Modern approaches for evaluating treatment effect heterogeneity from clinical trials and observational data

Ilya Lipkovich (Eli Lilly and Company)

BBS Workshop

29 August 2024

Joint work with David Svensson (AstraZeneca), Bohdana Ratitch (Bayer), and Alex Dmitrienko (Mediana)

Outline

- A causal framework for heterogeneous treatment effects (HTE)
- Four general approaches for estimating HTEs
- What to look at in papers on HTE evaluation?
- Post-selection inference on HTE
- Summary

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.

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

Received: 20 November 2023 | Revised: 28 May 2024 | Accepted: 21 June 2024

DOI: 10.1002/sim.10167

TUTORIAL IN BIOSTATISTICS

Statistics in Medicine WILEY

Modern approaches for evaluating treatment effect heterogeneity from clinical trials and observational data

Ilya Lipkovich[1] | David Svensson[2] | Bohdana Ratitch[3] | Alex Dmitrienko[4]

[1]Advanced Analytics and Access Capabilities, Eli Lilly and Company, Indianapolis, Indiana, USA

[2]Statistical Innovation, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

[3]Clinical Statistics and Analytics, Research & Development, Pharmaceuticals, Bayer Inc., Mississauga, Ontario, Canada

[4]Department of Biostatistics, Medtana, San Juan, Puerto Rico, USA

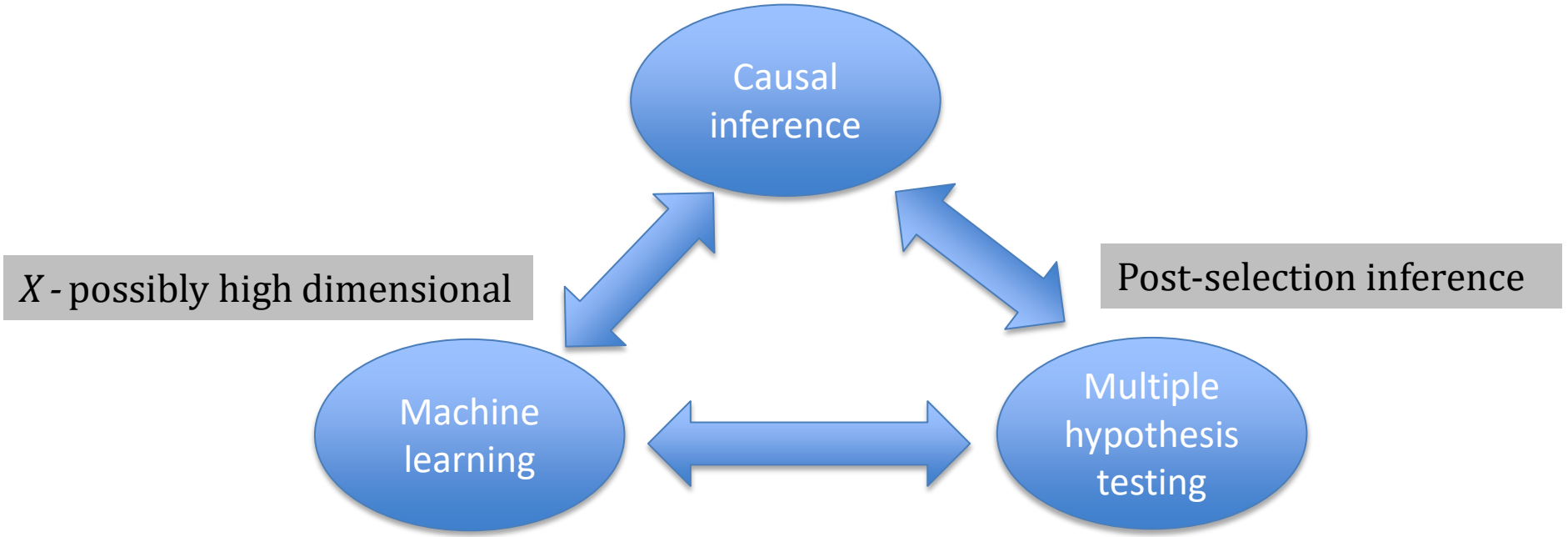
Correspondence
Ilya Lipkovich, Eli Lilly and Company, Indianapolis, IN 46285, USA.
Email: ilya.lipkovich@lilly.com

In this paper, we review recent advances in statistical methods for the evaluation of the heterogeneity of treatment effects (HTE), including subgroup identification and estimation of individualized treatment regimens, from randomized clinical trials and observational studies. We identify several types of approaches using the features introduced in Lipkovich et al (*Stat Med* 2017;36: 136-196) that distinguish the recommended principled methods from basic methods for HTE evaluation that typically rely on rules of thumb and general guidelines (the methods are often referred to as common practices). We discuss the advantages and disadvantages of various principled methods as well as common measures for evaluating their performance. We use simulated data and a case study based on a historical clinical trial to illustrate several new approaches to HTE evaluation.

KEYWORDS
individualized treatment regimen, personalized medicine, subgroup identification

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)

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)

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

- In studies with non-randomized treatments, we need to estimate propensity scores

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

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.

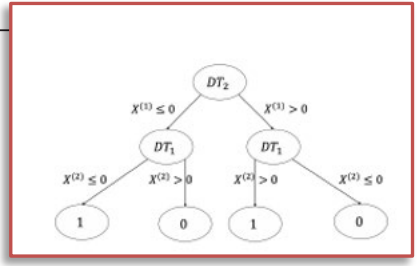
Keywords: Effect modification; Interaction; Optimal treatment selection; Precision medicine; Personalized treatment; Randomized trial; Subgroup

(*Epidemiology* 2019;30: 334–341)

treatment across sub
covariates.[1–6] Such an
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

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



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

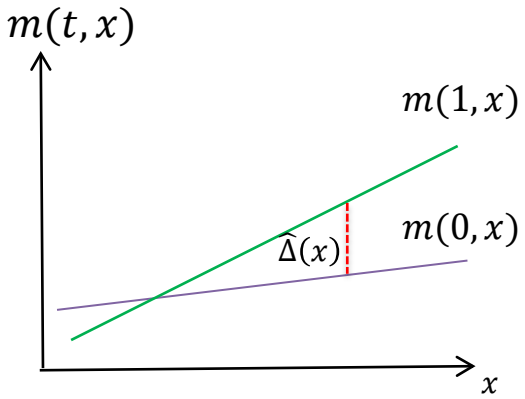
Henry W. J. Reeve, Timothy I. Cannings and Richard J. Samworth
University of Bristol, University of Edinburgh
and University of Cambridge

Abstract

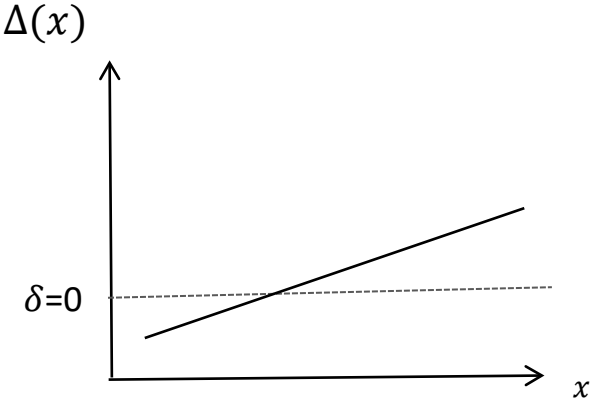
In clinical trials and other applications, we often see regions of the feature space that appear to exhibit interesting behaviour, but it is unclear whether these observed phenomena are reflected at the population level. Focusing on a regression setting, we consider the subgroup selection challenge of 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.

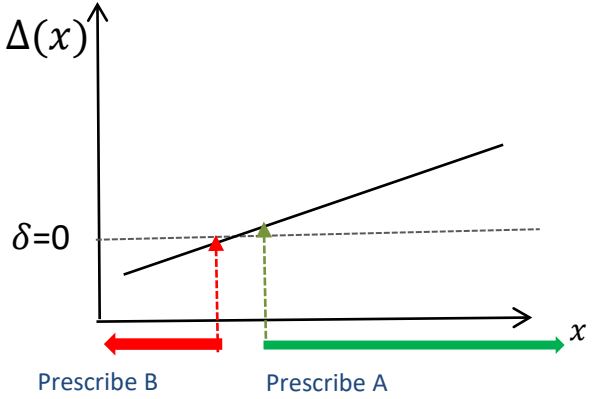
Typology of Subgroup Identification; Lipkovich et al. (2017)



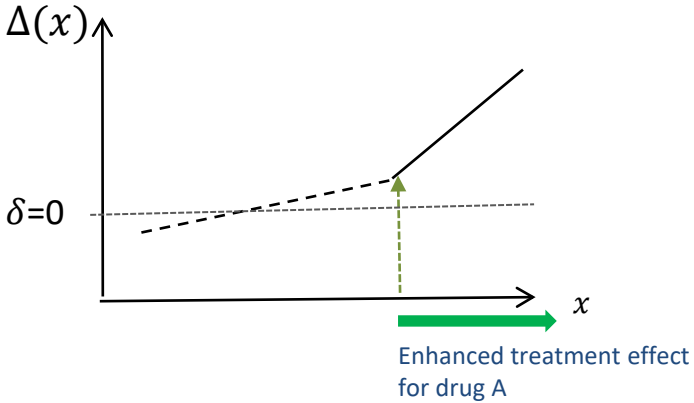
Global outcome modeling: Y



Direct treatment effect modeling



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



Local treatment effect modeling : Subgroup search

ITE scores vs CATE learners

- It is important to distinguish between estimators of CATE, $\widehat{\Delta}(x)$ (often presented as meta-learners, coined by Künzel et al) and an individual treatment effect (ITE) score $\widehat{\Delta}_i$ estimated for a given subject in observed data
  - $\widehat{\Delta}(x)$ predict Δ_i for any subject by plugging-in their X_i
  - Computing scores $\widehat{\Delta}_i$ require both X_i and Y_i for a given subject, they are consistent estimators of ITE, $E\{\widehat{\Delta}_i\} = \Delta_i$ and are used as pseudo-outcomes to model CATE
- Examples of ITE scores
  - Imputed/matched counterfactuals : $\widehat{\Delta}_i^{imp} = T_i\left(Y_i - \tilde{Y}_i(0)\right) + (1 - T_i)\left(\tilde{Y}_i(1) - Y_i\right)$
  - IPW score: $\widehat{\Delta}_i^{ipw} = \frac{T_i Y_i}{\hat{\pi}(X_i)} - \frac{Y_i(1-T_i)}{1-\hat{\pi}(X_i)}$, overoptimistic hopes: no need to fit prognostic effects
  - AIPW score: $\widehat{\Delta}_i^{aipw} = \hat{m}_1(X_i) - \hat{m}_0(X_i) + \frac{T_i(Y_i-\hat{m}_1(X_i))}{\hat{\pi}(X_i)} - \frac{(1-T_i)(Y_i-\hat{m}_0(X_i))}{1-\hat{\pi}(X_i)}$
  - Robinson's transformation: $\widehat{\Delta}_i^{rob} = \frac{Y_i-\hat{m}(X_i)}{T_i-\hat{\pi}(X_i)}$, $\hat{m}(X_i) = \hat{E}(Y_i|X_i)$, motivated R-learning

What to look for in papers on HTE?

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 (aka predictive biomarkers), making model selection more challenging
  - Confounders are predictive of both treatment T and outcome Y
  - Effect modifies are predictive of CATE, $\Delta(x)$

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; Han et al)
- $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, $\hat{S}(x) = \{x: \hat{\Delta}(x) > c\}$
  - Subgroups defined by simple biomarker signatures with up to 2 variables using a tree search, $\hat{S}(x) = \{x: X_1 \leq c_1, X_3 > c_3\}$
- Strategies often combine multiple steps and models. Fore example:
  - Compute ITE scores: $\hat{\Delta}_i$: Doubly robust score involving fitting outcome and propensity model, imputation of counterfactuals, e.g, by matching on propensity score or using ML, …
  - Fit a CART tree to $\hat{\Delta}_i$ as the pseudo-outcomes, and prune the tree
- How is model complexity controlled to prevent data overfitting?
  - Optimal tuning at each step does not guarantee optimal estimation of the targets causal estmand.

What output does the method produce?

- Individualized treatment contrast, $\widehat{\Delta}(x)$
- Biomarker signatures of promising subgroups
  - $\hat{\mathrm{S}}(x) = \{x: X_1 \leq c_1, X_3 > c_3\}$
- 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) e.g. selected by variable importance score.

What inference is done, if at all?

- Inference on presence of HTE: H_0: $\Delta(x) = \Delta$
- Inference on $\Delta(x)$
- Inference on subpopulations:
  - Controlling the probability of selecting the right subgroup, $\hat{S}(x)$ vs $S_{true}(x)$
  - Estimating "honest effect" in identified subgroup: $E\{\mathrm{Y}(1) - \mathrm{Y}(0) | \hat{S}(x)\}$
- Inference on ITR
  - Estimating the Value of ITR: $\mathrm{V}\left[\widehat{D}\right] = E\left\{Y\left(\widehat{D}(X)\right)\right\}$ (Qian and Murphy)
- Inference on selection of predictive biomarkers
  - E.g. controlling FDR via knockoffs (Sechidis et al.)

Inference on presence of HTE

- Best linear projection (BPL) of an ML proxy for CATE, $\widehat{\Delta}(X_i)$ (Chernozhukov et al, `GenericML`; Athey and Wager, `grf`)

$$Y_i - \hat{m}^{-i}(X_i) = \alpha\overline{\Delta}\left(T_i - \hat{\pi}^{-i}(X_i)\right) + \beta\left(T_i - \hat{\pi}^{-i}(X_i)\right)\left(\widehat{\Delta}^{-i}(X_i) - \overline{\Delta}\right)$$

  - Use cross-fitted versions of outcome and propensity models
  - $\beta > 0$ inidicates presence of heterogeneity of treatment effect
- GATE (Group ATE) testing (Chernozhukov et al; Imai and Li)
  - Null hypothesis: $E(\Delta(X)|G_1) \ldots = E(\Delta(X)|G_K)$, where G_K are groups induced by a generic ML method for estimating CATE.
  - Imai and Li developed cross-validation (cross-fitting) framework to test the homogeneity hypothesis (`evalITR`)
  - They derived the asymptotic variance for the test statistics under cross-fitting framework for an arbitrary ML algorithm for estimating CATE,
- Variation in CATE over covariate space, $VTE = var\{\Delta(X)\}$,
  - Levy et al. developed a cross-validated TMLE estimator with simultaneous inference for ATE and VTE

Inference on $\Delta(x)$

- Pointwise CI for $\Delta(x)$
  - based on post-selection inference from penalized regression, lasso with trt by covariate interaction terms (Ballarini et al)
  - based on causal random forests (Wager and Athey):
    - combining the ideas of R learning (Nie and Wager motivated by Double ML of Chernozhukov et al) with the inference for bagging and RF (Wager and Efron)
- Simultaneous confidence bands on $\Delta(x)$
  - by semi-parametric modeling, Guo at al.
  - using nonparametric kernel estimators of CATE, Lee et al. proposed 2 stage modeling:
    - 1 stage: High dimensional modeling of nuisance functions to compute DR ITE scores, $\widehat{\Delta}_i$
    - 2 stage: Use a much smaller number of candidate effect-modifiers $X_0 \in X$ to model CATE by regressing $\widehat{\Delta}_i$ on X_0.
- Bayesian approaches for inference on $\Delta(x)$
  - BART (Hill, `bartCausal`) and Bayesian causal forest (Hahn et al, `bcf`)

Inference on identified subgroups: what's the right subgroup?

- Controlling the probability of selecting the right subgroups (Schnell et al)
  - $S_{true}(X) = \{x: \Delta(x) > \delta\}$, e.g. $\delta = 0$
  - Bayesian credible subsets, $\Pr(\hat{S}_{lower} \subseteq S_{true} \subseteq \hat{S}_{upper}) > 1 - \alpha$
  - Bounding subgroups:
    - $\hat{S}_{lower}(X) = \{x: \widehat{\Delta}_{lower}(x) > \delta\}$, exclusive set
    - $\hat{S}_{upper}(X) = \{x: \widehat{\Delta}_{upper}(x) \geq \delta\}$, inclusive set
      - $\hat{S}_{lower}(X) \equiv \emptyset$ implies lack of heterogeneity
- Placing a guarantee on a set of subjects suggests testing for positive treatment effect at an individual patient level: $\mathrm{H_{0i}}$: $\Delta_i = 0$ (Duan et al.)
  - How do we interpret the collection of patients for whom we reject the null? Generalizability?

Inference on subgroups: what's effect within subgroup?

- Inference on treatment effect within identified subgroups, $E(\Delta(X)|\hat{S}(X))$
  - Bayesian shrinkage and Bayesian Model Averaging
- Resampling methods:
  - Correcting for overoptimism bias incurred by subgroup search with a ML algorithm. Subgroups identified in the resampled set may be different from those on the original set
  - Correcting for selection of the best subgroup within a pre-specified set of candidate subgroups: e.g.: $S(c) = \{X \leq c\}$ via bootstrap (Guo and He)
  - Combining the two frameworks: debiased lasso + bootstrap adjustment (Guo at al.)
- Inference on data-driven subgroups without resampling or a test data?
  - Subgroup search on the full sample while masking some aspects of the data
  - e.g. tree-based search based on squared ITE scores, $\widehat{\Delta}_i^2$ while using the known distribution of the sign ($\widehat{\Delta}_i$) under null for controlling Type 1 error/FDR. (Hsu et al; Karmakar et al)

Inference on ITR

- Estimating value of $\mathrm{V}\left[\widehat{D}\right] = E\{Y(\widehat{D}(X))\}$ is a challenging and irregular problem, even for a single stage ITR
  - Important distinction: inference for the value of estimated ITR, $\mathrm{V}\left[\widehat{D}\right]$ vs. inference for the value of true/optimal ITR, $\mathrm{V}\left[D_{opt}\right]$
  - TMLE estimator for the "Mean under Dynamic Treatment Regimen" by van der Laan et al. Inference is based on cross-fitted Efficient Influence Curves,
  - provides a guarantee that their 95% CI for the Value function covers the true $\mathrm{V}\left[D_{opt}\right]$
- The cross-validation (cross-fitting) framework for estimating Population Average Prescriptive Effect (PAPE) from randomized trials (Imai and Li)
  - PAPE contrasts the value of a regimen under budget constraint p with the benchmark of the value under randomly assigning $p\%$ patients to active treatment.

$$PAPE(p) = E\left\{Y\left(D_p(X)\right)\right\} - E\{pY(1) + (1-p)Y(0)\},$$

  - $D_p(X) = I(\widehat{\Delta}(X) > \delta(p))$, $\delta(p)$ is calibrated to ensure the budget constraint p (p =proportion treated) is met, and no patient is harmed, $\delta(p) \geq 0$.

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 Subgrou 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).

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 response surface $m(t, x)$) regularization bias can be large, as each step is optimized for prediction, rather than for the final estimation target. Doubly robust strategies for CATE are preferred.
- Post-selectin inference on HTE is challenging. We reviewed some recent methods, mostly within frequentist domain

References

Athey S, Tibshirani J, Wager S. (2019). Generalized random forests. The Annals of Statistics. 47(2): 1148–1178.

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 (2020). Generic machine learning inference on heterogenous treatment effects in randomized experiments. arXiv:1712.04802v4.

Duan B, Wasserman L, Ramdas A. (2023) Interactive identification of individuals with positive treatment effect while controlling false discoveries. arXiv: 2102.10778v2

Guo W, Zhou XH, 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. (2020). Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects. Bayesian Anal. 15(3), 965-1056.

Hsu JY, Zubizarreta JR, Small DS, Rosenbaum PR. (2015). Strong control of the familywise error rate in observational studies that discover effect modification by exploratory methods. Biometrika. 102(4):767-782.

Imai K, Li ML. (2022). Statistical inference for heterogeneous treatment effects discovered by generic machine learning in randomized experiments. arXiv:2203.14511v1 2022.

Imai K, Li ML. (2023) Experimental evaluation of individualized treatment rules. J Am Stat Assoc.;118(541):242-256.

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

References (cont.)

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.

Lee S, Okui R, Whang YJ. (2017) Doubly robust uniform confidence band for the conditional average treatment effect function. J Appl Econ. 32(7):1207-1225

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.

Lipkovich I, Svensson D, Ratitch B, Dmitrienko A. (2024) Modern approaches for evaluating treatment effect heterogeneity from clinical trials and observational data. Statistics in Medicine. 1-49. doi:10.1002/sim.10167

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

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

Sechidis K, Kormaksson M, Ohlssen D. (2021) Using knockoffs for controlled predictive biomarker identification. Stat Med.;40(25):5453-5473.

Qian M, Murphy S. (2011). Performance guarantees for individualized treatment rules. Annals of Statistics. 39(2): 1180–1210.

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

Thank you!

Q & A

Ilya.Lipkovich@lilly.com