Dynamic Treatment Regimes for Personalized Medicine

Personalized Medicine

A Quick View

- Personalized Medicine is a general medical paradigm referring to systematic use of individual patient information to optimize patient's health outcomes.

- It is important in treating chronic diseases including hypertension, obesity, diabetes, drug abuse, cancer, HIV infection, depression and schizophrenia.

- Growing interest from statistical community and other quantitative researchers is due to
  - advance of biotechnology to collect individual data (Volume, Variety, Velocity)
  - advance of computational algorithms and tools
  - advance of clinical designs and statistical methods

Definition of Personalized Medicine

- Personalized medicine refers to **"the tailoring of medical treatment to the individual characteristics of each patient".**

- Food and Drug Administration (FDA): characterization of heterogeneity of subject responses to treatment is a critical component of drug development and regulatory decision making.
- Pharmaceutical industry: targeting therapeutic intervention in a well-characterized subpopulation.
- Clinicians: use individual characteristics to guide optimal treatment and provide best clinical care for a patient.

Why Personalized Medicine?

- Heterogeneity across patients: what works for one may not work for another (e.g., MDD response rate 40%, Gaynes et al., 2009).
- Temporal variability within a patient: what works now may not work later (MDD relapse rate 50%, APA 2000).
- Different pathologies underlying a clinical syndrome.
- Multiple active treatments available (e.g., 6 main SSRIs for treating MDD).
- Potential heterogeneity of treatment delivery related to practical application (stage of disease, co-morbidity etc.)

Scientific Goals in Personalized Medicine

- To identify right patients for a given treatment.
- To identify right treatments for a given patient.
- Treatment is in a broad sense: type of drugs/interventions, dosage, schedules, treatment strategies and etc.

Methodology focus: To determine the optimal (dynamic) treatment regimen for a given patient using data-driven approaches (Evidence-Based Decision Making).

- Right data
- Right question
- Right method
- Right tool

Trial Evidence for Personalized Medicine

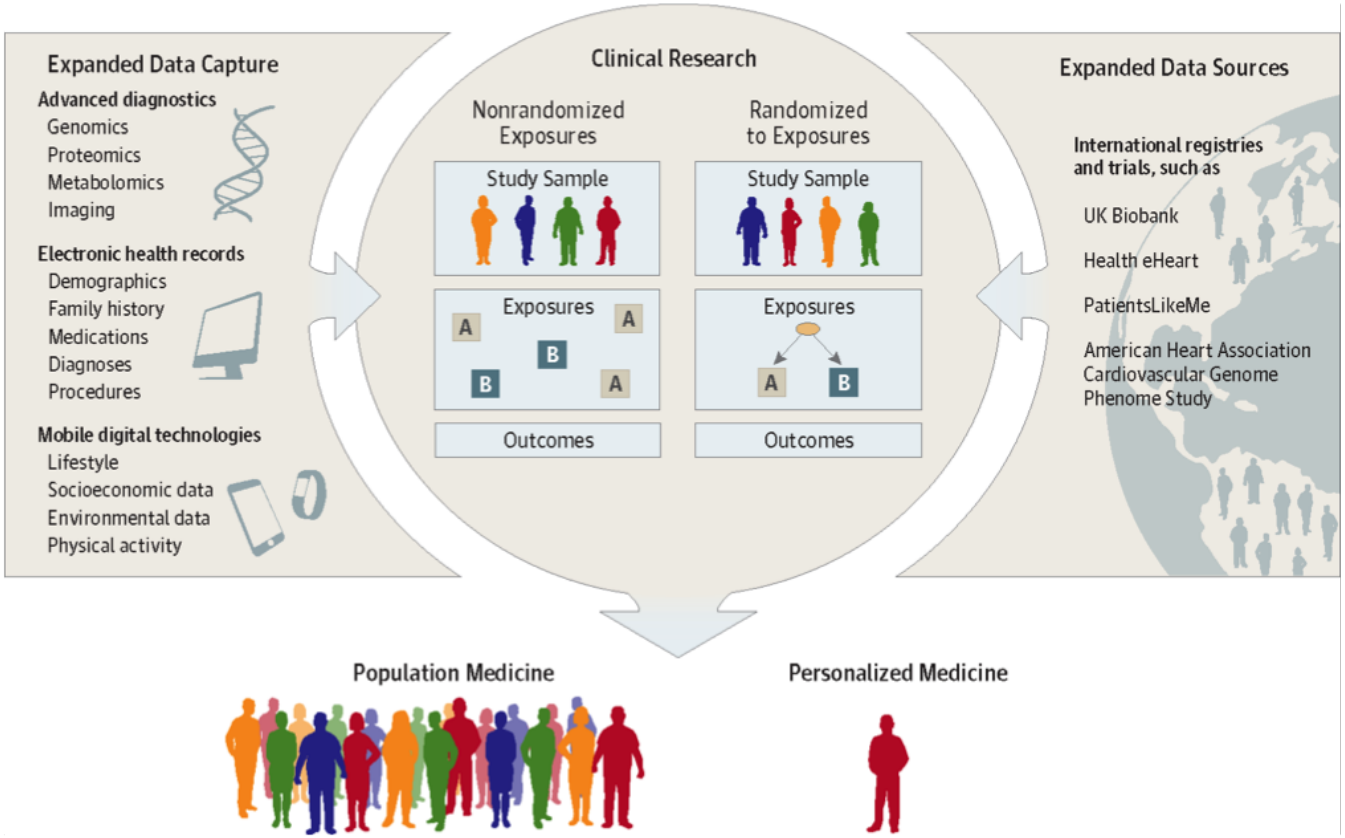
- Data from one or more clinical trials
- Data collected from sequential randomization trials
  - Adaptive Pharmacological and Behavioral treatments for ADHD (Pelham WE, 2002);
  - Sequenced Treatment Alternatives to Relieve Depression (STAR*D) (Rush, et al., 2004);
  - CATIE for schizophrenia (Schneider, et al., 2003);
  - ExTENd for alcohol dependence (Oslin, 2005);
  - Adaptive therapy for androgen independent prostate cancer (Thall et al. 2007)
- Trial evidence provides valid assessment of personalized medicine.

Challenges with Trial Evidence

- Trial data are limited in sample sizes so are usually lack of statistical power.
- Trials are not designed for personalized treatments .
- The generalizability to a broad population is questionable.

Observational Data for Personalized Medicine

- Rich data sources provide patient information: Demographics, Co-morbidity, Stage of disease, Genomic (oncology), Imaging (psychiatry), Mobile technology (neurology, psychiatry), Electronic Medical Records



Challenges with Observational Data

- Potential confounders can invalidate results.
- High-dimensional or dynamic tailoring variables
- Noisy and poor quality data
  (measurement errors, missing data)
- High redundancy within data
- Infinite choices of treatment rules
- Multivariate outcomes (benefit and risk outcomes)
- Big Data size

Decision Framework in Personalized Medicine

Medical Decision Making

- Personalized medicine (treatment) is essentially a medical decision making framework.
- Same as statistical learning, we need to address the following components:
  - What information are we going to make decision based on?
  - Where do we obtain such information for decision making?
  - What output is a decision going to give?
  - What is the loss if we make wrong decisions, or what is the value (utility) if we make correct ones?
  - How should we evaluate or validate the derived decision?
- Under this framework, it is not surprising why statistical learning can be useful for personalized medicine.

Single-stage decision making in personalized medicine

- In a single-stage decision, optimal treatment is given to an individual patient at one time.
- Decision is based on the static features of the patient, such as demographic variables, biomarkers measured at the time, genotypes, exposures and etc.
- The decision rule is a function from the features to one of the treatment options (discrete, continuous).
- The optimal decision (treatment) rule is expected to optimize a pre-defined value function or utility function.
- The evidence-based rule needs to be validated for significance in a confirmatory study.

Multi-stage decision making in personalized medicine

- In a multi-stage decision, sequential treatments are given to an individual patient at each stage (e.g., different lines of cancer therapy).
- At each stage, the decision not only is based on the static features, but may also vary with evolving features (tumor growth, side effect, cognitive decline).
- At each stage, the decision is a function mapping all collected features at this stage to one of the treatment options at this stage.
- The optimal decision is a sequence of decision functions (treatment regimens) and should maximize the total value in a long term, not an immediate outcome.

Infinite-stage decision making in personalized medicine

- This is similar to infinite-horizon problems in reinforcement learning.
- Decision is made over a large number of stages (often, real time) over the disease course of an individual patient.
- Each decision is based on the static features and time-sensitive features and the optimal decision rule aims to maximize the long-term positive outcome of an individual patient.
- This problem usually arises for long-term health management using mobile health.

Example 1. Treatment by Anticoagulation

- Anticoagulation is often given after stroke, pulmonary embolism or deep vein thrombosis.
- Treatment goal is to ensure that the patient's prothrombin time, measured by the international normalized ratio (INR), is within a safe range.
- Clinicians need to make decisions about increase or decrease the dose of anticoagulant during regular monitoring, based on patient's information collected over time.
- It is one multi-stage decision problem in personalized medicine.

Example 2. Treatment of Alcohol Addiction

- Clinical decisions are needed to manage alcohol dependent subjects.
- An an initial stage, they are given either an opiate-antagonist called naltrexone (NTX) or cognitive behavior therapy (CBT).
- In two months, each subject is assessed based on whether they have experienced a heavy-drinking day. If not, they are called responders or non-responders, otherwise.
- For a non-responder with the initial NTX treatment, clinicians decide whether to either switch to CBT or augment NTX with CBT and an enhanced motivational program.
- For a non-responder with the initial CBT treatment, decision is whether to either switch to NTX or augment CBT with NTX and the enhanced program.
- The goal is to maximize the presence of days abstinent over 1 12-month period.

Example 3: Treatment of Cancer

- Cancer patients are often treated with the first-line induction therapy (chemotherapy) for disease remission.
- The therapy is maintained if the patient responds; otherwise, the patient will be treated by the second line therapy.
- Decisions to be made include which first-line therapy and second-line therapy, when to initiate therapies and what dose to use.
- The goal is to maximize the patient's cancer survival.
- This is a 2-stage decision problem.

Potential Outcome Framework

Potential Outcome Framework in Personalized Medicine

For one treatment stage with only two treatment options $\{-1, 1\}$ and one single outcome R,

- Potential outcome $Y(a)$ refers to the outcome if treated by $a \in \{1, -1\}$.
- Personalized medicine is really interested in whether $Y(1)$ is larger than $Y(-1)$ and the size of $Y(1) - Y(-1)$.

For two treatment stages and one single outcome R,

- Potential outcome $Y(a_1, a_2)$ refers the outcome if treated by a_1 in the first stage then a_2 in the second stage.
- Personalized medicine is interested in the comparisons among $Y(a_1, a_2)$.

Impossibility of Personalized Medicine

In general, consider any treatment sequence $(a_1, a_2, ...)$. We let $Y(a_1, a_2, ...)$ be the corresponding potential outcome then personalize medicine aims to understand the difference among these potential outcomes using evidence.

- Only one possible sequence of treatments is observed for a given individual so in theory, this is an impossible task!
- One solution is to use other individuals to impute unobserved potential outcomes; however, imputation can be very questionable due to inevitable difference between individuals.

Practical Compromise: Approximation to Personalized Medicine

- Instead of seeking individual effects, we can aim for the average effects from homogeneous individuals.
- Average treatment effects:

$$E[Y(1)] - E[Y(-1)], \quad E[Y(a_1, a_2)] - E[Y(a_1', a_2')].$$

- Feature-specific average treatment effects (subgroup analysis):

$$E[Y(1) - Y(-1)|X], \quad E[Y(a_1, a_2) - Y(a_1', a_2')|X]$$

  where X is a pre-defined feature variable.
- Are they only comparators? The answer is NO!

Limitation of Traditional Treatment Comparison

- We can compare

$$E[Y(\text{sign}(X))] - E[Y(-\text{sign}(X))]$$

- Or,

$$E[Y(D_1(X))] - E[Y(D_2(X))],$$

  where D_1 and D_2 are both maps from X's domain to $\{-1, 1\}$.
- This can of interest in public health in terms of health policy decision.

From Treatment Options to Treatment Strategies

- Traditional comparison of average treatment effects is on fixed treatment options, i.e, what if we apply treatment option a_1 or a_2 to everyone in the study population.
- Essentially, we are comparing two treatment strategies:
  Strategy 1: Everyone is treated by a_1.
  Strategy 2: Everyone is treated by a_2.
- This is called one-size-fits-all strategies.
- Similarly, the subgroup comparison is comparing two fixed treatment strategies within the subgroup.
- However, personalized medicine is interested in even broader treatment strategies!

Broader Definitions of Treatment Strategies

- A broader treatment definition: treatment is no longer understood as some constant options but as a strategy which maps individual feature/characteristics to the domain of treatment options, i.e.,

$$\mathcal{D}: \text{ domain of } X \to \{-1, 1\}.$$

- Constant treatment options (one-size-fits-all strategies) are the special case where $\mathcal{D}(x)$ is a constant.
- An example of $\mathcal{D}$:
  "if patient's age is older than 50 or his/her BMI is higher than 26, we should use drug 1; otherwise, we should use drug 2."
- Hence, personalized medicine is interested in evaluation, comparison and selection of treatment strategies.

Remarks on Treatment Strategies

- Treatment strategies are also called treatment regime, treatment policies or treatment decisions.
- Conclusion: Nothing changes dramatically or crazily, but we need a broader view.

Treatment Strategies Can Be Dynamic: Examples

- Example 1. A patient with a particular cancer is given therapy 1 first. If his/her disease progresses, we immediately change to a more aggressive therapy 2; otherwise, we maintain the first therapy.

- Example 2. A patient with a particular cancer is given therapy 1 first. If his/her disease progresses, we wait to see his/her immune responses. Depending on the responses, we decide to change to a more aggressive therapy 2 if the responses are low, otherwise, maintain the early treatment.

Dynamic Treatment Strategies

- The average potential outcomes for the examples cannot be quantified in terms of constant treatment options, since the treatment choices depends the intermediate outcomes for different individuals.

- However, both are two different treatment strategies of practical interest and they are dynamic.

- They are called dynamic treatment strategies, dynamic treatment regime, dynamic treatment policies, or adaptive treatment strategies.

Random Treatment Strategies

- Even more broader treatment definition is to allow $\mathcal{D}$ to map into a probability distribution in the domain of treatment option, i.e., we allow random treatment strategies.
- Random treatment strategies are useful when new evidence continues to come in so randomness allows to explore better strategies.
- Random treatment strategies are useful for mobile health and artificial intelligence.
- However, they are not good for practice.
- We only focus on non-random treatment strategies in this course.

Goals and Challenges

Goals

- Evaluate a specific treatment strategy.
- Compare different treatment strategies.
- Select the optimal treatment strategies.
- Implement a specific treatment strategy.
- Validate a specific treatment strategy.

Challenges

- What designs can provide unbiased evidence?
- What assumptions ensure validity of the results?
- What methods meet the goals?
- What practical concerns remain about the results?

Definition of DTRs

- Contrast to static treatment that comes out of traditional clinical trials: once and for all treatment
- DTRs are sequential treatment decision rules, tailored at each stage by patients' time-varying features and intermediate outcomes in previous stages (Lavori & Dawson 1998, Lavori et al. 2000, Murphy et al. 2001).
- DTRs are different from subgroup treatment rules.

Why DTRs

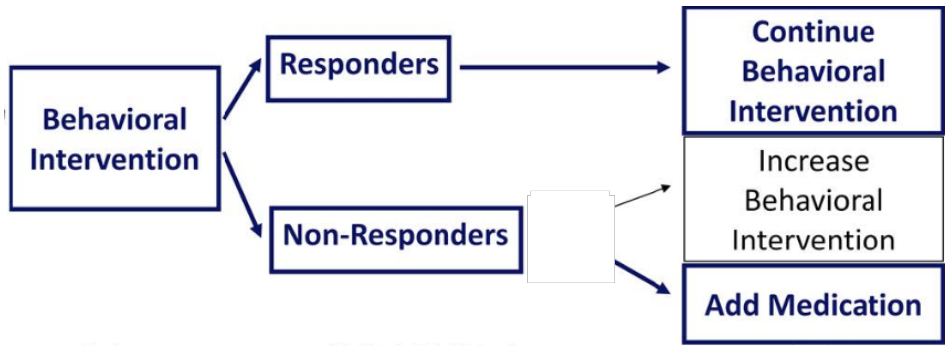
- Reflect clinical practice
- Patients respond heterogeneously to treatments
- Effect changes over time
- Comorbidity conditions, relapses and side effects
- High cost of intensive interventions (potential burden/side effects motivate intensity to be reduced when possible)
- Improve adherence rate

DTRs are often used in chronic diseases including hypertension, obesity, diabetes, mental illness, HIV infection, cancer and alcohl and drug abuse.

Example 1 of DTRs

Adaptive Pharmacological Behavioral Treatments for Children with Attention Deficit Hyperactive Disorder (ADHD, Pelham 2002).

- DTR1: Prescribe medication (MED) as initial treatment; if a child responds then continue; if a child does not respond then augment with behavioral modification (BMOD).
- DTR2: Prescribe BMOD as initial treatment; if a child responds then continue; if a child does not respond then augment with MED.



Example 2 of DTRs

- Cancer patients are initially treated with induction therapy:
  - if they show signs of remission, they are given maintenance therapy to intensify or augment the first line treatment;
  - if not, a second line therapy is prescribed to induce remission.
- Multiple line therapies for individual patients are clearly a dynamic treatment decision.

Components of DTRs

- Defined disease stages without ambiguity for treatment decisions
  (Stage/phase/critical decision points)
- Individual dynamic characteristics for treatment decisions
  (State/feature/covariate space)
- Feasible treatment options at each stage
  (Treatment space/domain)
- Treatment decisions
  (Decision functions/treatment rules/policies)

Formal DTR Definition

- Consider K stages (finite horizon)
- H_k: state/covariate at stage k, including previous treatment history and treatment responses
- $\mathcal{A}_k$: treatment domain at stage k
- $\mathcal{D}_k$: a map from H_k to $\mathcal{A}_k$

DTR consists of the sequence of decision functions $\mathcal{D} = (\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_K)$. Note that the value of $\mathcal{D}_k$ is part of the state space in $\mathcal{D}_{k+1}$.

Value Function

Evaluating DTRs

- To evaluate DTRs, we need to define reward/utility outcome, $R = \sum_{k=1}^{K} Y_k$, where Y_k is the intermediate reward at stage k after treatment:
  - It should be most relevant to scientific goals: to maximize cancer patient's survival or to main international normalized ratio (INR) within a target range within anticoagulation treatment.
  - Sometimes it can be defined with general consensus: survival time, distance from the target range.
  - Often it is user-defined and is subjective, especially when both benefit and risk are in consideration.
- The most important quantity to evaluate DTRs, $\mathcal{D}$, is Value function:

$$\mathcal{V}(\mathcal{D}) = E^{\mathcal{D}}[Y] = E[Y(\mathcal{D})],$$

  where $Y(\mathcal{D})$ is the potential reward outcome for the DTR $\mathcal{D}$.
- The value is the expected reward if the DTR implemented.

Comparing DTRs

- Given two DTRs: $\mathcal{D}_1$ and $\mathcal{D}_2$, we prefer $\mathcal{D}_1$ to $\mathcal{D}_2$ if

$$\mathcal{V}(\mathcal{D}_1) - \mathcal{V}(\mathcal{D}_2) > 0.$$

- When $\mathcal{D}_2$ takes the same treatment values no matter what covariates are at each stage, $\mathcal{D}_2$ is equivalent to once and for all treatment. Therefore, comparing $\mathcal{D}_1$ and $\mathcal{D}_2$ can support/reject the use of once and for all treatment.

Optimal DTRs

- The optimal DTRs: $\mathcal{D}^* = \max_{\mathcal{D}} \mathcal{V}(\mathcal{D})$.
- Finding the optimal DTRs is the ultimate goal of personalized medicine discovery.
- Sometimes, the optimal DTRs searching is restricted to a subset of DTRs (simple decision rules or practically feasible treatment rules).
- Related to the optimal DTRs, two quantities are also used in this context:
  Regret/loss: $\mathcal{L}(\mathcal{D}) = \mathcal{V}(\mathcal{D}^*) - \mathcal{V}(\mathcal{D})$
  Blip: $\mathcal{C}(\mathcal{D}) = \mathcal{V}(\mathcal{D}) - \mathcal{V}(\mathcal{D}_0)$ where $\mathcal{D}_0$ is some benchmark DTR.

Bellman Equations

Stage-Specific Value Function

- Given state variable at stage k, we can define state value function at stage k as

$$\mathcal{V}_k(\mathcal{D}; H_k) = E^{\mathcal{D}}[\sum_{j=k}^{K} Y_j | H_k],$$

  i.e., the expected reward increment if the future treatments follow the DTR, $\mathcal{D}$.

- Bellman equation

$$\mathcal{V}_k(\mathcal{D}; H_k) = E_{A_k=\mathcal{D}_k(H_k)} \left[Y_k + \mathcal{V}_{k+1}(\mathcal{D}; H_{k+1}) \middle| H_k \right].$$

- It shows a iterative way to compute each stage-specific value function BACKWARDS!

Stage-Specific Qaulity Function

- Given state variable at stage k and if treated by $A_k = a_k$, we can define treatment-state value function, also called quality (Q-)function at stage k as

$$Q_k(\mathcal{D}; H_k, a_k) = E^{\mathcal{D}}[\sum_{j=k}^{K} Y_j | A_k = a_k, H_k],$$

  i.e., the expected reward increment if treated as a_k at stage k but the future treatments follow the DTR, $\mathcal{D}$.

- Bellman equation

$$Q_k(\mathcal{D}; H_k, a_k) = Y_k +$$

$$E_{A_{k+1}=\mathcal{D}_{k+1}(H_{k+1})} \left[Q_{k+1}(\mathcal{D}; H_{k+1}, A_{k+1}) \middle| A_k = a_k, H_k \right].$$

- It shows a iterative way to compute each stage-specific quality function BACKWARDS!

Iterative Computation Formulae for Optimal Value Function

For the optimal DTR, $\mathcal{D}^* = (\mathcal{D}_1^*, \cdots, \mathcal{D}_K^*)$, it holds

- $$\mathcal{D}_k^*(H_k) = \text{argmax}_{a_k} E_{A_k=a_k}\left[Y_k + \mathcal{V}_{k+1}(\mathcal{D}^*; H_{k+1}) \middle| H_k\right].$$
- $$\mathcal{D}_k^*(H_k) = \text{argmax}_{a_k} Q_k(\mathcal{D}^*; H_k, a_k).$$
- These relationships are the key to many methods/algorithms for estimating the optimal DTRs.

Sequential Multiple Assignment Randomized Trials (SMARTs) for Evaluating DTRs

Introduction

- Consider simple DTRs: $\mathcal{D} = (\mathcal{D}_1, ..., \mathcal{D}_K)$

$$\mathcal{D}_k(H_k) = 1 \text{ or } -1 \ (\mathcal{A}_k = \{-1, 1\}).$$

That is, a fixed treatment is assigned to any patients at stage k.

- How shall we evaluate $\mathcal{V}(\mathcal{D})$?
Conduct a single-arm randomized trial.

- How shall we compare two and more different DTRs: $\mathcal{D}_1$ vs $\mathcal{D}_2$ (not necessarily simple ones)?
Conduct a multiple-arm randomized trial.

Randomization Is Key

- Why randomization is needed?
  - It provides representative coverage of the patient population.
  - It controls all potential confounders which may bias the effects of comparing treatments.
- One essential mathematical relationship links observed outcome to potential outcome:

$$\begin{aligned} E[Y(d)|H] &= E[Y(d)|A=d,H] \text{ by randomization} \\ &= E[Y|A=d,H] \text{ by consistency assumption.} \end{aligned}$$

Consistency/SUTVA: $R = Y(d)$ if $A = d$ and is not affected by the particular treatment assignments to the others.

SMART

SMART: Sequential Multiple Assignment Randomized Trial (Lavori & Dawson 2000, 2004; Murphy 2005)

- Patients are sequentially randomized at each critical decision stage.

- Randomization probability may depend on current states of patients.

- Practical SMART
  – Adaptive Pharmacological and Behavioral treatments for ADHD;
  – Sequenced Treatment Alternatives to Relieve Depression (STAR*D) ;
  – CATIE for schizophrenia;
  – ExTENd for alcohol dependence;
  – Adaptive therapy for androgen independent prostate cancer

Example: Two-Stage SMART Study

- The study (Kasari et al., 2014) was designed to study communication intervention for minimally verbal children with autism.
- The study aimed to test the effect of SGD, each stage lasting 12 weeks.
- SGD: speech-generating device; (JASP+EMT): blended developmental/behavioral intervention
- The second stage had another 12 week follow-up.
- The study started with 61 eligible children and 46 completed both stages.

Diagram of The Autism Study

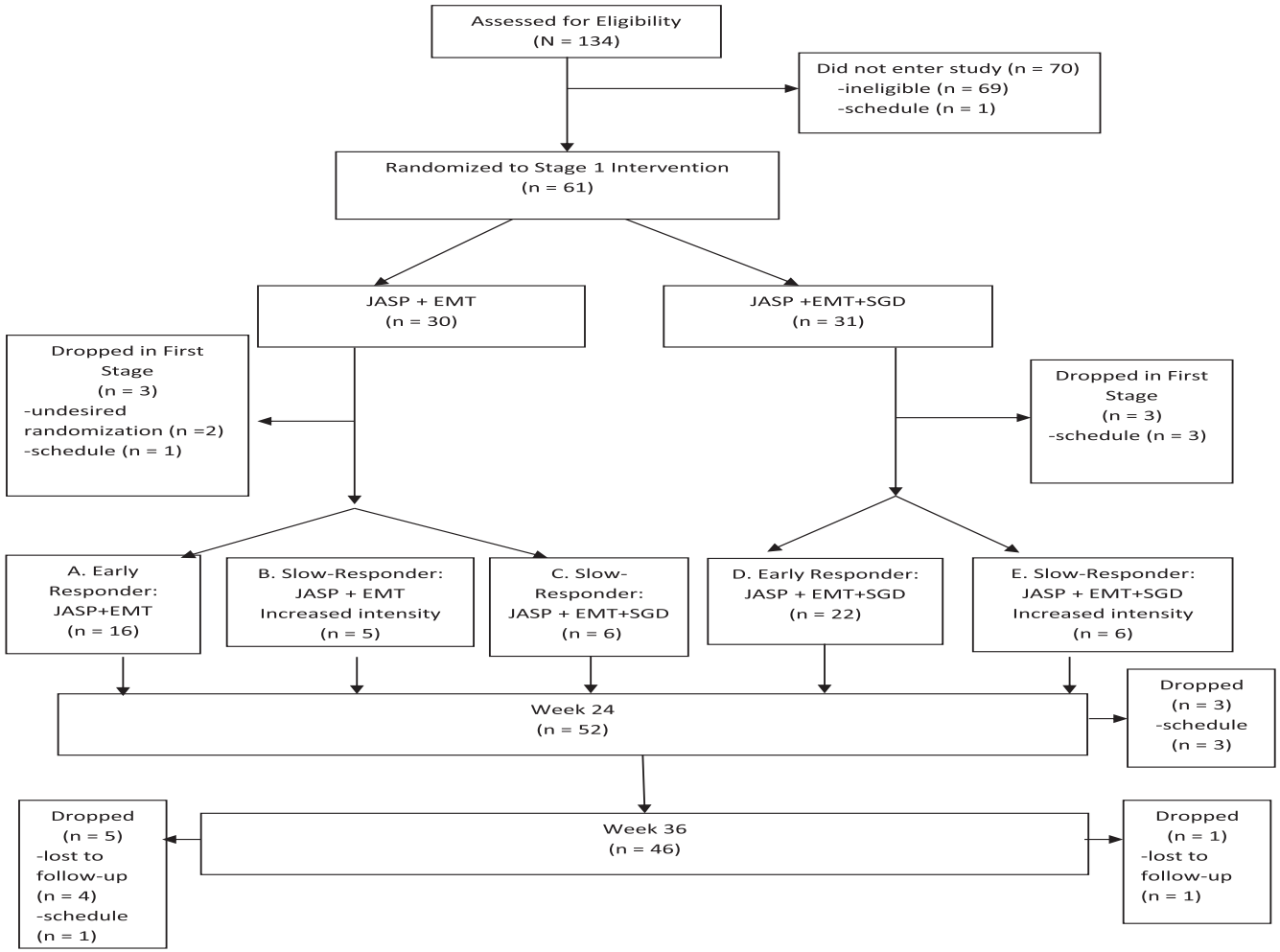


Figure: SMART Design of Autism Study (Kasari et al. 2014)

A list of SMART studies

- CALGB Protocol 8923 for treating elderly patients with leukemia
- CATIE trial for antipsychotic medication in Alzheimer's disease patients
- STAR*D trial for treating depression
- Smoking cessation study by the University of Michigan
- A trial of neurobehavioral treatments for patients with metastatic malignant melanoma
- SMART for the Adaptive Characterizing Cognition in Nonverbal Individuals with Autism (CCNIDA)
- Adaptive Pharmacological and Behavioral Treatments for children with attention deficit hyperactivity disorder (ADHD)
- Adaptive Reinforcement Based on treatment for pregnant drug abusers (RBT)
- the ExTENDd study for alcohol-dependent individuals

Advantages of SMART

Research questions to be answered from a SMART

- **Main effects of treatments**
  - What is the better initial treatment, JASP+EMT or JASP+EMT+SGD?
  - What about the slow-responders: intensify or not?
- **Effects of embedded DTR**
  - JASP+EMT→intensify vs JASP+EMT+SGD→ intensify vs JASP+EM→JASP+EMT+SGD
- **Exploring optimal treatment strategy (deep tailoring)**
  - intensify or not in the second stage dependent on additional intermediate outcomes?

General Advantages

- Valid comparisons of different treatment options at different stages due to the virtue of randomization.
- Discover adaptive treatment strategies that are embedded in the SMART trial.
- Inform the development of adaptive and deeply tailored treatments (using potentially high-dimensional biomarkers).

Comparison with single-stage randomized trials

- DTRs might be synthesized by pooling results from randomized trials independently conducted at each stage.
- There are some potential issues.
  - it fails to account for delayed effects of treatment so the resulting DTRs are not optimal.
  - SMART allows treatment decisions at later stages to be tailored to post-earlier stage outcomes (intermediate outcomes), which cannot be done using pooled single-stage trials.
  - Subjects enrolled into different single-stage trials may be different (cohort or selection effect).

Unbiased Value Estimation under SMART

Recall the value function associated with $\mathcal{D} = (\mathcal{D}_1, ..., \mathcal{D}_K)$

$$\mathcal{V}(\mathcal{D}) = E[Y(\mathcal{D})] = E[Y(\mathcal{D}_1, ..., \mathcal{D}_K)].$$

Let $\pi_k(a, h)$ be the randomization probability $P(A_k = a | H_k = h)$.

- Start from stage K:

$$\begin{aligned}
& E[Y(\mathcal{D}_1, ..., \mathcal{D}_{k-1}, \mathcal{D}_K)] \\
= \; & E\left\{E\left[Y(\mathcal{D}_1, ..., \mathcal{D}_K) \middle| H_K, A_K = \mathcal{D}_K(H_k)\right]\right\} \\
= \; & E\left\{E\left[Y(\mathcal{D}_1, ..., \mathcal{D}_{k-1}, A_K) \frac{I(A_K = \mathcal{D}_K(H_K))}{\pi_K(A_K, H_K)} \middle| H_K\right]\right\} \\
= \; & E\left[Y(\mathcal{D}_1, ..., \mathcal{D}_{k-1}, A_K) \frac{I(A_K = \mathcal{D}_K(H_K))}{\pi_K(A_K, H_K)}\right].
\end{aligned}$$

Continue to Prior Stages

- Now at stage $K-1$, we continue and repeat the same derivation to obtain

$$E[Y(\mathcal{D}_1, ..., \mathcal{D}_{k-1}, \mathcal{D}_K)]$$

$$= E\left[Y(\mathcal{D}_1, ..., A_{K-1}, A_K)\frac{I(A_{K-1} = \mathcal{D}_{K-1}(H_{K-1}), A_K = \mathcal{D}_K(H_K))}{\pi_{K-1}(A_{K-1}, H_{K-1})\pi_K(A_K, H_K)}\right].$$

- Continue backwards till stage 1:

$$\mathcal{V}(\mathcal{D}) = E\left[Y(A_1, ..., A_{K-1}, A_K)\frac{I(A_1 = \mathcal{D}_1(H_1), \cdots, A_K = \mathcal{D}_K(H_K))}{\prod_{k=1}^K \pi_k(A_k, H_k)}\right]$$

$$= E\left[Y\frac{I(A_1 = \mathcal{D}_1(H_1), \cdots, A_K = \mathcal{D}_K(H_K))}{\prod_{k=1}^K \pi_k(A_k, H_k)}\right].$$

- The value function for $\mathcal{D}$ can be estimated using the average reward outcomes from the patients whose treatments follow $\mathcal{D}$, weighted by their randomization probabilities.

An Alternative Interpretation

- In SMART, a particular treatment assignment takes the probability

$$\prod_{k=1}^{K} \pi_k(A_k, H_k).$$

- If we run a single-arm trial for a given $\mathcal{D}$, the treatment assignment is $\prod_{k=1}^{K} I(A_k = \mathcal{D}_k(H_k))$.
- To use SMART to estimate the expected reward in the $\mathcal{D}$-specified trial, important sampling theory gives

$$\mathcal{V}(\mathcal{D}) = E\left[Y \frac{I(A_1 = \mathcal{D}_1(H_1), \cdots, A_K = \mathcal{D}_K(H_K))}{\prod_{k=1}^{K} \pi_k(A_k, H_k)}\right].$$

- In conclusion, SMART provides unbiased estimation and comparisons between DTRs so potentially leads to valid estimation of optimal DTRs.

Apply missing data methods for evaluating or comparing DTRs

- Likelihood method $\equiv$ G-computation
- Semiparametric method $\equiv$ G-estimation (MSM, Nested structural models)
- However, they are not convenient to learn optimal DTRs (especially nonparametric DTRs).
- Machine learning methods are useful for the latter.