Comparison of recent approaches for subgroup identification from clinical and observational data

Ilya Lipkovich (Eli Lilly and Company), David Svensson (AstraZeneca)

Bohdana Ratitch (Bayer), Alex Dmitrienko (Mediana)

14th Annual Conference on Statistical Issues in Clinical Trials

April 12, 2022

Acknowledgements: Haoda Fu

Previous overview articles

Tutorial in Biostatistics

Statistics in Medicine

Received 4 August 2015, Accepted 5 July 2016 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.7064

Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials

Ilya Lipkovich,[a*†] Alex Dmitrienko[b] and Ralph B. D'Agostino Sr.[c]

It is well known that both the direction and magnitude of the treatment effect in clinical trials are often affected by baseline patient characteristics (generally referred to as biomarkers). Characterization of treatment effect heterogeneity plays a central role in the field of personalized medicine and facilitates the development of tailored therapies. This tutorial focuses on a general class of problems arising in data-driven subgroup analysis, namely, identification of biomarkers with strong predictive properties and patient subgroups with desirable characteristics such as improved benefit and/or safety. Limitations of ad-hoc approaches to biomarker exploration and subgroup identification in clinical trials are discussed, and the ad-hoc approaches are contrasted with principled approaches to exploratory subgroup analysis based on recent advances in machine learning and data mining. A general framework for evaluating predictive biomarkers and identification of associated subgroups is introduced. The tutorial provides a review of a broad class of statistical methods used in subgroup discovery, including global outcome modeling methods, global treatment effect modeling methods, optimal treatment regimes, and local modeling methods. Commonly used subgroup identification methods are illustrated using two case studies based on clinical trials with binary and survival endpoints. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: clinical trials; exploratory subgroup analysis; biomarker analysis; data mining; multiplicity control.

Chapter 3

Data-Driven and Confirmatory Subgroup Analysis in Clinical Trials

Alex Dmitrienko, Ilya Lipkovich, Aaron Dane, and Christoph Muysers

Abstract In this chapter we provide an overview of the principles and practice of subgroup analysis in late-stage clinical trials. For convenience, we classify different subgroup analyses into two broad categories: data-driven and confirmatory. The two settings are different from each other primarily by the scope and extent of pre-specification of patient subgroups. First, we review key considerations in confirmatory subgroup analysis based on one or more pre-specified patient populations. This includes a survey of multiplicity adjustment methods recommended in multi-population Phase III clinical trials and decision-making considerations that ensure clinically meaningful inferences across the pre-defined populations. Secondly, we consider key principles for data-driven subgroup analysis and contrast it with that for a guideline-driven approach. Methods that emerged in the area of principled data-driven subgroup analysis in the last 10 years as a result of cross-pollination of machine learning, causal inference and multiple testing are reviewed. We provide examples of recommended approaches to data-driven and confirmatory subgroup analysis illustrated with data from Phase III clinical trials. We also illustrate common errors, pitfalls and misuse of subgroup analysis approaches in clinical trials often resulting from employing overly simplistic or naive methods. Overview of available statistical software and extensive bibliographical references are provided.

N. Ting et al. (eds.), *Design and Analysis of Subgroups with Biopharmaceutical Applications*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-030-40105-4_3

The mythology of subgroup analysis in Pharma

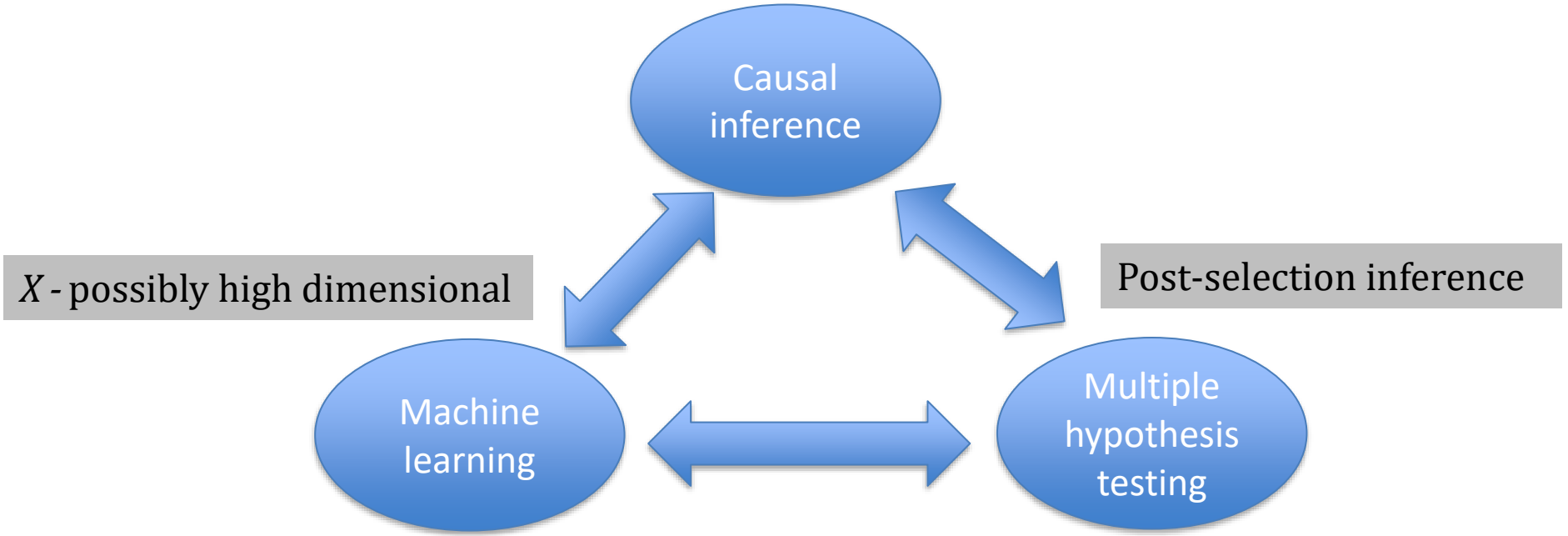
Common practices	”Good practices”
One covariate at a time strategy, (e.g test interactions at alpha=0.1)	Subgroups should be “pre-specified” and “biologically plausible”
Multiplicity does not need to be controlled since “it is for internal decision making”, “not for submission”	The central role of covariate-by-treatment interaction test, as a “gatekeeper” (no testing in subgroups unless passing the interaction test)
Accounting for uncertainty in the very last step of a multi-stage strategy, forgetting about “preliminary data looks”	No testing in subgroups unless the effect in the overall population is significant (consistency)
The subgroup search involves human interaction that is rarely reported	“Data-driven elements should be minimized”
“Null findings” rarely reported	Interpreting results “with caution”

Principled/disciplined data-driven subgroup analysis (SA)

- SA is a special case of statistical learning, rather than merely multiple testing problem
- A key challenge is estimating individual treatment effects (not observable on any subject)
- Intersection and cross-fertilization of different fields: causal inference, machine learning, multiple hypothesis testing.

Learning heterogeneity of TE from the data

$$CATE(x) = \Delta(x) = E(Y(1)|X = x) - E(Y(0)|X = x)$$



CATE: Conditional Average Treatment Effect (a.k.a ITE, PTE)

The set up: individual TE

- Each patient has two potential outcomes of Y, i.e. $Y_i(0), Y_i(1)$ corresponding to $T = 0,1$; only one outcome is observed (SUTVA)
- Outcome function, given pre-treatment covariates

$$m(t,x) = E(Y_i(t)|X = x), t \in \{0,1\}$$

- Under treatment ignorability, ensured by randomization in RCT, or "no unmeasured confounder" assumption in OC

$$m(t,x) = E(Y|T = t, X = x)$$

- Treatment contrast or conditional causal effect (CATE)

$$\Delta(x) = m(1,x) - m(0,x)$$

- We can write the response surface as

$$m(t,x) = h(x) + \frac{1}{2}\Delta(x)(2t - 1),$$

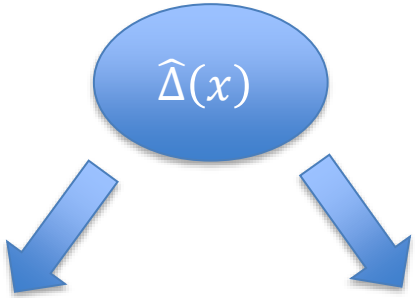
- $h(x)$ is the main covariate (prognostic) effect
- In studies with non-randomized treatments, we need to estimate propensity scores

$$\pi(x) = P(T = 1|X = x)$$

Defining subgroups based on $\Delta(x) = \mathrm{CATE}(x)$

- Assume we managed to estimate $\widehat{\Delta}(x)$
  - Perhaps, simply as $\widehat{\Delta}(x) = \hat{E}(Y|T=1, X=x) - \hat{E}(Y|T=0, X=x)$

Individualized treatment regimen/rule (ITR)
$\widehat{D}(x) = 1$ if $\widehat{\Delta}(x) > \delta$, $\widehat{D}(x) = 0$ if $\widehat{\Delta}(x) < -\delta$, otherwise treat randomly



May not ensure that each individual $\widehat{\Delta}(x_i) > \delta$, e.g. $E\{\widehat{\Delta}(x)\} > \delta$, for $x \in \hat{S}(x)$

$\hat{S}(x) = \{x: \widehat{\Delta}(x) > \delta)$
e.g. for δ=0

$\hat{S}(x)$ is learned from data as a biomarker signature:
e.g.
$\{x : X_1 > c_1 \ \& \ X_2 = c_2\}$

Literature on subgroup identification is diverse

ORIGINAL ARTICLE

OPEN

Selecting Optimal Subgroups for Treatment Using Many Covariates

Tyler J. VanderWeele,[a] Alex R. Luedtke,[b] Mark J. van der Laan,[c] and Ronald C. Kessler[d]

Abstract: We consider the problem of selecting the optimal subgroup to treat when data on covariates are available from a randomized trial or observational study. We distinguish between four different settings including: (1) treatment selection when resources are constrained; (2) treatment selection when resources are not constrained; (3) treatment selection in the presence of side effects and costs; and (4) treatment selection to maximize effect heterogeneity. We show that, in each of these cases, the optimal treatment selection rule involves treating those for whom the predicted mean difference in outcomes comparing those with versus without treatment, conditional on covariates, exceeds a certain threshold. The threshold varies across these four scenarios, but the form of the optimal treatment selection rule does not. The results suggest a move away from the traditional subgroup analysis for personalized medicine. New randomized trial designs are proposed so as to implement and make use of optimal treatment selection rules in healthcare practice.

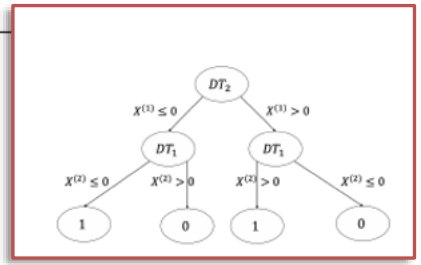
Keywords: Effect modification; Interaction; Optimal treatment selection; Precision medicine; Personalized treatment; Randomized trial; Subgroup

(*Epidemiology* 2019;30: 334–341)

treatment across sub
covariates.[1–6] Such ar
a treatment might be
or for younger versu
acteristic or variable
These types of analy
might vary across in
often referred to as "
be useful in deciding
sources are limited.
which of two treatme

While well-esta
decades to carry out st
by a single covariate,[1
sirable to make use of
the individual perspe
to best choose the ap
a particular set of ch
described as "personal

$$\hat{S}(x) = \{x : \hat{\Delta}(x) > \delta\}$$



CAPITAL: Optimal Subgroup Identification via Constrained Policy Tree Search

Hengrui Cai* [1], Wenbin Lu†[1], Rachel Marceau West†[2], Devan V. Mehrotra§[2], and Lingkang Huang¶[2]

[1]Department of Statistics, North Carolina State University
[2]Biostatistics and Research Decision Sciences, Merck & Co., Inc.

Abstract

Personalized medicine, a paradigm of medicine tailored to a patient's characteristics, is an increasingly attractive field in health care. An important goal of personalized medicine is to identify a subgroup of patients, based on baseline covariates, that benefits more from the targeted treatment than other comparative treatments. Most of the current subgroup identification methods only focus on obtaining a subgroup with an enhanced treatment effect without paying attention to subgroup size. Yet, a clinically meaningful subgroup learning approach should identify the maximum number of patients who can benefit from the better treatment. In this paper, we present an optimal subgroup selection rule (SSR) that maximizes the number of selected patients, and in the meantime, achieves the pre-specified clinically meaningful mean outcome, such as the average treatment effect. We derive two equivalent theoretical forms of the optimal SSR based on the contrast function that describes the treatment-covariates interaction in the outcome. We further propose a ConstrAined Polley Tree seArch aLgorithm (CAPITAL) to find the optimal SSR within the interpretable decision tree class. The proposed method is flexible to handle multiple constraints that penalize the inclusion of patients with negative treatment effects, and to address time to event data using the restricted mean survival time as the clinically interesting mean outcome. Extensive simulations, comparison studies, and real data applications are conducted to demonstrate the validity and utility of our method.

Optimal subgroup selection

Henry W. J. Reeve, Timothy I. Cannings and Richard J. Samworth
University of Bristol, University of Edinburgh
and University of Cambridge

Abstract

In clinical trials and other applications, we often see regions of the feature space that appear to exhibit interesting behaviour, but it is unclear whether these observed phenomena are reflected at the population level. Focusing on a regression setting, we consider the subgroup selection challenge of identifying a region of the feature space on which the regression function exceeds a pre-determined threshold. We formulate the problem as one of constrained optimisation, where we seek a low-complexity, data-dependent selection set on which, with a guaranteed probability, the regression function is uniformly at least as large as the threshold; subject to this constraint, we would like the region to contain as much mass under the marginal feature distribution as possible. This leads to a natural notion of regret, and our main contribution is to determine the minimax optimal rate for this regret in both the sample size and the Type I error probability. The rate involves a delicate interplay between parameters that control the smoothness of the regression function, as well as exponents that quantify the extent to which the optimal selection set at the population level can be approximated by families of well-behaved subsets. Finally, we expand the scope of our previous results by illustrating how they may be generalised to a treatment and control setting, where interest lies in the heterogeneous treatment effect.

the level τ on B. The p-values are then combined via Holm's procedure (Holm, 1979) to identify a finite union of hyper-cubes that satisfy our Type I error control property. Our final selection set $\hat{A}_{\text{OSS}}$ maximises the empirical measure among all elements of $\mathcal{A}$ that lie within this finite union of hyper-cubes.

What to look for in papers on Subgroup Identification?

The number of predictors the procedure can handle

- p=1
  - focus on selecting a cutoff for a single continuous biomarker (e.g. STEPP method by Bonetti and Gelber, 2000; Han et al, 2021)
- $p \approx$10-20
- $p \approx$100-1000
- $p \gg n$
  - Feature space grows with sample size

Model complexity

- What is the complexity of the "model space" where the subgroups reside?
  - Subgroups defined based on "black box" functions of covariates, $\hat{S}(x) = \{x: \widehat{\Delta}(x) > c\}$
  - Subgroups defined by simple biomarker signatures with up to 2 variables, $\hat{S}(x) = \{x: X_1 \leq c_1, X_3 > c_3\}$
- How is model complexity controlled to prevent data overfitting?

Does it apply only to RCT or to OS as well?

- For observational data, there is an interplay between confounders and modifiers of treatment effect, making model selection more challenging
  - Confounders are predictive of both treatment T and outcome Y
  - Effect modifies are predictive of CATE, $\Delta(x)$

What output does the method produce?

- Individualized treatment contrast, $\widehat{\Delta}(x)$
- Signatures of promising subgroups, $\widehat{\mathrm{S}}(x) = \{x: X_1 \leq c_1, X_3 > c_3\}$
- Optimal treatment assignment rule $\widehat{D}(x) = 1\ if\ \widehat{\Delta}(x) > \delta$, otherwise $\widehat{D}(x) = 0$
- Predictive biomarkers (a.k.a. effect modifiers ordered) by variable importance score.

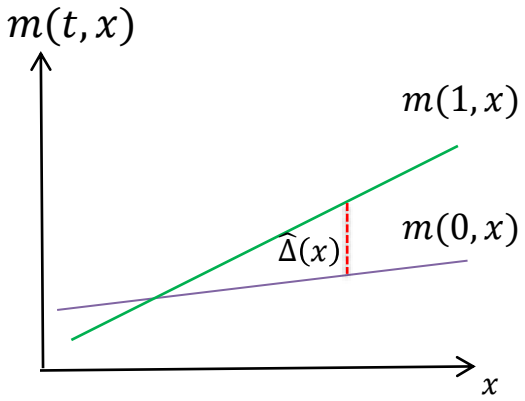
What inference is done, if at all?

- Inference on $\Delta(x)$
  - Pointwise CI for random forests (Wager and Athey, 2018), CI for $\widehat{\Delta}(x)$ estimated from LASSO (Ballarini et al, 2018),
  - Simultaneous bands on $\widehat{\Delta}(x)$ from semiparametrics (Guo at al., 2021)
- Inference on certain features of $\Delta(x)$
  - Testing for presence of treatment effect heterogeneity (via latent mixtures, Shen and He, 2015) or
  - Machine learning methods with cross-fitting (Chernozhukov, 2019)
- Controlling the probability of selecting the right subgroups, $\hat{S}(x)$ vs $S_{true}(x)$
  - Bayesian credible intervals, $\Pr(\hat{S}_{lower} \subseteq S_{true} \subseteq \hat{S}_{upper}) > 1 - \alpha$ (Schnell et al, 2018)

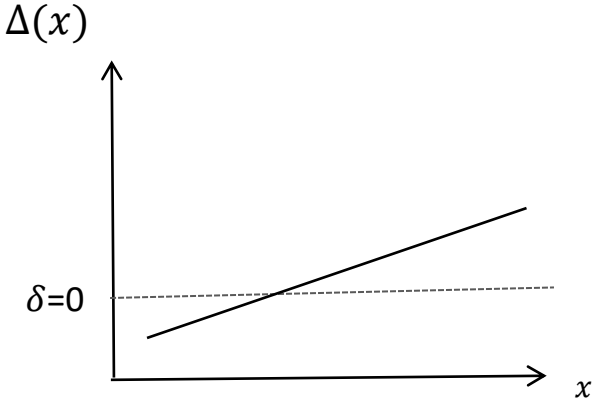
What inference is done, if at all (cont.)?

- Estimating "honest effect" in selected subgroup $\hat{S}(x)$
  - Using bootstrap correction for optimism bias (Foster et al, 2011; Guo and He, 2020)
  - Bayesian model averaging (Bornkamp et al, 2017)
- Inference on ITR, $D(X)$
  - Evaluating expected benefits if the regimen (rule) were applied to all patients,
  - $\text{Value} = E\{Y(\hat{D}(X))\}$ contrasted with the value of "always treat" or other strategy
- Controlling the False Discovery Rate
  - E.g., for selection of predictive biomarkers (Wei et al, 2021; Sechidis et al, 2021)

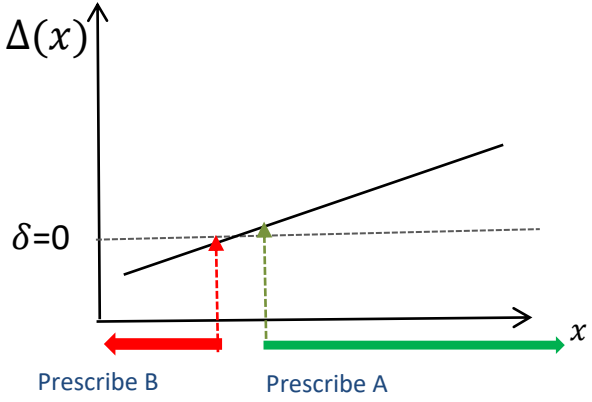
Typology of Subgroup Identification; Lipkovich et al. (2017)



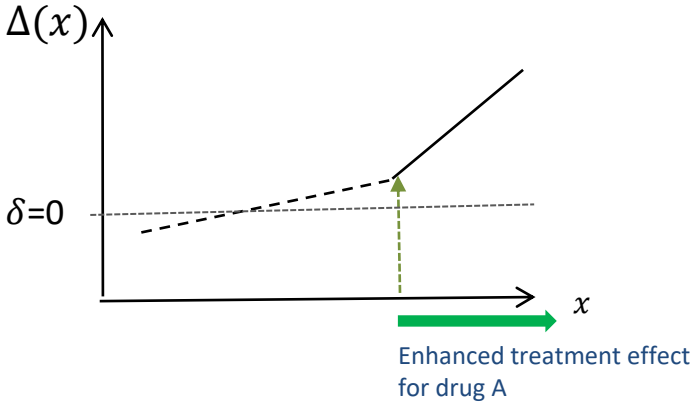
Global outcome modeling: Y



Direct treatment effect modeling

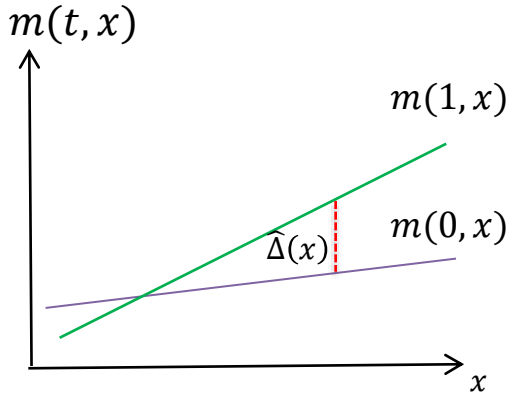


Individual treatment regimen modeling: $\mathbf{sign\{\Delta(x)\}}$



Local treatment effect modeling : Subgroup search

Global outcome modeling



Global outcome modeling: Y

A multi-stage (multi-model) process termed *meta-learning*

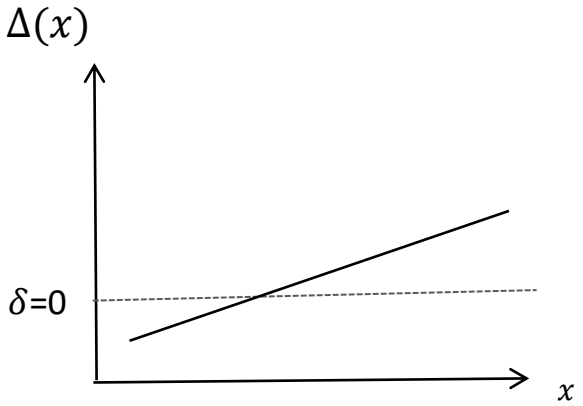
As a precursor, see Virtual Twins (VT) by Foster et al (2011)

- **T**-(*two*) learning:
  - Fit $m(t,x) = E(Y|T=t, X=x)$, separately by arms
  - Compute $\hat{\Delta}(x) = \hat{m}(1,x) - \hat{m}(0,x)$
- **S**-(*single*) learning:
  - Fit $m(t,x) = E(Y|T=t, X=x)$, in pooled data with X*T interactions added
  - Compute $\hat{\Delta}(x) = \hat{m}(1,x) - \hat{m}(0,x)$
- **X**-learning based on two version of CATE
  - $\hat{\Delta}_1(x)$ by modeling $Y(1) - \hat{m}(0,x)$, on treated subjects
  - $\hat{\Delta}_0(x)$ by modeling $\hat{m}(1,x) - Y(0)$, on control subjects
  - Compute $\hat{\Delta}(x) = \hat{\Delta}_0(x)\pi(x) + \hat{\Delta}_1(x)(1-\pi(x))$
- Regularization challenges when modeling CATE
  - Separate penalties for prognostic and predictive effects (Imai & Ratkovic, 2013)
  - Separate modeling of counterfactuals in X-learning (Künzel et al, 2019)
  - Separate penalties for prognostic and predictive effects in Bayesian causal forests (Hahn et al, 2020)

Direct treatment effect modeling

Directly evaluates $\Delta(x)$ obviating estimating main effects $h(x)$

- Adopt any tree-based method by modifying splitting criterion
  - Interaction trees, e.g. Su et al (2009) maximizing at every split $\left(\widehat{\Delta}_{left} - \widehat{\Delta}_{right}\right)^2$
  - Causal trees and causal forests (Athey and Imbens, 2016; Wager and Athey, 2018)
    - Local non-parametric estimates of $\Delta(x)$ by averaging treatment effects from terminal nodes across trees
    - "Honest trees": divide data into two halves, use one for splitting and the second for estimating $\Delta(x)$
    - Inference for random forests (Efron, 2013 and Wager et al. 2014)
- Modified outcome and covariate (functions) methods
  - Tian et al. (2014) and Chen et al. (2017), see next



Global treatment effect modeling

From Modified Outcome to Modified Covariate [function] Methods

- A broad framework for directly estimating $\Delta(x)$ for different types of outcomes/loss functions (R package **personalized**)
  - $A = 2T - 1 \in \{-1,1\}, \pi(x) = \Pr(T = 1|X = x)$, $\boxed{\pi(A|x) = A\pi(x) + \frac{1-A}{2}}$

Probability of receiving the treatment actually received

MOM $E\left(\left(\frac{AY}{\pi(A|x)} - g(x)\right)^2 \middle| X = x\right) \to min$ returns $g(x) = \Delta(x)$,

$E\left(\frac{1}{\pi(A|x)}(2AY - g(x))^2 \middle| X = x\right)$ has the same estimand $g(x) = \Delta(x)$, and so is

MCM $E\left(\frac{4}{\pi(A|x)}\left(Y - \frac{A}{2}g(x)\right)^2 \middle| X = x\right)$, **W-learning** in Chen et al. (2017)

- Choosing different loss functions allows for different outcomes types
- Options for modeling $g(x)$: linear (e.g. via penalized regression) reduces it to multiplying each covariate by *A/2* (modified covariate), gradient boosting, …

Treatment effect modeling: R-learning

- R-learning for estimation of $\Delta(x)$ (Zhao et al, 2018; Nie and Wager, 2021; inspired by Robinson's transformation and Double/Debiased Machine Learning of Chernozhukov, 2017)
- Note, $\Delta(x) = E\left(\frac{Y-m(x)}{T-\pi(x)}\right)$, where $m(x) = E(Y|X = x)$

$$\widetilde{\Delta}(\cdot) = argmin_{\Delta} \frac{1}{N} \sum_{i=1}^{N} [Y_i - m(x_i) - \{T_i - \pi(x_i)\}\Delta(x_i)]^2 + \Lambda_N\{\Delta(\cdot)\}$$

  - Prognostic effects and propensity (for non-randomized trials) need to be estimated at first step, but the focus is placed on the target $\Delta(x)$
  - $m(x_i)$ and $\pi(x_i)$ are estimated from off-the-shelf ML methods and their cross-fitted versions are plugged-in $\hat{m}^{-i}(x_i)$ and $\hat{\pi}^{-i}(x_i)$

Software for subgroup identification

- http://biopharmnet.com/subgroup-analysis-software/

Software for subgroup identification

SIDES method

R package SIDES implementing the regular SIDES method (Subgroup Identification Based on Differential Effect Search) based on Lipkovich et al. (2011) [last update: October 04, 2016]. The package is maintained by Marie-Karelle Riviere (eldamjh@gmail.com).

Download the SIDESxl package (an Excel add-in) which implements the regular SIDES and SIDEScreen methods [last update: March 25, 2016]. The package is maintained by Ilya Lipkovich (ilya.lipkovich@gmail.com).

Download the R functions, C++ functions (sides64.dll), and examples for the regular SIDES (Lipkovich et al, 2011), SIDEScreen (Lipkovich and Dmitrienko, 2014), and Stochastic SIDEScreen (Lipkovich et al, 2017) methods [last update: October 01, 2018]. The functions and examples are provided by Ilya Lipkovich (ilya.lipkovich@gmail.com), Alex Dmitrienko and Bohdana Ratitch.

Interaction Trees method

Download the R functions and examples for the Interaction Trees method [last update: Dec 30, 2014]. The functions and examples are provided by Xiaogang Su (Xiaogang Su's site). Download the R code for the Interaction Trees method [last update: Dec 30, 2014].

Virtual Twins method

Download the R code for the Virtual Twins method [last update: Dec 30, 2014]. The code is provided by Jared Foster (jaredcf@umich.edu).
R package aVirtualTwins that implements an adaptation of the Virtual Twins method by Foster et al. (2011)

GUIDE method

GUIDE package for classification and regression trees now includes methods for subgroup identification. The GUIDE package is maintained by Wei-Yin Loh (Wei-Yin Loh's site). For more information on the subgroup identification features, see Section 5.10 of the GUIDE User Manual [last update: September 25, 2018] and paper by Wei-Yin Loh, Xu He and Michael Man.

In addition, MrSGUIDE package implements the GUIDE method for randomized trials and observational studies.

QUINT method

Quint package for *QU*alitative *IN*teraction *T*rees. The package is maintained by Elise Dusseldorp (Elise Dusseldorp's site) and colleagues. Reference: Dusseldorp and Mechelen (2014).

FindIt method

FindIt package for finding heterogeneous treatment effects [last update: February 27, 2015]. Reference: Imai and Ratkovic (2013).

Blasso method

Download the R functions for the Bayesian two-stage Lasso strategy for biomarker selection for time-to-event endpoints [last update: December 16, 2014]. The code is provided by Xuemin Gu (xuemin.gu@bms.com). Reference: Gu, Yin and Lee (2013).

ROWSi method

Download the R code for the ROWSi method (Regularized Outcome Weighted Subgroup identification). Reference: Yu et al. (2015).

Model-based Recursive Partitioning

R partykit package: A Toolkit for Recursive Partytioning, which can perform subgroup analyses using the functions lmtree(), glmtree() (or more generally, mob()) and ctree()).

Recently a new package model4you has been created that specializes on stratified and personalized treatment effect estimation. The package is maintained by Heidi Seibold (heidi@seibold.co).

See examples of subgroup analysis in Seibold et al. (2015) and Seibold et al. (2016)

Other packages

R package *personalized* (maintained by Jared Huling) for subgroup identification and estimation of heterogeneous treatment effects. It is a general framework that encompasses a wide range of methods including ROWSi, outcome weighted learning, and many others. See documentation and article explaining the underlying methodology.

R package SubgrID implements several algorithms for developing threshold-based multivariate (prognostic/predictive) biomarker signatures via bootstrapping and aggregating of thresholds from trees (BATTing), Monte-Carlo variations of the Adaptive Indexing Method (AIM) by Huang X. et al. (2017) and and adaptation of Patient Rule Induction Method (PRIM) for subgroup identification by Chen G. et al. (2015).

Fu, Zhou and Faries (2016) developed a search approach that provides simple and interpretable rules defining subgroup of patients with maximizes average patients' benefit for different treatments within a general framework of outcome weighted learning (OWL). Here you can find the C++ implementation.

R package DynTxRegime implements methods to estimate dynamic treatment regimes using Interactive Q-Learning, Q- Learning, weighted learning, and value-search methods based on Augmented Inverse Probability Weighted Estimators and Inverse Probability Weighted Estimators.

R package listdtr constructs list-based rules (lists of if-then clauses) to estimate the optimal dynamic treatment regime based on the approach by Zhang et al. (2016).

The subtee R package implements method for bootstrap-corrected estimation after subgroup selection described in Rosenkranz (2016) and a model averaging approach from Bornkamp et al. (2016).

TSDT: Treatment-Specific Subgroup Detection Tool by Chakib Battioui, Brian Denton and Lei Shen (2018).

StratifiedMedicine by Thomas Jemielita is a broad toolkit for subgroup identification and stratified/precision medicine. The package also includes a novel algorithm PRISM (Patient Response Identifiers for Stratified Medicine) by Jemielita and Mehrotra (to appear).

Generalized Random Forests (grf) is a package for forest-based statistical estimation and inference. The package currently provides methods for non-parametric least-squares regression, quantile regression, survival regression and treatment effect estimation (optionally using instrumental variables), with support for missing values.

Policy learning via doubly robust empirical welfare maximization over trees (policytree) supports optimal policies via doubly robust empirical welfare maximization over trees. This package implements the multi-action doubly robust approach of Zhou, Athey and Wager (2018).

R package (debiased.subgroup) implements bootstrap-assisted desparsified Lasso and bootstrap-assisted R-split estimators on selected subgroup's treatment effect estimation. The implemented estimators remove the subgroup selection bias and the regularization bias induced by high-dimensional covariates. For more information, see Guo, Wei, Wu and Wang (2021).

R package (rlearner) supports quasi-oracle estimation of heterogeneous treatment effects based on Nie and Wager (2021).

R package (causalToolBox) is available to enable metalearners for estimating heterogeneous treatment effects using machine learning based on Künzel, Sekhona, Bickel and Yu (2019).

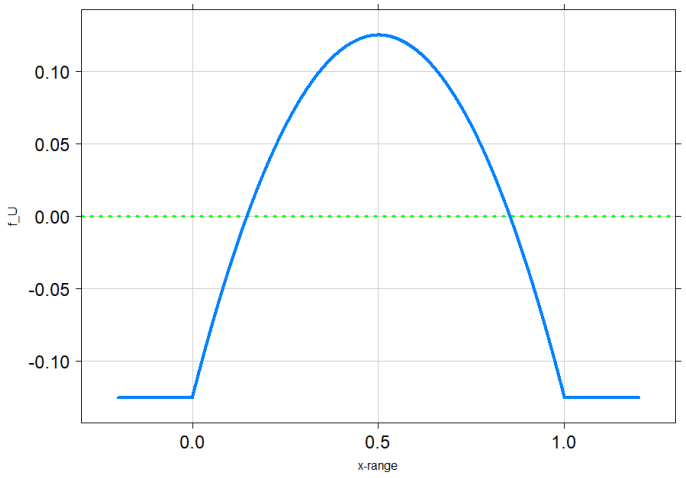
R code (CAPITAL) for the implementation of optimal subgroup identification via constrained policy tree search based on Cai, Lu, West, Mehrotra and Huang (2021).

R package (bcf) supports causal inference for a binary treatment and continuous outcome using Bayesian causal forests based on Hahn, Murray and Carvalho (2019).

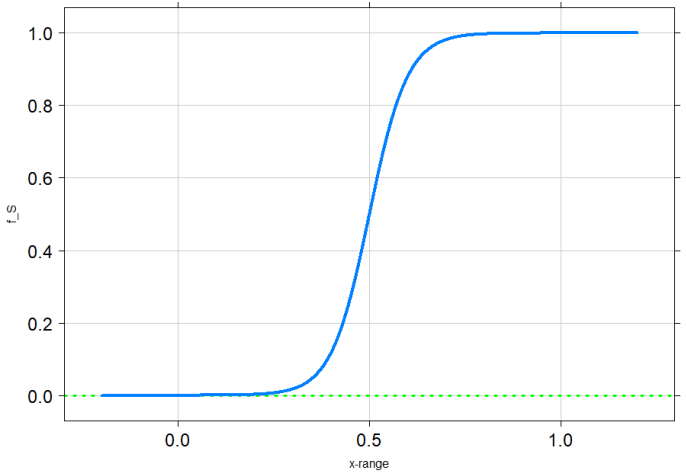
Simulation example: A single data set

- N=1000, randomization 3:1, $T \in \{0,1\}$
- $X_1, X_3, X_4 \sim N(0,1)$, $X_2 \in \{1,2,3\}$ with $p = 1/3$
- $Y = 100 - (X_1 + 5X_2) + T\{g_1(X_3) + g_2(X_4)\} + \varepsilon$, $\varepsilon \sim N(0,1)$
- $g_1(x) = a - b(x - 0.5)^2, 0 \le x \le 1$
- $g_2(x) = \frac{c}{1+e^{-d(x-0.5)}}, 0 \le x \le 1$; else $g_2(X) = 0$
- Add 16 noise biomarkers $X_5, \ldots, X_{20} \sim N(0,1)$ to the analysis data set

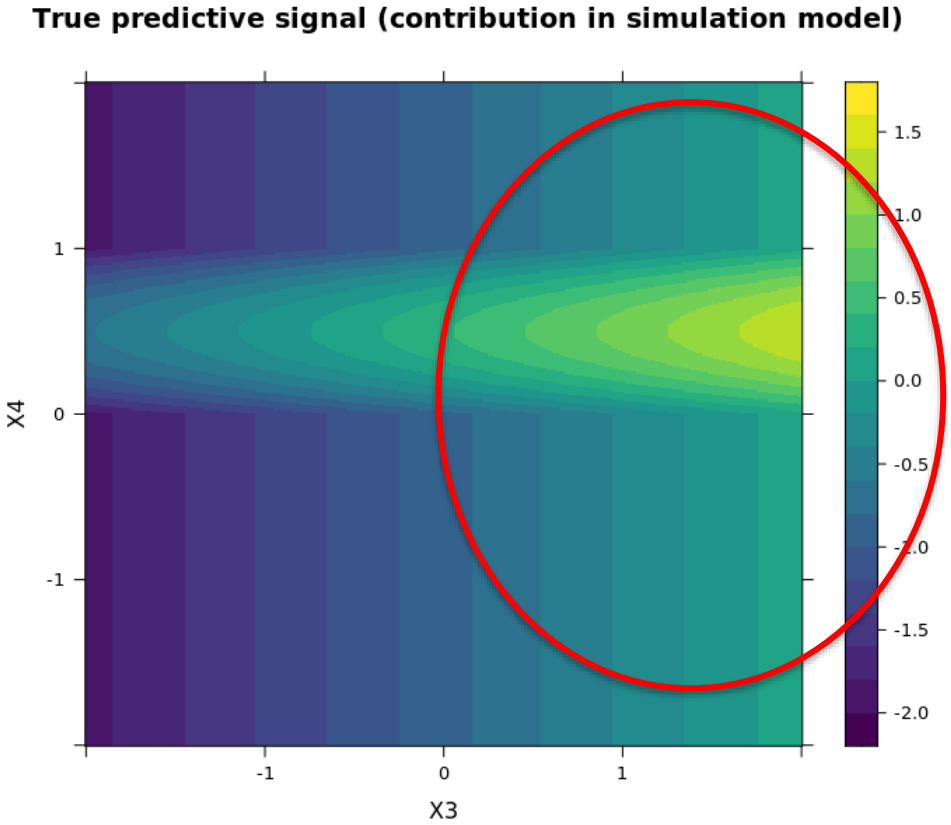
$g_1(x)$



$g_2(x)$



Treatment effect: $\Delta(x) = E(Y(1) - Y(0)|x)$



Overall ATE: Δ=0.21

TE within subgroup:

$E(\Delta(x)|\Delta(x) > 0) = 0.45$

Subgroup size:

$E\{I(\Delta(x) > 0)\} = 0.69$

Methods

- T, S, X - learning
- Causal forest (CF)
- Modified Outcome method (MOM) with RF and Xgboost
- R - learning

- For each method, identify *subgroup signature* on fitting training data: $\hat{S}(X) = \{x: I(\widehat{\Delta}_{train}(x) > 0)\}$
- Evaluate average treatment effect in identified subgroup $\hat{S}$ on independent test data (n=10,000),

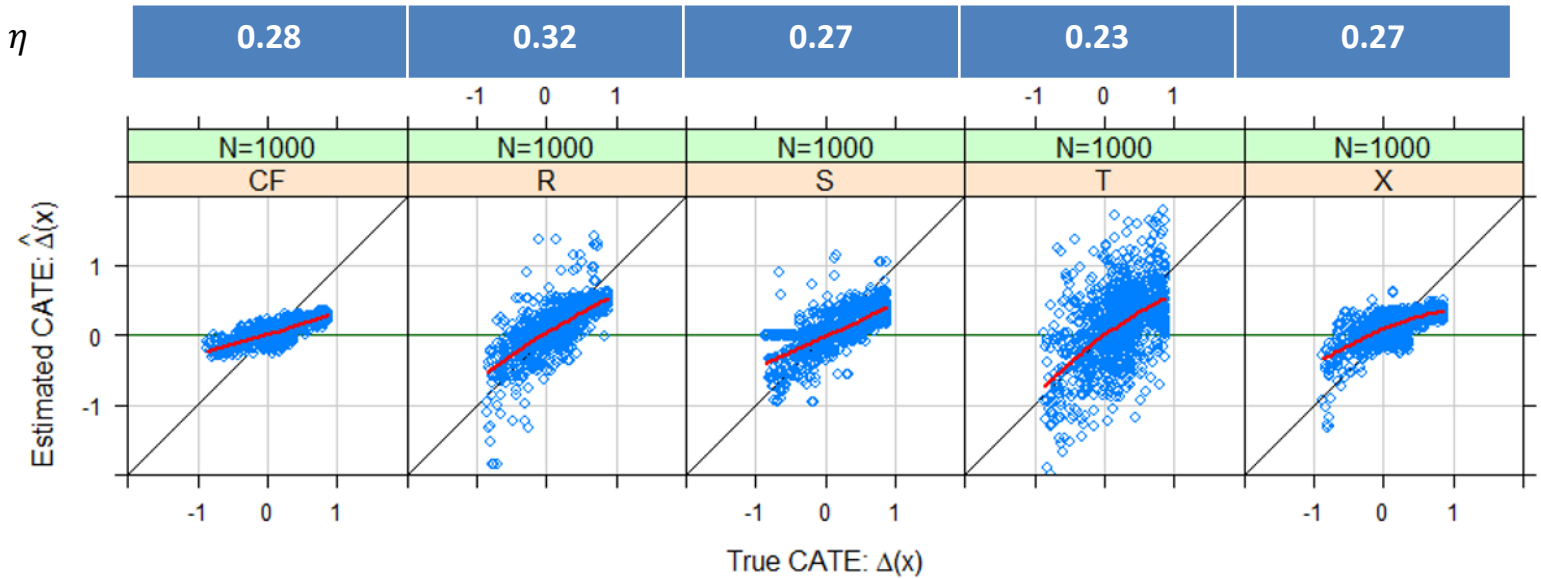
$$TE(\hat{S}) = E_X\{\Delta(X)|\widehat{\Delta}_{train}(X_{test}) > 0\}$$

- Compute subgroup utility index: Treatment effect per subject in overall population

$$\eta = TE(\hat{S}) \times \frac{n(\hat{S})}{n}$$

Estimating CATE with NO noise variables in the analysis set

- Regularization bias towards zero: largest in causal forest and smallest in T-learning
- large variability: MOM (not shown) and T learning

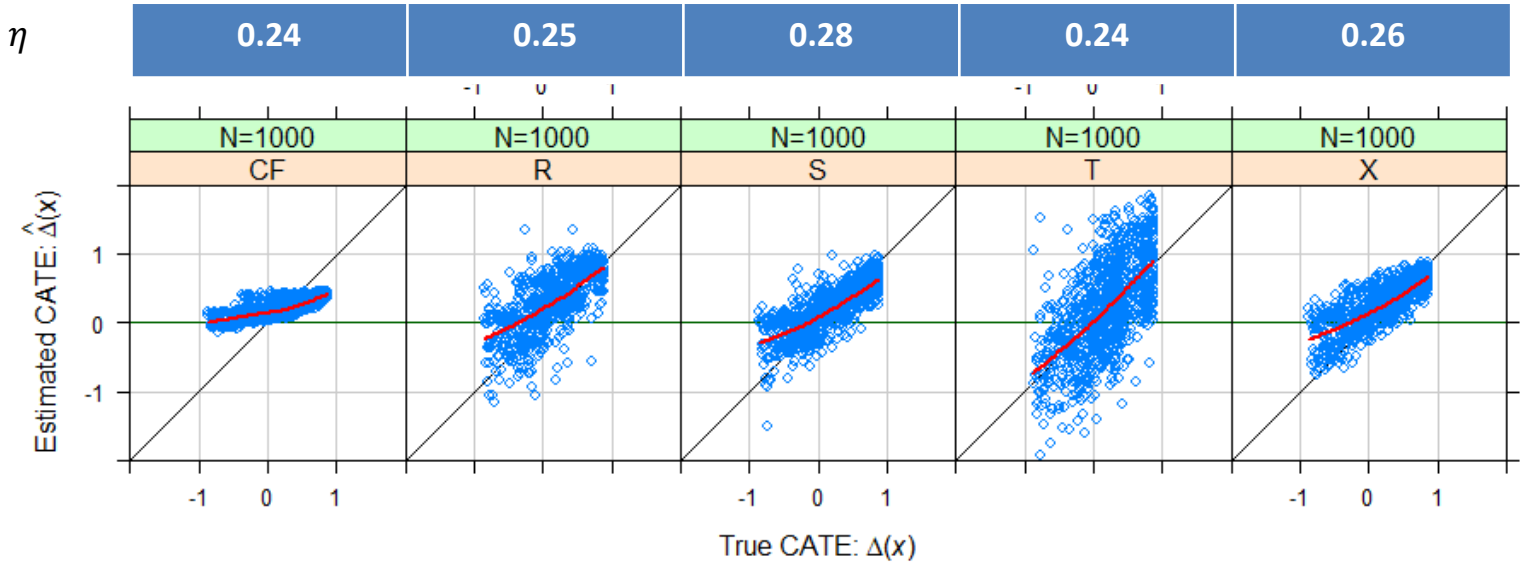


ATE=0.21

Red line is smoothed predicted CATE, black line 45 degrees

Estimating CATE, with noise variables in the analysis set

- Regularization bias towards zero: largest in causal forest and smallest in T-learning
- large variability: MOM (not shown) and T learning



Red line is smoothed predicted CATE, black line 45 degrees

Summary

- A shift from ad-hoc “subgroup chasing” methods towards principled methods of personalized/precision medicine utilizing ideas from causal inference, machine learning and multiple testing emerged in last 10 years producing a vast number of diverse approaches
- For naïve multistage methods (requiring fitting the response surface $m(t, x)$) regularization bias can be large, as each step is optimized for prediction, rather than for the final estimation target (Künzel et al, 2019; Chernozhukov, 2019; Nie and Wager, 2021)
- While methods that estimate $\Delta(x)$ obviating fitting main effects $h(x)$ are attractive, substantial efficiency can be gained by using doubly-robust methods, such as utilizing augmented inverse propensity weighted scores, even in the context of RCT where propensities are known (Athey and Wager, 2021; Kennedy, 2021)
- There is increasing interest in developing ITRs respecting constraints on costs, adverse events, sample size (Wang et al, 2018; Athey and Wager, 2021; Cai et al, 2021)
- There is a need in interpretable personalized solutions (ITR’s) within a pre-defined policy class, e.g tree-structured or boxes (Laber and Zhao, 2015; Cai et al, 2021; Doubleday et al., 2021)

References

Athey S and Wager S (2021) Policy learning with observational data. Econometrica. 89(1),133-161.

Ballarini NM, Rosenkranz GK, Jaki T, König F, Posch M (2018) Subgroup identification in clinical trials via the predicted individual treatment effect. PLoS One 13:e0205971

Bornkamp B, Ohlssen D, Magnusson BP, Schmidli H (2017) Model averaging for treatment effect estimation in subgroups. Pharm Stat 16,133–142.

Cai H, Lu Wenbin, West RM, Mehrotra DV, Huang L. (2021). CAPITAL: Optimal subgroup identification via constrained policy tree search. arXiv:2110.05636v1

Chen S, Tian L, Cai T, Yu M (2017) A general statistical framework for subgroup identification and comparative treatment scoring. Biometrics. 73(4),1199–1209.

Chernozhukov V, Demirer M, Duflo E, and Fernandez-val (2019) Generic machine learning inference on heterogenous treatment effects in randomized experiments. arXiv:1712.04802v4.

Doubleday K, Zhou H, Fu H, Zhou J. (2021) Risk controlled decision trees and random forests for precision medicine. Statistics in medicine. To appear

Guo X, He X. Inference on selected subgroups in clinical trials. J Am Stat Assoc 116(535),1498-1506.

Guo W, Zhou X-H, Ma S (2021) Estimation of optimal individualized treatment rules using a covariate-specific treatment effect curve with high-dimensional covariates. J Am Stat Assoc 116(533),309-321.

Hahn PR, Murray JS, Carvalho CM. (2019) Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects. Bayesian Anal. 15(3), 965-1056.

Han Y, Tang S-Y, Lin H-M, Hsu JC. (2021) Exact simultaneous confidence intervals for logical selection of a biomarker cut-point. Biometrical journal 1-18.

Hermansson E and Svensson D. (2021) On discovering treatment-effect modifiers using virtual twins and causal forest ML in the presence of prognostic biomarkers. In O. Gervasi et al. (Eds.): ICCSA, pp. 624–640.

References (cont.)

Huling JD and Yu M. (2021) Subgroup identification using the personalized package. Journal of statistical software. 98(5) 1-60.

Jemielita TO and Mehrotra DV (2019) PRISM: Patient Response Identifiers for Stratified Medicine, arXiv:1912.03337

Kennedy EH (2021). Optimal doubly robust estimation of heterogeneous causal effects. arXiv:2004.14497v2.

Künzel SR, Sekhona JS, Bickel PJ and Yu B. (2019) Metalearners for estimating heterogeneous treatment effects using machine learning. Proceedings of the National Academy of Sciences, 116(10), 4156-4165.

Kitagawa T, and Tetenov A. (2018) Who should be treated? Empirical Welfare Maximization methods for treatment choice. Econometrica 86(2), 591–616.

Laber EB, Zhao YQ (2015) Tree-based methods for individualized treatment regimes. Biometrika 102,501–514.

Lipkovich I, Dmitrienko A (2014) Strategies for identifying predictive biomarkers and subgroups with enhanced treatment effect clinical trials using SIDES. J Biopharm Stat 24,130–153.

Lipkovich I, Dmitrienko A, D'Agostino BR (2017) Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. Stat Med 36,136–196.

Loh WY, Fu H, Man M, Champion V, Yu M (2016) Identification of subgroups with differential treatment effects for longitudinal and multiresponse variables. Stat Med 35,4837–4855.

Nie X and Wager S. (2021) Quasi-oracle estimation of heterogeneous treatment effects. Biometrika 108(2), 299–319.

Qi Z, Lui B, Fu H, Lui Y. (2020). Multi-armed angle-based direct learning for estimating optimal individualized treatment rules with various outcomes. J Am Stat Assoc 115(530), 678-691.

Reeve HW, Cannings TI, and Samworth RJ. (2021) Optimal subgroup selection. arXiv:2109.01077.

Schnell PM, Müller P, Tang Q, Carlin BP (2018) Multiplicity-adjusted semiparametric benefiting subgroup identification in clinical trials. Clinical trials 15(1),75-86.

References (cont.)

Sechidis K, Kormaksson M, Ohlssen D (2021) Using knockoffs for controlled predictive biomarker identification. Statistics in Medicine 40,5453–5473.

Shen C, Li X, and Jeong J. (2016) Estimation of treatment effect in a sub-population: an Empirical Bayes approach. Journal of Biopharmaceutical Statistics 26(3),507–518.

Shen J, He X. (2015). Inference for subgroup analysis with a structured logistic-normal mixture model. J Am Stat Assoc 110,303–312.

Thomas M, Bornkamp B, Seibold H. (2018) Subgroup identification in dose-finding trials via model-based recursive partitioning. Statistics in Medicine 37(10),1608–1624.

VanderWeele TY, Luedtke AR, van der Laan MJ, Kessler RC. (2019) Selecting optimal subgroups for treatment using many covariates. Epidemiology 30,334-341.

Wager S, Athey S (2018) Estimation and inference of heterogeneous treatment effects using random forests. J Am Stat Assoc 113,1228–1242.

Wang Y, Fu H, and Zeng D. (2018) Learning optimal personalized treatment rules in consideration of benefit and risk: with an application to treating type 2 diabetes patients with insulin therapies. J Am Stat Assoc 113(521), 1-13.

Wei Y, Hsu JC, Chen W, Chew EY, Ding Y. (2021) Identification and inference for subgroups with differential treatment efficacy from randomized controlled trials with survival outcomes through multiple testing. Statistics in Medicine. 2021;1–18.

Zhang Y, Schnell P, Song C, Huang B, Lu B. (2021) Subgroup causal effect identification and estimation via matching tree. Comp Stat and Data Analysis. https://doi.org/10.1016/j.csda.2021.107188

Zhao Y, Zheng D, Rush AJ, Kosorok MR. (2012) Estimating individualized treatment rules using outcome weighted learning. Journal of the American Statistical Association 107,1106–1118.

Thank you!

Q & A

Ilya.Lipkovich@lilly.com

Backup

Direct estimation of CATE: A-learning

- It is easy to see that the following modified covariate function method has as population minimizer, $g(x) = \Delta(x)$

$$E\left(\left(Y - (T - \pi(x))g(x)\right)^2 \middle| X = x\right) \rightarrow min$$

- Like with W-method, this generalizes to different types of outcomes by replacing squared loss with appropriate loss functions (See Chen et al., 2017)
  - Chen S, Tian L, Cai T, Yu M. (2017) A general statistical framework for subgroup identification and comparative treatment scoring. Biometrics, 73(4), 1199–1209.

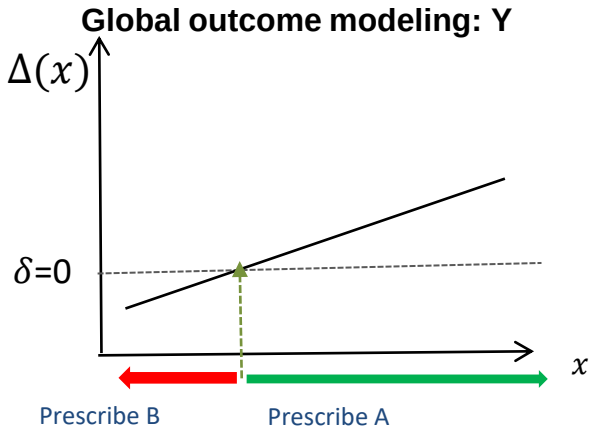
Doubly robust estimators of CATE

- Combines estimators of expected PO's with IPW (Athey and Wager, 2019; Kennedy, 2021)
- Consistent if eat least PO or propensity model is correct
- Reduces variability compared with direct methods
- Let $\widehat{\Delta}(x) = \widehat{m}(1,x) - \widehat{m}(0,x)$ be an estimator of CATE from any metalearner, causal forest, etc

$$\widehat{\Delta}_{DR}(x) = \widehat{\Delta}(x) + \frac{T - \hat{\pi}(x)}{\hat{\pi}(x)\big(1 - \hat{\pi}(x)\big)}\left[Y - \widehat{m}(x) - \big(T - \hat{\pi}(x)\big)\widehat{\Delta}(x)\right]$$

where, $m(x) = E(Y|X = x)$

Modeling ITRs (outcome weighted learning)



Individual treatment regimen modeling: sign{Δ(*x*)}

While ITR can be estimated based on methods of outcome modeling (1) or treatment effect modeling (2), some methods estimate directly the sign of $\Delta(x)$ by casting it as a classification problem (Zhao et al, 2012)

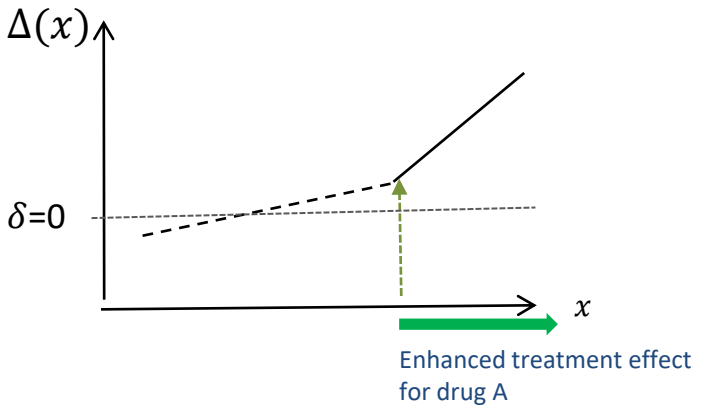
- One approach is to write the expected value of ITR
  $E\{Y(D(X))\} = E\left[\frac{I(D(\boldsymbol{X})=T)Y}{\Pr(T|x)}\right] \rightarrow max$
- This is equivalent to minimizing weighted classification loss $E\left[\frac{I(D(\boldsymbol{X})\neq T)Y}{\Pr(T|x)}\right] \rightarrow \min$
- Minimizing 0-1 loss is an NP problem so typically we modify it using a smooth convex surrogate loss function. E.g hinge, or exponential loss: $E[L_w(T, f(x))]$
- This allows using off-the-shelf packages to identify ITRs, e.g. logistic regression with lasso penalty and weights $w_i = Y_i/\Pr(T = t_i|\mathrm{X} = x_i)$

Modeling ITRs: Recent advances

- Treatment allocation based on simultaneous confidence band estimated from semiparametric modeling of $\Delta(x)$ (Guo et al, 2021)
- Multi armed angle-based direct learning for ITR (Qi et al, 2020)
- Learning optimal ITR adopting risk/costs constraints (Wang et al, 2018)
- Risk controlled decision trees and random forests for precision medicine (Doubleday et al, 2021)
- Searching treatment policies within a restricted class of fixed depth trees. Uses doubly robust estimator of treatment effect function. Athey and Wager (2021), **policytree** R package (by Sverdrup et al.)
  - Extending work on maximizing empirical welfare (value) of policies within restricted classes from randomized studies by Kitagawa and Tetenev (2018).
  - Recent application/extension: CAPITAL: Optimal subgroup identification via constrained policy tree search (Cai et al, 2021)

Direct subgroup search (local treatment effect modeling)

- Instead of estimating the response function $\Delta(x)$ on the entire covariate space and then carving out segments, search directly for such regions
- Recent methods
  - SIDEScreen (Lipkovich and Dmitrienko, 2014)
  - Adaptation of PRIM method in Chen et al, 2015
  - Sequential-BATTing (Huang et al, 2017) implemented in R package **SubgrID**



Local treatment effect modeling : Subgroup search