Penn Medicine

CHIPS RESEARCH TRAINING INSTITUTE 2024

Practical Design and Analytic Considerations for SMARTs

Nicholas J. Seewald, PhD
Assistant Professor of Biostatistics

May 13, 2024

Slides are online!

`slides.nickseewald.com/chips2024.pdf`

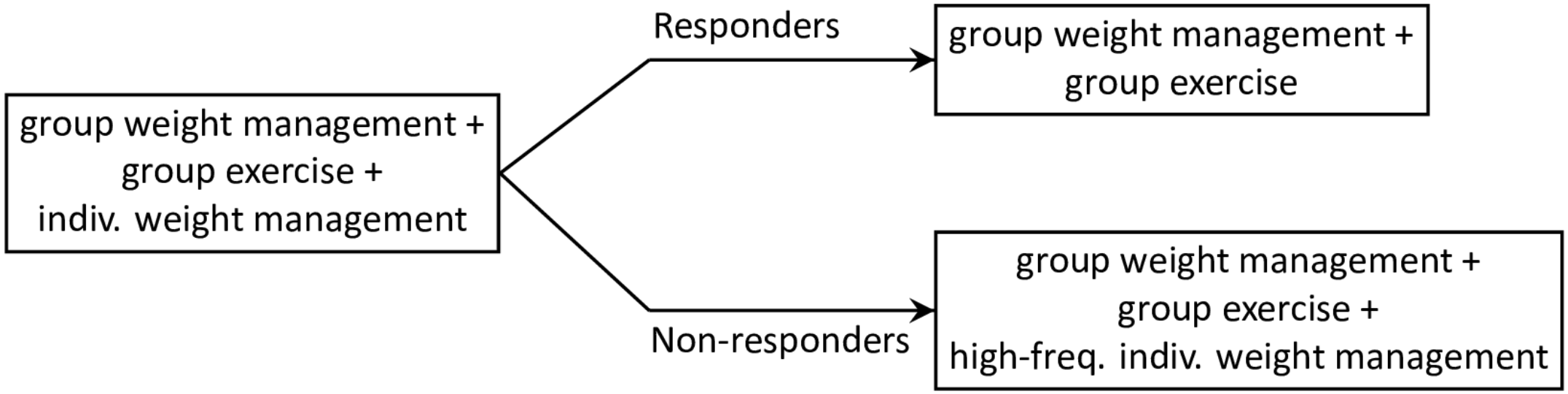
Dynamic Treatment Regimens (DTRs)

SMARTs are one experimental design that is useful for answering questions about the construction of high-quality **dynamic treatment regimens**

A DTR is a sequence of decision rules that maps ongoing participant information to a recommendation for subsequent treatment.

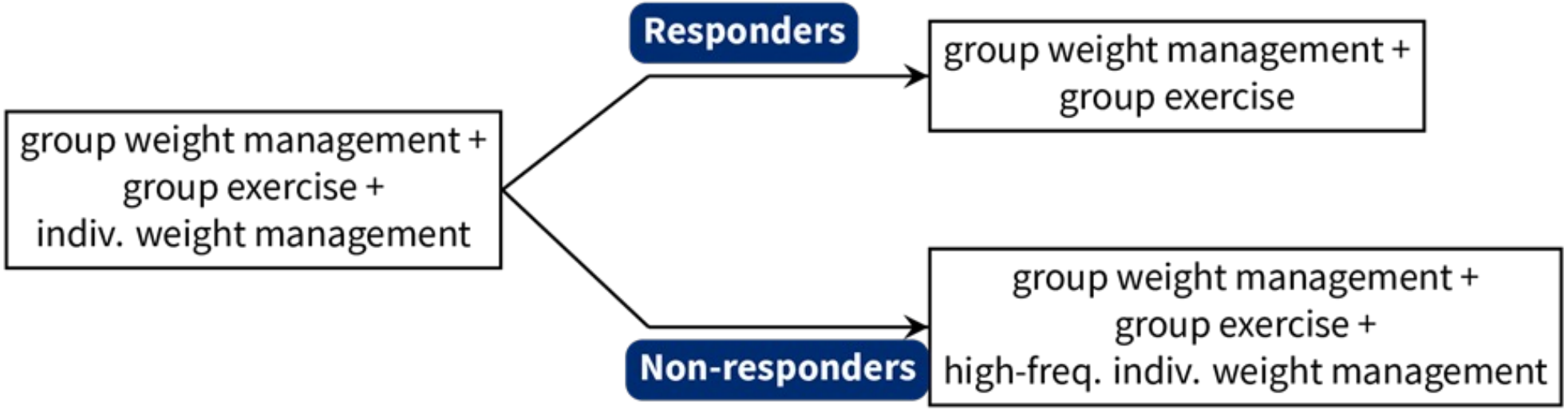
Also called adaptive interventions (AI), individualized treatment strategies (ITS), etc.

An example DTR



An individual is a *responder* if they have lost 5+ pounds
in the first 6 months, and a *non-responder* otherwise.

Tailoring Variables



- DTRs recommend treatments for each level of the tailoring variable.
- Tailoring variables should be well-defined and chosen based on scientific, ethical, or practical grounds.

Some notes on tailoring variables

Tailoring variables should be *pre-specified* and *well-defined*.

Tailoring variables are **part of the intervention**!

Should be based on *practical*, *ethical*, or *scientific* considerations.

- **Practical**: You might save more intense or costly intervention options for those who need it most (i.e., “non-responders”).
- **Ethical**: You might have an ethical obligation to modify treatment for a particular subset of individuals
- **Scientific**: You might have empirical evidence suggesting that “responders” need a different type of intervention than do “non-responders”

Why would we be interested in DTRs?

High heterogeneity in need for, or response to, a particular intervention

What works for one person may not work for another.

Need to:

- Detect early signs of intervention failure,
- Modify the intervention, and
- Work to prevent ultimate intervention failure.

Why would we be interested in DTRs?

Intervention is burdensome

Interventions can be burdensome when participant required to invest significant time or effort.

Burden leads to non-adherence, reducing the likelihood of a positive intervention effect

Need to identify:

- Signs of burden
- How to modify intervention intensity based on signs of burden

Why would we be interested in DTRs?

Intervention is Costly

Certain treatments can be very expensive

Resources are often limited

Expensive interventions can be difficult to scale

May need to:

- Try less expensive intervention first, saving more costly intervention for those who need it
- Try most costly intervention up front and reduce intervention over time

Scientific Questions about DTRs

There are often unanswered questions about how to sequence and adapt interventions! These are typically related to

- relative effectiveness of different intervention options
- how intervention options work with/against each other
- relative effectiveness of different adaptive interventions

For example:

- Which treatment option should the adaptive intervention begin with?
- How should we modify treatment for initial non-responders?
- How should we modify treatment for initial responders?
- How do we define response/non-response?
- How should we time decision points?

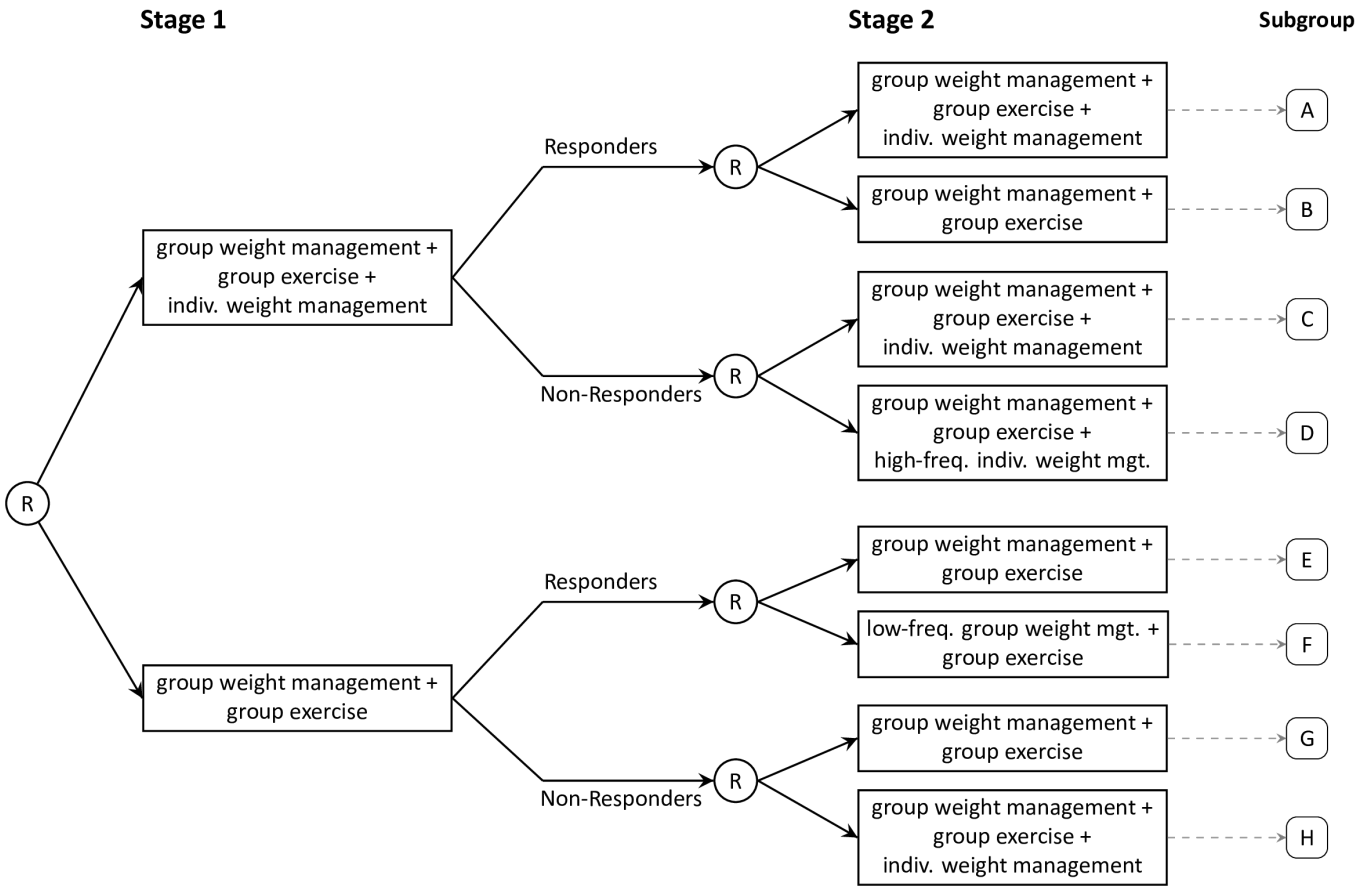
Sequential, Multiple-Assignment Randomized Trials

One trial design useful for constructing high-quality DTRs

- Allows researchers to answer questions about **multiple stages of the DTR**

The key feature is that **some or all participants are randomized more than once**

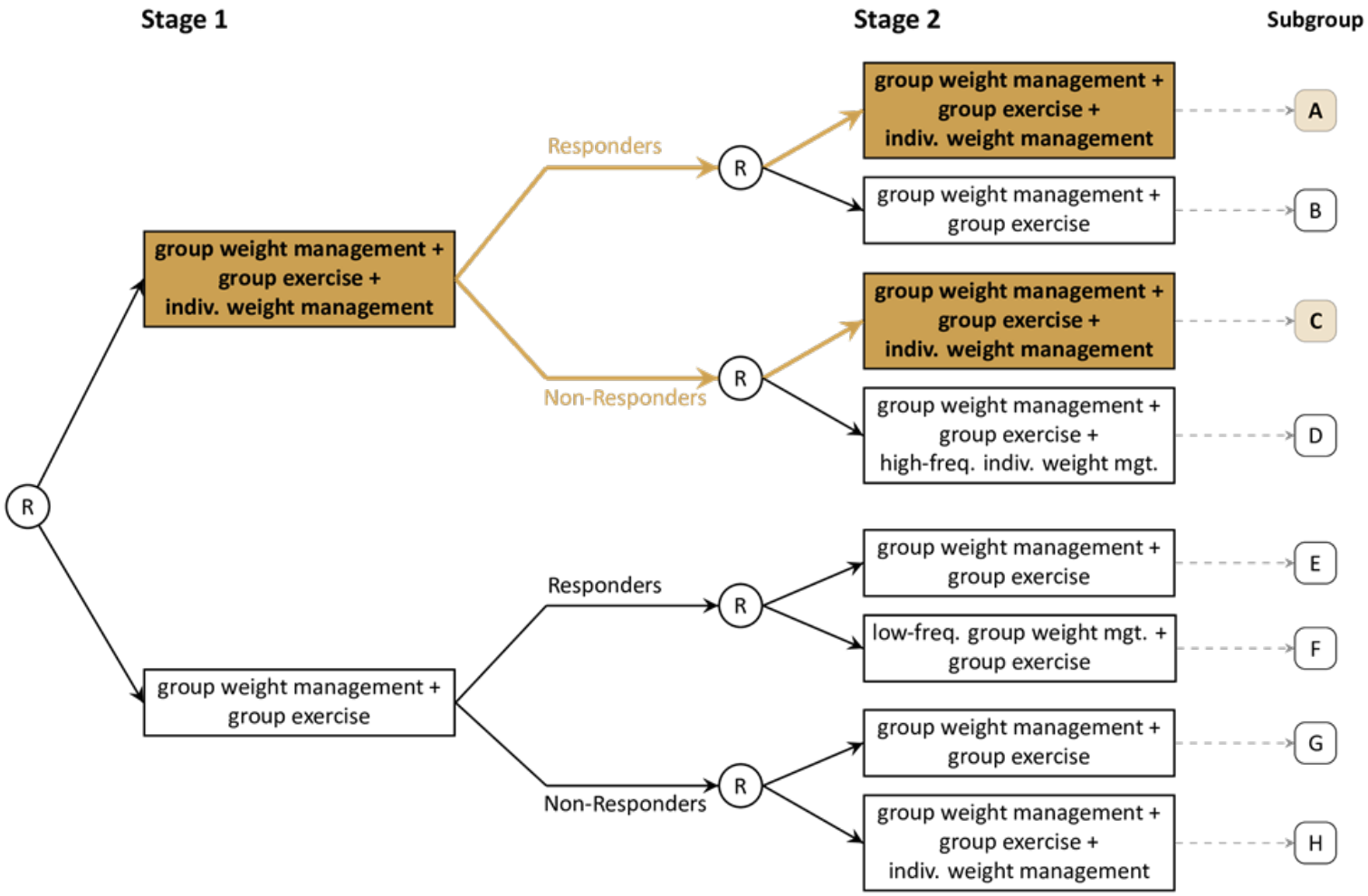
Hypothetical example: weight loss program for individuals with serious mental illness



This is a hypothetical example

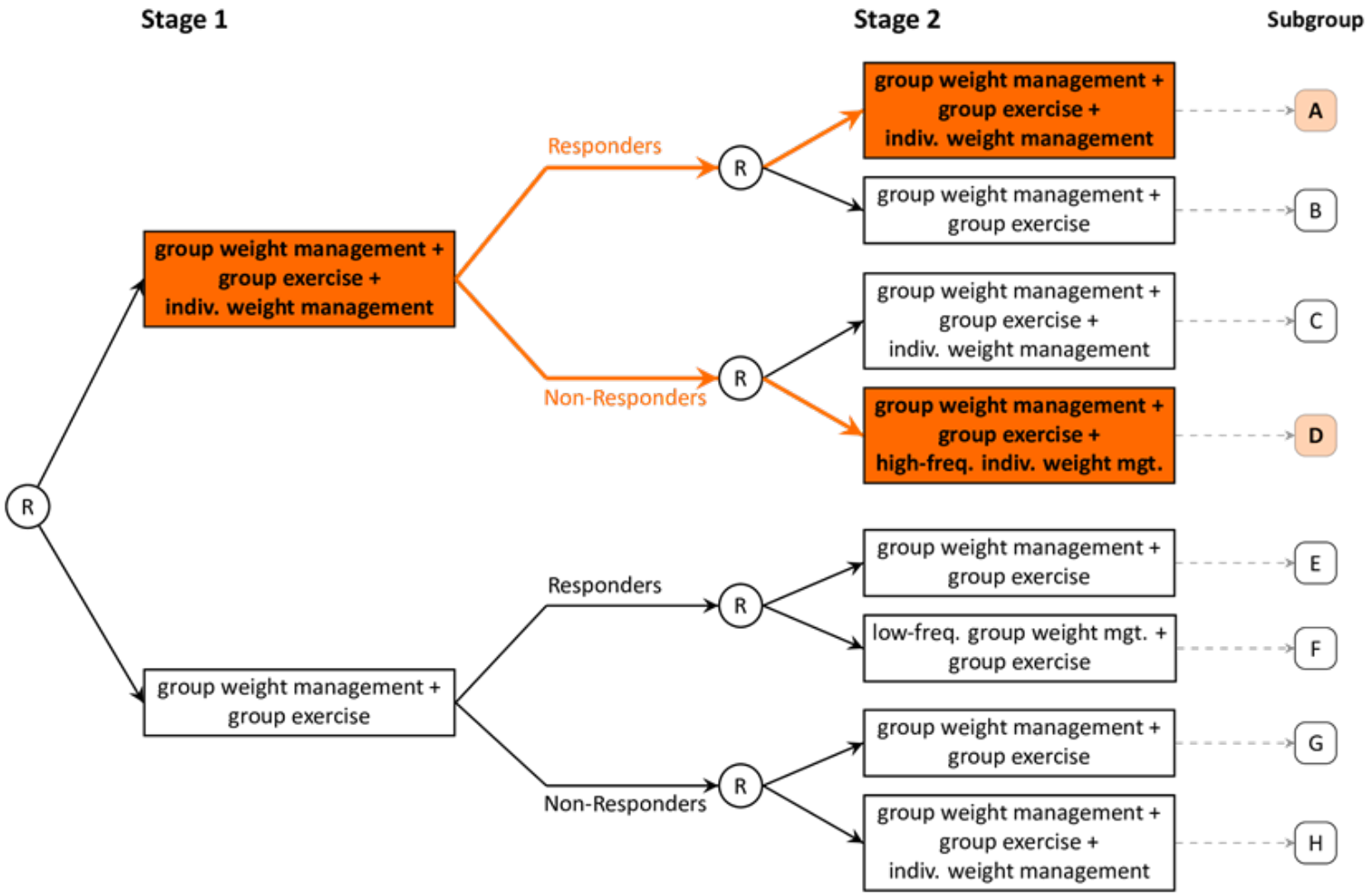
This is a hypothetical example

8 “embedded” DTRs



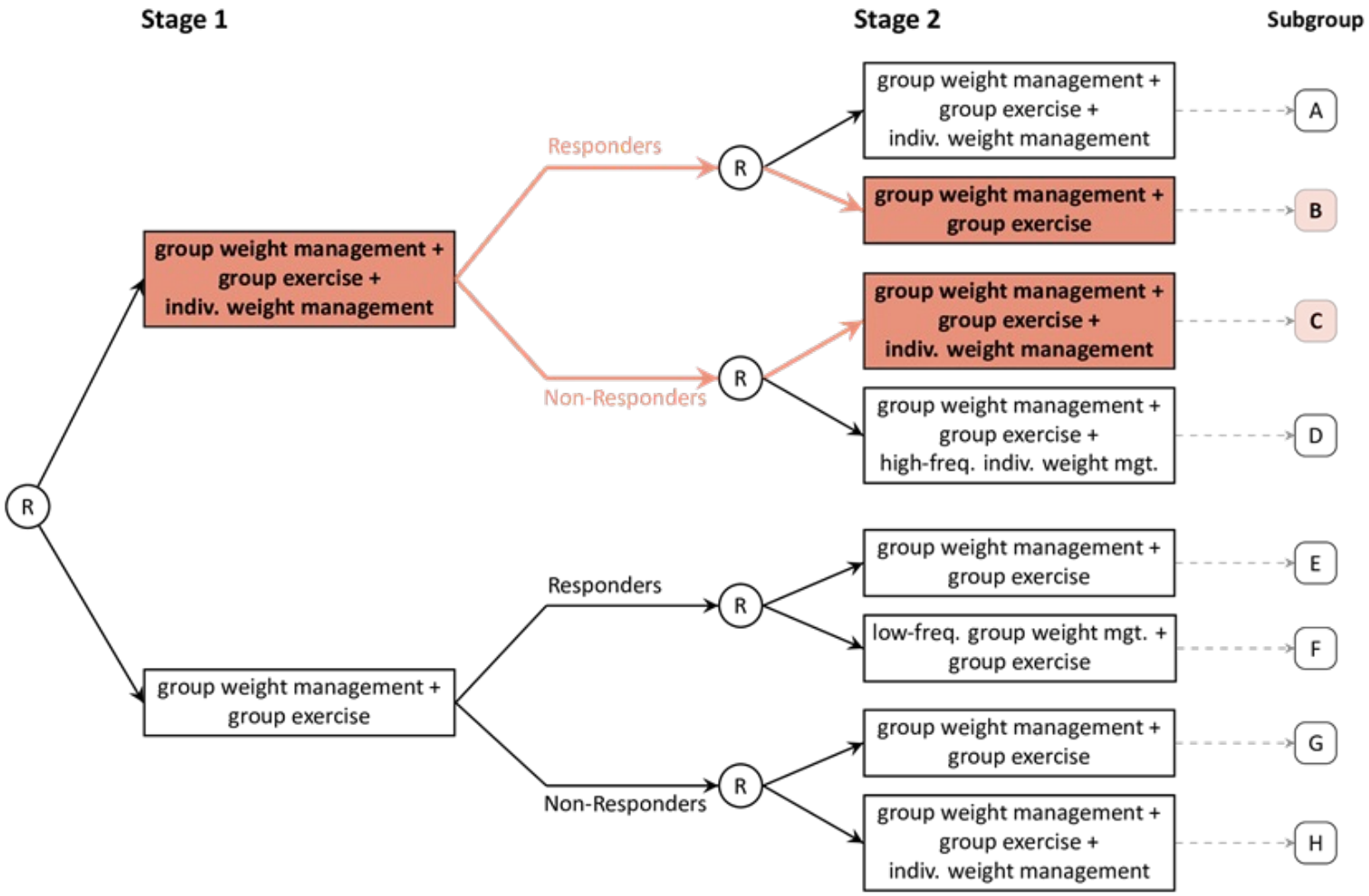
This is a hypothetical example

8 “embedded” DTRs



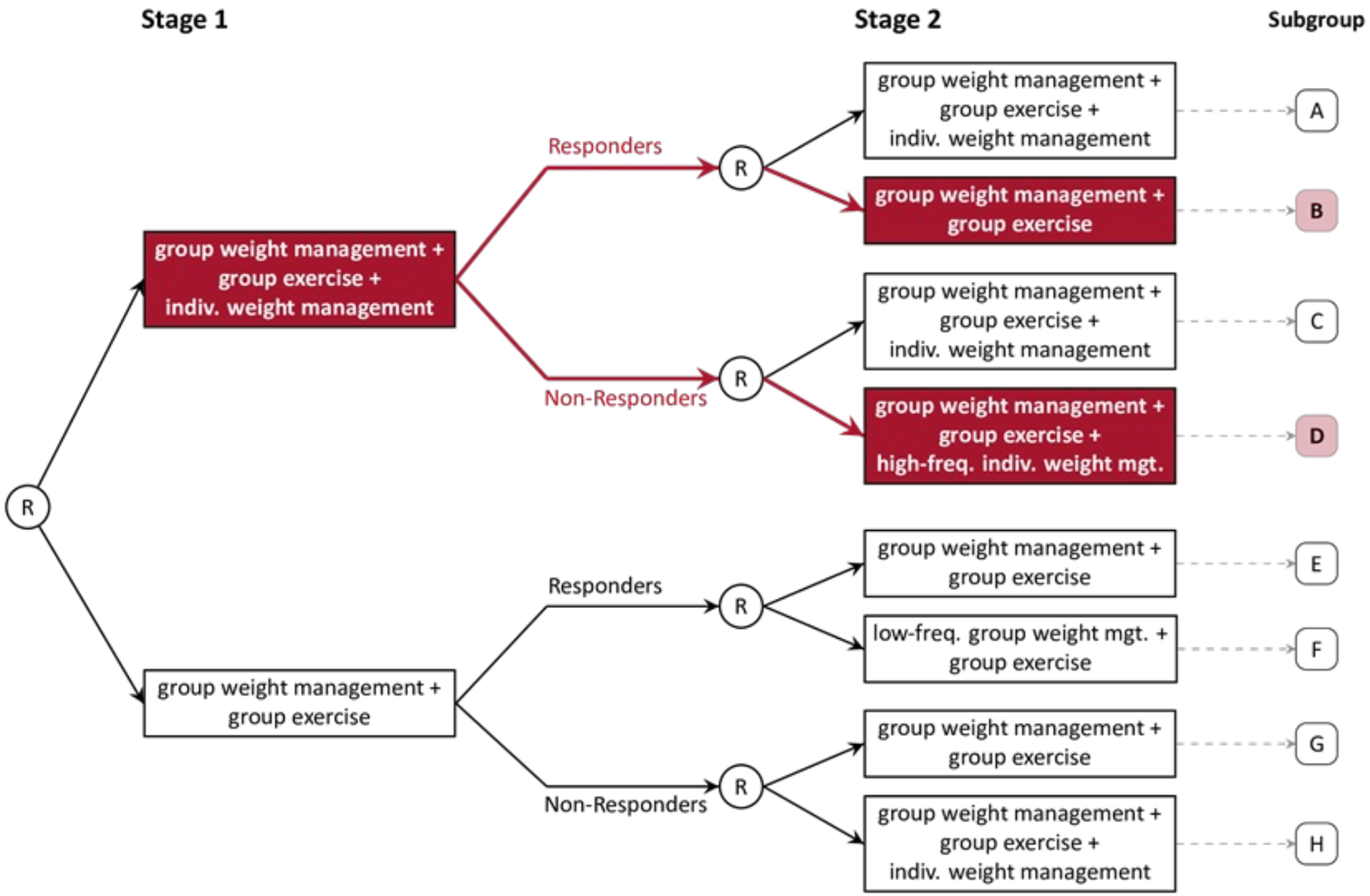
This is a hypothetical example

8 “embedded” DTRs



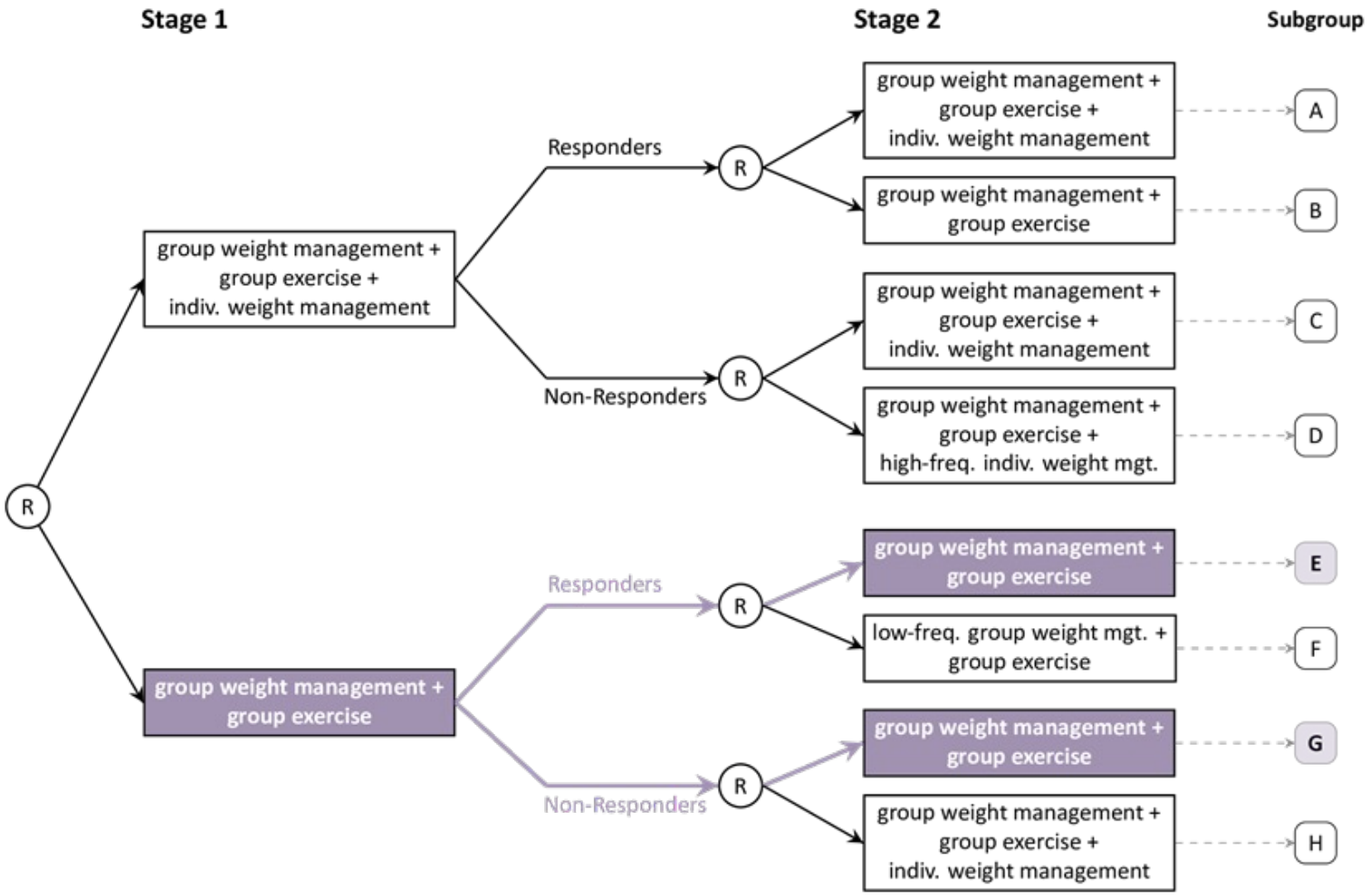
This is a hypothetical example

8 “embedded” DTRs



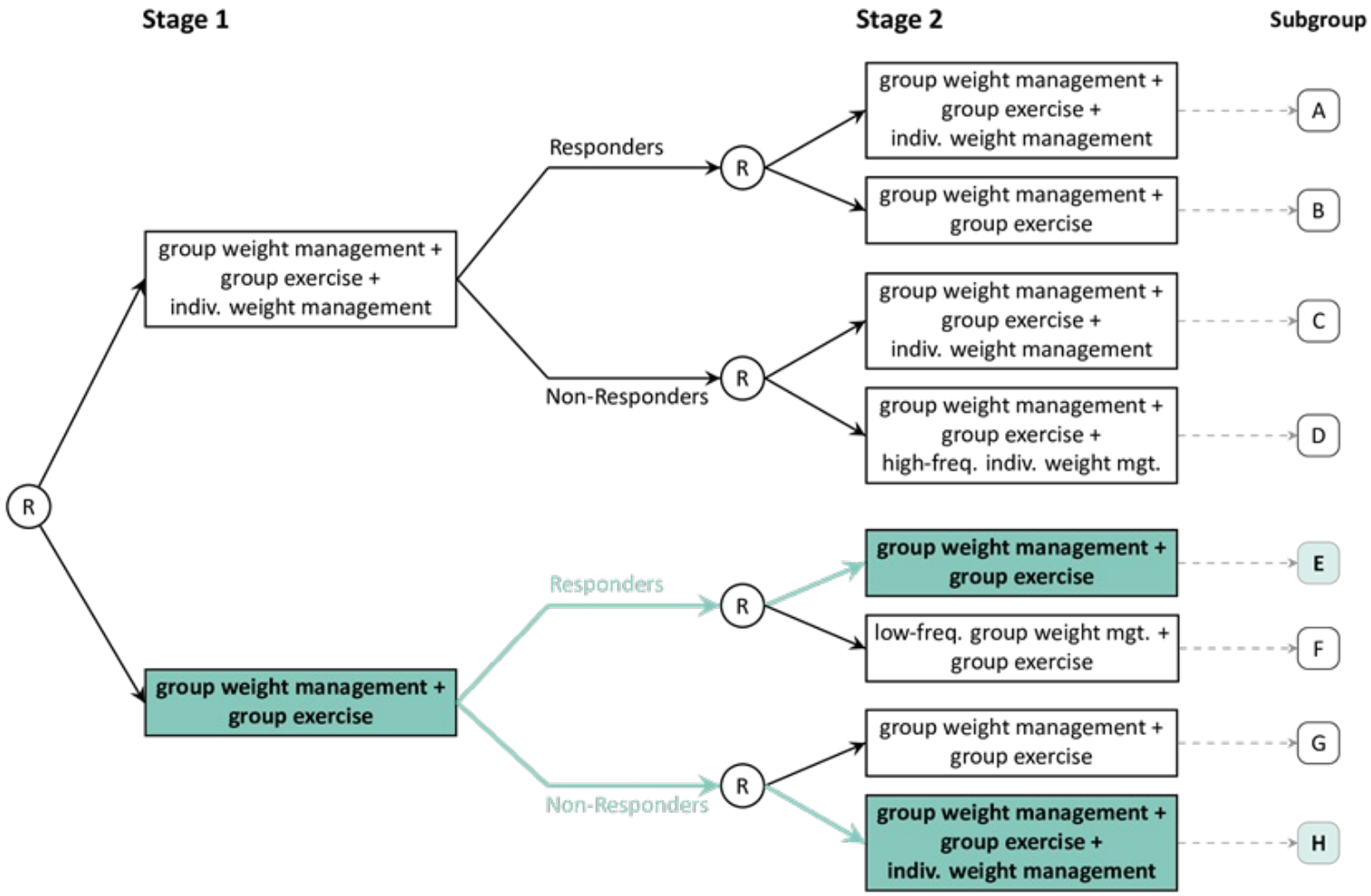
This is a hypothetical example

8 “embedded” DTRs



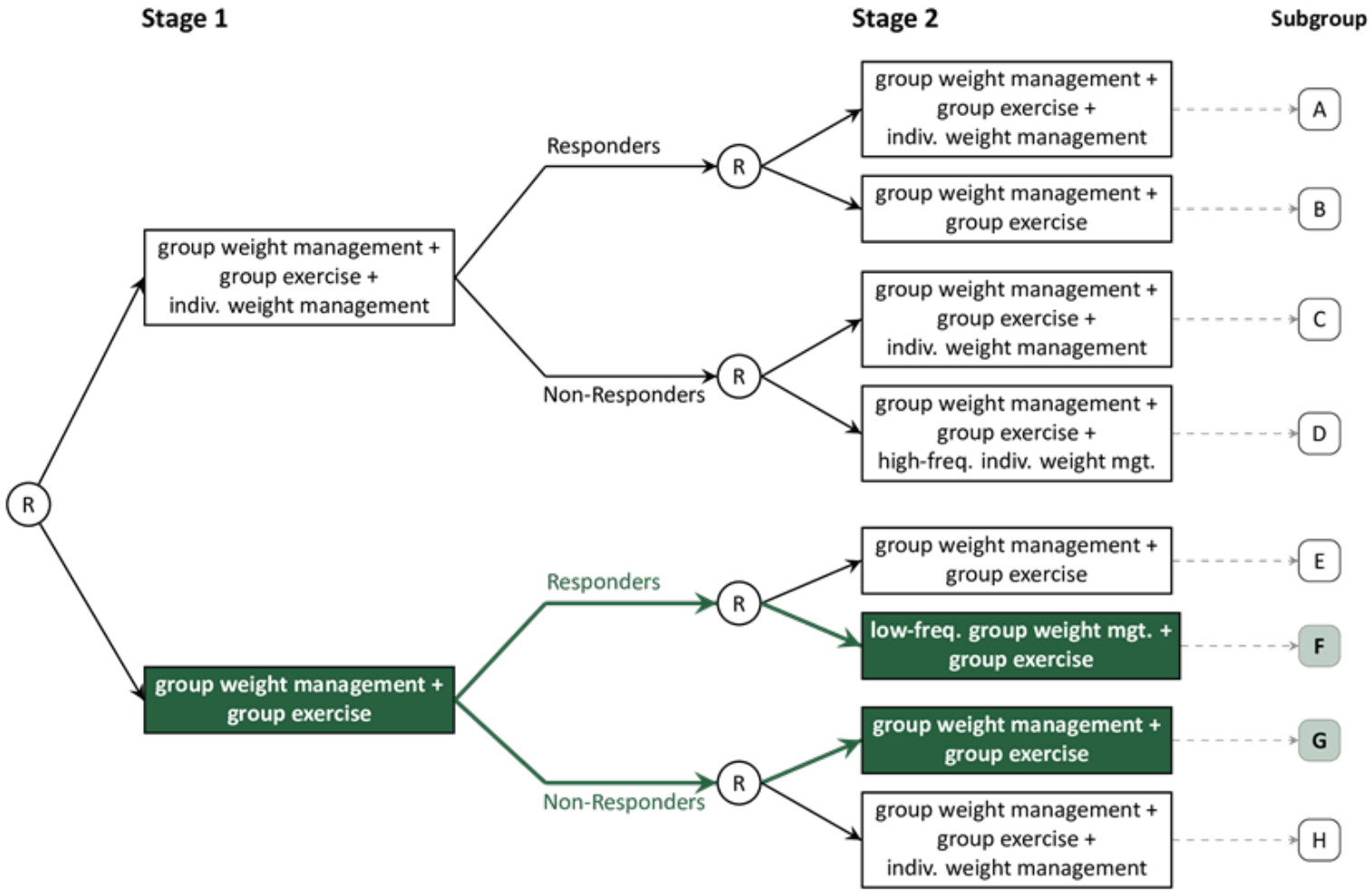
This is a hypothetical example

8 “embedded” DTRs



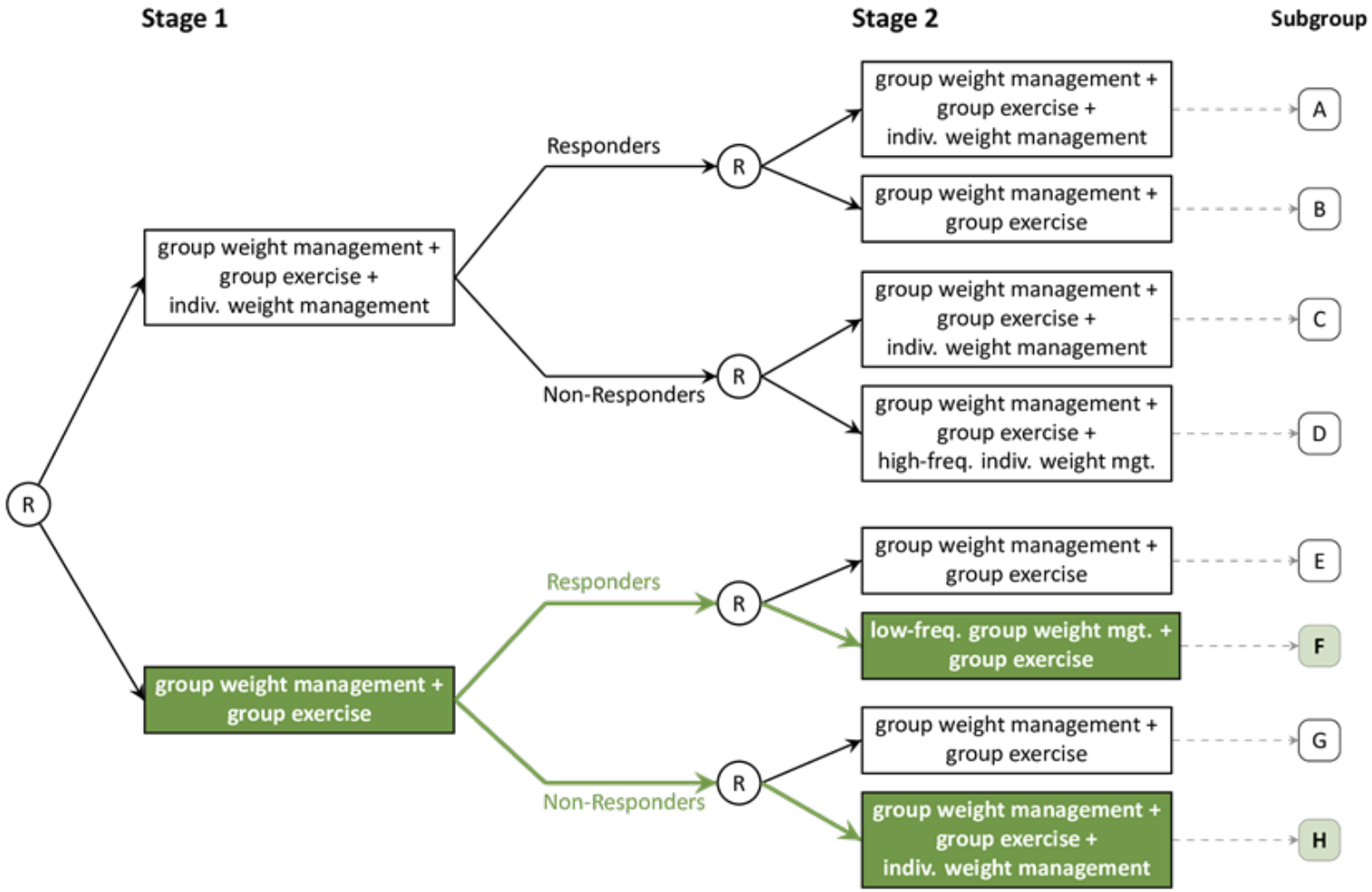
This is a hypothetical example

8 “embedded” DTRs



This is a hypothetical example

8 “embedded” DTRs



Do you need a SMART?

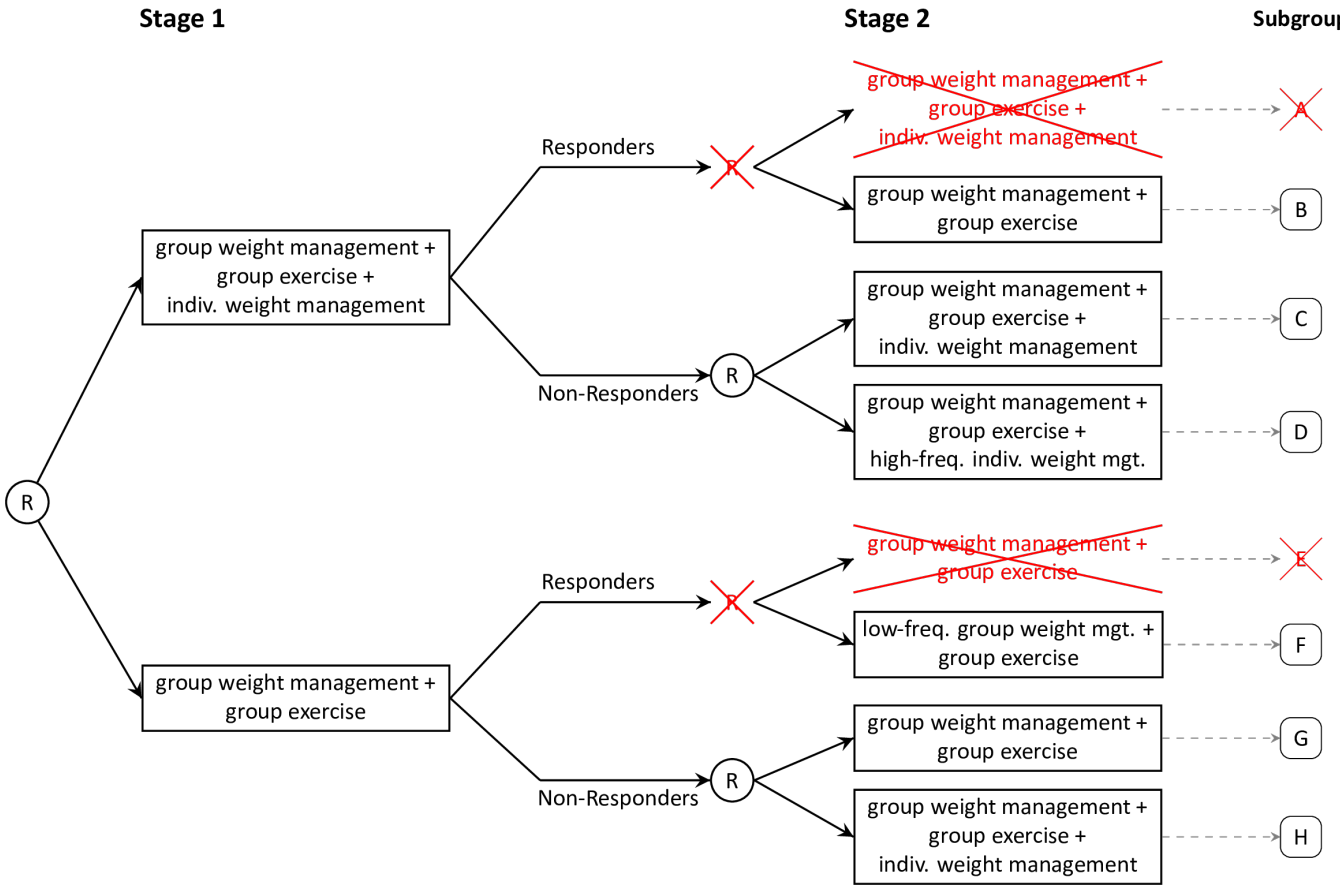
SMARTs are designed to answer questions about the development of high-quality DTRs.

You might consider a SMART if…

- you want to develop a DTR,
- there are open questions preventing the construction of an effective DTR, *and*
- there are open questions at **multiple decision points** within a DTR

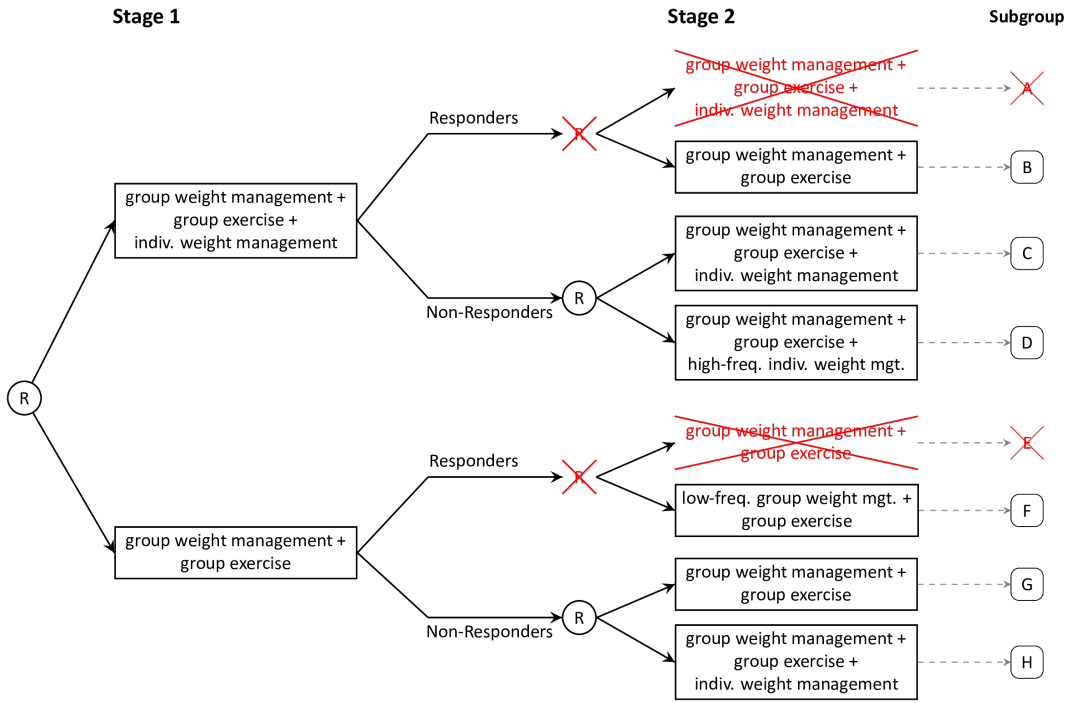
If any of the above are *not* true, you do not need a SMART!

Do you need a SMART if you know what to do for responders?



This is a hypothetical example

Do you need a SMART if you know what to do for responders?

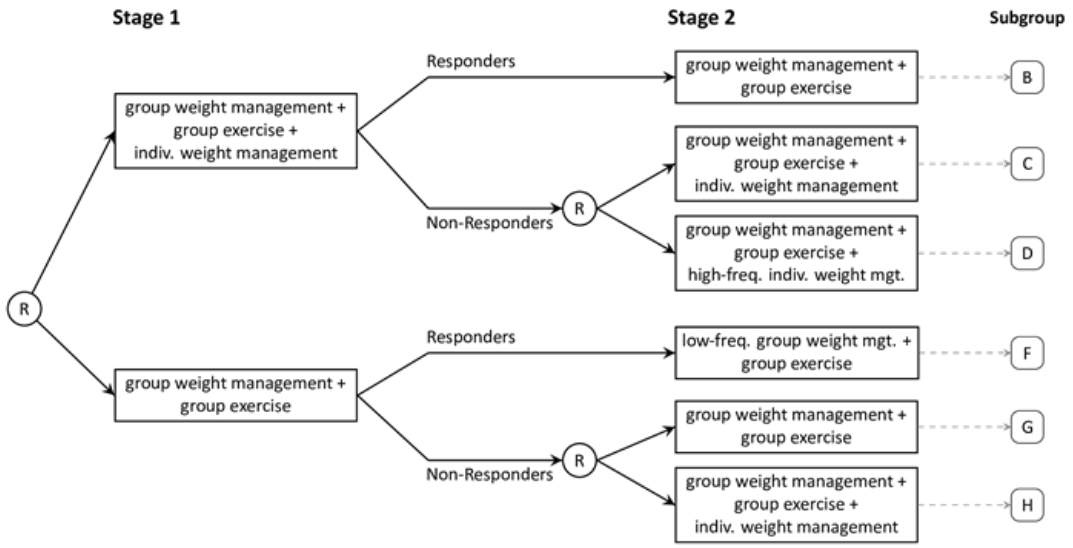


Still questions about *multiple stages* of a DTR:

- What should we do first?
- What should we do for non-responders?

This is a hypothetical example

Do you need a SMART if you know what to do for responders?



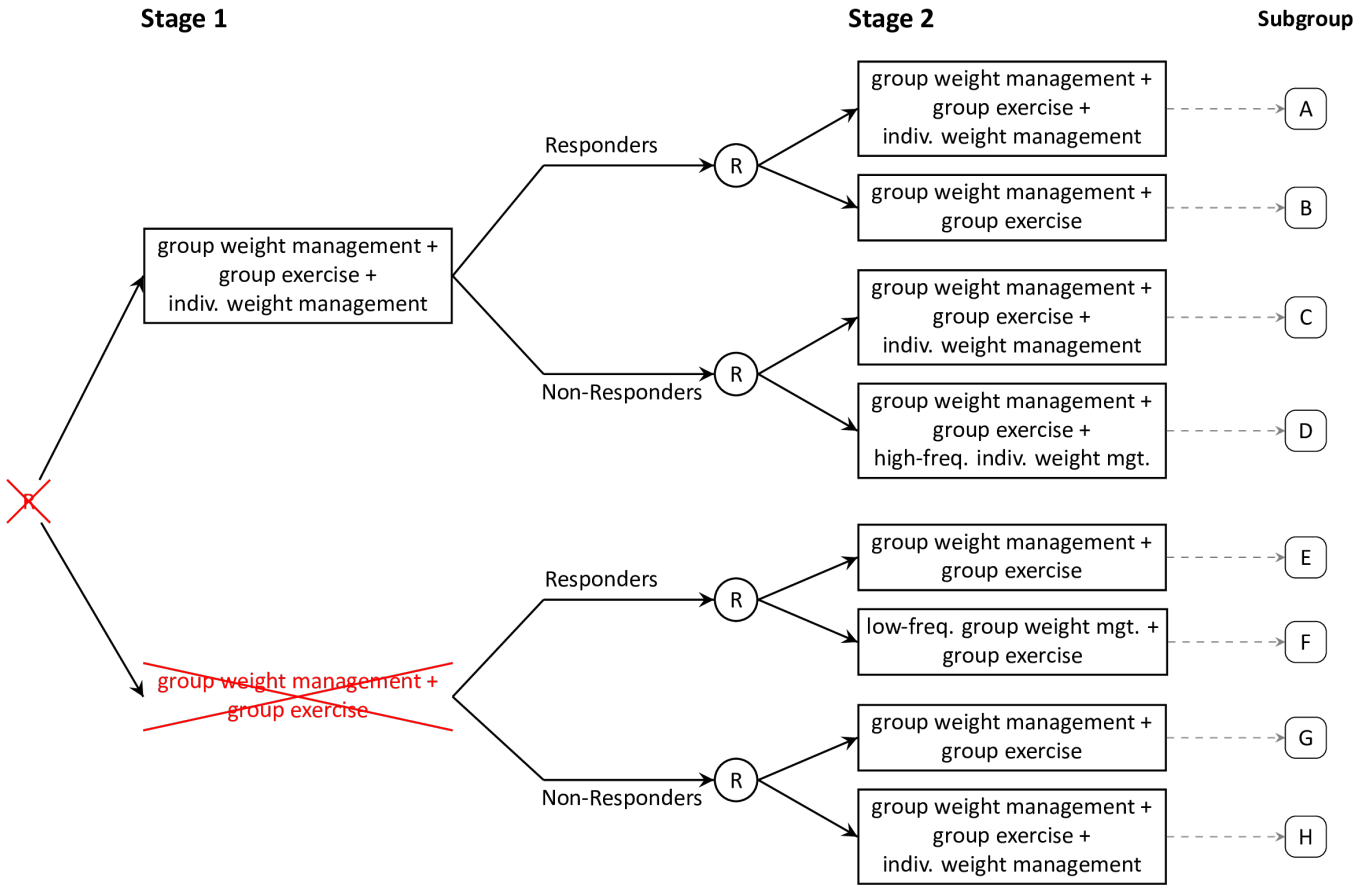
Still questions about *multiple stages* of a DTR:

- What should we do first?
- What should we do for non-responders?

A SMART is appropriate here

- **Some** participants are randomized more than once

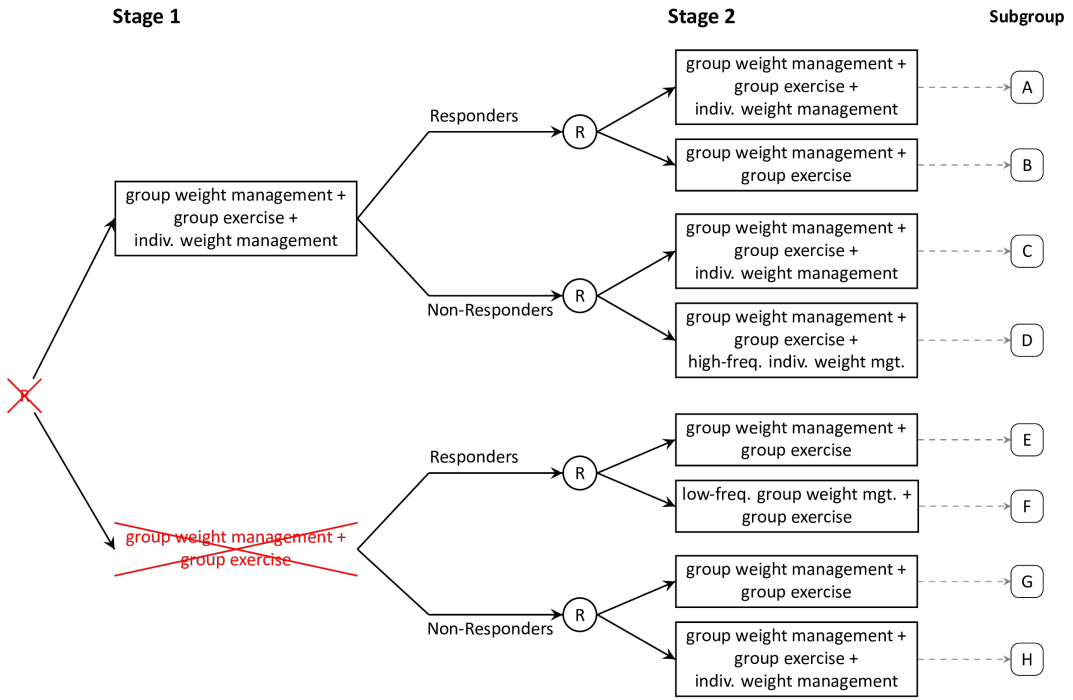
Do you need a SMART if you know what to do initially?



This is a hypothetical example

This is a hypothetical example

Do you need a SMART if you know what to do initially?

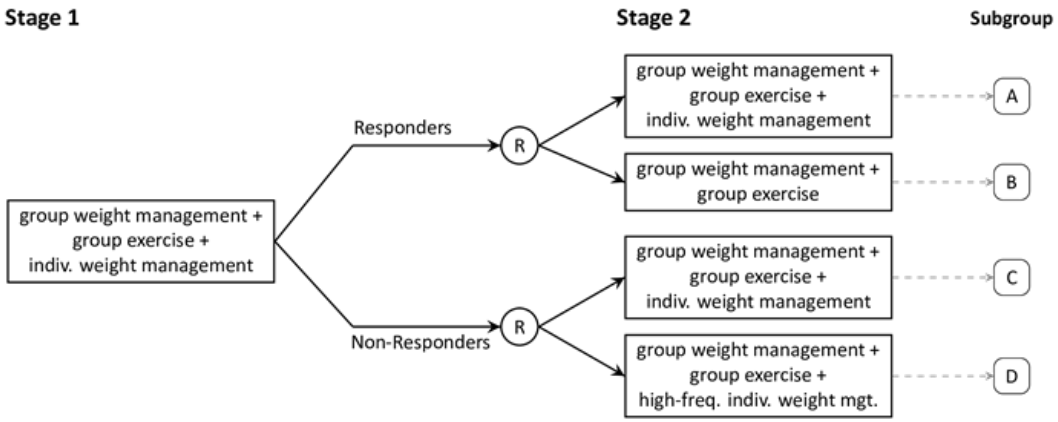


There are not questions about **multiple stages** of an adaptive intervention.

If there is no scientific question about how to initiate an adaptive intervention, we do not need the initial randomization.

This is a hypothetical example

Do you need a SMART if you know what to do initially?



We could instead design a trial with a run-in period on the initial intervention.

Randomization is still tailored, but there is only one.

- This is not a SMART!

Do you need a SMART?

Not all research on DTRs requires a SMART

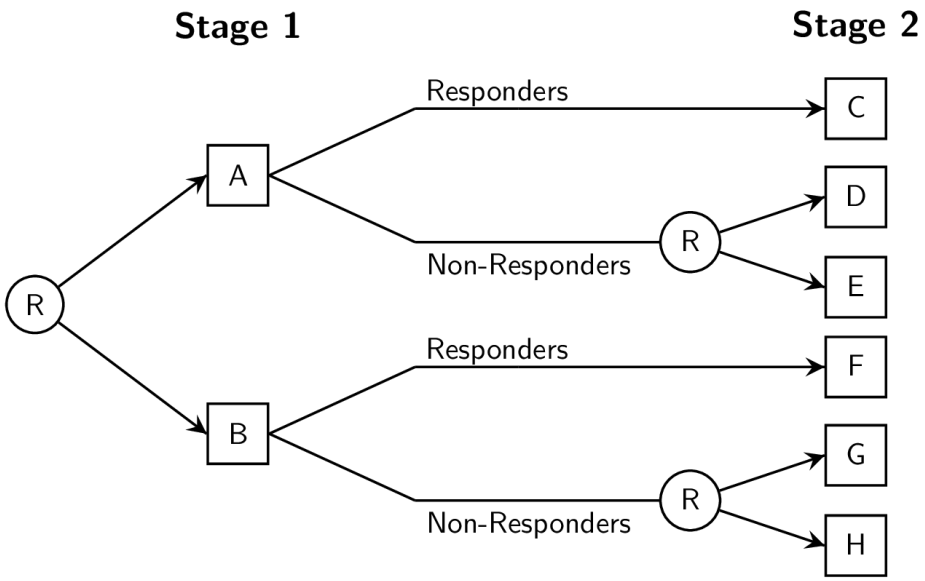
It may be appropriate to consider a “singly-randomized” alternative to a SMART

- See Almirall et al. (2018) for examples.

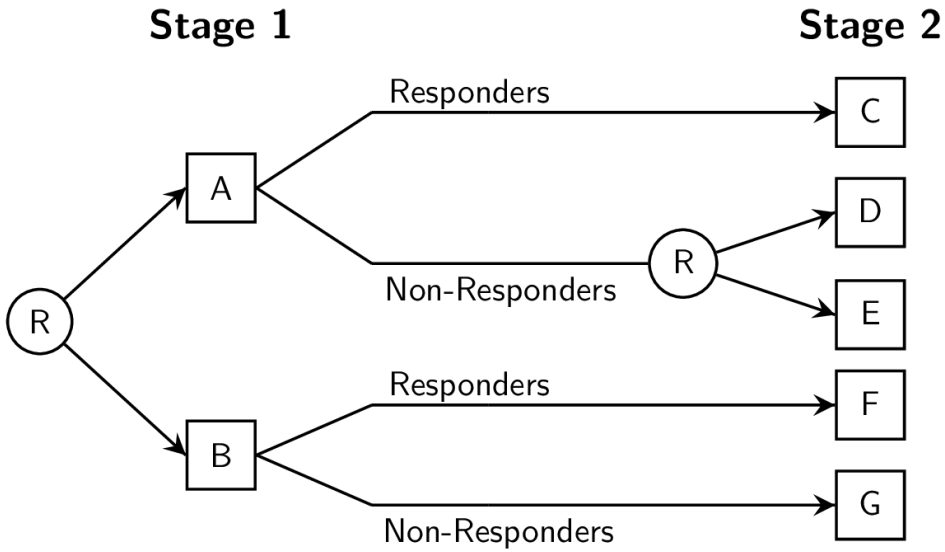
Almirall, D., et al. 2018. “Experimental Designs for Research on Adaptive Interventions: Singly and Sequentially Randomized Trials.” In Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: Advanced Topics, edited by Linda M. Collins and Kari C. Kugler, 89–120. Statistics for Social and Behavioral Sciences. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-91776-4_4.

Other SMART Designs

Only (non-)responders re-randomized

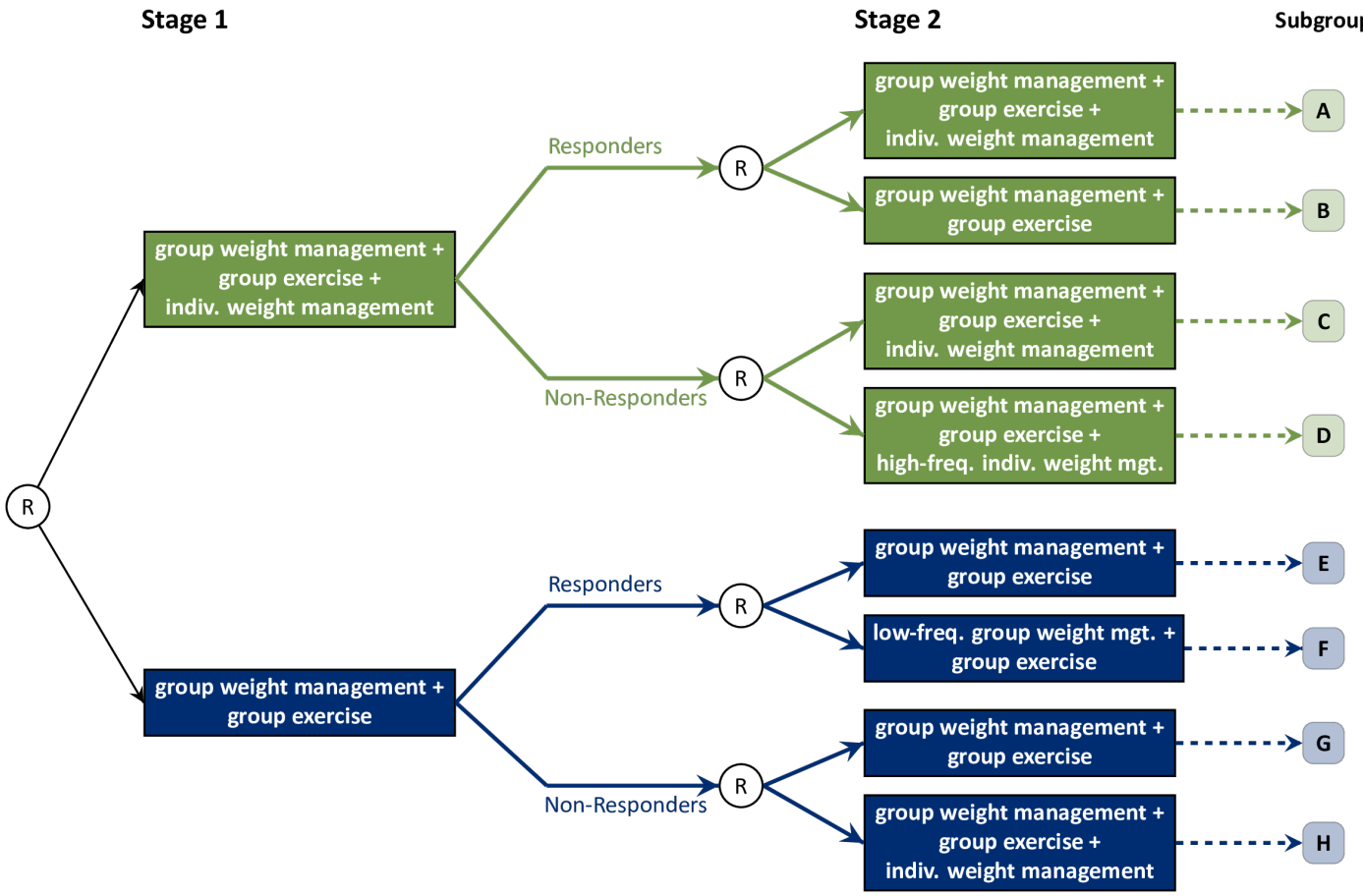


Only some (non)-responders re-randomized



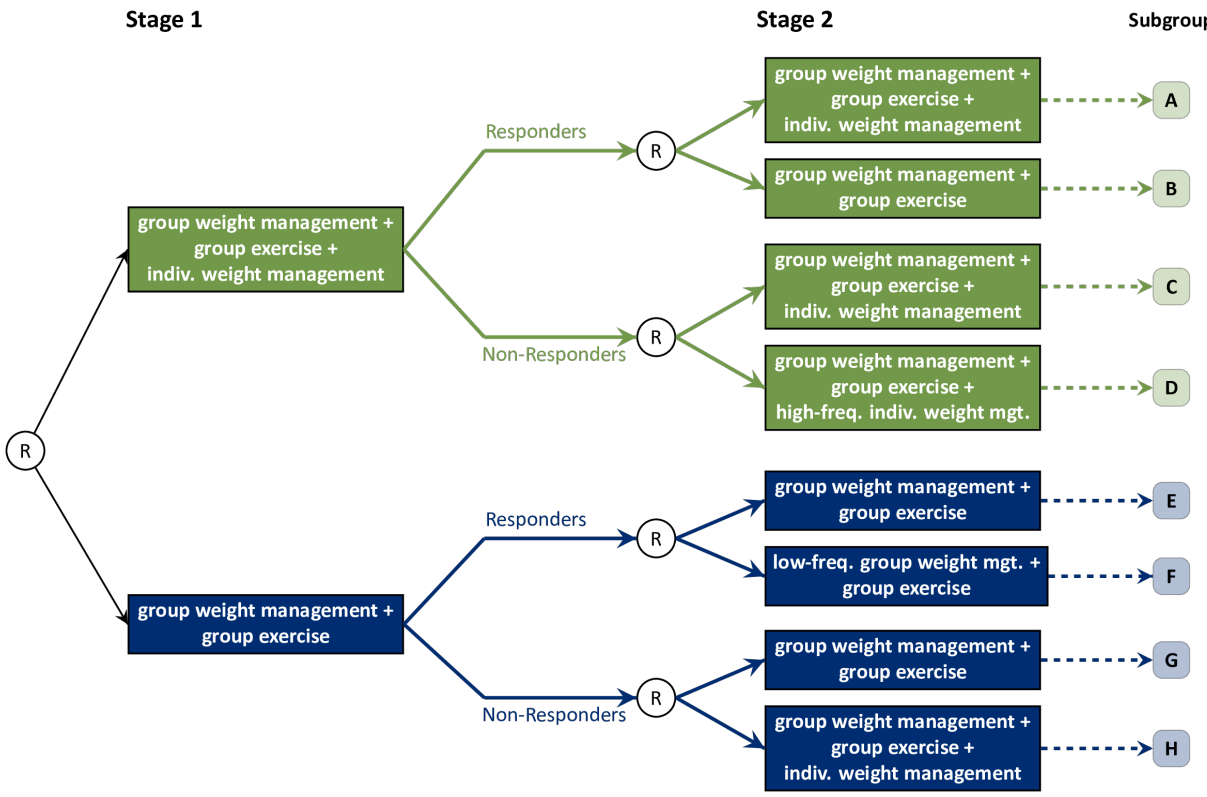
Common Primary Aims for SMARTs

Compare initial intervention options in the context of a DTR



Common Primary Aims for SMARTs

Compare initial intervention options in the context of a DTR



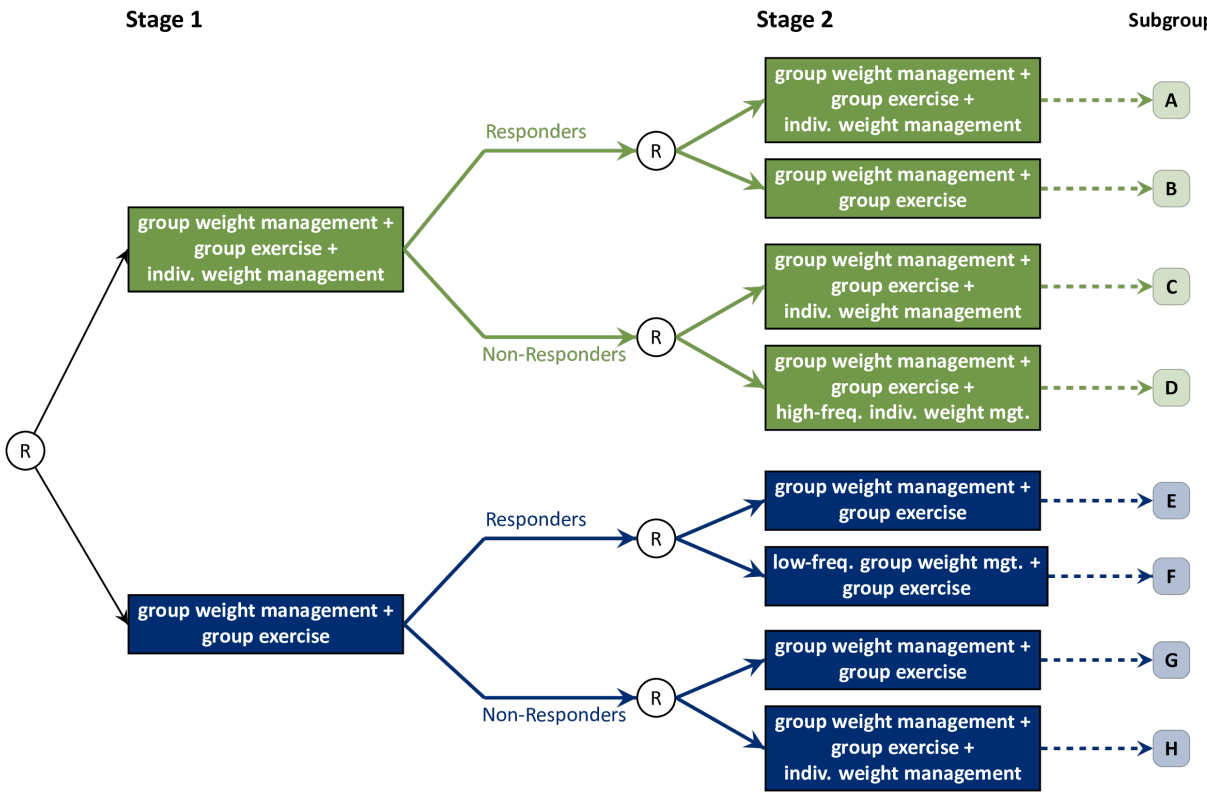
Hypothetical hypothesis:

> "Individuals who receive an adaptive weight-loss intervention which initially includes individual weight management sessions will lose more weight at 18 months, on average, than individuals who receive an adaptive weight-loss intervention that involves only group sessions."

Notice that the hypothesis is *in the context of adaptive interventions:* it "averages over" future treatment.

Common Primary Aims for SMARTs

Compare initial intervention options in the context of a DTR



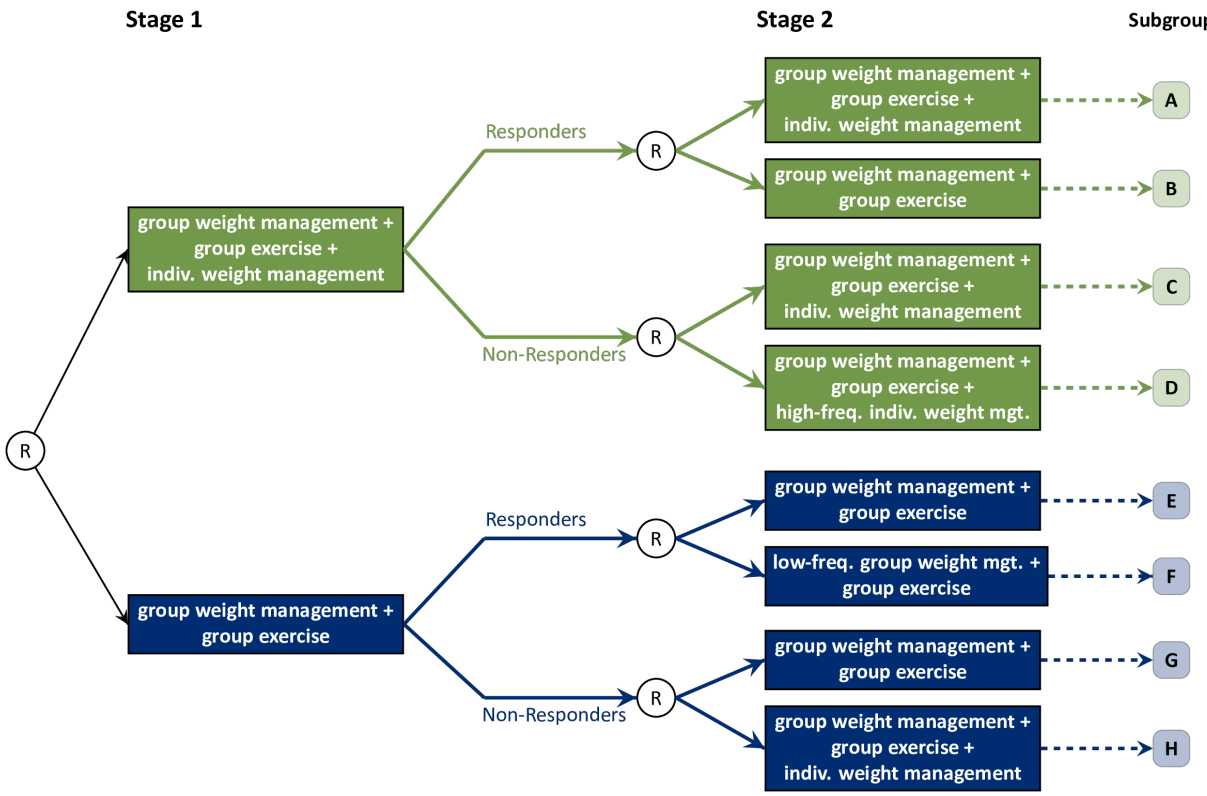
Analysis is a comparison of subgroups A, B, C, D vs. subgroups E, F, G, H.

- A two-group comparison!
- Can use standard methods (t-test, linear regression, etc.)

Sample size requirements are the same as for a two-arm trial.

Common Primary Aims for SMARTs

Compare initial intervention options in the context of a DTR



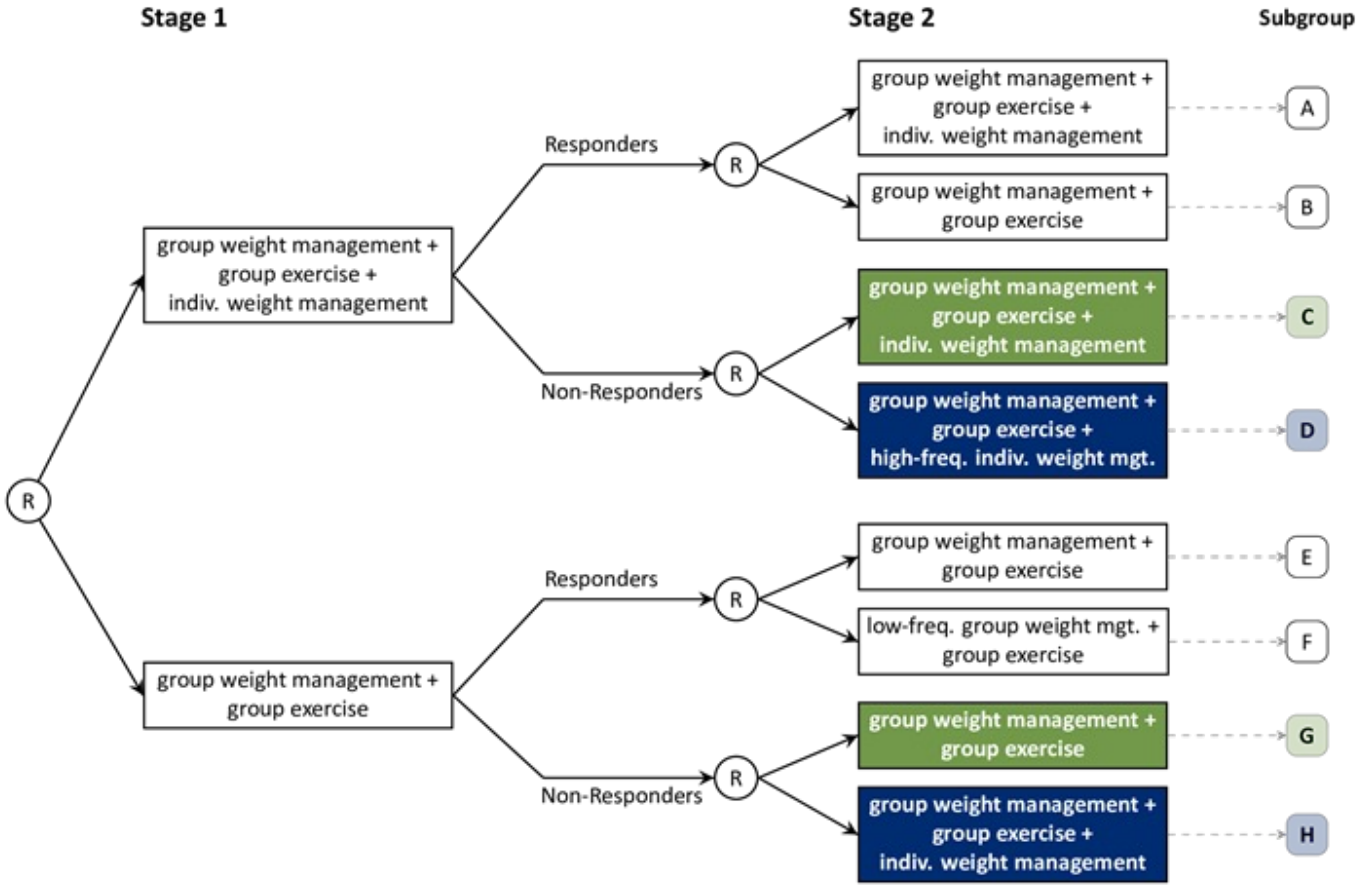
Sample size requirements are the same as for a two-arm trial.

$$n \geq \frac{4\left(z_{1-\frac{\alpha}{2}} + z_{1-\gamma}\right)^2}{\delta^2}$$

	80% power	90% power
$\delta = .3$	$n = 351$	$n = 469$
$\delta = .5$	$n = 128$	$n = 171$
$\delta = .8$	$n = 52$	$n = 68$

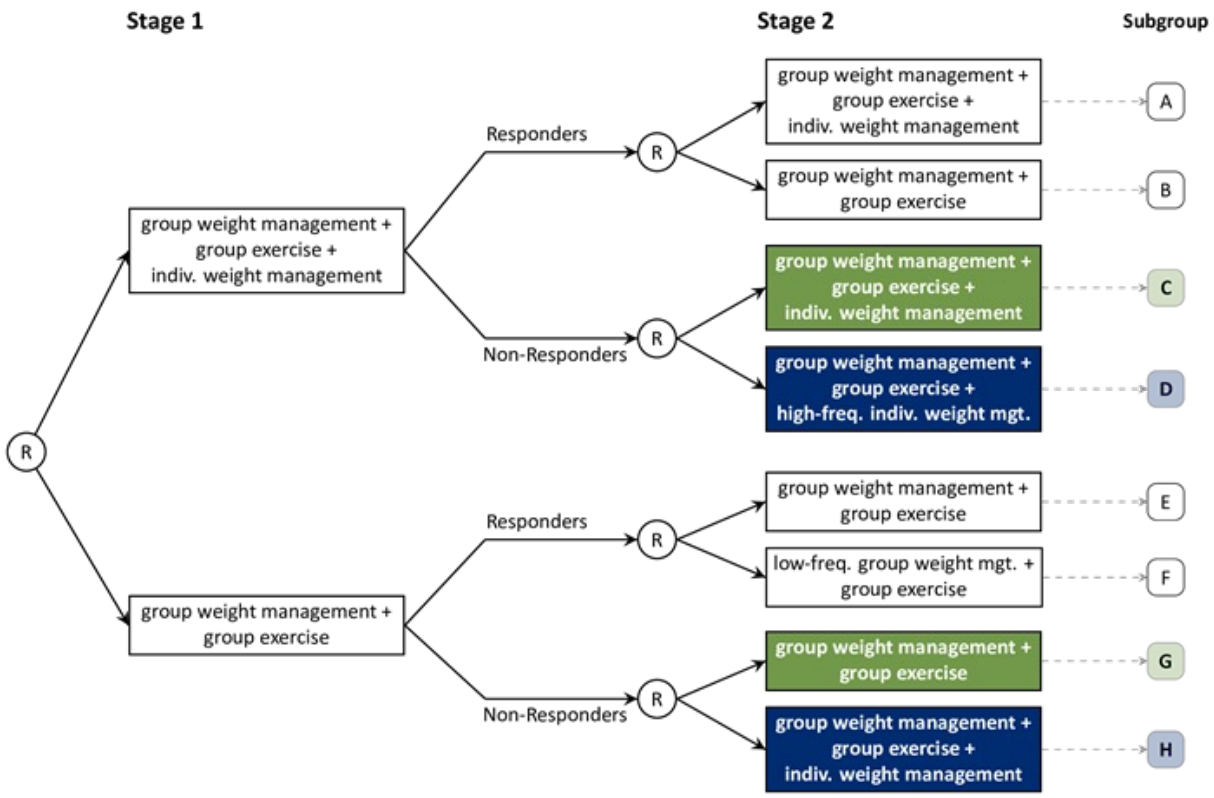
Common Primary Aims for SMARTs

Compare second-stage intervention options among (non-)responders



Common Primary Aims for SMARTs

Compare second-stage intervention options among (non-)responders



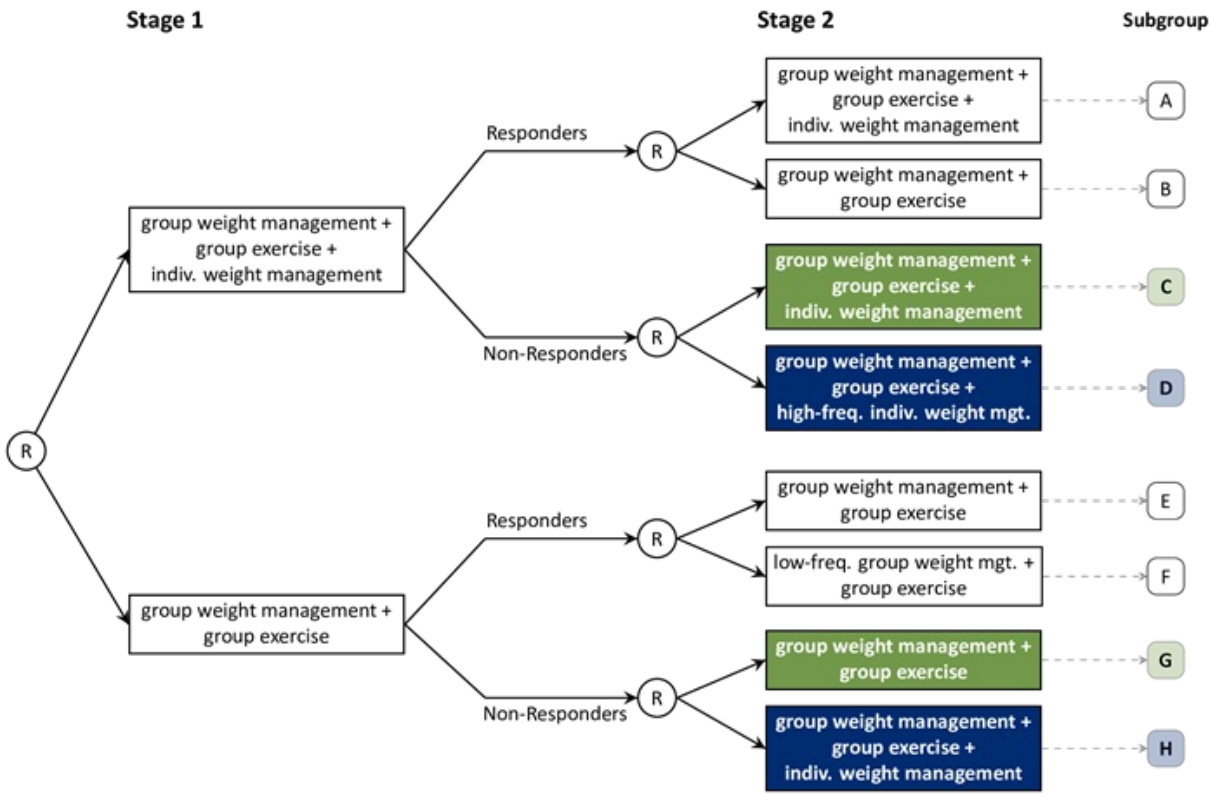
Hypothetical hypothesis:

> "Individuals who do not lose ≥ 5 lbs in the first 6 months of a weight-loss intervention will lose more weight at 18 months, on average, if their initial intervention is stepped up, compared to if they continued on the existing intervention."

Notice that the hypothesis is *in the context of DTRs:* it "averages over" past treatment.

Common Primary Aims for SMARTs

Compare second-stage intervention options among (non-)responders



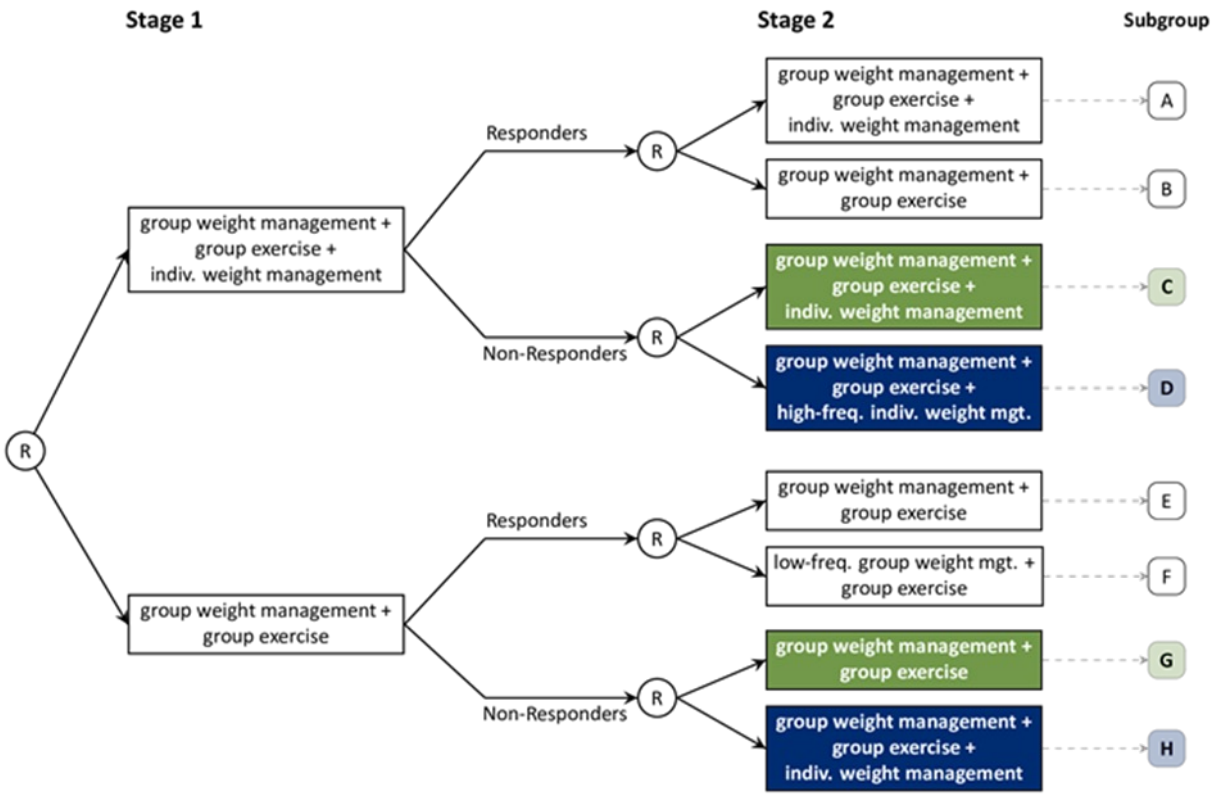
Analysis is a comparison of subgroups C & G vs. subgroups D & H.

- A two-group comparison among non-responders!
- Can use standard methods

Sample size requirements are the same as for a two-arm randomized trial, upweighted by (non-)response rate.

Common Primary Aims for SMARTs

Compare second-stage intervention options among (non-)responders



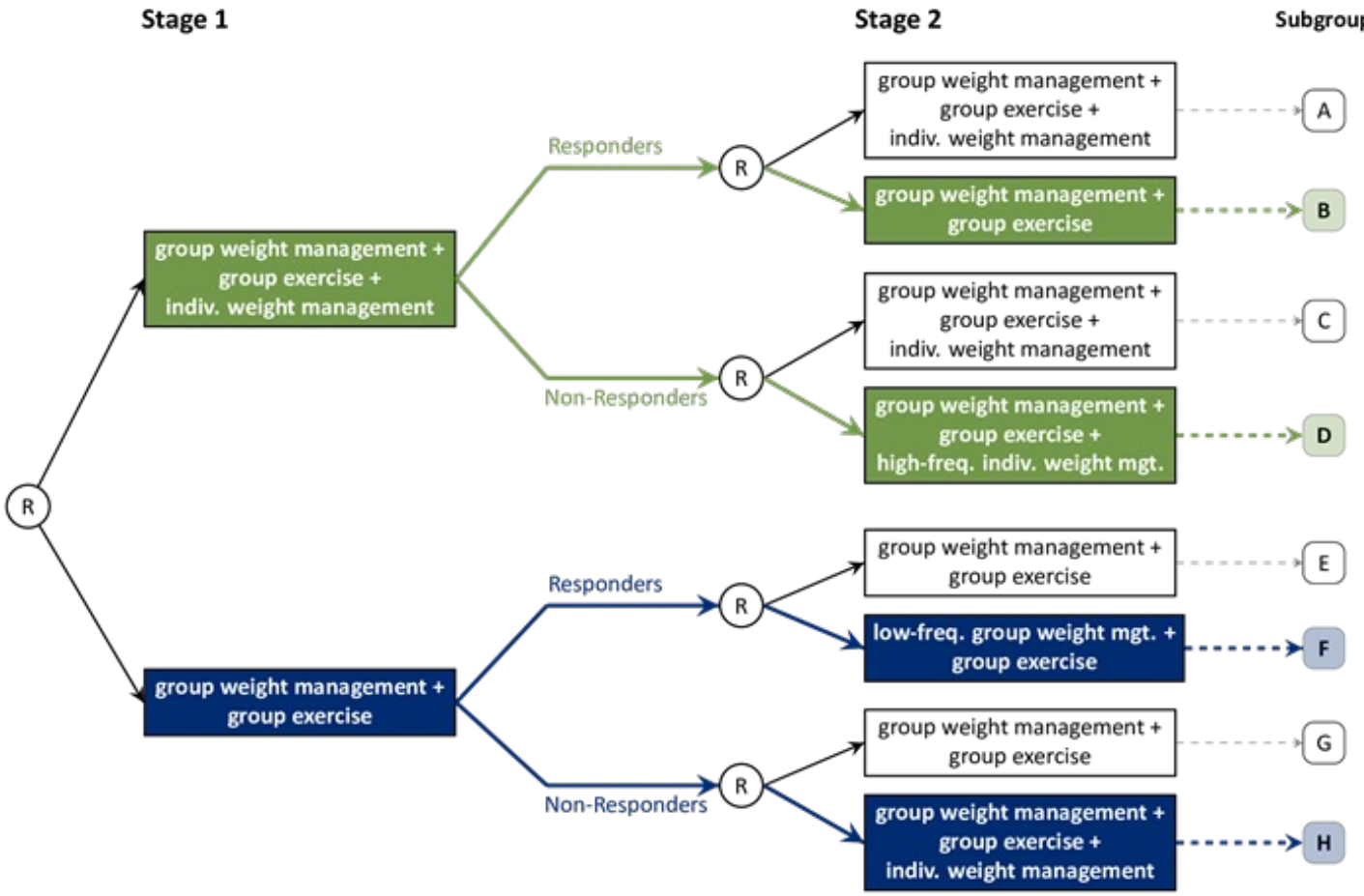
Sample size requirements are the same as for a two-arm trial.

$$n \geq \frac{4\left(z_{1-\frac{\alpha}{2}} + z_{1-\gamma}\right)^2}{\delta^2} \cdot \frac{1}{1 - P(R = 1)}$$

	80% power	90% power
$\delta = .3$	$n = 351/(1-r)$	$n = 469/(1-r)$
$\delta = .5$	$n = 128/(1-r)$	$n = 171/(1-r)$
$\delta = .8$	$n = 52/(1-r)$	$n = 68/(1-r)$

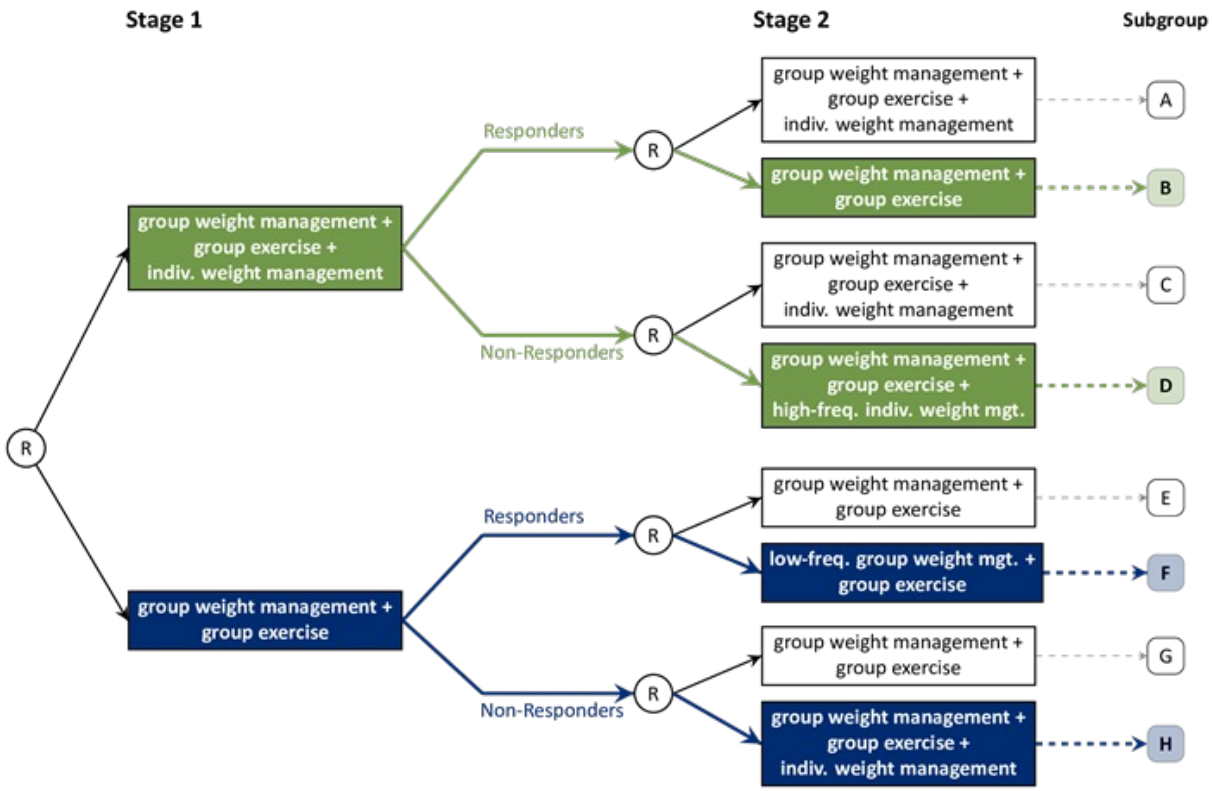
Common Primary Aims for SMARTs

Compare embedded DTRs



Common Primary Aims for SMARTs

Compare embedded DTRs

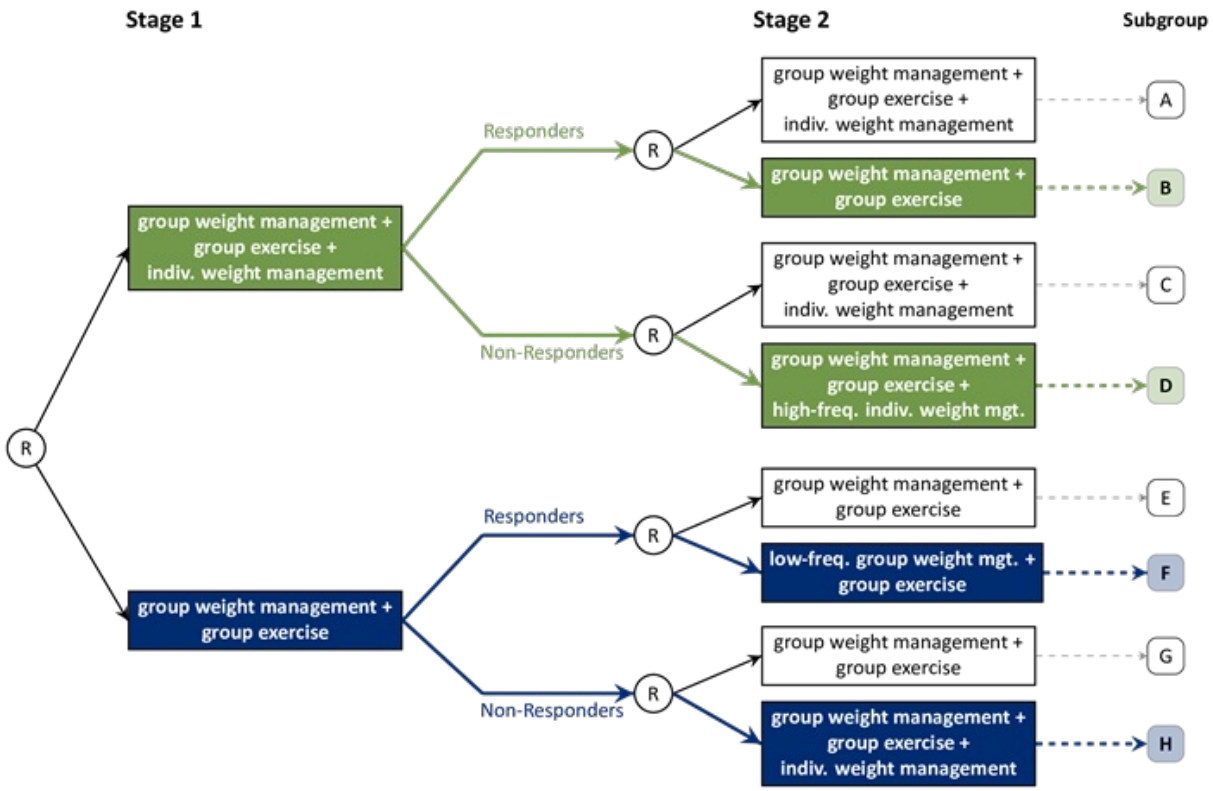


Hypothetical hypothesis:

"Individuals who receive treatment according to the green DTR will lose more weight after 18 months, on average, compared to those treated according to the blue DTR."

Common Primary Aims for SMARTs

Compare embedded DTRs

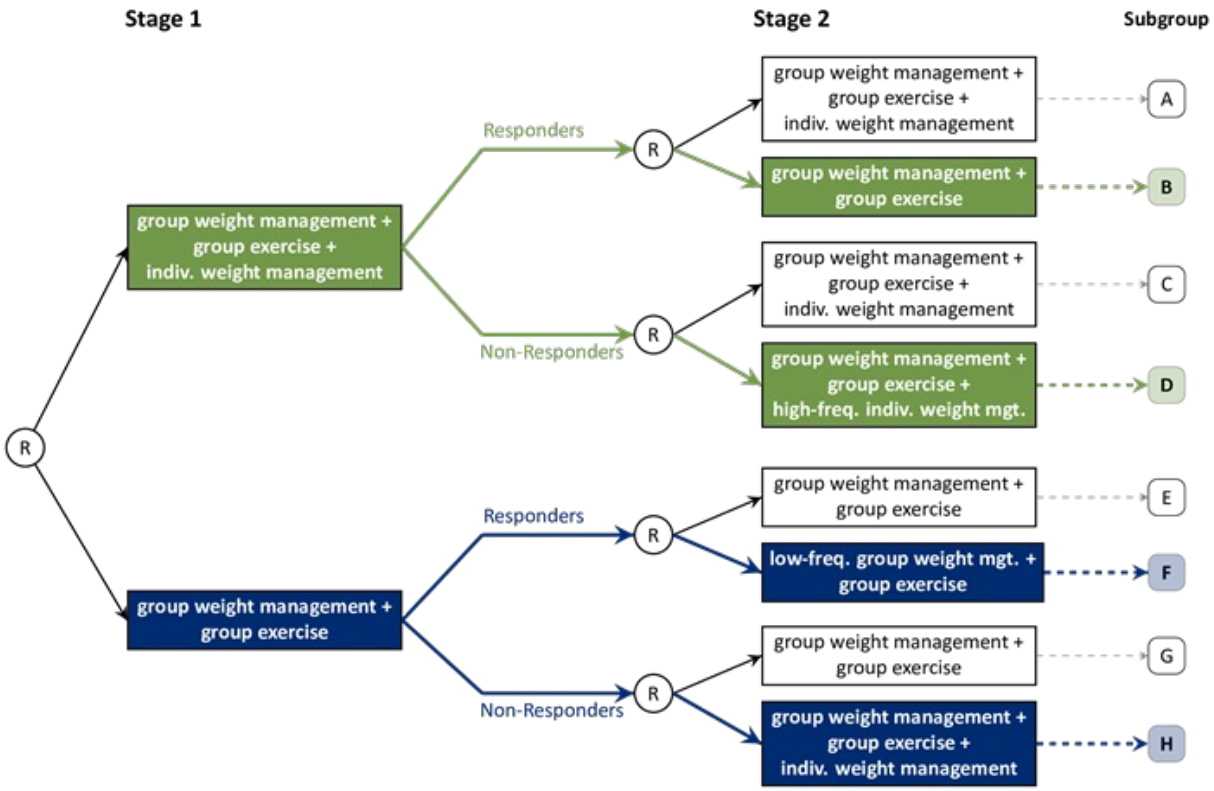


Analysis is a comparison of subgroups B & D vs. subgroups F & H.

- In general, need to account for unique design features of a SMART when comparing DTRs.

Common Primary Aims for SMARTs

Compare embedded DTRs



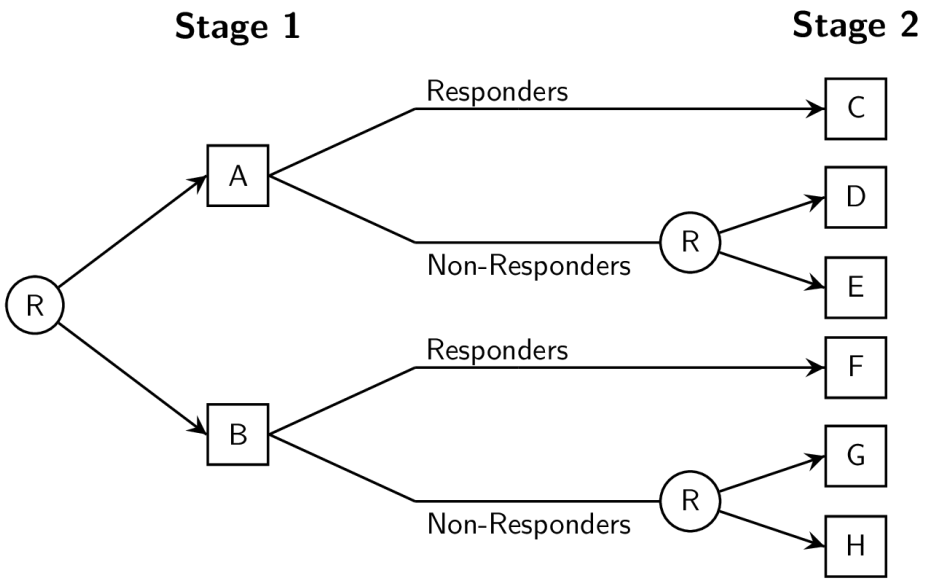
Sample size formulae available for many outcome types. For *this* design with a continuous outcome:

$$n \geq \frac{4\left(z_{1-\frac{\alpha}{2}} + z_{1-\gamma}\right)^2}{\delta^2} \cdot 2$$

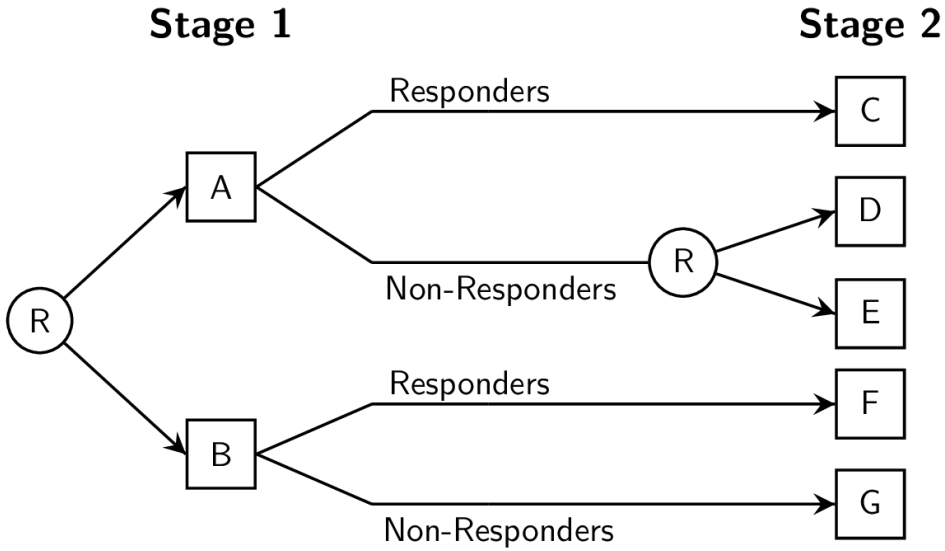
	80% power	90% power
$\delta = .3$	$n = 702$	$n = 938$
$\delta = .5$	$n = 256$	$n = 342$
$\delta = .8$	$n = 104$	$n = 136$

*If response rate is different between A & B, design effect becomes $\left(2-\frac{r_A+r_B}{2}\right)$

Sample Size for Comparing Embedded DTRs



$$n \geq \frac{4\left(z_{1-\frac{\alpha}{2}}+z_{1-\gamma}\right)^2}{\delta^2}\times(2-r)^*$$



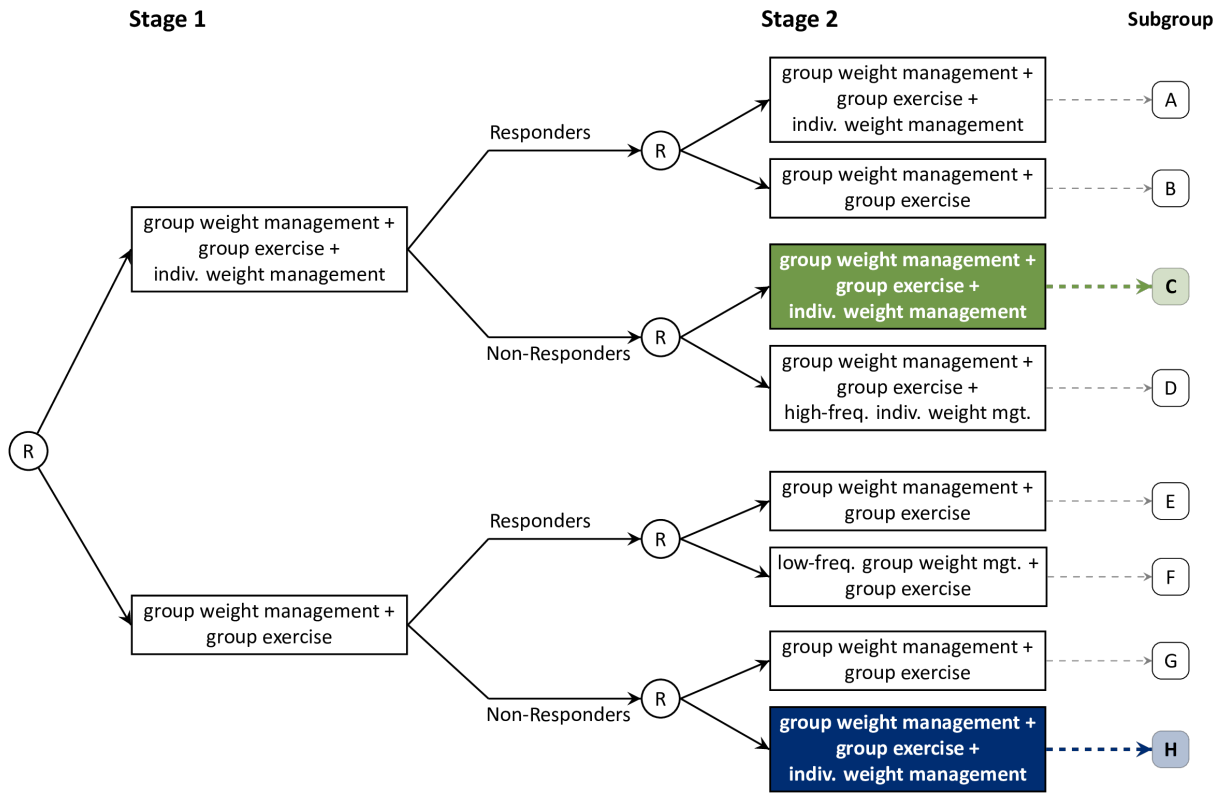
$$n \geq \frac{4\left(z_{1-\frac{\alpha}{2}}+z_{1-\gamma}\right)^2}{\delta^2}\times\left(\frac{3}{2}-r\right)$$

Questionable Primary Aims for SMARTs

Compare individual subgroups or experimental conditions

DTRs recommend treatments for every level of the tailoring variable.

This is not a question about DTRs and is not strong motivation for a SMART.

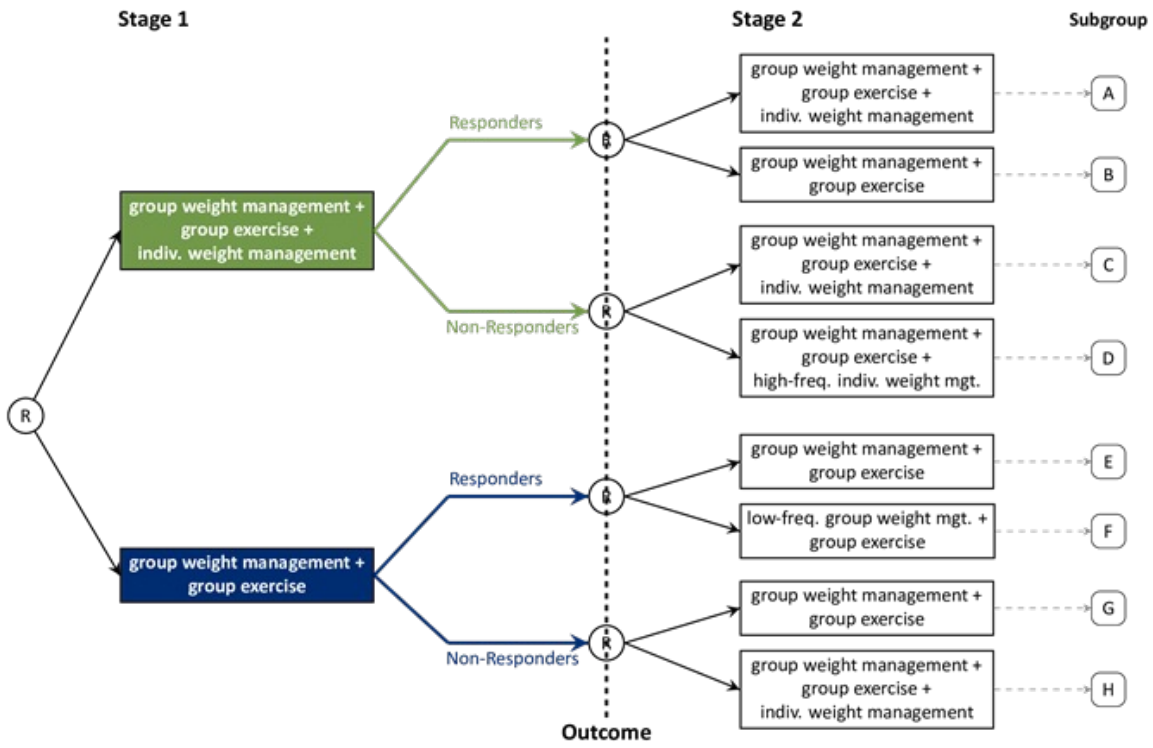


Questionable Primary Aims for SMARTs

Compare response rates to first-stage interventions

Not about adaptive interventions: ignores stage-2 treatment

Maybe an interesting secondary analysis, but is not strong motivation for a SMART.



Questionable Primary Aims for SMARTs

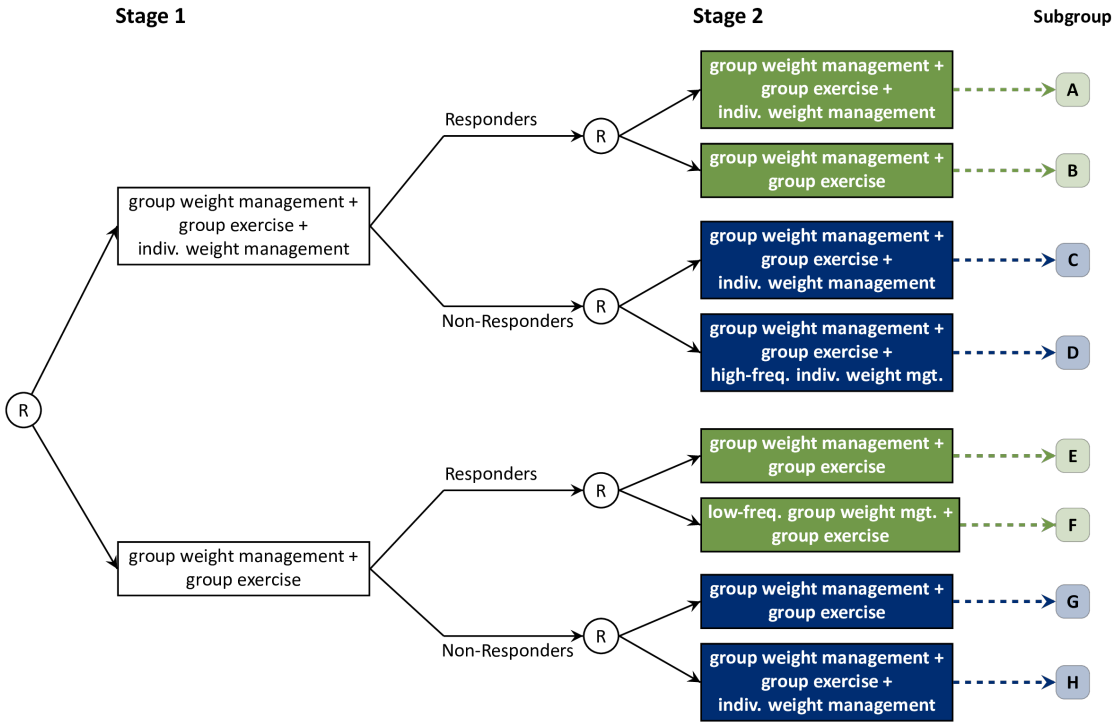
Compare responders to non-responders

This is a non-randomized comparison: we did not experimentally assign response status

Not really a question about DTRs

- DTRs recommend treatments for *both* responders and non-responders

A non-randomized comparison does not motivate a randomized trial.

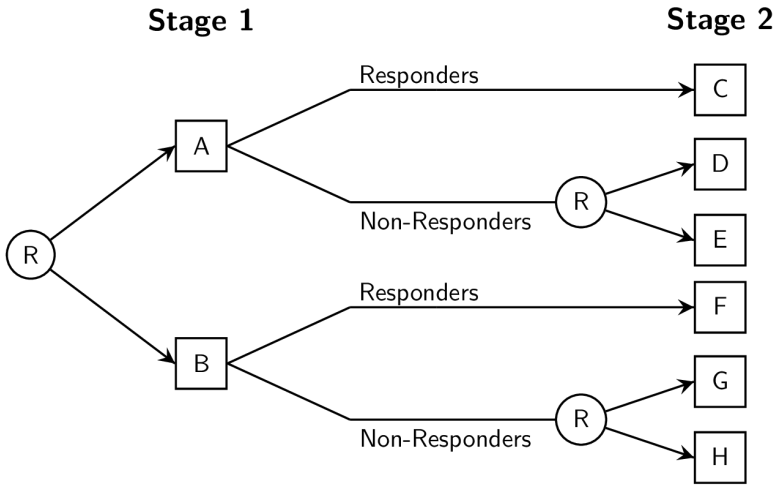


Why a SMART and not _____?

Not all research on DTRs requires a SMART. We've seen some examples already.

When a SMART is an option, why might you choose to use it over something else?

Why a SMART and not multiple separate trials?



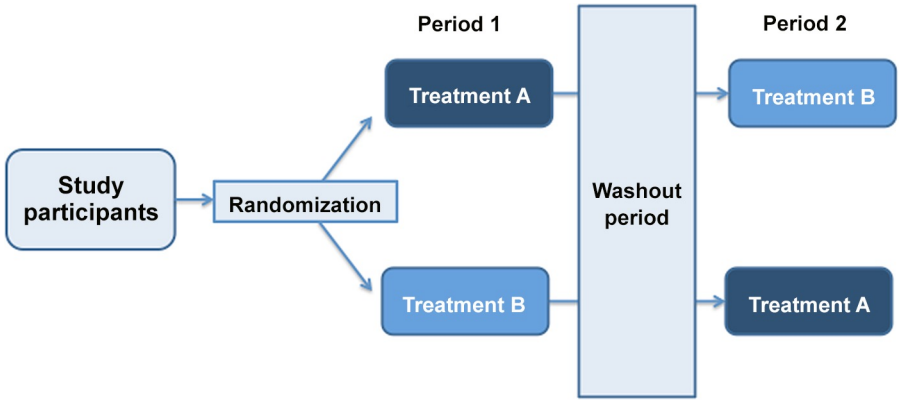
VS.

1. DELAYED EFFECTS
2. DROP-OUT
3. SELECTION EFFECTS
4. RICH DATA

Why a SMART and not a crossover trial?

In a **crossover trial**, participants start on one treatment then switch to another after a *washout period*.

The goal of a crossover trial is typically to evaluate the effects of standalone treatments – generally want to wash out any carryover effects.



Li, et al. (2015) *PLOS One*. https://doi.org/f8zws8

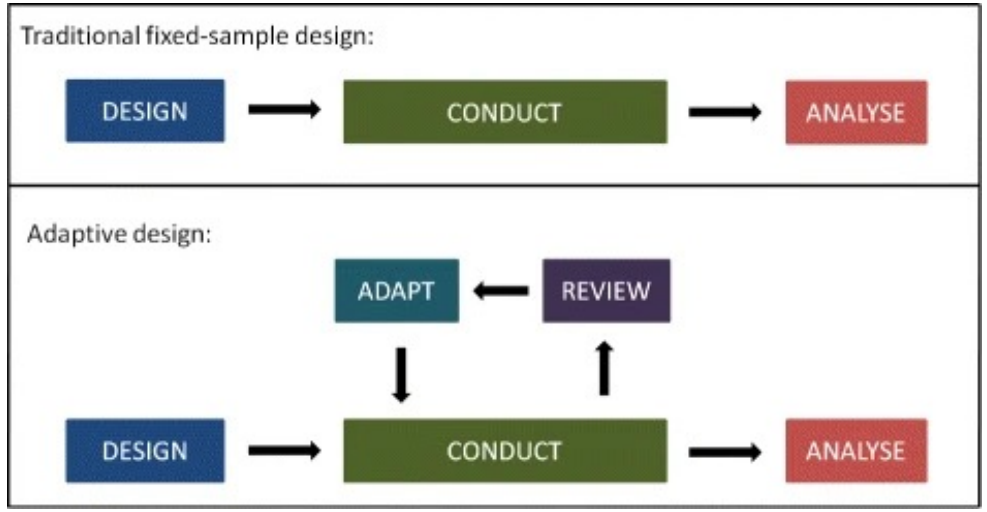
Why a SMART and not an adaptive trial?

An **adaptive trial** is a multistage study in which data collected throughout the trial is used to *modify features of the trial itself.*

- e.g., early stopping, dropping arms, modifying randomization probabilities, etc.

SMARTs are typically fixed designs: all participants move through every stage of the trial as it was initially designed.

In adaptive trials, the *trial* is adaptive. SMARTs are designed to address questions about *interventions* that are adaptive.



Pallmann, et al. (2018) *BMC Medicine*. https://doi.org/gc6jrz

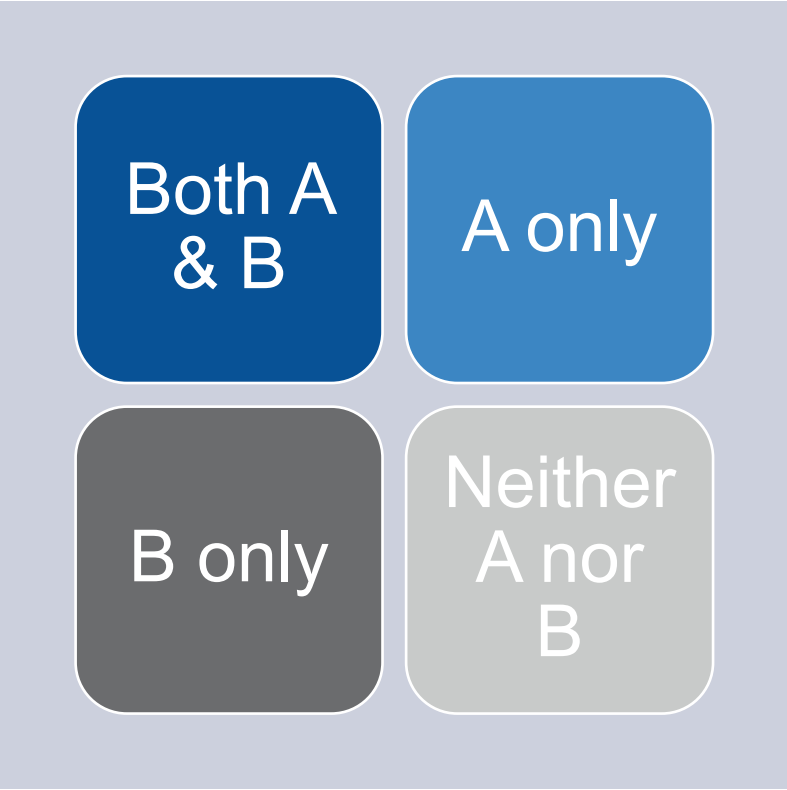
Seewald, N.J., et al. 2021. “Sequential, Multiple Assignment, Randomized Trials (SMART).” In Principles and Practice of Clinical Trials, edited by Steven Piantadosi and Curtis L. Meinert, 1–19. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-52677-5_280-1.

Why a SMART and not a factorial trial?

In a factorial trial, two or more *factors* (each with 2+ *levels*) are crossed to create different experimental conditions.

SMARTs are conceptually similar to (fractional) factorial designs in which treatments are delivered *sequentially*.

- A *fractional factorial design* does not fully cross all levels of all factors



Seewald, N.J., et al. 2021. “Sequential, Multiple Assignment, Randomized Trials (SMART).” In Principles and Practice of Clinical Trials, edited by Steven Piantadosi and Curtis L. Meinert, 1–19. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-52677-5_280-1.

Why a SMART and not a factorial trial?

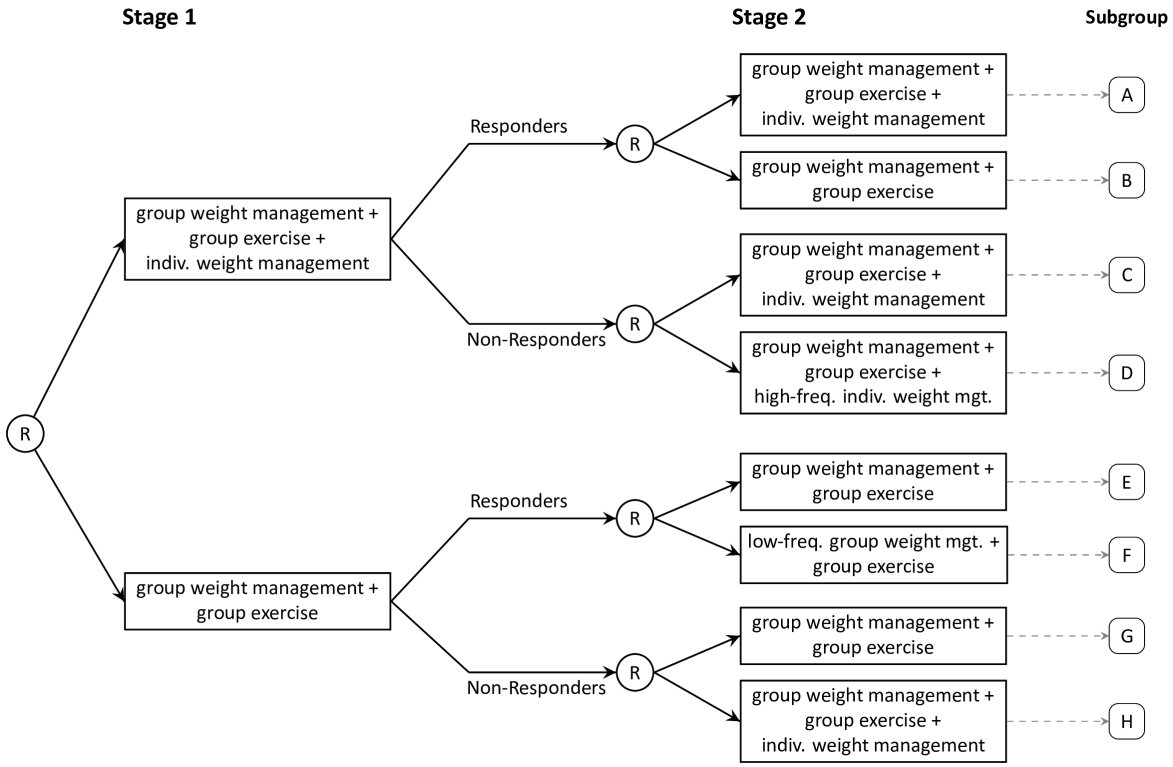
Our hypothetical SMART is similar to a (2 x 2 x 2) (fractional) factorial trial.

Factor 1: first-stage treatment options

Factor 2: second-stage tactic *for responders*

Factor 3: second stage tactic *for non-responders*

Factors 2 and 3 are restricted by the tailoring variable: a key difference from standard factorials!



Seewald, N.J., et al. 2021. “Sequential, Multiple Assignment, Randomized Trials (SMART).” In Principles and Practice of Clinical Trials, edited by Steven Piantadosi and Curtis L. Meinert, 1–19. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-52677-5_280-1.

Resources and References

Sample size for comparing embedded adaptive interventions

- **Continuous Outcomes**
  - Oetting, A. I., et al. 2011. “Statistical Methodology for a SMART Design in the Development of Adaptive Treatment Strategies.” In *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*, edited by P.E. Shrout, K.M. Keyes, and K. Ornstein, 179–205. New York: Oxford University Press.
  - Ogbagaber, S.B., J. Karp, and A.S. Wahed. 2016. “Design of Sequentially Randomized Trials for Testing Adaptive Treatment Strategies.” *Statistics in Medicine* 35 (6): 840–58. https://doi.org/10.1002/sim.6747.
- **Continuous Longitudinal Outcomes**
  - Seewald, N.J, K.M. Kidwell, I. Nahum-Shani, T. Wu, J.R. McKay, and D. Almirall. 2020. “Sample Size Considerations for Comparing Dynamic Treatment Regimens in a Sequential Multiple-Assignment Randomized Trial with a Continuous Longitudinal Outcome.” *Statistical Methods in Medical Research* 29 (7): 1891–1912. https://doi.org/10/gf85ss.
- **Binary Outcomes**
  - Kidwell, K.M., N.J. Seewald, Q. Tran, C. Kasari, and D. Almirall. 2018. “Design and Analysis Considerations for Comparing Dynamic Treatment Regimens with Binary Outcomes from Sequential Multiple Assignment Randomized Trials.” *Journal of Applied Statistics* 45 (9): 1628–51. https://doi.org/10.1080/02664763.2017.1386773.

Resources and References

Sample size for comparing embedded adaptive interventions

- **Survival / Time-to-Event Outcomes**
  - Feng, W., and A.S. Wahed. 2009. “Sample Size for Two-Stage Studies with Maintenance Therapy.” *Statistics in Medicine* 28 (15): 2028–41. https://doi.org/10.1002/sim.3593.
  - Li, Z., and S.A. Murphy. 2011. “Sample Size Formulae for Two-Stage Randomized Trials with Survival Outcomes.” *Biometrika* 98 (3): 503–18. https://doi.org/10.1093/biomet/asr019.
- **Continuous Outcomes in a Cluster-Randomized SMART**
  - NeCamp, T., A. Kilbourne, and D. Almirall. 2017. “Comparing Cluster-Level Dynamic Treatment Regimens Using Sequential, Multiple Assignment, Randomized Trials: Regression Estimation and Sample Size Considerations.” *Statistical Methods in Medical Research* 26 (4): 1572–89. https://doi.org/10.1177/0962280217708654.
- **Find the Best Embedded Adaptive Intervention**
  - Artman, W.J., I. Nahum-Shani, T. Wu, J.R. Mckay, and A. Ertefaie. 2018. “Power Analysis in a SMART Design: Sample Size Estimation for Determining the Best Embedded Dynamic Treatment Regime.” *Biostatistics*. https://doi.org/10/ggth75.

Resources and References

In-depth book on adaptive interventions and SMARTs

- Kosorok, M.R., and E.E.M. Moodie, eds. 2015. Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine. Philadelphia, PA: Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611974188.

Overview of a variety of SMARTs in the field

- Lei, H., et al. "A 'SMART' Design for Building Individualized Treatment Sequences." Annual Review of Clinical Psychology 8, no. 1 (2012): 21–48. https://doi.org/10.1146/annurev-clinpsy-032511-143152.

Clear explanations of primary aim analysis in SMARTs

- Nahum-Shani, I., et al. 2012. "Experimental Design and Primary Data Analysis Methods for Comparing Adaptive Interventions." *Psychological Methods* 17 (4): 457–77. https://doi.org/10.1037/a0029372.

Resources and References

Analysis of Longitudinal Outcomes in SMARTs

- Lu, X., Nahum-Shani, I., Kasari, C., Lynch, K. G., Oslin, D. W., Pelham, W. E., Fabiano, G., & Almirall, D. (2016). Comparing dynamic treatment regimes using repeated-measures outcomes: Modeling considerations in SMART studies. *Statistics in Medicine*, *35*(10), 1595–1615. https://doi.org/10/gg2gxc

Noninferiority and Equivalence Testing in SMARTs

- Ghosh, P., Nahum-Shani, I., Spring, B., & Chakraborty, B. (2020). Noninferiority and equivalence tests in sequential, multiple assignment, randomized trials (SMARTs). *Psychological Methods*, *25*(2), 182–205. https://doi.org/10/ggtmgv

Example of Cluster-Randomized SMART

- Kilbourne, A. M., et al. (2014). Protocol: Adaptive Implementation of Effective Programs Trial (ADEPT): cluster randomized SMART trial comparing a standard versus enhanced implementation strategy to improve outcomes of a mood disorders program. *Implementation Science*, *9*(1), 132. https://doi.org/10/f6q9fc

Get in touch!

www.nickseewald.com

seewaldn@pennmedicine.upenn.edu

Penn Medicine