Using Data Arising from a SMART to Address Primary Aims (Part II)

Module 4

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

Review

- ADHD SMART study
- Weighted regression approach for estimating the mean outcome under one AI

Learn

- Use weighted regression to compare the mean outcomes for two AIs that begin with different treatments
- Use weighted-and-replicated regression to simultaneously compare all embedded AIs in a SMART

Outline

Review

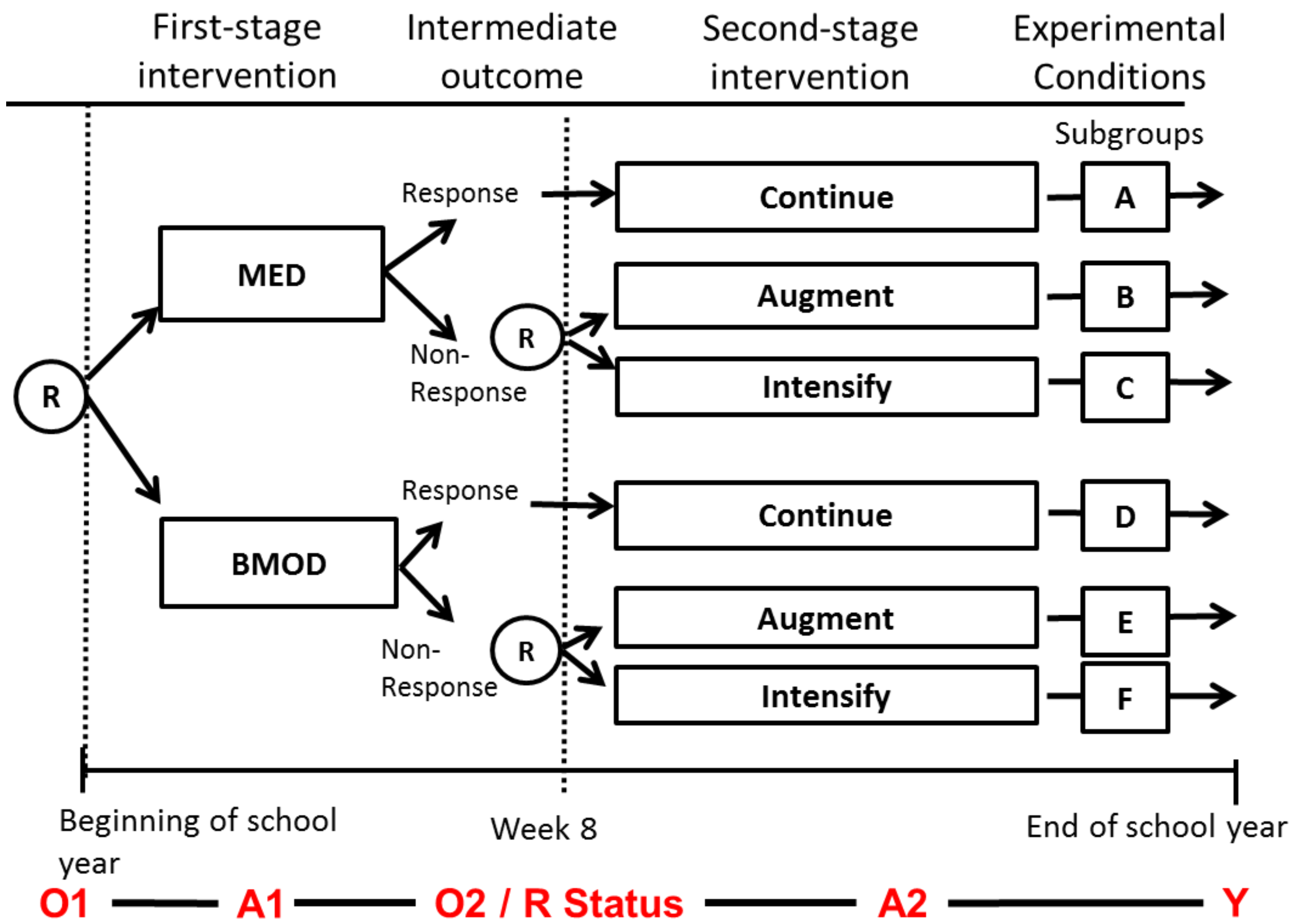
- **ADHD SMART study**
- **Weighted regression approach for estimating the mean outcome under one AI**

Learn

- Use weighted regression to compare the mean outcomes for two AIs that begin with different treatments
- Use weighted-and-replicated regression to simultaneously compare all embedded AIs in a SMART

ADHD SMART

PI: Pelham



ADHD SMART

PI: Pelham

4 embedded adaptive interventions

AI #1:
Start with MED;
if non-responder AUGMENT, else CONTINUE

AI #2:
Start with BMOD;
if non-responder AUGMENT, else CONTINUE

AI #3:
Start with MED;
if non-responder INTENSIFY, else CONTINUE

AI #4:
Start with BMOD;
if non-responder INTENSIFY, else CONTINUE

Recall Typical Primary Aim 3

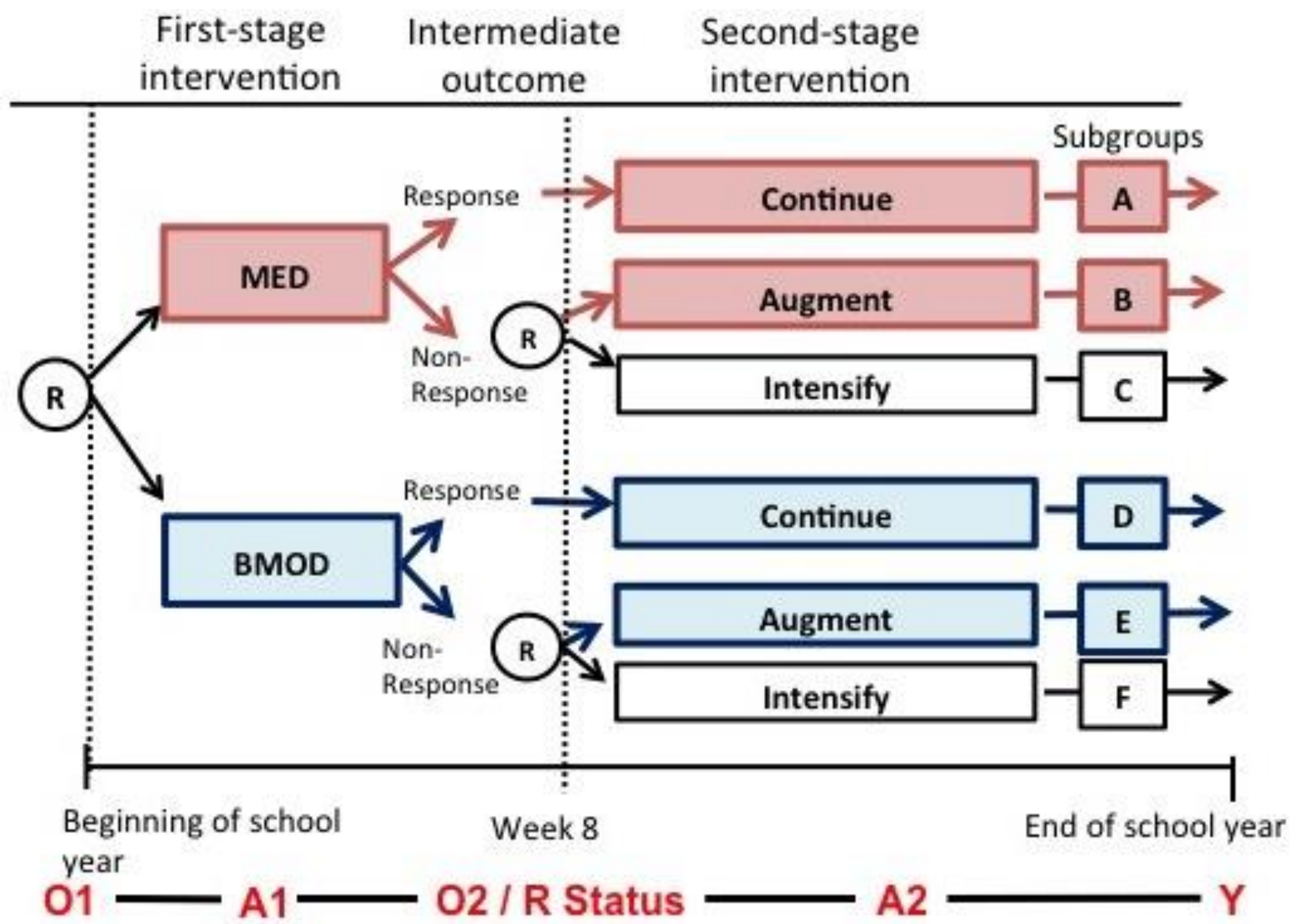
Compare 2 embedded adaptive interventions

AI #1:
Start with MED;
if non-responder AUGMENT,
else CONTINUE

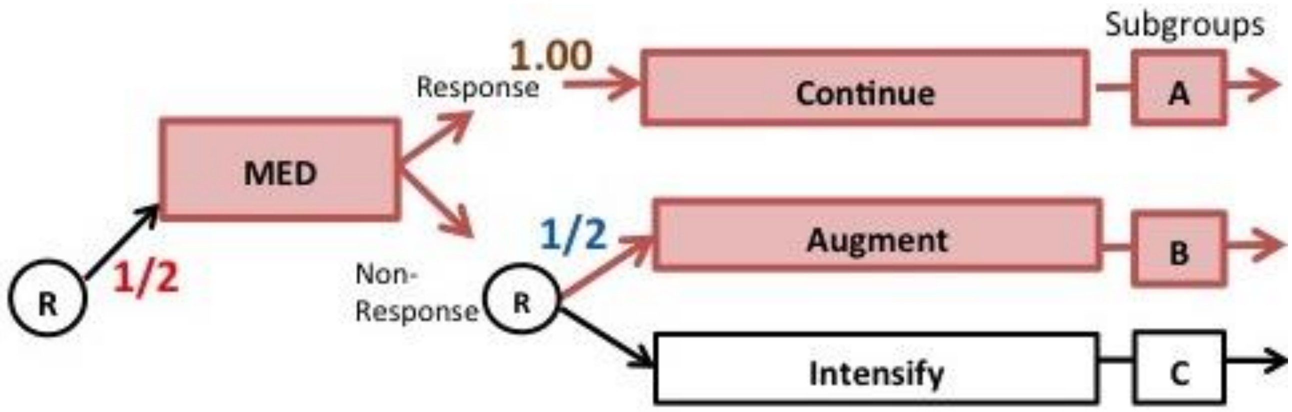
vs.

AI #2:
Start with BMOD;
if non-responder AUGMENT,
else CONTINUE

This Aim is a Comparison of Mean Outcome Under AI #1 vs. mean outcome under AI #2



We Know How to Account for the Imbalance in Non-Responders Following AI #1

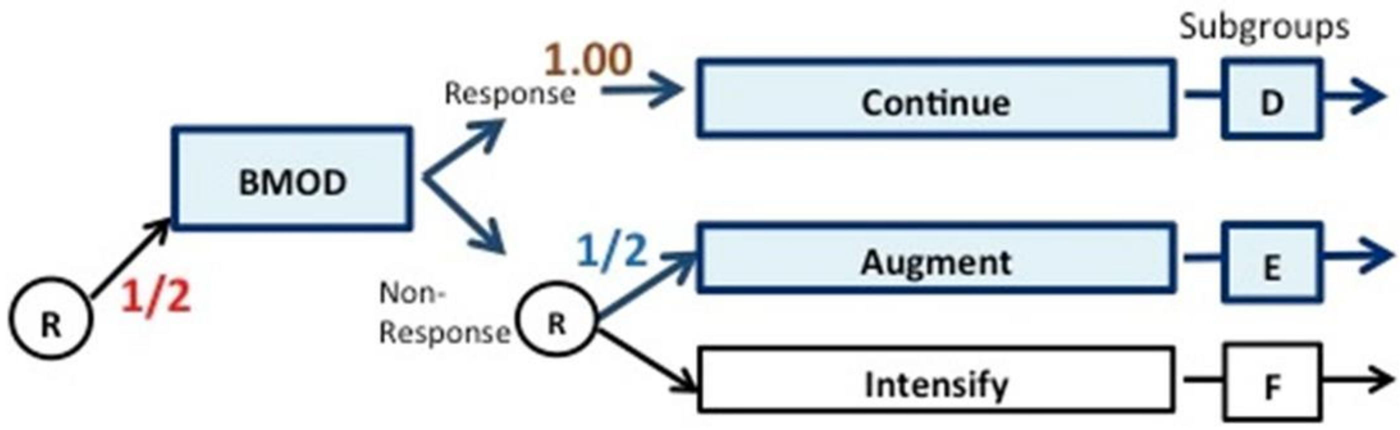


Assign W = weight = 2 to responders to MED: $2 \times \frac{1}{2} = 1$

Assign W = weight = 4 to non-responders to MED: $4 \times \frac{1}{4} = 1$

Then we take W-weighted mean of sample who ended up in boxes A & B.

A Similar Approach (and SAS Code) Can be Used to Obtain Mean Under AI #2



Assign W = weight = 2 to responders to MED: $2 \times \frac{1}{2} = 1$

Assign W = weight = 4 to non-responders to MED: $4 \times \frac{1}{4} = 1$

Then we take W-weighted mean of sample who ended up in boxes D & E.

Results for Estimated Mean Outcome had All Participants Followed AI#2 (BMOD, AUGMENT)

Analysis Of GEE Parameter Estimates			
Parameter	**Estimate**	**Standard Error**	**Pr > |Z|**
Intercept	3.0982	0.1070	<.0001
Z1	0.4085	0.1070	0.0001

Contrast Estimate Results					
		95% Confidence Limits			
Label	**Mean Estimate**	**Lower**	**Upper**	**Standard Error**	**Pr > ChiSq**
Mean Y under AI #2 (BMOD, AUGMENT)	**3.5067**	3.1643	3.8490	**0.1747**	**<.0001**

Interpretation: The estimated mean school performance score for children consistent with AI #2 is ~3.51 (95% CI: (3.16, 3.85)).

This analysis is with simulated data.

d³lab

Outline

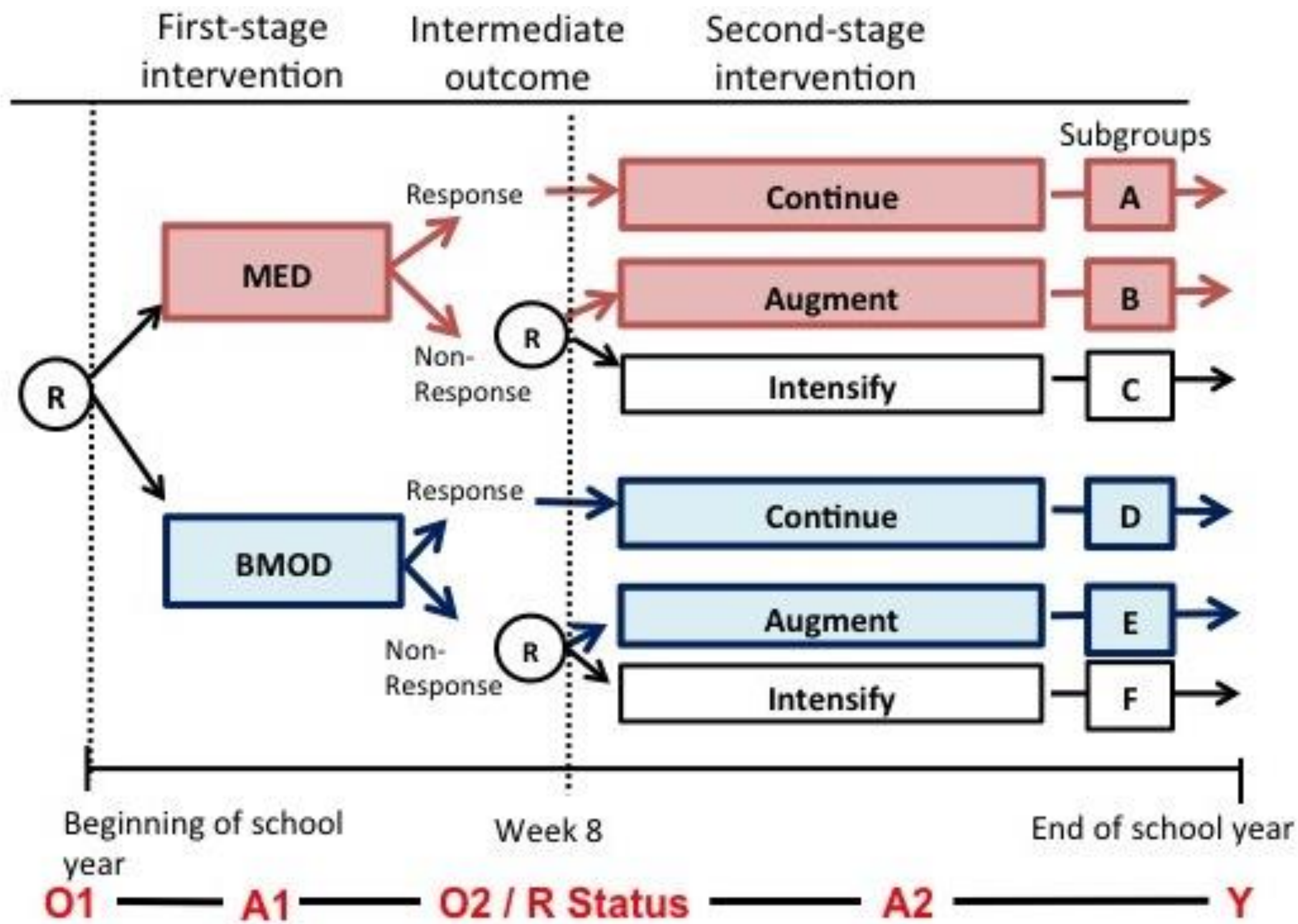
Review

- ADHD SMART study
- Weighted regression approach for estimating the mean outcome under one AI

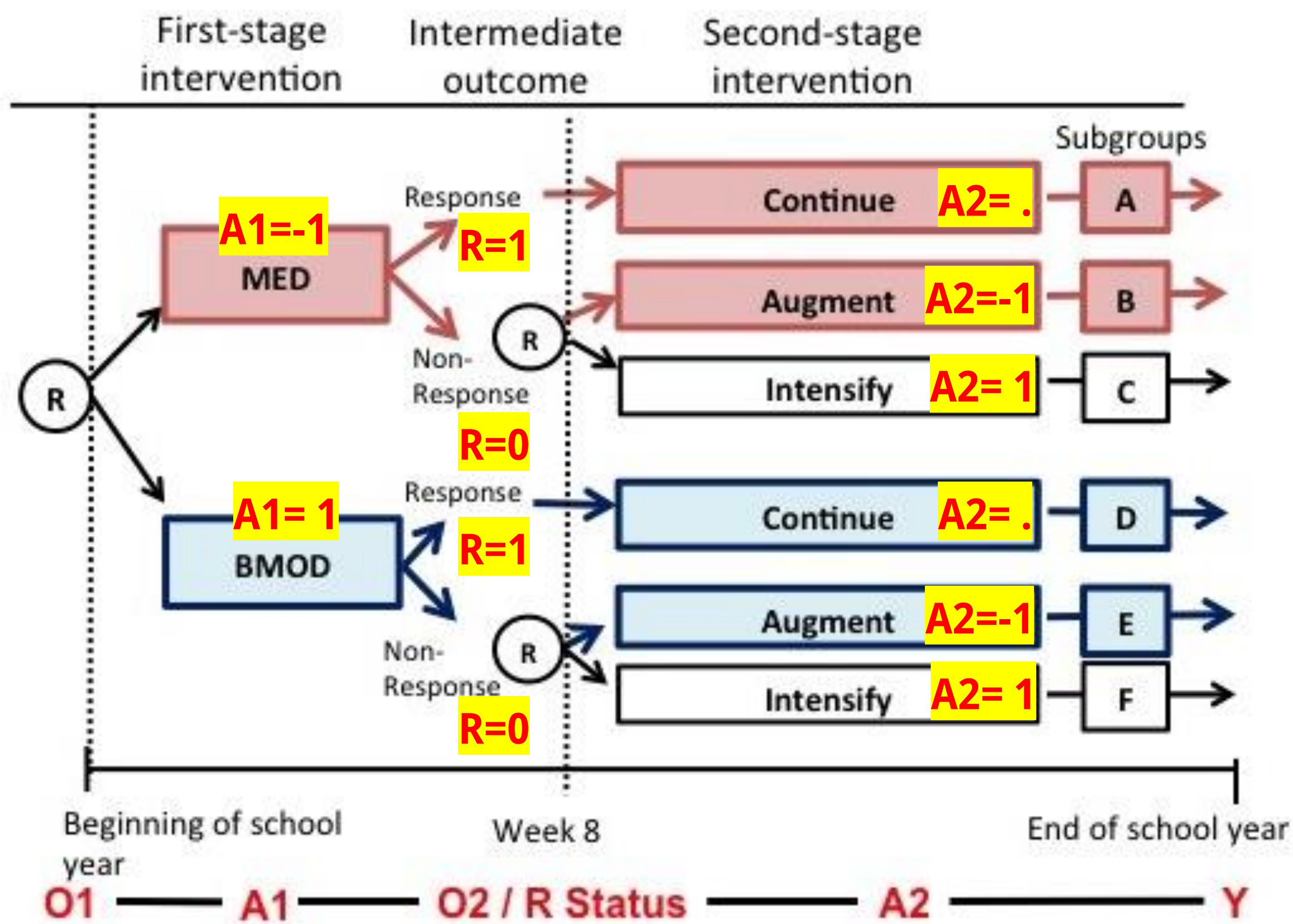
Learn

- **Use weighted regression to compare the mean outcomes for two AIs that begin with different treatments**
- Use weighted-and-replicated regression to simultaneously compare all embedded AIs in a SMART

An Intuitive (Yet Less Efficient) Approach to Comparing AI#1 vs AI#2



Reminder of Coding Scheme



An Intuitive (Yet Less Efficient) Approach to Comparing AI#1 vs AI#2

```
data dat7; set dat1;
 Z1=-1;
  if A1=-1 and R=1              then Z1=1;
  if A1=-1 and R=0 and A2=-1    then Z1=1;
 Z2=-1;
  if A1= 1 and R=1              then Z2=1;
  if A1= 1 and R=0 and A2=-1    then Z2=1;
 W=2*R + 4*(1-R);
run;


data dat8;
  set dat7; if Z1=1 or Z2=1;
run;
```

An Intuitive (Yet Less Efficient) Approach to Comparing AI#1 vs AI#2

```
data dat7; set dat1;
 Z1=-1;
  if A1=-1 and R=1              then Z1=1;
  if A1=-1 and R=0 and A2=-1    then Z1=1;
 Z2=-1;
  if A1= 1 and R=1              then Z2=1;
  if A1= 1 and R=0 and A2=-1    then Z2=1;
 W=2*R + 4*(1-R);
run;


data dat8;
  set dat7; if Z1=1 or Z2=1;
run;
```

Create Z1***:***
Indicator for whether or not the person is consistent with AI#1

An Intuitive (Yet Less Efficient) Approach to Comparing AI#1 vs AI#2

```
data dat7; set dat1;
 Z1=-1;
  if A1=-1 and R=1              then Z1=1;
  if A1=-1 and R=0 and A2=-1  then Z1=1;
 Z2=-1;
  if A1= 1 and R=1              then Z2=1;
  if A1= 1 and R=0 and A2=-1  then Z2=1;
 W=2*R + 4*(1-R);
run;


data dat8;
  set dat7; if Z1=1 or Z2=1;
run;
```

Create **Z2:**
Indicator for whether or not the person is consistent with AI#2

d³lab

An Intuitive (Yet Less Efficient) Approach to Comparing AI#1 vs AI#2

```
data dat7; set dat1;
 Z1=-1;
  if A1=-1 and R=1            then Z1=1;
  if A1=-1 and R=0 and A2=-1  then Z1=1;
 Z2=-1;
  if A1= 1 and R=1            then Z2=1;
  if A1= 1 and R=0 and A2=-1  then Z2=1;
 W=2*R + 4*(1-R);
run;
```

Assign weights:
2 for responders
4 for non-responders

```
data dat8;
  set dat7; if Z1=1 or Z2=1;
run;
```

An Intuitive (Yet Less Efficient) Approach to Comparing AI#1 vs AI#2

```
data dat7; set dat1;
 Z1=-1;
  if A1=-1 and R=1              then Z1=1;
  if A1=-1 and R=0 and A2=-1  then Z1=1;
 Z2=-1;
  if A1= 1 and R=1              then Z2=1;
  if A1= 1 and R=0 and A2=-1  then Z2=1;
 W=2*R + 4*(1-R);
run;
```

```
data dat8;
  set dat7; if Z1=1 or Z2=1;
run;
```

Subset Data:
Keep only participants consistent with either AI#1 or AI#2

An Intuitive (Yet Less Efficient) Approach to Comparing AI#1 vs AI#2

The Regression and Contrast Coding Logic:

Recall:

Z_1 is now an indicator for whether the person is consistent with AI#1 or with AI#2:
→ $Z_1 = 1$ = AI#1
→ $Z_1 = -1$ = AI#2

To compare the 2 AIs, we can fit the Model:

$$E(Y|Z_1) = \beta_0 + \beta_1 Z_1$$

Overall Mean Y under AI#1 = $\boldsymbol{\beta_0 + \beta_1 \times 1}$

Overall Mean Y under AI#2 = $\boldsymbol{\beta_0 + \beta_1 \times -1}$

Diff Between AIs = $\boldsymbol{\beta_0 + \beta_1 - (\beta_0 - \beta_1) = 2\beta_1}$

d³lab

An Intuitive (Yet Less Efficient) Approach to Comparing AI#1 vs AI#2

```
proc genmod data = dat8;
  class id;
  model Y = Z1;
  weight W;
  repeated subject = id / type = ind;
  estimate 'Mean Y AI#1(MED, Add BMOD)' intercept 1 Z1  1;
  estimate 'Mean Y AI#2(BMOD, Add MED)' intercept 1 Z1 -1;
  estimate 'Diff: AI#1 - AI#2'                      Z1  2;
run;
```

Mean Y under AI#1	= $\boldsymbol{\beta_0 + \beta_1 \times 1}$
Mean Y under AI#2	= $\boldsymbol{\beta_0 + \beta_1 \times -1}$
Diff Between AIs	= $\boldsymbol{2\beta_1}$

An Intuitive (Yet Less Efficient) Approach to Comparing AI#1 vs AI#2

Analysis Of GEE Parameter Estimates

Parameter	Estimate	Standard Error	Pr > |Z|
Intercept	3.1858	0.1221	<.0001
Z1	-0.3209	0.1221	0.0086

Contrast Estimate Results

Label	Mean Estimate	95% Confidence Limits Lower	Upper	Standard Error	Pr > ChiSq
Mean Y under AI #1 (MED, AUGMENT)	**2.8649**	2.5305	3.1992	**0.1706**	**<.0001**
Mean Y under AI #2 (BMOD, AUGMENT)	**3.5067**	3.1643	3.8490	**0.1747**	**<.0001**
Diff: AI#1 – AI#2	**-0.6418**	-1.1203	-0.1633	**0.2442**	**0.0086**

This analysis is with simulated data.

Notice SE

An Intuitive (Yet Less Efficient) Approach to Comparing AI#1 vs AI#2

```
proc genmod data = dat8;
  class id;
  model Y = Z1 O12c O14c;
  weight w;
  repeated subject = id / type = ind;
  estimate 'Mean Y AI#1(MED, AUGMENT)' intercept 1 Z1  1;
  estimate 'Mean Y AI#2(BMOD,AUGMENT)' intercept 1 Z1 -1;
  estimate 'Diff: AI#1 - AI#2'                     Z1  2;
run;
```

Add baseline control covariates

An Intuitive (Yet Less Efficient) Approach to Comparing AI#1 vs AI#2

Analysis Of GEE Parameter Estimates

Parameter	Estimate	Standard Error	Pr > |Z|
Intercept	3.1858	0.1221	<.0001
Z1	-0.2442	0.1122	0.0295
O12c	-0.5153	0.0971	<.0001
O14c	0.4905	0.2774	0.0770

Contrast Estimate Results

Label	Mean Estimate	95% Confidence Limits		Standard Error	Pr > ChiSq
		Lower	Upper		
Mean Y under AI #1	**2.8842**	2.5919	3.1765	**0.1491**	**<.0001**
Mean Y under AI #2	**3.3727**	3.0542	3.6912	**0.1625**	**<.0001**
Diff: AI#1 – AI#2	**-0.4884**	-0.9282	-0.0487	**0.2244**	**0.0295**

This analysis is with simulated data.

Notice SE: Slightly smaller compared to the analysis without control covariates

Outline

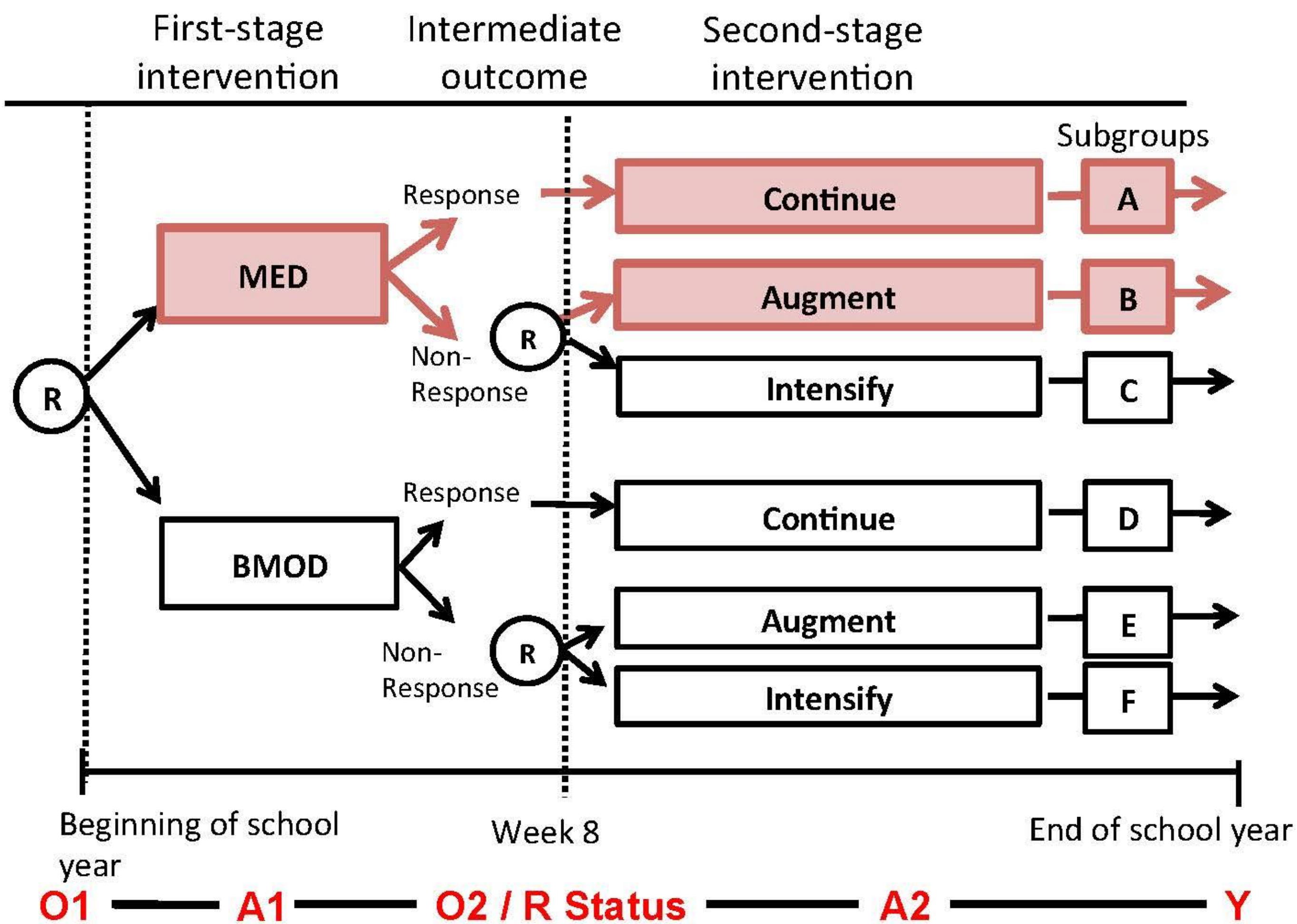
Review

- ADHD SMART study
- Weighted regression approach for estimating the mean outcome under one AI

Learn

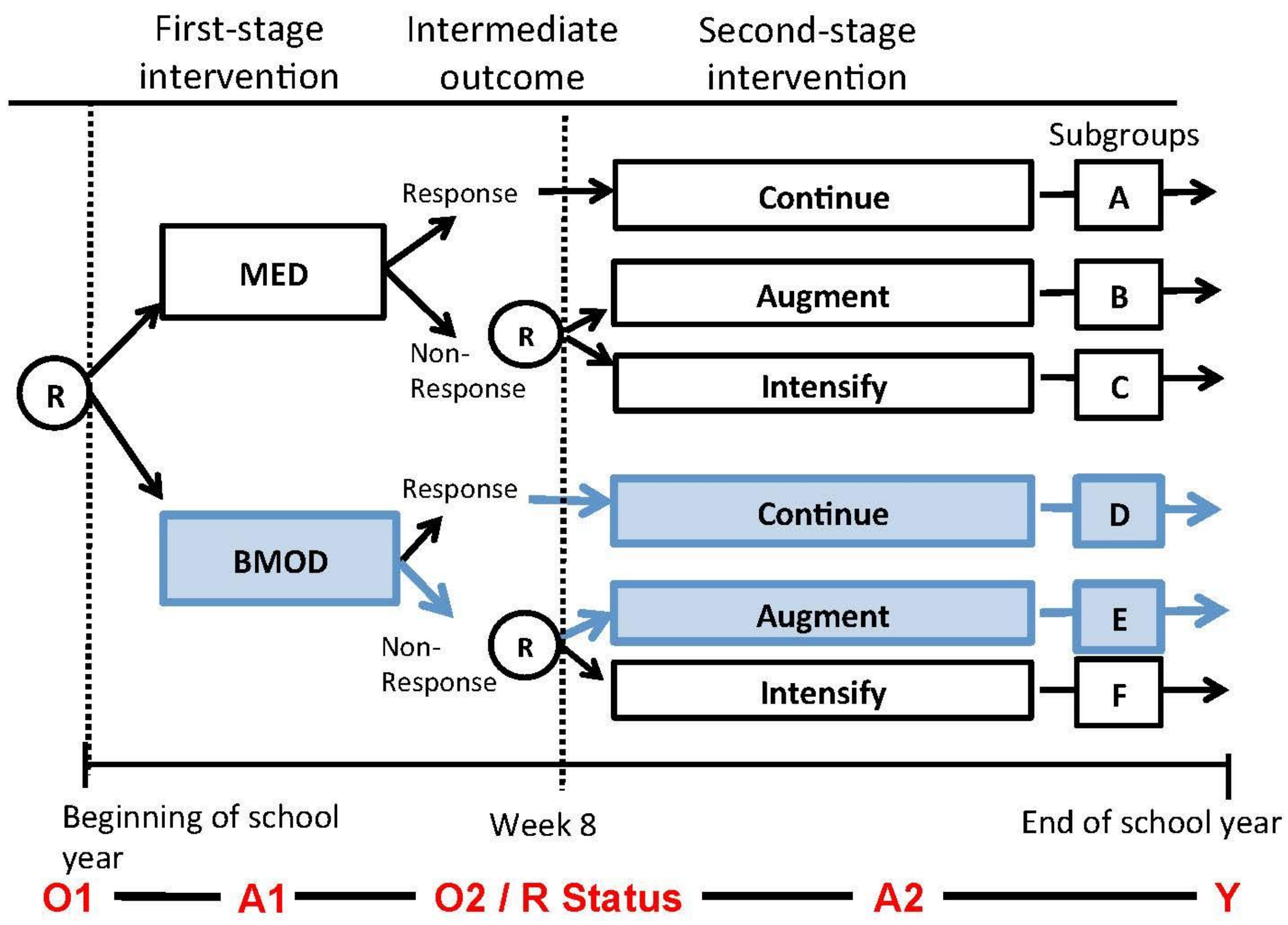
- Use weighted regression to compare the mean outcomes for two AIs that begin with different treatments
- **Use weighted-and-replicated regression to simultaneously compare all embedded AIs in a SMART**

What about a Regression to Compare AI#1 (MED, AUGMENT) vs…

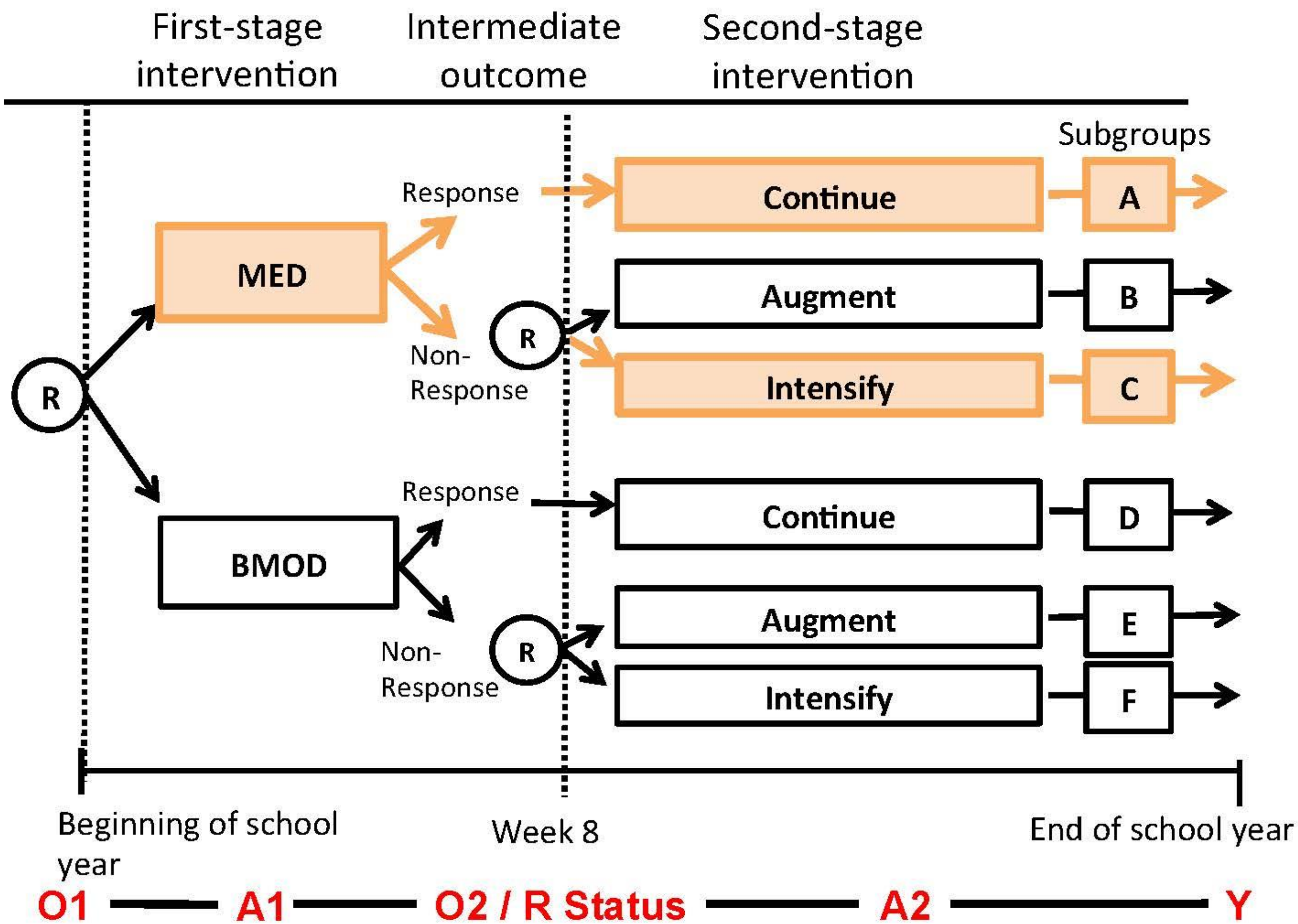


d³lab

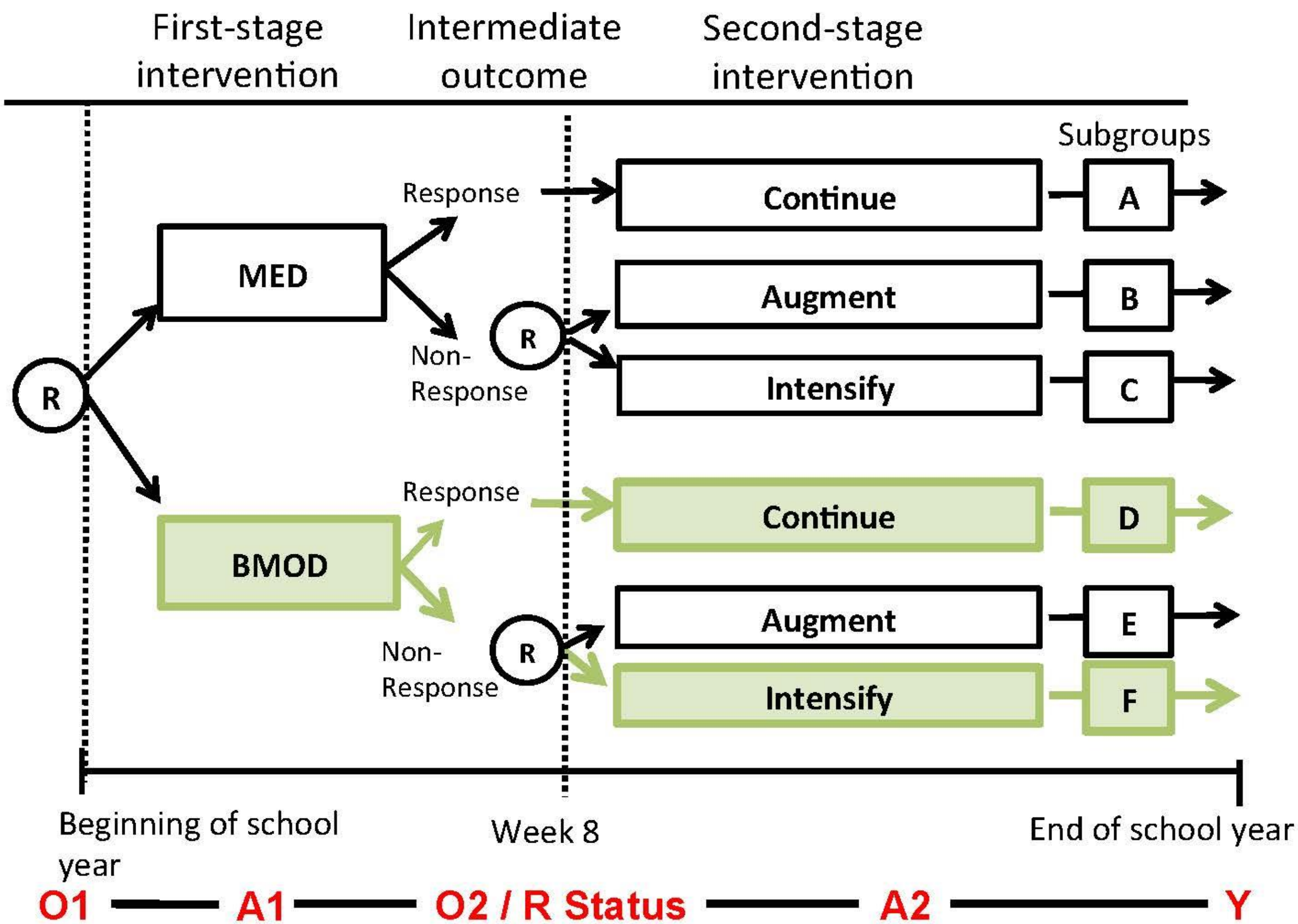
… AI#2 (BMOD, AUGMENT) vs…



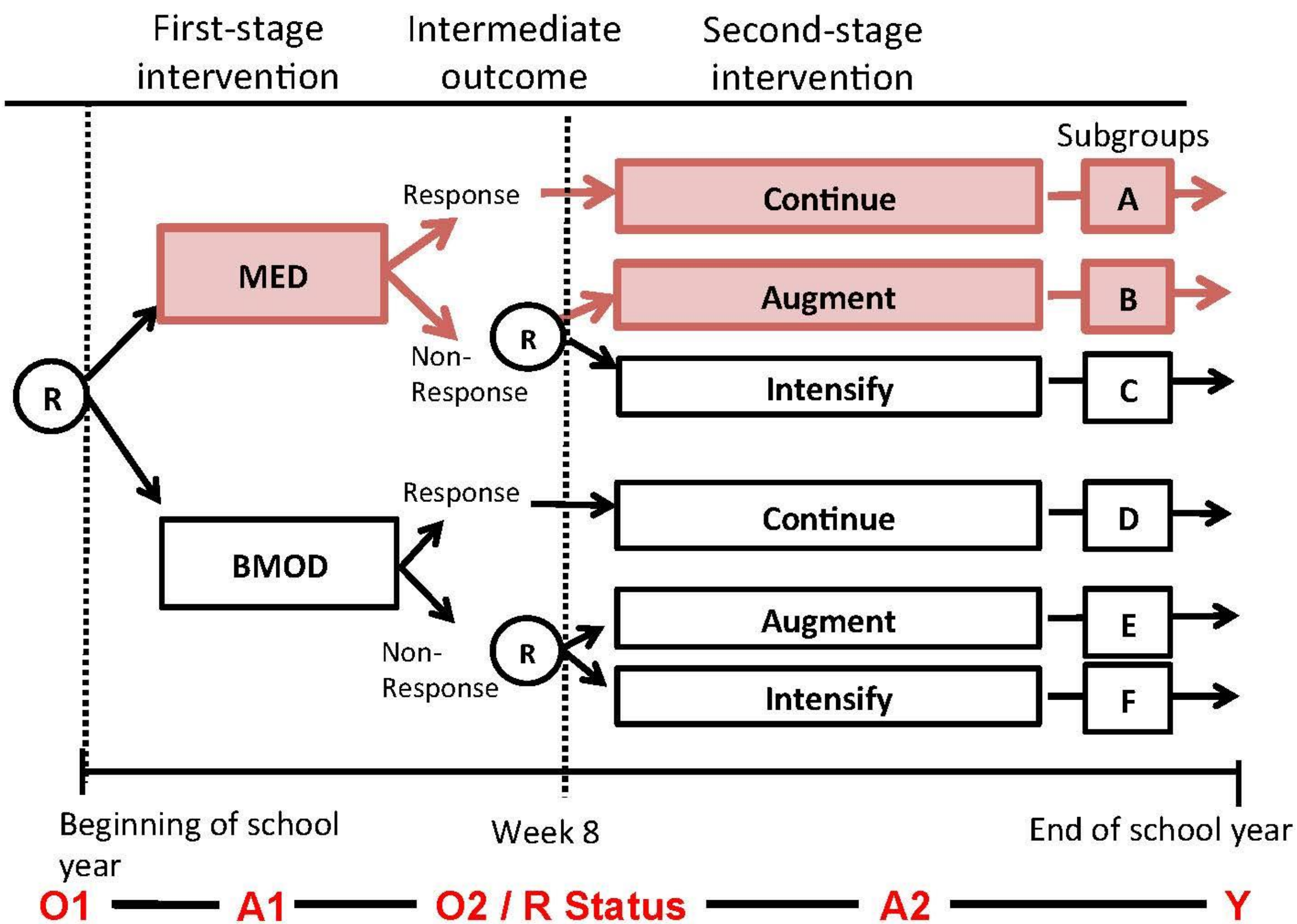
… AI#3 (MED, INTENSIFY) vs…



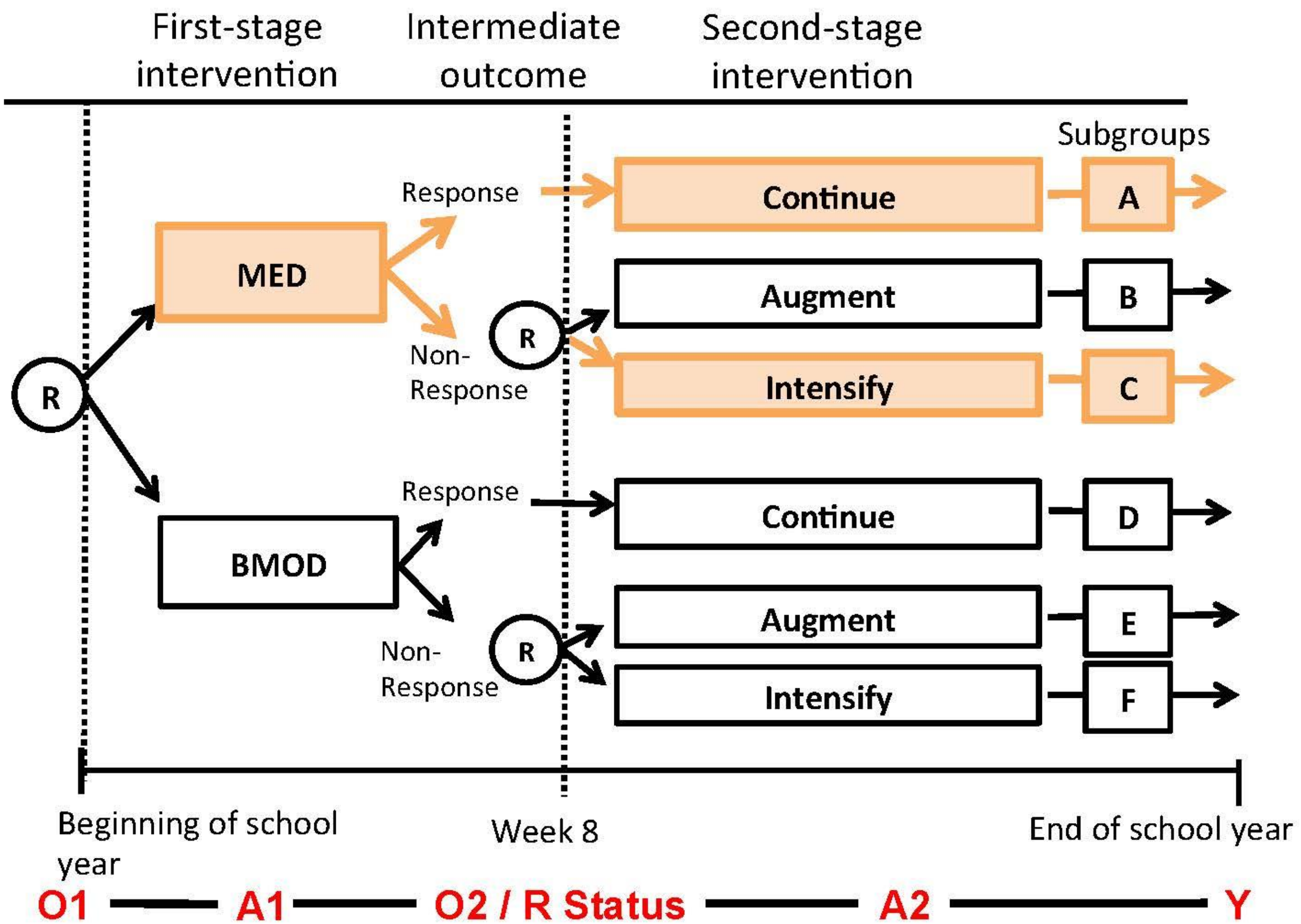
… AI#4 (BMOD, INTENSIFY), All In One Swoop?



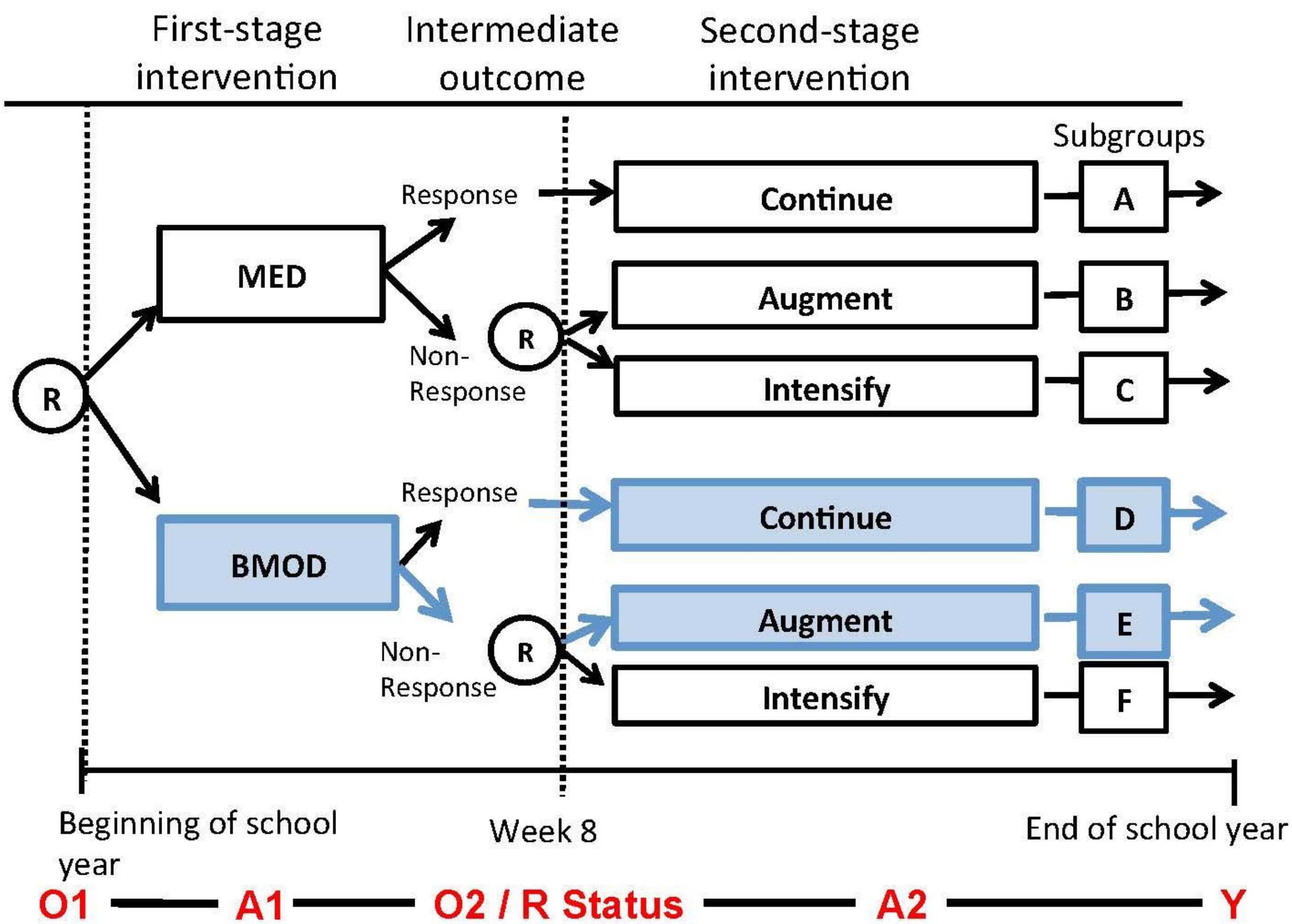
Notice that AI#1 and AI#3 (start MED) Share Responders (Box A)



Notice that AI#1 and AI#3 (start MED) Share Responders (Box A)

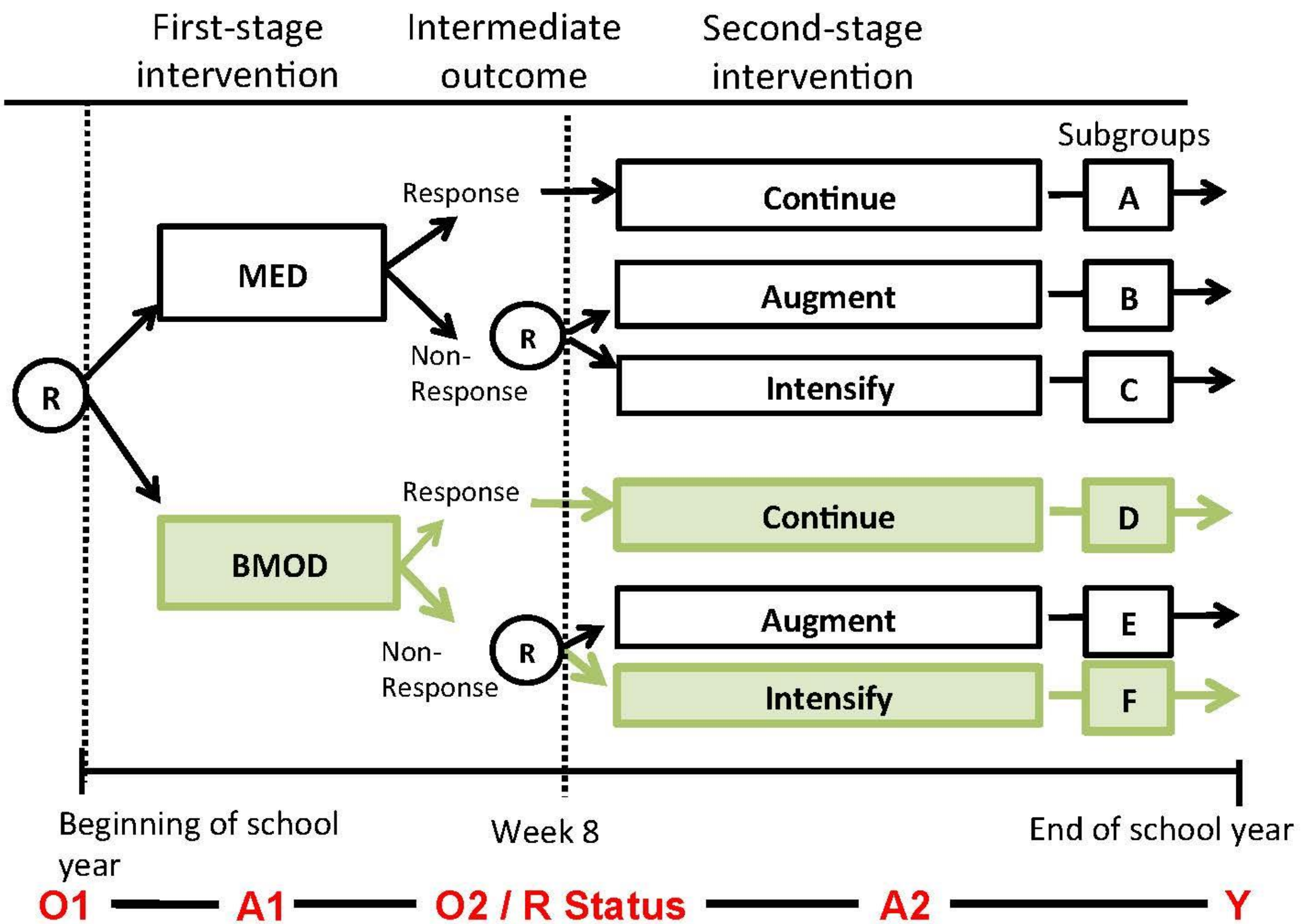


Similarly: AI#2 and AI#4 (start BMOD) Share Responders (Box D)



d3lab

Similarly: AI#2 and AI#4 (start BMOD) Share Responders (Box D)



So, What's Going On?

In this SMART, all responders are consistent with two AIs

- Responders to MED are part of AI#1 and AI#3
- Responders to BMOD are part of AI#2 and AI#4

If our goal is to estimate the mean outcome under all AIs simultaneously,

We must share responders somehow.

- But how?

What Do We Do?

We "trick" SAS into using the responders twice

We do this by replicating responders:

- Create 2 observations for each responder
- We assign half of them A2=1, the other half A2=-1

`W=2` to responders and `W=4` to non-responders

Robust standard errors account for weighting and the fact that responders are "re-used". No cheating here!

d³lab

Weighting and Replicating Serve Different Purposes

Weighting

- Accounts for over/underrepresentation of responders or non-responders
- Because of the randomization scheme

Replicating

- Allows us to use standard software to do simultaneous estimation and comparison
- Because participants are consistent with more than one AI

SAS Code for Weighting & Replicating to Compare Means Under All Four AIs

```
data dat9; set dat1;
 if R=1 then do;
    ob = 1; A2 =-1; weight = 2; output;
    ob = 2; A2 = 1; weight = 2; output;
    end;

 else if R=0 then do;
    ob = 1; weight = 4; output;
    end;
run;
```

Replicated Data

Obs	ID	A1	R	A2	Y	o11c	o12c	o13c	o14c	ob	weight
45	32	1	1	−1	5	−0.35333	−2.73889	−0.31333	0.19333	1	2
46	32	1	1	1	5	−0.35333	−2.73889	−0.31333	0.19333	2	2
47	33	1	0	1	3	0.64667	−1.07820	0.68667	0.19333	1	4
48	34	1	0	1	[illegible]	[illegible]	[illegible]	−0.31333	0.19333	1	4
49	35	1	0	−1	[illegible]	[illegible]	[illegible]	−0.31333	0.19333	1	4
50	36	−1	0	1	1	0.64667	−0.03527	−0.31333	0.19333	1	4
51	37	−1	1	−1	1	−0.35333	0.99556	−0.31333	0.19333	1	2
52	37	−1	1	1	1	−0.35333	0.99556	−0.31333	0.19333	2	2
53	38	−1	0	−1	3	−0.35333	0.14034	0.68667	−0.80667	1	4
54	39	−1	1	[illegible]	3	0.64667	1.64983	0.68667	0.19333	1	2
55	39	−1	1	1	3	0.64667	1.64983	0.68667	0.19333	2	2

Responders
are replicated!

Non-Responders
aren't!

After Weighting & Replicating: SAS Code for the Weighted Regression

The Regression and Contrast Coding Logic:

Recall:

Our goal is to compare all 4 embedded AIs
We have 2 indicators: A_1, A_2:

A_1	**A_2**
1 BMOD	**1** INTENSIFY
-1 MED	**-1** AUGMENT

To compare all 4 AIs, we can fit the following model:

$$E(Y|A_1, A_2) = \beta_0 + \beta_1 A_1 + \beta_2 A_2 + \beta_3 A_1 A_2$$

d³lab

After Weighting & Replicating: SAS Code for the Weighted Regression

The Regression and Contrast Coding Logic:

$$E(Y|A_1, A_2) = \beta_0 + \beta_1 A_1 + \beta_2 A_2 + \beta_3 A_1 A_2$$

AI		Mean Y Under AI
1	(MED, AUGMENT)	$\boldsymbol{\beta_0 + \beta_1(-1) + \beta_2(-1) + \beta_3(-1)(-1)}$
2	(BMOD, AUGMENT)	$\boldsymbol{\beta_0 + \beta_1(\ \ 1) + \beta_2(-1) + \beta_3(\ \ 1)(-1)}$
3	(MED, INTENSIFY)	$\boldsymbol{\beta_0 + \beta_1(-1) + \beta_2(\ \ 1) + \beta_3(-1)(\ \ 1)}$
4	(BMOD, INTENSIFY)	$\boldsymbol{\beta_0 + \beta_1(\ \ 1) + \beta_2(\ \ 1) + \beta_3(\ \ 1)(\ \ 1)}$

A_1		**A_2**	
1	BMOD	**1**	INTENSIFY
-1	MED	**-1**	AUGMENT

After Weighting & Replicating: SAS Code for the Weighted Regression

The Regression and Contrast Coding Logic:

$$E(Y|A_1, A_2) = \beta_0 + \beta_1 A_1 + \beta_2 A_2 + \beta_3 A_1 A_2$$

AI		Mean Y Under AI
1	(-1, -1)	$\beta_0 + \beta_1(-1) + \beta_2(-1) + \beta_3(-1)(-1)$
2	(1, -1)	$\beta_0 + \beta_1(\ \ 1) + \beta_2(-1) + \beta_3(\ \ 1)(-1)$
3	(-1, 1)	$\beta_0 + \beta_1(-1) + \beta_2(\ \ 1) + \beta_3(-1)(\ \ 1)$
4	(1, 1)	$\beta_0 + \beta_1(\ \ 1) + \beta_2(\ \ 1) + \beta_3(\ \ 1)(\ \ 1)$

A_1	A_2
1 BMOD	1 INTENSIFY
-1 MED	-1 AUGMENT

After Weighting & Replicating: SAS Code for the Weighted Regression

The Regression and Contrast Coding Logic:

$$E(Y|A_1, A_2) = \beta_0 + \beta_1 A_1 + \beta_2 A_2 + \beta_3 A_1 A_2$$

AI	Mean Y Under AI
1 (-1, -1)	$\boldsymbol{\beta_0 - \beta_1 - \beta_2 + \beta_3}$
2 (1, -1)	$\boldsymbol{\beta_0 + \beta_1 - \beta_2 - \beta_3}$
3 (-1, 1)	$\boldsymbol{\beta_0 - \beta_1 + \beta_2 - \beta_3}$
4 (1, 1)	$\boldsymbol{\beta_0 + \beta_1 + \beta_2 + \beta_3}$

The difference between AI#1 and AI#2:

$$\boldsymbol{(\beta_0 - \beta_1 - \beta_2 + \beta_3) - (\beta_0 + \beta_1 - \beta_2 - \beta_3) = -2\beta_1 + 2\beta_3}$$

d³lab

After Weighting & Replicating: SAS Code for the Weighted Regression

```
proc genmod data = dat9;
 class id;
 model Y = A1 A2 A1*A2;
 weight weight;
 repeated subject = id / type = ind;
 estimate 'MeanY:AI#1(MED,AUGMENT) ' int 1 A1 -1 A2 -1 A1*A2  1;
 estimate 'MeanY:AI#2(BMOD,AUGMENT)' int 1 A1  1 A2 -1 A1*A2 -1;
 estimate 'MeanY:AI#3(MED,INTENSFY)' int 1 A1 -1 A2  1 A1*A2 -1;
 estimate 'MeanY:AI#4(BMOD,INTNSFY)' int 1 A1  1 A2  1 A1*A2  1;
 estimate '  Diff: AI#1 - AI#2     ' int 0 A1 -2 A2  0 A1*A2  2;
 estimate '  Diff: AI#1 - AI#3     ' int 0 A1  0 A2 -2 A1*A2  2;
 estimate '  Diff: AI#1 - AI#4     ' int 0 A1 -2 A2 -2 A1*A2  0;
*etc...;
run;
```

After Weighting & Replicating:
SAS Code for the Weighted Regression

```
proc genmod data = dat9;
 class id;
 model Y = A1 A2 A1*A2;
 weight weight;
 repeated subject = id / type = ind;
 estimate 'MeanY:AI#1(MED,AUGMENT) ' int 1 A1 -1 A2 -1 A1*A2  1;
 estimate 'MeanY:AI#2(BMOD,AUGMENT)' int 1 A1  1 A2 -1 A1*A2 -1;
 estimate 'MeanY:AI#3(MED,INTENSFY)' int 1 A1 -1 A2  1 A1*A2 -1;
 estimate 'MeanY:AI#4(BMOD,INTNSFY)' int 1 A1  1 A2  1 A1*A2  1;
 estimate '  Diff: AI#1 - AI#2     ' int 0 A1 -2 A2  0 A1*A2  2;
 estimate '  Diff: AI#1 - AI#3     ' int 0 A1  0 A2 -2 A1*A2  2;
 estimate '  Diff: AI#1 - AI#4     ' int 0 A1 -2 A2 -2 A1*A2  0;
*etc...;
run;
```

Estimate Difference:

Diff AI #1 – AI # 2 = $-2\beta_1 + 2\beta_3$

Results for Weighted & Replicated Regression: Comparing Mean Outcome for all AIs Simultaneously

Contrast Estimate Results

Label	Mean Estimate	95% Confidence Limits Lower	95% Confidence Limits Upper	Standard Error
Mean Y under AI #1 (MED, AUGMENT)	2.8649	2.5305	3.1992	0.1706
Mean Y under AI #2 (BMOD, AUGMENT)	3.5067	3.1643	3.8490	0.1747
Mean Y under AI #3 (MED, INTENSIFY)	2.7895	2.4644	3.1145	0.1658
Mean Y under AI #4 (BMOD, INTENSIFY)	2.6533	2.2515	3.0552	0.2050
Diff: AI#1 – AI#2	-0.6418	-1.1203	-0.1633	0.2442
Diff: AI#1 – AI#3	0.0754	-0.3106	0.4614	0.1969
Diff: AI#1 – AI#4	0.2115	-0.3112	0.7343	0.2667

NOTE: We get the exact same results as before when we compared AI#1 vs AI#2, but now we can simultaneously make inference for all the comparisons.

This analysis is with simulated data.

But wait!...
There’s More to Weighted & Replicated Regression Than Just Convenience!

Weighted & Replicated Regression is More Efficient Statistically

Improve power:
Adjusting for baseline covariates that are associated with outcome leads to more efficient estimates (lower standard error = more power = smaller p-value).

```
proc genmod data = dat9;
 class id;
 model Y = A1 A2 A1*A2 O12c O14c;
 weight weight;
 repeated subject = id / type = ind;
 estimate ‘MeanY:AI#1(MED,AUGMENT) ’ int 1 A1 -1 A2 -1 A1*A2  1;
 estimate ‘MeanY:AI#2(BMOD,AUGMENT)’ int 1 A1  1 A2 -1 A1*A2 -1;
 estimate ‘MeanY:AI#3(MED,INTENSFY)’ int 1 A1 -1 A2  1 A1*A2 -1;
 estimate ‘MeanY:AI#4(BMOD,INTNSFY)’ int 1 A1  1 A2  1 A1*A2  1;
 estimate ‘  Diff: AI#1 - AI#2     ’ int 0 A1 -2 A2  0 A1*A2  2;
 estimate ‘  Diff: AI#1 - AI#3     ’ int 0 A1  0 A2 -2 A1*A2  2;
 estimate ‘  Diff: AI#1 - AI#4     ’ int 0 A1 -2 A2 -2 A1*A2  0;
*etc...;
 run;
```

d³lab

Weighted & Replicated Regression is More Efficient Statistically

Improved efficiency: Adjusting for baseline covariates resulted in lower standard error and tighter confidence intervals. Point estimates remained about the same, as expected.

Contrast Estimate Results

Label	Mean Estimate	95% Confidence Limits Lower	95% Confidence Limits Upper	Standard Error
Mean Y under AI #1 (MED, AUGMENT)	2.8801	2.5869	3.1733	0.1496
Mean Y under AI #2 (BMOD, AUGMENT)	3.3854	3.0689	3.7018	0.1614
Mean Y under AI #3 (MED, INTENSIFY)	2.8149	2.5163	3.1135	0.1524
Mean Y under AI #4 (BMOD, INTENSIFY)	2.7338	2.3596	3.1081	0.1909
Diff: AI#1 – AI#2	-0.5053	-0.9401	-0.0704	0.2219
Diff: AI#1 – AI#3	0.0652	-0.2811	0.4115	0.1767

SE in unadjusted model was **0.2442**

This analysis is with simulated data.

d³lab

Weighted & Replicated Regression is More Efficient Statistically

Contrast Estimate Results

Label	Mean Estimate	95% Confidence Limits		Standard Error
		Lower	Upper	
Mean Y under AI #1 (MED, AUGMENT)	2.8801	2.5869	3.1733	0.1496
Mean Y under AI #2 (BMOD, AUGMENT)	3.3854	3.0689	3.7018	0.1614
Mean Y under AI #3 (MED, INTENSIFY)	2.8149	2.5163	3.1135	0.1524
Mean Y under AI #4 (BMOD, INTENSIFY)	2.7338	2.3596	3.1081	0.1909
Diff: AI#1 – AI#2	-0.5053	-0.9401	-0.0704	0.2219
Diff: AI#1 – AI#3	0.0652	-0.2811	0.4115	0.1767

SE in unadjusted model was **0.2442**

SE in adjusted model including only data from participants in AI #1 and AI # 2 was **0.2244**

This analysis is with simulated data.

Citations

- Murphy, S.A. (2005). An experimental design for the development of adaptive interventions. *Statistics in Medicine*, 24, 1455-1481.

- 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), 457-477.