Overview of methods for subgroup and biomarker identification from clinical data

Ilya Lipkovich, PhD
Eli Lilly and Company

OUTLINE

- Data mining/machine learning vs classical statistics
- Why data mining/machine learning for clinical data?
- Predictive versus prognosis effects
- Data-driven versus “guidance driven” subgroup analysis
- Taxonomy of biomarker identification methods
- Software for subgroup identification. What features to look at?
- Summary

DATA MINING/MACHINE LEARNING VS CLASSICAL STATISTICS

Classical statistics	Data mining/Machine learning
Relatively small data sets collected from designed experiments or by sampling from well–defined populations	Large (sometimes dispersed and heterogeneous) data sets, often collected for purposes other than data mining
Assumes data generation mechanism: $y = f(\boldsymbol{x}) + \varepsilon$, where $f(\boldsymbol{x})$ has a simple structure and the error term (ε) is modeled by parametric distributions	Aims at recovering unknown function $f(\boldsymbol{x})$ as a "black box" while the presence of the "error term" is often ignored
Objective is to estimate parameters for the entire population from available sample(s)	Objective is to obtain predictions for new (future) cases [supervised] or extract useful features that reveal unknown structure [unsupervised]. Analysis data often represent the entire population
Focus on hypothesis testing: a single test or a small number of pre-specified tests with clearly defined multiplicity control procedures	Hypothesis generation/knowledge discovery rather than formal hypothesis testing, less emphasis on statistical significance (often focusing on controlling the false discovery rate)

DMML LERANING FOR CLINICAL DATA: SUPERVISED LEARNING

- Patient diagnostics
  - Example: tree-based decision rules that allowed clinicians of an emergency unit to make a quick assessment whether a patient with chest pain can be diagnosed with a myocardial infraction
- Predictive models for patients' future outcomes
  - Prediction models for safety or efficacy outcomes, informed by assigned treatment, biomarkers available prior to treatment initiation, and evolving (early) patient outcomes
- Modeling as part of treatment evaluation strategies
  - Examples: modeling to account for selection bias due to post randomization/intercurrent events (e.g. modeling dropouts or implementing multiple imputation)

DMML FOR CLINICAL DATA: UNSUPERVISED LEARNING

- Clustering to identify patients with similar efficacy outcomes
  - Especially relevant for diseases where the patients' well-being is described by a set of variables representing complementary and sometimes conflicting clinical criteria and scales (neuroscience)
- Identifying patients with distinct response profiles (or trajectories) over time
  - Response profiles may represent different types of patients (e.g., "early responders who later fail," "relapsers," "gradual responders," "sustained responders," etc.)
- Methods for association learning
  - Example: in pharmacovigilance to uncover drug-adverse event relationships and drug-drug interactions in spontaneous reporting systems and large healthcare databases
- Detecting outliers and unusual patterns
  - Often used in the context of fraudulent assessment of outcomes (e.g. see O'Kelly, 2004)

DMML FOR CLINICAL DATA: UNSUPERVISED LEARNING

- Clustering to identify patients with similar efficacy outcomes
  - Especially relevant for diseases where the patients' well-being is described by a set of variables representing complementary and sometimes conflicting clinical criteria and scales (neuroscience)
- Identifying patients with distinct response profiles (or trajectories) over time
  - Response profiles may represent different types of patients (e.g., "early responders who later fail," "relapsers," "gradual responders," "sustained responders," etc.)
- Methods for association learning
  - Example: in pharmacovigilance to uncover drug-adverse event relationships and drug-drug interactions in spontaneous reporting systems and large healthcare databases
- Detecting outliers and unusual patterns
  - Often used in the context of fraudulent assessment of outcomes (e.g. see O'Kelly, 2004)

DMML FOR CLINICAL DATA: SEMI-SUPERVISED LEARNING

- Subgroup identification
  - Identifying subgroups of patients with differential treatment effect from clinical trials data (e.g. from failed Phase 3 trials or from early phase trials with the idea of using for enrichment in subsequent stages of clinical programs)
- Estimating optimal individual treatment regimes
  - Construction of optimal dynamic treatment regimes (DTRs) utilizing information on patient's characteristics and accumulated patient's outcomes at each decision point

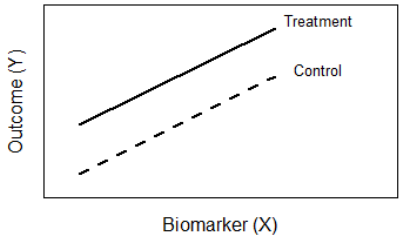
*Unlike in supervises learning the outcome here is **individual** treatment difference/contrast that is not observable. Unless the same patient is taken all candidate treatments, only one potential treatment outcome is observed per patient.*

PREDICTIVE VERSUS PROGNOSTIC BIOMARKERS

- The task of personalized medicine can be "translated" into statistical language as constructing predictive biomarker signature that would allow identifying patients with differential treatment response
- The schematic plots show four types of relationships between the outcome and a single biomarker

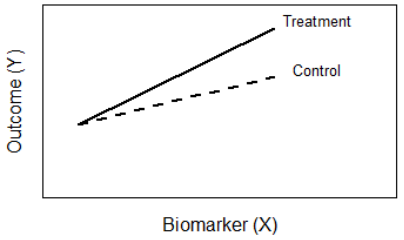
X is prognostic but not predictive

(a)



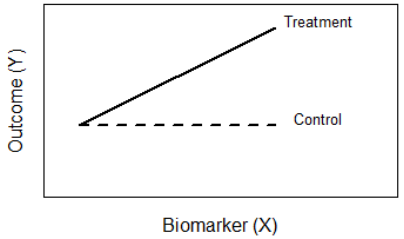
X is prognostic and predictive

(b)



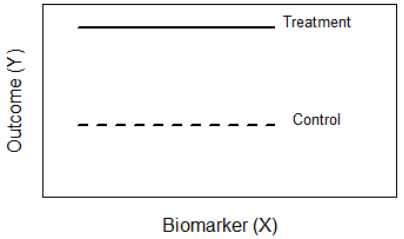
X is predictive but not prognostic

(c)



X is neither prognostic nor predictive

(d)



SUBGROUP ANALYSIS GUIDELINES

- Subgroup analyses are often (rightfully) viewed as data dredging
- Many authors came up with various “checklists” of principles for Subgroup Analyses
  - NHS R&D HTA Programme (Brookes et al. 2001) provides a list of 25 recommendations
  - Rothwell (2005) proposed a guideline with 21 rules
  - Sun et al (2009) listed the existing 7 plus 4 additional criteria for assessing credibility of subgroup analysis
- EMA Guideline on the Investigation of Subgroups in Confirmatory Clinical Trials (Draft, Jan 2014)
  - Recognizes issues with current SA practices that “create disincentive to properly plan the investigation of subgroups”
- The Guidelines encourage to “exercise caution” when conducting subgroup analyses, which is hard to operationalize …

DATA-DRIVEN VS. "GUIDELINE-DRIVEN" APPROACH

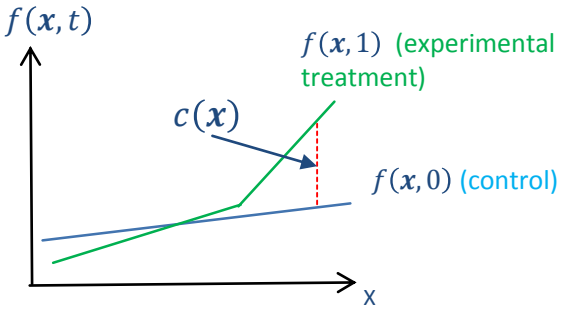
- "Guideline-driven" approach fails to encompass modern scientific approaches to statistical learning and the need for evidence-based personalized/stratified/precision medicine
- A different view: subgroup identification/analysis is framed as a special case of model selection
- This helps link subgroup identification efforts with the wealth of statistical methodology on model selection
- Pre-specified is the entire biomarker/subgroup selection strategy, not specific subgroup(s)

WHAT MAKES DATA-DRIVEN SA STRATEGIES "PRINCIPLED"?

- "Complexity control" to prevent data overfitting
  - Tuning parameters controlling the search process need to be determined often in a data-driven fashion, e.g., via cross-validation
  - E.g., penalized regression, a.k.a. shrinking, regularization
- Evaluating the type I error rate for the entire subgroup search strategy
  - E.g., using resampling under null
- Obtaining "honest" estimates of treatment effect in subgroups (i.e. treatment effect expected in identified subgroups if applied to future studies)
  - E.g., by using resampling methods or Bayesian model averaging/empirical Bayes
  - Uncertainty associated with the entire strategy should be accounted for

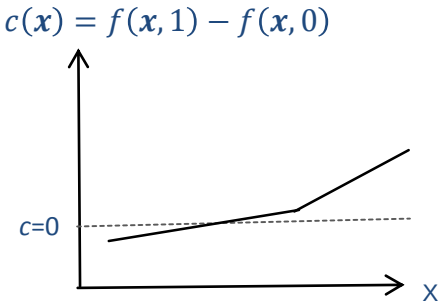
GLOBAL OUTCOME MODELING

- Modeling underlying outcome function $f(\boldsymbol{x}, t) = E(Y|\boldsymbol{X} = \boldsymbol{x}, T = t)$, where Y is an outcome, X is a collection of biomarkers and T=0,1 is a treatment indicator
  - computing individual treatment differences $\hat{c}_i = \hat{f}(\boldsymbol{x}_i, 1) - \hat{f}(\boldsymbol{x}_i, 0), i = 1, \dots, N$, that can be further modeled as an outcome
  - allows constructing a predictive score as a function of biomarkers, a biomarker signature: $c(\boldsymbol{x})$
- Some recent methods
  - Virtual Twins by Foster, Taylor and Ruberg (2011) [combining Random Forest for $f(\boldsymbol{x}, t)$ and CART for $c(\boldsymbol{x})$]
  - Penalized regression (FindIT) by Imai and Ratkovic (2013)
  - Bayesian hierarchical modeling (Jones et al, 2011 extending Dixon and Simon, 1991)



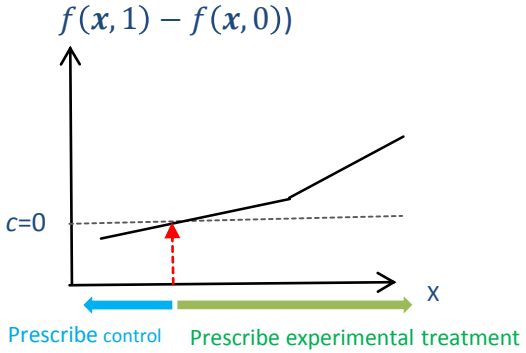
GLOBAL TREATMENT EFFECT MODELING

- Directly modeling underlying treatment effect, $c(\boldsymbol{x})$
  - Classification and regression tree methods can be adopted by incorporating treatment variable in the splitting criterion, resulting in piecewise constant fit for $c(\boldsymbol{x})$
  - Parametric models were proposed that obviate the need for fitting in prognostic effects
- Some recent methods
  - Interaction trees (Su et al., 2005)
  - Gi method (Loh et al., 2015) (implemented within GUIDE suite)
  - Model-based recursive partitioning (Seibold et al., 2014).
  - Modified covariate method by Tian et al. (2014)



MODELING INDIVIDUAL TREATMENT REGIMES

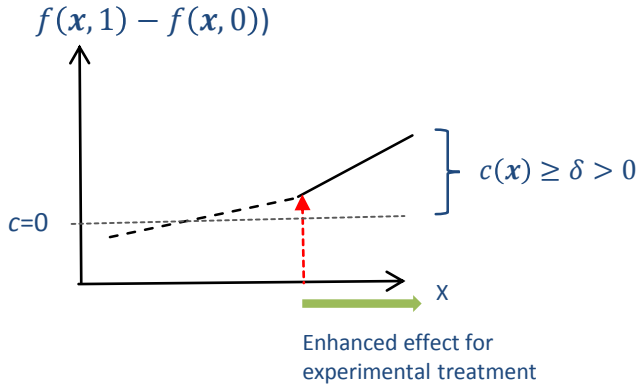
- Estimating optimal treatment regime $sign[c(\boldsymbol{x})]$
  - Obviates the need to fit-in prognostic (main) effects, estimates optimal treatment regime by fitting a weighted classifier for treatment as a "response" with outcome-based weights $w(y, \boldsymbol{x}) = \frac{y}{Pr(T=t|\boldsymbol{X}=\boldsymbol{x})}$
  - Weights incorporate the probabilities of treatment which are known in RCT and can be obtained by modeling propensity of treatment assignment in observational (non-randomized) studies



- Some recent methods
  - Outcome weighted learning (OWL) introduced by Zhao et al. (2012); ROWSi method (Xu et al. 2015)

LOCAL TREATMENT EFFECT MODELING (SUBGROUP SEARCH)

- Identifying subgroups S with enhanced treatment effect $c(\boldsymbol{x}) > \delta$ for $\boldsymbol{x} \in S$
  - Instead of estimating the response function $c(\boldsymbol{x})$ in the entire covariate space first an then carving out the interesting part where $c(\boldsymbol{x}) > \delta$, these methods would directly search for such interesting regions
- Some recent methods:
  - Subgroup search methods of Kehl and Ulm (2006), Chen et al. (2015) (inspired by Bump Hunting a.k.a. PRIM by Fisher and Friedman, 1999)
  - SIDES (by Lipkovich et al., 2011) and SIDEScreen (Lipkovich and Dmitrienko, 2014)



WHAT FEATURES OF A SA METHOD WE SHOULD LOOK FOR?

- What is the number of candidate predictors that can be processed in efficient manner (p=1, 20, 100, 1000)?
- What is the "model space" induced by the procedure and how model complexity is controlled to prevent overfitting?
- What outputs does the method produce?
  - Signatures of promising subgroups
  - Personalized treatment contrast
  - Optimal treatment assignment
  - Predictive biomarkers ordered by predictive strength.
- How the false discovery is controlled, if at all (type I error control, FDR)
- Does the method provide "honest" estimates (point estimates, SE, CI) of treatment effect in identified subgroups corrected for over-optimism?
  - E.g. using cross-validation, bootstrap, Bayesian model averaging

SUMMARY OF SUBGROUP IDENTIFICATION METHODS

Method	Modeling type (1)	Dimen-sionality (2)	Biomarker selection (3)	Control of false positive rate (4)	Complexity control (5)	Selection control (6)	Honest estimate of treatment effect (7)	Software implemen-tation (8)
Global outcome modeling								
Virtual Twins (Foster et al. [42])	Freq/NP	High	P,S	No	Yes	No	Yes	B
Cai et al. [38]	Freq/NP	Low	P	Yes	No	No	Yes	
FindIt (Imai and Ratkovic [37])	Freq/P	High		No	Yes	No	No	C
STIMA (Dusseldorp, Concersano and Van Os, 2010)	Freq/NP	Medium	S	No	Yes	No	No	C
Bayesian approaches (Dixon and Simon [45]; Hodges et al. [46])	Bayes/P	Low	P	No	Yes	No	Yes	
Global treatment effect modeling								
Interaction Trees (Su et al. [70]; Negassa et al. [68])	Freq/NP	High	S	No	Yes	No	No	B
Gi as part of GUIDE (Loh et al. [71])	Freq/NP	Medium	S	No	Yes	Yes	Yes	C
Modified covariate method (Tian et al. [75])	Freq/P	High	P	No	Yes	No	No	
QUINT (Dusseldorp and Mechelen [74])	Freq/NP	Medium	S	No	Yes	No	No	C
Optimal treatment regimes								
Biomarker selector (Gunter et al. [21])	Freq/P	High	B	Yes	Yes	No	No	
Qian and Murphy [78]	Freq/P	High	P,T	No	Yes	No	No	
Zhao et al. [85], Xu et al. [88]	Freq/P	High	P,T	No	Yes	No	No	B
Zhang et al. [86]	Freq/SP	High	T	No	Yes	No	No	
Local modeling								
Adaptation of PRIM (Chen et al [93]; Kehl and Ulm [90])	Freq/NP	High	S	No	Yes	No	No	P
SIDES (Lipkovich et al. [91]) and SIDEScreen (Lipkovich et al. [94])	Freq/NP	Medium	B,S	Yes	Yes	Yes	Yes	B
Berger et al. [25], Sivaganesan et al. [95]	Bayes/NP	Medium	S	Yes	Yes	No	Yes	P

Lipkovich, Dmitrienko, D'Agostino. Tutorial in biostatistics… 2016

SOFTWARE FOR SUBGROUP IDENTIFICATION

- Site maintained by QSPI Subgroup analysis industry group sponsored by the society of clinical trials
- 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 package

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.

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 Tool**kit** for Recursive **Party**tioning, 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 sources:

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.

SUMMARY

- Data mining/machine learning methods are becoming an integral part of data analysis at all stages of clinical drug development, which can be contrasted with its primary use in pre-clinical stage of “drug discovery” in the past
- We emphasize principled or disciplined use of subgroup identification (and data mining) as opposed to haphazard data-dredging and treat subgroup identification as a special case of model selection and contrast data-driven with guideline-driven approach
- Unlike standard predictive modeling methods that aim at identifying subgroups with heterogeneous outcome, using methods for tailoring/personalized medicine requires modeling individual treatment differences targeting subgroups with heterogeneous treatment effect
- Methods for subgroup identification and analysis borrow from diverse literature in machine learning, multiple testing and causal inference
- A feature of subgroup identification (and data mining in general) in drug development is the need to control the Type I error (or false discovery) rates which is a new trend in the area of machine learning
- Once subgroups have been identified, analyst is facing the challenge of obtaining “honest” estimates for associated effects that should be expected in the future data

REFERENCES

- Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G. Subgroup analyses in randomized controlled trials: quantifying the risks of false-positives and false-negatives. Health Technology Assessment 2001, 5(33)
- Chen, G., Zhong, H, Belousov, A. Viswanath, D. PRIM Approach to Predictive-signature Development for Patient Stratification. Statistics in medicine 2015. 34:317-342.
- Foster JC, Taylor JMC, Ruberg SJ, Subgroup identification from randomized clinical trial data, Statistics in medicine 2011, 30:2867–2880.
- Jones HE, Ohlssen DI, Neuenschwander B, Racine A and Branson M. Bayesian models for subgroup analysis in clinical trials. Clinical Trials 2011; 8:129–143.
- Kehl V and Ulm K. Responder identification in clinical trials with censored data. Computational Statistics & Data Analysis 2006, 50:1338-1355.
- Lipkovich I, Dmitrienko A, Denne J, Enas G. Subgroup identification based on differential effect search (SIDES): A recursive partitioning method for establishing response to treatment in patient subpopulations. Statistics in Medicine 2011, 30:2601-2621.
- Lipkovich I, Dmitrienko A. Strategies for identifying predictive biomarkers and subgroups with enhanced treatment effect in clinical trials using SIDES. Journal of Biopharmaceutical Statistics 2014, 24:130-153.
- Lipkovich I, Dmitrienko A, D'Agostino BR. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. Statistics in Med 2017. 36:136-196.
- Loh, W-Y., He, X., Man, M. A regression tree approach to identifying subgroups with differential treatment effects. Statistics in medicine 2015. 34:1818-1833.
- Rothwell PM. Subgroup analysis in randomized controlled trials: importance, indications, and interpretation. Lancet 2005; 365:176–86
- Seibold H, Zeileis A, Hothorn T. Model-based recursive partitioning for subgroup analyses. International journal of Biostatistics 2016; 12:45-63.
- Su X, Tsai CL, Wang H, Nickerson DM, Li B. Subgroup analysis via recursive partitioning. Journal of Machine Learning Research 2009; 10:141-158.
- Sun X, Briel M, Walter SD, Guyatt GH. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses, *BMJ* 2010; 340:c117doi: 10.1136/bmj.c117
- Tian L, Alizaden AA, Gentles AJ, Tibshirani R. A Simple Method for Detecting Interactions between a Treatment and a Large Number of Covariates *Journal of the American Statistical Association* 2014; **109**:1517–1532.
- Xu Y, Yu M, Zhao YQ, Li Q, Wang S, Shao J. Regularized outcome weighted subgroup identification for differential treatment effects. Biometrics 2015. 71:645-653.
- Zhao Y, Zheng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. Journal of the American Statistical Association 2012, 107:1106-1118.

THANK YOU!

Lipkovich_Ilya_A@Lilly.com