Sequential, Multiple Assignment, Randomized Trials

Module 2

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

Billie Nahum-Shani, Daniel Almirall
Outline

- What are Sequential Multiple Assignment Trials (SMARTs)?
- Why SMART experimental designs?
  - “new” clinical trial design
- Trial Design Principles and Analysis
- Summary & Discussion
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.

*The goal is to inform the construction of adaptive Interventions*

In stat. people may call these multistage trials (the randomization at each stage is assumed)
Motivating Questions

• Hypothetical Aim: AI for treating minecraft addiction
• Insufficient empirical evidence/theories to determine
  (a) What is the best way to initiate treatment (A vs. B)?
  (b) How to modify treatment for early non-responders
      (switch vs. augment)
  (c) How to maintain minecraft abstinence among early
      responders (relapse prevention vs. monitoring)
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.
Other names are  dynamic treatment regimes, treatment algorithms, stepped care models, expert systems, adaptive treatment strategies, treatment protocols. Structured treatment interruptions in the treatment of AIDS are a form of adaptive txt strategy

Individualized interventions
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.

Or, 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 treatment strategy.
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

(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.
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).

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!!!
treatment of psychosis: a medication may result in many immediate responders but Some patients are not helped and/or experience abnormal movements of the voluntary muscles (TDs). The class of subsequent medications is greatly reduced.

Or the kind of response produced may not be sufficiently strong so that patients can take advantage of maintenance care.

A negative delayed effect would occur if the initial treatment overburdens an individual, resulting 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.
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 bind 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 to me.

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.
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 bind 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 to me.

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.
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.

**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.
Summary

- When evaluating and comparing initial treatments, *in a sequence of treatments*,
  
  - We need to take into account, e.g. control, the effects of the subsequent treatments
    
    ▪ hence the need for a SMART
  
  - Keep in mind: standard single-stage randomized trials may yield information about different populations from SMARTs.
Other names are dynamic treatment regimes, treatment algorithms, stepped care models, expert systems, adaptive treatment strategies, treatment protocols. Structured treatment interruptions in the treatment of AIDS are a form of adaptive strategy.

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

Equal randomization
Note we considered different options for the responders as compared to the nonresponses.

You can use an endless number of intermediate outcomes to restrict the class of options. But then the decision tree will be over complicated to justify and implement.

But it is important that you keep it simple: use a low dimensional summery (e.g., response status) and then specify how it is operationalized, namely how do 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.
Note we considered different txt’s for the responders as compared to the nonresponders.

In mental illness studies feasibility considerations may force us to use preference in this low dimensional summary.
SMART design Principals

- Choose primary hypotheses
  - that are both scientifically important and aids in developing the AI
  - Power trial to address these hypotheses

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

- **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
  - Trial is not necessarily powered to address these hypotheses.

Confounding: alternative explanations other than treatment effect for the observed difference
Example of Primary Aims

1. Comparison of initial options

- When sample size is highly constrained:
- **H1:** The initial intervention option A results in lower symptoms than the initial intervention option B.
  - Controlling for second-stage intervention options

Here, we are controlling for second-stage treatment by design.

These are main effects a la’ ANOVA

The second would be appropriate if you initially wanted to run a trial for non-responders and are now considering SMART

Example 1: Effects of secondary treatments are controlled by experimental design – not by statistical analysis

Example 2: Effects of first-stage treatments are controlled by experimental 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) or adherence as alternative explanation: people who do not adhere will be switched, so all switched people are non-adherent (for the second-stage)
Here, we are controlling for second-stage treatment by design.

These are main effects a la’ ANOVA
The second would be appropriate if you initially wanted to run a trial for non-responders and are now considering SMART

Example 1: Effects of secondary treatments are controlled by experimental design – not by statistical analysis

Example 2: Effects of first-stage treatments are controlled by experimental 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) or adherence as alternative explanation: people who do not adhere will be switch, so all switched people are non-adherent (for the second-stage)
Example of Primary Aims

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

- When sample size is less constrained:
- **H2:** Among non-responders a switch to C results in lower symptoms than augmenting with D
  - Controlling for first-stage intervention options

Here, we are controlling for second-stage treatment by design.

These are main effects a la’ ANOVA

The second would be appropriate if you initially wanted to run a trial for non-responders and are now considering SMART

Example 1: Effects of secondary treatments are controlled by experimental design – not by statistical analysis

Example 2: Effects of first-stage treatments are controlled by experimental 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) or adherence as alternative explanation: people who do not adhere will be switch, so all switched people are non-adherent (for the second-stage)
Here, we are controlling for second-stage treatment by design.

These are main effects a la’ ANOVA
The second would be appropriate if you initially wanted to run a trial for non-responders and are now considering SMART

Example 1: Effects of secondary treatments are controlled by experimental design – not by statistical analysis
Example 2: Effects of first-stage treatments are controlled by experimental 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) or adherence as alternative explanation: people who do not adhere will be switch, so all switched people are non-adherent (for the second-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.
A study of non-responders in which one controls the initial intervention option to which people don’t respond to.
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 treatment

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.

One of these AIs is indicated in red.
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.*

These are main effects a la’ ANOVA
Sigma for example 1 is the std of primary outcome of patients initially assigned to intervention option A (or B)

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

Throughout working assumptions are equal variances, normality and equal number in each of the two groups being compared.

** 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-responder, 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

For H1 with 0.3 effect size, if non-response rate=0.3, then N=1340; if non-response rate=0.5, then N=804; if non-response rate=0.6, then N=670
For the analysis you will need a weight and replicate approach which we will discuss later.

We have a sample size formula that specifies the sample size necessary to detect a standardized mean outcome difference (\( \delta \) standard deviations) between two AIs beginning with different intervention options.

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.

For the only NR randomized setting I assumed 0.3 NR rate – because only NRs are split into two groups – the sample size needed will be lower to the extent that NR rate is lower (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).

** 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.

We also have sample size formula that specify the sample size for comparing AIs in time-to-event studies.

Sample Size

- **Primary Aim that is less useful in developing an AI:**
  - Compare initial treatments (A vs. B) in terms of intermediate outcomes (proportion of early responders).
  - Why less useful?
    - Focus short-term effects—Reading the first half of a book
    - Not considering the implications of second-stage options
  - This makes sense only if all second-stage options are equally effective

This makes sense only if all second-stage options are equally effective
SMART Designing Principles

- Choose secondary hypotheses
  - That further develop the AI.
  - Example:
    
    **H4:** non-adhering non-responders will exhibit lower symptoms if their initial treatment is augmented with D as compared to switching to C (e.g. augment D includes MI).
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 Trials (SMARTs)?
- Why SMART experimental designs?
  - “new” clinical trial design
- Trial Design Principles and Analysis
- 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