Response to

Review of “Issues in developing multivariate molecular signatures for guiding clinical care decisions”

Michael C Sachs and Lisa M. McShane

This is a very nicely and easy to read paper on an important topic. Although the results are not new, I think the paper provides a very nice review of some important principles. I just have a few relatively minor comments.

1. I found the discussion of pre-validation on page 7 a bit hard to follow. More discussion is necessary to clarify.

*We added the following to page 7 for clarity:*

“This process is repeated to obtain a set of pre-validated signature estimates . Then some measure of the association between these signatures estimates and the withheld true outcomes, , is estimated to serve as a measure of the signature performance. The association measure  could be estimated as a coefficient in a linear, Cox proportional hazards, or logistic regression model, with or without adjustment for other covariates, as appropriate for the situation. The idea behind this pre-validation approach is that if the signature has good performance, then it should produce values that are strongly associated with the true outcomes.”

1. It would be useful to provide some intuition about how the results might change depending on the dimensions of the potential multivariate predictors? For example, the biases reported would be smaller if we only had 10 potential predictors. How about if we had 50 or 100?

*We expanded the simulation results to include 10, 100, 500, and 5000 predictors and also 200 observations. We included these in figures 1 and 2, and describe the results. The biases are smaller with smaller number of predictors, but they are still biased. Typical omics studies involve very large numbers of predictors, in the thousands or tens of thousands.*

1. The authors develop a predictor using a simple logistic regression and a compound predictor. What about for fancier methods such as lasso, support vector machines, and others? I presume that the same types of biases occur, but it would be good to say this in the discussion.

*We added the following sentences to the discussion:*

*“In our simulations and data example, we used a simple variable selection procedure follow by multivariable regression. More complex statistical methods for developing signatures such as the lasso, random forests, and others are not immune to this type of bias. In fact, more complex methods often involve tuning parameters, which give more opportunities for bias due to overfitting. “*

1. When doing cross-validation or other valid approaches, it would be useful to guide the reader about what predictor to actually report. For example, the cross-validation estimate provides an unbiased estimate of the predictive accuracy measure for the procedure and not for a particular predictive model. Should one report some “average predictor” or the predictor using the entire data?

*We added some discussion of this at the end of page 7:*

*“These resampling-based procedures described above provide unbiased*

*estimates of predictive accuracy for the procedure, that is* EP[φF(S)]  *as opposed to the accuracy for a particular signature that others can use. Often, we will provide*

*the signature developed on the entire dataset as the one for future use,*

*as using the full data will estimate signature coefficients with the*

*most precision. Still, it is important to avoid the temptation to report*

*the estimate of performance on this full data signature, as it can be*

*severely biased.”*

1. There are a few typos. For example, lines 122 and 133 on page 5.

*Thanks, we fixed these typos.*