Exploratory Subgroup Analysis: Subgroup Identification Approaches in Clinical Trials

Ilya Lipkovich and Alex Dmitrienko, Quintiles

JSM, August 6, 2013

Outline

- Introduction
  - Principles and standards for “Subgroup Analysis” in clinical research
  - Data-driven versus guidance-driven approaches to subgroup analysis
  - Taxonomy of data mining methods for subgroup identification
- SIDES approach
  - Basic SIDES algorithm and SIDES parameters
  - Clinical trial example
  - SIDEScreen, an extension of basic SIDES method
- Summary and Discussion
- References

Principles and Standards for Subgroup Analysis in Clinical Research

- Subgroup analyses are often (rightfully) viewed by regulatory agencies as a way for the trial sponsor to make unsubstantiated claims about efficacy in subpopulations by *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
- General theme
  - Subgroups need to be pre-specified, biologically plausible, significance tests multiplicity adjusted, "*no testing in subgroup unless interaction test is significant*"; sometimes stating that "*no testing in subgroup if the overall effect is not significant*", …, and finally "*interpreted with caution*"
- From D. Berry 1990 note on Subgroup Analyses: "... *there's something unscientific about requiring hypotheses to be specified in advance. Science would proceed very slowly if scientists never took data at face value*."

Data-driven versus Guidance –Driven Approach for Subgroup Identification

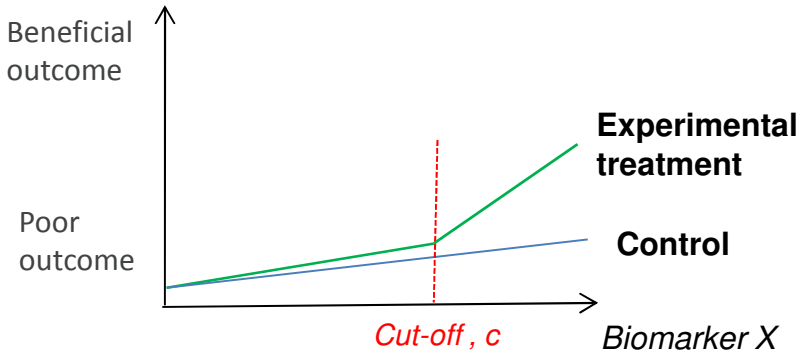
- Subgroup Identification is framed as a special case of Model Selection
- Pre-specified is the entire biomarker/subgroup selection strategy, not the specific subgroup(s) which need to be found!

Data Mining Methods for Subgroup Identification

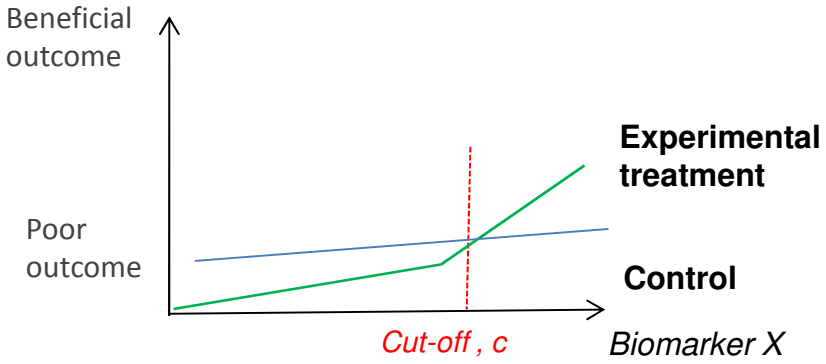
- Notation
  - Outcome (Y), treatment (T) (e.g. t_i=1 if i-th subject is assigned to "Drug" and t_i=0 if assigned to "Placebo"), various subject characteristics measured prior to treatment, $X_1,..,X_p$ combined in $\mathbf{x}=\{x_{i1},..,x_{ip}\}$
  - Expected response for the i-th subject is $f(\mathbf{x}_i,t_i)$
  - Expected potential outcomes: $f(\mathbf{x}_i,1)$ and $f(\mathbf{x}_i,0)$
  - Treatment contrast $z(\mathbf{x}_i) = g(f(\mathbf{x}_i,1), f(\mathbf{x}_i,0))$
    - E.g. $z(\mathbf{x}_i) = f(\mathbf{x}_i,1) - f(\mathbf{x}_i,0)$

Two Frameworks of Personalized Medicine

- Identifying the right subject for a given treatment
  - therapy provides minimal or no benefit in the overall population
  - find subpopulations that will have enhanced benefit from the treatment vs. control
- Identifying the right treatment for a subject
  - finding optimal treatment regime or policy for a given subpopulation.



The case of quantitative interaction



The case of qualitative interaction

Taxonomy of Data Mining Methods for Subgroup Identification

- Global outcome modeling
  - modeling underlying outcome function $f(\mathbf{x},t)$
- Global treatment effect modeling
  - modeling underlying treatment effect, $z(\mathbf{x})$.
- Local treatment effect modeling (subgroup search)
  - identifying subgroups $\{\mathbf{x} \in S\}$ with higher values of $z(\mathbf{x})$

Taxonomy of Subgroup Identification: Global outcome modeling $f(\mathbf{x},t)$

- Typically requires a non-parametric "black box" model for f(**x**,t)
- Examples:
  - Virtual Twins method of Foster et al., 2011: estimating f(**x**,t) via Random Forest at the first stage, then using CART to predict estimated treatment contrast z(**x**)= f(**x**,1)-f(**x**,0) at the second stage.
  - Parametric approaches of subgroup analysis via positing a model with main effects and interactions and shrinking coefficients associated with $T*X_j$ interaction terms.
    - Bayesian hierarchical models (e.g. Jones et al., 2011)
    - frequentist analysis via penalized likelihood (LASSO, elastic net, etc., e.g. Imai & Ratcovic, 2013)

Taxonomy of Subgroup Identification: Global treatment effect modeling, z(**x**)

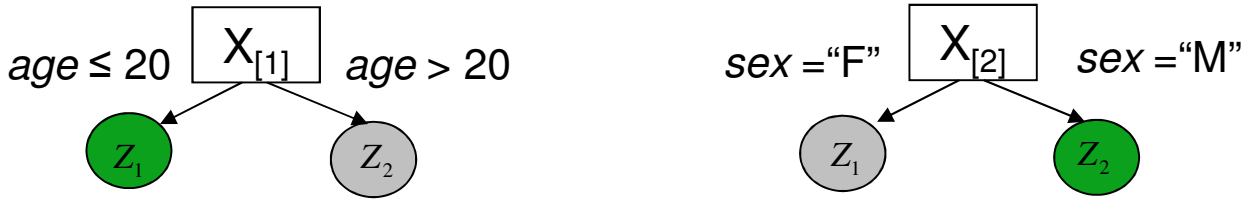
- Advantage: obviates the need to model prognostic (main) effects and focuses on predictive (T*X interaction) effects
- Examples:
  - Interaction tree method of Su et al., 2009.
  - "Modified covariates" method of Tian et al., 2012.
  - Estimating optimal treatment regimes (OTR) via "outcome-weighted" classification.
    - Zhao et al., 2012 showed that minimizing weighted misclassification error for predicting treatment labels (T=0/1) leads to OTR, if weights are taken as $w_i=y_i/\pi_i$ for treated subjects and $w_i=y_i/(1-\pi_i)$ for control subjects
      - π_i = Prob(T_i =1), Y is continuous outcome with larger values indicating beneficial outcome

Taxonomy of Subgroup Identification: Local treatment effect modeling

- Identifying subgroups {**x** $\in$ S} with higher values of z(**x**)
  - Advantage: obviates the need to estimate the response function over the entire covariate space and focuses on identifying specific regions with a large differential treatment effect.
  - Examples:
    - Bump hunting approach proposed by Kehl and Ulm ,2006 (extending the PRIM methodology by Friedman and Fisher, 1999).
    - SIDES method by Lipkovich et al., 2011.
    - Bayesian subgroup analysis via model averaging (Berger et al, MCP 2011,2013).

SIDES - **S**ubgroup **I**dentification based on **D**ifferential **E**ffect **S**earch

- All (p) candidate covariates are ordered from "best" to "worst" in terms of a treatment effect-based splitting criterion
- Candidate splits per covariate are exhaustively searched and the criterion is evaluated at optimal split into two child sub-groups resulting (for each covariate) in one "promising subgroup" and one "non-promising"
- Retained are first k (e.g. k=2) covariates (rather than only the top one)

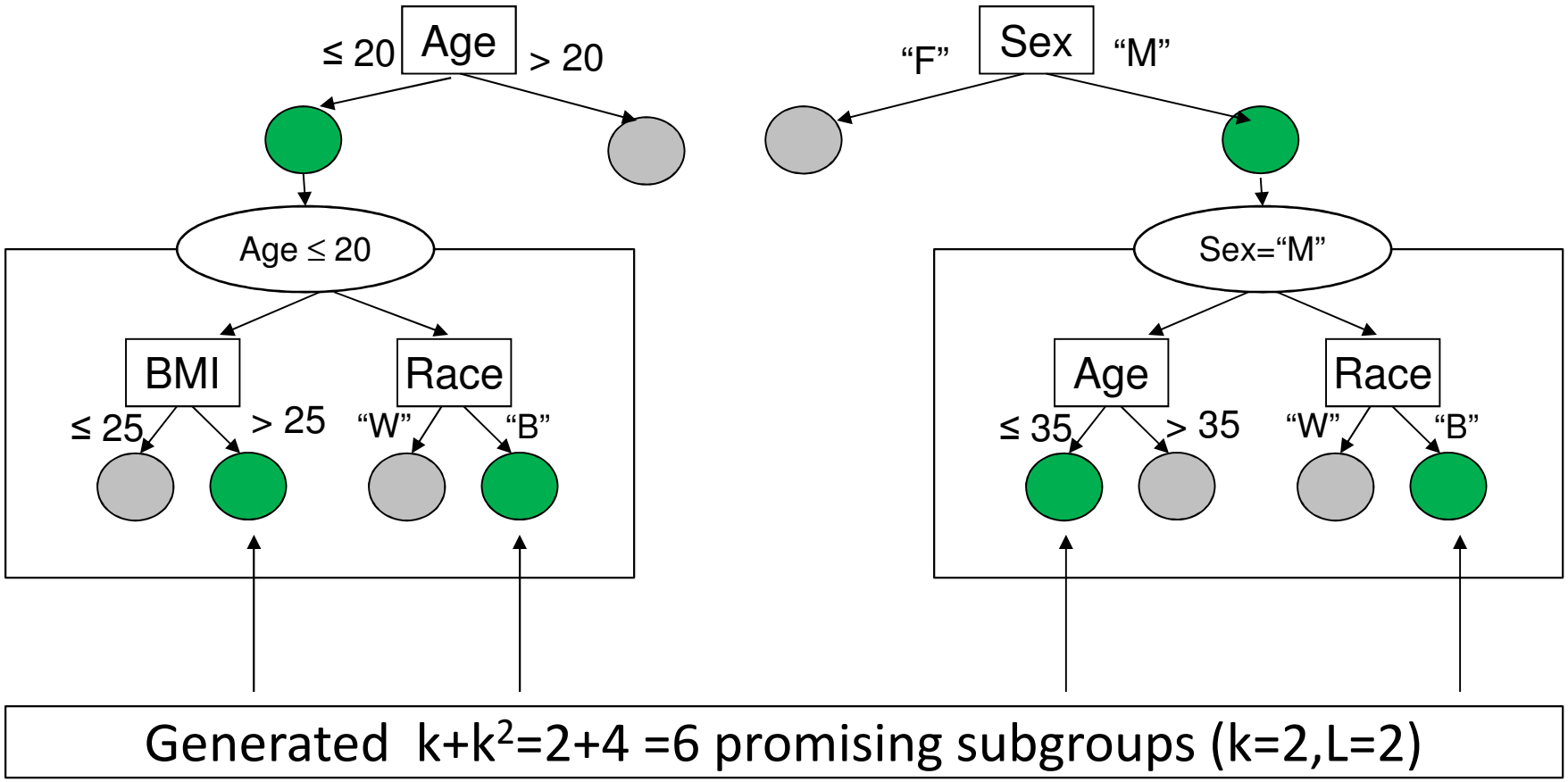


Z_1 and Z_2 are test statistics for testing H_0: treatment effect=0 in child nodes.

Differential splitting criterion = $1 - \Phi\left(|Z_1 - Z_2| / \sqrt{2}\right)$

Recursive Partitioning Algorithm

- Apply the same procedure recursively **up to L** times



SIDES Parameters

- Number of best candidate covariates pursued, "width" (e.g. k=3)
- Splitting criteria evaluated when forming a split (e.g. treatment by split interaction)
- The "depth", or number of levels in a subgroup (e.g. subgroup X1=0 & X2=0 has 2 levels), L=2
  - Note a subgroup of depth 3 corresponds to a 4-level treatment by covariate interaction!
- Complexity (tuning) parameter: relative improvement in treatment effect in child versus parent node, P(child)<γ P(parent) required to make a split. γ can be calibrated using cross-validation
- Type I error control
  - Construct reference distribution by permuting treatment labels and computing the proportion of null sets with smallest P-value $\leq$ P-value$_{obs}$
  - Multiplicity adjustment is not the end goal but is used in the process of subgroup selection in conjunction with tuning parameters to control the size of the search space and increase chances of subgroup replication in future trials

Running Example. Severe Sepsis Trial

- Testing for the difference in proportions in subjects who died within the first 28 days (assuming lower proportion is preferable)

Subgroup	Subgroup size	Test statistic (z)	1-tailed p value	treatment event rate	control event rate	treatment sample size	control sample size	effect size
All subjects (Analysis data)	578	-1.1266	0.8700	0.367	0.319	387	191	-0.100

Potential Covariates	Description
TIMFIRST	TIME FROM FIRST SEPSIS-ORGAN FAIL TO START DRUG
AGE	
BLLPLAT	BASELINE LOCAL PLATELETS
blSOFA	SUM OF BASELINE SOFA (CARDIOVASCULAR, HEMATOLOGY, HEPATICRENAL, RESPIRATION SCORES)
BLLCREAT	BASELINE CREATININE
ORGANNUM	NUMBER OF BASELINE ORGAN FAILURES
PRAPACHE	PRE-INFUSION APACHE-II SCORE
BLGCS	BASELINE GLASGOW COMA SCALE SCORE
BLIL6	BASELINE SERUM IL-6 CONCENTRATION
BLADL	BASELINE ACTIVITY OF DAILY LIVING SCORE

Applying SIDES to Data Example

- Apply SIDES with width (k=3), depth=up to 2 levels (L=2), and child-to-parent ratio (γ)=0.2
- Note the last column contains multiplicity-adjusted p-values

Subgroup	Subgroup size	Test statistic (z)	1-tailed p value	treatment event rate	control event rate	treatment sample size	control sample size	effect size	multiplicity -adjusted 1-sided p-value
TIMFIRST<=39 & AGE>59.126	195	2.4056	0.0081	0.380	0.561	129	66	0.364	0.416
TIMFIRST<=35.53 & AGE>59.126	185	2.2363	0.0127	0.376	0.550	125	60	0.351	0.474
TIMFIRST<=35.53 & BLADL>0	132	2.1336	0.0164	0.363	0.561	91	41	0.401	0.492
BLLPLAT>228	134	1.4422	0.0746	0.210	0.321	81	53	0.255	0.620

Evaluating Variable Importance

- Apply SIDES with very loose constrains to generate a large number of candidate subgroups (here m=55)
- Variable importance is computed for each biomarker as the average of its "contributions" over all subgroups where X was involved as the splitter
  - "X-Contributions" are –log(p) or = 0, if X is not involved
  - where p= trt-by-split interaction p-value

Subgroup	Subgroup size	Test statistic (z)	1-tailed p value	treatment event rate	control event rate	treatment sample size	control sample size	effect size
. < TIMFIRST<=39 & AGE>59.126 & blSOFA <=10	154	2.9138	0.0018	0.343	0.592	105	49	0.504
. < TIMFIRST<=39 & AGE>59.126 & ORGANNUM <=3	166	2.8063	0.0025	0.336	0.566	113	53	0.467
. < TIMFIRST<=35.53 & AGE>59.126 & blSOFA <=10	146	2.7346	0.0031	0.337	0.578	101	45	0.490
. < TIMFIRST<=39 & AGE>59.126 & BLLPLAT>92	157	2.6922	0.0035	0.366	0.589	101	56	0.449
. < TIMFIRST<=39 & AGE>59.126 & blSOFA>5	166	2.6414	0.0041	0.391	0.607	110	56	0.434
. < TIMFIRST<=35.53 & AGE>59.126 & ORGANNUM <=3	156	2.6104	0.0045	0.330	0.553	109	47	0.456
. < TIMFIRST<=35.53 & AGE>56.388 & ORGANNUM <=3	165	2.6101	0.0045	0.316	0.529	114	51	0.440
AGE>59.126 & ORGANNUM <=3 & blSOFA>5	191	2.4863	0.0065	0.359	0.550	131	60	0.388
. < BLLBILI<=2.5 & . < TIMFIRST<=39 & AGE>59.126	159	2.4816	0.0065	0.377	0.585	106	53	0.417
AGE>59.126 & BLLPLAT>92 & blSOFA>5	170	2.3986	0.0082	0.393	0.586	112	58	0.388
. < TIMFIRST<=35.53 & AGE>59.126 & BLLPLAT>111	133	2.3790	0.0087	0.360	0.574	86	47	0.432
AGE>59.126 & BLLPLAT>92 & PRAPACHE>25	100	2.3697	0.0089	0.383	0.625	60	40	0.484
. < TIMFIRST<=35.53 & ORGANNUM <=3 & PRAPACHE>24	133	2.3050	0.0106	0.348	0.561	92	41	0.433
. < BLLCREAT<=3.1 & . < TIMFIRST<=32.92 & BLIL6>567.4	136	2.2968	0.0108	0.305	0.512	95	41	0.429
. < BLLBILI<=1.5 & . < TIMFIRST<=35.53 & AGE>59.126	129	2.2716	0.0116	0.349	0.558	86	43	0.424
AGE>45.193 & BLGCS <=14 & BLIL6>567.4	115	2.2685	0.0116	0.453	0.675	75	40	0.444
. < BLADL<=5 & AGE>59.126 & BLLPLAT>92	138	2.2247	0.0130	0.289	0.479	90	48	0.398
AGE>39.89 & BLIL6>567.4 & PRAPACHE>19	136	2.2197	0.0132	0.438	0.638	89	47	0.400
. < TIMFIRST<=39 & BLGCS <=14 & BLIL6>567.4	108	2.1905	0.0142	0.382	0.600	68	40	0.436
AGE>39.89 & BLIL6>567.4 & PRAPACHE>17	155	2.1843	0.0145	0.446	0.630	101	54	0.368
. < TIMFIRST<=32.92 & BLIL6>567.4 & BLLPLAT>92	122	2.1632	0.0153	0.333	0.537	81	41	0.415
AGE>39.89 & BLIL6>567.4 & blSOFA>6.25	154	2.1600	0.0154	0.431	0.615	102	52	0.368
. < TIMFIRST<=35.53 & BLADL>0	132	2.1336	0.0164	0.363	0.561	91	41	0.401
. < TIMFIRST<=32.92 & AGE>39.89 & BLIL6>567.4	134	2.1060	0.0176	0.375	0.565	88	46	0.383
. < TIMFIRST<=32.92 & BLIL6>567.4 & PRAPACHE>17	135	2.0900	0.0183	0.367	0.556	90	45	0.382
AGE>59.126 & ORGANNUM <=3 & PRAPACHE>19	171	2.0870	0.0184	0.368	0.537	117	54	0.343
AGE>39.89 & BLGCS <=14 & BLIL6>567.4	120	2.0682	0.0193	0.455	0.651	77	43	0.394
AGE>59.126 & BLLCREAT>0.8 & BLLPLAT>92	182	2.0663	0.0194	0.380	0.541	121	61	0.324
. < TIMFIRST<=35.53 & AGE>56.388 & BLIL6>93.9	158	2.0403	0.0207	0.359	0.527	103	55	0.341
. < BLLBILI<=1.5 & . < TIMFIRST<=35.53 & PRAPACHE>22	132	2.0300	0.0212	0.318	0.500	88	44	0.375
. < TIMFIRST<=42.5 & BLLPLAT>228	106	2.0087	0.0223	0.203	0.381	64	42	0.399
AGE>39.89 & BLIL6>567.4 & BLLBILI>0.6	137	2.0065	0.0224	0.430	0.614	93	44	0.367
AGE>59.126 & BLIL6>93.9 & ORGANNUM <=3	176	2.0033	0.0226	0.368	0.525	117	59	0.320
. < TIMFIRST<=35.53 & BLIL6>93.9 & ORGANNUM <=3	238	1.9969	0.0229	0.286	0.416	161	77	0.277
. < BLLBILI<=2.5 & AGE>59.126 & ORGANNUM <=3	194	1.9067	0.0283	0.356	0.500	132	62	0.294
. < TIMFIRST<=35.53 & ORGANNUM <=3 & blSOFA>5	240	1.9008	0.0287	0.304	0.431	168	72	0.268
. < BLLBILI<=2.5 & AGE>59.126 & blSOFA>5	190	1.9000	0.0287	0.426	0.574	129	61	0.295
AGE>59.126 & PRAPACHE <=29 & blSOFA>5	159	1.8896	0.0294	0.360	0.521	111	48	0.326
. < BLLCREAT<=2.8 & BLLPLAT>228	110	1.8489	0.0322	0.171	0.325	70	40	0.366
. < BLLBILI<=1.5 & . < TIMFIRST<=35.53 & blSOFA>5	181	1.8297	0.0336	0.309	0.448	123	58	0.291
. < TIMFIRST<=35.53 & BLIL6>93.9 & PRAPACHE>24	134	1.7974	0.0361	0.371	0.533	89	45	0.329
. < BLLCREAT<=3.5 & . < TIMFIRST<=35.53 & BLIL6>93.9	251	1.7836	0.0372	0.294	0.407	170	81	0.241
. < TIMFIRST<=35.53 & BLIL6>93.9 & blSOFA>5	241	1.7342	0.0414	0.329	0.446	167	74	0.242
. < BLLPLAT<=247 & BLGCS <=14 & BLIL6>567.4	116	1.6977	0.0448	0.434	0.600	76	40	0.332
. < BLLBILI<=1.5 & . < TIMFIRST<=35.53 & ORGANNUM <=3	212	1.6507	0.0494	0.266	0.377	143	69	0.242
AGE>59.126 & BLIL6>567.4	109	1.6455	0.0499	0.515	0.674	66	43	0.322
BLGCS <=14 & BLIL6>567.4 & BLLBILI>0.5	115	1.6345	0.0511	0.440	0.600	75	40	0.320
AGE <=71.408 & BLLPLAT>228	104	1.5862	0.0564	0.143	0.268	63	41	0.318
BLIL6>567.4 & ORGANNUM <=3 & PRAPACHE>17	141	1.5785	0.0572	0.411	0.549	90	51	0.277
BLIL6>567.4 & BLLBILI>0.8	138	1.5631	0.0590	0.427	0.571	96	42	0.289
. < BLLCREAT<=3.1 & BLGCS <=14 & BLIL6>567.4	111	1.4948	0.0675	0.414	0.561	70	41	0.294
. < BLLBILI<=1.5 & . < TIMFIRST<=35.53 & BLIL6>93.9	184	1.4502	0.0735	0.306	0.413	121	63	0.225
BLIL6>567.4 & BLLCREAT>1.1 & PRAPACHE>17	134	1.4271	0.0768	0.452	0.585	93	41	0.268
BLIL6>567.4 & BLLBILI>0.6 & PRAPACHE>17	144	1.2377	0.1079	0.443	0.553	97	47	0.220
BLGCS <=8	125	1.1795	0.1191	0.388	0.500	85	40	0.226

Evaluating Variable Importance

- Apply SIDES with very loose constrains to gene large number of c subgroups (here n
- Variable importan computed for eac biomarker as the of its "contributio all subgroups whe involved as the sp
  - "X-Contributio –log(p) or = 0, involved
  - where p= trt-by-split interaction p-value

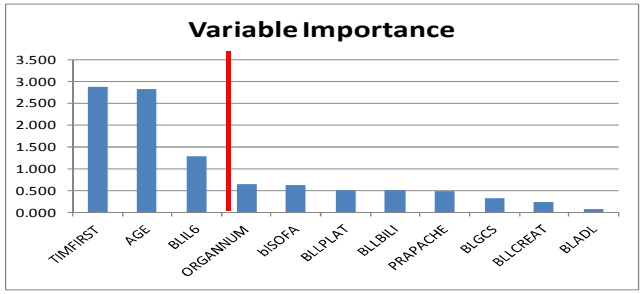
Subgroup	Subgroup size	Test statistic (z)	1-tailed p value
. < TIMFIRST<=39 & AGE>59.126 & bISOFA <=10	154	2.9138	0.0018
. < TIMFIRST<=39 & AGE>59.126 & ORGANNUM <=3	166	2.8063	0.0025
. < TIMFIRST<=35.53 & AGE>59.126 & bISOFA <=10	146	2.7346	0.0031
. < TIMFIRST<=39 & AGE>59.126 & BLLPLAT>92	157	2.6922	0.0035
. < TIMFIRST<=39 & AGE>59.126 & bISOFA>5	166	2.6414	0.0041
. < TIMFIRST<=35.53 & AGE>59.126 & ORGANNUM <=3	156	2.6104	0.0045
. < TIMFIRST<=35.53 & AGE>56.388 & ORGANNUM <=3	165	2.6101	0.0045
AGE>59.126 & ORGANNUM <=3 & bISOFA>5	191	2.4863	0.0065
. < BLLBILI<=2.5 & . < TIMFIRST<=39 & AGE>59.126	159	2.4816	0.0065
AGE>59.126 & BLLPLAT>92 & bISOFA>5	170	2.3986	0.0082
. < TIMFIRST<=35.53 & AGE>59.126 & BLLPLAT>111	133	2.3790	0.0087
AGE>59.126 & BLLPLAT>92 & PRAPACHE>25	100	2.3697	0.0089
. < TIMFIRST<=35.53 & ORGANNUM <=3 & PRAPACHE>24	133	2.3050	0.0106
. < BLLCREAT<=3.1 & . < TIMFIRST<=32.92 & BLIL6>567.4	136	2.2968	0.0108
. < BLLBILI<=1.5 & . < TIMFIRST<=35.53 & AGE>59.126	129	2.2716	0.0116
AGE>45.193 & BLGCS <=14 & BLIL6>567.4	115	2.2685	0.0116

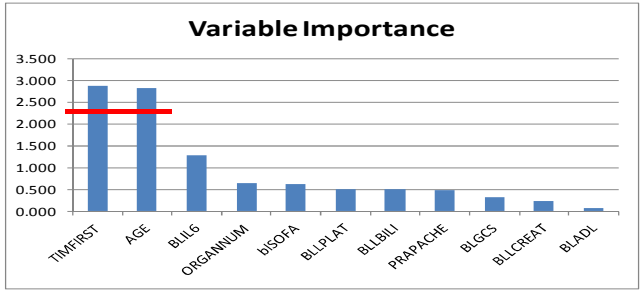
Subgroup	Subgroup size	Test statistic (z)	1-tailed p value	treatment event rate	control event rate	treatment sample size	control sample size	effect size
. < TIMFIRST<=39 & AGE>59.126 & bISOFA <=10	154	2.9138	0.0018	0.343	0.592	105	49	0.504
. < TIMFIRST<=39 & AGE>59.126 & ORGANNUM <=3	166	2.8063	0.0025	0.336	0.566	113	53	0.467
								0.490
								0.449
								0.434
								0.456
								0.440
								0.388
								0.417
								0.388
								0.432
								0.484
								0.433
								0.429
								0.424
								0.444
								0.398
								0.400
								0.436
								0.368
								0.415
								0.368
								0.401
								0.383
								0.382
								0.343
								0.394
								0.324
								0.341
								0.375
								0.399
								0.367
								0.320
								0.277
								0.294
								0.268
								0.295
								0.326
								0.366
								0.291
. < TIMFIRST<=35.53 & BLIL6>93.9 & PRAPACHE>24	134	1.7974	0.0361	0.371	0.533	89	45	0.329
. < BLLCREAT<=3.5 & . < TIMFIRST<=35.53 & BLIL6>93.9	251	1.7836	0.0372	0.294	0.407	170	81	0.241
. < TIMFIRST<=35.53 & BLIL6>93.9 & bISOFA>5	241	1.7342	0.0414	0.329	0.446	167	74	0.242
. < BLLPLAT<=247 & BLGCS <=14 & BLIL6>567.4	116	1.6977	0.0448	0.434	0.600	76	40	0.332
. < BLLBILI<=1.5 & . < TIMFIRST<=35.53 & ORGANNUM <=3	212	1.6507	0.0494	0.266	0.377	143	69	0.242
AGE>59.126 & BLIL6>567.4	109	1.6455	0.0499	0.515	0.674	66	43	0.322
BLGCS <=14 & BLIL6>567.4 & BLLBILI>0.5	115	1.6345	0.0511	0.440	0.600	75	40	0.320
AGE <=71.408 & BLLPLAT>228	104	1.5862	0.0564	0.143	0.268	63	41	0.318
BLIL6>567.4 & ORGANNUM <=3 & PRAPACHE>17	141	1.5785	0.0572	0.411	0.549	90	51	0.277
BLIL6>567.4 & BLLBILI>0.8	138	1.5631	0.0590	0.427	0.571	96	42	0.289
. < BLLCREAT<=3.1 & BLGCS <=14 & BLIL6>567.4	111	1.4948	0.0675	0.414	0.561	70	41	0.294
. < BLLBILI<=1.5 & . < TIMFIRST<=35.53 & BLIL6>93.9	184	1.4502	0.0735	0.306	0.413	121	63	0.225
BLIL6>567.4 & BLLCREAT>1.1 & PRAPACHE>17	134	1.4271	0.0768	0.452	0.585	93	41	0.268
BLIL6>567.4 & BLLBILI>0.6 & PRAPACHE>17	144	1.2377	0.1079	0.443	0.553	97	47	0.220
BLGCS <=8	125	1.1795	0.1191	0.388	0.500	85	40	0.226

Data Example. Screening Noise Variables Using Variable Importance

- Interpretation →Look at sharp decline in VI
- A more formal procedure for selecting variables
  - Fixed number of top covariates, e.g. 3
  - Select only those that are above a benchmark (from the dummy data sets)

Variable	Importance Based on Splitting Criterion
TIMFIRST	2.858
AGE	2.833
BLIL6	1.267
ORGANNUM	0.650
blSOFA	0.612
BLLPLAT	0.510
BLLBILI	0.504
PRAPACHE	0.479
BLGCS	0.316
BLLCREAT	0.225
BLADL	0.087
Permutation-based Mean(Max VI)	1.895
Permutation-based Std(Max VI)	0.562
Variable Screening Rule: Choose covariates above Mean(MaxVI)+1Std(MaxVI)	2.457





SIDEScreen: Screening Noise Variables Using Variable Importance

- A more formal procedure for selecting variables
  - A Fixed number of top covariates, e.g. 3
  - Select only those covariates that are above a data-dependent benchmark (calibrated from generated null data sets)
- SIDEScreen – an extension of basic SIDES procedure
  - Step 1: apply basic SIDES with loose complexity control, generating a large number of subgroups
  - Step 2: apply SIDES again to only selected covariates with largest Vl exceeding a cut-off. Calibrate the significance cut-offs for the final subgroups using data resampling, so as to ensure desired type I error rate of the entire procedure
    - This involves replicating entire 2-stage procedure on muliple null dat asets

SIDEScreen vs. SIDES. Data Example

- **Strategy I.** Apply SIDES with width=3 (pursuing 3 top candidate from each parent node), depth=up to 2 levels, and child-two-parent ratio (γ) =0.2

Subgroup	Subgroup size	Test statistic (z)	1-tailed p value	treatment event rate	control event rate	treatment sample size	control sample size	effect size	multiplicity-adjusted 1-sided p-value
TIMFIRST<=39 & AGE>59.126	195	2.4056	0.0081	0.380	0.561	129	66	0.364	0.416
TIMFIRST<=35.53 & AGE>59.126	185	2.2363	0.0127	0.376	0.550	125	60	0.351	0.474
TIMFIRST<=35.53 & BLADL>0	132	2.1336	0.0164	0.363	0.561	91	41	0.401	0.492
BLLPLAT>228	134	1.4422	0.0746	0.210	0.321	81	53	0.255	0.620

- **Strategy II.** Apply SIDES with a more loose "search space", with less restrictions (width=5, depth=3 and child-two-parent ratio =1) and apply SIDES again only to those covariates whose VI is above benchmark

Second Stage Analysis: Using Variables Selected in Stage I

Subgroup	Subgroup size	Test statistic (z)	1-tailed p value	treatment event rate	control event rate	treatment sample size	control sample size	effect size	multiplicity-adjusted 1-sided p-value
TIMFIRST<=39 & AGE>59.126	195	2.4056	0.0081	0.380	0.561	129	66	0.364	0.064
TIMFIRST<=35.53 & AGE>59.126	185	2.2363	0.0127	0.376	0.550	125	60	0.351	0.072

The same subgroup has been selected but its adjusted p-value is smaller because it passed through a more stringent two-stage screening process

Conclusions and Discussion

- SIDES is a novel method for subgroup identification incorporating 3 elements
  - Flexible direct-search algorithm
  - Complexity control
  - Multiplicity control.
- Using the SIDEScreen – a multistage subgroup identification strategies with preliminary screening of noise variables by variable importance – may substantially boost the probability of indentifying the true predictors, especially for data sets with larger number of covariates
- Using data-dependent reference cutoffs for VI substantially improves detection of true subgroup and filtering out noise covariates
- Performance of SIDESscreen substantially deteriorated when all noise covariates were correlated with true predictors
- Further improvement may be achieved by introducing randomness in generating candidate subgroups for the first step of SIDEScreen:

References
(Guidelines for subgroup analyses)

- Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G. (2001). Subgroup analyses in randomized controlled trials: quantifying the risks of false-positives and false-negatives. Health Technology Assessment. 5:1-33
- Rothwell PM. (2005). Subgroup analysis in randomized controlled trials: importance, indications, and interpretation. *Lancet*. 365:176–86
- Sun X, Briel M, Walter SD, Guyatt GH. (2010) Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses, *British Medical Journal*. 340:850-854

References (data-mining methods for subgroup analysis)

- Kehl V and Ulm K. (2006). Responder identification in clinical trials with censored data. *Computational Statistics & Data Analysis*. 50:1338-1355
- Su X, Tsai CL, Wang H, Nickerson DM, Li B. (2009). Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research*. 10:141-158.
- Lipkovich I, Dmitrienko A, Denne J, Enas G. (2011). Subgroup Identification based on Differential Effect Search (SIDES) – A recursive partitioning method for establishing response to treatment in patient subpopulations. *Statistics in Medicine*. 30:2601-2621
- Foster JC, Taylor JMC, Ruberg SJ. (2011). Subgroup identification from randomized clinical trial data, *Statistics in medicine*. 30:2867–2880
- Berger J and Scott J. Bayesian Adjustment for Multiplicity. (2011). 7th International Conference on Multiple Comparison Procedures. Washington, DC
- Jones HE, Ohlssen DI, Neuenschwander B, Racine A and Branson M. (2011). Bayesian models for subgroup analysis in clinical trials. Clinical Trials. 8:129–143
- Zhang B, Tsiatis AA, Davidian M, Zhang M, Laber E. (2012). Estimating optimal treatment regimes from a classification perspective. Statistics. 1, 103-114.
- 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.
- Tian L, Alizaden AA, Gentles AJ, Tibshirani R. (2012). A Simple Method for Detecting Interactions between a Treatment and a Large Number of Covariates. http://arxiv.org/abs/1212.2995

Thank You!

Back Up

Comparing SIDES with Classification and Regression Tree Methodology (CART)

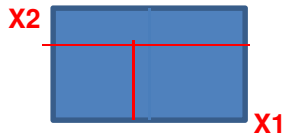
Prognostic variables

Predictors of treatment effect

$$Y_i = f(\boldsymbol{X}_i) + \{\varphi_1 I(T_i = A) - \varphi_2 I(T_i = B)\} I(\boldsymbol{X}_i \in S) + \varepsilon_i$$

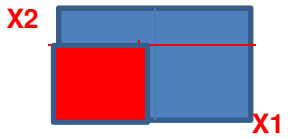
CART

- Looks for subgroups with high/low level of outcome: Y(S)→ high/low
- "Hidden multiplicity". Always chooses the covariate with the top split, other competing splits are not shown"
- Partitions entire covariate space into non-overlapping subsets



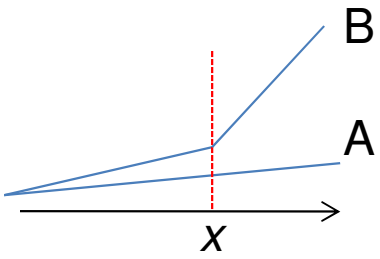
SIDES

- Directly looks for subgroups with high level of treatment effect, $\Delta(S)=Y_A(S)-Y_B(S) \rightarrow$ high
- Generates multiple "promising" subgroups, allowing for "transparent multiplicity"
- Focuses only on "interesting" segments of the covariate space



Which Splitting Criteria?

- **Differential Effect (Treatment By Split Interaction).** Maximize difference in treatment effects for the two child subgroups formed by the split (converted to a probability scale)
  - $C_1 = 2(1 - \Phi\{|Z_1 - Z_2|/\sqrt{2}\})$, $\Phi()$ is Normal CDF
  - May point in the wrong direction, selecting a covariate because it defines a subgroup where the control (B) is superior



- **Maximal Effect.** Looks at maximal treatment effect in the direction favorable for active drug A vs. B in one of the two child subgroups
  - $C_2 = 2\min\{1 - \Phi(Z_1), 1 - \Phi(Z_2)\}$

- **Hybrid Effect** – take the worst of the two
  - the split is justifiable only when the resulting subgroup
    - (A) contains treatment effect and
    - (B) reveals heterogeneity of TE in the parent group
  - $C_3 = \max(C_1, C_2)$

Computing Test Statistic in Candidate Subgroups

- For continuous outcome: standardized difference in treatment means

$$z_j = (\bar{y}_{j,T} - \bar{y}_{j,C}) / (\bar{\sigma}_j \sqrt{1/n_{j,T} + 1/n_{j,C}}), \;\; j = 1,2$$

- For binary response: standardized difference in treatment proportions

$$z_j = (p_{j,T} - p_{j,C}) / (\bar{p}_j (1 - \bar{p}_j) \sqrt{1/n_{j,T} + 1/n_{j,C}}), \;\; j = 1,2$$

- For time to event: log-rank test:

$$Z_j = \sum_{i=1}^{D} [d_{iC}^j - Y_{iC}^j (\frac{d_i^j}{Y_{iC}^j + Y_{iT}^j})] / \sqrt{\sum_{i=1}^{D} \frac{Y_{iC}^j}{Y_{iC}^j + Y_{iT}^j} (1 - \frac{Y_{iC}^j}{Y_{iC}^j + Y_{iT}^j}) (\frac{Y_{iC}^j + Y_{iT}^j - (d_{iC}^j + d_{iT}^j)}{Y_{iC}^j + Y_{iT\,i}^j - 1}) (d_{iC}^j + d_{iT}^j)}$$

Y_i – number of subjects at risk t_i-0 (just prior to time t_i), i=1,..,D
d_i – number of events occurring at time t_i

Evaluating the Overall Type I Error Rates for the Entire Search Procedure

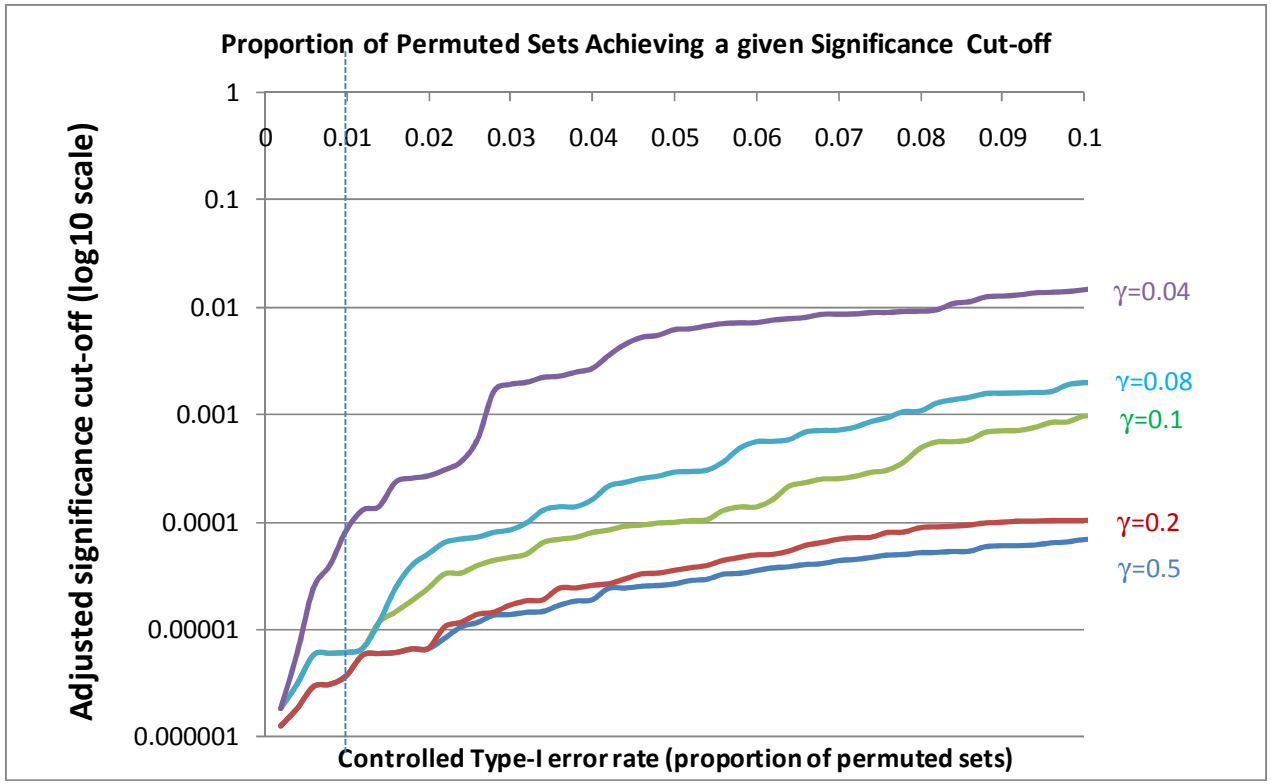
- Construct re-sampling reference distribution for the **min** observed significance value **across all subgroups within the space Σ induced by the search algorithm**

$$\Pr\left(\min_{S \in \Sigma} \{Pvalue(S)\} < ? \right) < \alpha_0$$

- Permute outcome Y and T against X's and apply the search algorithm to each permuted set
- Each permuted set
  - Retains the overall treatment effect
  - Retains correlation across covariates
  - Makes covariates uninformative for identifying subgroups with improved treatment effect

Y	T	X1	X2	X3	…	Xk
1	A	0	4	10.2		0
1	B	1	3	12.1		2
0	A	2	2	23.4		1
0	B	3	7	12.2		1
1	B	2	5	19.3		1

Overall Type I Error Rates Associated with Procedures Using Different Search Space



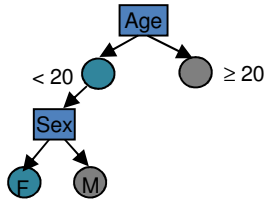
Given the same fixed type I error rate (on horizontal axis) , one can get a larger (more "relaxed") probability cut-off (vertical axis) by choosing a smaller γ (more "stringent" continuation rule)

Smaller γ effectively constrains the subgroup space, hence we pay less penalty in multiplicity adjustment

Using k-fold Cross-validation For Choosing Optimal Levels of Tuning Parameter(s)

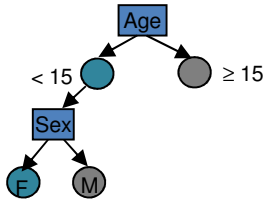
Best subgroups on training folds

Exploratory (training) data



Train | Train | Train | Train | **Test** → $P_1(\gamma)$

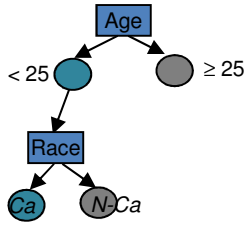
P-value for TE



Train | Train | Train | **Test** | Train → $P_2(\gamma)$

P-value for TE

. . . . .



Test | Train | Train | Train | Train → $P_5(\gamma)$

P-value for TE

Compute *average Pvalue*(γ) across 5 validation folds

Choose optimal $\gamma^* = \textbf{\textit{argmin}}\{AvP(\gamma)\}$