Competitive Modeling of Outcomes for Prediction

`rstudio::conf`

Max Kuhn[1] and Greg Steeno[2]

Rstudio[1], Pfizer WRD[2]

Chemical Synthesis

When a candidate drug is designed it is synthesized and purified.

These steps require experiments on a compound–to–compound basis

Once a compound is synthesized, a number of different purification *protocols* are tested to get a pure substrate.

For some types of compounds, there is a portfolio of 10 different purification protocols that *might* work and each is executed for each compound

Some Examples

Comp	A	B	C	D	E	F	G	H	I	J
40	0.0	2.4	2.7	1.8	0.0	0.0	0	0.7	0.9	6.8
97	20.6	0.0	0.0	0.0	0.0	0.0	0	0.0	0.0	0.0
99	18.9	10.1	0.0	1.1	0.0	4.0	0	13.5	11.5	0.0
114	0.0	0.0	0.0	0.0	0.0	0.0	0	0.0	0.0	0.0
262	11.9	12.0	0.0	22.5	15.3	8.7	0	8.2	23.9	0.0

Converted to Binary Classes

Comp	A	B	C	D	E	F	G	H	I	J
40	×	✓	✓	✓	×	×	×	✓	✓	✓
97	✓	×	×	×	×	×	×	×	×	×
99	✓	✓	×	✓	×	✓	×	✓	✓	×
114	×	×	×	×	×	×	×	×	×	×
262	✓	✓	×	✓	✓	✓	×	✓	✓	×

Predicting Success

We would like to predict which protocols would be successful before the compound is synthesized.

Our main approach for doing this is quantitative structure-activity relationships (QSAR)

We can compute thousands of *chemical descriptors* based on the equation for the molecule. Examples are: size, charge, greasiness, permeability,

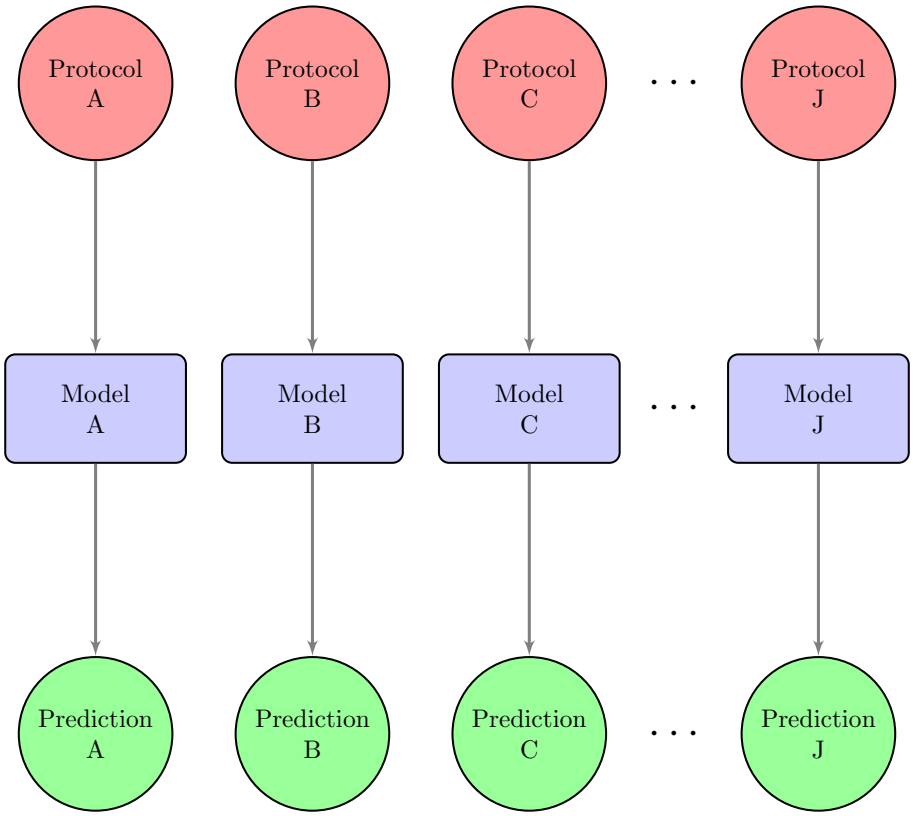
These characteristics of *structure* can be modeled in an effort to predict which protocols should be used.

For this analysis, we have 120 compounds in the training set and 60 in the test set.

Strategy 1: Direct Response Modeling

For this strategy, 10 different models are created to predict the probability that each of the protocols would be successful (since more than one might work).

Once we have the set of 10 models, each new compound can be scored on the likelihood of success


Protocol A
Protocol B
Protocol C
Protocol J
Model A
Model B
Model C
Model J
Prediction A
Prediction B
Prediction C
Prediction J

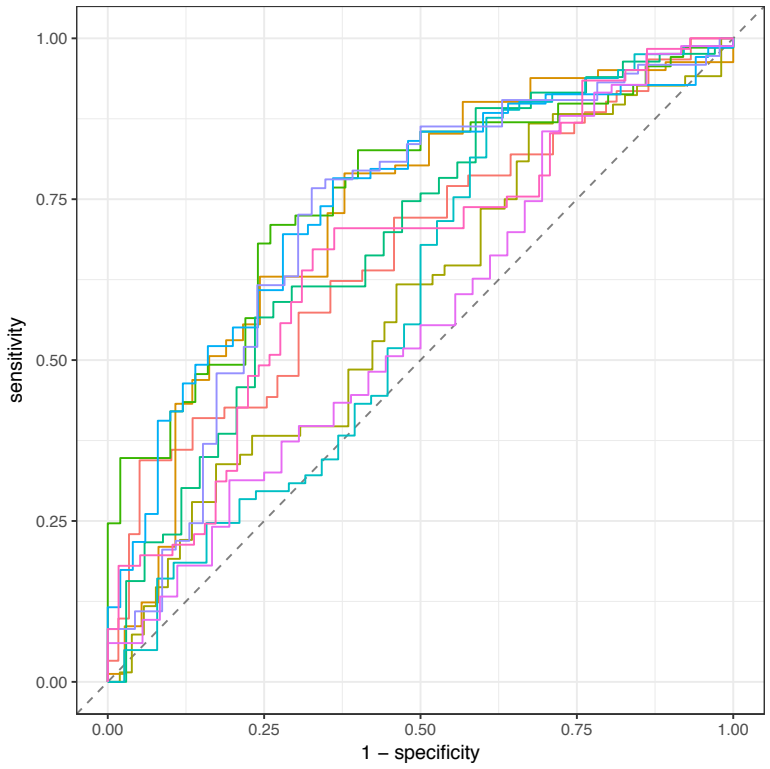
The Bad News

Molecular structure has limited utility in predicting which protocols are successful even using 1028 separate characteristics of the chemical.

A large number of models were evaluated using the `caret` package. Boosted trees and regularized logistic regression (using glmnet) seemed to do best and the latter was used.

We used ROC curves to quantify the performance of the models.

ROC Curves

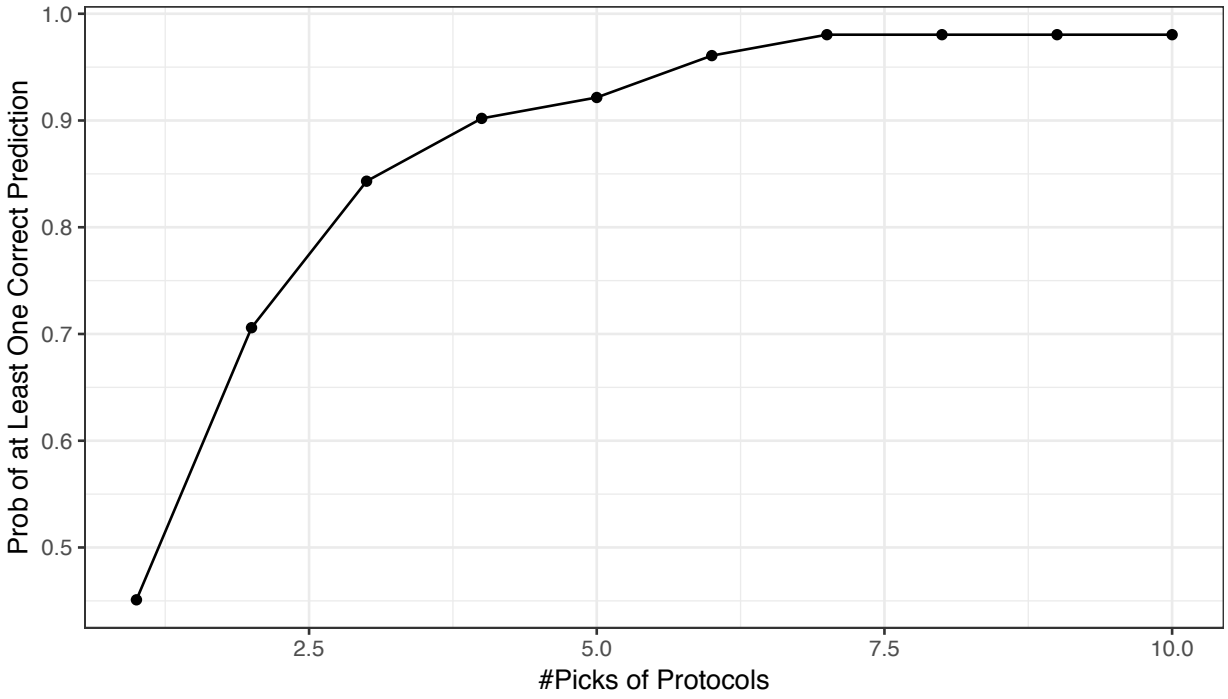

1.00
0.75
0.50
0.25
0.00
sensitivity
0.00
0.25
0.50
0.75
1.00
1 – specificity

Overall Performance

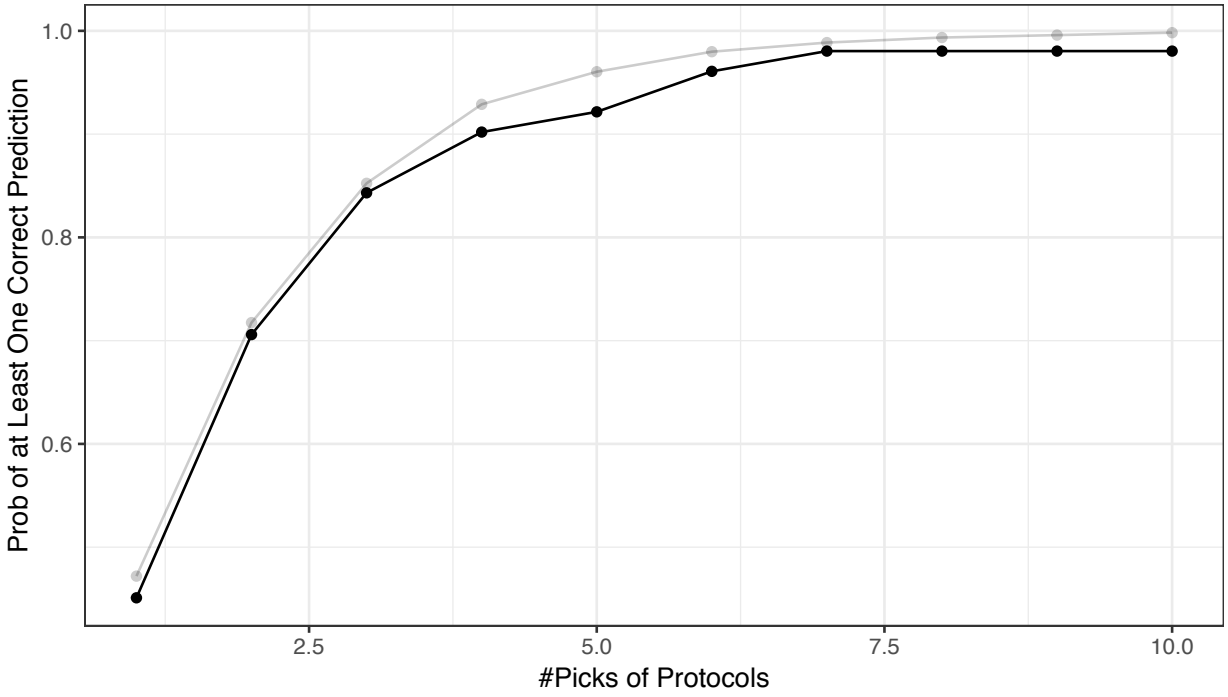
Taking this set of models, we can use the predicted class probabilities to prioritize the protocols. Performance is estimated using the test set of 60 compounds.

To quantify the set of models, we can look how well we do if we evaluate the to T protocols We can calculate the proportion of times we correctly predict that at least one good protocol.

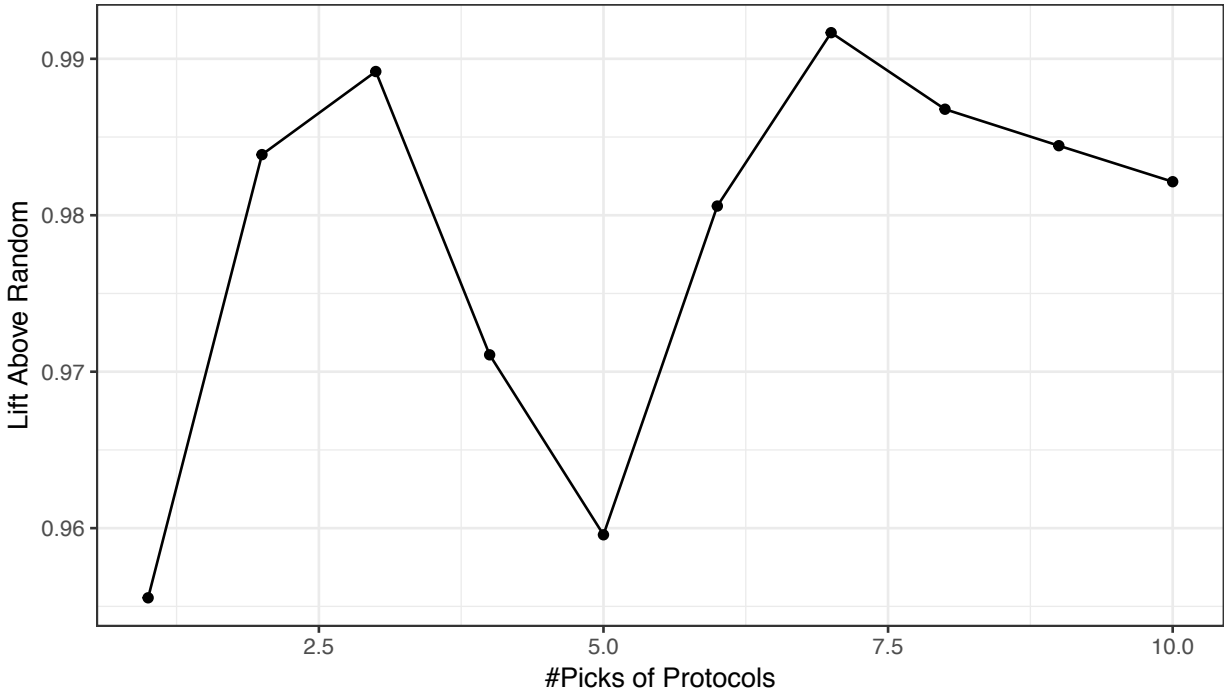
Gain Chart


1.0
0.9
0.8
0.7
0.6
0.5
Prob of at Least One Correct Prediction
2.5
5.0
7.5
10.0
#Picks of Protocols

Gain Chart with the Permutation Probability



Life Above Baseline



Strategy 2: Competitive Modeling

Instead of predicting each protocol separately, we can try to predict the best ordering of protocols to evaluate.

To do this, we can treat the data like sports.

- each protocol is a *team*
- each compound is a *tournament* or *contest*

Using this approach, we can rank the protocols into a testing sequence.

Competitions

Comp	A	B	C	D	E	F	G	H	I	J
40	0.0	2.4	2.7	1.8	0.0	0.0	0	0.7	0.9	6.8
97	20.6	0.0	0.0	0.0	0.0	0.0	0	0.0	0.0	0.0
99	18.9	10.1	0.0	1.1	0.0	4.0	0	13.5	11.5	0.0
114	0.0	0.0	0.0	0.0	0.0	0.0	0	0.0	0.0	0.0
262	11.9	12.0	0.0	22.5	15.3	8.7	0	8.2	23.9	0.0

For each compound, there are `choose(10, 2)`=45 competitions between protocols.

The Bradley–Terry Model

The Bradley–Terry model is a logistic regression model:

$$logit[Pr(Y_{ij} > Y_{ij'})] = \lambda_j - \lambda_{j'} = \beta_j$$

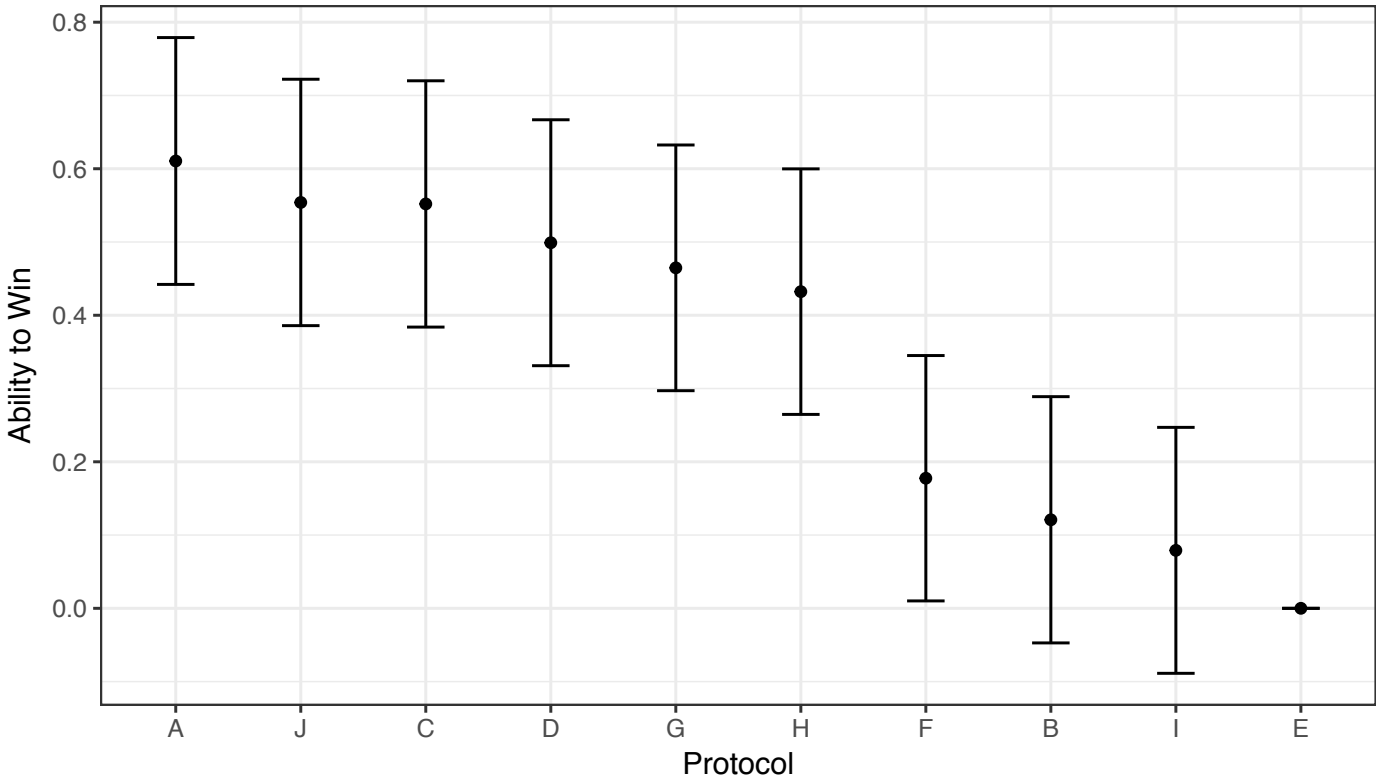
$i = 1 \ldots 120$ compounds and $j = 1 \ldots 10$ protocols.

Given a reference protocol j', this model estimates the ability of each protocol to win over the reference.

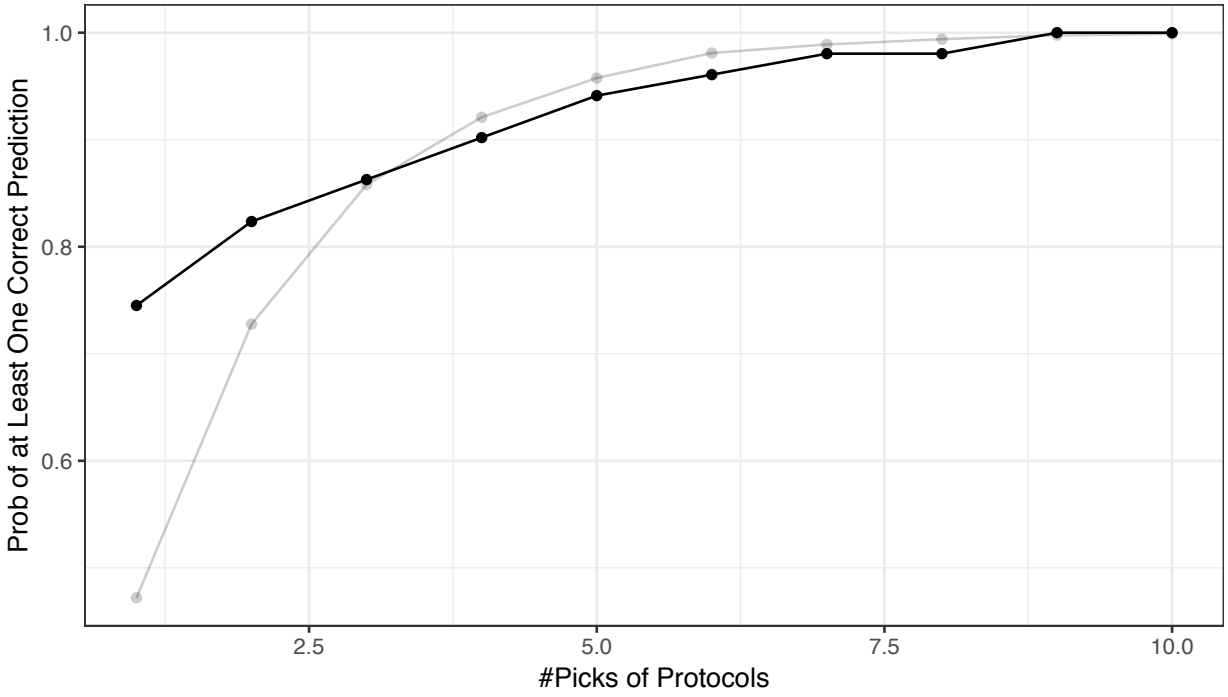
The estimated contrasts $\widehat{\beta}_j$ can be interpreted as the log–odds that protocol j has a better *ability* to purify compared to the reference protocol of j'.

We can prioritize the protocols based on their ability estimates.

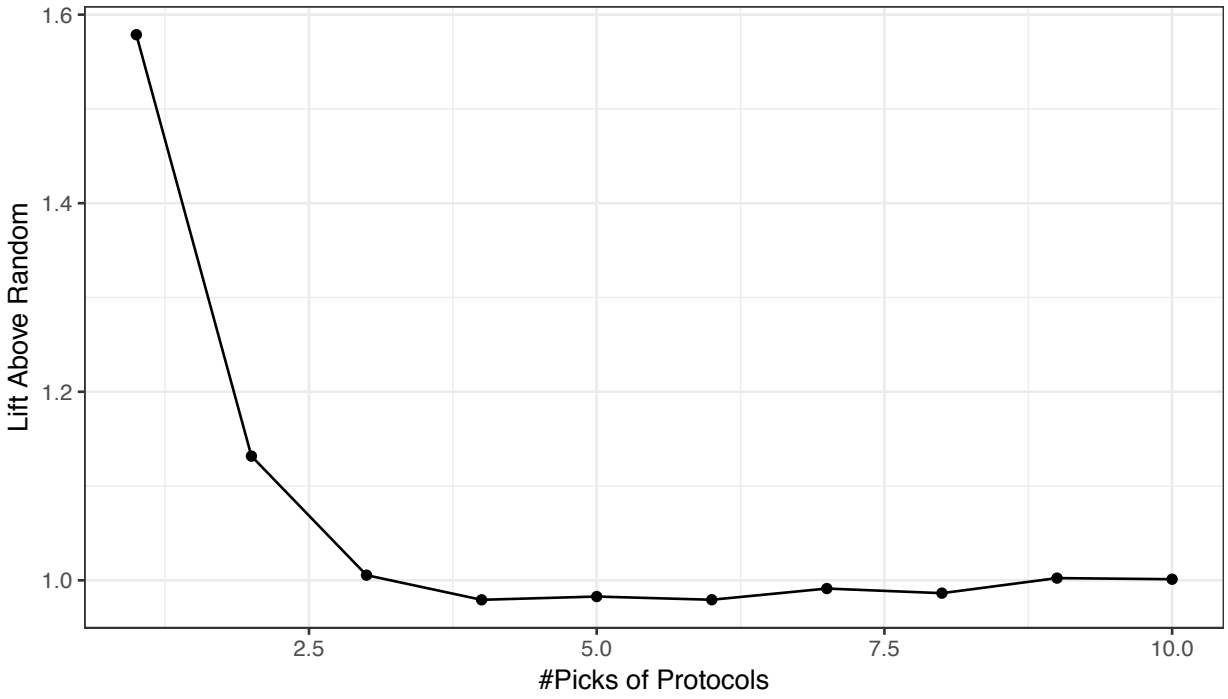
Model Results



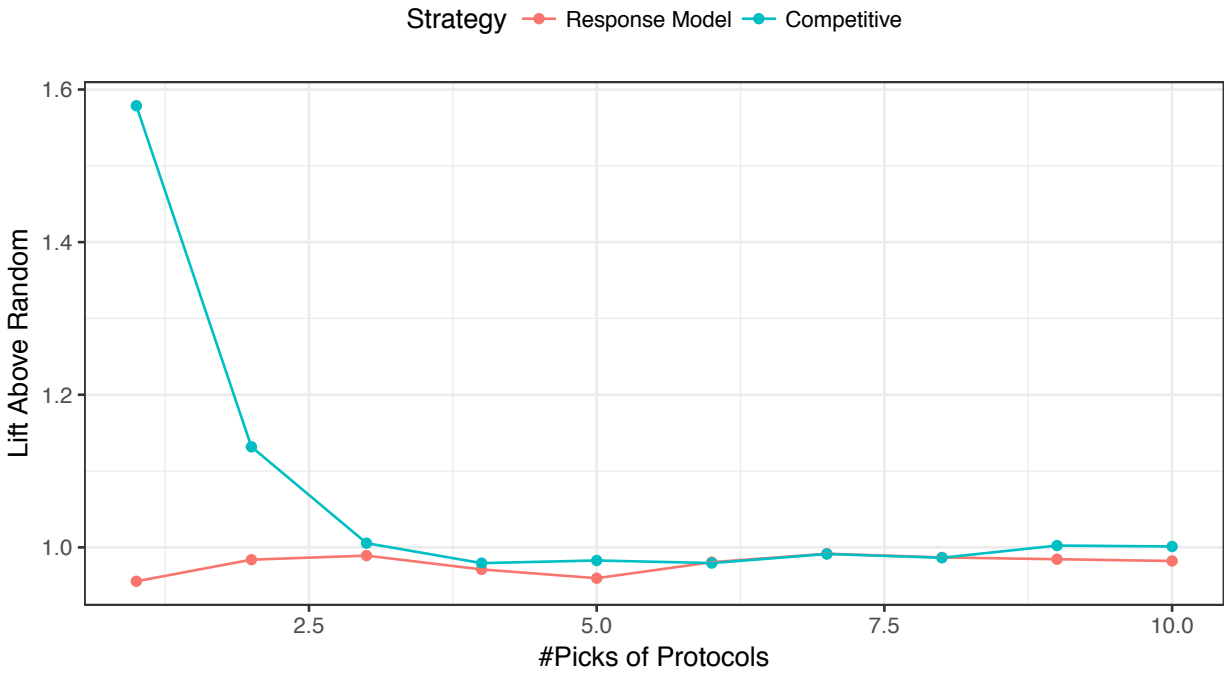
Gain Chart with the Permutation Probability



Lift Above Baseline



Lift Above Baseline



Summary

- Our predictors weren't very good for this application*
- Treating the problem as a competition helped prioritize the protocols
- Assessing the lift over random choice was the most realistic way to assess performance

*The Bradley-Terry model can also use predictors to produce compound-specific abilities but these models did worse than the simple competitive model.

Thanks

- Pfizer Medicinal Chemistry
- Heather Turner and David Firth for the `BradleyTerry2` package
- Hadley for the invitation to speak

Backup Slides

Success Rates