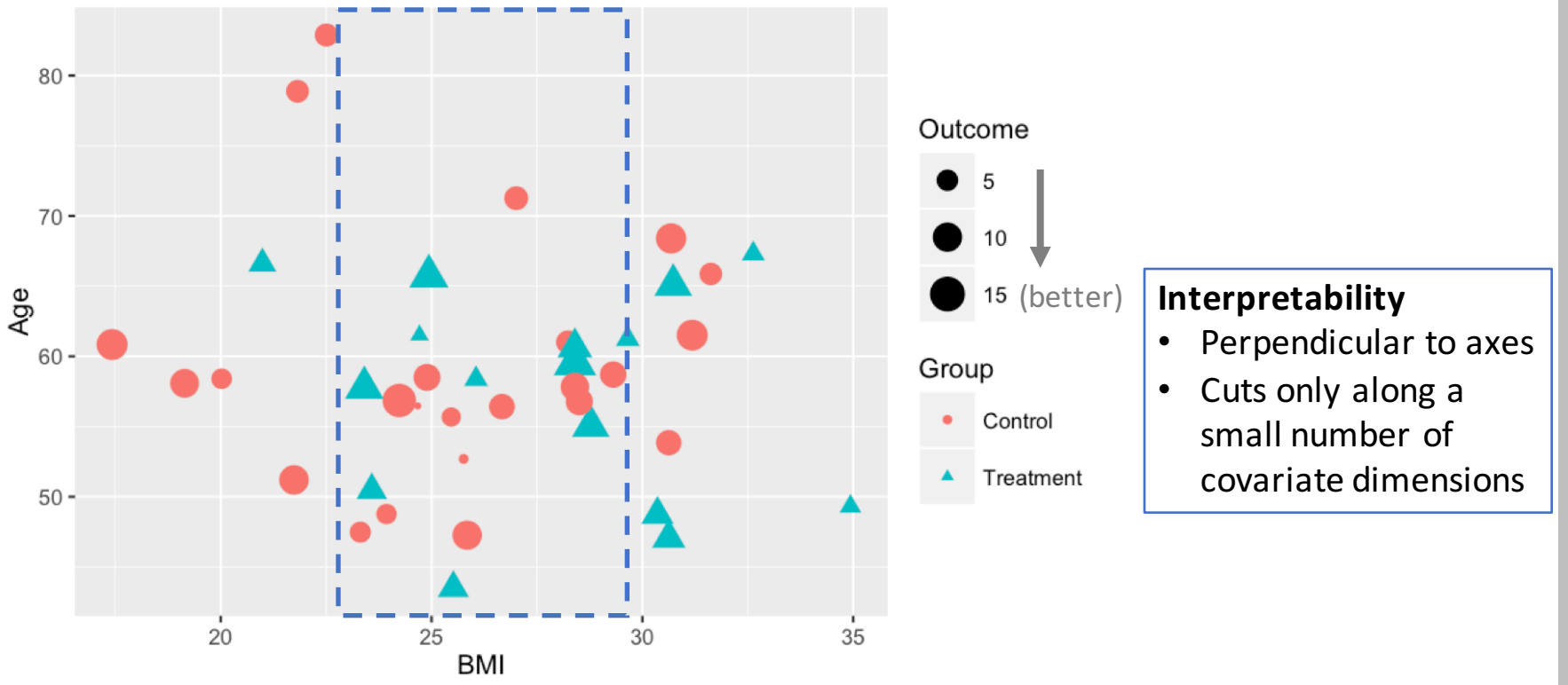
# Lecture 16: Identifying Exceptional Responders in Randomized Trials

## 15.095 Machine Learning via a Modern Optimization Lens

*Find interpretable subset with highest (or lowest) average treatment effect.*

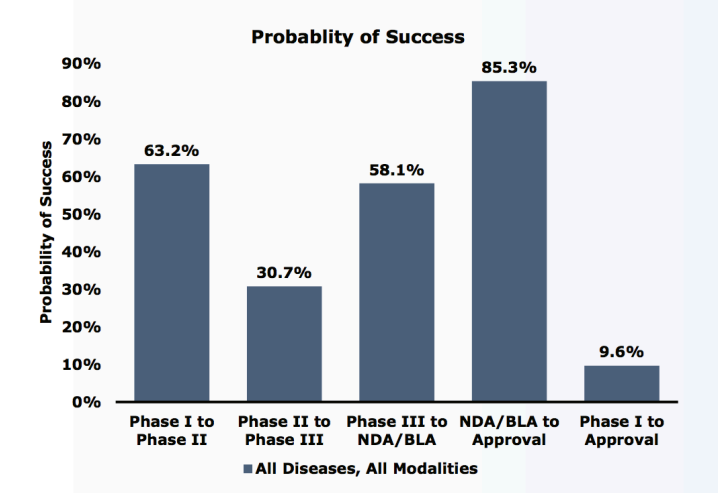



**Interpretability**

- Perpendicular to axes
- Cuts only along a small number of covariate dimensions

# Practical Impact & Opportunity

- Revisit use of failed drugs in promising subgroups.
- Avoid further testing in high-risk subgroups due to adverse events.
- Re-market drugs whose patents are expiring.




Source: Clinical Development Success Rates 2006-2015, BIO. June 2016.

# Identifying Exceptional Responders

| | |
|---|---|
| **Cox proportional hazards with treatment-covariate interaction** | Schemper, 1988. |
| **Probabilistic approaches** | Kehl & Ulm, 2006; Foster et al., 2011. |
| **Recursive partitioning** | Su et al., 2009; Hardin et al., 2013. |

With ***mixed-integer optimization*** (MIO) modeling, we can:

- Find optimal, interpretable solutions directly.
- Solve a single global formulation.
- No need for recursion or iteration.

# Identifying Exceptional Responders

1. Algorithm: Modeling & Validation

   **Contributions**: Efficient formulation despite fractional objective; significance testing approach for validation.

2. Computational Results

   **Contributions**: Effective on real & simulated data.

3. Extensions

   **Contributions**: Fast heuristic; find multiple subsets.

# Methodological challenges

1. **How can we identify the interpretable subset with highest ATE?**

2. **Once we identify a subset, how do we know if it is truly exceptional?**
   Every trial has a *best* subset. Not every trial has an *interesting* best subset.

2. **Once we identify a subset, how do we know if it is truly exceptional?**
   Every trial has a *best* subset. Not every trial has an *interesting* best subset.

   - Common approach: Split data into training and validation sets.
     - Problem: Not enough data in subset.

   - Our approach: Test for "robustness" of ATE in optimal subset.
     - Non-parametric hypothesis test based on the bootstrap.
       - Efron and Tibshirani, 1994.
     - Test statistic: **trimmed ATE** in optimal subset.
       - Trim 10% highest and lowest treatment responses.

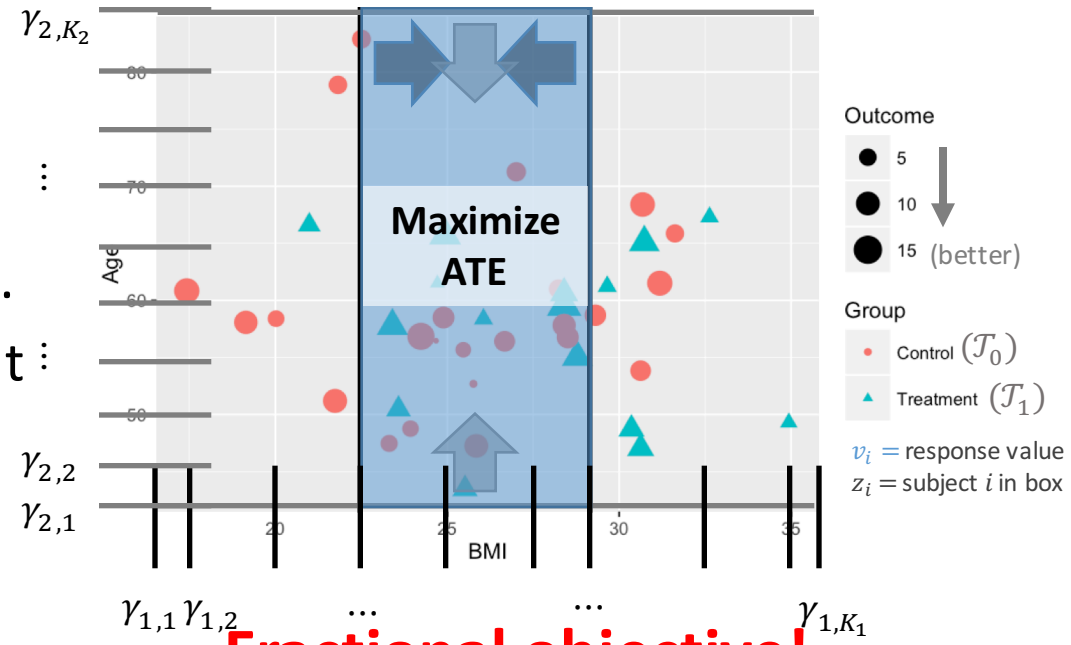
## 1. How can we identify the interpretable subset with highest ATE?

*Use integer optimization.*

**Modeling approach:**

- Choose a lower and upper bound for the box along each dimension $s = 1, \ldots, S$, from a set of "cut-points" $\gamma_{s1}, \ldots, \gamma_{sK_s}$.
- To maximize average treatment effect for points in the box:

$$\max_{\boldsymbol{z},\boldsymbol{q},\boldsymbol{L},\boldsymbol{U}} \frac{\sum_{i\in\mathcal{T}_1} v_i z_i}{\sum_{i\in\mathcal{T}_1} z_i} - \frac{\sum_{i\in\mathcal{T}_0} v_i z_i}{\sum_{i\in\mathcal{T}_0} z_i}$$




**Fractional objective!**

1. **How can we identify the interpretable subset with highest ATE?**

   *Use integer optimization.*

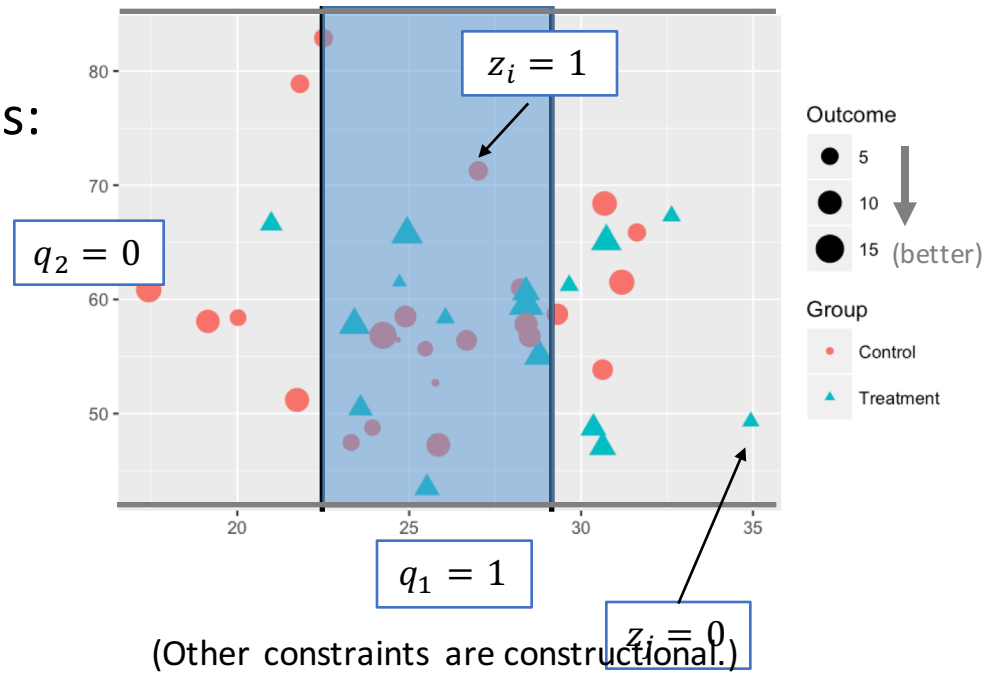
**Constraints (Practical)**

- At most $S_0$ restricted dimensions:

$$\sum_{s=1}^{S} q_s \leq S_0$$

- Restrict cardinality of box from each treatment group $\mathcal{T}_t$:

$$\underline{N} \leq \sum_{i \in \mathcal{T}_t} z_i \leq \overline{N}, \qquad \forall t = 0, 1$$




(Other constraints are constructional.)

**Non-fractional objective!**
**But still nonlinear…**

Transform fractional objective

$$\frac{\sum_{i\in\mathcal{T}_1} v_i z_i}{\sum_{i\in\mathcal{T}_1} z_i} - \frac{\sum_{i\in\mathcal{T}_0} v_i z_i}{\sum_{i\in\mathcal{T}_0} z_i}$$

- Define auxiliary dummy variables, $\theta_j^{(t)},\ j = \underline{N}, \dots, \bar{N},\ t = 0,1$, such that:

$$\theta_j^{(t)} = 1 \Leftrightarrow \sum_{i\in T_t} z_i = j$$

- Then, by construction,

$$\sum_{j=\underline{N}}^{\bar{N}} \frac{1}{j}\theta_j^{(t)} = \frac{1}{\sum_{i\in T_t} z_i}, \qquad t = 0,1$$

- Objective function becomes:

$$\sum_{i\in T_1}\sum_{j=\underline{N}}^{\bar{N}} \frac{1}{j} v_i z_i \theta_j^{(1)} - \sum_{i\in T_0}\sum_{j=\underline{N}}^{\bar{N}} \frac{1}{j} v_i z_i \theta_j^{(0)}$$

## Linear objective!

Transform nonlinear objective

$$\sum_{i\in\mathcal{T}_1}\sum_{j=\underline{N}}^{\overline{N}}\frac{1}{j}v_i z_i\theta_j^{(1)} - \sum_{i\in\mathcal{T}_0}\sum_{j=\underline{N}}^{\overline{N}}\frac{1}{j}v_i z_i\theta_j^{(0)}$$

- Introduce auxiliary variables:

$$\zeta_{ij} \coloneqq z_i \cdot \theta_j^{(T_i)},$$
$$i = 1,\dots,n, j = \underline{N},\dots,\overline{N}$$

- Objective becomes **linear** in $\zeta_{ij}$:

$$\sum_{i\in T_1}\sum_{j=\underline{N}}^{\overline{N}}\frac{1}{j}v_i\zeta_{ij} - \sum_{i\in T_0}\sum_{j=\underline{N}}^{\overline{N}}\frac{1}{j}v_i\zeta_{ij}$$

- Add constraints to model $\zeta_{ij} = z_i \cdot \theta_j^{(T_i)}$:

$$\zeta_{ij} \le \theta_j^{(T_i)}, \qquad \forall i = 1,\dots,n,\ j = \underline{N},\dots,\overline{N},$$
$$\zeta_{ij} \le z_i, \qquad \forall i = 1,\dots,n,\ j = \underline{N},\dots,\overline{N},$$
$$\zeta_{ij} \ge \theta_j^{(T_i)} + z_i - 1, \qquad \forall i = 1,\dots,n,\ j = \underline{N},\dots,\overline{N},$$
$$\sum_{i\in\mathcal{T}_t}\sum_{j=\underline{N}}^{\overline{N}}\frac{1}{j}\zeta_{ij} = 1, \qquad \sum_{j=\underline{N}}^{\overline{N}}\theta_j^{(t)} = 1, \qquad \forall t = 0,1$$

# Transformed MIO formulation

- Number of variables: $O(n^2)$.
  - Can include $\zeta$ as continuous variables on [0,1].
  - Leaves $O(n)$ binary variables.
- Time to provable optimality with $S = 10$:
  - $n = 100$: 33s
  - $n = 1{,}000$: 4hrs

$$
\begin{aligned}
\max_{\mathbf{z},\mathbf{q},\mathbf{L},\mathbf{U},\boldsymbol{\zeta},\boldsymbol{\theta}} \quad & \sum_{i\in\mathcal{T}_1}\sum_{j=\underline{N}}^{\overline{N}} \frac{1}{j} v_i \zeta_{ij} - \sum_{i\in\mathcal{T}_0}\sum_{j=\underline{N}}^{\overline{N}} \frac{1}{j} v_i \zeta_{ij} \\
\text{s.t.} \quad & z_i + \sum_{s=1}^{S}\left[\sum_{k:\gamma_{sk}>x_{is}} L_{sk} + \sum_{k:\gamma_{sk}<x_{is}} U_{sk}\right] \geq 1, && \forall i = 1,\dots,n, \\
& z_i + L_{sk} \leq 1, && \forall s = 1,\dots,S,\ k = 1,\dots,K_s,\ i : x_{is} < \gamma_{sk}, \\
& z_i + U_{sk} \leq 1, && \forall s = 1,\dots,S,\ k = 1,\dots,K_s,\ i : x_{is} > \gamma_{sk}, \\
& \sum_{k=1}^{K_s} L_{sk} = 1, && \forall s = 1,\dots,S, \\
& \sum_{k=1}^{K_s} U_{sk} = 1, && \forall s = 1,\dots,S, \\
& q_s + L_{s1} \geq 1, && \forall s = 1,\dots,S, \\
& q_s + U_{sK_s} \geq 1, && \forall s = 1,\dots,S, \\
& q_s + L_{s1} + U_{sK_s} \leq 2, && \forall s = 1,\dots,S, \\
& \sum_{s=1}^{S} q_s \leq S_0, \\
& \underline{N} \leq \sum_{i\in\mathcal{T}_t} z_i \leq \overline{N}, && \forall t = 0,1, \\
& \zeta_{ij} \leq \theta_j^{(T_i)}, && \forall i = 1,\dots,n,\ j = \underline{N},\dots,\overline{N}, \\
& \zeta_{ij} \leq z_i, && \forall i = 1,\dots,n,\ j = \underline{N},\dots,\overline{N}, \\
& \zeta_{ij} \geq \theta_j^{(T_i)} + z_i - 1, && \forall i = 1,\dots,n,\ j = \underline{N},\dots,\overline{N}, \\
& \sum_{i\in\mathcal{T}_t}\sum_{j=\underline{N}}^{\overline{N}} \frac{1}{j}\zeta_{ij} = 1, && \forall t = 0,1, \\
& \sum_{j=\underline{N}}^{\overline{N}} \theta_j^{(t)} = 1, && \forall t = 0,1, \\
& 0 \leq \zeta_{ij} \leq 1, && \forall i = 1,\dots,n,\ j = \underline{N},\dots,\overline{N}, \\
& \mathbf{z},\mathbf{q},\mathbf{L},\mathbf{U},\boldsymbol{\theta} \in \{0,1\}.
\end{aligned}
$$

**Primary Decision Variables**
$L_{sk}, U_{sk}$ - whether cut-point $k$ chosen as *lower/upper* bound, respectively, for dimension $s$.

# Identifying Exceptional Responders

1. Algorithm: Modeling & Validation

   **Contributions**: Efficient formulation despite fractional objective; significance testing approach for validation.

2. Computational Results

   **Contributions**: Effective on real & simulated data.

3. Extensions

   **Contributions**: Fast heuristic; find multiple subsets.

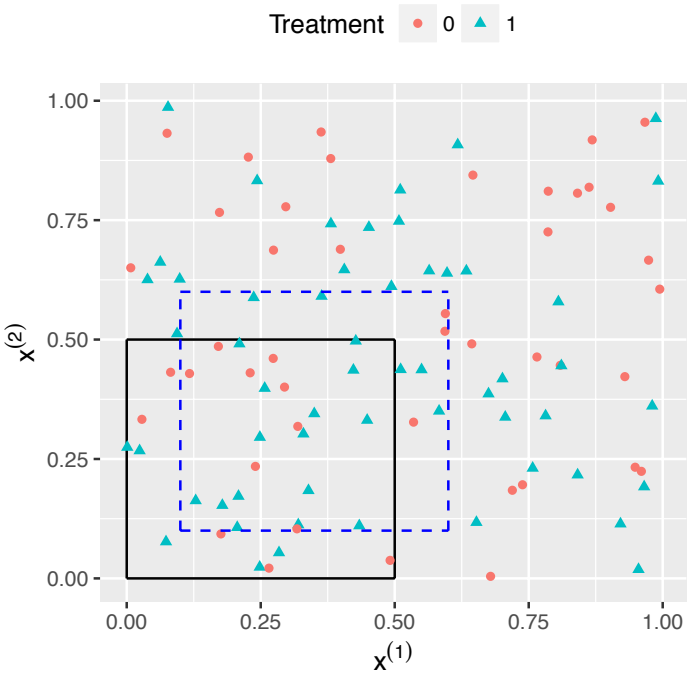
# Simulation experiments

- *Base case*: $n = 100,\ S = 4,\ S_0 = 2,$ $\delta_0 = 2,\ \ \varepsilon_i \sim N(0,1).$
- 250 unique random samples.
  - Found valid subset: 76.8%.
  - True positive rate (PPV threshold): 68.0%.
  - True positive rate (PPV-weighted): 60.7%.
- False positive rate: 12.4%.

$y_i$ = indicator subject $i$ in treatment group $\mathcal{T}_1$.

*Response model:*

$$v_i = 2 + \delta_0 \cdot y_i \cdot \mathbb{I}\{x_i^{(1)} \leq 0.5\} \cdot \mathbb{I}\{x_i^{(2)} \leq 0.5\} + \varepsilon_i.$$



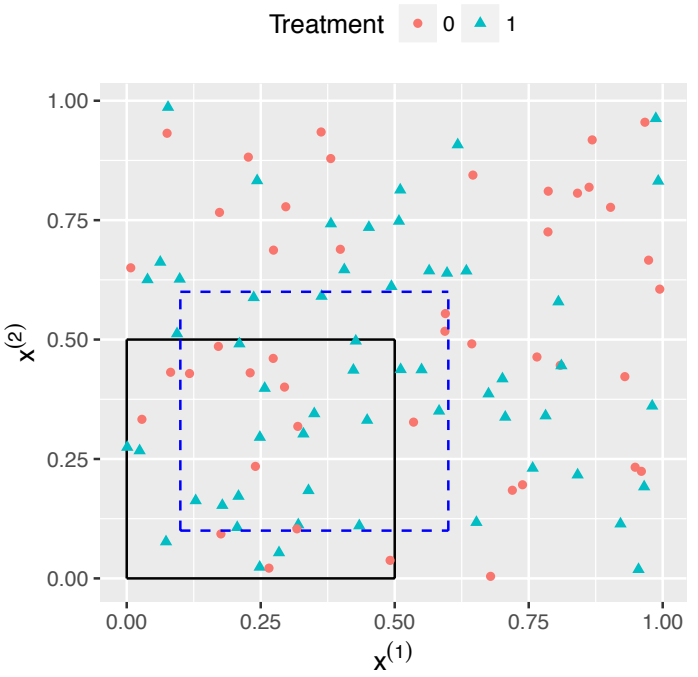
# Sensitivity analysis

- True positive rate increases with:
  - Sample size $n$
  - Added effect size in box $\delta_0$
- False positive rate increases slightly with:
  - Covariate dimension $S$
  - Box dimensional parameter $S_0$

$y_i$ = indicator subject $i$ in treatment group $\mathcal{T}_1$.

*Response model:*

$$v_i = 2 + \delta_0 \cdot y_i \cdot \mathbb{I}\{x_i^{(1)} \leq 0.5\} \cdot \mathbb{I}\{x_i^{(2)} \leq 0.5\} + \varepsilon_i.$$



# Real-world example:

## RCT of estrogen treatment for late-stage prostate cancer[1]

| | 5.0 mg dose of diethylstilbestrol | Placebo | Total/ Difference |
|---|---|---|---|
| $n$ | 125 | 127 | 252 |
| **Avg. survival** (months) | 35.0 | 35.3 | -0.3 |

OVERALL SAMPLE

[1]Byar and Green (1979). Data available at http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets.

*Optimal box*: Stage 4 cancer, no hist. of cardiovascular disease, & diastolic blood pressure $\geq 70$ mmHg.

| | | | |
|---|---|---|---|
| $n$ | 30 | 29 | 59 |
| **Avg. survival** (months) | 42.5 | 24.3 | +18.2* |

OPTIMAL BOX

* Significant at p=0.001

# Extensions

**Fast heuristic**

- Greedy, tree-like.
- Optimal or near-optimal solutions in seconds.
- True positive rate: 66.4%
  - vs. 68.0% for MIO.
- Provides good warm-start solutions for MIO.

# Extensions

**Multiple subsets**

- Re-solve optimization with constraint on intersection between points in box.
- Use Holm procedure to control for multiple significance.
- High true positive rate in simulations.

# Contributions:

## Identifying Exceptional Responders via MIO

- **MIO modeling**: Transform fractional objective to efficient formulation.
- **Validation**: Significance testing approach for validity of solutions.
- **Practical impact**: Extract added value from randomized trials.
- **Extensions:** Fast heuristic; ability to find multiple subsets.