

# Thresholding for Biomarker Selection in Multivariate Data using Higher Criticism Supplementary Material: Simulation Results

Ron Wehrens

Pietro Franceschi

Biostatistics and Data Management

Research and Innovation Centre, Fondazione Edmund Mach

Via E. Mach 1, I-38010 San Michele all'Adige, Italy

March 26, 2012

## 1 Setting

Simulated data have been obtained from the experimental data in two steps:

1. remove the “true” biomarker variables from all eight data sets (controls, and groups one to three)
2. for each of the six spike-in sets, randomly draw 22 variables, and multiply these with factors of 1.2, 1.4 and 2.0, again randomly chosen.

One hundred spike-in sets were generated for each of the six spike-in sets. Putative biomarkers are selected using the Higher Criticism criterion for a number of different techniques: the  $t$  test, PCLDA, PLSDA and the VIP. For the multivariate methods, 2, 4 and 10 components are used. In addition, for the  $t$  test and for the VIP statistic the standard cutoff values of 0.05 and 1 are used, as alternatives to the HC-based thresholds. For calculating  $p$  values for the PCLDA, PLSDA and VIP coefficients, 10,000 permutations are used.

## 2 Standard cutoff values

Figures S1 and S2 show the results of  $t$  testing and VIP analysis using the “standard” cutoff values of 0.05 and 1, respectively, for all simulated data sets. In this setting, the VIP statistic selects many more variables than the  $t$  test. Even though the number of true positives in the selection is also highest, all VIP selections based on a threshold of 1 achieve very low efficiencies. This is true for both ionization forms and all three comparisons.

In these figures, the VIP shows no dependence on the number of latent variables at all. There are two reasons for this. The first is that the VIP is relatively insensitive to the number of components, because in the VIP these are weighted by the corresponding amount of variance, which get rapidly smaller after the first couple of components. The second aspect, in this case more important, is that any difference due to a different number of components is completely swept away by the inappropriate choice of the selection threshold: far too many variables are selected.

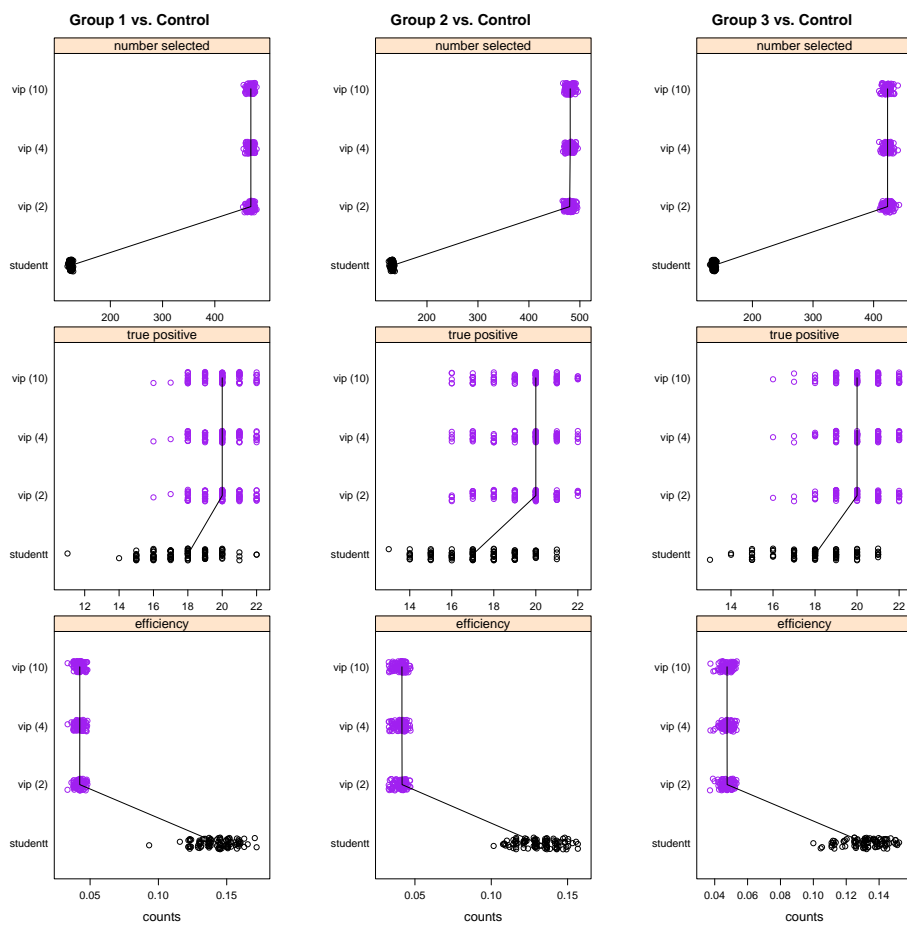


Figure S1: Biomarker selection using  $t$  tests and VIP values for the simulated **positive** ionization mode data. Standard selection thresholds of 0.05 and 1 are used for the  $t$  tests and the VIP, respectively.

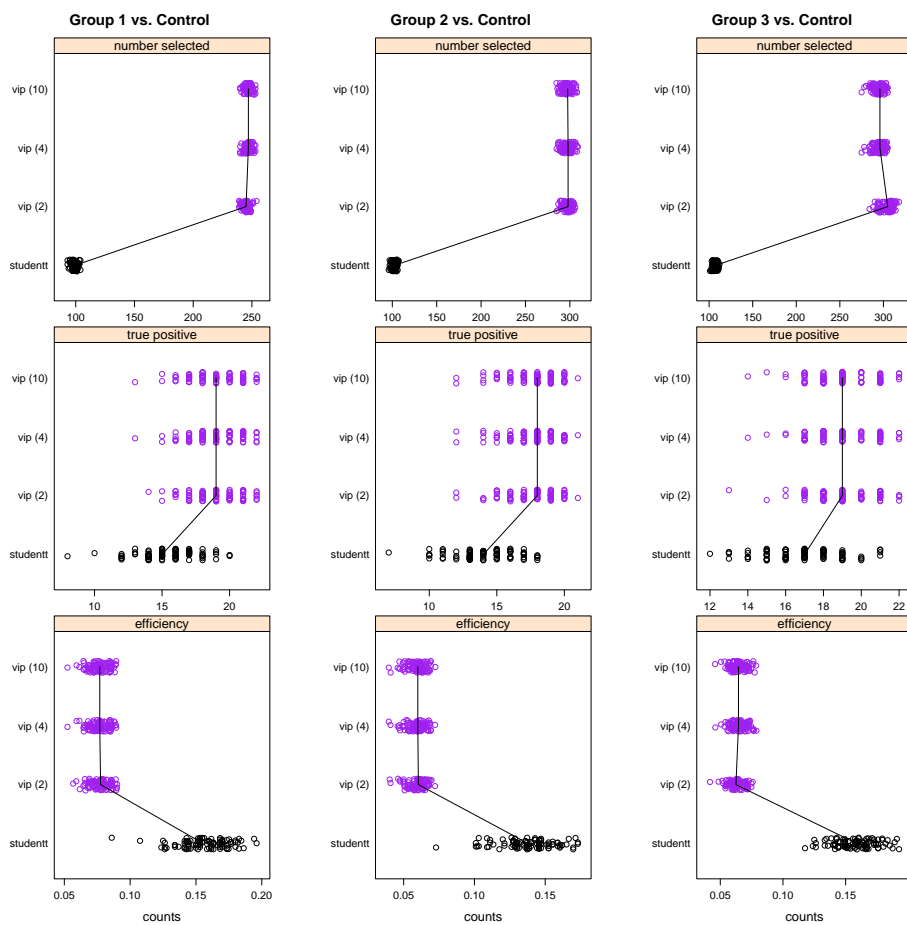


Figure S2: Biomarker selection using  $t$  tests and VIP values for the simulated **negative** ionization mode data. Standard selection thresholds of 0.05 and 1 are used for the  $t$  tests and the VIP, respectively.

### 3 HC-based cutoff values

Using “optimal” thresholds based on the HC criterion, it becomes possible to compare different biomarker selection methods in a fair way. The results are shown in Figures S3 and S4.

Overall, PLSDA, shown in blue, selects the lowest number of variables. Its efficiency, the fraction of “true” biomarkers in the selection, is the highest, and ranges between 25% and 50%. The VIP (purple in the figures) shows a very similar behaviour, as could be expected from the common roots. In general, the VIP selects slightly more variables, also succeeds in getting more true positives but the overall efficiency is slightly lower than that of the PLSDA regression coefficients. The *t* test in general has the highest number of true positives, but also a relatively high number of false positives, leading to efficiency values that are inferior to those of the VIP and PLSDA. Finally, PCLDA (shown in red) clearly performs less well than the other methods. It shows a big dependence on the number of components – choose too few components and the selection is more or less random. In our simulations, one needs ten PC components to obtain results to have any meaningful selection with the HC criterion; fewer components lead to a HC selection that is just equal to the maximum value of 10% of the variables. For both PLSDA and the VIP, virtually no dependence on the number of components can be seen.

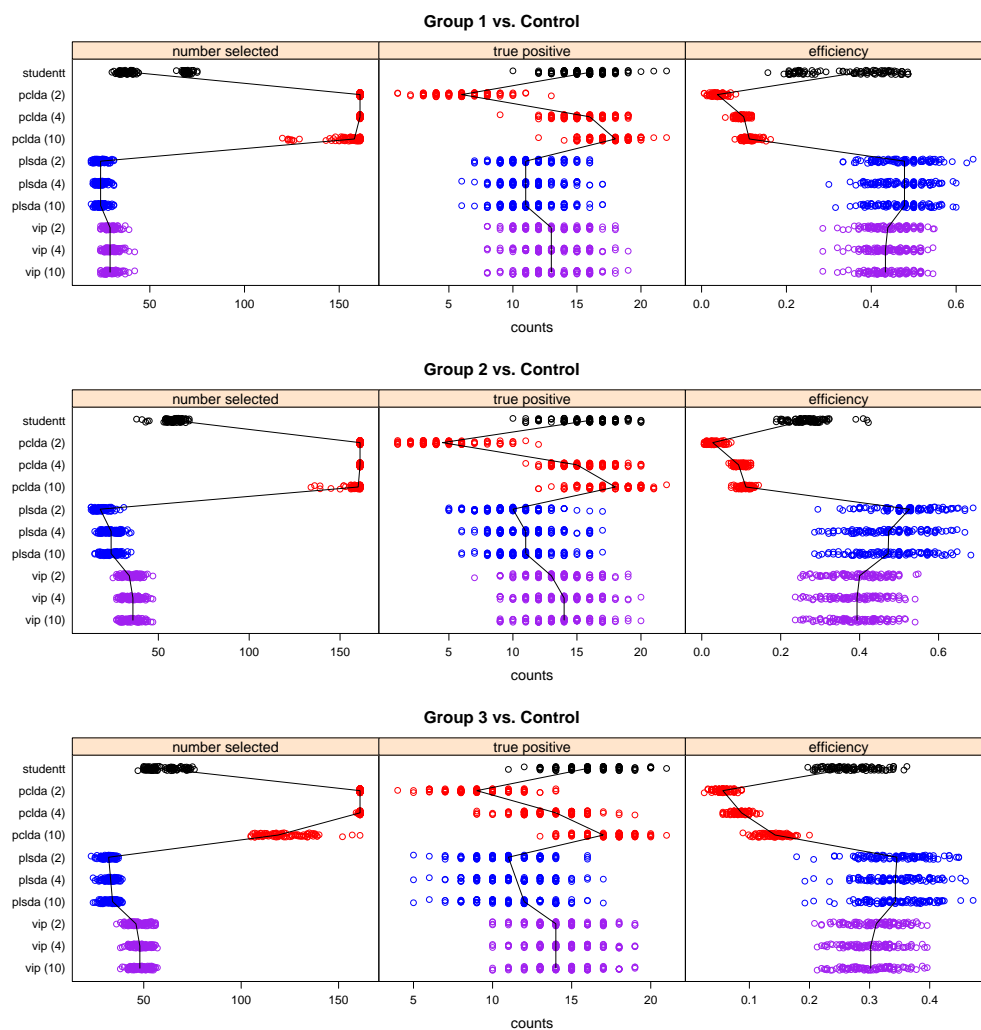


Figure S3: Biomarker selection using HC-based cutoff values for all biomarker selection methods considered, simulated **positive** ionization mode data. The top figure corresponds to Figure 5 in the main paper.

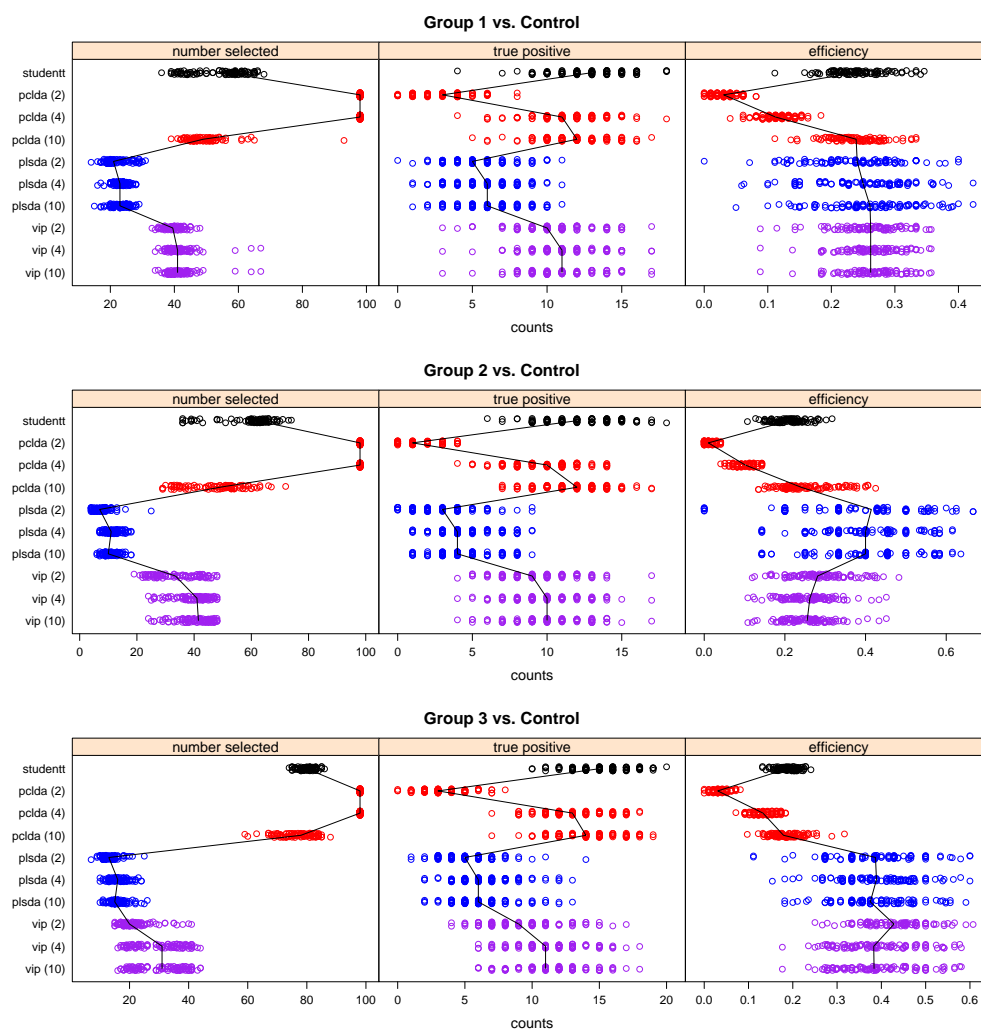


Figure S4: Biomarker selection using HC-based cutoff values for all biomarker selection methods considered, simulated **negative** ionization mode data.

## 4 Direct comparison between standard and HC-based cutoff values

Figures S5 to S10 summarize the information in Figures S1 to S4, concentrating on the comparison between standard thresholds and HC-based thresholds for  $t$  tests and the VIP. In each figure, the first column shows the number of selections. This goes down in all cases. The decrease in the number of true biomarkers that are selected (second column) is also visible, but much less strong, leading to an overall increase in selection efficiencies for all methods considered (third column). As expected, the highest efficiencies are achieved for the comparison of Group 1 with the controls – there, the VIP leads to an efficiency of approximately 50% in positive mode as well as in negative mode, going down to approximately 30% for the more difficult group 3.



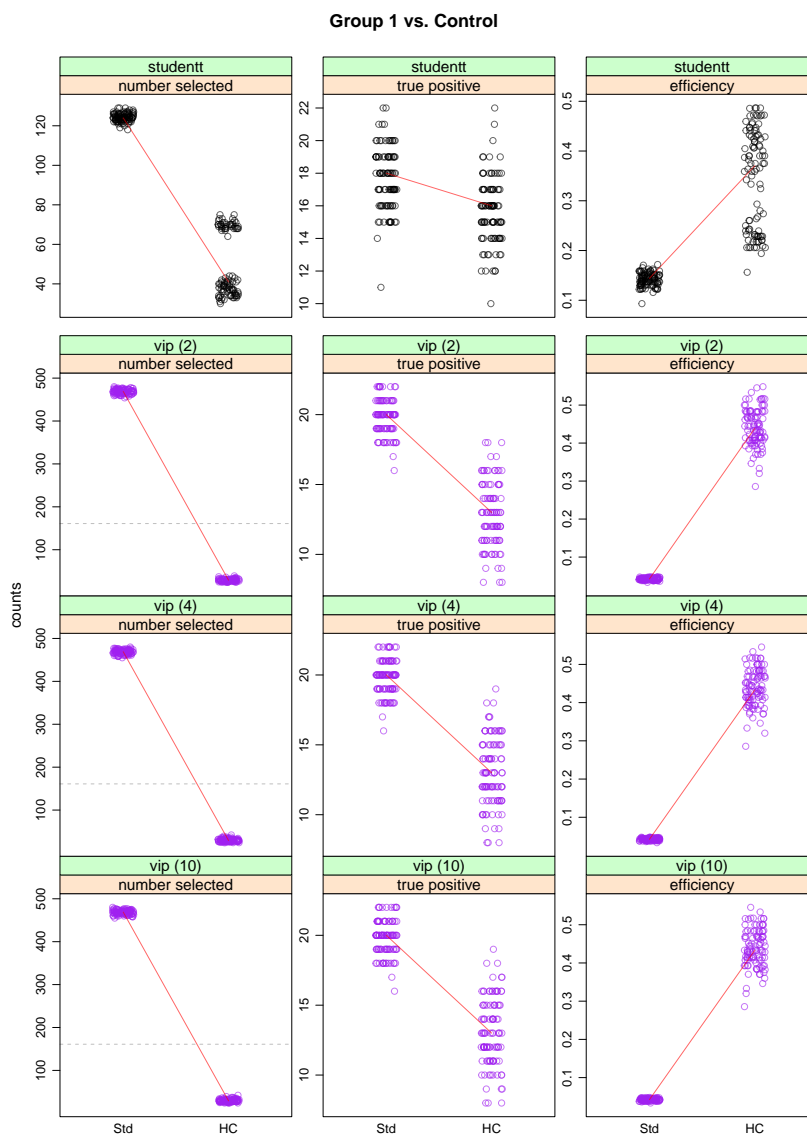


Figure S5: Comparison of standard and HC-based thresholds for  $t$  tests and VIPs. Simulated **positive** ionization mode data for group one, extracted from Figures S1 and S3.

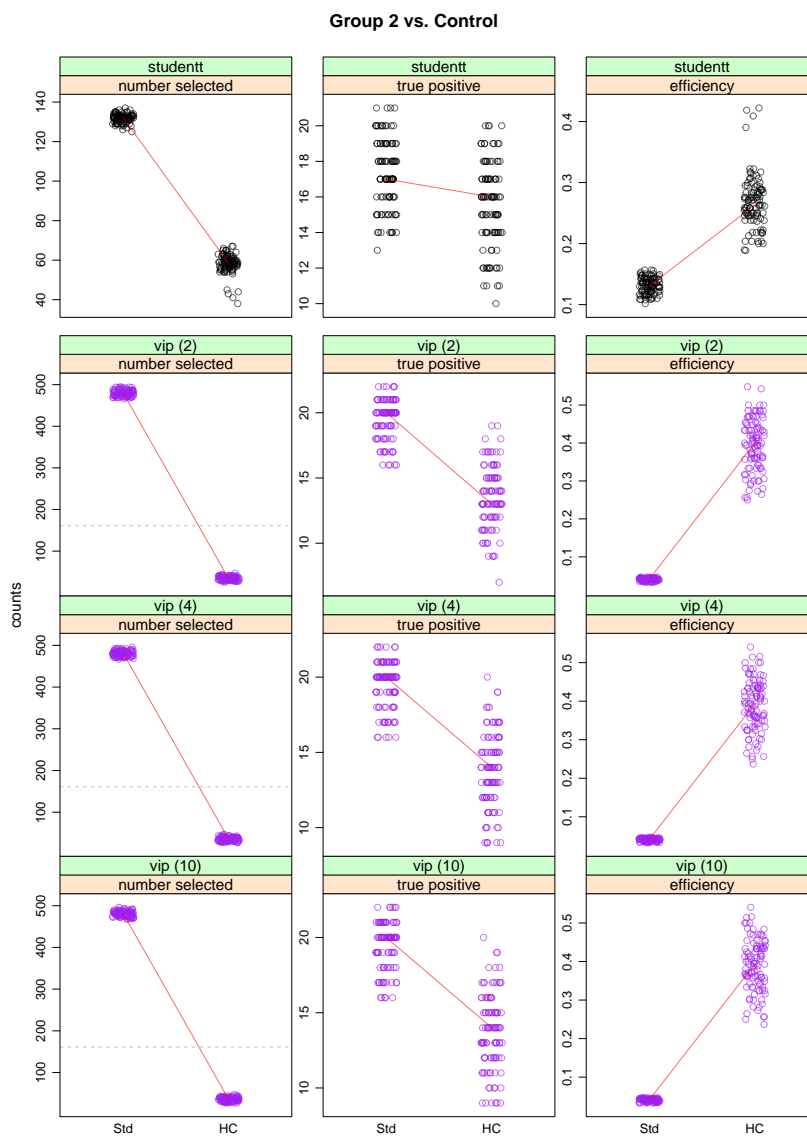


Figure S6: Comparison of standard and HC-based thresholds for  $t$  tests and VIPs. Simulated **positive** ionization mode data for group two, extracted from Figures S1 and S3.

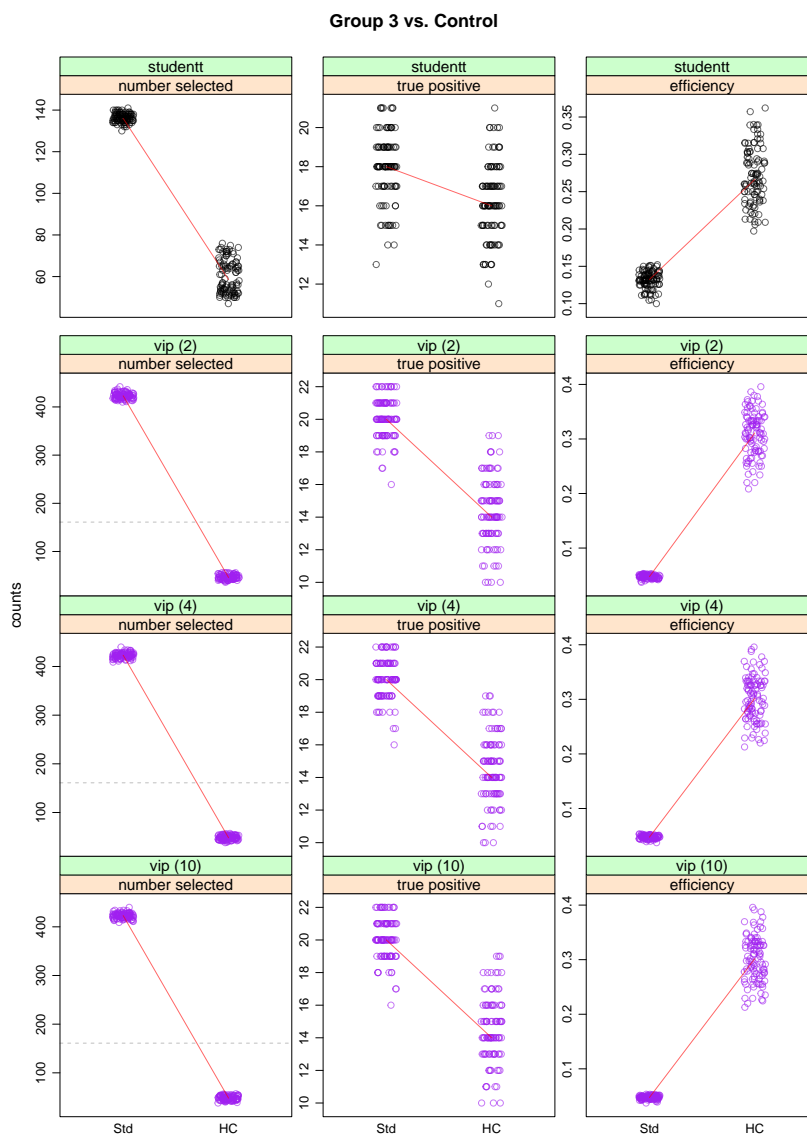


Figure S7: Comparison of standard and HC-based thresholds for  $t$  tests and VIPs. Simulated **positive** ionization mode data for group three, extracted from Figures S1 and S3.

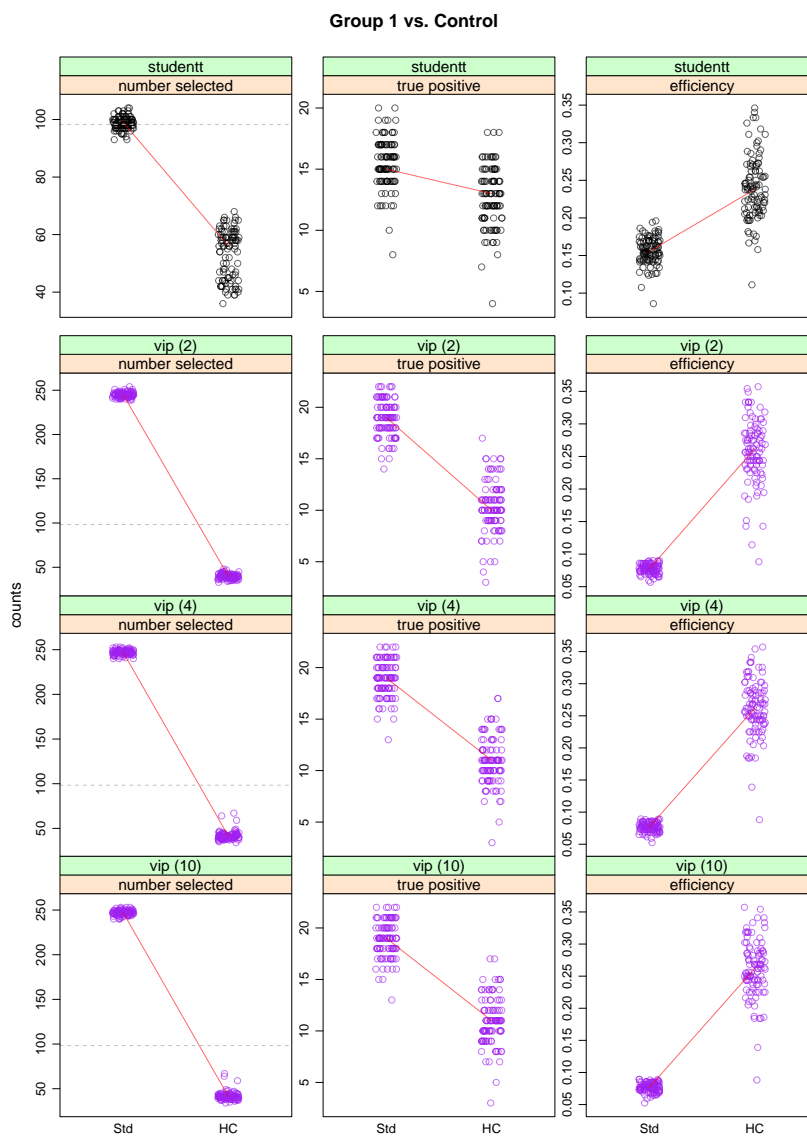


Figure S8: Comparison of standard and HC-based thresholds for *t* tests and VIPs. Simulated **negative** ionization mode data for group one, extracted from Figures S1 and S3.

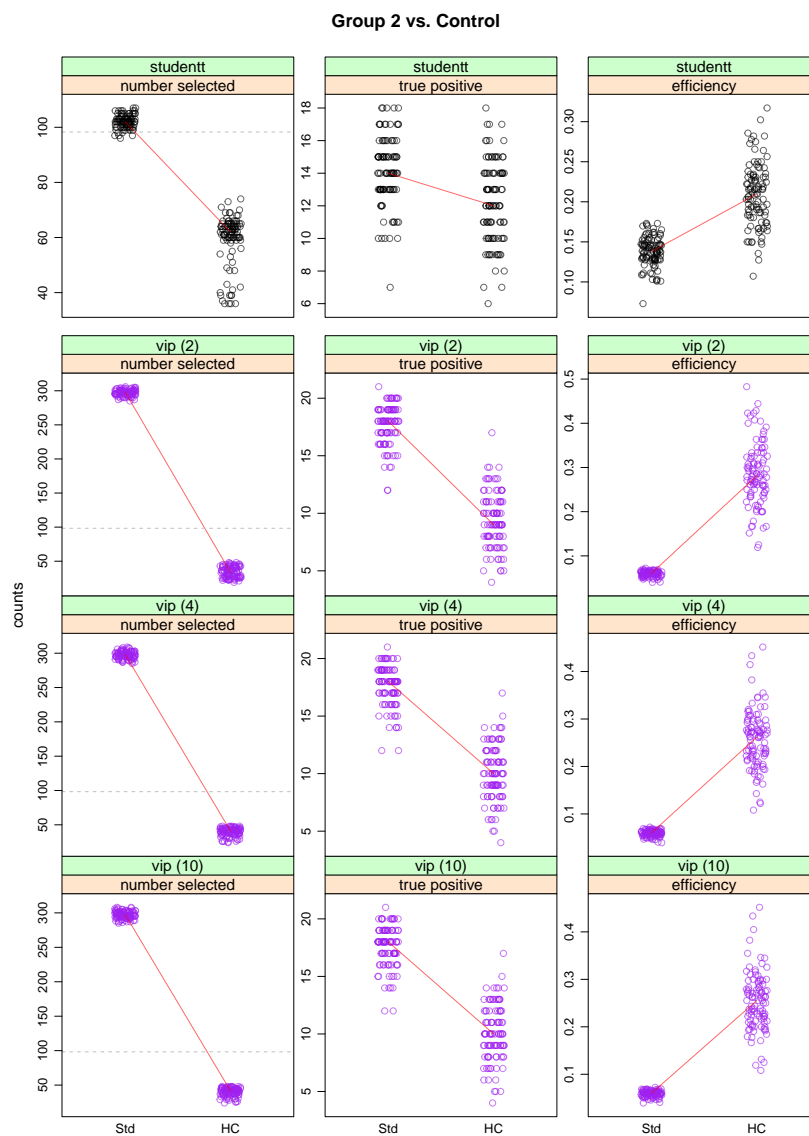


Figure S9: Comparison of standard and HC-based thresholds for  $t$  tests and VIPs. Simulated **negative** ionization mode data for group two, extracted from Figures S1 and S3.

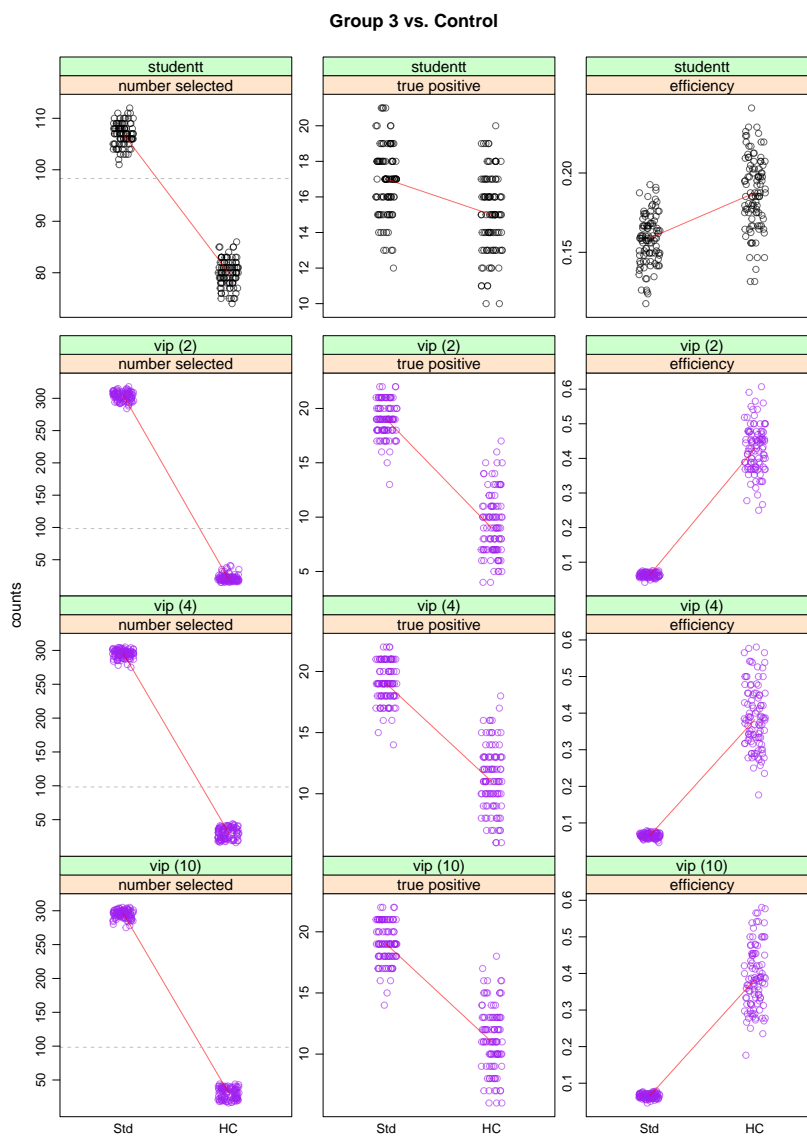


Figure S10: Comparison of standard and HC-based thresholds for  $t$  tests and VIPs. Simulated **negative** ionization mode data for group three, extracted from Figures S1 and S3.