Diagnosis-Group-Specific Transitional Care Program Recommendations for Thirty-Day Rehospitalization Reduction

Jared D. Huling

Division of Biostatistics
School of Public Health
University of Minnesota

Joint work with

Chensheng Kuang, Menggang Yu, and Maureen Smith

ENAR, 2021

Complex Patients: High Need and High Cost

- Early unplanned hospital readmission is common and costly, particularly among elderly and high-risk patients.
- One in five Medicare beneficiaries is readmitted within 30 days at a cost of more than $26 billion per year (Betancourt et al, 2015, Jencks et al, 2009), with *avoidable* readmissions estimated to cost as much as $17 billion per year (Rau, 2014)
- Under the Hospital Readmission Reduction Program, incentives in place to reduce avoidable readmissions
- A variety of **health system interventions** have been designed to help not just the problem of readmissions but to help provide better, more coordinated care for complex patients

Transitional Care Interventions

Transitional Care: a broad range of services and environments designed to promote the safe and timely passage of patients between levels of health care and across different care settings

- Multidisciplinary care team
- Structured patient support
  - Patient followup, education, or self-care management/training (or combinations thereof) after discharge using telephone technology
- Telemonitoring

Heterogeneity in who benefits from TC

- *On average*, TC interventions often do not work in improving readmission rates
- As with many health system interventions, *one size does not fit all*...interventions attempt to target specific problems in the delivery of care which may affect only some types of patients
- Hospital readmission can depend on many factors: hospital environment, policy environment, social determinants, patient lifestyle, characteristics, and more
- **Goal**: identify what drives who benefits from TC and then use this to help decide who is enrolled based on who benefits the most

Heterogeneity of the TC intervention

- Patients are only enrolled in TC upon a hospitalization
  - There is enormous heterogeneity in why people end up hospitalized
- We categorize patients into two large groups: medically complicated and medically uncomplicated
- Complicated and uncomplicated patients by and large receive very different health care and have very different needs and risks
- Patients in these two groups end up in TC for differing reasons *and* TC can address different problems for medically "complicated" than for "uncomplicated" patients
- Yet, some elements of TC remain the same or similar across these groups

Medically Complicated versus Uncomplicated

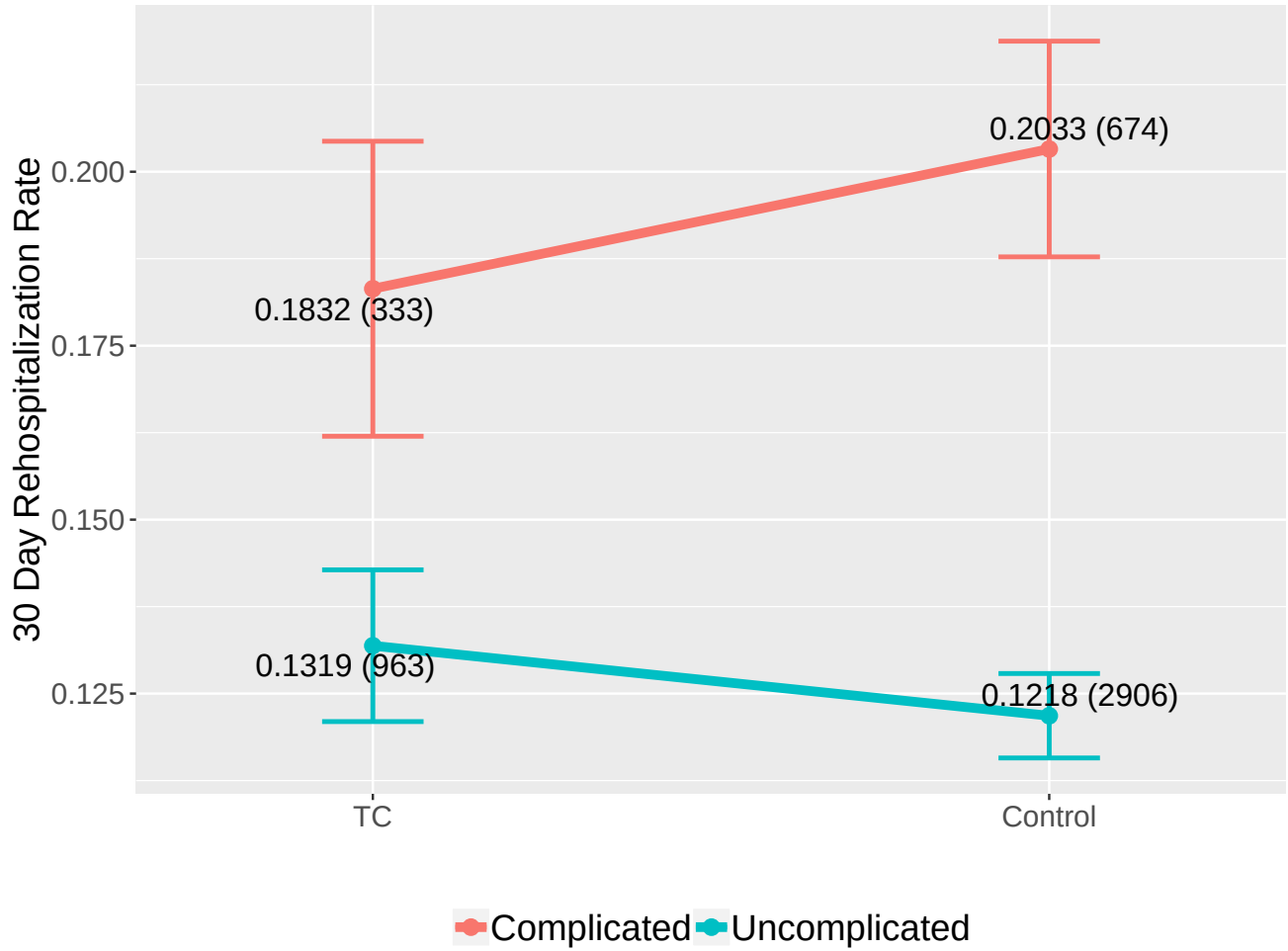
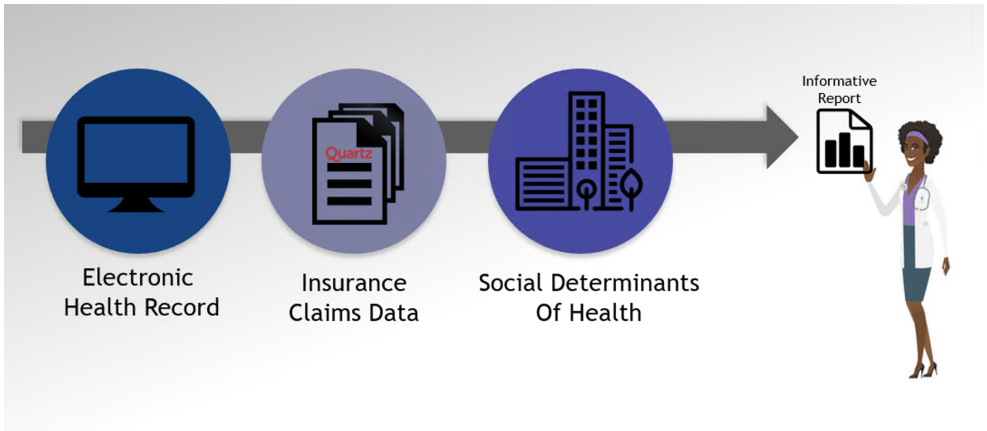


Figure: Unadjusted outcomes by intervention group (TC vs control) for medically complicated vs uncomplicated patients.

Data Points

- More than 20,000 data points available for each patient
- We often focus on a subset of these for modeling (still in the hundreds)



Examples of Data Points

- Congestive heart failure
- A1C values over the prior year
- Chronic liver disease
- Anxiety
- Kidney disease without failure
- Inpatient stays
- Incontinence, falls, dementia
- Prescription orders

Problems and Challenges for the TC Intervention

Problems:

1. TC has heterogeneous effects and does not work on average

Problems and Challenges for the TC Intervention

Problems:

1. TC has heterogeneous effects and does not work on average
   - $\rightarrow$ Estimate optimal mapping from patient characteristics to enrollment decisions (ITRs)

Problems and Challenges for the TC Intervention

Problems:

1. TC has heterogeneous effects and does not work on average
   → Estimate optimal mapping from patient characteristics to enrollment decisions (ITRs)
2. Health system populations are highly heterogeneous. Complicated vs uncomplicated patients are different and TC is itself different for these groups

Problems and Challenges for the TC Intervention

Problems:

1. TC has heterogeneous effects and does not work on average
   → Estimate optimal mapping from patient characteristics to enrollment decisions (ITRs)
2. Health system populations are highly heterogeneous. Complicated vs uncomplicated patients are different and TC is itself different for these groups
   → Allow estimated ITRs to be different for different patient groups (complicated vs uncomplicated)

Problems and Challenges for the TC Intervention

Problems:

1. TC has heterogeneous effects and does not work on average
   → Estimate optimal mapping from patient characteristics to enrollment decisions (ITRs)
2. Health system populations are highly heterogeneous. Complicated vs uncomplicated patients are different and TC is itself different for these groups
   → Allow estimated ITRs to be different for different patient groups (complicated vs uncomplicated)
3. ITR estimation is highly challenging, especially in high dimensions.

Problems and Challenges for the TC Intervention

Problems:

1. TC has heterogeneous effects and does not work on average
   - → Estimate optimal mapping from patient characteristics to enrollment decisions (ITRs)
2. Health system populations are highly heterogeneous. Complicated vs uncomplicated patients are different and TC is itself different for these groups
   - → Allow estimated ITRs to be different for different patient groups (complicated vs uncomplicated)
3. ITR estimation is highly challenging, especially in high dimensions.
   - There is some similarity/overlap across medically complicated/uncomplicated patients

Problems and Challenges for the TC Intervention

Problems:

1. TC has heterogeneous effects and does not work on average
   → Estimate optimal mapping from patient characteristics to enrollment decisions (ITRs)
2. Health system populations are highly heterogeneous. Complicated vs uncomplicated patients are different and TC is itself different for these groups
   → Allow estimated ITRs to be different for different patient groups (complicated vs uncomplicated)
3. ITR estimation is highly challenging, especially in high dimensions.
   - There is some similarity/overlap across medically complicated/uncomplicated patients
   → Leverage these similarities in a data-driven manner based on penalization techniques

Notation

- First consider single study setting:
  - p covariates $x_1, \ldots, x_p$
  - treatment $A \in \{0, 1\}$
  - response Y, potential outcomes under $A = 0, 1$: $Y(0), Y(1)$
- Let $g(X)$ be a treatment rule that maps X to $\{0, 1\}$.
- Our aim is to find $g(X)$ such that

$$g(X) = \arg\max_g E\{Y(g)\}$$

where $E\{Y(g)\}$ is the expectation of response under treatment rule g.

Weighted Contrast Classification Framework

- Denote $C(X) \triangleq E(Y|A=1,X) - E(Y|A=0,X)$ (called contrast function). We can write

$$E\{Y(g)\} = E_X[g(X)C(X) + E(Y|A=0,X)]$$

- Hence, we only need to find $g(X) = \arg\max_g E_X[g(X)C(X)]$.

- It can be further shown that

$$g(X)C(X) = \mathbb{1}\{C(X) > 0\}|C(X)| - |C(X)|\left[\mathbb{1}\{C(X) > 0\} - g(X)\right]^2$$

- Hence, the original problem is now transformed into a weighted classification problem.

$$g^{\text{opt}} = \arg\min_{g} E\left\{|C(X)|[\mathbb{1}\{C(X) > 0\} - g(X)]^2\right\}$$

- $C(X)$ is unknown. By estimating $C(X)$, the problem becomes

$$\hat{g}^{opt}(X) = \arg\min_{g} E\left\{|\hat{C}(X)|[\mathbb{1}\{\hat{C}(X) > 0\} - g(X)]^2\right\} \quad (1)$$

- We use whether $\hat{g}^{opt}(X) > 0.5$ to decide whether to assign treatment or not since $\hat{g}^{opt}(X) > 0.5 \iff C(X) > 0$

From *Zhang, et al (2012)*

Augmented Inverse Probability Weighted Estimator

- $C(X)$ can be estimated by an inverse probability weighted estimator (IPWE):

$$\hat{C}_{\text{IPWE}}(X) = \frac{AY}{P} - \frac{(1-A)Y}{1-P}$$

where P is the probability of assigning treatment (Here we assume treatment is assigned randomly for simplicity).

- Alternatively, for any function $a(X)$, one can work with $\tilde{Y} \equiv Y - a(X)$.

$$\begin{aligned}\hat{C}^a_{IPWE}(X) &= \frac{A\tilde{Y}}{P} - \frac{(1-A)\tilde{Y}}{1-P} \\ &= \hat{C}_{IPWE}(X) - a(X)\frac{A}{P} + a(X)\frac{1-A}{1-P}\end{aligned}$$

Hence, the estimator $\hat{C}^a_{IPWE}(X)$ is still unbiased.

Augmented Inverse Probability Weighted Estimator

- $\hat{C}^a_{IPWE}(X)$ is called augmented inverse probability weighted estimator (AIPWE).
- The optimal $a(X)$ minimizing the variance of $\hat{C}^a_{IPWE}(X)$ is

$$a_{opt}(X) = (1-P)E(Y|X, A=1) + PE(Y|X, A=0)$$

- Notice here $a(X)$ is only to increase estimation efficiency, can be a rough (even misspecified) estimate.
- Robins, Rotnitzky, and Zhao (1994); Robins (1999); Scharfstein, Rotnitzky, and Robins (1999)

Multiple study problem

- With multiple studies/subpopulations (DRGs in our setting), we borrow strength across different studies for subgroup identification by the following.
- $$\min_{g_1,\ldots,g_K} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \frac{1}{W_k} L(\hat{C}_{ik}, g_k(x_{ik})) + h(g_1, \ldots, g_k)$$

  where
  - g_k is the classifier for study K
  - $x_{ik} = (x_{i1k}, \ldots, x_{ipk})$.
  - L is a loss function measuring how well the functions $g_1, \ldots, g_K$ fits the data
  - h is a penalty function that induces sparsity and similarity between $g_1, \ldots, g_K$.
  - the weighting W_k may be needed to "standardize" each term in some way to take account for different scales in $\hat{C}_k$.

Our formulation

- We use linear model for $g_1, \ldots, g_K$ and use the same weighted loss function in (1) as g, that is,

$$L(\hat{C}_{ik}, g_k(x_{ik})) = |\hat{C}_k(x_{ik})|\{\mathbb{1}\{S_{ik} - g_k(x_{ik})\}^2$$

where

$$g_k(x_{ik}) = \beta_{0k} + \beta_{1k}x_{i1k} + \cdots + \beta_{pk}x_{ipk}$$

and

$$S_{ik} = \mathbb{1}\{\hat{C}_k(x_{ik}) > 0\}.$$

Our formulation

- For h, **ideally** we would propose to use a combination of group lasso penalty, lasso penalty and fused lasso penalty.

$$h(g_1,\ldots,g_K) = \lambda_1 \sum_{j=1}^{p} \sqrt{\beta_{j1}^2 + \cdots + \beta_{jK}^2} + \lambda_2 \sum_{j=1}^{p} \sum_{k=1}^{K} |\beta_{jk}| + \lambda_3 \sum_{j=1}^{p} \sum_{1 \le a < b \le K} |\beta_{ja} - \beta_{jb}|$$

Our formulation

- We view the coefficients for the same covariate X_j in different studies/DRGs as a group.
- All three terms are important:
  - group penalty is needed for group-wise selection of a covariate across studies/DRGs
  - lasso penalty is needed for covariate-wise selection within each group
  - fused lasso penalty is needed to encourage similarity between different studies.
- **Problem:** far too computationally challenging for applications using large/high dimensional EHR data

Our formulation

- We first decompose the study/DRG-specific contributions to the treatment rule as

$$\beta_{jk} = \mu_j + \delta_{jk}$$

  - Using the penalty $|\delta_{jk}|$ encourages the effect of the kth study for the jth covariate to be similar to the common effect μ_j – akin to a fused lasso penalty, i.e. $|\mu_j - \beta_{jk}|$
  - Using a group lasso penalty $\sqrt{\mu_j^2 + \sum_{k=1}^{K} \delta_{jk}^2}$ encourages all effects of the jth to be selected or removed simultaneously
  - Adding $|\mu_j|$ completes the set of selection possibilities

Our formulation

- Recall the effect decomposition of the study-specific contributions to the treatment rule:

$$\beta_{jk} = \mu_j + \delta_{jk}$$

$$h(g_1, \ldots, g_K) = (1-\alpha)\lambda_1 \sqrt{K} \left\{ \sum_{j=1}^{p} \|(\mu_k, \boldsymbol{\tau} \odot \boldsymbol{\delta}_{j\cdot})\|_2 \right\} + \alpha\lambda_1 \left\{ \|\boldsymbol{\mu}\|_1 + \sum_{k=1}^{K} \tau_j \|\boldsymbol{\delta}_{\cdot k}\|_1 \right\},$$

Here $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)$, $\boldsymbol{\delta}_{j\cdot} = (\delta_{j1}, \ldots, \delta_{jK})^T$, $\boldsymbol{\delta}_{\cdot k} = (\delta_{1k}, \ldots, \delta_{pk})^T$, and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_K)^T$ modify the penalty on study-specific terms.

Computation via data transformation

- The problem can be reformulated into a least squares problem with various penalties and solved with standard sparse group lasso software (in the same vein as Ollier and Viallon (2017))

- We can construct a matrix $\widetilde{\mathbf{X}}$ and a working response vector $\widetilde{\mathbf{S}}$ to pass to existing software such as `SGL` to optimize our proposed criterion

Computation via data transformation

Define the transformed design matrix as

$$\widetilde{\mathbf{X}} = \begin{pmatrix} \check{\mathbf{X}}^1 & \check{\mathbf{X}}^1/\tau_1 & 0 & \dots & 0 \\ \check{\mathbf{X}}^2 & 0 & \check{\mathbf{X}}^2/\tau_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \check{\mathbf{X}}^q & 0 & \dots & 0 & \check{\mathbf{X}}^q/\tau_q, \end{pmatrix}$$

where $\check{\mathbf{X}}^j$ is the a standardized design matrix for group j with ith row:

$$\check{\boldsymbol{x}}_i^j \equiv \sqrt{\frac{|\widehat{C}_i^j|}{W^j}} \left\{ \boldsymbol{x}_i^j - \frac{\sum_{i=1}^{n_j} |\widehat{C}_i^j| \boldsymbol{x}_i^j}{\sum_{i=1}^{n_j} |\widehat{C}_i^j|} \right\}$$

Computation via data transformation

Define the transformed working response matrix as $\widetilde{\mathbf{S}} = (\check{\mathbf{S}}^{1\,T}, \ldots, \check{\mathbf{S}}^{q\,T})^T$, where $\check{\mathbf{S}}^j = (\check{S}^j_1, \ldots, \check{S}^j_{n_j})$ with

$$\check{S}^j_i \equiv \sqrt{\frac{|\widehat{C}^j_i|}{W^j}} \left\{ S^j_i - \frac{\sum_{i=1}^{n_j} |\widehat{C}^j_i| S^j_i}{\sum_{i=1}^{n_j} |\widehat{C}^j_i|} \right\},$$

Analysis of Transitional Care Data

- The analysis data set had 3869 medically uncomplicated subjects and 1007 medically complicated subjects.
- 301 covariates (subset of a much larger list) screened for use in estimation
- Propensity score models to adjust for confounding fit within each patient group

Performance evaluation for ITRs

- ITRs estimated on 75% of data (training) and treatment effects conditional on estimated treatment assignments evaluated on remaining 25% (process repeated 100 times)
- For any estimated ITR $\hat{g}$, we evaluated different methods using the following statistic evaluated on the test data, for $a, b \in \{0, 1\}$

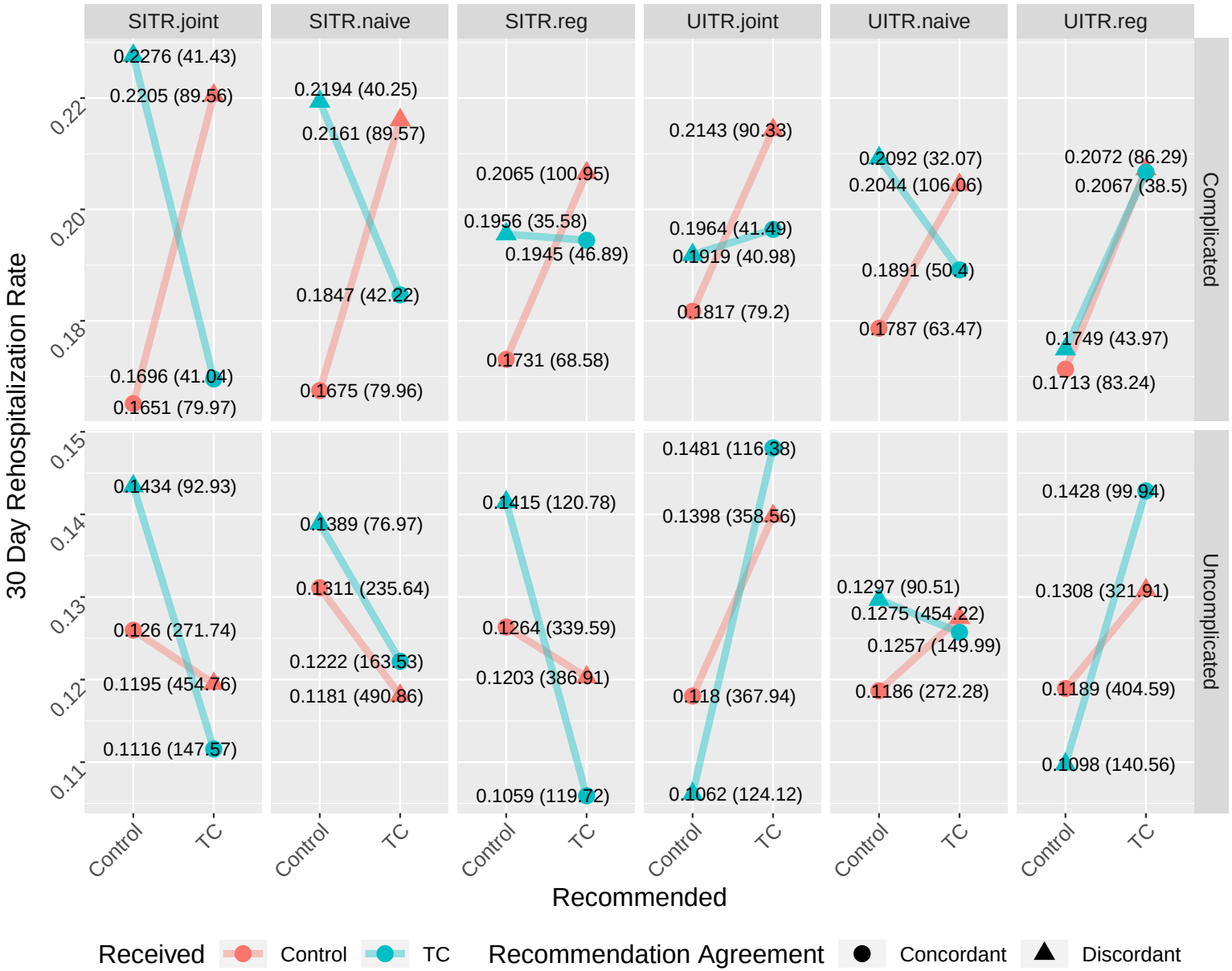
$$\bar{Y}_{a,b}(\hat{g}) = \frac{\sum_{i=1}^{n} Y_i \mathbb{1}(A_i = a, \hat{g}(X_i) = b)/P(A_i = a|X_i)}{\sum_{i=1}^{n} \mathbb{1}(A_i = a, \hat{g}(X_i) = b)/P(A_i = a|X_i)}$$

Under the usual causal assumptions,

$$\bar{Y}_{a,b}(\hat{g}) \xrightarrow{p} E(Y(a)|\hat{g}(X_i) = b)$$

Subgroup-cond. treatment effect: $E(Y(1) - Y(0)|\hat{g}(X_i) = b)$

Performance evaluation for ITRs on TC Data



Results

- 23 variables were selected into the estimated treatment rule for the complicated group, 40 were selected for the uncomplicated group, 22 of which were selected for both groups and 19 of the 22 had same sign

- Increased benefit of TC for both groups: those who have lymph node swelling; nephritis, nephrosis, or renal sclerosis; those with fluid and electrolyte disorders; those with immune disorders; gastrointestinal disorders; those who had a claim with a provider whose specialty is medical oncology

- Decreased benefit of TC for both groups: those with symptoms involving nervous of musculoskeletal systems

Summary

- The contrast framework is very flexible. It does not need perfect estimate of $E(Y|X, A)$. However, a *good* estimate will help increase the efficiency.
- When the true underlying models in different studies have common features, adding group penalty will help variable selection and classification.
- Our framework also dealt with different treatment effect scales and different assignment mechanisms across studies.
- Our framework happens to work well in our setting of individualized health system intervention assignment based on EHR data