Variable Selection for Decision-Making in Individualised Treatments

Professor Howard Bondell

School of Mathematics and Statistics
University of Melbourne
howard.bondell@unimelb.edu.au

28 May 2020

Outline

- Variable Selection for Decision-Making in Individualised Treatments
  - What is Personalised Medicine?
  - Optimal Decision Rules
  - Test-Based Selection of Variables for Decision-Making
  - Application to Clinical Trial

Personalised Medicine

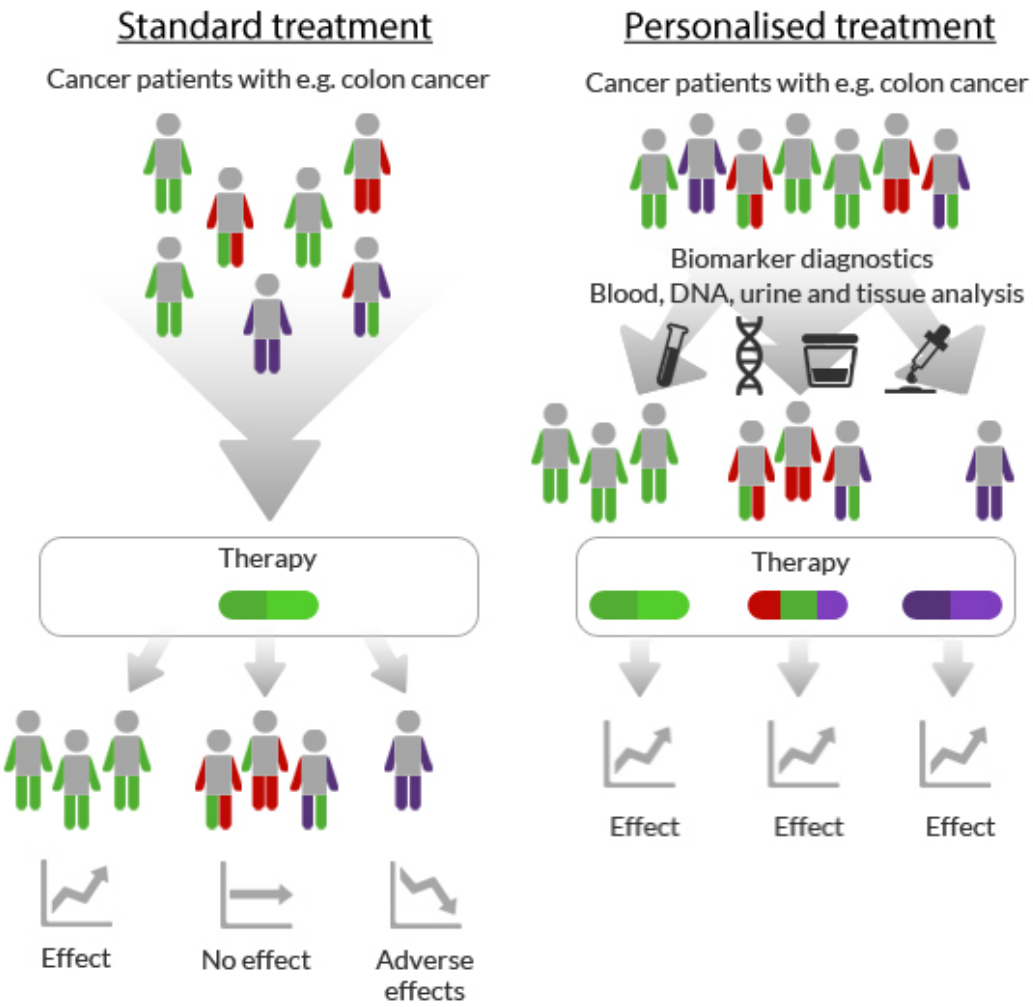


*Image adapted from the DNA Research Center

Personalised Medicine

- Same treatment is not always best for all patients.
- Which treatment should be assigned to which patient?

Personalised Medicine

- Same treatment is not always best for all patients.
- Which treatment should be assigned to which patient?
  - Depends on the patient's characteristics:
    - Age.
    - Sex.
    - Medical history.
    - Genetic information.
    - ...

Personalised Medicine

- Same treatment is not always best for all patients.
- Which treatment should be assigned to which patient?
  - Depends on the patient's characteristics:
    - Age.
    - Sex.
    - Medical history.
    - Genetic information.
    - ...
  - Collecting and storing all information can complicate decision process and add cost.
- **Goal:** Identify characteristics important for treatment decisions.

Motivating Example

- AIDS Clinical Trials Group (ACTG) 175:

  Randomized, double-blind, placebo-controlled trial to compare efficacy among treatments in HIV-infected patients.

- ACTG175 dataset contains 16 baseline covariates on 1,100 patients including: age, sex, previous treatment history, etc.

Motivating Example

- AIDS Clinical Trials Group (ACTG) 175:

  Randomized, double-blind, placebo-controlled trial to compare efficacy among treatments in HIV-infected patients.

- ACTG175 dataset contains 16 baseline covariates on 1,100 patients including: age, sex, previous treatment history, etc.

Mean (standard error) of treatment effect at 96 weeks

Treatment	CD4 Cell Count Ratio to Baseline
Zidovudine (AZT)	0.798 (0.025)
AZT + Didanosine (ddI)	1.009 (0.028)
AZT + Zalcitabine (ddC)	1.000 (0.025)

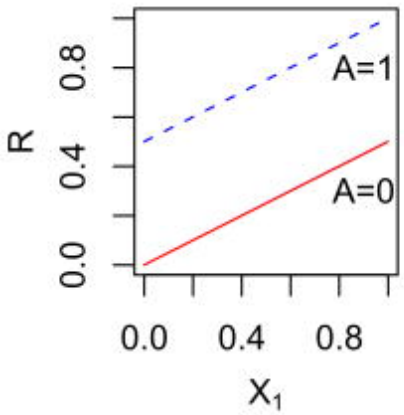
- Overall, combination of AZT + ddI is best (not significantly). But can we identify subgroups of patients who would benefit from the other combination, AZT + ddC?

Variable Selection for Treatment Decisions

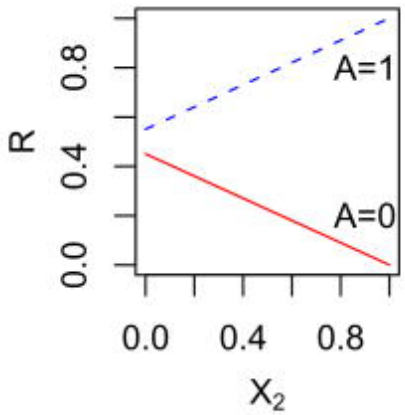
- Typical variable selection considers variables that affect the response, i.e. those that help predict higher/lower CD4 count.
- Here, we seek variables to help determine treatment to employ.

Variable Selection for Treatment Decisions

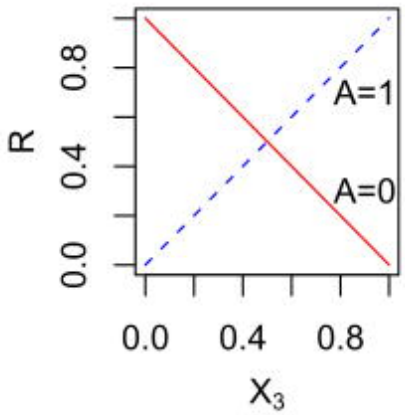
- Typical variable selection considers variables that affect the response, i.e. those that help predict higher/lower CD4 count.
- Here, we seek variables to help determine treatment to employ.



(a) No interaction



(b) Non-qualitative interaction



(c) Qualitative Interaction

Formal Objective

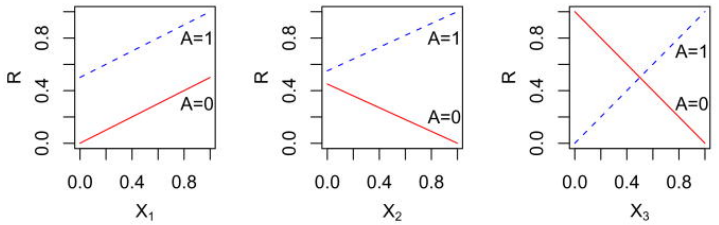
- *Optimal treatment* for a patient: treatment allocation which maximizes expected response.
- Decision rule will be a function of the variables $\boldsymbol{X}$.
- Want decision rule to yield optimal treatment for each patient.

Formal Objective

- *Optimal treatment* for a patient: treatment allocation which maximizes expected response.
- Decision rule will be a function of the variables $\boldsymbol{X}$.
- Want decision rule to yield optimal treatment for each patient.
- **Goal:** Find smallest subset of variables, $\boldsymbol{X}_S$, that results in rule that matches optimal treatment allocation.
  - These are only variables relevant to decision-making.

Simple Example

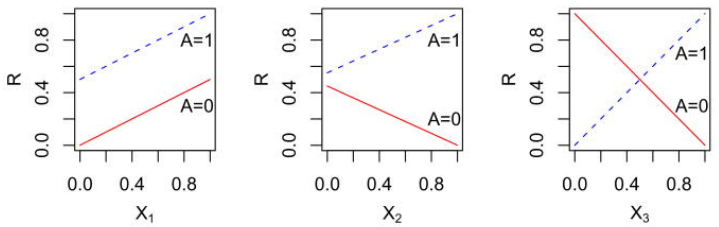
- Decision rule here is simple:
  - If $X_3 > 0.5$ then give Treatment 1. Otherwise Treatment 0.



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

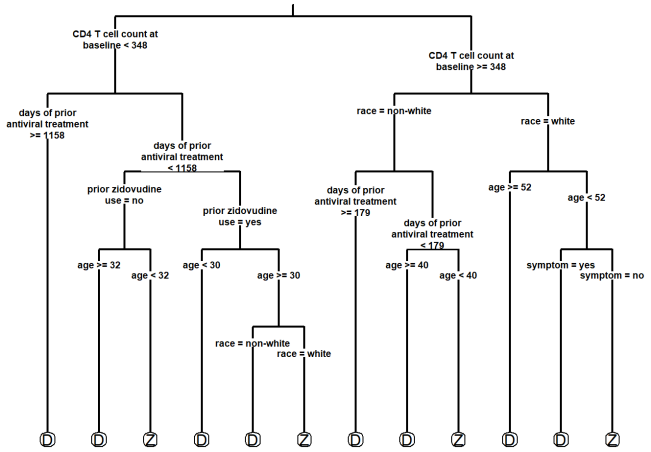
Simple Example

- Decision rule here is simple:
  - If $X_3 > 0.5$ then give Treatment 1. Otherwise Treatment 0.



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

- In general, it would be more complex.



Common Approaches

1. Variable selection as usual:
   Regularisation, Cross-Validation, AIC/BIC.
   - Targets on prediction error, not decision-making subset.

Common Approaches

1. Variable selection as usual:
   Regularisation, Cross-Validation, AIC/BIC.
   - Targets on prediction error, not decision-making subset.
2. Include interactions between treatment and each variable.
   - Perform variable selection only on interaction terms.
   - Better, but still targets prediction error.
   - Variable may interact with treatment, but not be qualitative interaction.

Common Approaches

1. Variable selection as usual:
   Regularisation, Cross-Validation, AIC/BIC.
   - Targets on prediction error, not decision-making subset.
2. Include interactions between treatment and each variable.
   - Perform variable selection only on interaction terms.
   - Better, but still targets prediction error.
   - Variable may interact with treatment, but not be qualitative interaction.
3. Univariate scores based on magnitudes and locations of crossing in interaction plots.
   - Rank variables based on individual scores.
   - Decide on where to make cutoff.

Proposed Approach: "No Regret"

- Assume we know the true model and parameters.
- Let $a^\star(\boldsymbol{X})$ denote optimal treatment rule.
- Then $a^\star(\boldsymbol{X}) = 1$ if $E[Y|\boldsymbol{X}, A = 1] > E[Y|\boldsymbol{X}, A = 0]$.
  - Note: $a^\star(\boldsymbol{X})$ is a function of true parameters.

Proposed Approach: "No Regret"

- Consider "Regret" from using subset $\boldsymbol{X}_S$ instead of full optimal.
- **Idea:** For every patient, compare response obtained using policy based only on $\boldsymbol{X}_S$ to optimal response.

Proposed Approach: "No Regret"

- Consider "Regret" from using subset $\boldsymbol{X}_S$ instead of full optimal.
- **Idea:** For every patient, compare response obtained using policy based only on $\boldsymbol{X}_S$ to optimal response.
- Formally define parameter, R_S:

$$R_S = E_{\boldsymbol{X}} \left\{ E[Y|\boldsymbol{X}, A = a^\star(\boldsymbol{X})] - E[Y|\boldsymbol{X}, A = a^*_S(\boldsymbol{X}_S)] \right\},$$

$a^\star(\boldsymbol{X})$ is optimal treatment rule.
$a^*_S(\boldsymbol{X}_S)$ is best available treatment rule using only subset $\boldsymbol{X}_S$.

Proposed Approach: "No Regret"

- Consider "Regret" from using subset $\boldsymbol{X}_S$ instead of full optimal.
- **Idea:** For every patient, compare response obtained using policy based only on $\boldsymbol{X}_S$ to optimal response.
- Formally define parameter, R_S:

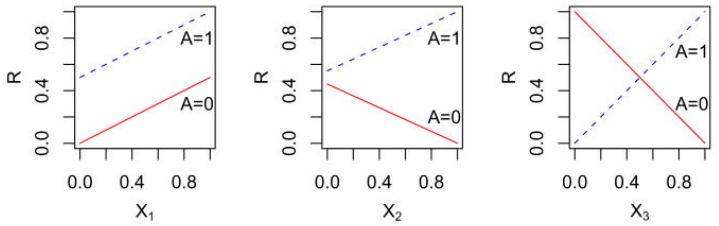
$$R_S = E_{\boldsymbol{X}} \left\{ E[Y|\boldsymbol{X}, A = a^\star(\boldsymbol{X})] - E[Y|\boldsymbol{X}, A = a_S^*(\boldsymbol{X}_S)] \right\},$$

  $a^\star(\boldsymbol{X})$ is optimal treatment rule.
  $a_S^*(\boldsymbol{X}_S)$ is best available treatment rule using only subset $\boldsymbol{X}_S$.

- **Goal:** No Regret!
  - Formal hypothesis test on univariate parameter, $H_0 : R_S = 0$.

Simple Example of Regret

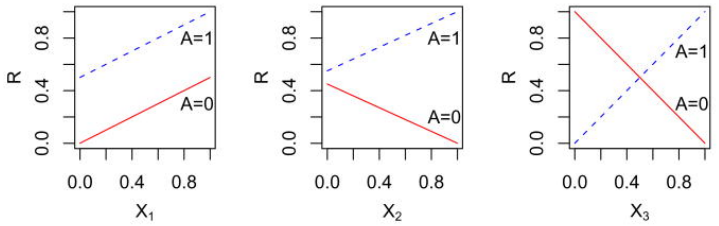
- Here, decision based on full model is
  - If $X_3 > 0.5$ then give Treatment 1. Otherwise Treatment 0.



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

Simple Example of Regret

- Here, decision based on full model is
  - If $X_3 > 0.5$ then give Treatment 1. Otherwise Treatment 0.

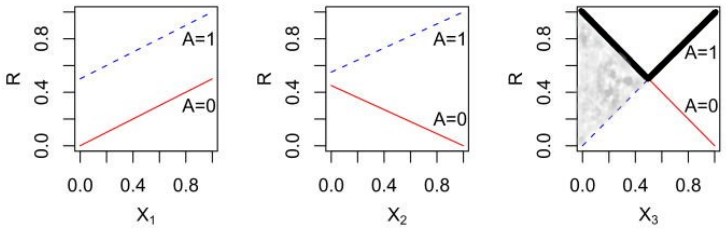


(a) No interaction (b) Non-qualitative interaction (c) Qualitative Interaction

- Consider subset with X_1 and X_2 only.
  - Then decision rule is: ALWAYS give treatment 1.

Simple Example of Regret

- Here, decision based on full model is
  - If $X_3 > 0.5$ then give Treatment 1. Otherwise Treatment 0.

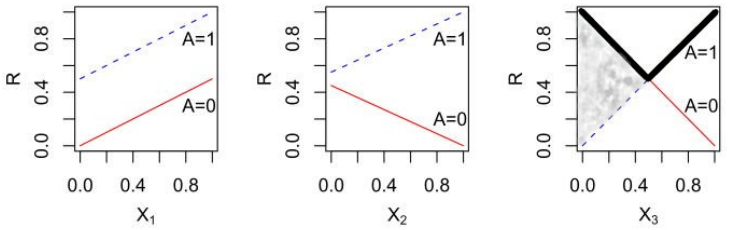


(a) No interaction (b) Non-qualitative interaction (c) Qualitative Interaction

- Consider subset with X_1 and X_2 only.
  - Then decision rule is: ALWAYS give treatment 1.
  - Apply this rule.
  - Regret is (weighted) difference between curves!

Simple Example of Regret

- Here, decision based on full model is
  - If $X_3 > 0.5$ then give Treatment 1. Otherwise Treatment 0.



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

- Consider subset with X_1 and X_2 only.
  - Then decision rule is: ALWAYS give treatment 1.
  - Apply this rule.
  - Regret is (weighted) difference between curves!

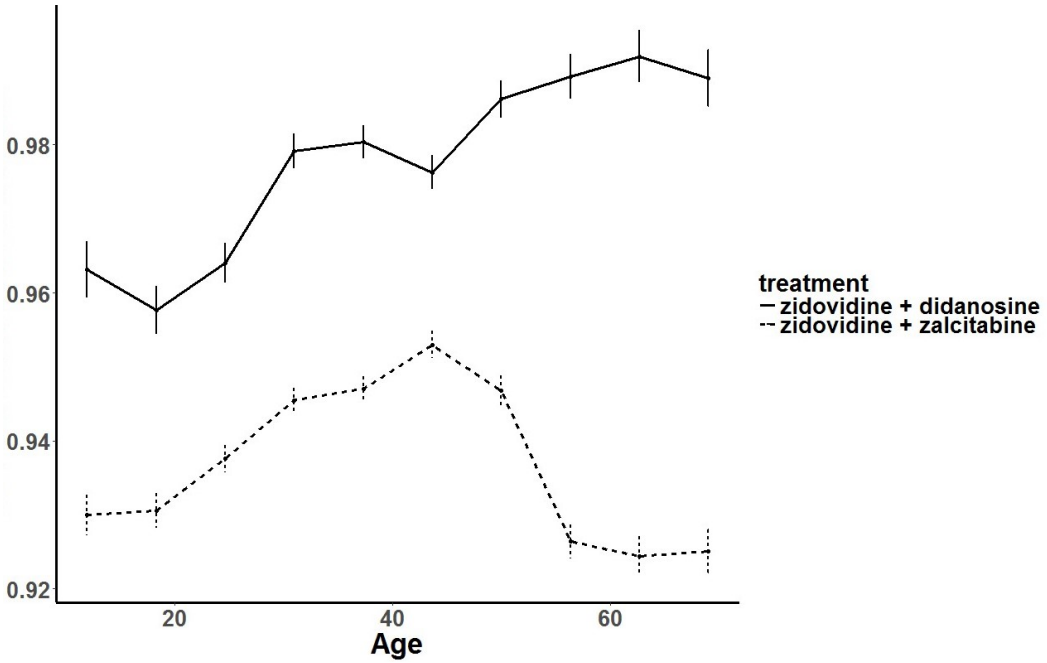
- If we throw out X_1 and X_2 and just keep X_3, NO REGRET.

Procedure

- To conduct hypothesis test, $H_0 : R_S = 0$, need model.

Procedure

- To conduct hypothesis test, $H_0 : R_S = 0$, need model.
- To allow for flexible model, use nonparametric regression for response given each treatment condition.
  - Any is suitable, choose your favourite!



Procedure

- Can now test any pair of nested subsets.
- $R_S = 0$ denotes that reduced subset is sufficient for decision-making.
- How to test $H_0 : R_S = 0$?

Procedure

- Can now test any pair of nested subsets.
- $R_S = 0$ denotes that reduced subset is sufficient for decision-making.
- How to test $H_0 : R_S = 0$?
- Univariate parameter, R_S, is function of parameters for full and reduced models.
- Can estimate by $\widehat{R_S}$ via plugging in parameter estimates.

Procedure

- Can now test any pair of nested subsets.
- $R_S = 0$ denotes that reduced subset is sufficient for decision-making.
- How to test $H_0 : R_S = 0$?
- Univariate parameter, R_S, is function of parameters for full and reduced models.
- Can estimate by $\widehat{R_S}$ via plugging in parameter estimates.
- Using (asymptotic) joint distribution of estimated parameters, transform to obtain distribution of $\widehat{R_S}$.

Backward Elimination Procedure

Application: Imbed testing in backward elimination.

Backward Elimination Procedure

Application: Imbed testing in backward elimination.

- Step 1: Estimate parameters in model.

Backward Elimination Procedure

Application: Imbed testing in backward elimination.

- Step 1: Estimate parameters in model.
- Step 2: For each patient, now have predicted response under each treatment. So have estimated treatment rule.

Backward Elimination Procedure

Application: Imbed testing in backward elimination.

- Step 1: Estimate parameters in model.
- Step 2: For each patient, now have predicted response under each treatment. So have estimated treatment rule.
- Step 3: Remove one predictor to get reduced subset, $\boldsymbol{X}_S$, and redo estimation under reduced model.

Backward Elimination Procedure

Application: Imbed testing in backward elimination.

- Step 1: Estimate parameters in model.
- Step 2: For each patient, now have predicted response under each treatment. So have estimated treatment rule.
- Step 3: Remove one predictor to get reduced subset, $\boldsymbol{X}_S$, and redo estimation under reduced model.
- Step 4: Using the two decision rules, can directly compute plug-in estimate of Regret.

Backward Elimination Procedure

Application: Imbed testing in backward elimination.

- Step 1: Estimate parameters in model.
- Step 2: For each patient, now have predicted response under each treatment. So have estimated treatment rule.
- Step 3: Remove one predictor to get reduced subset, $\boldsymbol{X}_S$, and redo estimation under reduced model.
- Step 4: Using the two decision rules, can directly compute plug-in estimate of Regret.
- Step 5: Repeat for all predictors, and choose the resulting subset with smallest Regret.

Backward Elimination Procedure

Application: Imbed testing in backward elimination.

- Step 1: Estimate parameters in model.
- Step 2: For each patient, now have predicted response under each treatment. So have estimated treatment rule.
- Step 3: Remove one predictor to get reduced subset, $\boldsymbol{X}_S$, and redo estimation under reduced model.
- Step 4: Using the two decision rules, can directly compute plug-in estimate of Regret.
- Step 5: Repeat for all predictors, and choose the resulting subset with smallest Regret.
- Step 6: Repeat treating current model as full model.

Backward Elimination Procedure

Application: Imbed testing in backward elimination.

- Step 1: Estimate parameters in model.
- Step 2: For each patient, now have predicted response under each treatment. So have estimated treatment rule.
- Step 3: Remove one predictor to get reduced subset, $\boldsymbol{X}_S$, and redo estimation under reduced model.
- Step 4: Using the two decision rules, can directly compute plug-in estimate of Regret.
- Step 5: Repeat for all predictors, and choose the resulting subset with smallest Regret.
- Step 6: Repeat treating current model as full model.
- Continue until desired stopping criterion met.

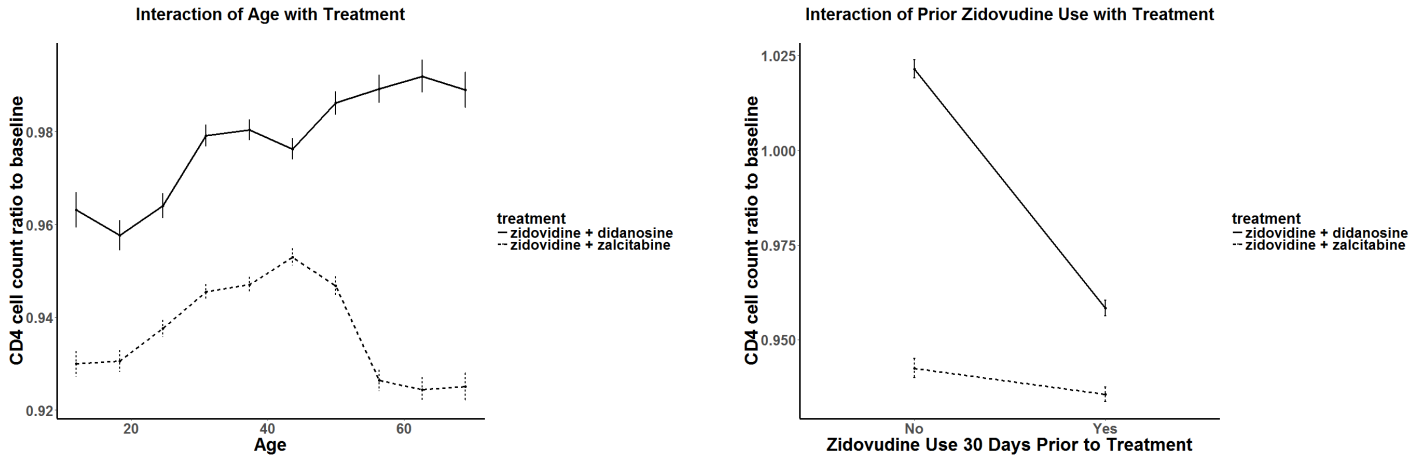
ACTG175 Analysis

- Two treatments for comparison:
  1. AZT + ddI (Best on average across all patients).
  2. AZT + ddC.
- Using 16 candidate predictors in Backward Elimination removes 10 and keeps 6.
  - Next one has (Bonferroni corrected) p-value around 0.20.

Variables selected as important for decision making. Values in table represent p-values from using full model of size 6, and seeking to reduce to model of size 5.

Covariate	Estimated Regret	p-value	Corrected p-value
number of days of previous antiretroviral therapy	0.01659	0.0029	0.0174
symptomatic indicator	0.02172	0.0004	0.0020
CD4 T cell count at baseline	0.03062	≤ 0.0001	≤ 0.0001
race	0.03787	≤ 0.0001	≤ 0.0001
indicator of prior AZT use	0.07317	≤ 0.0001	≤ 0.0001
age	0.07393	≤ 0.0001	≤ 0.0001

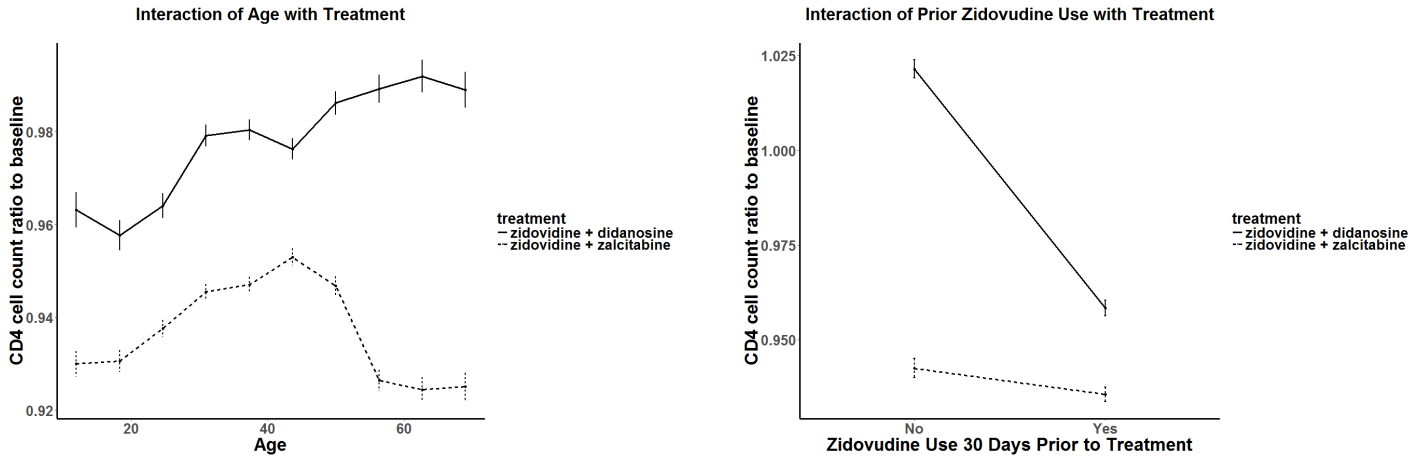
Marginal Covariates-Treatment Interaction



Marginal Covariates-Treatment Interaction Plot

Treatment response surfaces are predicted based on regression for each treatment, and marginal interaction plots are generated via Monte Carlo integration over the other covariates.

Marginal Covariates-Treatment Interaction

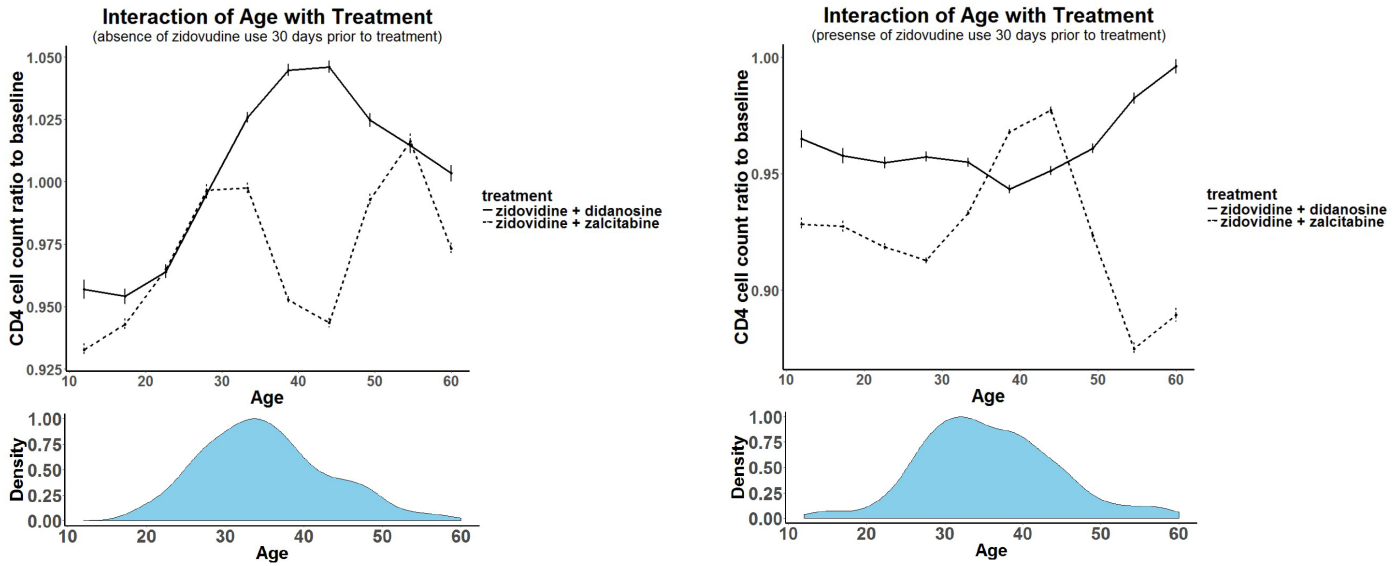


Marginal Covariates-Treatment Interaction Plot

Treatment response surfaces are predicted based on regression for each treatment, and marginal interaction plots are generated via Monte Carlo integration over the other covariates.

- Looks like no importance of *Age* or *AZT Usage* for decision-making?

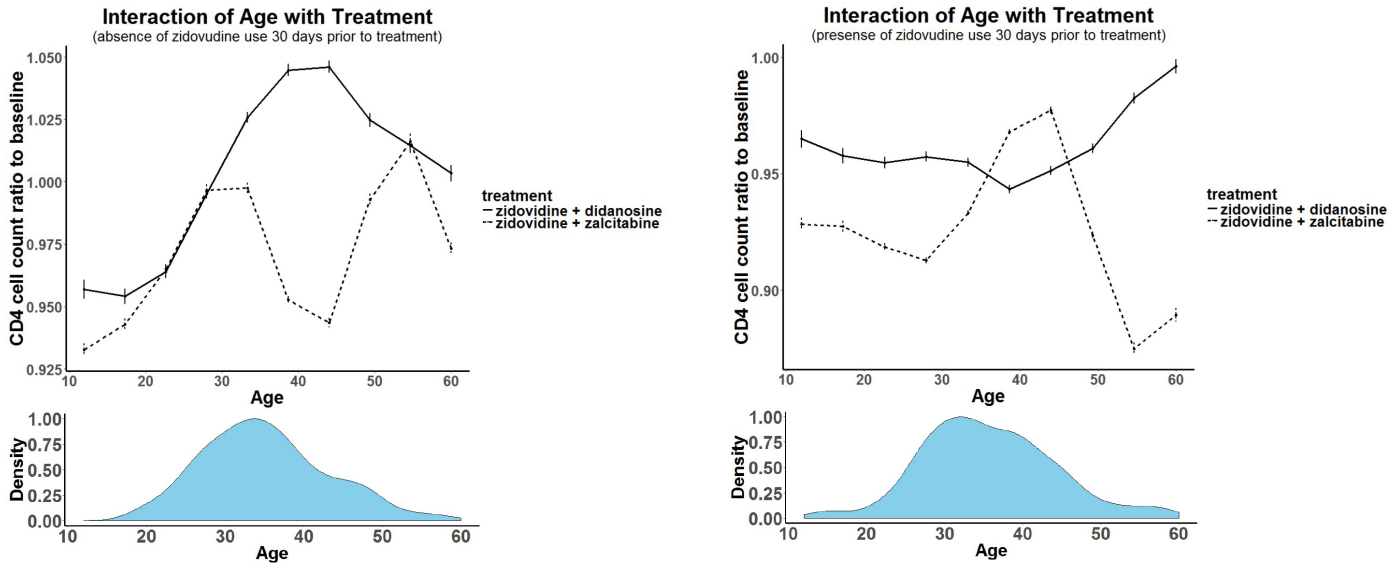
Conditional Covariates-Treatment Interaction



Conditional Covariates-Treatment Interaction Plot

Interaction of *Age* and *Treatment* stratified by *AZT Usage*, along with age distribution in each group.

Conditional Covariates-Treatment Interaction



Conditional Covariates-Treatment Interaction Plot

Interaction of *Age* and *Treatment* stratified by *AZT Usage*, along with age distribution in each group.

- Looks like a subgroup may benefit from alternate treatment!

Treatment Policy Value

- Cross-validation to evaluate the treatment policy.
- Value is estimated by computing mean response on holdout data for those observations consistent with the treatment policy found by each method.
  - Averaged over 100 splits of data (with standard deviation).

Value increase compared to AZT + ddC

Treatment Policy	Percent Patients Treated with AZT+ddI	Value Increase
All treated with AZT+ddI	100%	9.4 (3.1)
Policy from SA	51.1%	9.5 (5.3)
Policy from NP	50.4%	19.1 (4.5)

SA: sequential advantage (Fan et al., 2016). NP: Nonparametric regression variable selection procedure.

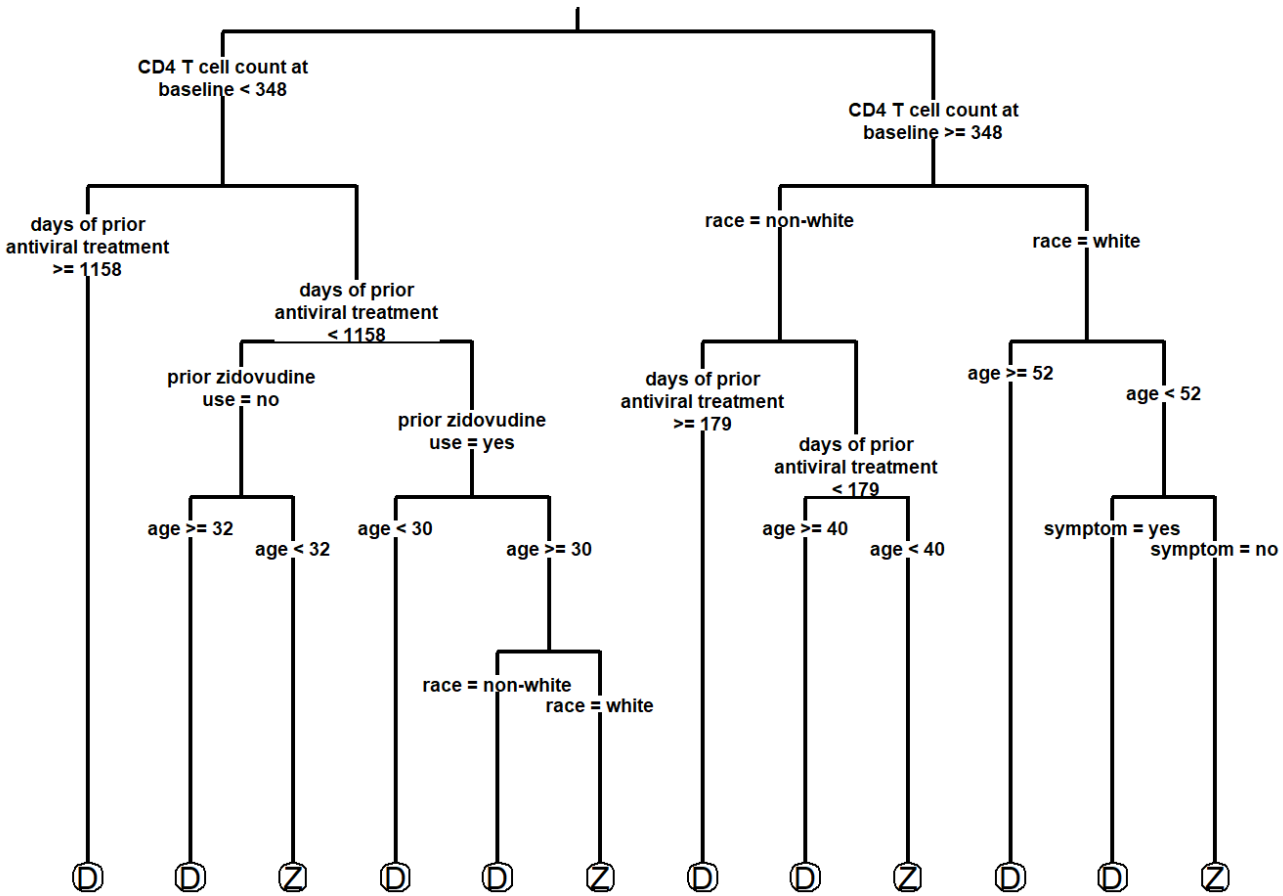
Implementation of Treatment Assignments

- How would medical professional use these results?
- Input patient values, have computer run procedure and spit out predicted best treatment.
  - May not appreciate "Black Box".

Implementation of Treatment Assignments

- How would medical professional use these results?
- Input patient values, have computer run procedure and spit out predicted best treatment.
  - May not appreciate "Black Box".
- Instead, take results from this estimated policy and provide friendly interface.
  - Now have binary response data, i.e. predicted best treatment for each patient in data.
  - Construct decision tree from this new dataset.
  - Provide this visual representation of treatment rule.

Treatment Policy Tree



Decision tree approximation to the optimal treatment policy.
Z: AZT+ddC. D: AZT+ddI.

Further Issues

- Two (related) questions you may ask:
  1. What happens with a large number of predictors? Backward elimination may not be well-suited.
  2. Why backward elimination, not forward selection?

Further Issues

- Two (related) questions you may ask:
  1. What happens with a large number of predictors? Backward elimination may not be well-suited.
  2. Why backward elimination, not forward selection?
- First is straightforward.
  - Use screening method to reduce number of predictors.
  - Here, screen predictors in each treatment group separately.
  - Only variables deemed to have non-zero effect in at least one group are kept.

Further Issues

- Two (related) questions you may ask:
  1. What happens with a large number of predictors? Backward elimination may not be well-suited.
  2. Why backward elimination, not forward selection?
- First is straightforward.
  - Use screening method to reduce number of predictors.
  - Here, screen predictors in each treatment group separately.
  - Only variables deemed to have non-zero effect in at least one group are kept.
- Second, a bit more subtle.

Why Backward Elimination?

- For forward selection, we consider addition of each possible variable, one-at-a-time.
  - Need to fit model for each candidate subset obtained by adding one predictor.
  - Need to do this at every step.
  - Depending on method, fitting can be expensive, although can be done in parallel.

Why Backward Elimination?

- For forward selection, we consider addition of each possible variable, one-at-a-time.
  - Need to fit model for each candidate subset obtained by adding one predictor.
  - Need to do this at every step.
  - Depending on method, fitting can be expensive, although can be done in parallel.
- For backward elimination, same issue.
  - **So what's the difference?**

Why Backward Elimination?

- Can use approximation to backward elimination.
  - Start with full fit.
  - Using joint Gaussian distribution of estimated parameters, can approximate results for any subset.

Why Backward Elimination?

- Can use approximation to backward elimination.
  - Start with full fit.
  - Using joint Gaussian distribution of estimated parameters, can approximate results for any subset.
- Removing X_j from model is equivalent to setting some parameter(s) to zero.

Why Backward Elimination?

- Can use approximation to backward elimination.
  - Start with full fit.
  - Using joint Gaussian distribution of estimated parameters, can approximate results for any subset.
- Removing X_j from model is equivalent to setting some parameter(s) to zero.
  - Obtain distribution for any subset by conditioning on the appropriate set of parameters being zero.
  - Once again, multivariate Gaussian with known conditional distribution.
- **No need to refit at all!**

Benefits:

- Selection of quantitative interactions represents variables relevant for decision-making.
- Approach directly targets difference from optimal decision.
- Can apply to multiple treatments, not just two.

Optimal Treatment Decisions in Personalized Medicine

Benefits:

- Selection of quantitative interactions represents variables relevant for decision-making.
- Approach directly targets difference from optimal decision.
- Can apply to multiple treatments, not just two.

Additional work:

- Adapt to dynamic treatments, i.e. multiple decision points.
  - Which variables?
  - Which time point?
  - What is the trigger to make a switch?