Part [2.1]: Evaluation of Markers for Treatment Selection – Linking Clinical and Statistical Goals

- Patrick J. Heagerty
- Department of Biostatistics
- University of Washington

Session Outline

- **Multivariate marker analysis**
  - Definition of goal
  - Model search methods / strategies
  - Marker-by-treatment
  - Treatment selection marker

Multiple Genes?

Analysis with multiple markers

- Data:
  - Outcome of interest: Y_i
  - Treatment group (dose): X_i
  - Generic (**genetic**) markers: G_{ij} for $j = 1, \ldots, M$.
- Questions:
  - **Q**: How to use multiple G_{ij} to predict outcome?
  - **Q**: How to use multiple G_{ij} to "score" subjects with respect to treatment benefit?
  - **Q**: How to use multiple G_{ij} to create treatment decision function, $A(\boldsymbol{G}_i) = a$?

Regression Framework

- Generalized Linear Model:

$$\begin{aligned} E[Y_i \mid X_i, \boldsymbol{G}_i] &= \mu_i \\ g(\mu_i) &= \beta(\boldsymbol{G}_i) + \gamma(\boldsymbol{G}_i) \cdot X_i \end{aligned}$$

- "Varying coefficient model" (Hastie and Tibshirani, 1993) – simple example:

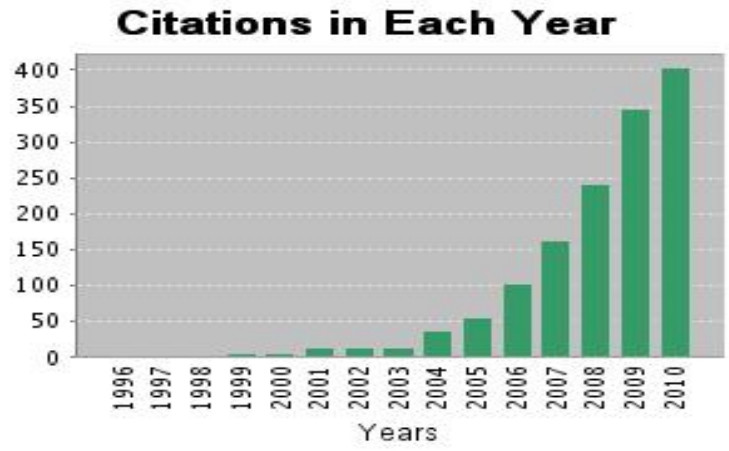
$$g(\mu_i) = \overbrace{(\beta_0 + \beta_1 \cdot G_{i1} + \ldots \beta_M \cdot G_{iM})}^{\beta(\boldsymbol{G}_i)} + \underbrace{(\gamma_0 + \gamma_1 \cdot G_{i1} + \ldots \gamma_M \cdot G_{iM})}_{\gamma(\boldsymbol{G}_i)} \cdot X_i$$

Major Challenges

- **Q**: How to select markers to include as part of $\beta(\boldsymbol{G}_i)$ and $\gamma(\boldsymbol{G}_i)$?
- **Q**: Should we also consider gene-gene interactions, $G_{ij} \times G_{ik}$, or higher order interactions (epistasis)?
- **Q**: We often have M that is $O(10^6)$ and that is much larger than the number of subjects, n, so how can we fit a model?
- **Q**: How do model choice criteria reflect the ultimate clinical goal of the model (e.g. prediction versus treatment selection)?

Regularization Methods

Tibshirani R (1996): Regression shrinkage and selection via the lasso. *JRSS-B*, 58(1): 267-288.



- Estimation: Regularization methods maximize an objective function (e.g. likelihood) subject to constraints / penalty.

$$\widehat{\boldsymbol{\theta}} = \underset{\theta}{\operatorname{argmax}} \left[\sum_i \log \; Pr(Y_i \mid \boldsymbol{G}_i, X_i; \boldsymbol{\theta}) - \lambda \cdot \sum_j |\theta_j|^p \right]$$

Regularization Methods

name		penalty
LASSO (Tibshirani 1996)	p=1	$\lambda \cdot \sum_j \lvert\theta_j\rvert$
Ridge regression (Hoerl 1962)	p=2	$\lambda \cdot \sum_j \lvert\theta_j\rvert^2$
Elastic net (Zou & Hastie 2005)	p=1,2	$\lambda_1 \cdot \sum_j \lvert\theta_j\rvert + \lambda_2 \cdot \sum_j \lvert\theta_j\rvert^2$

Regularization Methods: Comments

- Lasso tends to "select" variables by keeping $\widehat{\beta}_j = 0$.
- Ridge regression tends to include all variables, but with small coefficients (no selection).
- Lasso will not estimate a model with $m > n$ (e.g. m is number of non-zero coefficients).
- Fast algorithms exist and allow calculation of regularization paths.
- Lasso tends to select only one variable among a group of highly correlated predictors.

Regularization: example # 1

- Friedman, Hastie & Tibshirani (2010)
- R package — `glmnet`
- Example using data from Golub et al. (1999)
  - $n = 72$ observations
  - $m = 3571$ genes (expression)
  - binary outcomes (leukemia AML vs. ALL)

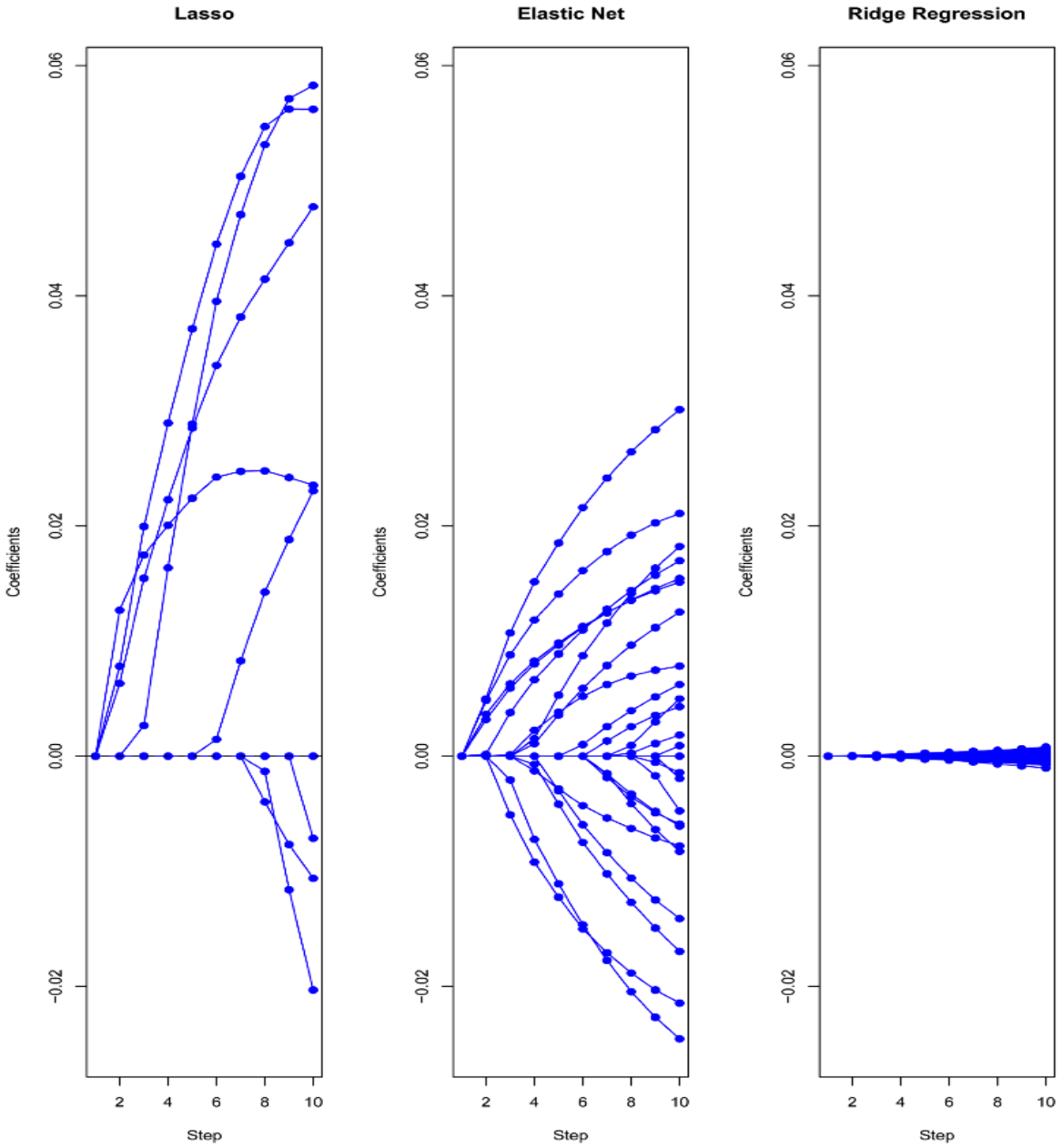


Figure 1.
Leukemia data: profiles of estimated coefficients for three methods, showing only first 10 steps (values for λ) in each case. For the elastic net, $\alpha = 0.2$.

Regularization: example # 2

- Kooperberg, LeBlanc, Obenchain (2010)
- Disease risk prediction
- Separate development and validation
- Data:
  - WTCCC Crohn's Disease
  - **Training**: (1045 cases / 1763 conrols)
  - **Test**: (703 cases / 1175 controls)
  - Also used NIDDK Crohn's data as test data
  - $m = 500K$ SNPs

TABLE I. Number of SNPs used in the prediction models with non-zero coefficients, log-likelihood, and AUC for the test data ($n = 1{,}878$: 1,175 controls and 703 cases) for the WTCCC Crohn's disease data using the lasso

Method	SNPs used	Log-likelihood	AUC
No SNPs used	0	−1,241.77	0.500
GLM	18	−1,223.12	0.606
Filtered GLM[a]	26	−1,224.94	0.626
Stepwise GLM AIC	38	−1,287.68	0.631
Stepwise GLM BIC	14	−1,236.35	0.614
Lasso top 1 SNPs considered	1	−1,239.33	0.528
Lasso top 2 SNPs considered	2	−1,231.84	0.551
Lasso top 5 SNPs considered	4	−1,232.09	0.569
Lasso top 10 SNPs considered	6	−1,232.71	0.568
Lasso top 25 SNPs considered	14	−1,207.98	0.612
Lasso top 50 SNPs considered	25	−1,197.59	0.630
Lasso top 100 SNPs considered	33	−1,193.20	0.637
Lasso top 250 SNPs considered	91	−1,196.24	0.634
Lasso top 500 SNPs considered	155	−1,195.61	0.635
Lasso top 1,000 SNPs considered	176	−1,195.04	0.636
Lasso top 2,000 SNPs considered	177	−1,194.78	0.637

WTCCC, Welcome Trust Case Control Consortium; AUC, area under the curve; GLM, generalized linear models; SNPs, single nucleotide polymorphisms.

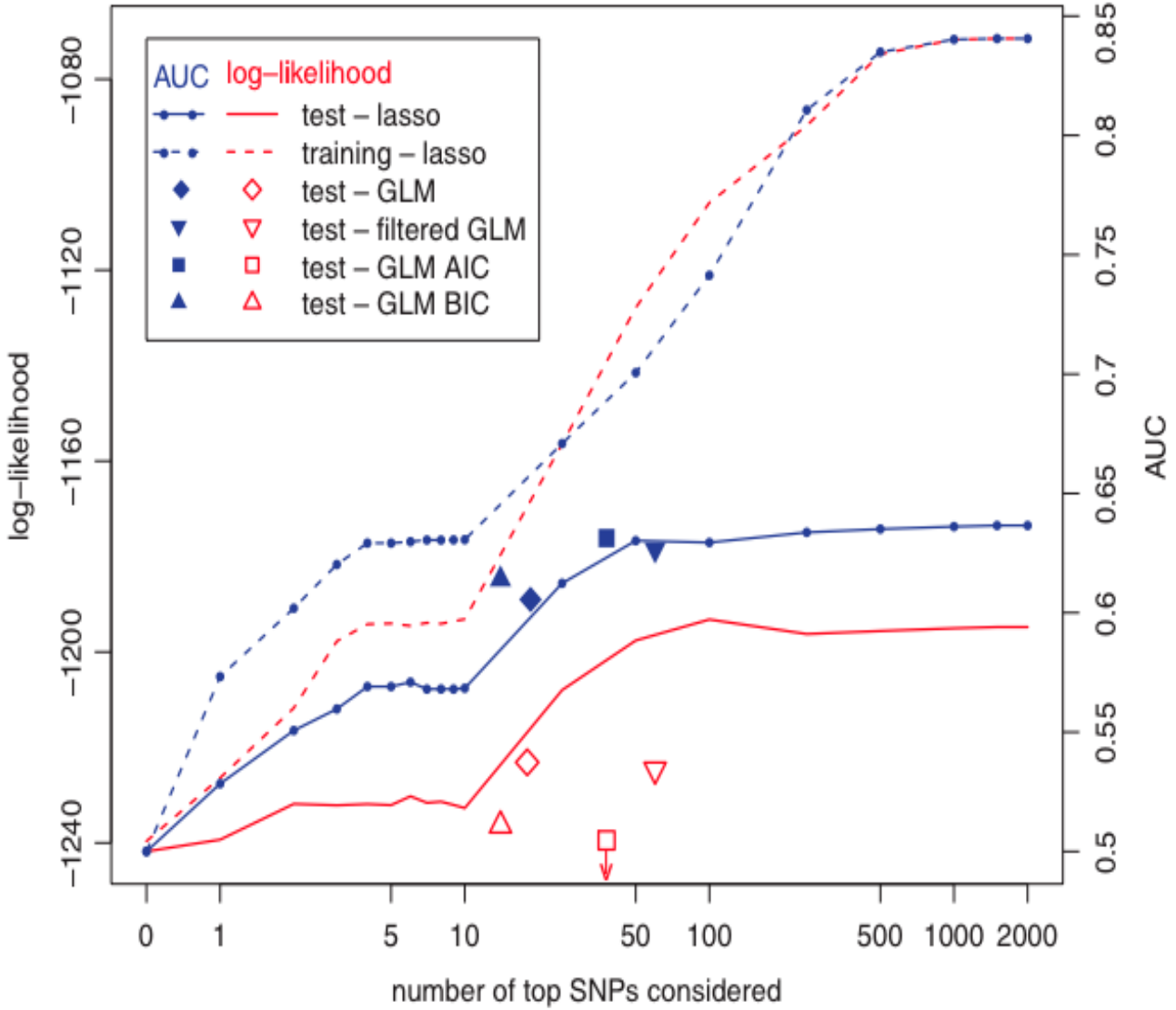


Fig. 2. Log-likelihood and AUC for the WTCCC Crohn's disease data for prediction models for test and training data. The training data log-likelihood was rescaled by a factor of 1,878/2,808 to be on the same scale as the test data log-likelihood. Note that not all SNPs considered have nonzero coefficients, see Table I. The log-likelihood for stepwise GLM using AIC (GLM-AIC) is –1,287.7. The insert figure at the left bottom vertically expands the curves for the models with 100 SNPs or less. WTCCC, Welcome Trust Case Control Consortium; AUC, area under the curve; GLM, generalized linear models; SNPs, single nucleotide polymorphisms.

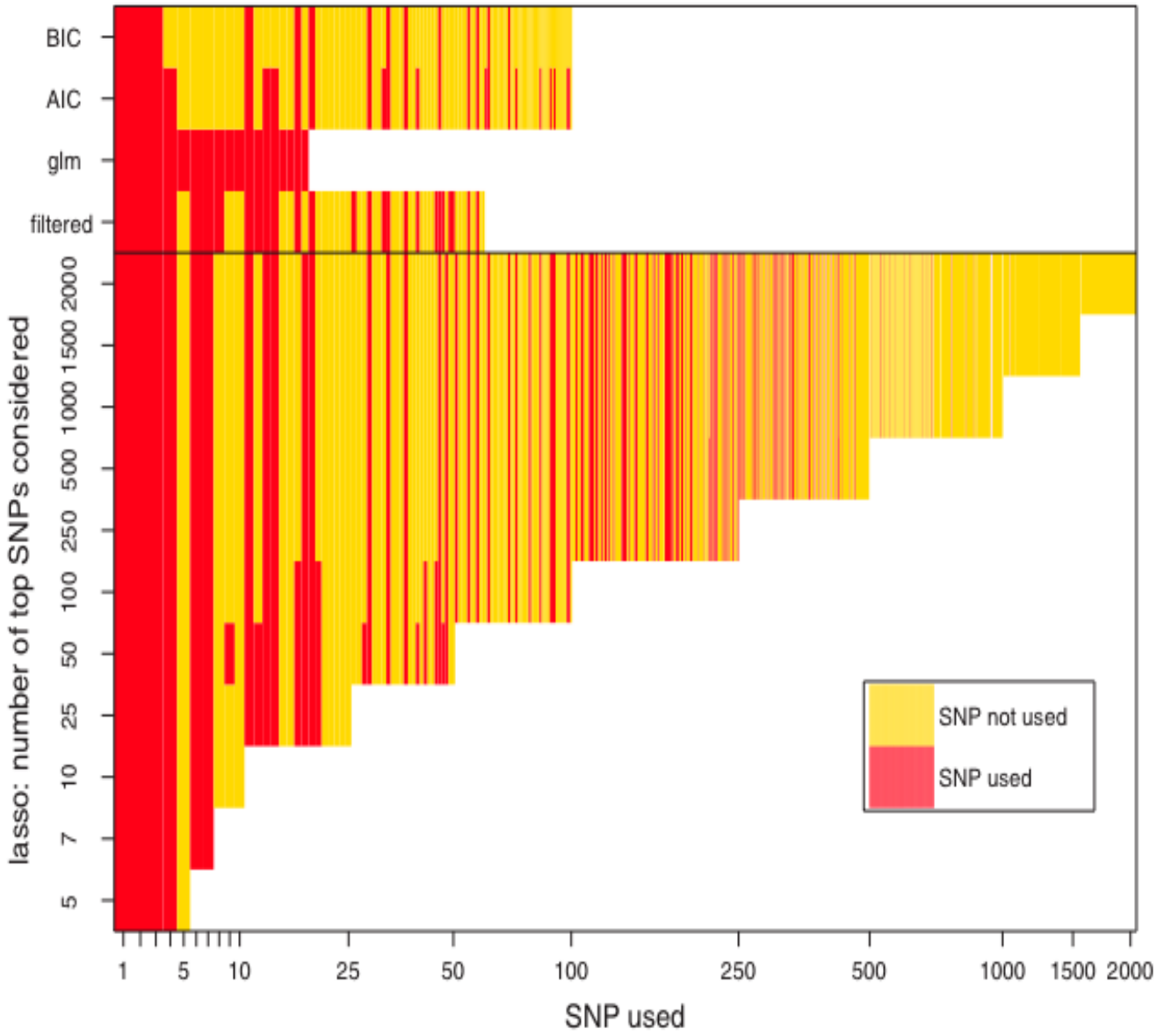


Fig. 3. Which SNPs are and are not used with nonzero coefficients for the lasso model and other prediction models for the WTCCC Crohn's disease data. The SNPs are ordered on the horizontal axis by significance. The vertical stripes suggest that frequently the same SNPs are selected. WTCCC, Welcome Trust Case Control Consortium; SNPs, single nucleotide polymorphisms.

Regularization: Summary

- Computationally **feasible**!
- Model with p predictors is "best" in terms of (fit-penalty).
- If **prediction** is desired then can be measured using cross-validation.
- Can be used with variables $= G_{ij} \cdot X_i$ to estimate $\gamma(\boldsymbol{G}_i)$ – treatment effect as a function of genetic profile.

Regularization: Summary

- May be used to identify patient subset(s) with large treatment response:

$$\begin{aligned} \widehat{\gamma}(\boldsymbol{G}_i) &= \widehat{\boldsymbol{\gamma}}^T \boldsymbol{G}_i^S \\ &= \widehat{\gamma}_0 + \widehat{\gamma}_1 \cdot G_{i(1)} + \ldots + \widehat{\gamma}_m \cdot G_{i(m)} \end{aligned}$$

high response : $[\ \widehat{\gamma}(\boldsymbol{G}_i) > c\ ]$

where $\boldsymbol{G}_i^S$ is the subset of $\boldsymbol{G}_i$ selected, denoted by $G_{i(j)}$ for $j = 1, 2, \ldots, m$.

- No direct linkage between standard lasso/eNet and use of the markers for treatment selection.

Illustration of A Benefit Score

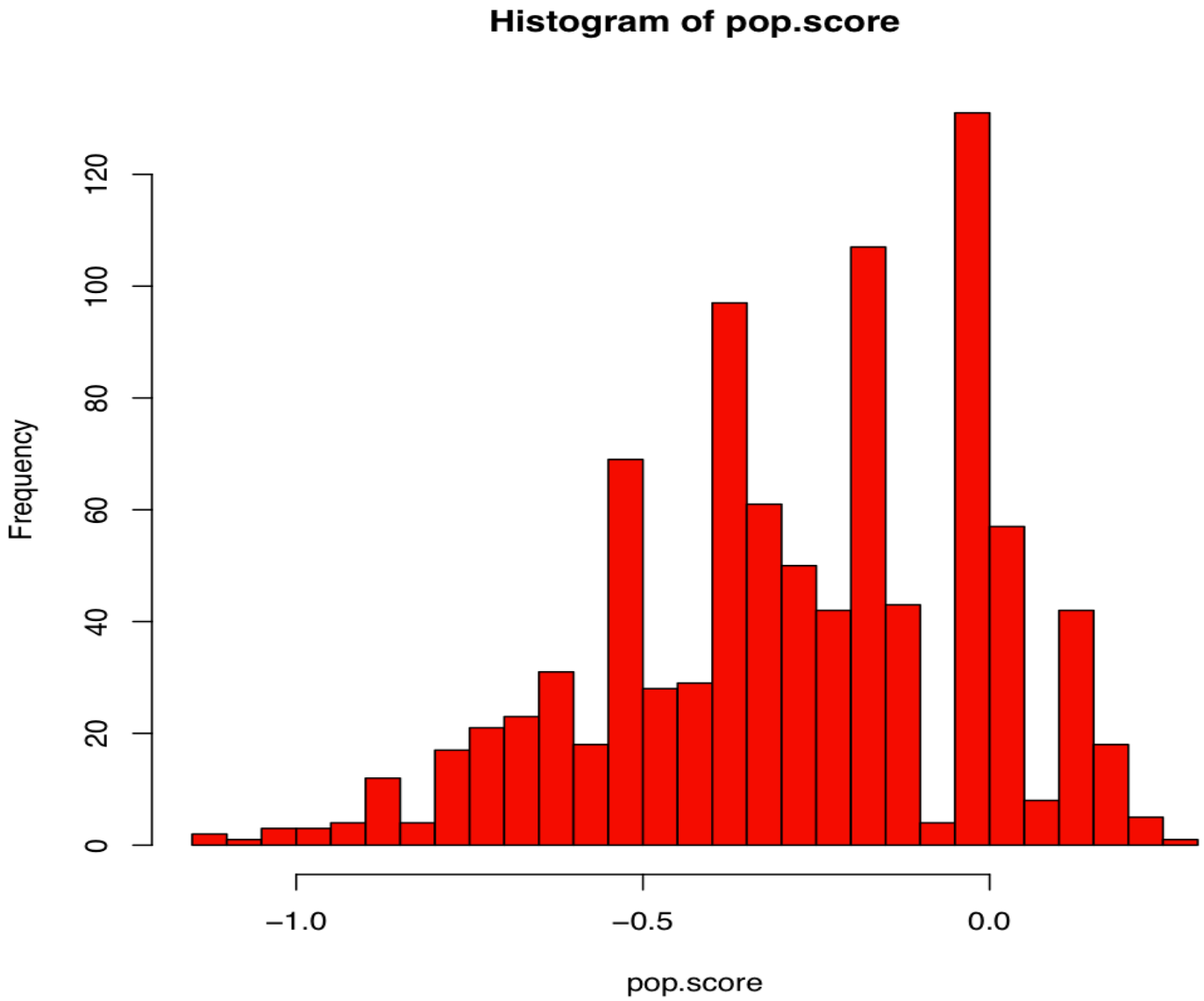
- Suppose (5) SNPs / logistic regression

	G_1	G_2	G_3	G_4	G_5
MAF	0.15	0.25	0.18	0.27	0.17
γ_j	-0.26	-0.35	+0.05	-0.15	+0.11

- The benefit score is:

$$S_i = \gamma_0 + \sum_{j=1}^{5} \gamma_j \cdot G_{i,j}$$

- PLOT: Assumes HW and unlinked markers


Histogram of pop.score
Frequency
0
20
40
60
80
100
120
−1.0
−0.5
0.0
pop.score

Treatment Selection Guideline

- **Q**: If the goal of genetic marker analysis is to develop a scoring that would be used to **select treatment** then what approach could/should be used?
- **A**: In order to identify "good properties" of a treatment selection scheme the goal or objective needs to be stated in statistical terms.
- Two approaches:
  - ▷ Population result of using a guideline
  - ▷ Accuracy of guideline in classifying those who benefit vs. do not

Treatment Selection Guideline

- Gunter, Zhu, Murphy (2007)
- One approach is to formulate an **action function** and then state the resulting **population mean** outcome:

$$\textbf{action function} \quad : \quad A(\boldsymbol{G}_i) = a$$

$$\textbf{population result} \quad : \quad E[Y_i(a) \mid A(\boldsymbol{G}_i) = a] = \mu_A$$

- Here we consider $Y_i(a)$ as the potential outcome for subject i if treated with choice a.
- Here the action a may be 1="treat", 0="do not treat", or may be a dose etc.

Treatment Selection Guideline

- Example with a single $G_i = 0, 1, 2$; Tx = 0 / 1:

genotype	$\overline{Y}_i(0)$	$\overline{Y}_i(1)$	A_0	A_1	A^*
0 (30%)	10	20	0	1	**1**
1 (50%)	10	10	0	1	**0**
2 (20%)	15	5	0	1	**0**
pop. mean			11	12	14

- Note that $A^*(G_i)$ given above is optimal in terms of maximizing μ_A over all possible functions $A(G_i)$.

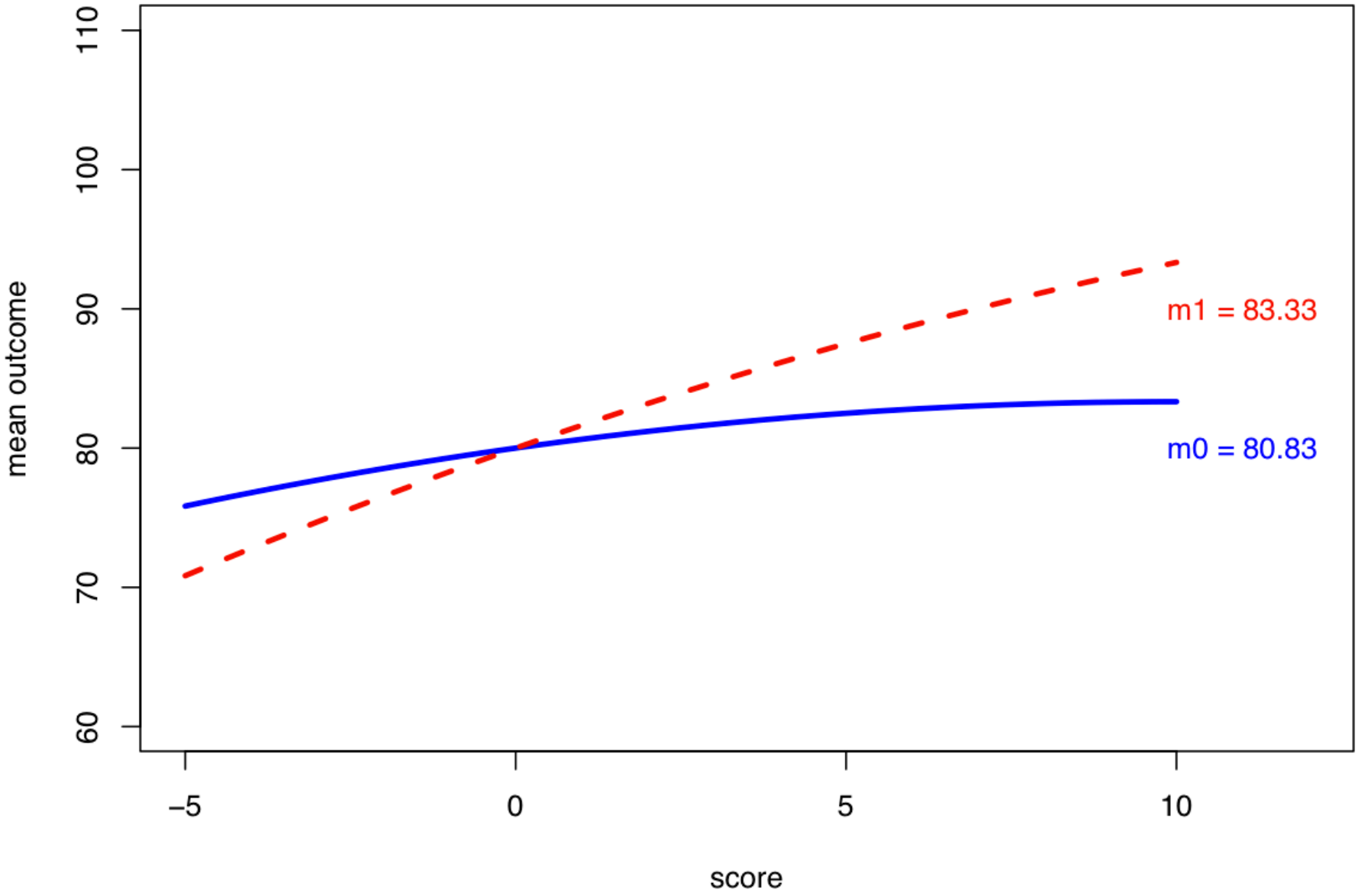
Treatment Selection Guideline

- **Q**: What about a quantitative variable or score: S?
- Define:

$$\begin{aligned} \mu_0(S) &= E[Y(0) \mid S] \quad \text{untreated mean curve} \\ \mu_1(S) &= E[Y(1) \mid S] \quad \text{treated mean curve} \end{aligned}$$

- Overall Means:

$$\begin{aligned} \overline{\mu}_0 &= \text{average}\,[\mu_0(S)] = \int \mu_0(s) \cdot f(s)\, ds \\ \overline{\mu}_1 &= \text{average}\,[\mu_1(S)] = \int \mu_1(s) \cdot f(s)\, ds \end{aligned}$$


mean outcome
110
100
90
80
70
60
−5
0
5
10
score
m1 = 83.33
m0 = 80.83

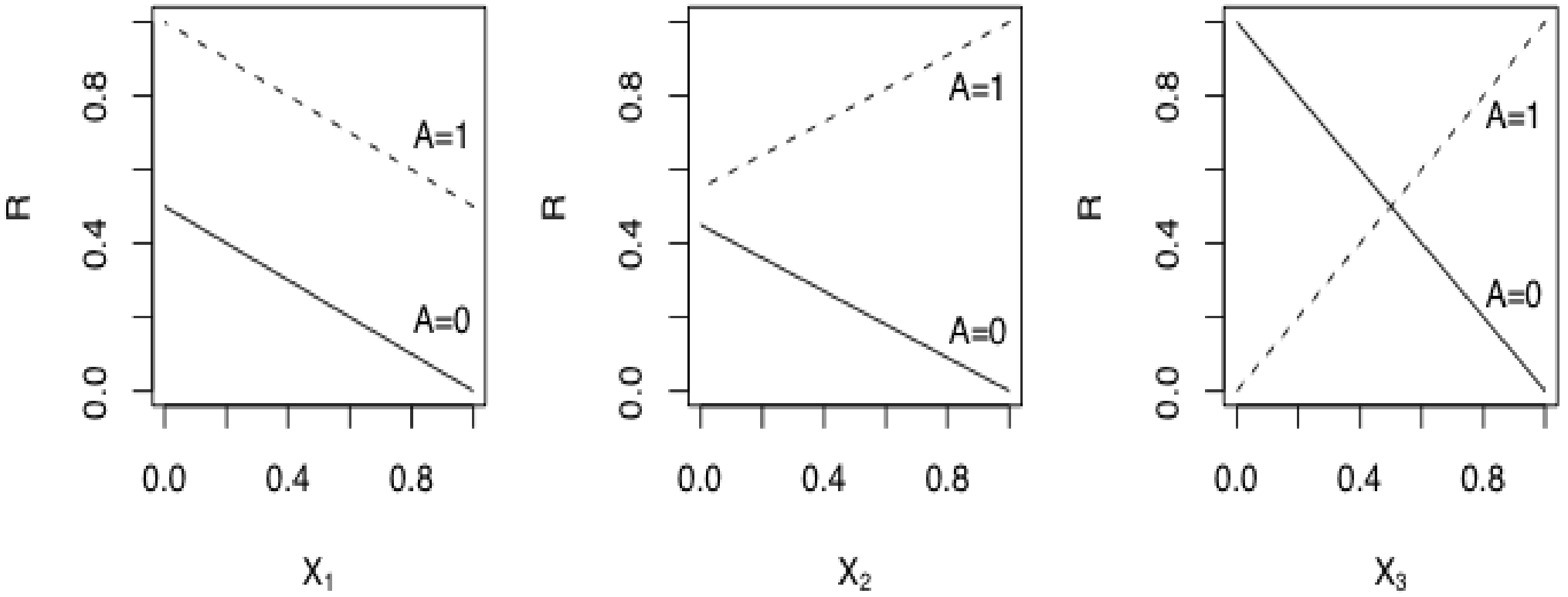
Treatment Selection Guideline

- Define: **Treatment Effect Function**

$$\Delta(S) \quad = \quad \mu_1(S) \; - \; \mu_0(S)$$

- Optimal Action: $A^*(S)$

$$A^*(S) = \begin{cases} 0 \;\text{ if }\; \Delta(S) \leq 0 \\ 1 \;\text{ if }\; \Delta(S) > 0 \end{cases}$$

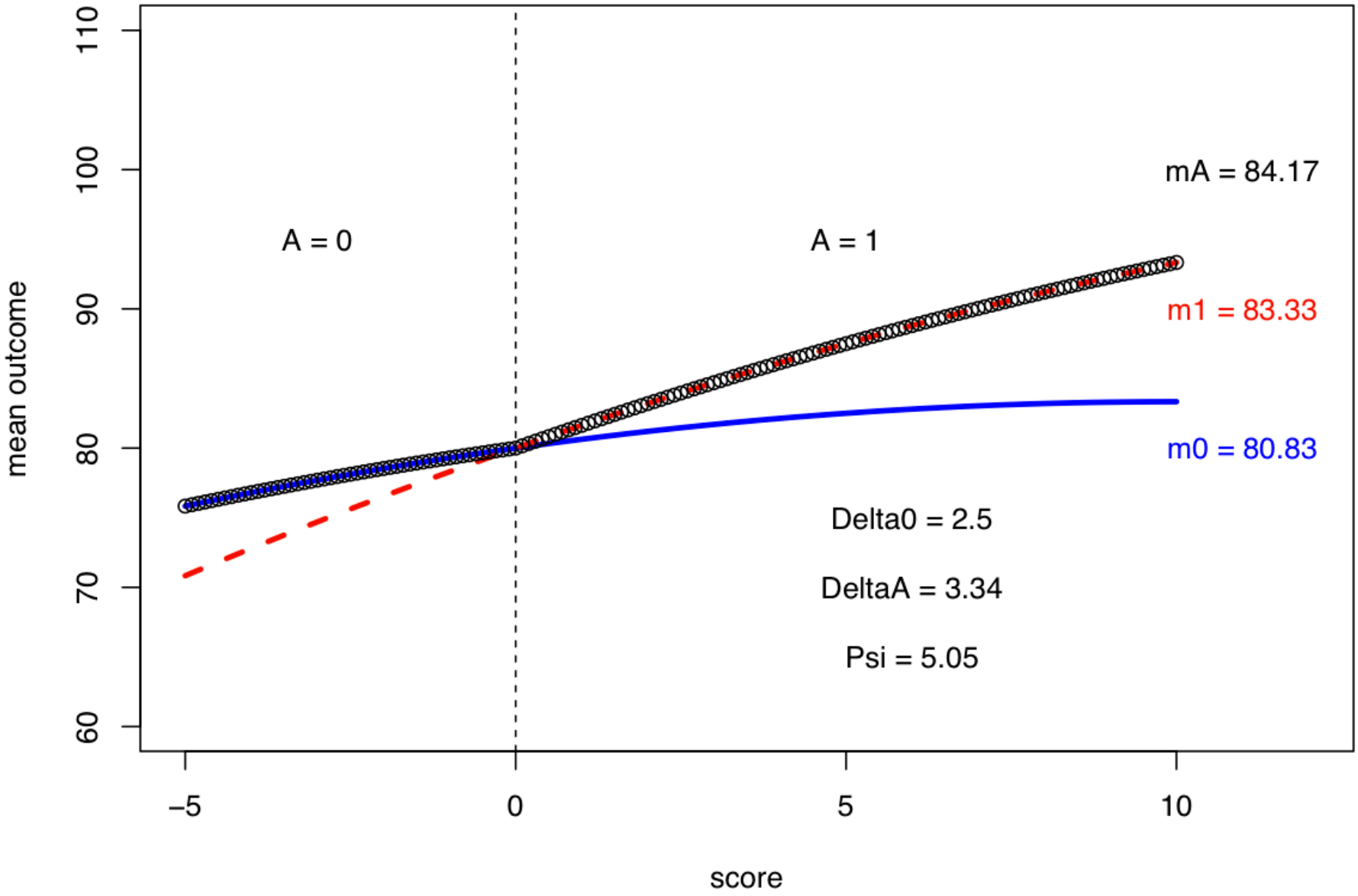


(a) No interaction (b) Non-qualitative interaction (c) Qualitative Interaction

Fig. 1. Plots demonstrating qualitative and non-qualitative interactions

Characteristics of Treatment Markers

- **Population Mean using Marker**:
  - What would be the average in the population if subjects were given treatment that was best for each of them individually?
- **Treatment Effect using A^***:
  - What is the treatment effect that is obtained comparing optimal treatment to control?
- **Treatment Effect Among Treated**:
  - What is the treatment effect among those subjects that get assigned to treatment (using marker)?


mean outcome
110
100
90
80
70
60
A = 0
A = 1
mA = 84.17
m1 = 83.33
m0 = 80.83
Delta0 = 2.5
DeltaA = 3.34
Psi = 5.05
−5
0
5
10
score

Characteristics of Treatment Markers

- **Population Mean using Marker**:

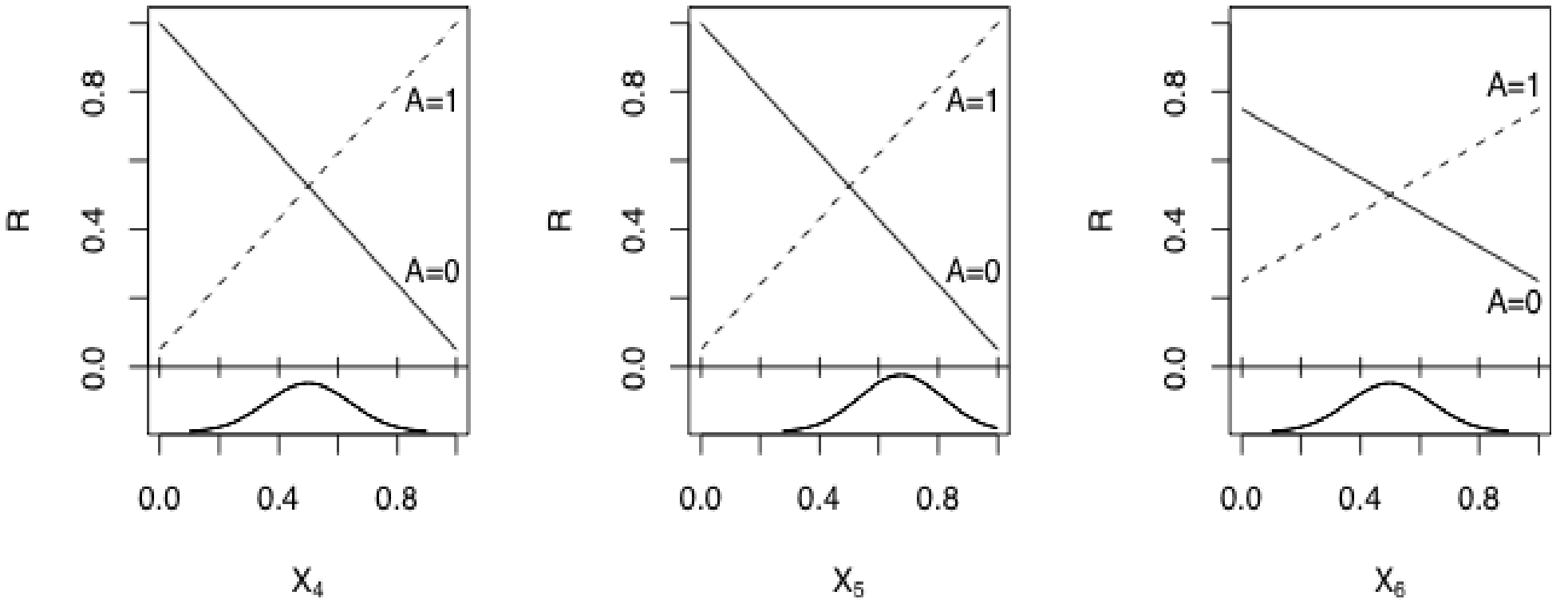
$$\mu_{A^*} = \int \{\mu_0(s) \cdot \mathbf{1}[\Delta(s) \leq 0] + \mu_1(s) \cdot \mathbf{1}[\Delta(s) > 0]\} \ f(s) \, ds$$

- **Treatment Effect using A^***:

$$\Delta^* = \int \Delta(s) \cdot \mathbf{1}[\Delta(s) > 0] \ f(s) \, ds$$

- **Treatment Effect Among Treated**:

$$\Psi = \int \Delta(s) \cdot f(s)/P[\Delta(S) > 0] \, ds$$



(a) Large interaction, large portion of change in action

(b) Large interaction, small portion of change in action

(c) Small interaction, large portion of change in action

Fig. 2. Plots demonstrating usefulness factors of qualitative interactions

Genetic Markers

- With a vector of genetic markers, $\boldsymbol{G}_i$ the goal would be:

$$A^*(\boldsymbol{G}_i) \quad := \underset{A}{\operatorname{argmax}}\, E[Y_i(a) \mid A(\boldsymbol{G}_i) = a]$$

- Here the goal is to determine which components of $\boldsymbol{G}_i$ are **prescriptive** markers – those with qualitative interactions rather than simply having quantitative interactions with treatment.
- Space of functions $\{\ A(\boldsymbol{x})\ \}$ is of order $O(10^M)$ — 3^M genotypes with $(3^M)^2$ binary actions a.

Key Steps to Move Forward

- The key idea is to not try to estimate $A^*(\boldsymbol{G}_i)$ but rather to focus on a simpler version:
  - ▷ **Scalar**: $A^*(S_i)$

$$S_i = \gamma_0 + \sum_j \gamma_j \cdot G_{i,j}$$

- **If your interaction model was correct** then:

$$S_i \;=\; \Delta(S_i) \;\text{ and } A^*(S) = \mathbf{1}(S > 0)$$

Key Steps to Move Forward

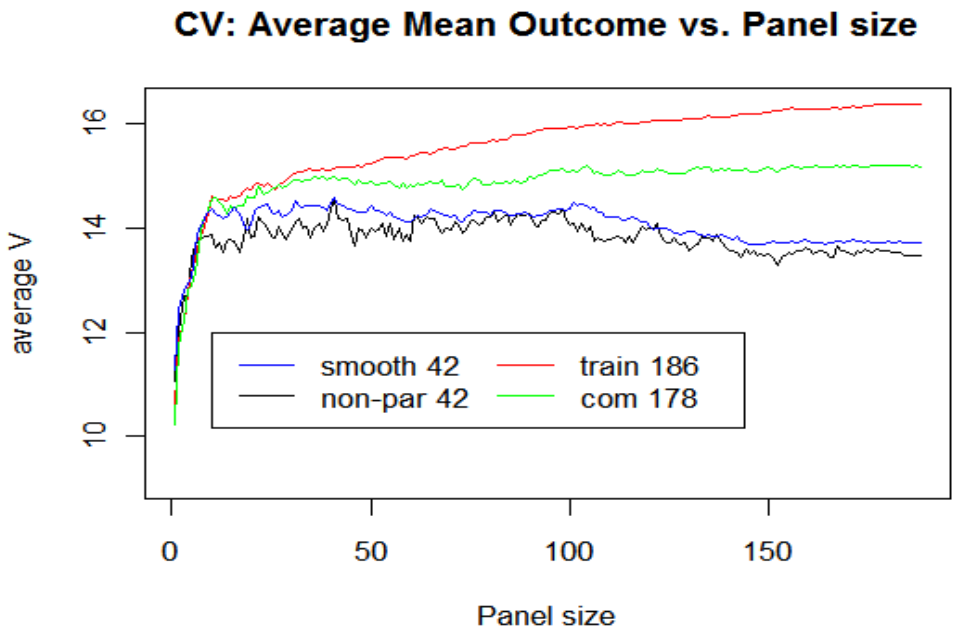
- **If your interaction model was incorrect** (it was!) then we can still:
  - ▷ Use **validation data** to evaluate $\Delta(S)$.
  - ▷ Use **validation data** to estimate μ_{A^*}.
  - ▷ Use **validation data** to estimate Δ^* and Ψ.
- The score may still be useful.
- We can use the above estimates to see if one score perfoms better than another...

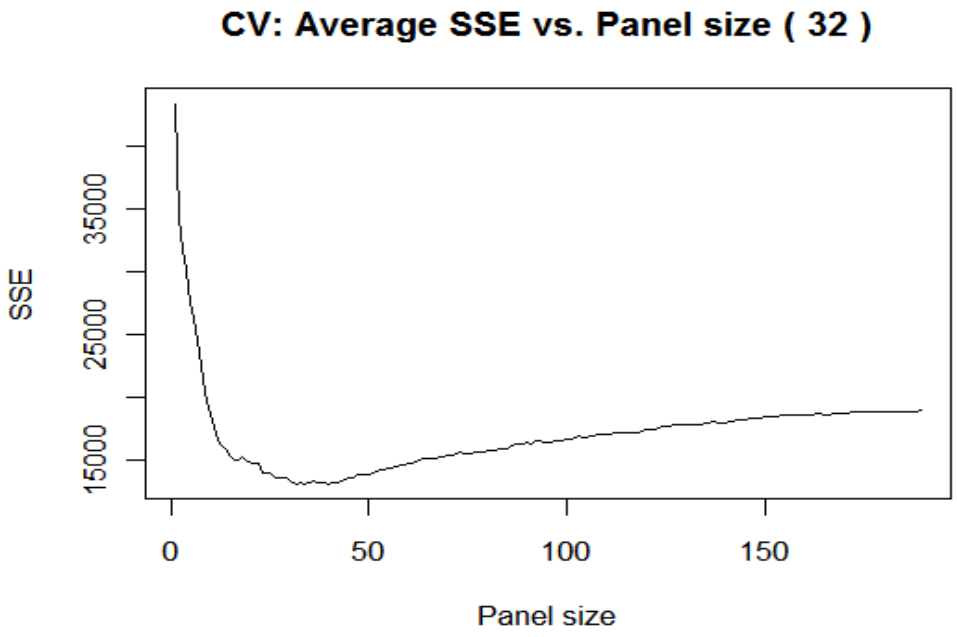
Our Approach (so far)

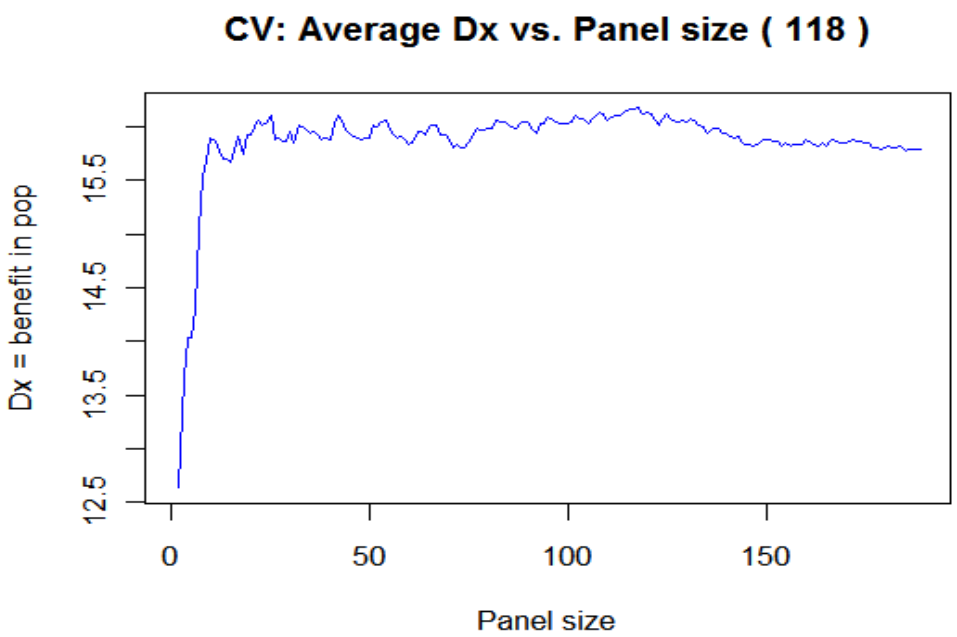
- Interaction terms: $X_i \cdot G_{i,j}$ and $(1 - X_i) \cdot G_{i,j}$
- Use 10-fold cross-validation
- For $p = 1, 2, \ldots, m$ use Lasso (or alternative) to generate a sequence of treatment benefit scores:

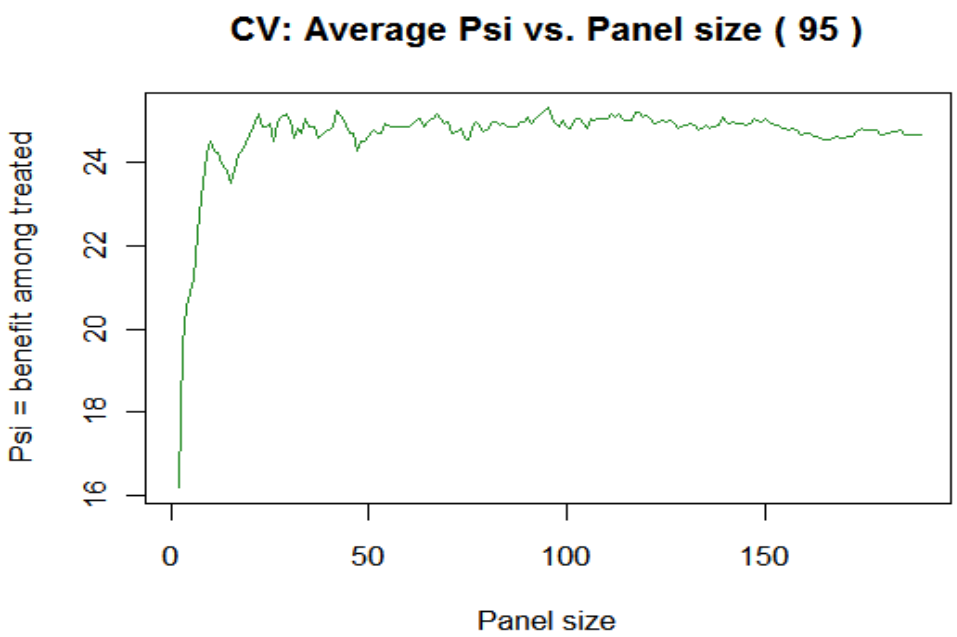
$$S_i^p = \gamma_0^p + \sum_k \gamma_k^p \cdot G_{i,(j)}$$

- Use the 10-fold validation data sets to measure performance in terms of mean outcome (μ_{A^*}), and average guided treatment effects (Δ^* and Ψ).
- Choose a marker panel size (p) that is best.


CV: Average Mean Outcome vs. Panel size
average V
10
12
14
16
0
50
100
150
Panel size
smooth 42
non-par 42
train 186
com 178


CV: Average SSE vs. Panel size (32)
SSE
15000
25000
35000
0
50
100
150
Panel size


CV: Average Dx vs. Panel size (118)
Dx = benefit in pop
12.5
13.5
14.5
15.5
0
50
100
150
Panel size

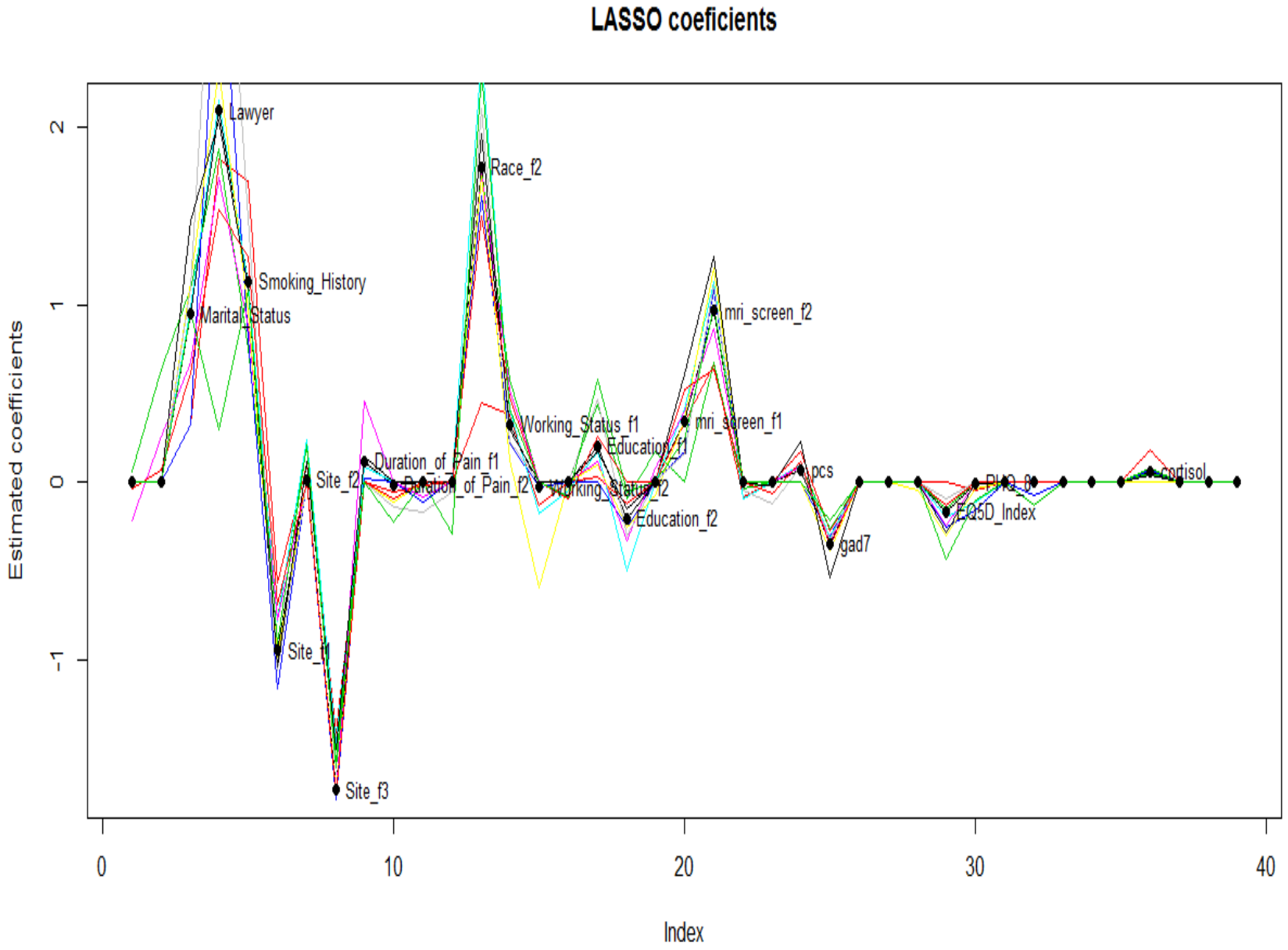

CV: Average Psi vs. Panel size (95)
Psi = benefit among treated
16
18
20
22
24
0
50
100
150
Panel size

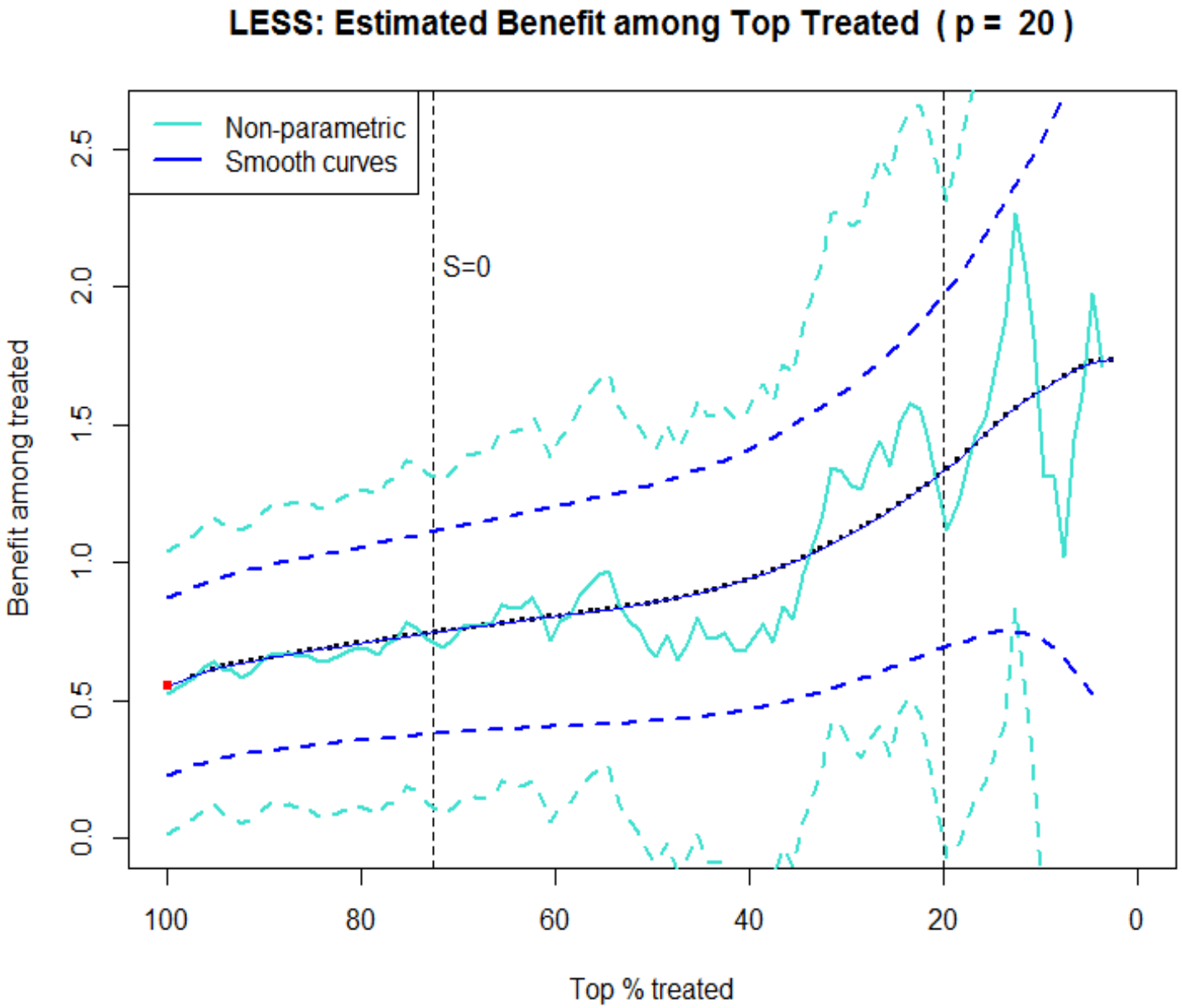
Example: LESS Trial

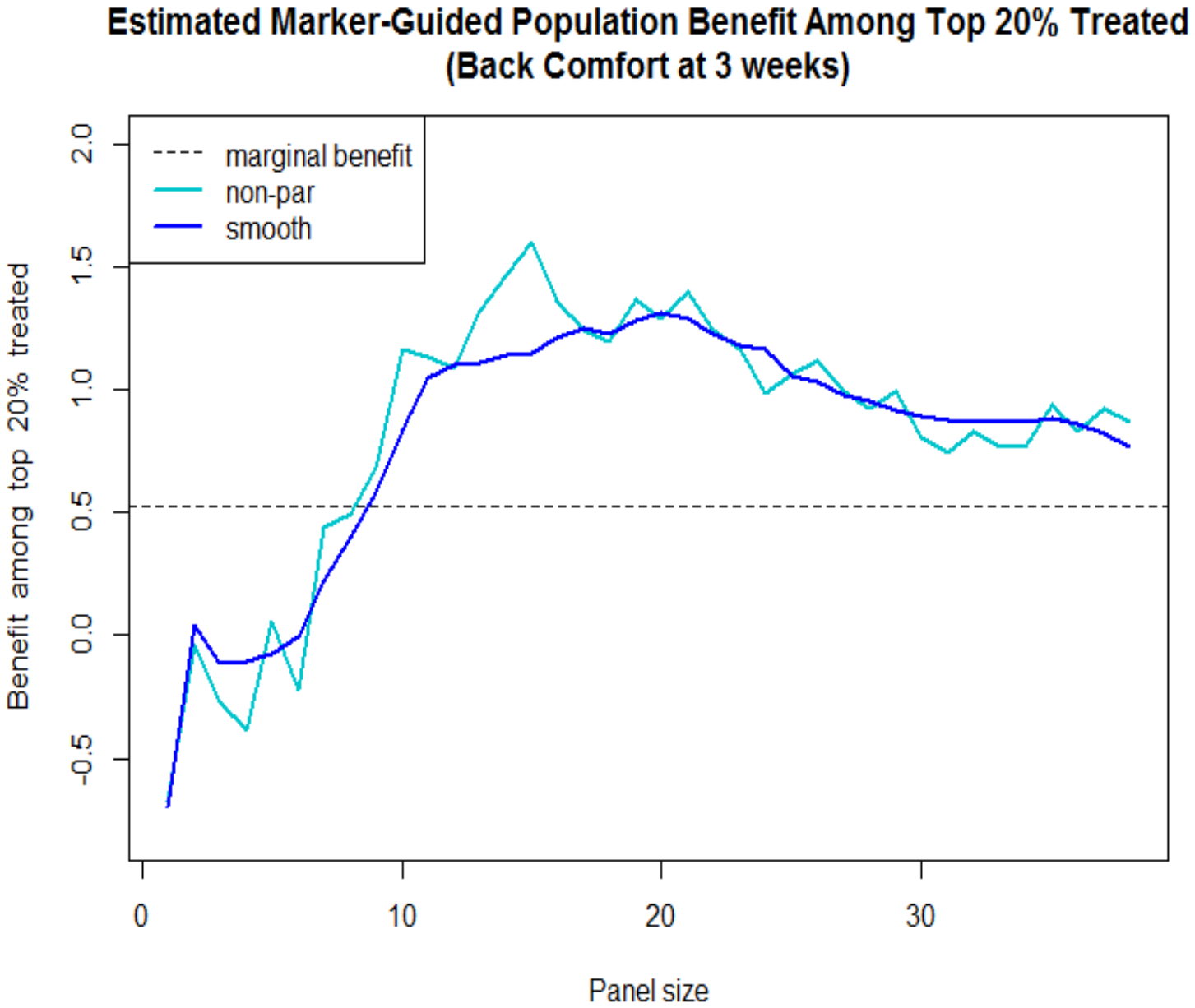
- Friedly et al. (2014) $NEJM$
- N=400 subjects randomized to epidural steroid injection or lidocaine
- Overall trial result: no differential benefit
- **Q**: subgroup of responders?

Example: LESS Trial

- Strategy:
  - Use lasso to develop sequence of treatment response scores ("predictors")
  - Evaluation use of score toward optimizing the mean population outcome using a guided treatment strategy ("policy")
  - Evaluation of treatment benefit among treated using various thresholds for treatment ("enrichment")


LASSO coeficients
Estimated coefficients
Index
Lawyer
Smoking_History
Marital_Status
Site_f2
Site_f1
Site_f3
Duration_of_Pain_f1
Duration_of_Pain_f2
Race_f2
Working_Status_f1
Working_Status_f2
Education_f1
Education_f2
mri_screen_f2
mri_screen_f1
pcs
gad7
EQ5D_Index
cortisol


LESS: Estimated Benefit among Top Treated (p = 20)
Non-parametric
Smooth curves
S=0
Benefit among treated
0.0
0.5
1.0
1.5
2.0
2.5
100
80
60
40
20
0
Top % treated


Estimated Marker-Guided Population Benefit Among Top 20% Treated
(Back Comfort at 3 weeks)
marginal benefit
non-par
smooth
Benefit among top 20% treated
2.0
1.5
1.0
0.5
0.0
-0.5
0
10
20
30
Panel size

Summary

- We have defined **statistical measures** that reflect the intended **clinical goals**.
- Algorithm / comparison
- Illustration using genetic marker data (in process).
- **Q**: What about alternative and specific selection strategies?
- **Veronika Skrivankova**
- U01 HG005157, P01 CA053996-34, U54 RR024379