Presented by
Daniel Almirall, PhD

Primary Aim Analyses in a SMART

Part I

Module 4

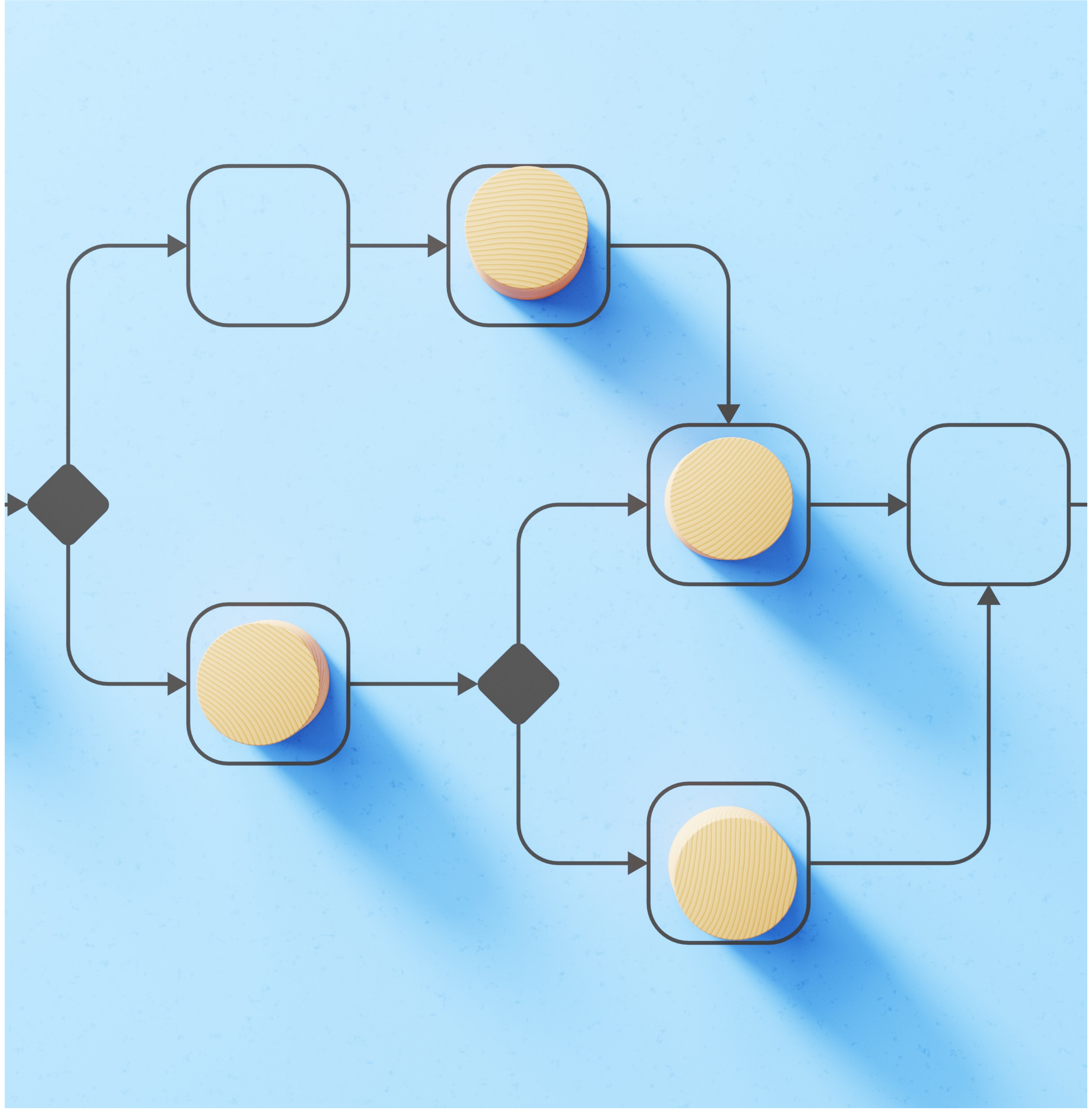
70 min

Learning Objectives

You will have a better understanding, and will continue learning how to frame, the typical Primary Aims in a SMART

You will learn about key statistical considerations in Primary Aim analyses in a SMART

You will learn how to interpret the output for the different Primary Aim Analyses in a SMART



Outline

Illustrative Example: ADHD SMART Study (PI: Pelham)

Data Analytics 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): Estimate the mean outcome under each of the embedded AIs (separately) using weighted least squares

Note About SAS Code

My slides include SAS Code, which will be available on our website on April 1, 2023

The goal is to provide the intuition for the data analysis and to help you learn how to interpret output from regression, not to make you experts on SAS

In the upcoming virtual half-days, you will learn how to do your own analysis in R

Outline

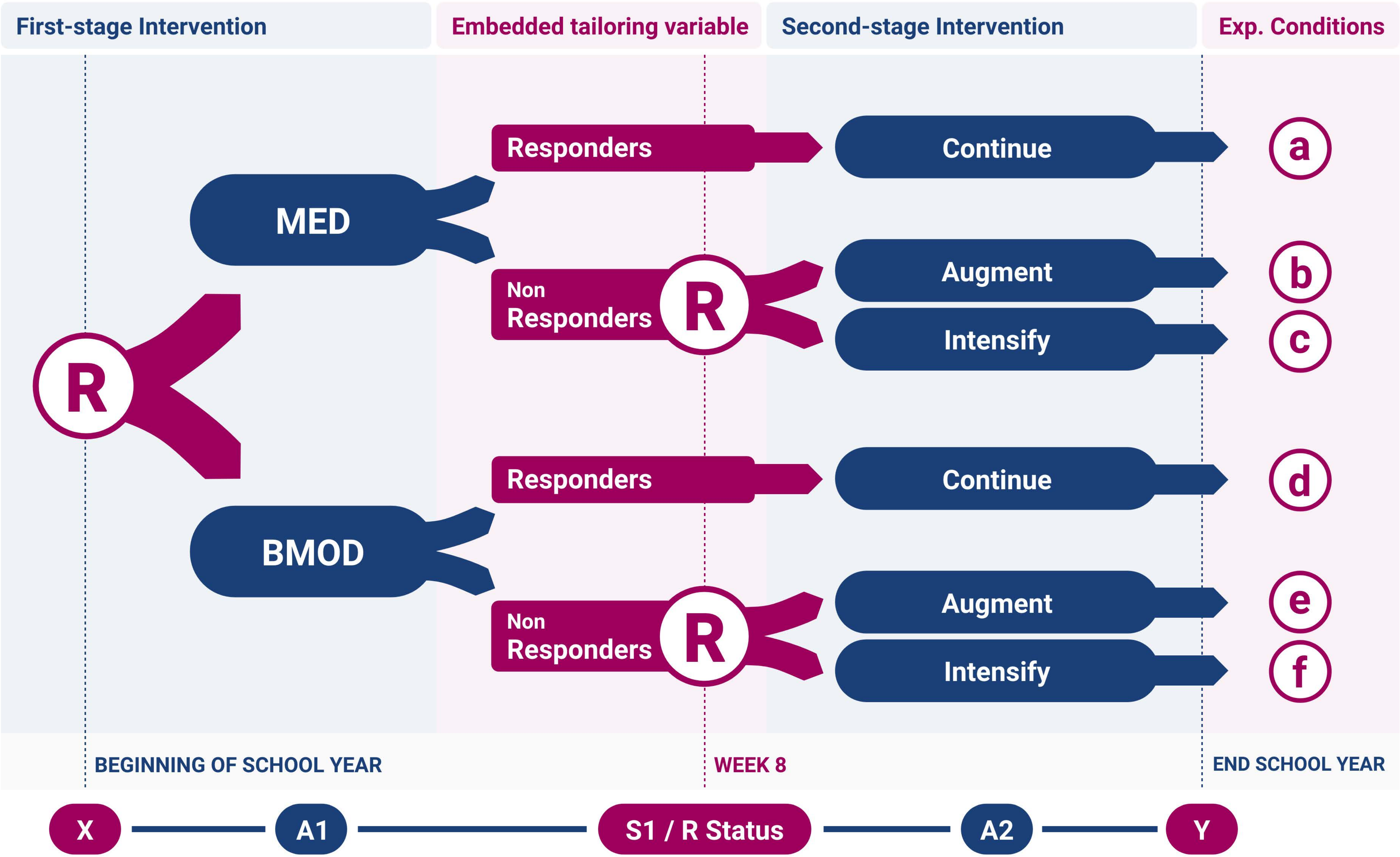
Illustrative Example: ADHD SMART Study (PI: Pelham)

Data Analytics 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): Estimate the mean outcome under each of the embedded AIs (separately) using weighted least squares

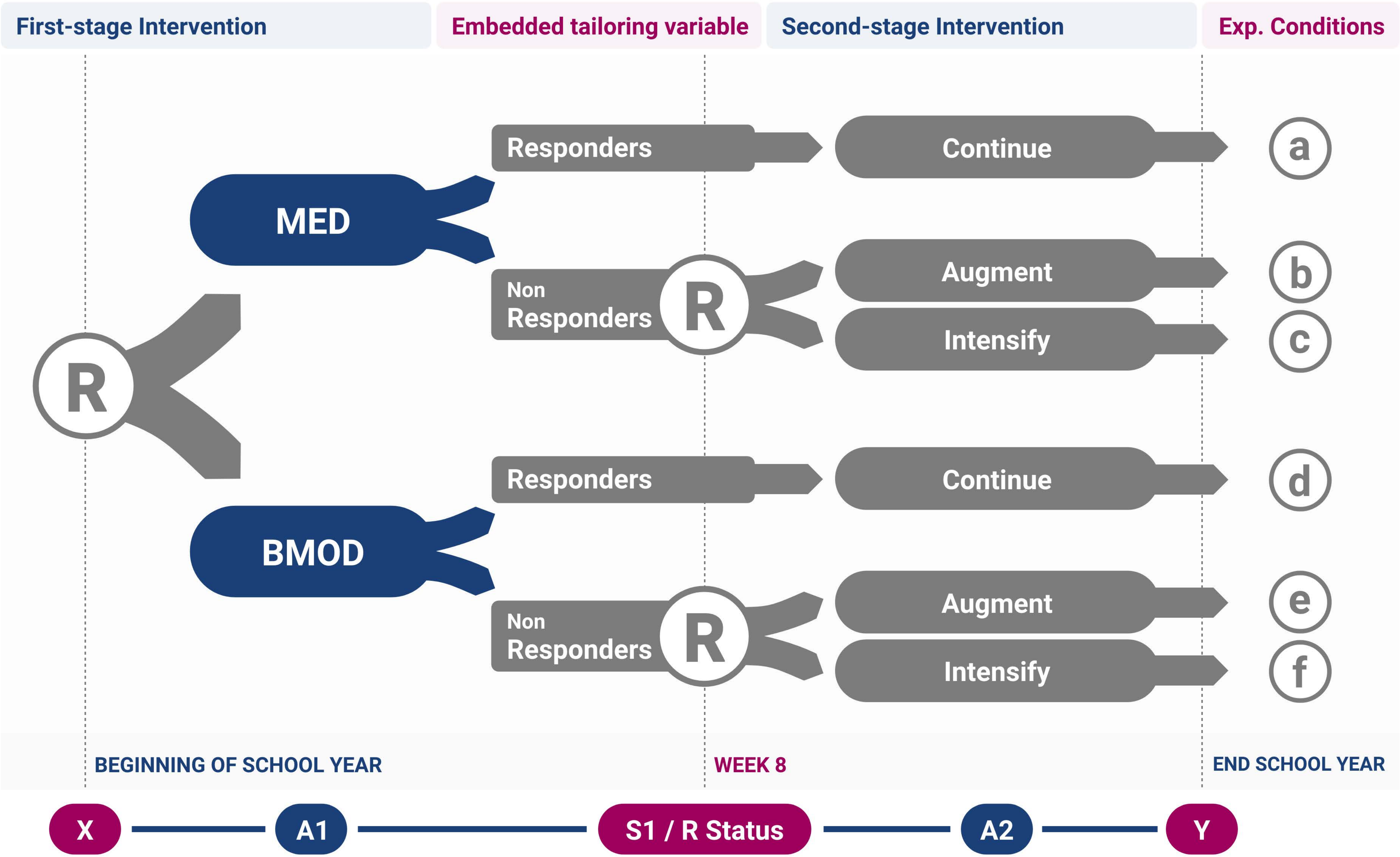
SMART Example ADHD Study

PI: Pelham



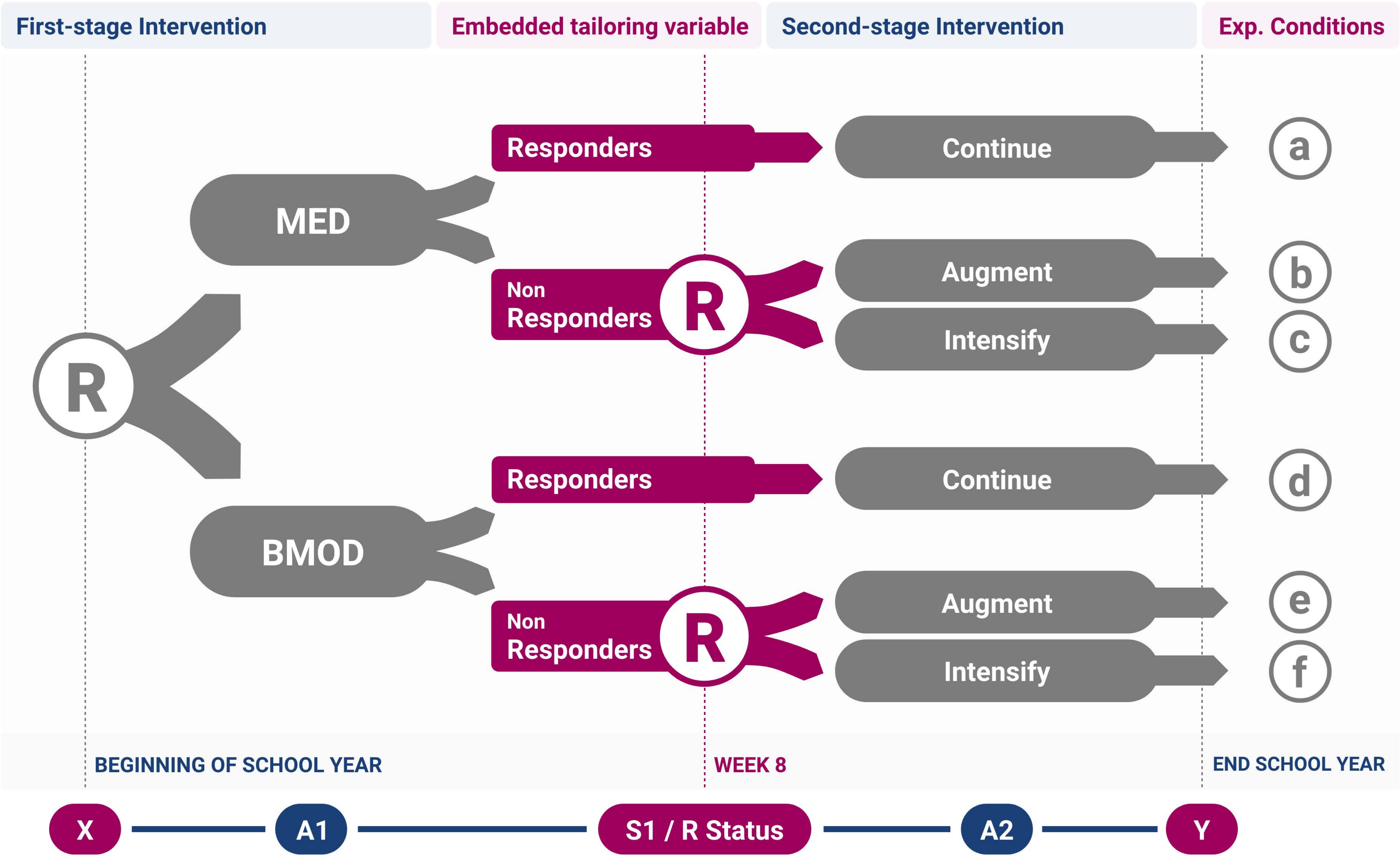
SMART Example ADHD Study

PI: Pelham



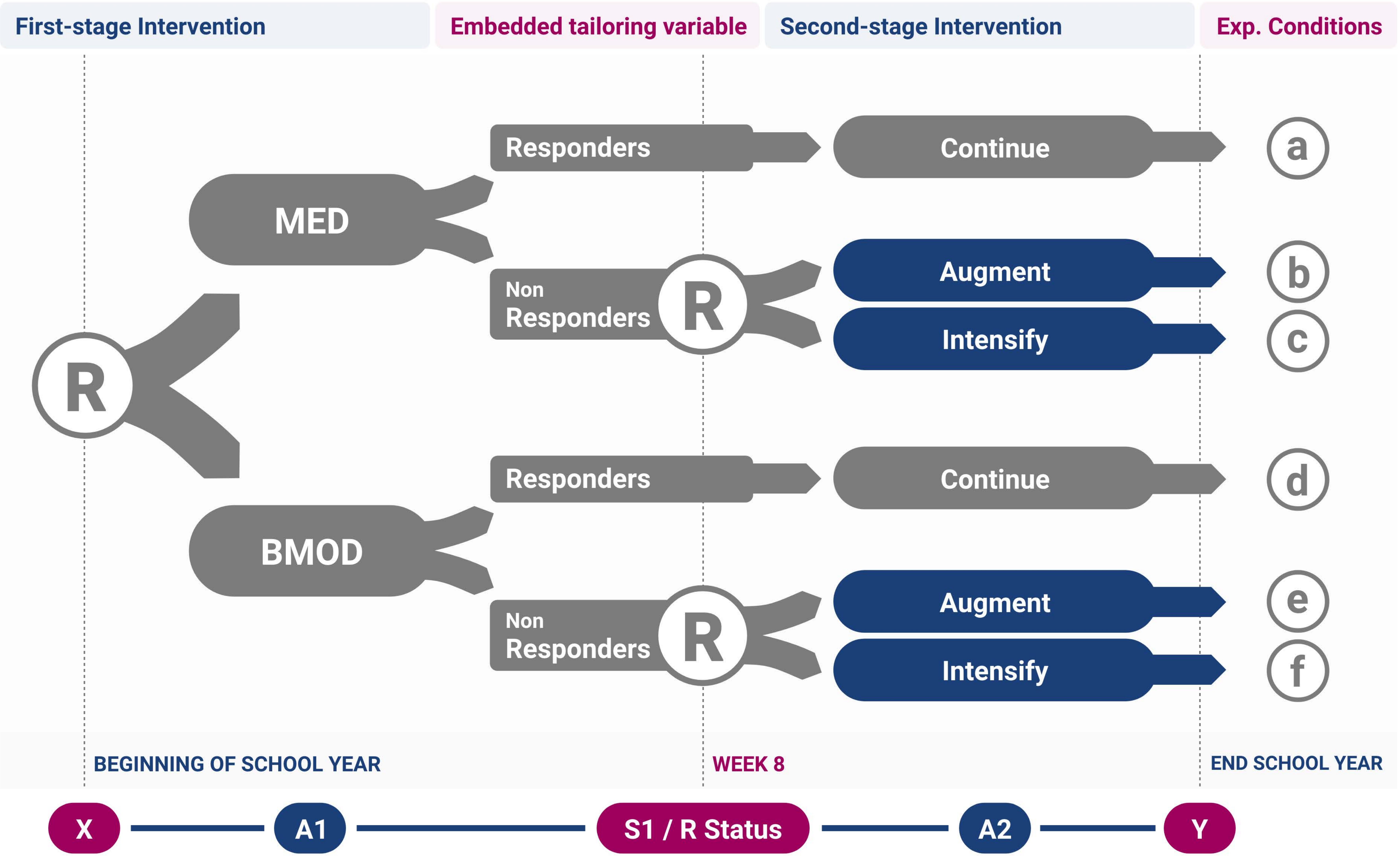
SMART Example ADHD Study

PI: Pelham



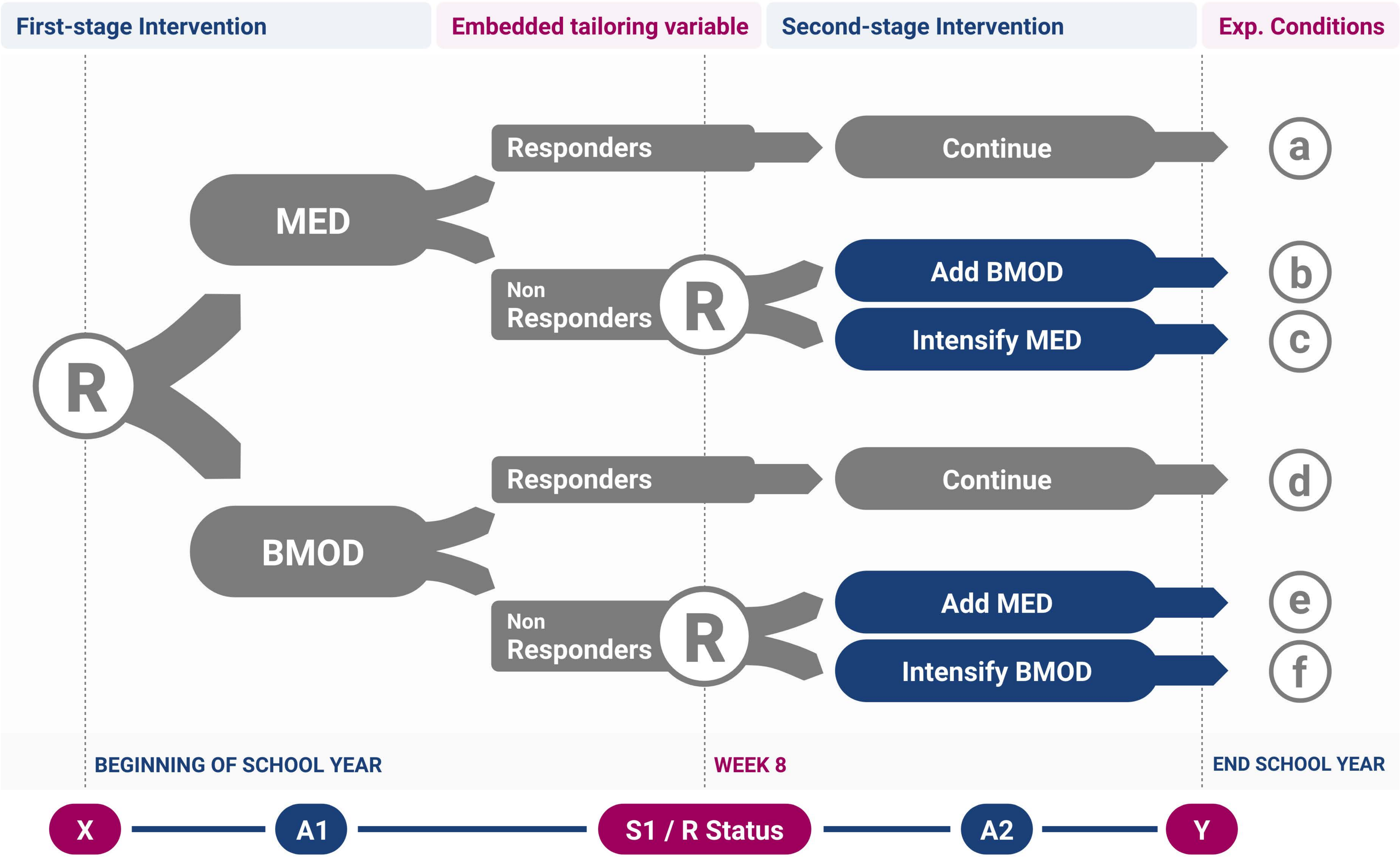
SMART Example ADHD Study

PI: Pelham



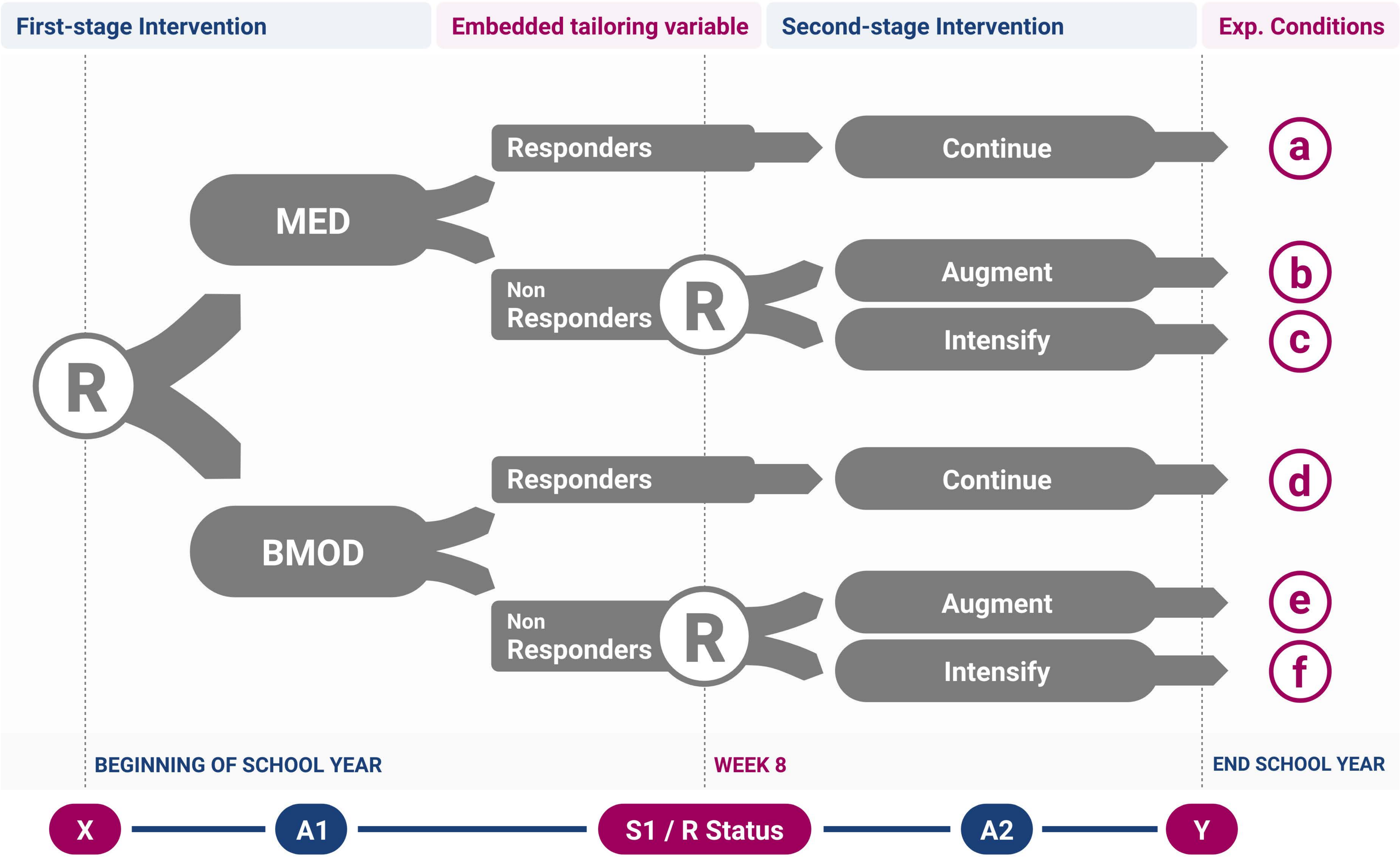
SMART Example ADHD Study

PI: Pelham



SMART Example **ADHD Study**

PI: Pelham



SMART Example ADHD Study

PI: Pelham

4 Embedded Adaptive Interventions

Adaptive Intervention 1

At the beginning of the school year

Stage 1 = {MED};

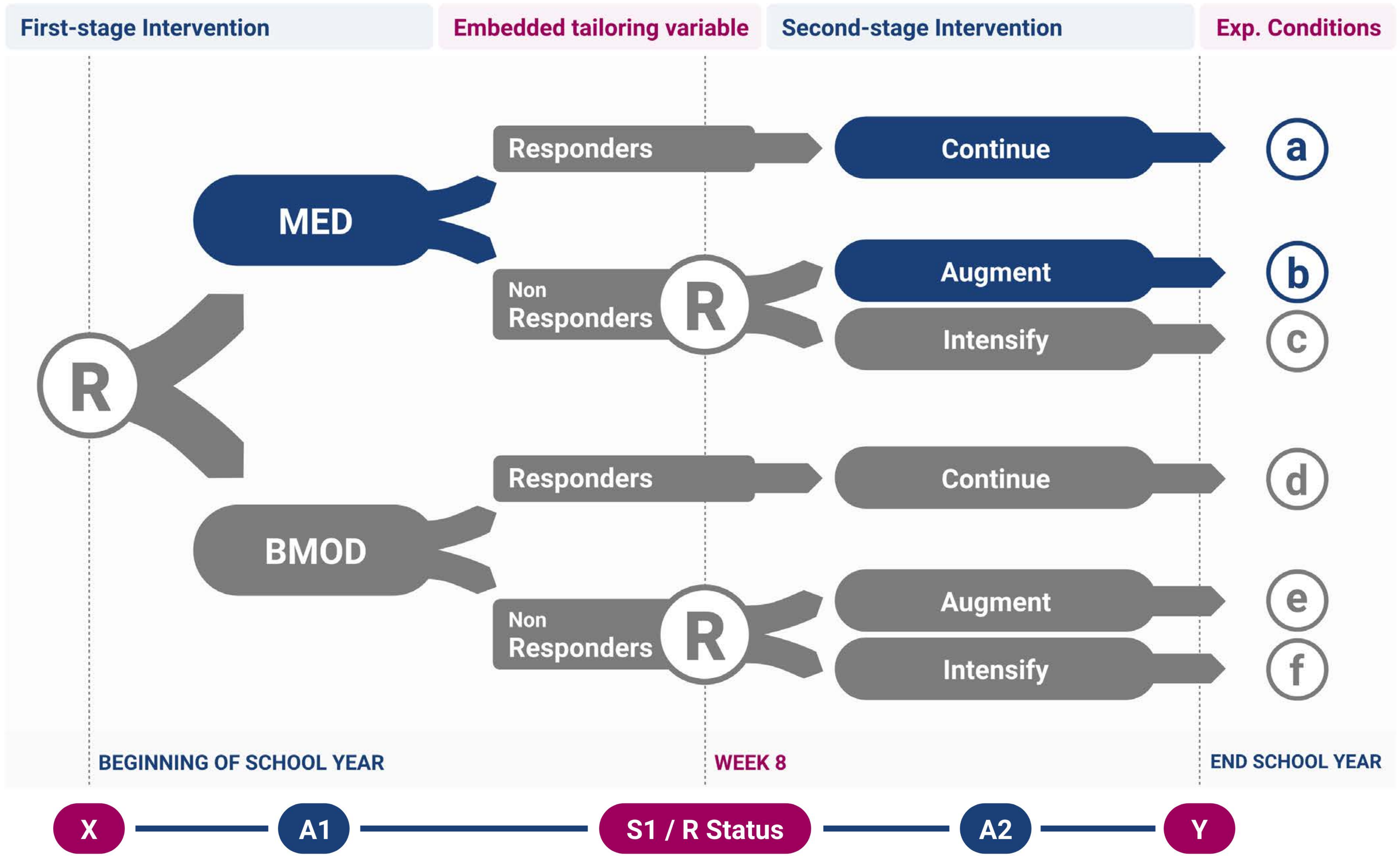
then, every month, starting week 8

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.

SMART Example ADHD Study

PI: Pelham

4 Embedded Adaptive Interventions

Adaptive Intervention 2

At the beginning of the school year

Stage 1 = {BMOD};

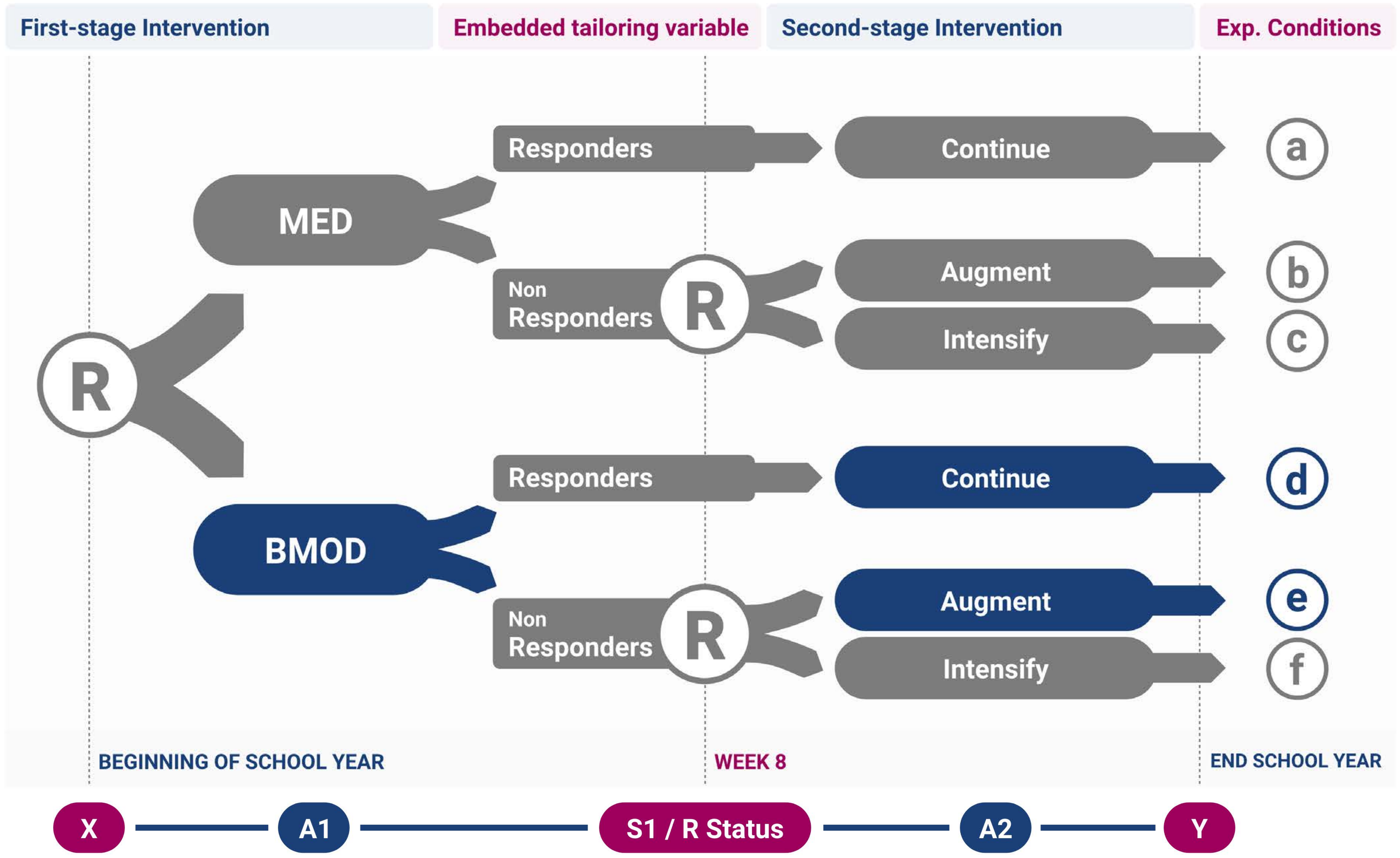
then, every month, starting week 8

if response status = {NR},

then, Stage 2 = {AUGMENT};

else if response status = {R},

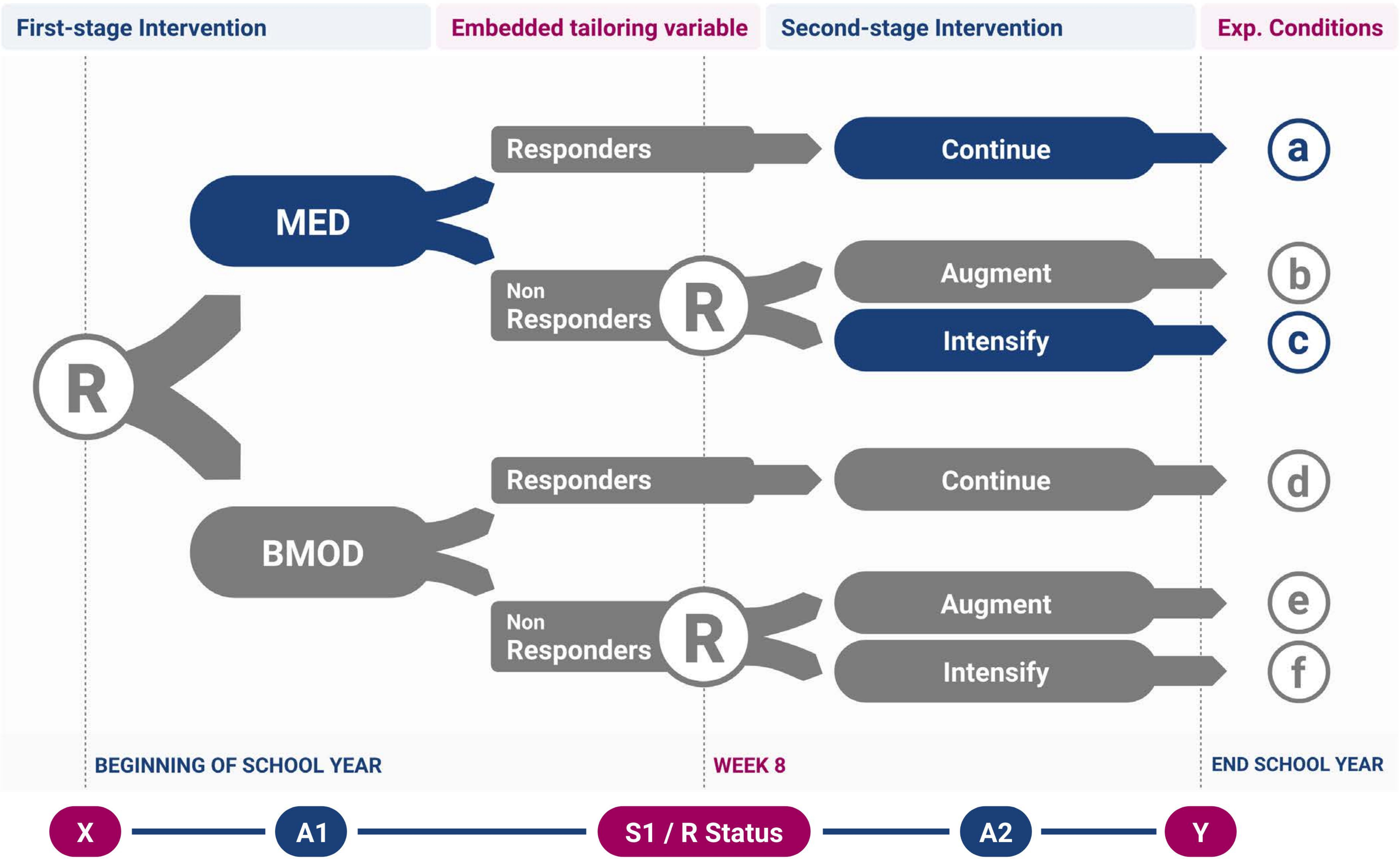
then, Continue Stage 1



4 Embedded Adaptive Interventions

Adaptive Intervention 3

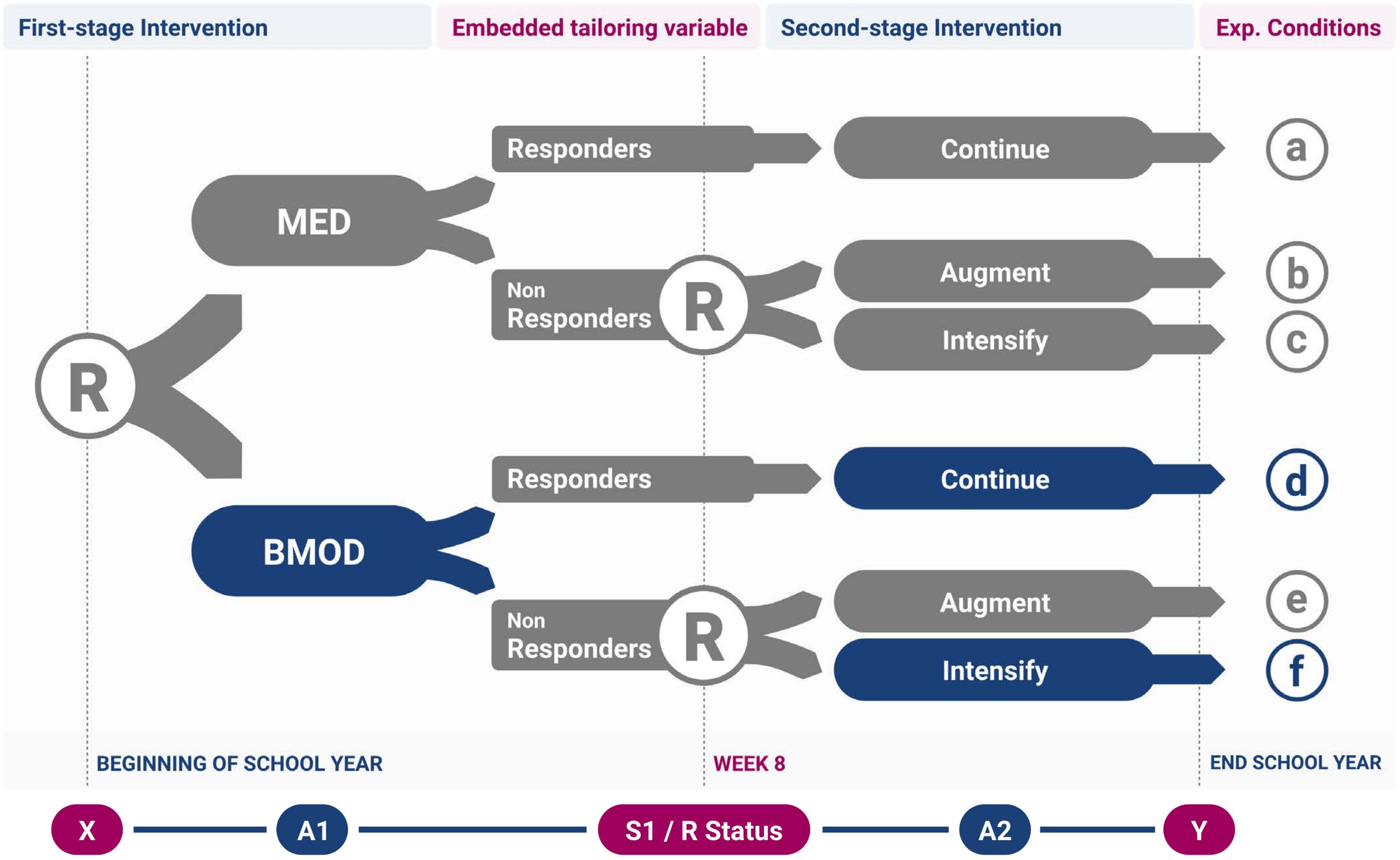
At the beginning of the school year
Stage 1 = {MED};
then, every month, starting week 8
if response status = {NR},
then, Stage 2 = {INTENSIFY};
else if response status = {R},
then, Continue Stage 1



4 Embedded Adaptive Interventions

Adaptive Intervention 4

At the beginning of the 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**

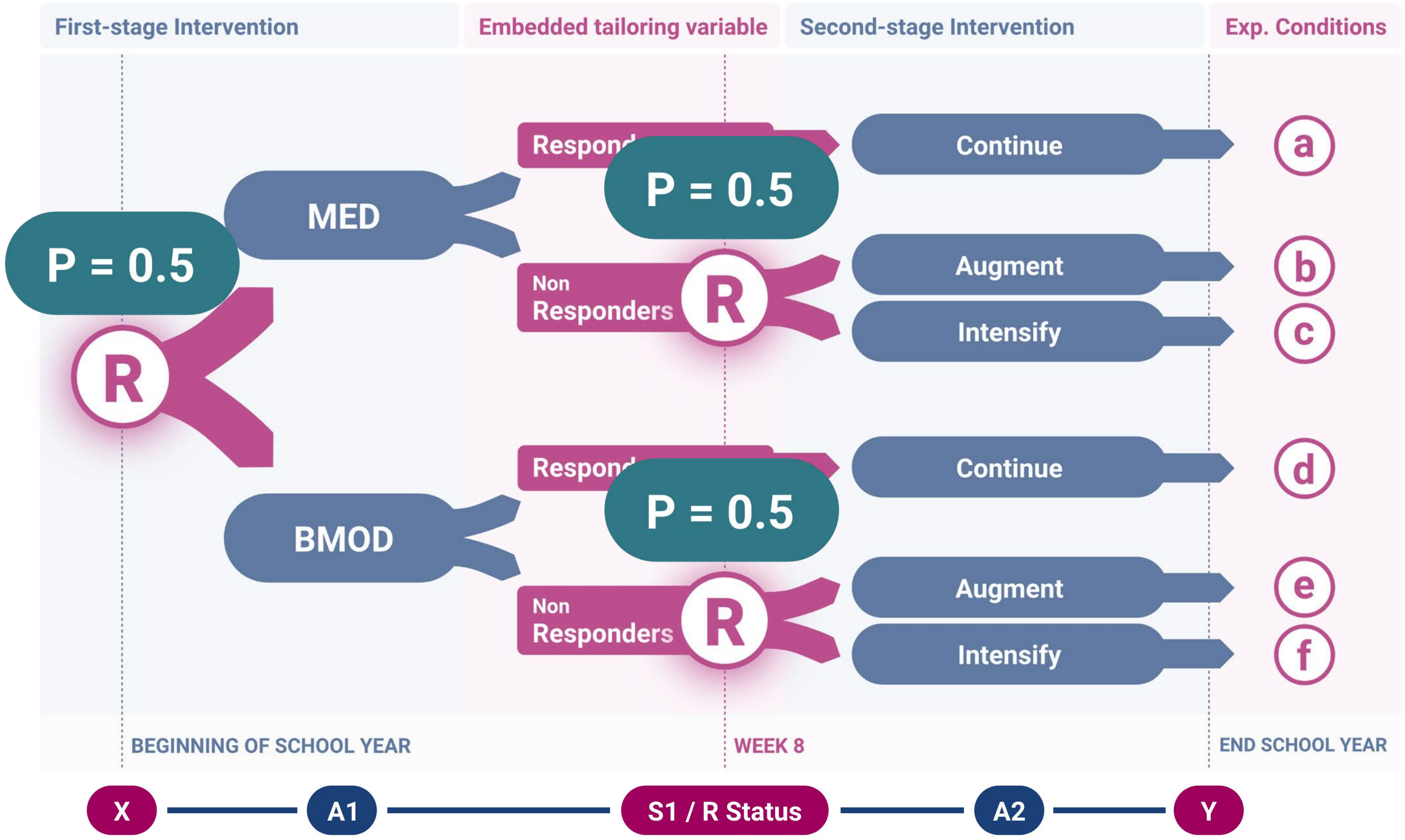


SMART Example ADHD Study

PI: Pelham

Sequential Randomizations

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



This data is simulated.

What the data looks like, Part I:

	ODD at baseline?	Baseline ADHD Score	Prior Med?	Race	Stage 1 Option
ID	odd	severity	priormed	race	A1
2	0	4.1	0 (NO)	0 (other)	1
6	0	5.5	0	1 (white)	1
7	0	6.8	0	1	1 (BMOD)
54	1 (YES)	2.6	0	1	-1 (MED)
59	0	3.5	0	1	-1
119	0	4.7	0	1	-1
…	…	…	1 (YES)	…	…

This data is simulated.

What the data looks like, Part II:

	Response/ Non-Response	Time until NR (months)	Adherence	Stage 2 Tactic	School Perfm
`ID`	`R`	`NRtime`	`adherence`	`A2`	`Y`
2	1 (R)	NA	0 (NO)	NA	4.3
6	0 (NR)	3	0	1 (INTSFY)	2.1
7	0	7	1 (YES)	1 (INTSFY)	2.6
54	1	NA	0	NA	2.9
59	0	5	1	-1 (AUG)	1.2
119	0	5	1	-1 (AUG)	0.9
…	…	…	…	…	…

Outline

Illustrative Example: ADHD SMART Study (PI: Pelham)

Data Analytics 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): Estimate the mean outcome under each of the embedded AIs (separately) using weighted least squares

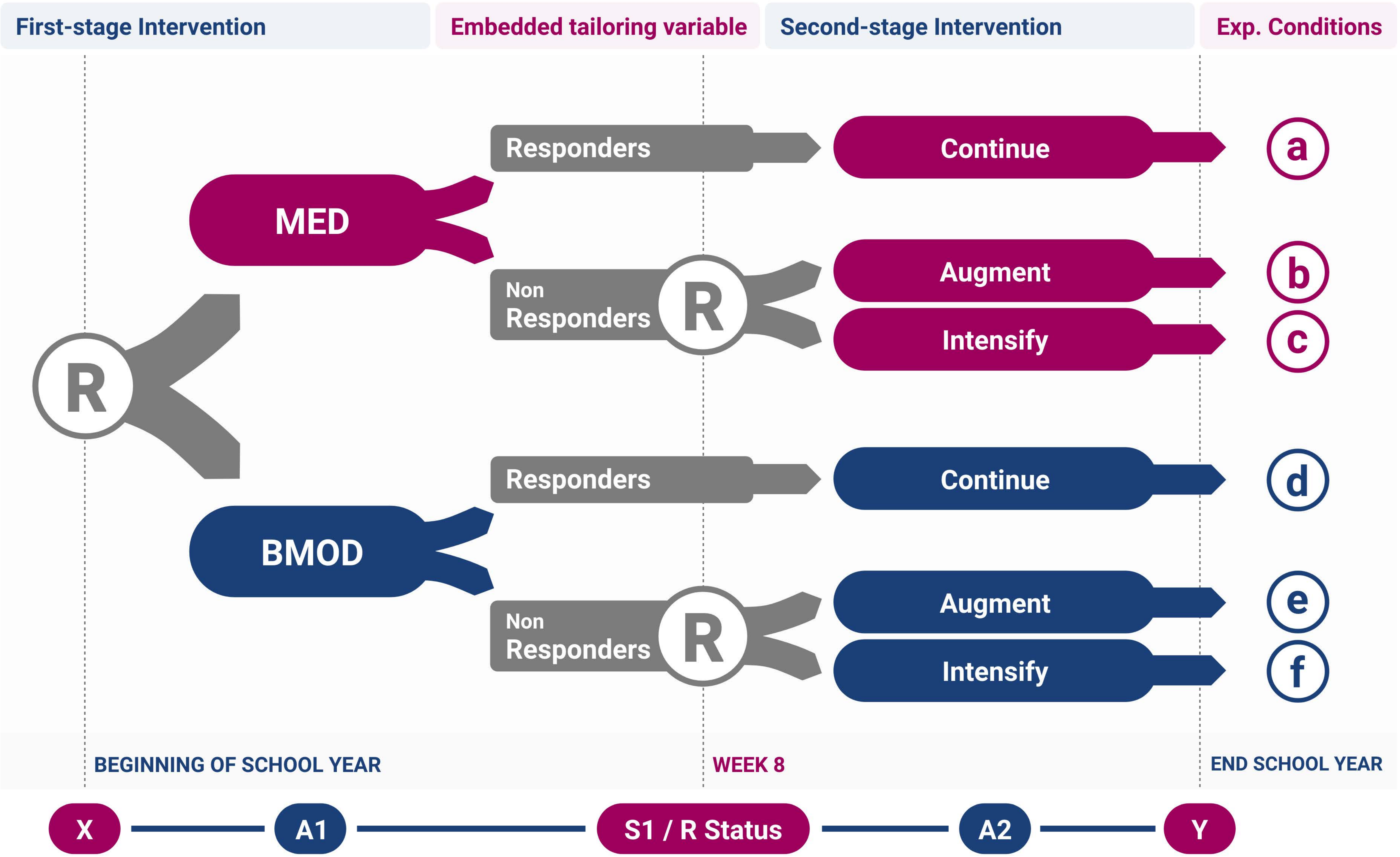
How to frame the question?

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?

Main Effect of Stage 1 Options

Simply a comparison of two groups:

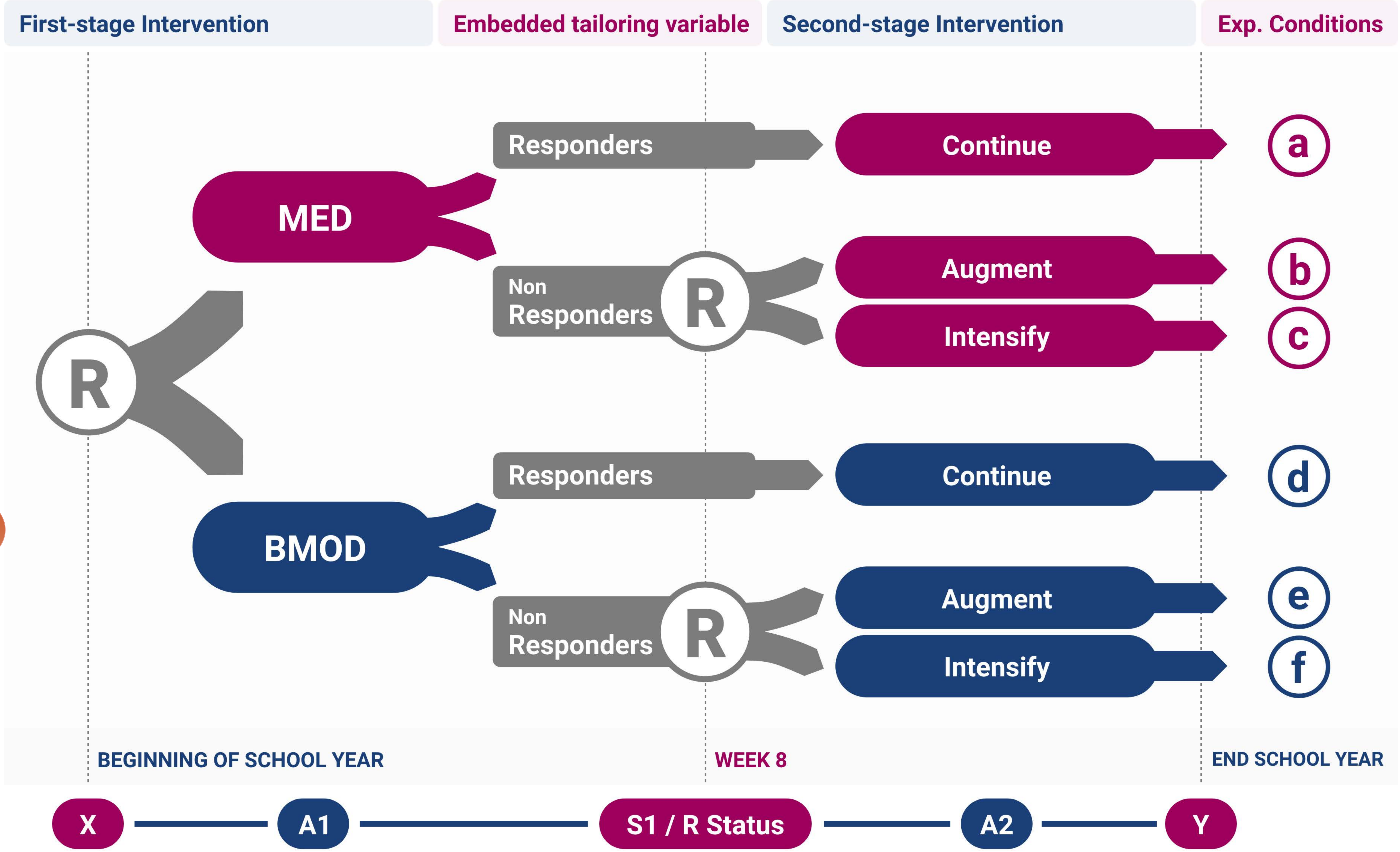
A two-sample t-test



Main Effect of Stage 1 Options

Simply a comparison of two groups:

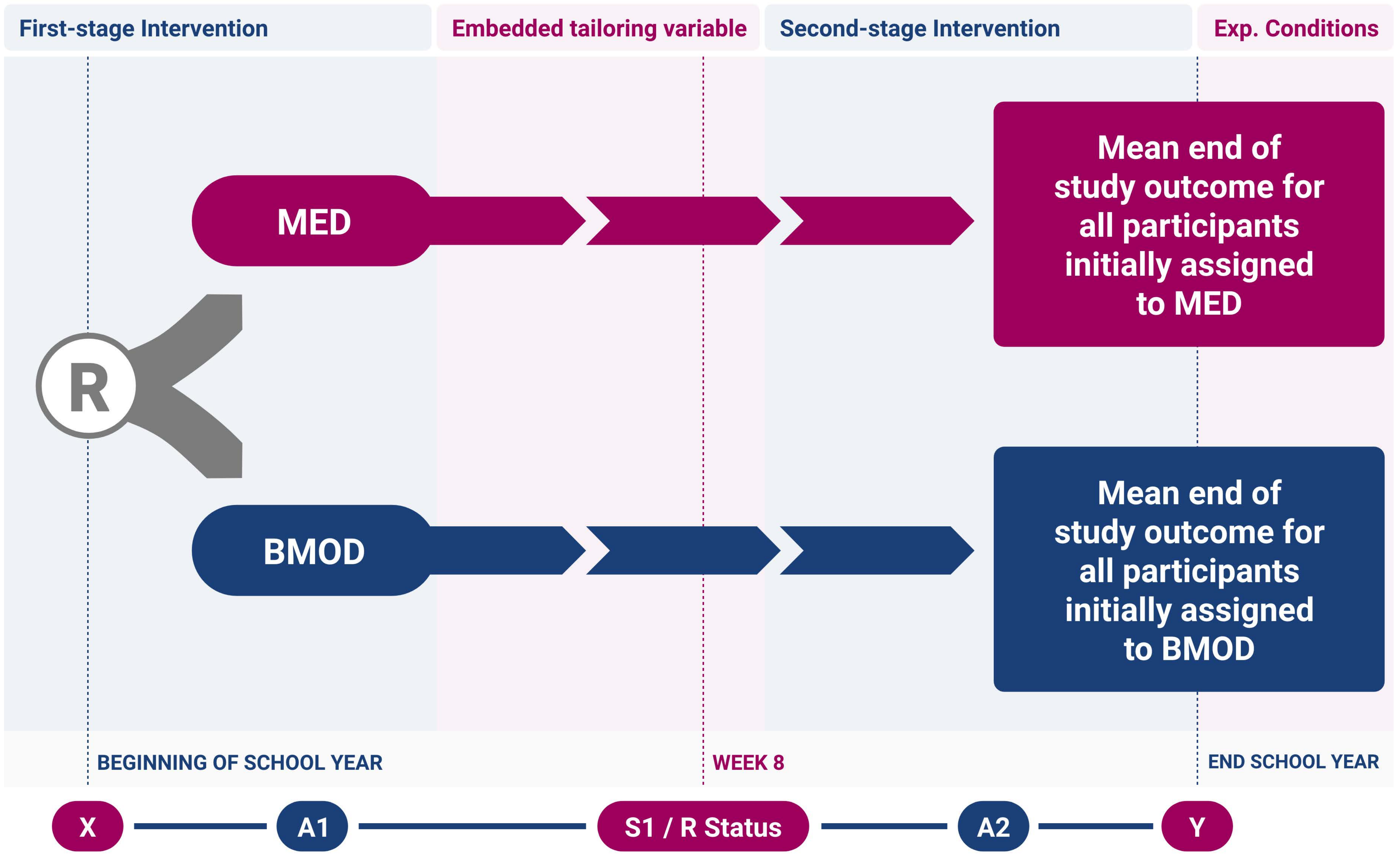
A two-sample t-test



You are ignoring subsequent treatments

Main Effect of Stage 1 Options

Think about a standard RCT, where “things happen” after treatment is offered…

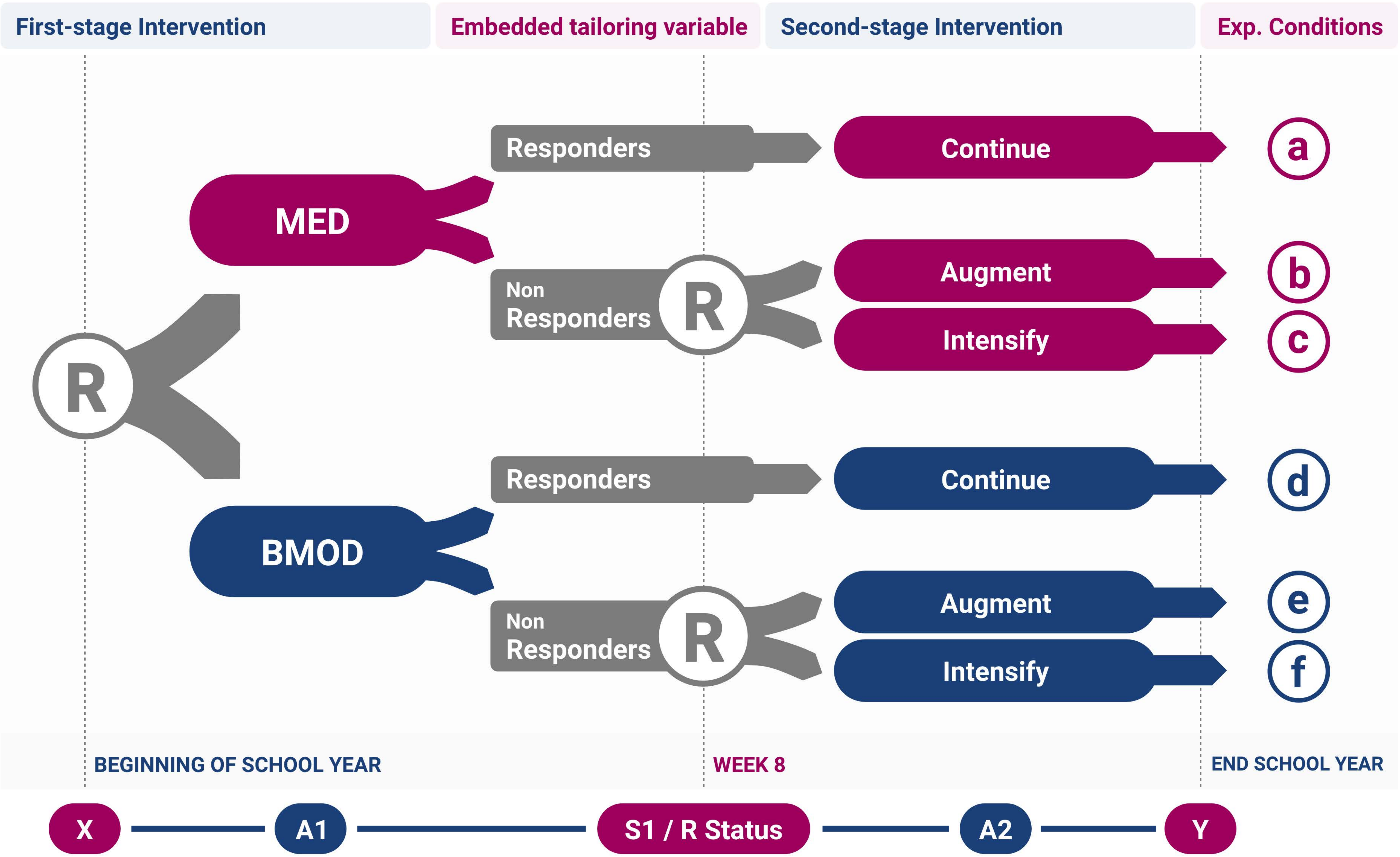


Typical Primary Aim 1

Main Effect of Stage 1 Options

PI: Pelham

Not ignoring; averaging over!



Before we show you SAS code…

Review Coding Scheme

Recall A1 = 1 => BMOD

A1 = -1 => MED

The Regression and Contrast Coding Logic:

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

or you can fit

c for centered

$$E[Y|A_1, \boldsymbol{X}] = b_0 + b_1 A_1 + b_2 X_{1c} + b_3 X_{2c} + b_4 X_{3c} + b_5 X_{4c}$$

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

$$E[Y|A_1, \boldsymbol{X}] = b_0 + b_1 A_1 + b_2 X_{1c} + b_3 X_{2c} + b_4 X_{3c} + b_5 X_{4c}$$

```
proc genmod data = dat1;
  model Y = A1    X1c X2c X3c X4c;
  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    X1c X2c X3c X4c;
  estimate 'Mean Y under BMOD' intercept 1 A1  1 X1c 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    X1c X2c X3c X4c;
  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:

$$E[Y|A_1, \boldsymbol{X}] = b_0 + b_1 A_1 + b_2 X_{1c} + b_3 X_{2c} + b_4 X_{3c} + b_5 X_{4c}$$

- 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    X1c X2c X3c X4c;
  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:

$$E[Y|A_1, \boldsymbol{X}] = b_0 + b_1 A_1 + b_2 X_{1c} + b_3 X_{2c} + b_4 X_{3c} + b_5 X_{4c}$$

- 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    X1c X2c X3c X4c;
  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:

$$E[Y|A_1, \boldsymbol{X}] = b_0 + b_1 A_1 + b_2 X_{1c} + b_3 X_{2c} + b_4 X_{3c} + b_5 X_{4c}$$

- 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

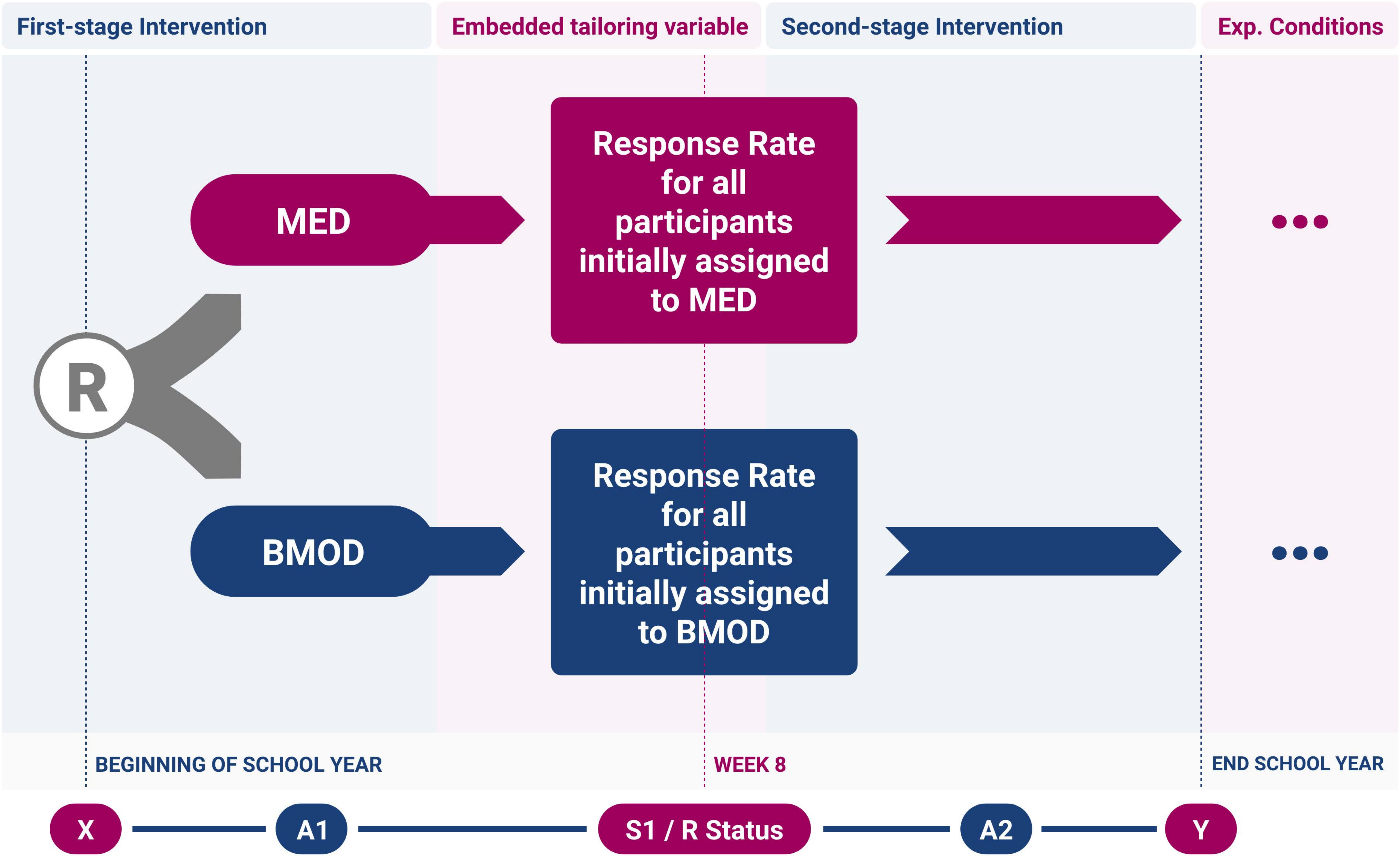
Results are from simulated data.

Contrast Estimate Results					
	Mean	95% Confidence Limits		Standard	
Label	Estimate	Lower	Upper	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).

Side Analysis

Effect of Stage 1 Options on NR Rate



Results of Side Analysis

Effect of Stage 1 Options on NR Rate

Results are from simulated data.

```
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)	45 62.67%	26 **36.1%**	72
1 (BMOD)	55 69.33%	23 **29.5%**	78
Total	101	49	150

In terms of early response rate, initial MED is slightly better (vs. BMOD) by ~7%

Outline

Illustrative Example: ADHD SMART Study (PI: Pelham)

Data Analytics 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): Estimate the mean outcome under each of the embedded AIs (separately) using weighted least squares

How to frame the question?

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

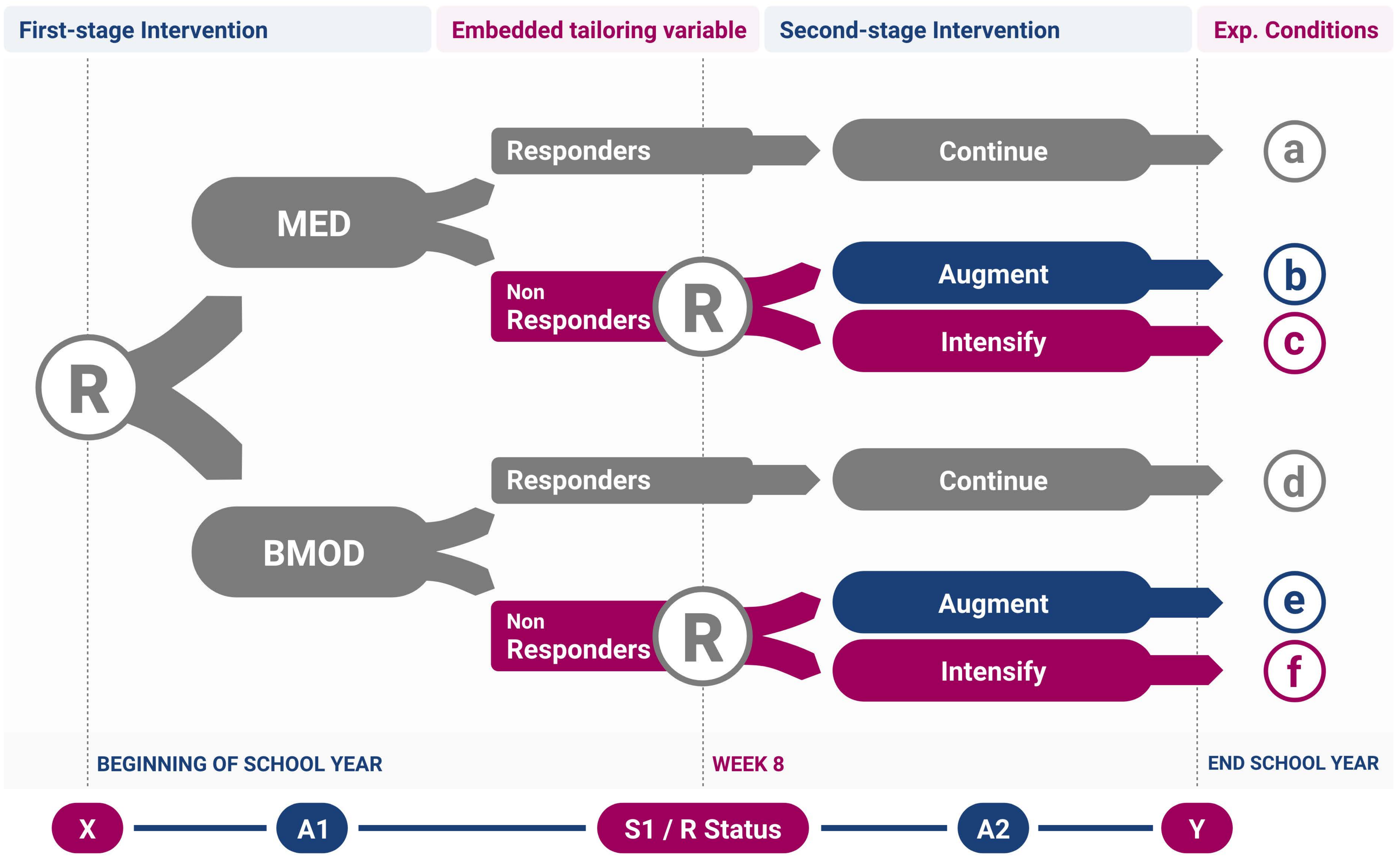
…in terms of end of study school performance

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

Typical Primary Aim 2

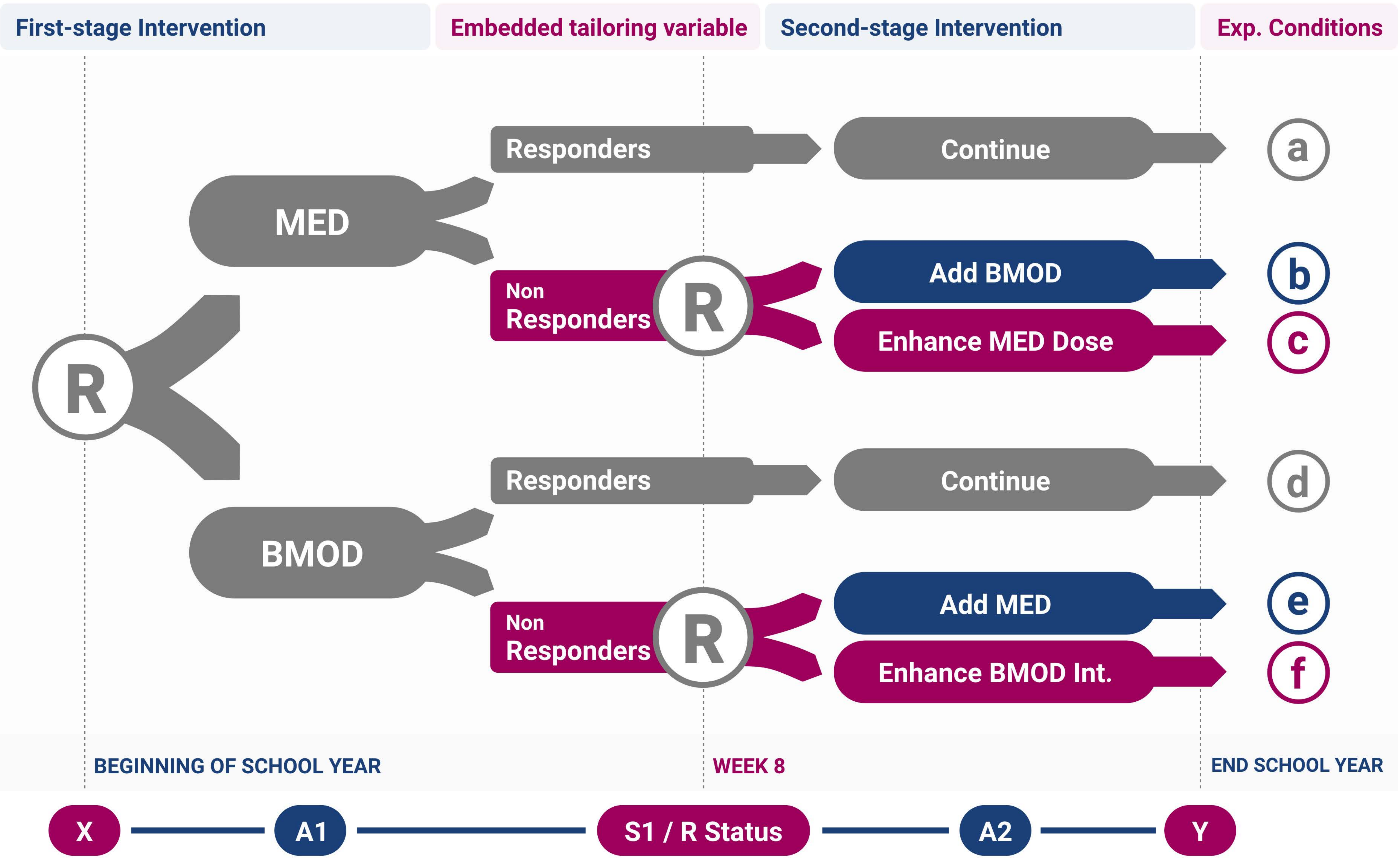
Main Effect of Stage 2 Tactics

PI: Pelham



Typical Primary Aim 2 — Main Effect of Stage 2 Tactics

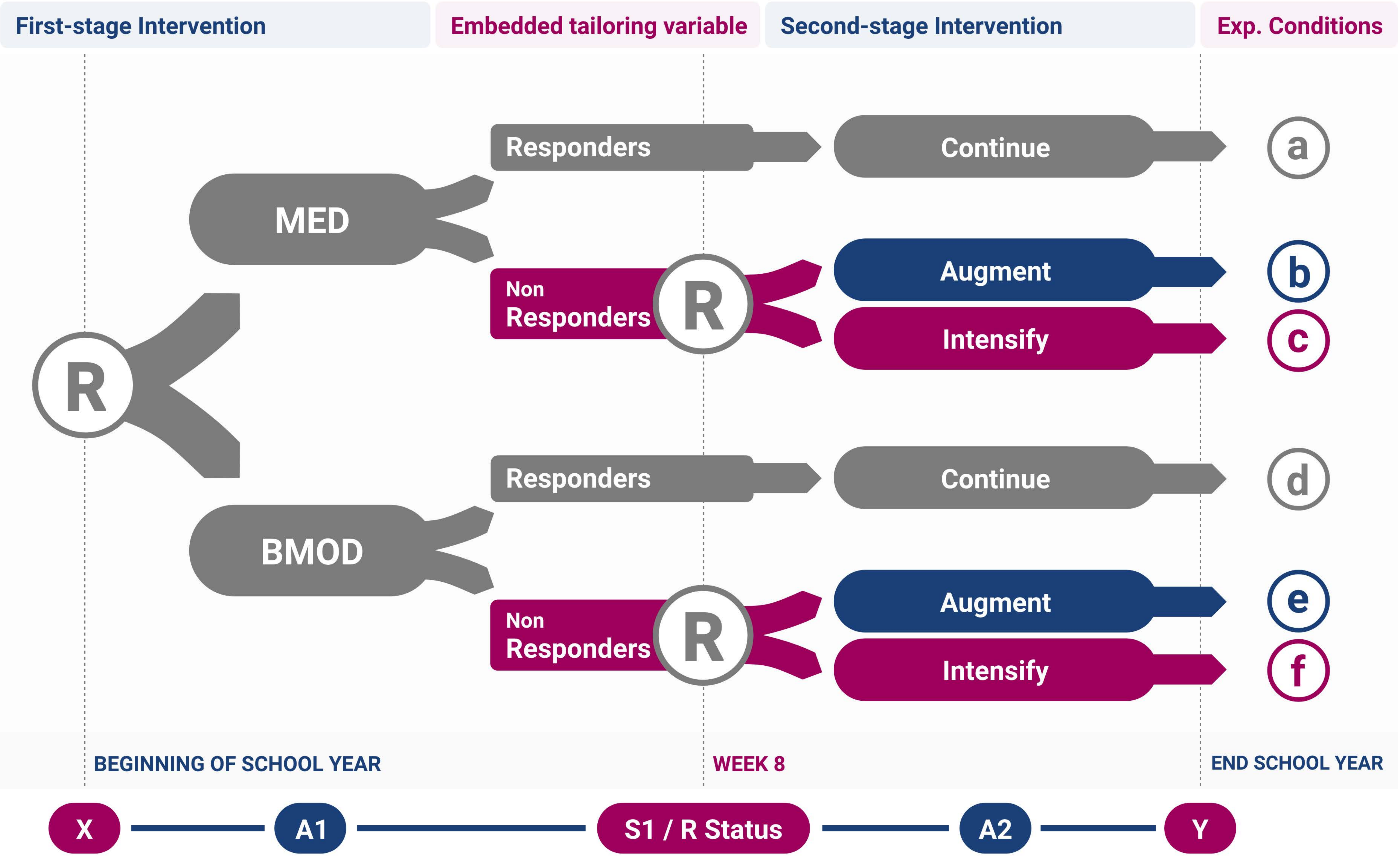
PI: Pelham

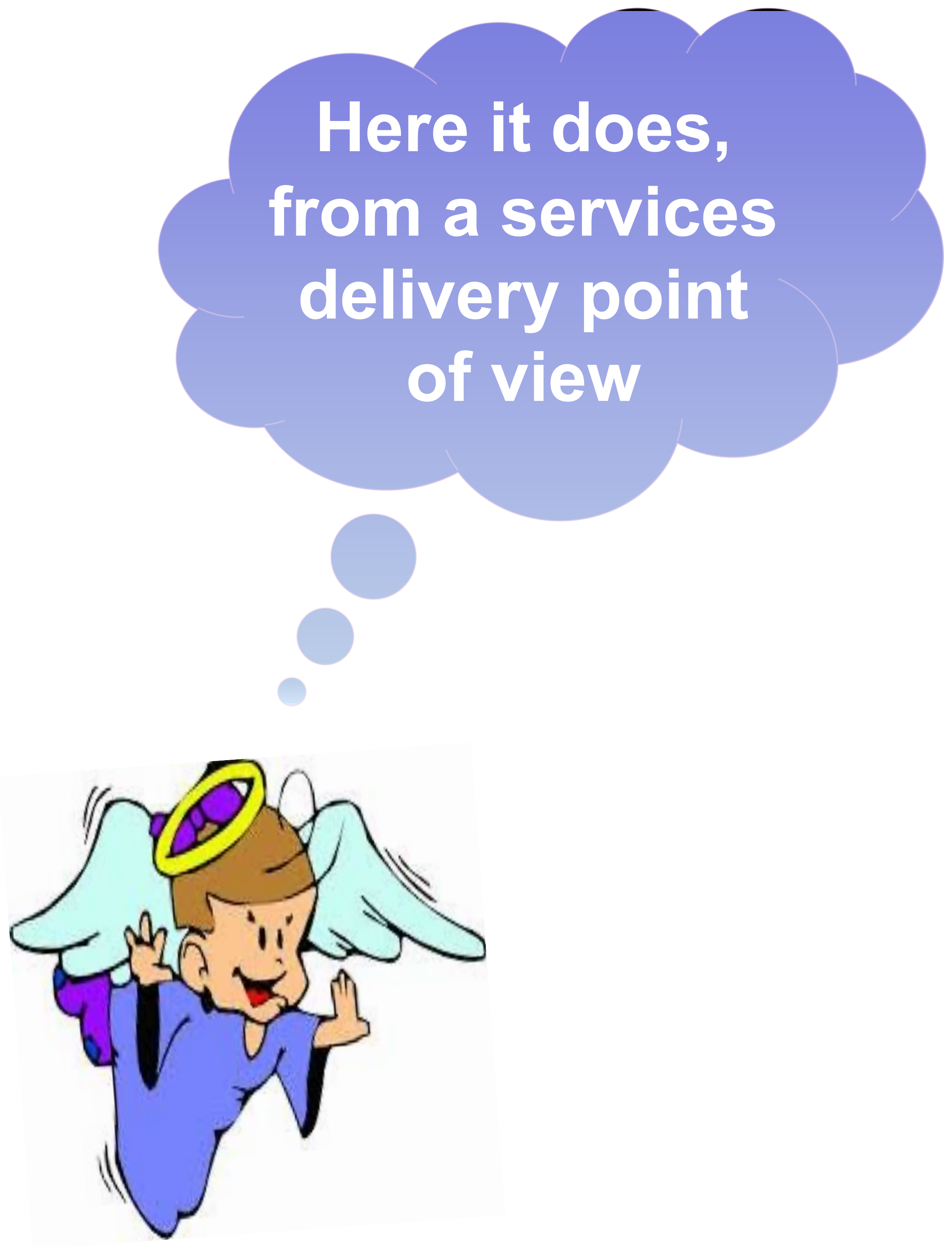


Typical Primary Aim 2

Main Effect of Stage 2 Tactics

PI: Pelham





Before we show you SAS code…

Review Coding Scheme

Recall A2 = 1 → INTENSIFY
A2 = -1 → AUGMENT

The Regression and Contrast Coding Logic:

$$E[Y|A_2, R = 0] = b_0 + b_1 A_2$$

This regression is among Non-responders only.

or you can fit with covariates

$$E[Y|A_2, \boldsymbol{X}, \boldsymbol{S_1}, R = 0] = b_0 + b_1 A_2 + b^T covariates$$

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

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

* run the regression;
proc genmod data = dat3;
           model Y = A2 Y0c oddc severityc priormedc adherencec NRtimec ;
           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;
```

This regression is among non-responders only

'c' means we center covariates around the mean [among non-responders]

SAS Code for Aim 2

```
proc genmod data = dat3;
    model Y = A2 Y0c oddc severityc priormedc adherencec NRtimec;
    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:

- $E[Y|A_2, \boldsymbol{X}, \boldsymbol{S_1}, R = 0] = b_0 + b_1 A_2 + bT covariates$
- 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

Results are from simulated data.

Contrast Estimate Results					
		95% Confidence Limits			
Label	Mean Estimate	Lower	Upper	Standard Error	Pr > ChiSq
Mean Y INTENSIFY tactic	2.316	1.9499	2.6838	0.187	<.0001
Mean Y AUGMENT tactic	3.111	2.6886	3.5336	0.216	<.0001
Between groups difference	**-0.7942**	-1.3658	-0.2227	**0.292**	**0.0065**

- 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

Outline

Illustrative Example: ADHD SMART Study (PI: Pelham)

Data Analytics 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): Estimate the mean outcome under each of the embedded AIs (separately) using weighted least squares

Best of 2 Design-Embedded AIs

Adaptive Intervention 1

At the beginning of the 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

Adaptive Intervention 2

At the beginning of the 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**

How to frame this question?

To investigate whether and AI that recommends to

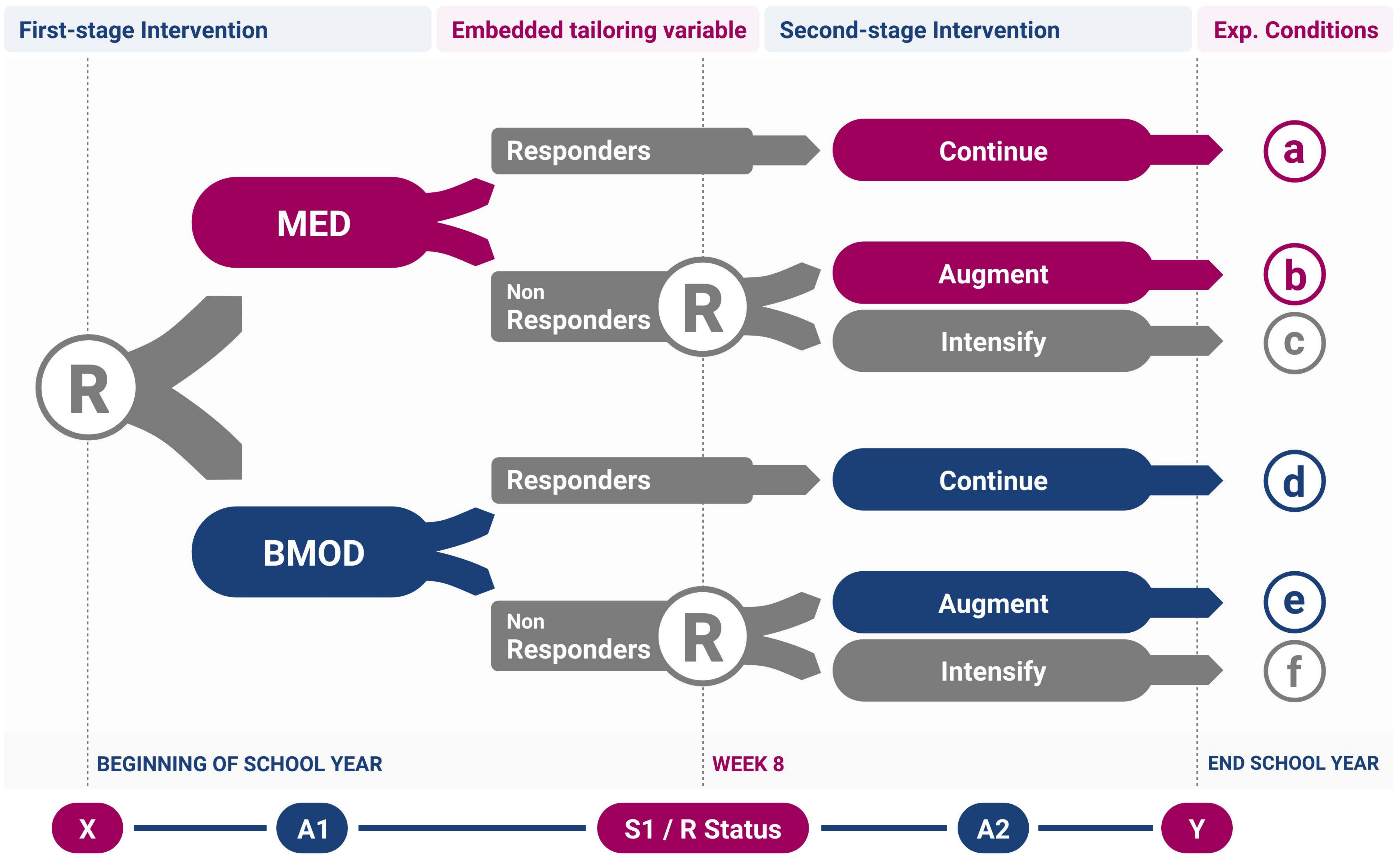
Start with BMOD; if non-responder AUGMENT [BMOD + MED], else continue [BMOD]

is better than an AI that recommends to

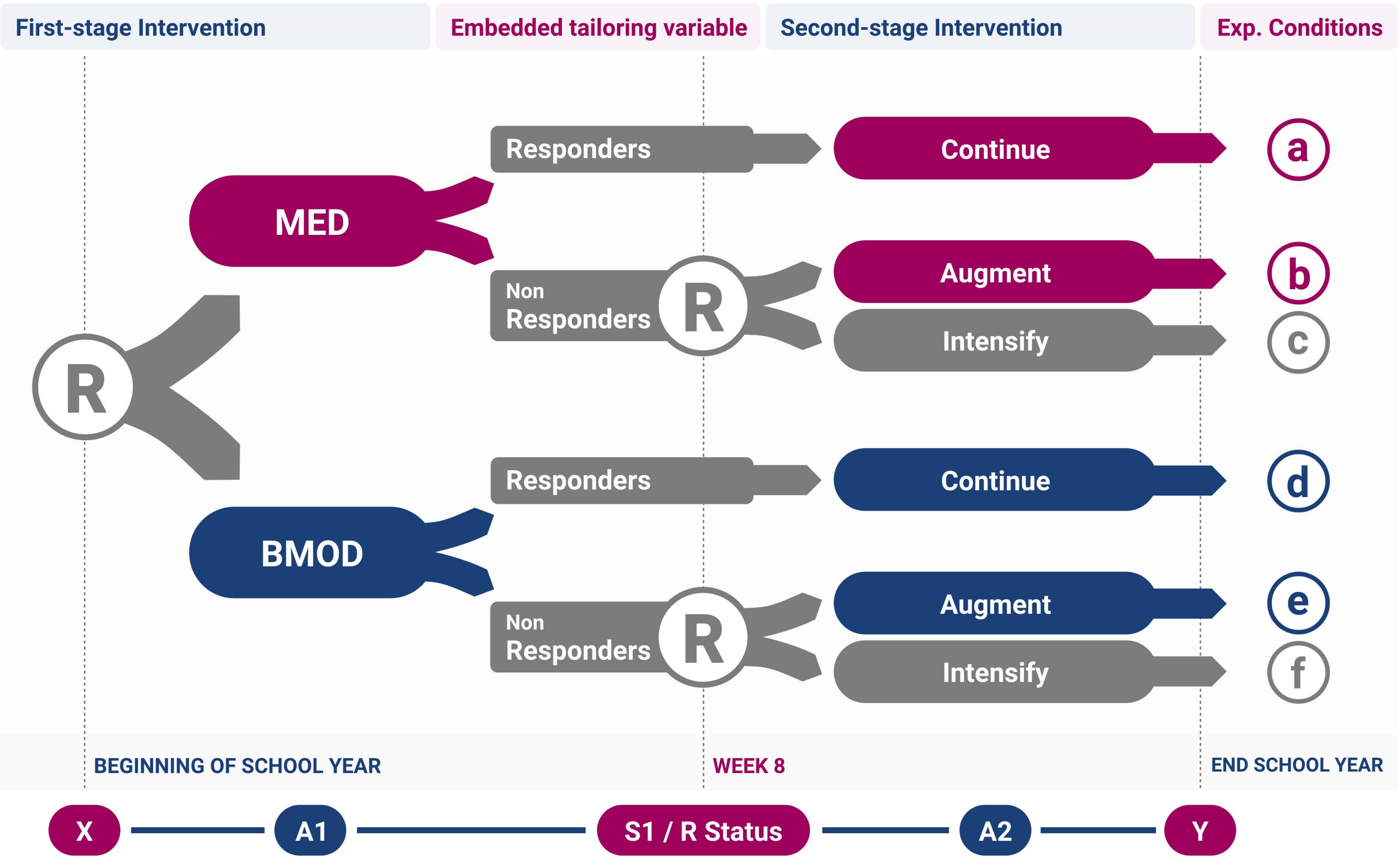
Start with MED; if non-responder AUGMENT [BMOD+MED], else continue [BMOD]

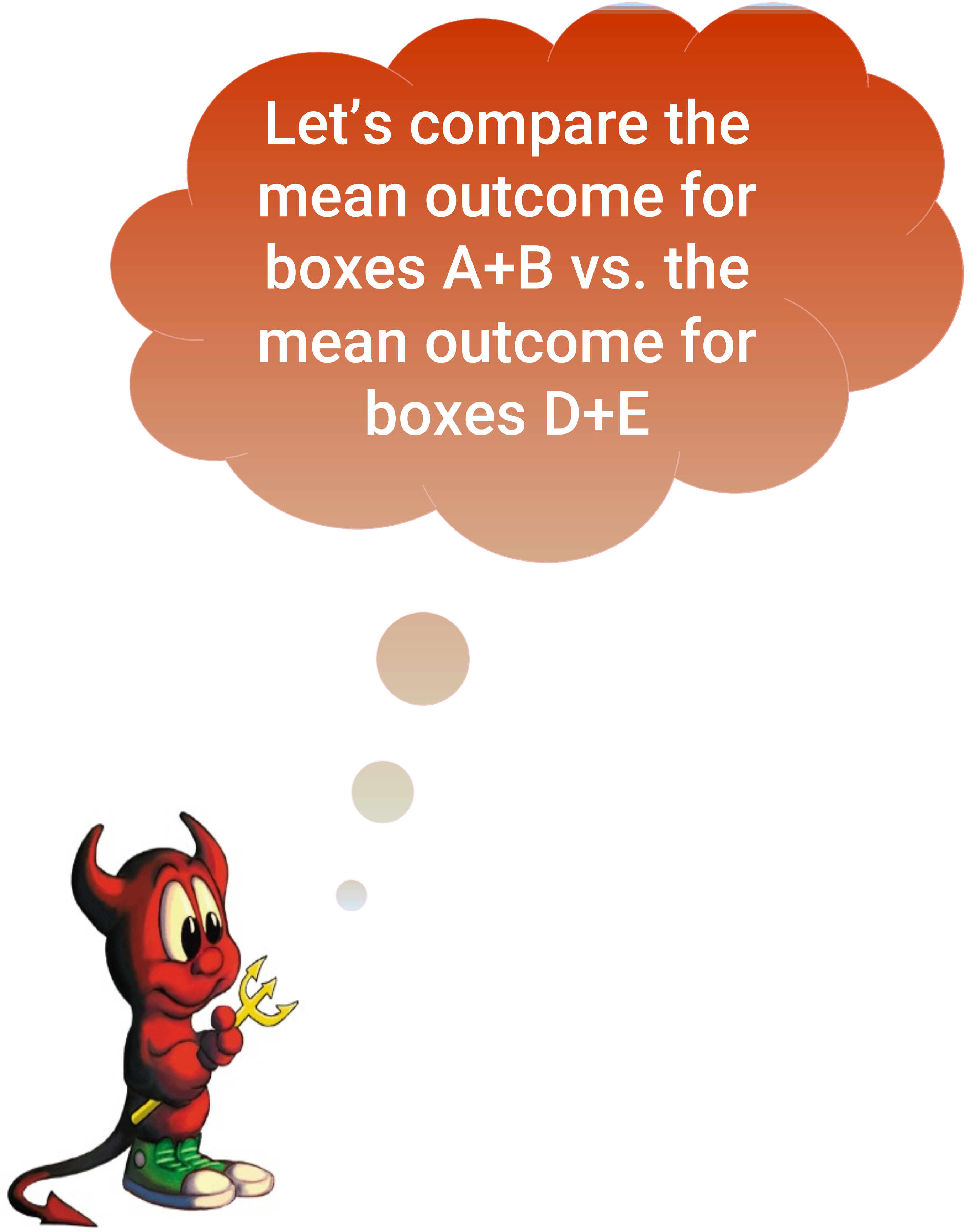
…in terms of end of study school performance.

Comparison of Mean Outcome Had Population Followed AI#1 vs. AI#2

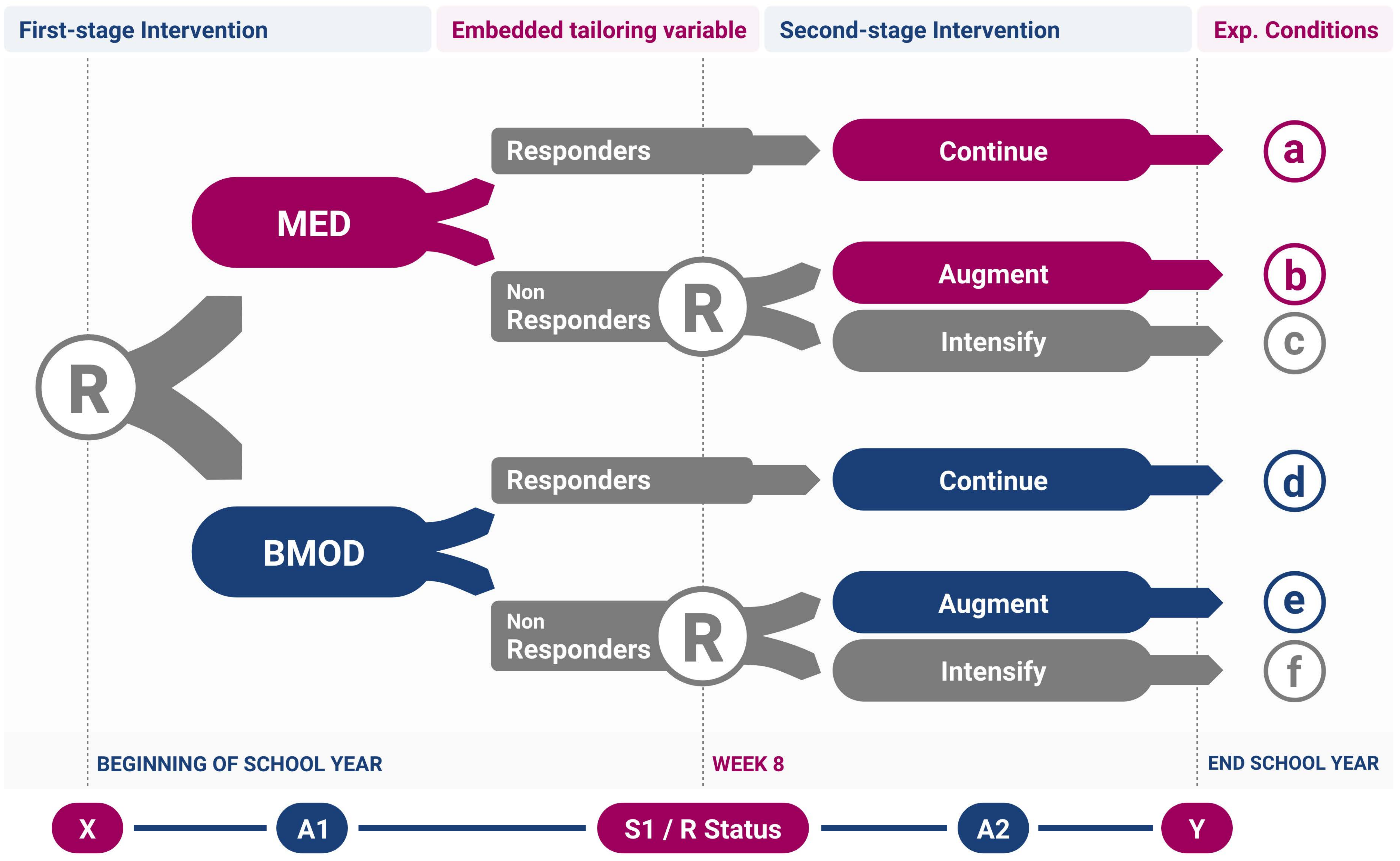


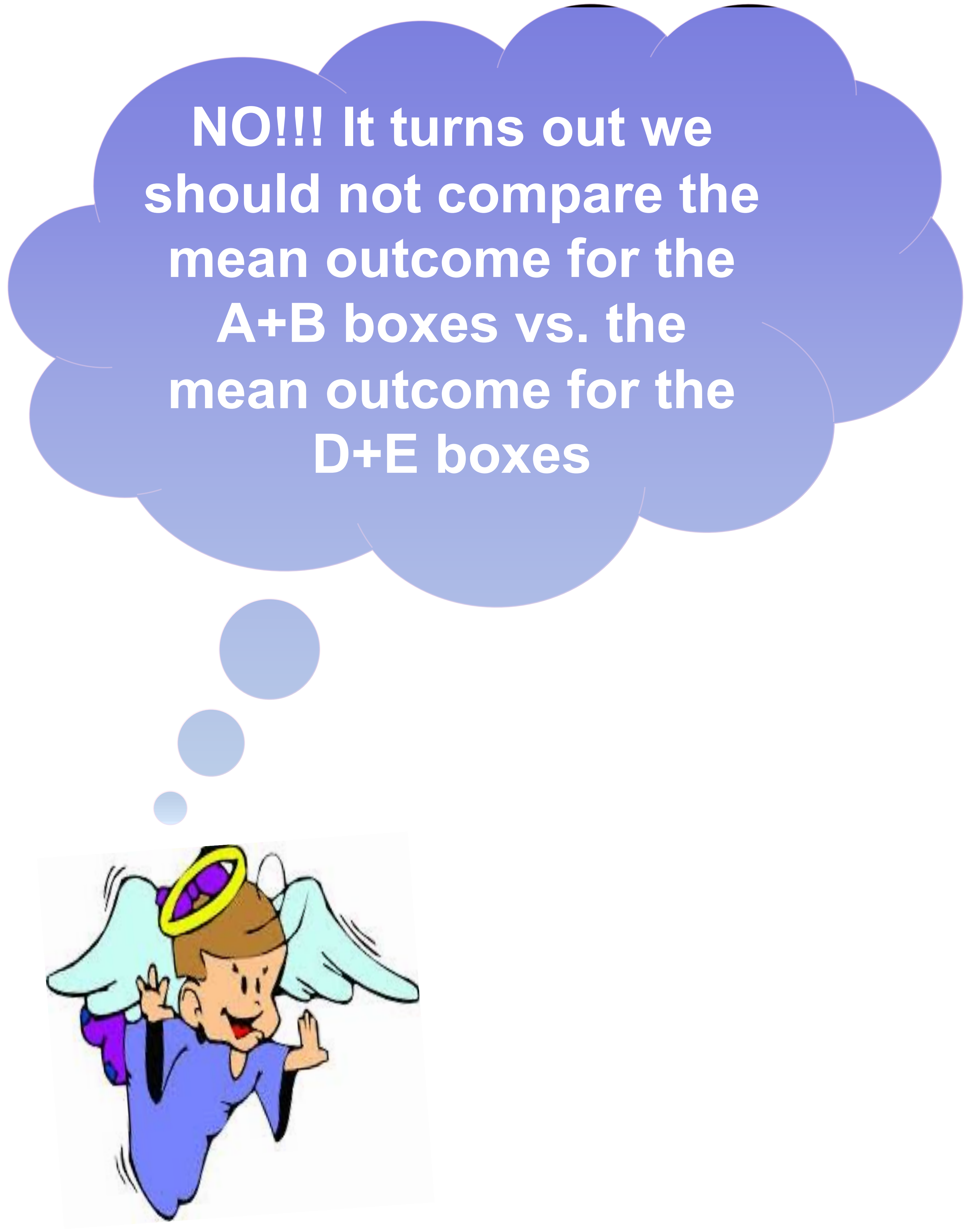
Comparison of Mean Outcome Had Population Followed AI#1 vs. AI#2





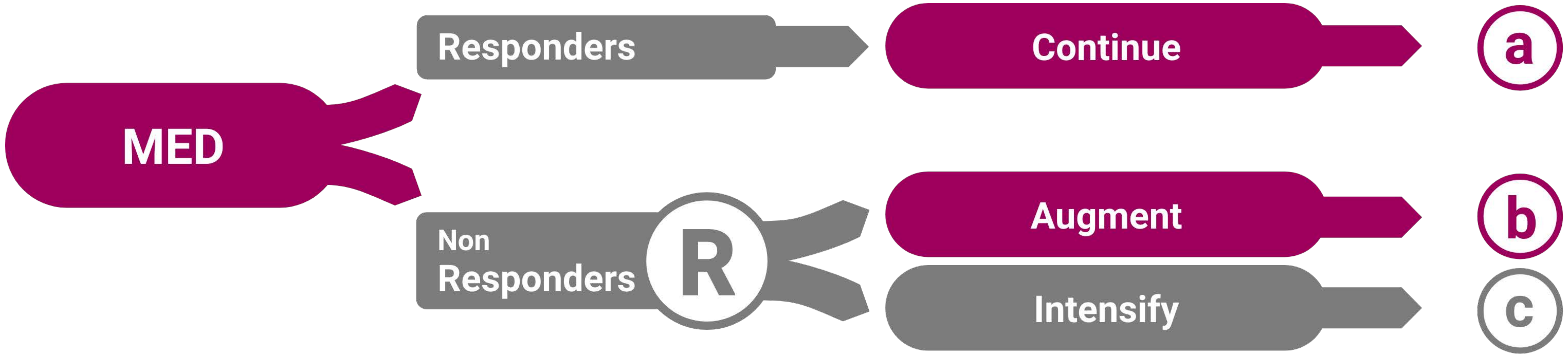
Comparison of Mean Outcome Had Population Followed AI#1 vs. AI#2



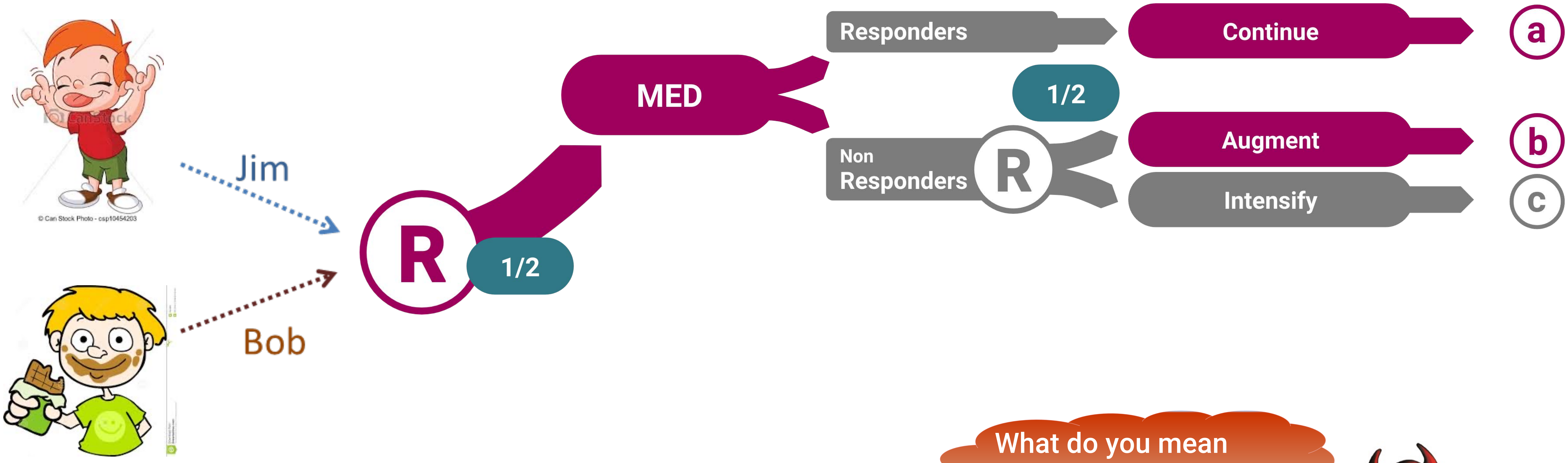


Had Population Followed AI#1 vs. AI#2

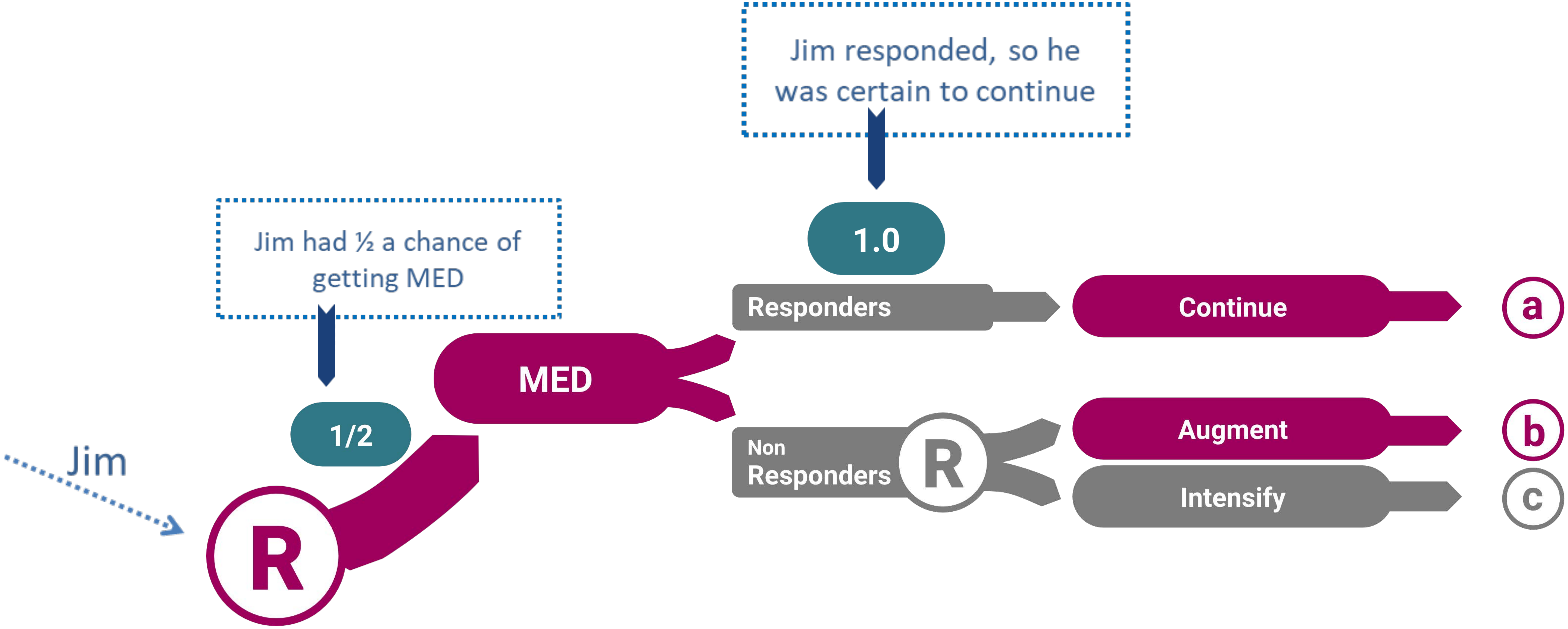
To understand this, we first, we learn how to obtain mean outcome under AI#1 (MED, AUGMENT)



Comparison of Mean Outcome Had Population Followed AI#1 vs. AI#2



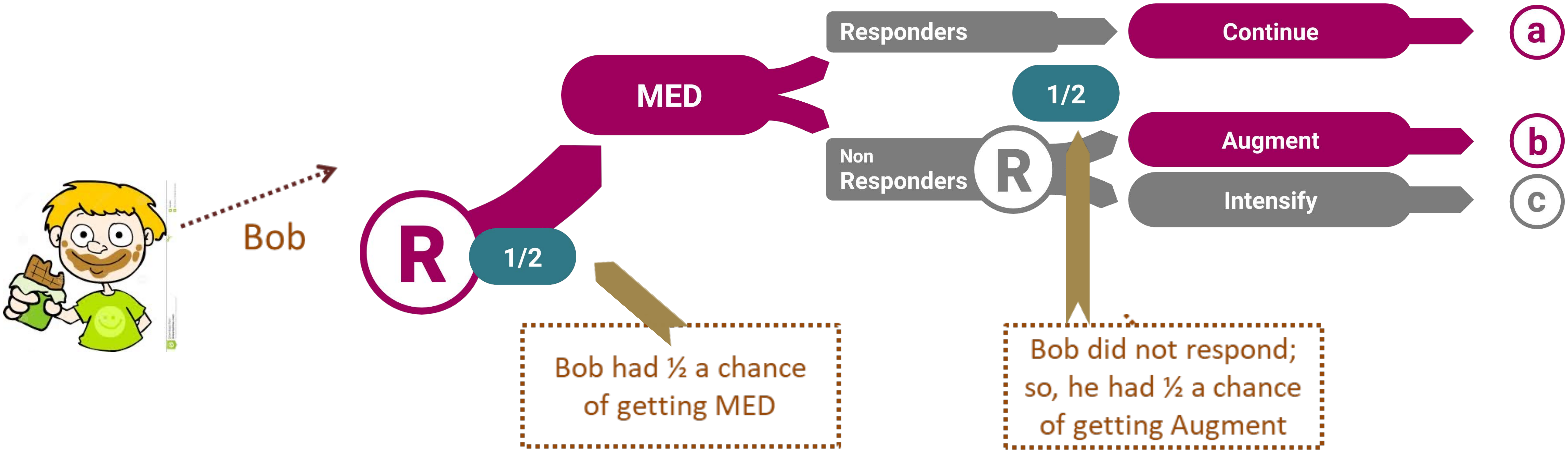
Comparison of Mean Outcome Had Population Followed AI#1 vs. AI#2



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

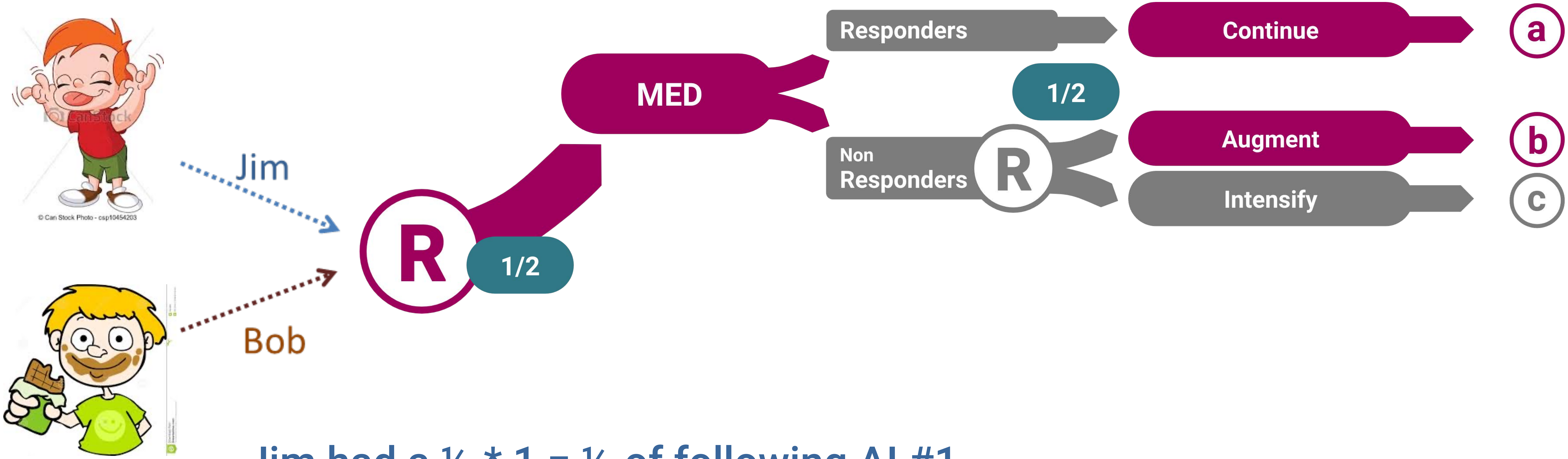
Had Population Followed AI#1 vs. AI#2

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



Comparison of Mean Outcome

Had Population Followed AI#1 vs. AI#2



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

Bob had ½ * ½ = 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: ½ a chance of following AI #1

Non-Responders: ¼ chance of following AI #1

Imbalance

What can we do? We can fix this imbalance by

- Assigning W = weight = 2 to responders to MED $2 \times \frac{1}{2} = 1$
- Assign W = weight = 4 to non-responders to MED $4 \times \frac{1}{4} = 1$
- 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)

Results are from simulated data.

Analysis Of GEE Parameter Estimates			
Parameter	Estimate	Standard Error	Pr > |Z|
Intercept	2.7790	0.146	<.0001
Z1	-0.1129	0.146	0.4392

Contrast Estimate Results

Label	Mean Estimate	95% Confidence Limits		Standard Error	Pr > ChiSq
		Lower	Upper		
Mean Y under AI #1 (MED, AUGMENT)	**2.66**	2.243	3.089	**0.216**	**<.0001**

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.

Q&A

10 min