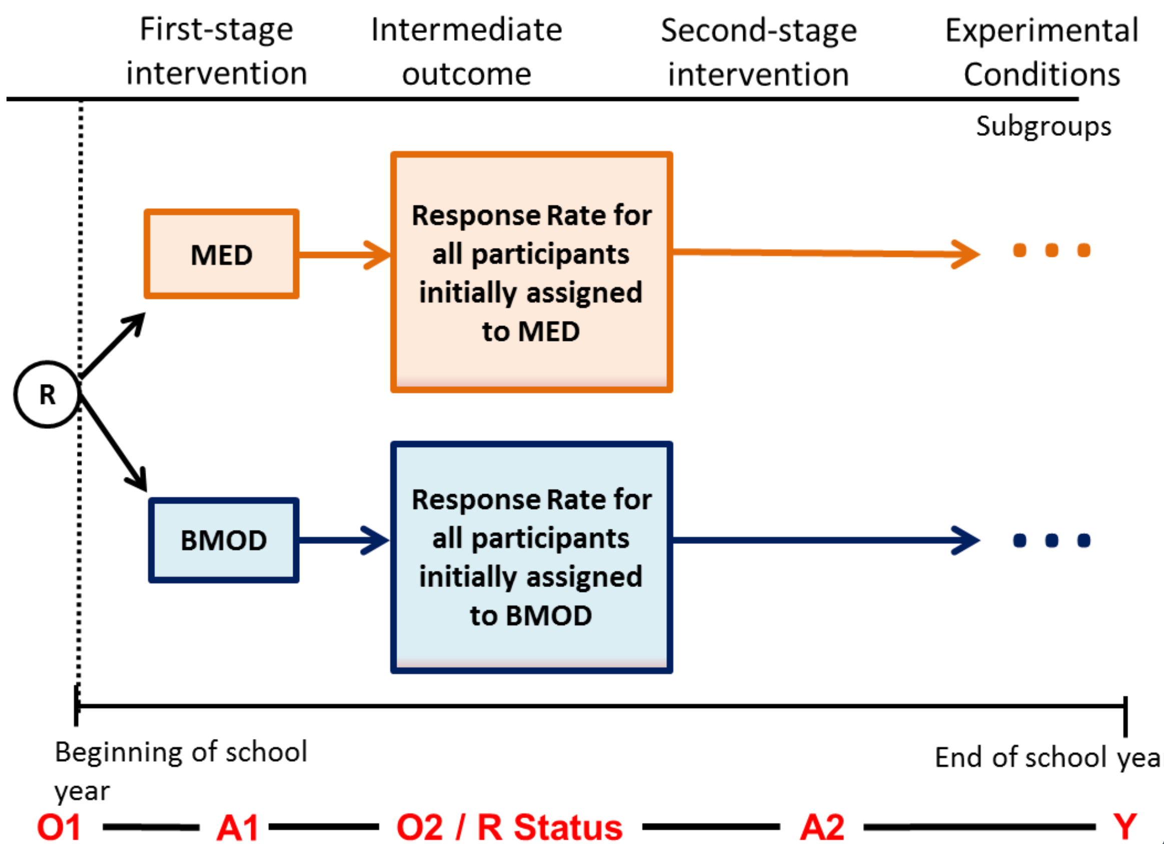
Contrast Estimate Results					
		95% Confidence Limits			
Label	Mean Estimate	Lower	Upper	Standard Error	Pr > ChiSq
Mean Y under BMOD	3.0459	2.7859	3.3059	0.1326	<.0001
Mean Y under MED	2.8608	2.6008	3.1208	0.1326	<.0001
Between groups diff	**0.1851**	-0.1849	0.5551	**0.1888**	**0.3269**

- Results are from simulated dataset
- Slightly better to begin with BMOD (vs MED) in terms of school performance at end of study, but not statistically significant (p-value = 0.33).

***Results are from simulated data

d³lab

Side Analysis:
Effect of Stage 1 Options on NR Rate



Results of Side Analysis
Effect of Stage 1 Options on NR Rate

```
proc freq data=dat1;
  table A1*R / chisq nocol nopercent;
run;
```

Table of A1 by R

A1	R: 0 (non-response)	R: 1 (Response)	Total
-1 (MED)	47 62.67%	28 37.33%	75
1 (BMOD)	52 69.33%	23 30.67%	75
Total	99	51	150

In terms of early response rate, initial MED is slightly better (vs. BMOD) by 7%, but *NS* (p-value = 0.39) .

***This data is simulated

Outline

- ADHD SMART study
- Learn how to analyze data from SMART to address two typical primary research questions

  (a): Main effect of first-stage options

  (b): Main effect of second-stage options/tactics
- Prepare for a third primary aim analysis by

  (c): Learning to estimate the mean outcome under each of the embedded AIs (separately) using an easy-to-use weighting approach.

Typical Primary Aim 2:
Main effect of Stage 2 Tactics

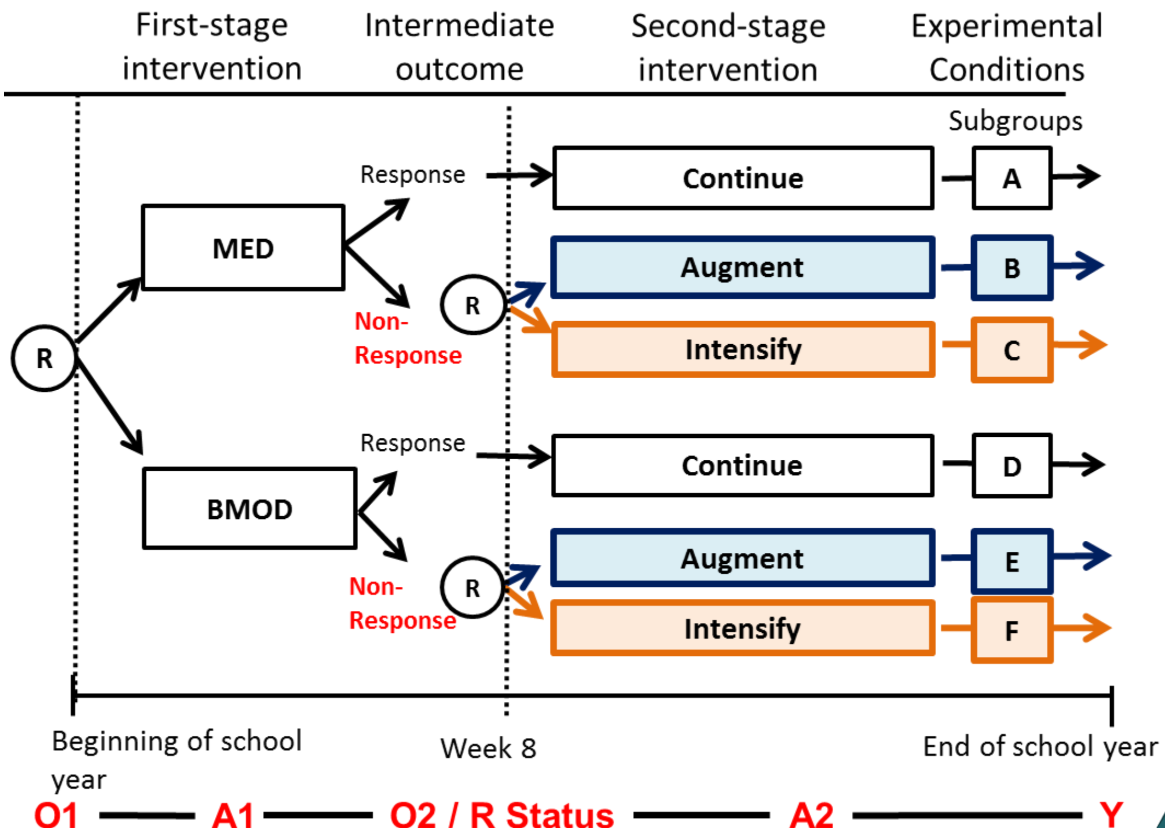
How to frame this question?

To investigate whether, among children who do not respond to (either) first-line treatments, it is better to **INTENSIFY** or **AUGMENT** the initial treatment

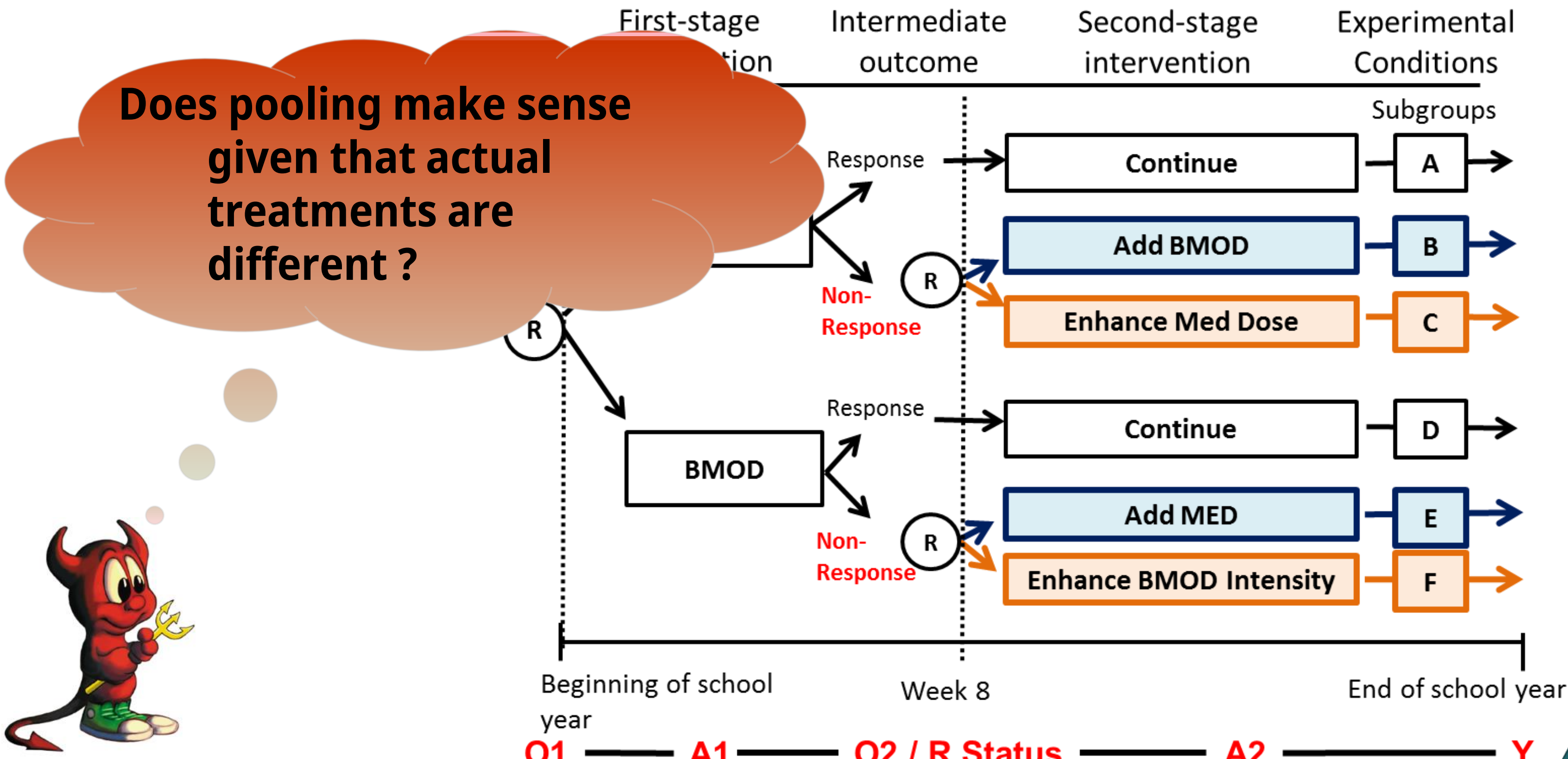
…in terms of end of study school performance

- Regardless of history of treatment
- Controlling for first-stage intervention options

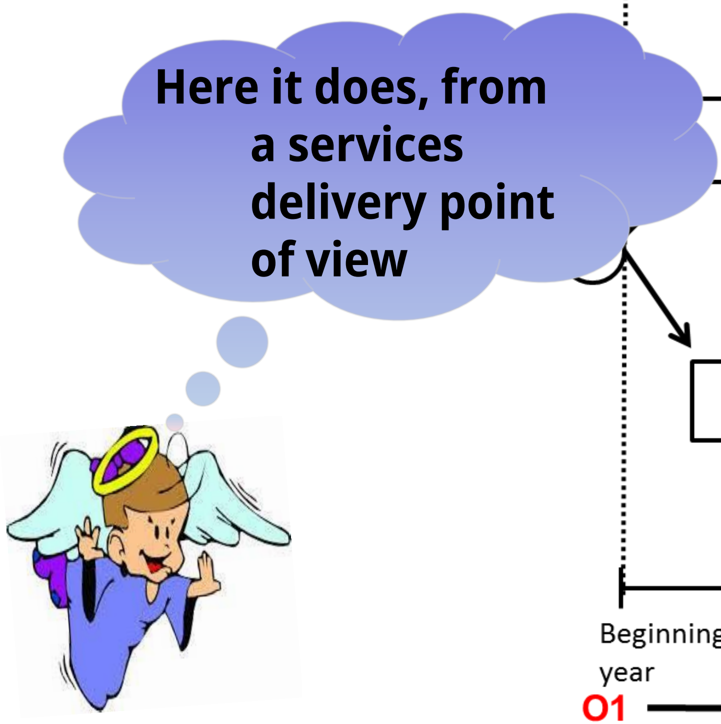
Typical Primary Aim 2:
Main effect of Stage 2 Tactics

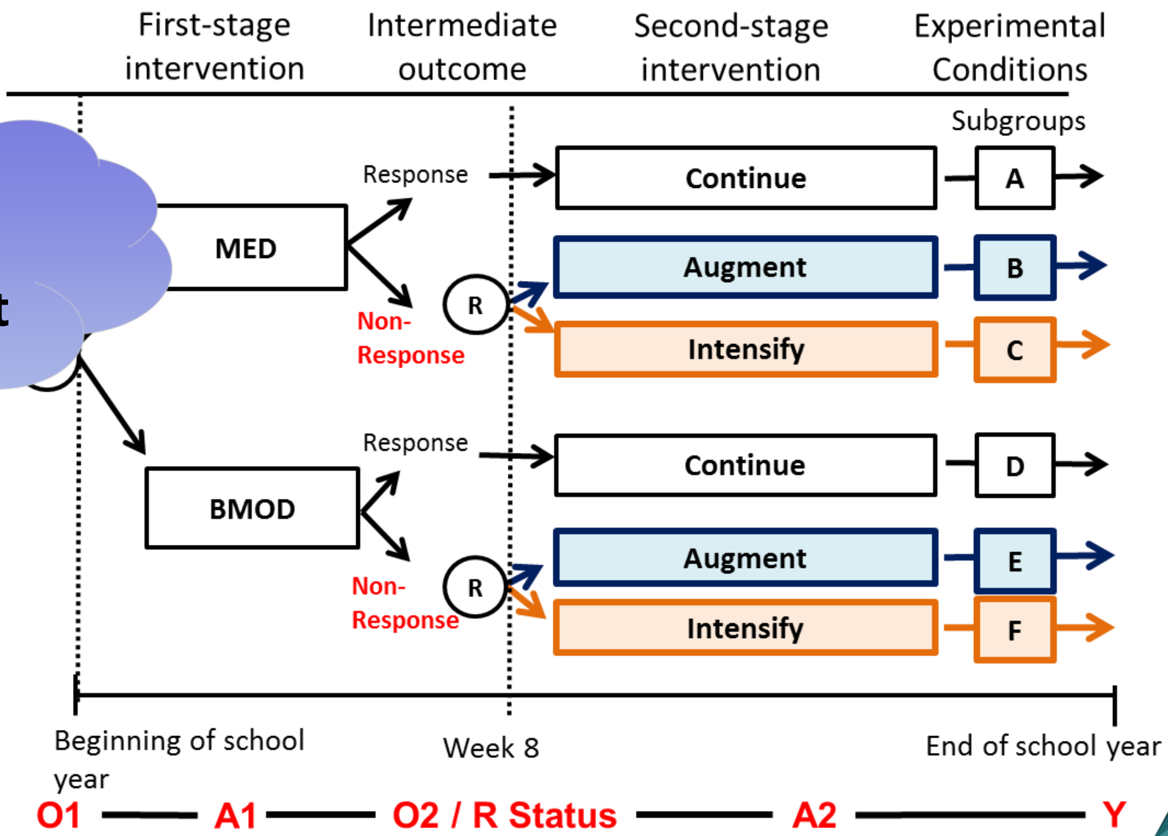


Typical Primary Aim 2:
Main effect of Stage 2 Tactics



Typical Primary Aim 2:
Main effect of Stage 2 Tactics





Before we show you SAS code:
Review Coding Scheme

Recall

$A_2 = 1 \rightarrow$ INTENSIFY

$A_2 = -1 \rightarrow$ AUGMENT

The Regression and Contrast Coding Logic:

$$E[Y|A_2] = b_0 + b_1 A_2$$

or you can fit

$$E[Y|A_2, \boldsymbol{O_1}, \boldsymbol{O_2}] = b_0 + b_1 A_2 + b_2 O_{11c} + b_3 O_{12c} + \cdots + b_6 O_{21c} + b_7 O_{22c}$$

Overall Mean Y under INTENSIFY $= b_0 + b_1 \times 1$

Overall Mean Y under AUGMENT $= b_0 + b_1 \times (-1)$

Between groups diff $= (b_0 + b_1) - (b_0 - b_1) = 2b_1$

SAS Code for Aim 2

NR–> center around the mean of non-responders

```
* use only non-responders;
data dat3;
  set dat2; if R=0;
run;

* run the regression;
proc genmod data = dat3;
       model Y = A2  O11cNR O12cNR O13cNR O14cNR O21cNR O22cNR;
       estimate 'Mean Y INTENSIFY tactic'      intercept 1 A2  1;
       estimate 'Mean Y AUGMENT tactic'        intercept 1 A2 -1;
       estimate 'Between groups difference'                A2  2;
run;
```

SAS Code for Aim 2

```
proc genmod data = dat3;
     model Y = A2 O11cNR O12cNR O13cNR O14cNR O21cNR O22cNR;
     estimate 'Mean Y INTENSIFY tactic'    intercept 1 A2  1;
     estimate 'Mean Y AUGMENT tactic'      intercept 1 A2 -1;
     estimate 'Between groups difference'              A2  2;
run;
```

The Regression Logic:

$$E[Y|A_2, \boldsymbol{O_1}, \boldsymbol{O_2}] = b_0 + b_1 A_2 + b_2 O_{11c} + b_3 O_{12c} + \cdots + b_6 O_{21c} + b_7 O_{22c}$$

Overall Mean Y under INTENSIFY = $b_0 + b_1 \times 1$

Overall Mean Y under AUGMENT = $b_0 + b_1 \times (-1)$

Between groups diff = $(b_0 + b_1) - (b_0 - b_1) = 2b_1$

Aim 2 Results

Contrast Estimate Results					
		95% Confidence Limits			
Label	Mean Estimate	Lower	Upper	Standard Error	Pr > ChiSq
Mean Y INTENSIFY tactic	2.6181	2.3181	2.9182	0.1531	<.0001
Mean Y AUGMENT tactic	3.2060	2.9028	3.5092	0.1547	<.0001
Between groups difference	-0.5879	-1.0206	-0.1552	0.2208	0.0077

- Results are from simulated dataset
- On average, AUGMENT is a better tactic (vs. INTENSIFY) for non-responders to either MED or BMOD in terms of school performance at end of study.
- Difference is statistically significant

***Results are from simulated data

d3lab

Outline

- ADHD SMART study
- Learn how to analyze data from SMART to address two typical primary research questions

  (a): Main effect of first-stage options

  (b): Main effect of second-stage options/tactics
- **Prepare for a third primary aim analysis by**

  (c): Learning to estimate the mean outcome under each of the embedded AIs (separately) using an easy-to-use weighting approach.

Typical Primary Aim 3:
Best of 2 design-embedded AIs

At the beginning of school year

Stage 1 = {**MED**},

Then, every month, starting week 8

IF response status = {NR}

THEN Stage 2 = {**AUGMENT**}

ELSE IF response status = {R}

THEN **CONTINUE** Stage 1

vs.

At the beginning of school year

Stage 1 = {**BMOD**},

Then, every month, starting week 8

IF response status = {NR}

THEN Stage 2 = {**AUGMENT**}

ELSE IF response status = {R}

THEN **CONTINUE** Stage 1

Typical Primary Aim 3:
Best of 2 design-embedded AIs

How to frame this question?

To investigate whether, an AI that recommends to

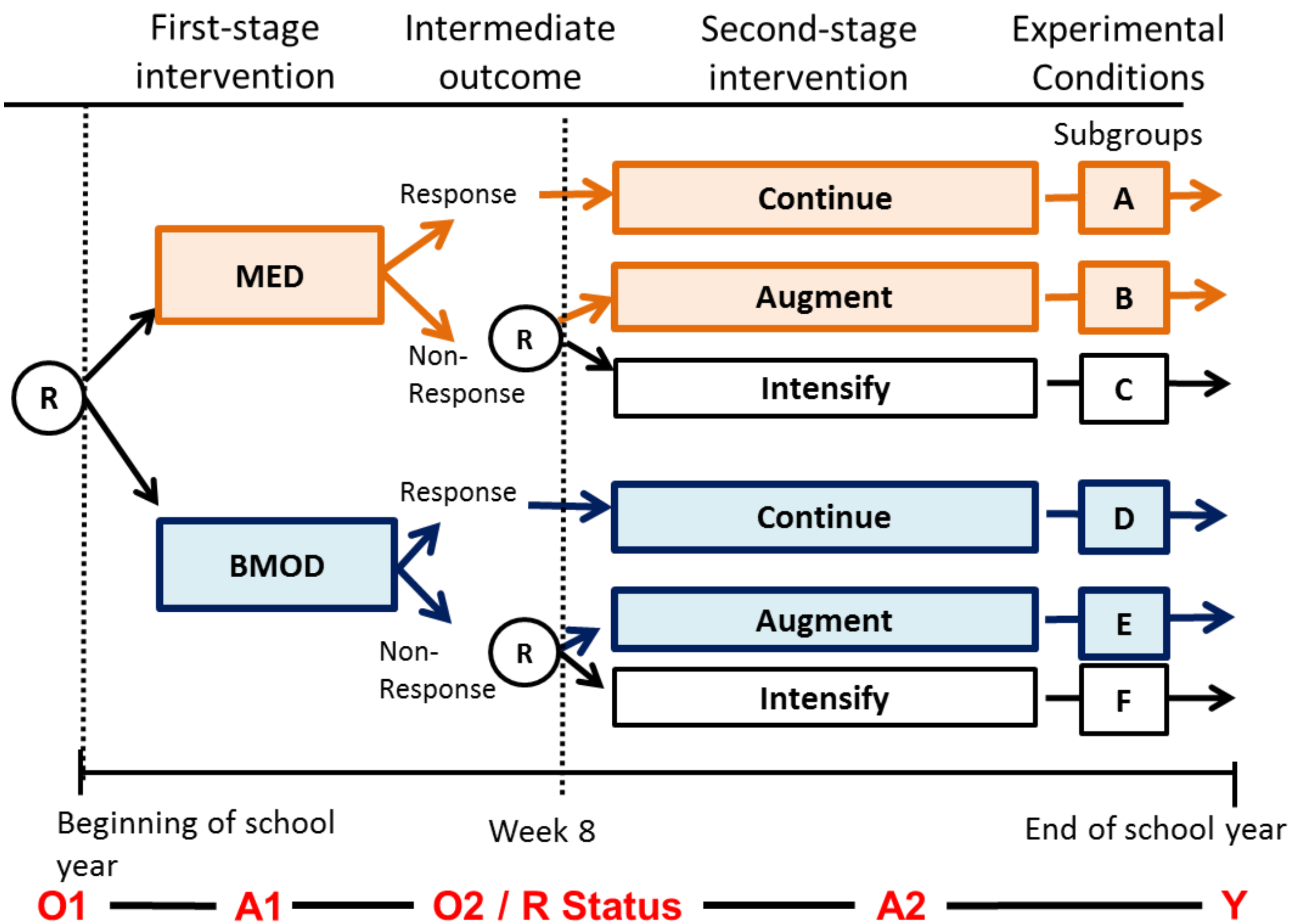
start with BMOD; if non-responder AUGMENT (BMOD+MED), else continue (BMOD),

is better than an AI that recommends to

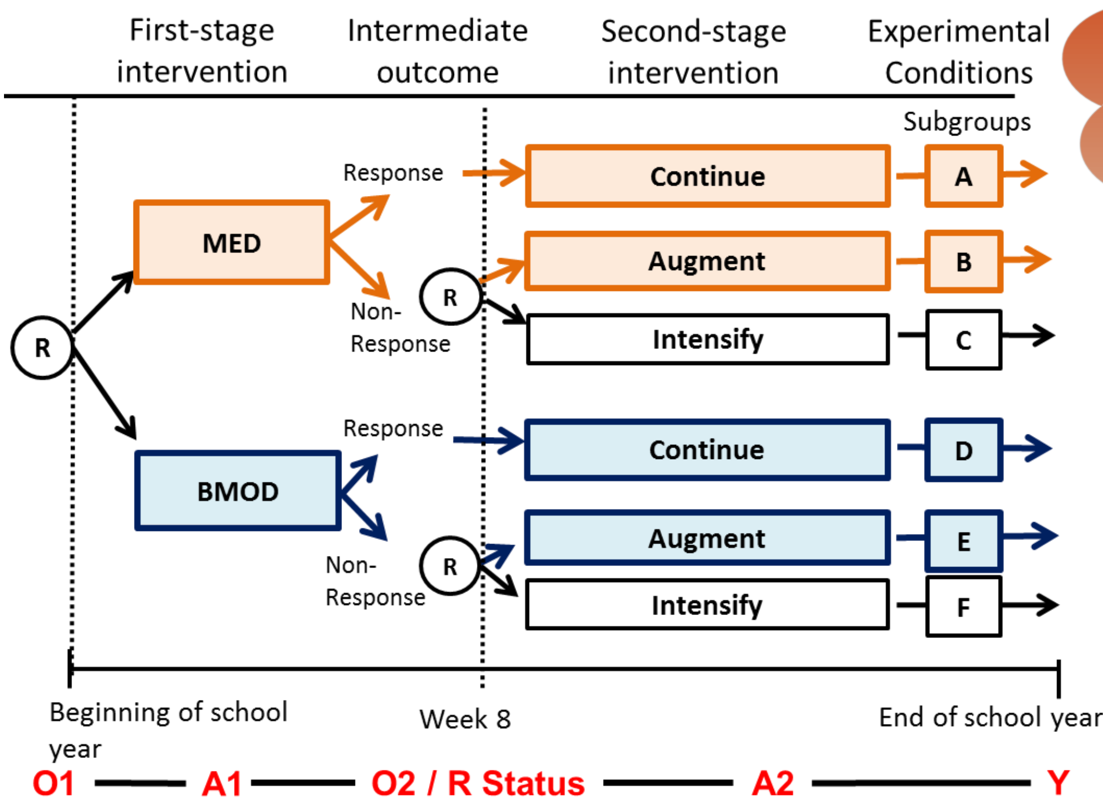
start with MED; if non-responder AUGMENT (BMOD+MED), else continue (MED),

…in terms of end of study school performance.

This is a Comparison of Mean Outcome had Population Followed AI#1 vs. AI#2

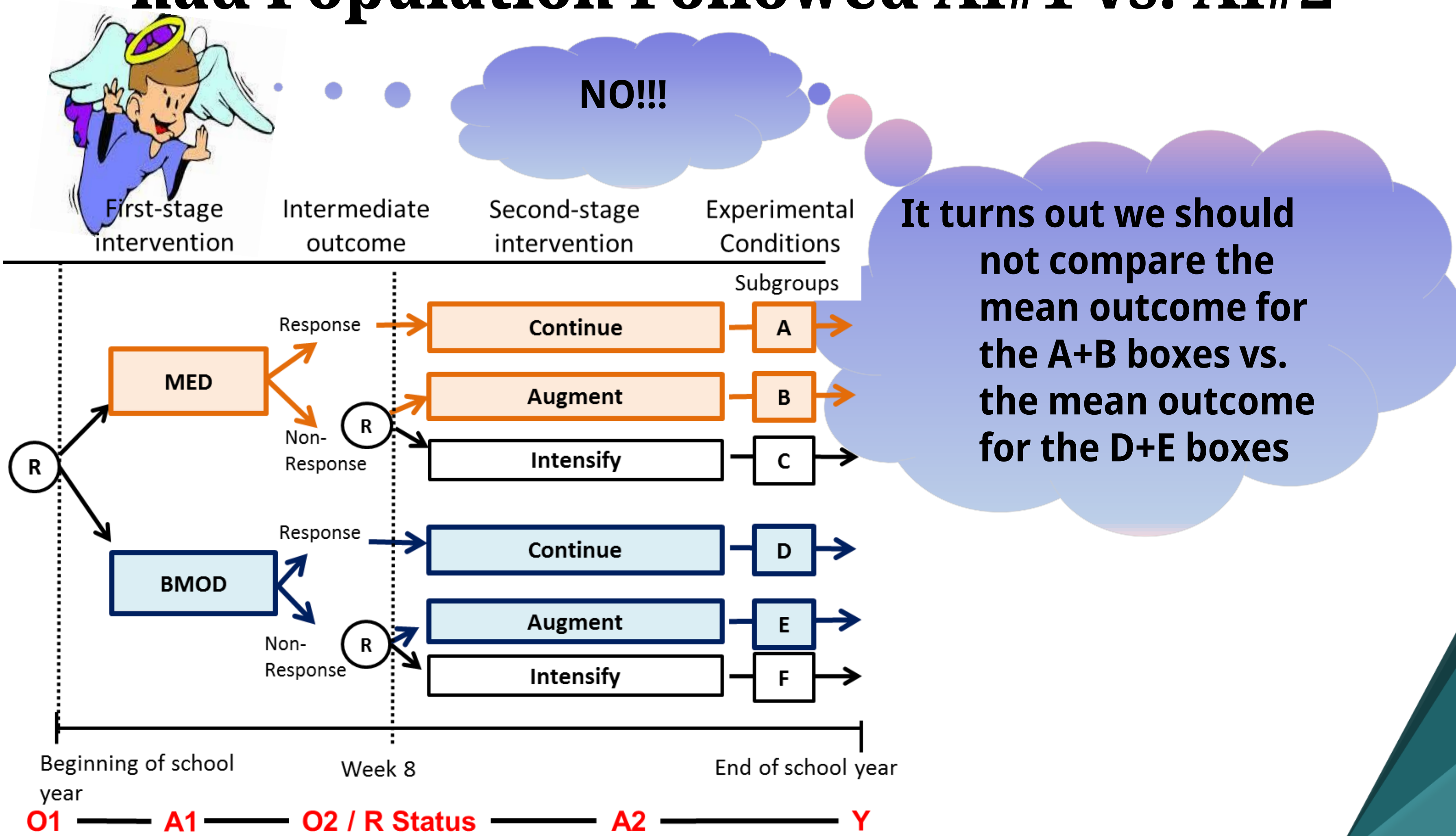


This is a Comparison of Mean Outcome had Population Followed AI#1 vs. AI#2



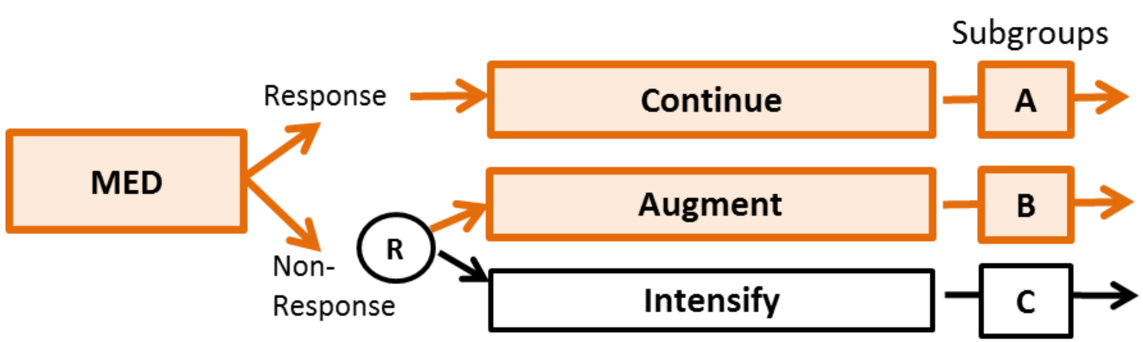
Let's compare the mean outcome for boxes A+B vs. the mean outcome for boxes D+E

This is a Comparison of Mean Outcome had Population Followed AI#1 vs. AI#2

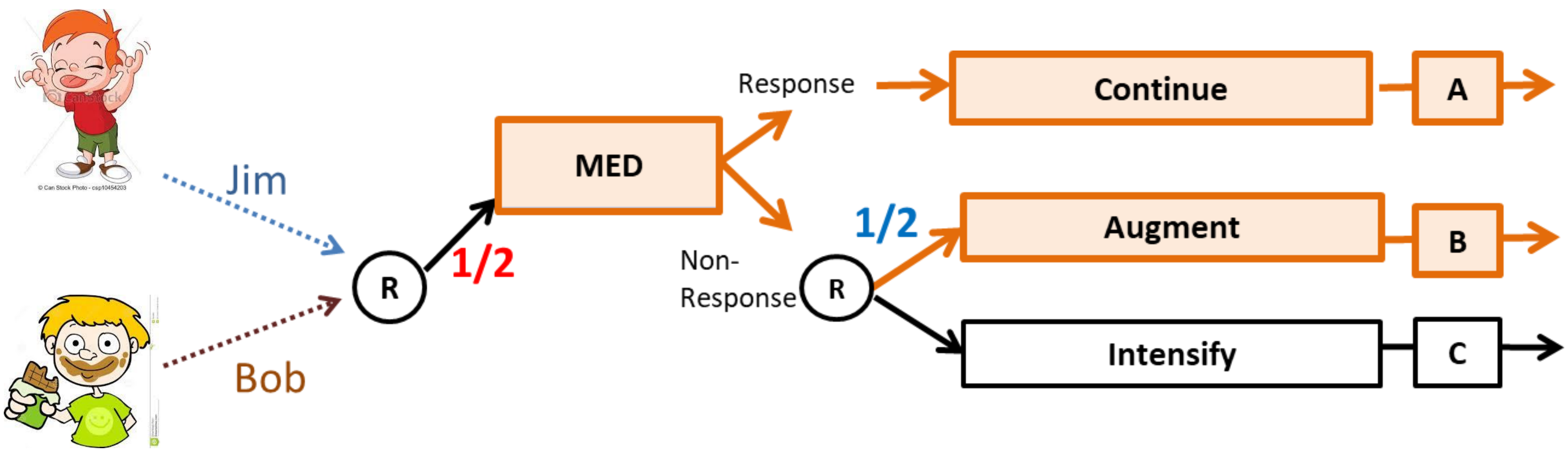


This is a Comparison of Mean Outcome had Population Followed AI#1 vs. AI#2

To understand this, we first, we learn how to obtain mean outcome under AI#1 (MED, AUGMENT)

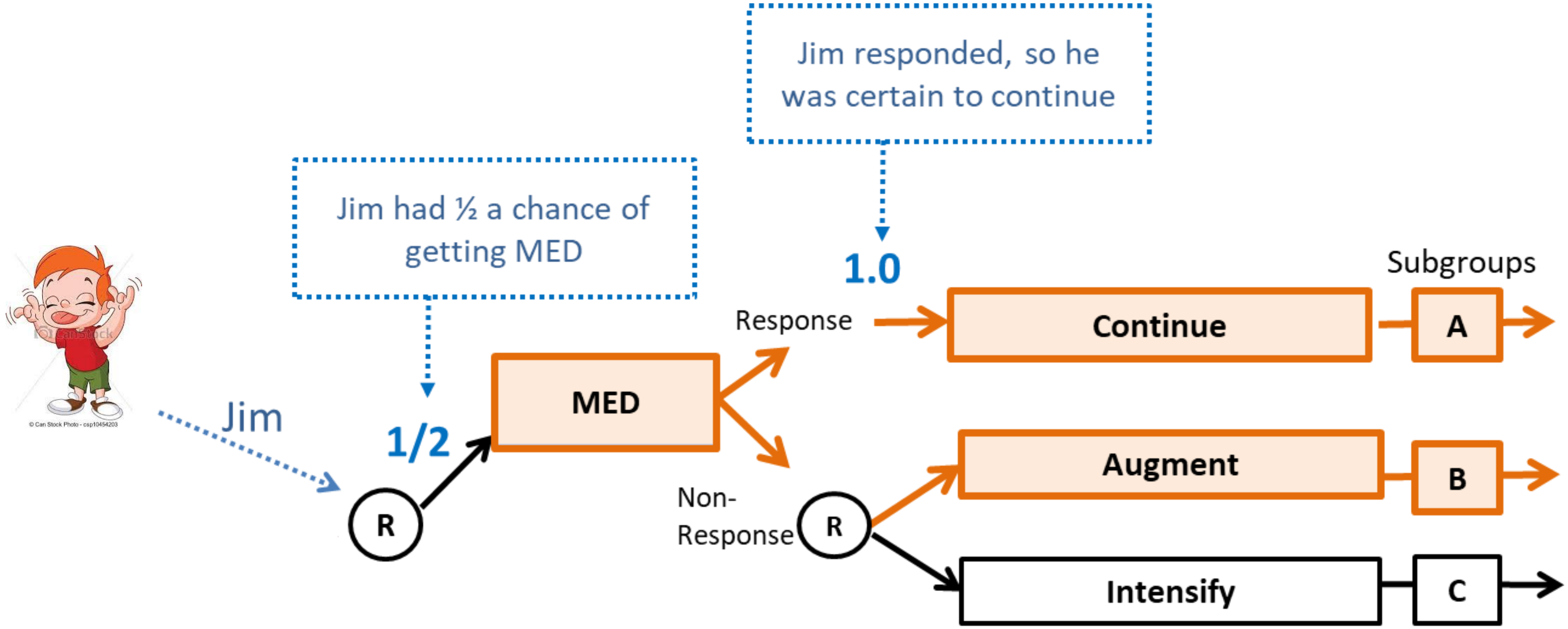


There is Imbalance in the Non/Responding Participants Following this AI



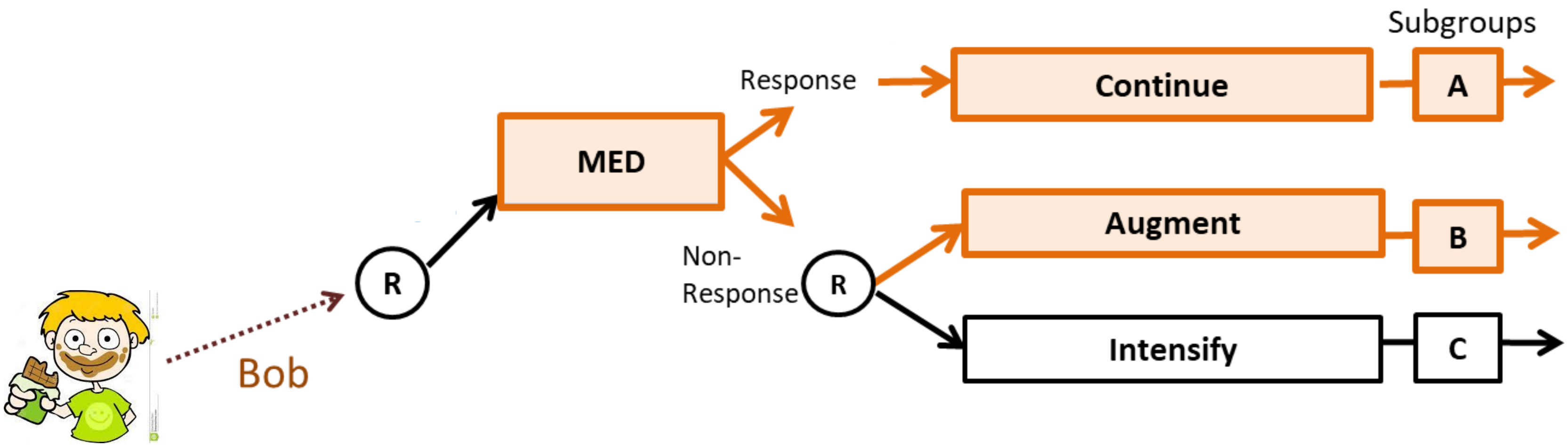
What do you mean by ‘imbalance’?

There is Imbalance in the Non/Responding Participants Following this AI

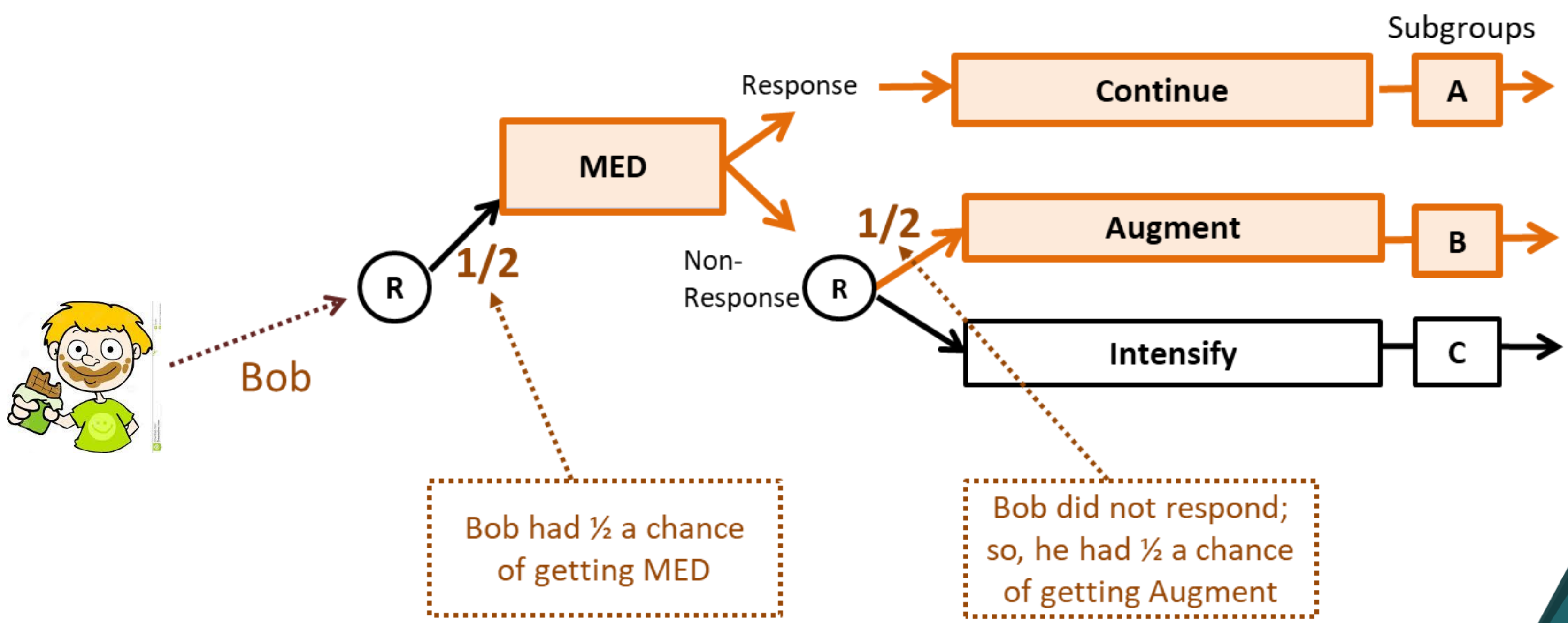


Jim had ½ * 1 = ½ a chance of following AI #1

There is Imbalance in the Non/Responding Participants Following this AI

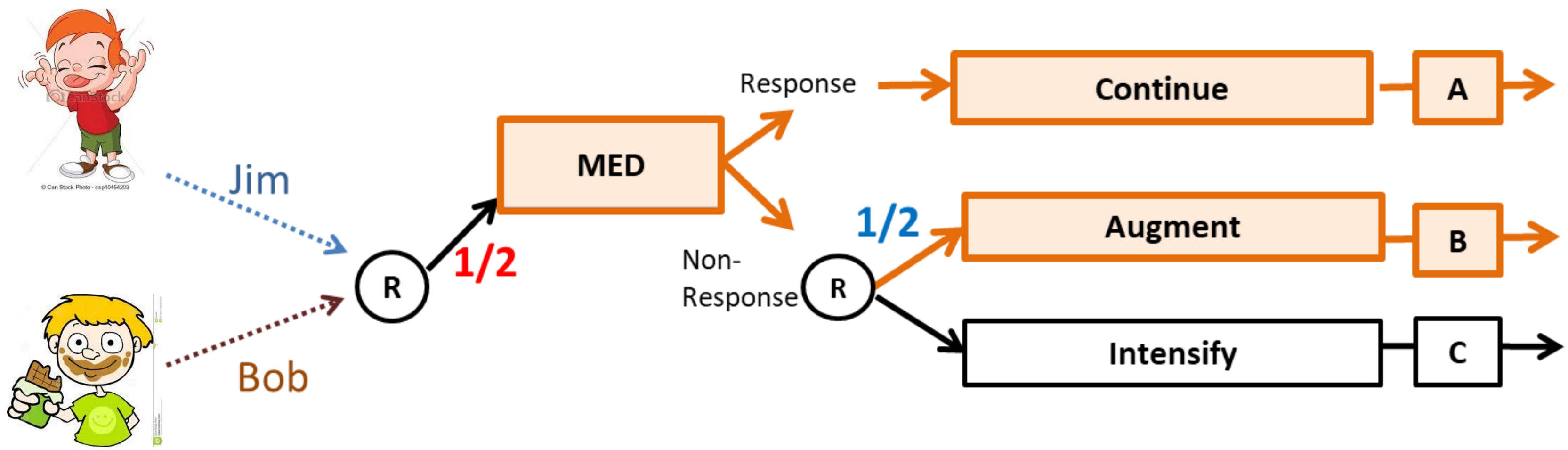


There is Imbalance in the Non/Responding Participants Following this AI



Bob had ½ * ½ = ¼ chance of following AI #1

There is Imbalance in the Non/Responding Participants Following this AI



Jim had ½ * 1 = ½ a chance of following AI #1

Bob had ½ * ½ = ¼ chance of following AI #1

Imbalance

There is Imbalance in the Non/Responding Participants Following this AI

Jim: ½ a chance of following AI #1

Bob: ¼ chance of following AI #1

Imbalance

This imbalance occurs by design,

- Responders had a $^{1}/_{2}$ chance of following AI #1, whereas
- Non-responders had a $^{1}/_{2} \times {}^{1}/_{2} = {}^{1}/_{4}$ chance of following AI #1
- So, we want to estimate mean outcome had all participants followed AI#1
- But, responders are over-represented in this data, by design.
- We want all participants to be equally represented in this data

There is Imbalance in the Non/Responding Participants Following this AI

Responders:
<u>½</u> a chance of following AI #1

Non-Responders:
<u>¼</u> chance of following AI #1

→ **Imbalance**

What can we do? We can fix this imbalance by

- Assigning W = weight = 2 to responders to MED → $2 \times {}^{1}/_{2} = \mathbf{1}$
- Assign W = weight = 4 to non-responders to MED → $4 \times {}^{1}/_{4} = \mathbf{1}$
- This "balances out" the responders and non-responders.
- Then we take W-weighted mean of sample who ended up in the 2 boxes.

SAS Code to Estimate Mean Outcome had all Participants Followed AI#1 (MED, AUGMENT)

First, create an indicator for AI#1 and assign weights.

```
data dat5; set dat2;
  Z1=-1;
  if A1=-1 and R=1             then Z1=1;
  if A1=-1 and R=0 and A2=-1   then Z1=1;
  W=2*R + 4*(1-R);
run;
```

- The indicator Z1 differentiates between participants who followed AI#1 (Z1 = 1) and those who did not (Z1 = -1)
- W will equal 2 if R=1 (responder) and 4 if R=0 (non-responder)

SAS Code to Estimate Mean Outcome had all Participants Followed AI#1 (MED, AUGMENT)

Second, run W-weighted regression: $E[Y|Z_1] = b_0 + b_1 Z_1$.

Mean Y under AI#1: $b_0 + b_1 \times 1$

```
proc genmod data = dat5;
  class id;
  model Y = Z1;
  weight W;
  repeated subject = id / type = ind;
  estimate 'Mean Y under AI#1' intercept 1 Z1  1;
run;
```

<u>This is how we ask SAS to provide robust standard errors</u>:
Why do we need that?
Weights depend on response status, which is unknown ahead of time.
Robust SE account for this uncertainty (i.e., for sampling error in the "estimation" of the weights).

Results for Estimated Mean Outcome had All Participants Followed AI#1 (MED, AUGMENT)

Analysis Of GEE Parameter Estimates

Parameter	Estimate	Standard Error	Pr > |Z|
Intercept	2.9153	0.1084	<.0001
Z1	-0.0504	0.1084	0.6417

Contrast Estimate Results

Label	Mean Estimate	95% Confidence Limits		Standard Error	Pr > ChiSq
		Lower	Upper		
Mean Y under AI #1 (MED, AUGMENT)	**2.8649**	2.5305	3.1992	**0.1706**	**<.0001**

***This analysis is with simulated data

d³lab

Citations

- Murphy, S. A. (2005). An experimental design for the development of adaptive intervention. *Statistics in Medicine*, 24, 455-1481.
- Nahum-Shani, I., Qian, M., Almirall, D., Pelham, W. E., Gnagy, B., Fabiano, G. A., … & Murphy, S. A. (2012). Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychological methods*, *17*(4), 457.