Sequential, Multiple Assignment, Randomized Trials

Module 2

Getting SMART: Experimental Design and Analysis Methods for Developing Adaptive Interventions

In this module, we'll talk about an *experimental design* that lets us develop effective adaptive interventions, called a sequential multiple assignment randomized trial, or SMART.

Outline

- **What are Sequential Multiple Assignment Randomized Trials (SMARTs)?**
- Trial Design Principles and Analysis
- Advantages of a SMART
- Summary & Discussion

"Ignorance of whether or how to change psychotherapies is a major and persisting gap in psychiatric knowledge."

John Markowitz, Barbara Milrod (2015).
The Lancet Psychiatry, 2(2), 186-190.

Quote clearly justifies the need for adaptive interventions and highlights the fact that there are many open scientific questions which prevent the development of a high-quality one.

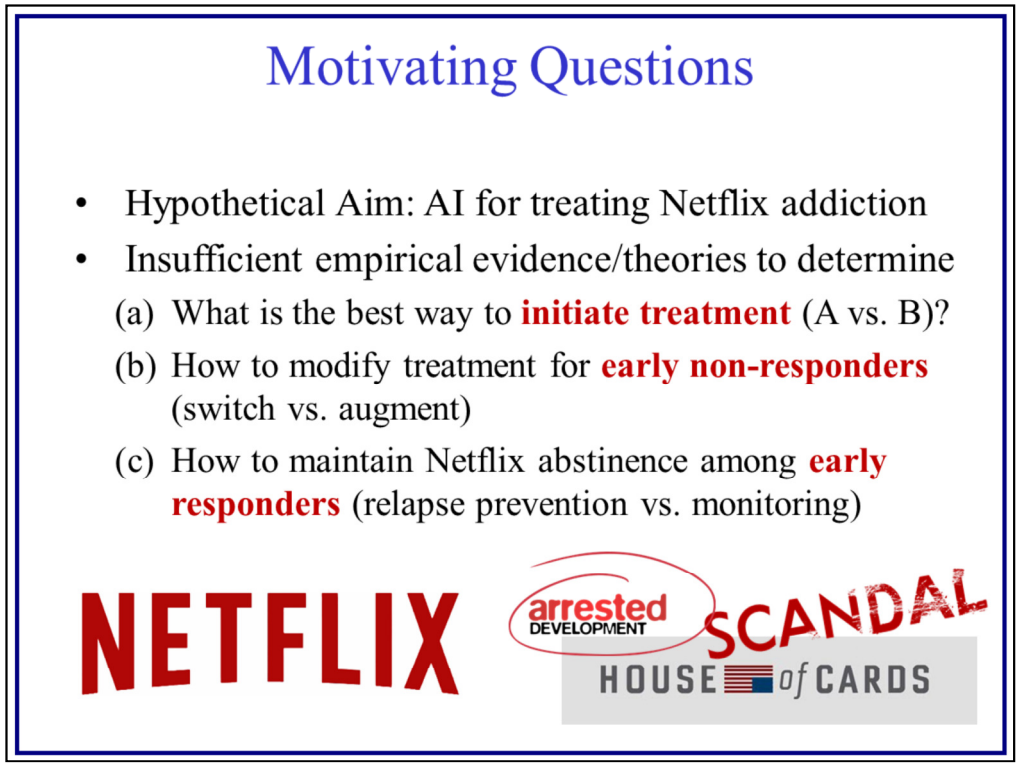
What is a SMART?

- A Multi-Stage Randomized trial
- Each stage corresponds to a critical decision point
- A randomization takes place at each critical decision
- Some (or all) participants are randomized more than once, often based on earlier covariates

The goal is to inform the construction of effective adaptive interventions

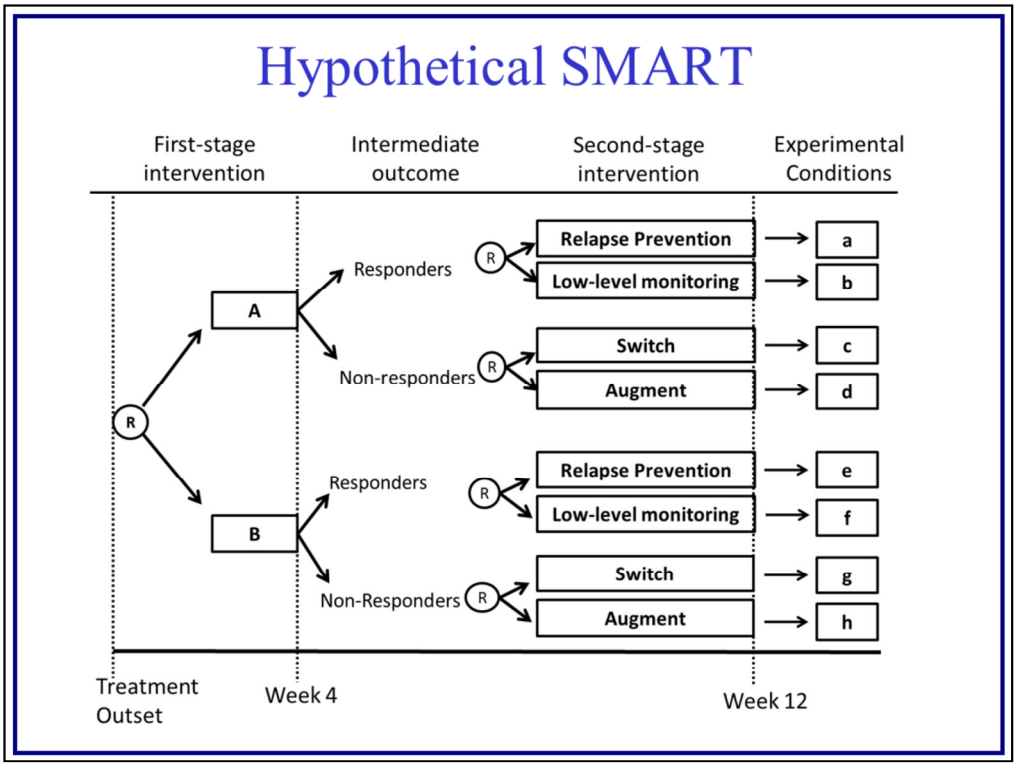
- The key feature of a SMART is that some or all participants can be randomized more than once. You'll see this in examples later on.
- SMARTs are always motivated by scientific questions regarding adaptive interventions. Let's see how open questions can be addressed by SMARTs.

- In statistics, people may call these multistage trials (the randomization at each stage is assumed)



- We (hypothetically) want to develop an adaptive intervention for Netflix addition, but there are three open *scientific questions* that are preventing us from developing a high-quality adaptive intervention.
- We have empirical evidence suggesting that both A and B are effective treatments, but there is debate as to which is better to start with. Maybe one is more expensive or has worse side effects, etc.
- We know that a fairly large proportion of people don't respond well to either A or B, and we can identify them early on. We need to prevent early non-responders from failing, but we don't know the best way to modify treatment for these people: do we switch them to the other option, or do we augment their existing first-stage therapy?
- We also know that even among responders to A and B, risk of relapse is pretty high. So we have to do something to maintain abstinence, but we don't know what: should we give relapse prevention therapy, or just some low-level monitoring?

Notice the scientific justification for the restriction of subsequent intervention options. Non-responders need more or different type of treatment; whereas responders need some sort of maintenance strategy, but we are not sure what kind.



Hypothetical trial: Outcome is not shown but is on far right. The randomizations can take place up front.

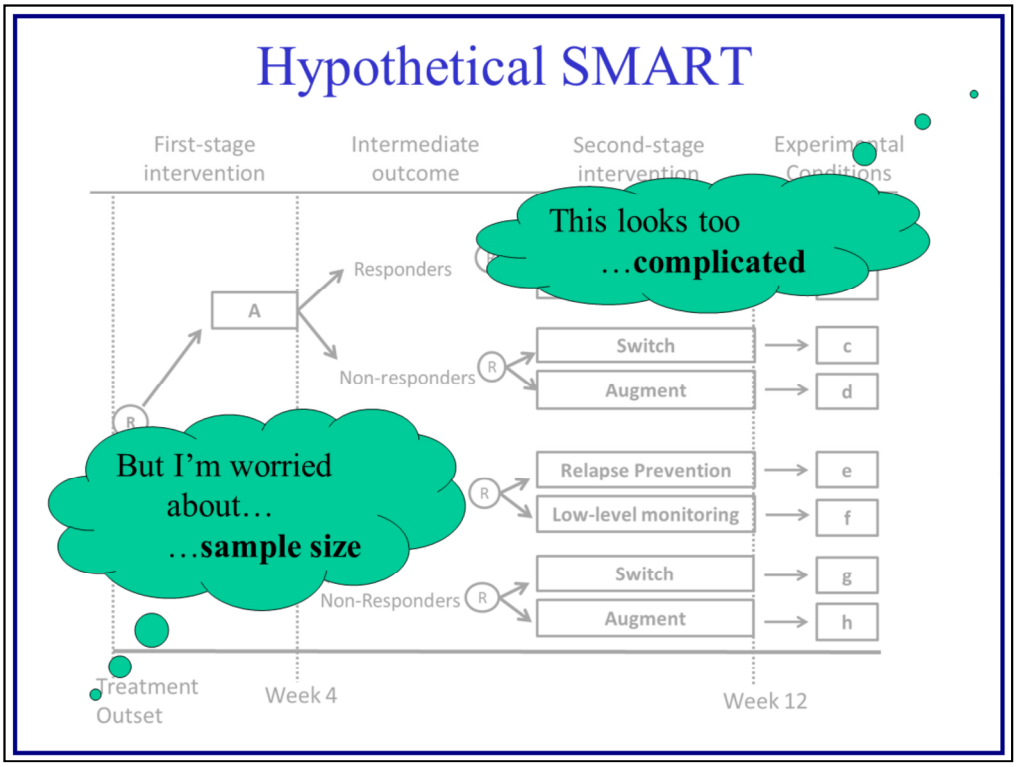
Equal randomization.

Usual reaction is

(1) I'm worried about sample size and

(2) This looks awfully complicated.

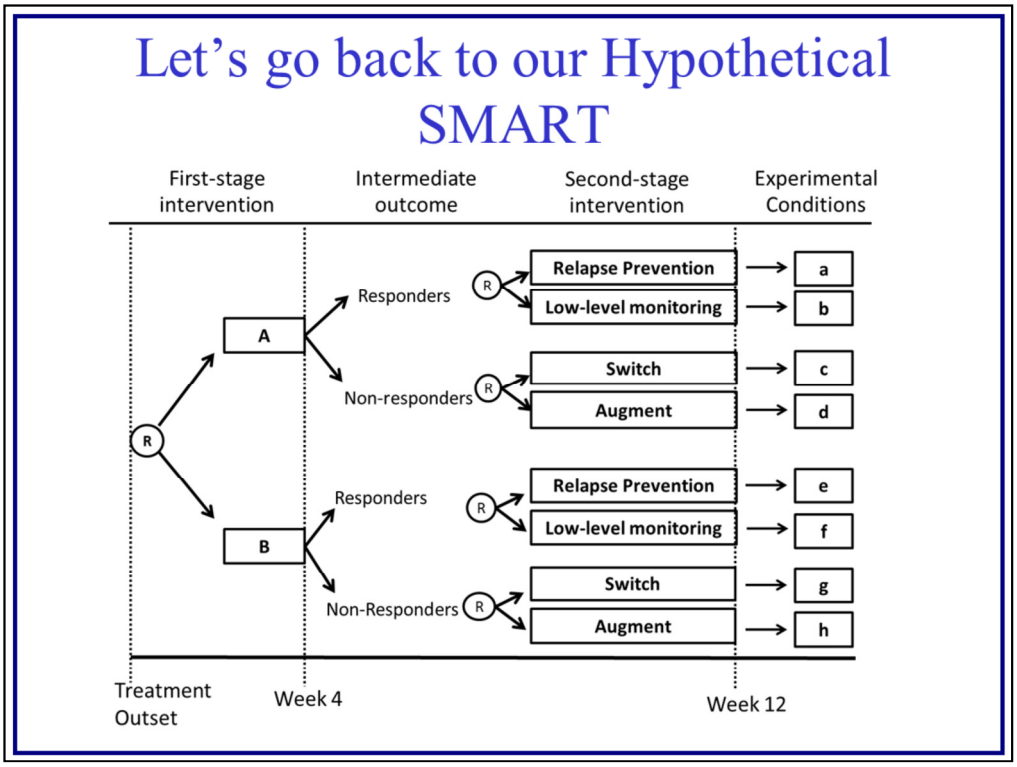
In reality both of these problems are less worrisome than one might think—see following slides.



- Two common worries people have are about sample size and the design's *perceived* complexity.
- You might be thinking "There are eight experimental conditions here! How am I ever going to get a big enough sample size to be able to compare them?" Well, as we'll see later in this module, we size SMARTs to compare *groups* of experimental conditions. We never compare them individually, and this helps alleviate that concern.
- You might also be looking at this and thinking "This looks really complicated. How am I going to be able to explain and justify it to readers and reviewers?" Something we want to do with this module is show you that SMARTs *aren't* complicated. What's complex about a SMART is the way we talk about it. An RCT can seem very complex if you talk about it in a complex way, and the same holds for SMARTs. Later in this module, we'll talk about a set of core design principles that help reduce this perceived complexity.

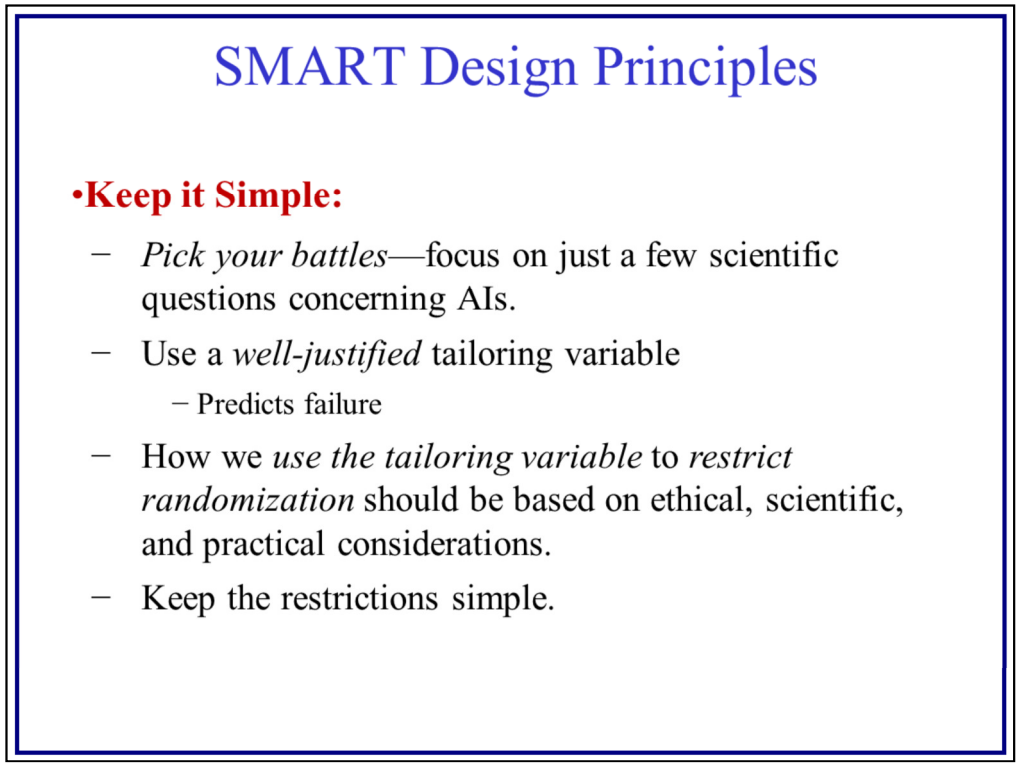
Outline

- What are Sequential Multiple Assignment Randomized Trials (SMARTs)?
- **Trial Design Principles and Analysis**
- Advantages of a SMART
- Summary & Discussion



Hypothetical trial: Outcome is not shown but is on far right. The randomizations can take place up front.

Equal randomization



For more information about a well-justified tailoring variable, see module 1 (remember, three kinds: obvious, predictor, moderator).

Considerations for restricting randomization:

Ethical: A situation where a subset of treatment options is not appropriate for a subset of participants for ethical reasons (e.g., intensifying already-intense chemotherapy). So, restrict randomization in a way that avoids unethical assignments.

Scientific: Based on empirical evidence. We might have established treatment protocols for responders, i.e., we know what to do for them, so we won't re-randomize them. But, there may be some doubt about what to give non-responders, so they're re-randomized.

Practical: For example, a stepped-care approach. Save the most intense, most expensive treatments for the people who need them (re-randomize non-responders to these), and keep responders at the same intensity, or step them down (re-randomize responders to these).

Keeping restrictions simple:

You can use an endless number of intermediate outcomes to restrict the class of second-stage options. But then the decision tree will be over complicated to justify and implement (e.g., non-compliant non-

responders, compliant non-responders, non-compliant responders, compliant responders, etc.)

But it is important that you keep it simple: use a low dimensional summary (e.g., response status) and then specify how it is operationalized; namely, clearly state how you define responders and non-responders via intermediate outcomes. In mental illness studies feasibility considerations may force us to use preference in this low dimensional summary.

SMART Design Principles

- **Plan to collect intermediate outcomes needed to ascertain response status.**
  - But also consider collecting other information that might be useful in ascertaining for whom each treatment works best
  - Information that might enter later into the AI
  - Namely, candidate tailoring variables

SMART Design Principles

- **Choose a primary hypothesis**
  - that is both scientifically important and aids in developing the AI
  - Choose sample size to address this hypothesis

SMART Design Principles

- **Choose secondary hypotheses**
  - that further develop the AI and use the randomization to eliminate confounding.
    - Confounding: alternative explanations other than treatment effect for the observed difference
  - Sample size doesn't have to be determined based on these hypotheses.

Confounding: alternative explanations other than treatment effect for the observed difference

Example of Primary Aims

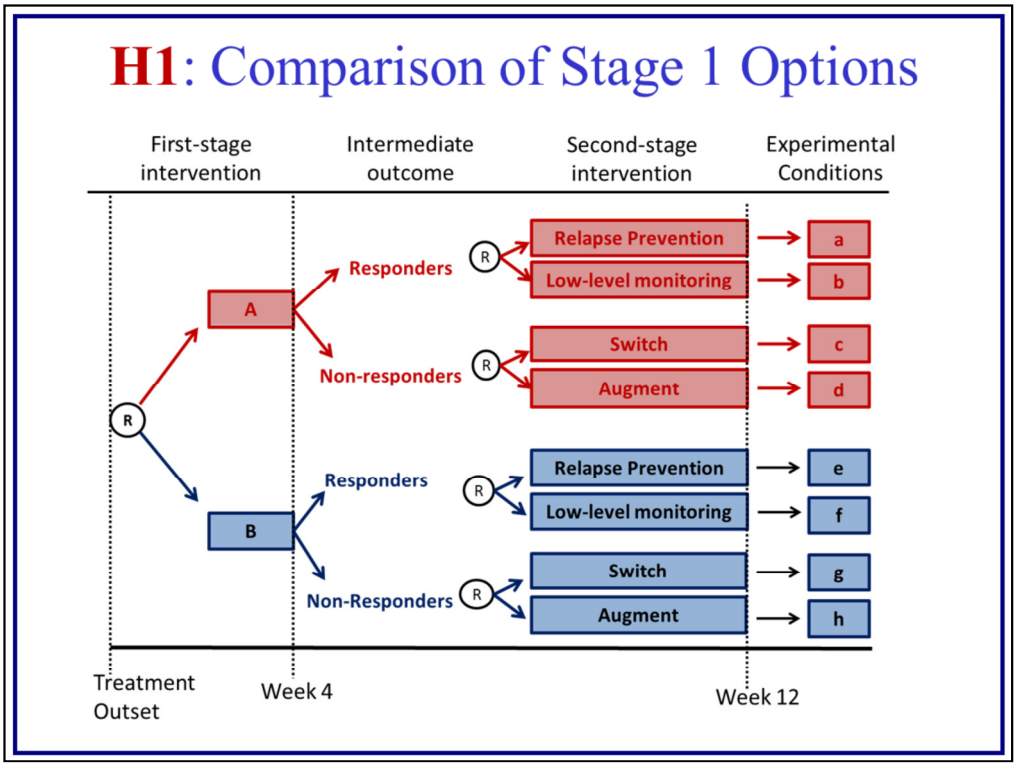
1. *Comparison of initial options*

- **H1:** The initial intervention option A results in lower symptoms than the initial intervention option B.
  - Controlling for second-stage intervention options

This is the main effect of the initial intervention options a la' ANOVA.

Here, we are controlling for second-stage treatment by design –not by statistical analysis.

Because of the randomizations, we are ruling out alternative explanations like severity at baseline (for the effect of first stage).



A study of initial intervention options in which subsequent intervention options are controlled.

Here you can use a variety of analyses, growth curve models, survival analysis, etc.

Example of Primary Aims

2. *Comparison of second stage options for non-responders*

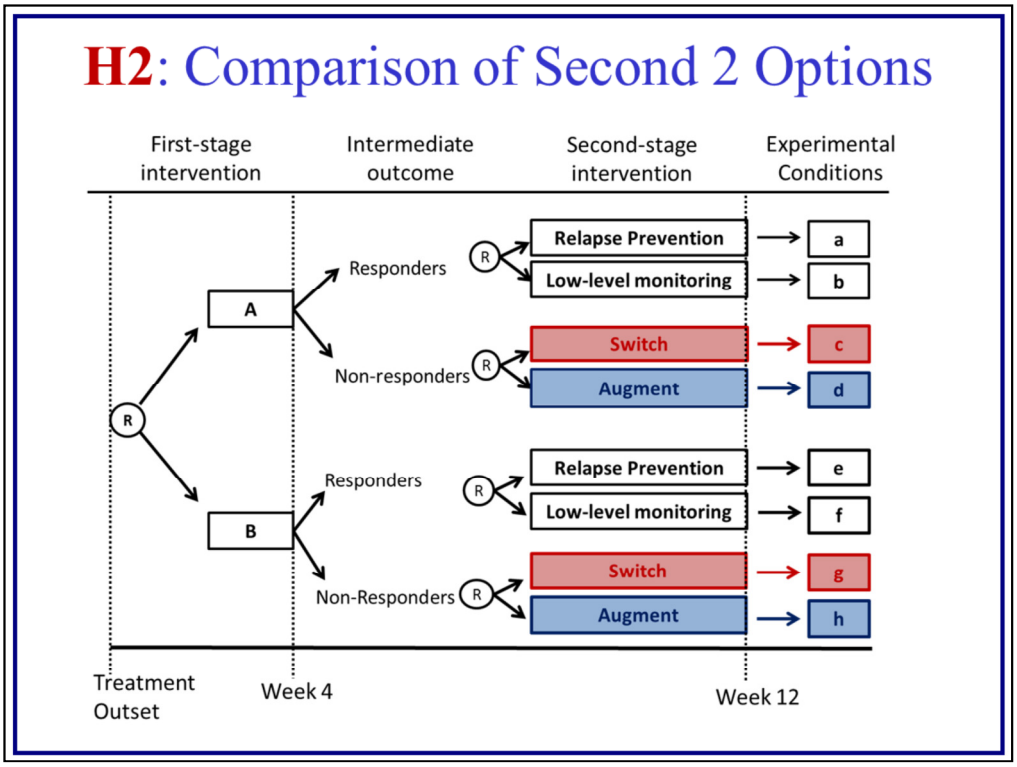
- **H2:** Among non-responders, switching treatments results in lower symptoms than augmenting existing treatment
  - Controlling for first-stage intervention options

This is the main effect of the second-stage intervention options among non-responders, again a la' ANOVA.

Here, we are controlling for first-stage treatment by design-– not by statistical analysis.

This primary hypothesis would be appropriate if you initially wanted to run a trial for non-responders and are now considering a SMART.

Because of the re-randomizations, we are ruling out alternative explanations like adherence: people who do not adhere will be switch, so all switched people are non-adherent (for the second-stage).

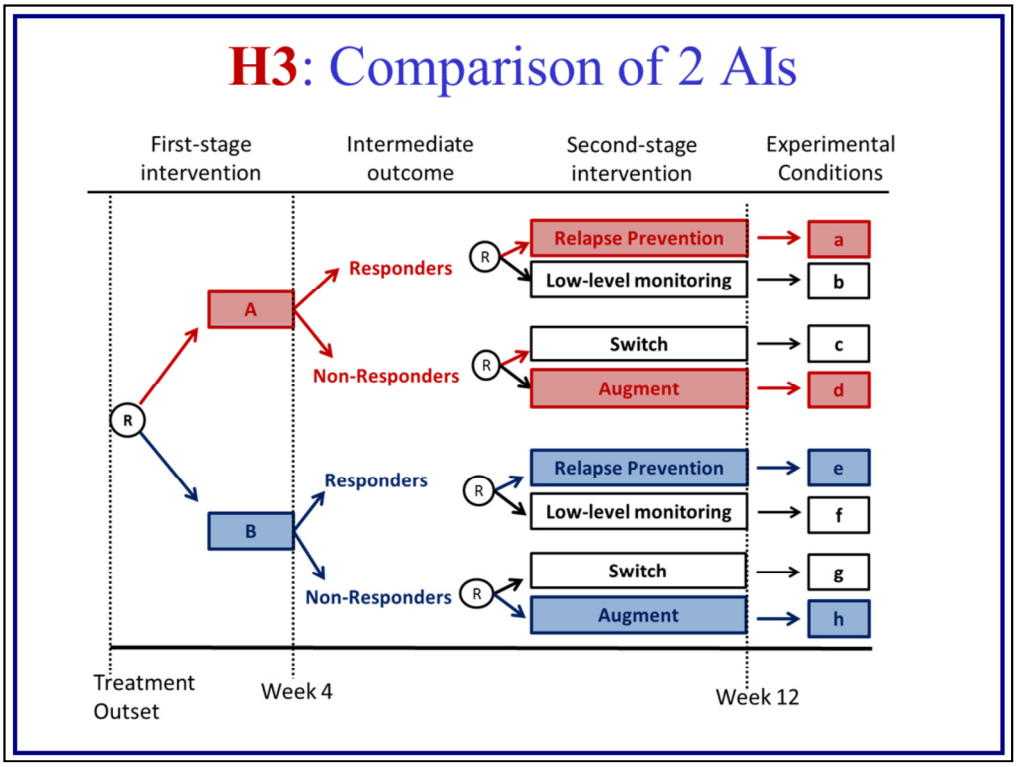


A study of non-responders in which one controls the initial intervention option to which people don't respond to.

Example of Primary Aims

3. *Comparison of embedded adaptive interventions*

- **H3:** Adaptive intervention #1 results in improved symptoms compared to adaptive intervention #2



There are two ways to think about this comparison:

(1)Comparison of AI that begin with different options (and continue with the same) – framing is around the AI

(2) assuming that we will treat non-responders with relapse prevention and non-responders with augment, is it better to start with A or B) – framing is around the initial intervention options.

In every SMART design there are several (more than 2) embedded AIs.

Here, there are 8 embedded AIs.

Participants in subgroups a and d are consistent with these AI, because participants in these two subgroups experience this sequence of treatments.

The AI operationalizes the intervention options for both responders and non-responders and hence both responders and non-responders are consistent with each AI.

Sample Size

H1: The initial intervention option A results in lower symptoms than the initial intervention option B.

- *Sample size formula is same as for a **two group comparison**.*

H2: Among non-responders, a switch to C results in lower symptoms than augmenting with D

- *Sample size formula is same as a **two group comparison of non-responders**.*

Again, these are main effects a la' ANOVA.

Sample Size

N= sample size for the ***entire*** trial

	H1	H2
$\Delta\mu/\sigma$ =.3	*N* = 352	*N* = 352/ NR rate
$\Delta\mu/\sigma$ =.5	*N* = 128	*N* = 128/ NR rate

α = .05 (two sided), power =1 – β =.80

*Assumptions: equal variances, normality, equal # in each group, no dropout.

Example sample sizes for entire trial for example primary aims H1 and H2, assuming a *continuous outcome*. We're able to use a standard online calculator for a two-group comparison with continuous outcomes (see below). If you don't have a continuous outcome, you can use other standard calculators which accommodate that.

Sigma for example 1 is the standard deviation of primary outcome of patients initially assigned to intervention option A (or B).

Sigma for example 2 is the standard deviation of primary outcome of non-responding patients who are assigned a switch (or augment).

Throughout working assumptions are equal variances, normality, equal number in each of the two

groups being compared, and no dropout or loss to follow-up.

** What if I have very small rate of non-responders in one of the arms (say 4 non-responders to B) how does this influence my power? (1) it will not influence your power for H1; it will influence your power for H2 (which is only based on information from non-responders, and you have very few); and most importantly this implies that you don't need to re-randomize non-responders to B because you anticipate very few of them, so this has implications for how you design the study.

Sample sizes calculated on the website (David A. Schoenfeld):
http://hedwig.mgh.harvard.edu/sample_size/js/js_parallel_quant.html

Sample Size

H3: AI #1 results in improved symptoms compared to AI #2

- Analysis is non-standard (so sample size calculation is too)
- Sample size formula depends on who gets re-randomized

Type I error rate (2-sided)	Power	Standardized Difference	N	Randomization
0.05	80%	0.3	697	Both R and NR are re-randomized
		0.5	251	

- *Continuous Outcomes:* Oetting, A.I., et al. (2011)
- *Survival Outcomes:* Feng, W. and Wahed, A., (2009); Li, Z. and Murphy, S.A., (2011)
- *Binary Outcomes:* Kidwell, K.M., et al. (In preparation)

Analysis for this primary aim is nonstandard (a weighted and replicated approach)—we'll talk about that in more detail in modules 4 and 5. Because the analysis is nonstandard, we can't use a standard sample size calculator. Susan Murphy's group developed a sample size formula for SMARTs with a continuous outcome in which the primary aim is to compare two embedded AIs. These sample sizes were computed using that method (described in the cited book chapter).

Here, sample size is dependent on the design: namely, who gets re-randomized. Remember that tailoring variables are used to restrict randomization options in the second stage, and it's possible that we know what to do for responders, for example (e.g., have them continue on initial therapy). Sample size is lower for designs that only re-randomize responders. Remember, though, that the choice of who to re-randomize should be made based on ethical, scientific, or practical considerations. See below for more details.

Full Citations:

- Oetting, A.I., Levy, J.A., Weiss, R.D. Murphy, S.A. (2011), Statistical Methodology for a SMART Design in the Development of Adaptive Treatment Strategies (book chapter)
- Z. Li and S.A. Murphy, Sample Size Formulae for Two-Stage Randomized Trials with Survival Outcomes. Biometrika 2011; 98(3):503-518.
- Feng W, Wahed AS. Sample size for two-stage studies with maintenance therapy. Stat Med 2009;28:2028-41.

The results are for comparing AIs in a setting where both responders and non-responders are split

into two groups. You will need a much lower sample size to compare AIs in a setting where only 1 sub-group (e.g., non-responders) are re-randomized.

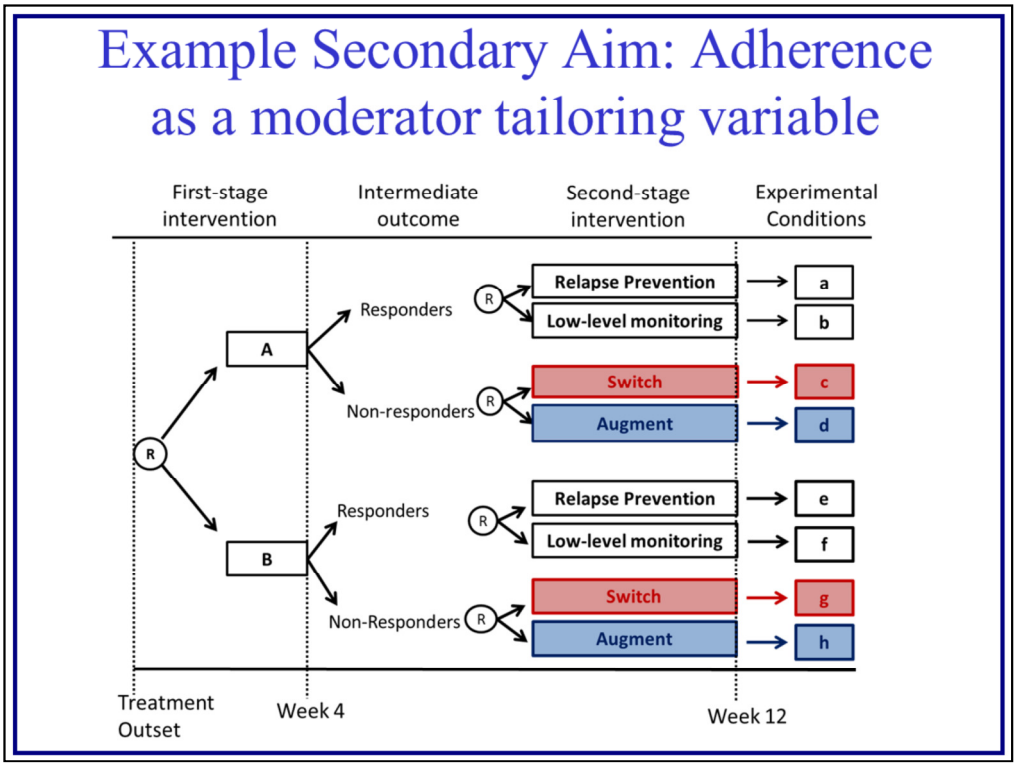
In case studies, we'll see an example of a SMART that re-randomized only non-responders (ADHD study). Responders were assigned "continue". This was done because if initial therapy worked, then there was no reason to modify treatment. To size studies of this kind, we need to hypothesize a non-response rate, since only non-responders are split into two groups. Assuming 30% non-response, we need N=453 to detect a standardized effect size of 0.3, and 163 for a standardized effect size of 0.5 The sample size needed for this comparison will be lower than in a trial in which both responders and non-responders are re-randomized to the extent that NR rate is lower. This is because I will have more people in the sub-group that is not split into two– I can use info from only half of these subjects in the comparison of AIs. Sample size needed will increase with non-response rate.

** What about the comparison of AIs that begin with the same initial treatment – we rarely see investigators interested in comparing AIs that begin with the same treatment. Tomorrow we will provide a way to compare AIs that begin with same and different treatment.

Example of Secondary Aims

- **Choose secondary hypotheses**
  - That further develop the AI.
  - Example:

    H4: *non-adhering* non-responders will exhibit lower symptoms if their initial treatment is augmented as compared to switching treatments (e.g. augment includes motivational interviewing).



NRs ARE HETEROGENEOUS

I'm basically proposing to explore whether adherence is a moderator of the second-stage intervention options. The second-stage intervention options for non-responders are randomized, I can test whether the second-stage intervention effect for non-responders varies depending on the level of adherence to first-stage.

Outline

- What are Sequential Multiple Assignment Randomized Trials (SMARTs)?
- Trial Design Principles and Analysis
- **Advantages of a SMART**
- Summary & Discussion

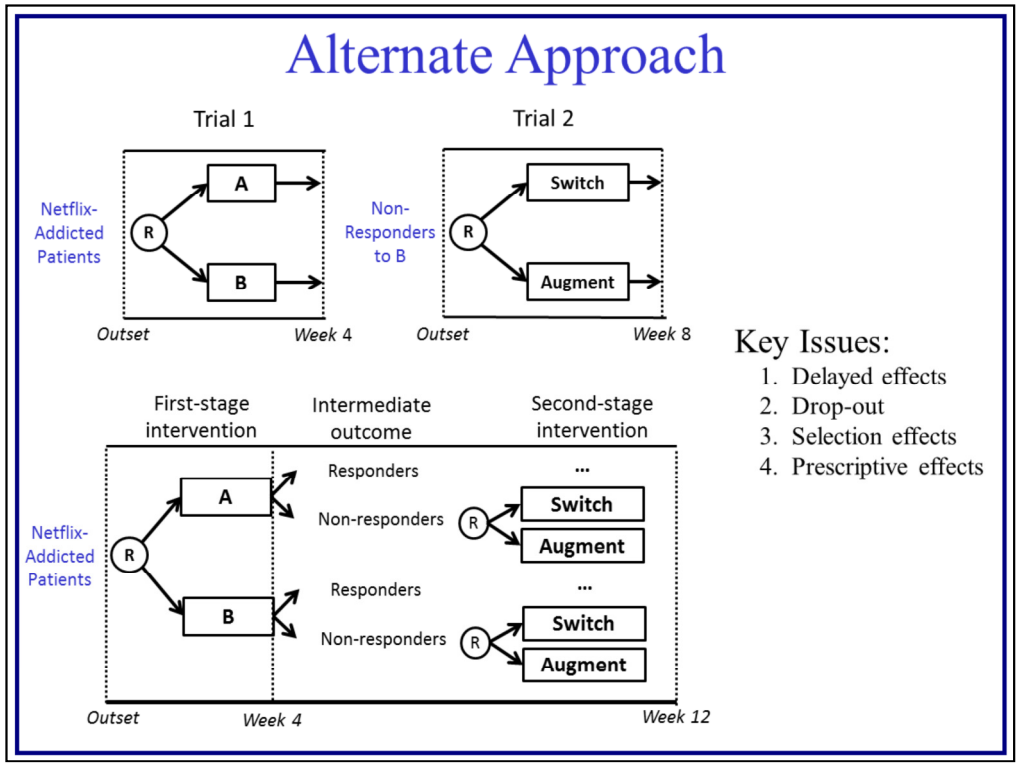
Alternate Approach to Constructing an Adaptive Intervention

- Why not use data from multiple trials to construct the AI?
- The single-stage at a time approach
  - **Trial 1**: Randomized trial of initial intervention options → Choose the best stage 1 option.
  - **Trial 2**: Randomized trial of secondary intervention options → Choose the best stage 2 option.

- People usually try to think of alternatives to SMARTs. Why, for instance, can't we use data from multiple trials to develop an adaptive intervention? This is something we call the single-stage-at-a-time approach.
- This approach goes something like this: we conduct two trials. In the first, we randomize between first-stage intervention options and pick the best one. Then, we do another trial to compare second-stage options among people who got the best first-stage treatment, and pick the best one from that. Then, we stick those two treatments together and make our "optimal" AI.

Particularly attractive since potential initial treatment may have been evaluated in prior trials. So you propose a responder study or you propose a nonresponder study.

Why choosing the best initial treatment on the basis of a randomized trial of initial treatments and choosing the best secondary treatment on the basis of a randomized trial of secondary treatments is not the best way to construct an adaptive intervention?



What happens in reality is that investigators make decisions about the initial options, based on available preliminary evidence/ tradition in their field. Then they might go to clinics where B is provided and they will recruit non-responders to B.

The Single stage approach might have several disadvantages compared to SMART

(1) Cant detect delayed effects: positive synergies (you are not collecting info about effect of A in stage 2 so you cant observe its effectiveness when followed by augment); negative synergies (B is better initially, but is highly burdensome, and this burden accumulates when you augment or switch which reduces overall effectiveness compared to A– with the single stage you might be able to see that both subsequent approaches are not effective, but you will not be able to understand why because you are not looking at the entire sequence– you cant see that burden accumulates during first stage and you wont be able to compare to A.

(2) Selection effect: people who enroll in SMART differ from single stage trials: (a) in SMART more motivation to enroll because they know you will offer something if they fail; (b) non-responders to B in single stage may not represent the population of non-responders because demoralized people (who got discouraged because B didn't work) will not join the study. In a SMART both the demoralized and motivated are included and get re-randomized and you can learn that the demoralized people need more support (e.g., augment) in order to re-engage.

(3) Retention: participant are les likely to drop out from a SMART because you catch them if they show early signs of failure. In the single stage they have no choice but to drop-out of they are not improving.

(4) Prescriptive information: although A is not so good initially, it provides information that can help you better tailor the treatment (e.g., adherence). It is possible that people who do

not adhere to A do very well on augment: they just need more support to engage – you will not be able to see this if you are only focusing on non-responders to B in Trial 2. So with single stage your ability to more deeply tailor treatment might be limited.

Reason # 1: Delayed Therapeutic Effects

Delayed effect:

- Short term effectiveness of initial treatment does not capture its long-term effectiveness when followed by subsequent treatment
  - i.e., when considered part of a sequence of treatments
  - Might happen when there are:

a. Positive synergies

b. Negative Synergies

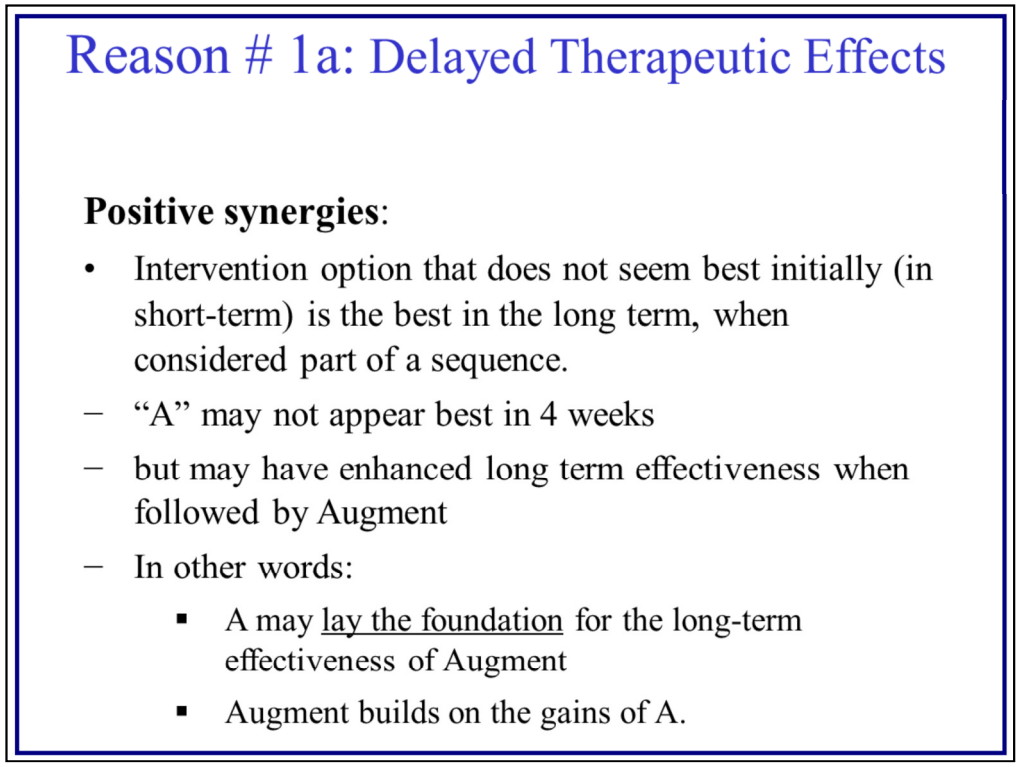
Delayed effects: it's a setting in which the effect that appears best initially (in the short-term) is not best when considered as part of a sequence.

A consequence is that comparing two initial therapies based on a proximal outcome may produce different results from the comparison of two initial therapies when followed by one of several maintenance therapies based on longer term outcomes.

Additionally, restricting comparisons to longer term outcomes, a comparison of two initial therapies followed by usual care or no therapy may yield different results from the comparison of two initial therapies when followed by one of several maintenance therapies.

We can expect that in an optimized AI, the best subsequent therapy will build on the gains achieved by prior therapies and thus these delayed effects should be common.

We want big positive delayed effects. We want profound positive cross-over effects!!!



This happens with behavioral interventions. Sometime it may take time for a behavioral intervention to work (for the approach to really sink) – so what we see is that there are no short-term gains. But then, when we add something to the intervention or provide a different context for the person to utilize skills, we see a huge gain. This is a very known concept in skill transfer (what you learn initially will sink only when you are exposed to a different context/setting, or a different type of intervention).

Reason # 1b: Delayed Therapeutic Effects

Negative synergies:

- Intervention option that appears best initially (in short-term) is not best in the long term, when considered as part of a sequence.
  - "B" may produce a higher proportion of responders
  - But, the burden imposed by B may be sufficiently high so that non-responders are less likely to adhere to subsequent treatments

A negative delayed effect would occur if the initial treatment overburdens an individual, resulting in decreased responsivity to future treatment; see Thall et al. (2007), Bembom and van der Laan (2007) for an example of the latter in cancer research.

Reason #2: Adherence/Drop-out

Subjects more likely to adhere/remain in SMART

- In the alternate trial of the initial treatments subjects are assigned a fixed treatment.
  - Those who are not improving have no other option besides non-adherence or drop-out.
- SMART, by design, provides alternates for non-improving subjects,
  - Enhancing motivation to adhere and remain in the study.

Consider the issue of adherence; in many historical trials subjects were assigned a fixed treatment, that is, there were no options besides non-adherence for subjects who were not improving. This often leads to higher than expected drop-out or non-adherence. This is particularly the case in longer studies where continuing treatments that are ineffective is likely associated with high non-adherence especially if the subject doesn't know if they are receiving treatment such as in a double blind study. As a result the subjects who remained in the historical trial may be quite different from the subjects that remain in a SMART trial, which by design provides alternates for non-improving subjects. David Oslin made this point.

Reason #3: Selection Effect

A Non-Responders study might not represent the population of non-responders

- In the alternate trial 2, I may choose to recruit non-responders to B and randomize to subsequent options.
  - Only non-responders who are highly motivated will select to participate in this study
  - Because of this selection bias I might not realize that I need to provide more support to encourage demoralized non-responders to start treatment again.

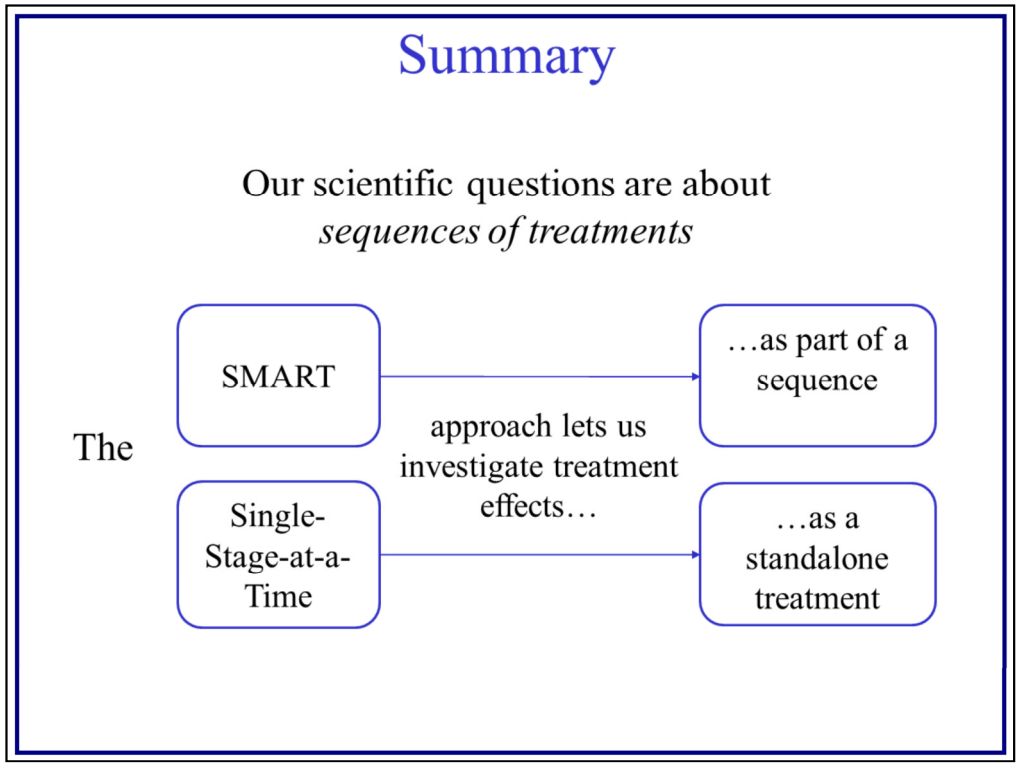
Consider the issue of motivation. Nonresponder trials recruit individuals who are not responding to their present treatment, say Med A. An important consideration is whether these nonresponders represent the population of individuals who do not respond to Med A or whether the nonresponders recruited into the trial are more motivated (because non-responders who gave up because the initial treatment did not work will not be motivated to enroll in another study). Such selection bias will prevent us from realizing that we might need a behavioral intervention to encourage nonresponders to start again with treatment.

Reason #4: Prescriptive Effects

Single stage provides limited options to explore ways to more deeply tailor the AI.

- Treatment A may not produce as high a proportion of responders as treatment B
- but treatment A may elicit symptoms (e.g., non-adherence) that allow you to better match the subsequent treatment to the patient and thus achieve improved response to the sequence of treatments as compared to initial treatment B.

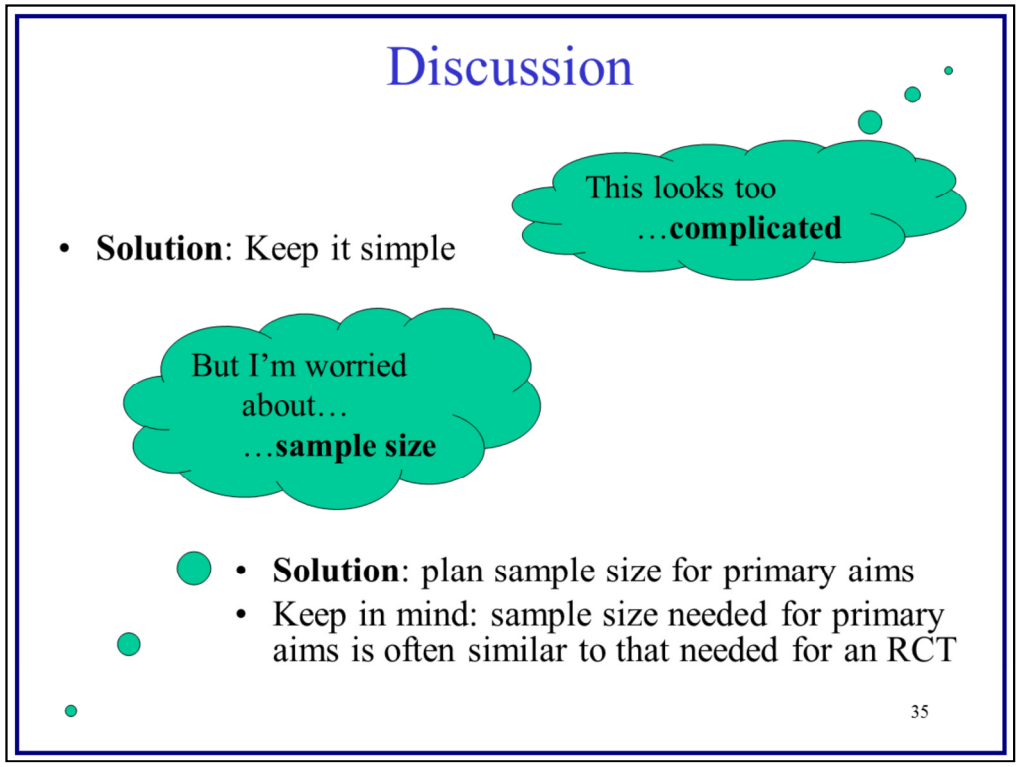
Consider the issue of motivation as expressed via adherence; if treatment A provides less social support than B, then patients who require the social support will exhibit adherence problems during A but not during B. This is useful information as we then know that these patients, even if they respond will potentially need an enhancement of social support during the maintenance or aftercare phase.



Using the single-stage approach is like reading the first half of a book without being able to know what will happen at the end.

Outline

- What are Sequential Multiple Assignment Randomized Trials (SMARTs)?
- Trial Design Principles and Analysis
- Advantages of a SMART
- **Summary & Discussion**



Keep it clear and simple:

1)Focus on a few important open scientific questions.

2)Order questions– primary and secondary.

3)Choose well-defined tailoring variable to restrict the randomization based on well-justified ethical, scientific and practical considerations.