Criteria for Building and Selecting Optimal Risk Models

Frank E Harrell Jr
Department of Biostatistics
Vanderbilt University School of Medicine
Nashville TN
f.harrell@vanderbilt.edu
biostat.mc.vanderbilt.edu

DEFINING RISK IN ACS: STATISTICAL ISSUES IN RISK MODELS WASHINGTON, DC
2 June 2005

Outline

- General Issues in Building Models
- Goodness of Fit
- Measures of Predictive Accuracy
- Judging Usefulness of Model for Individual Patients
- Validation Methods

General Issues in Building Models

- Predictions (prob., time), not classifications
  - Avoid dichotomizing predictions or collapsing risk scores
  - Leave classification up to the possessor of the utility function (usually not the analyst)
  - Optimum decisions involve continuous outputs (e.g. $\hat{P}$) and patient-specific utilities
- Adequate sample size and patient diversity
- Predictors specified by clinical knowledge, not P-values

Respect Continuous Variables

- Huge loss of power and precision if dichotomize a continuous predictor or response variable
- Categorization assumes a discontinuous relationship
- Results in estimates applicable only to groups, not individuals
- Gives other variables and interactions artificially high weights

Avoid Overfitting

- Fitting model more complex than information content supports
- **Many** published models are overfitted; be skeptical
- Other authors failing to validate a published model falsely assume non-transportability
- An unbiased validation would have revealed poor fit in the original analysis
- Use shrinkage (discounting, penalization) or data reduction

- Don't underfit by assuming linearity
- Nonlinearity is the most common cause of lack of fit
- Additivity holds more often than not
- Instead of spending effort assessing goodness of fit, fit flexible models from the start

Problems with Extreme Flexibility

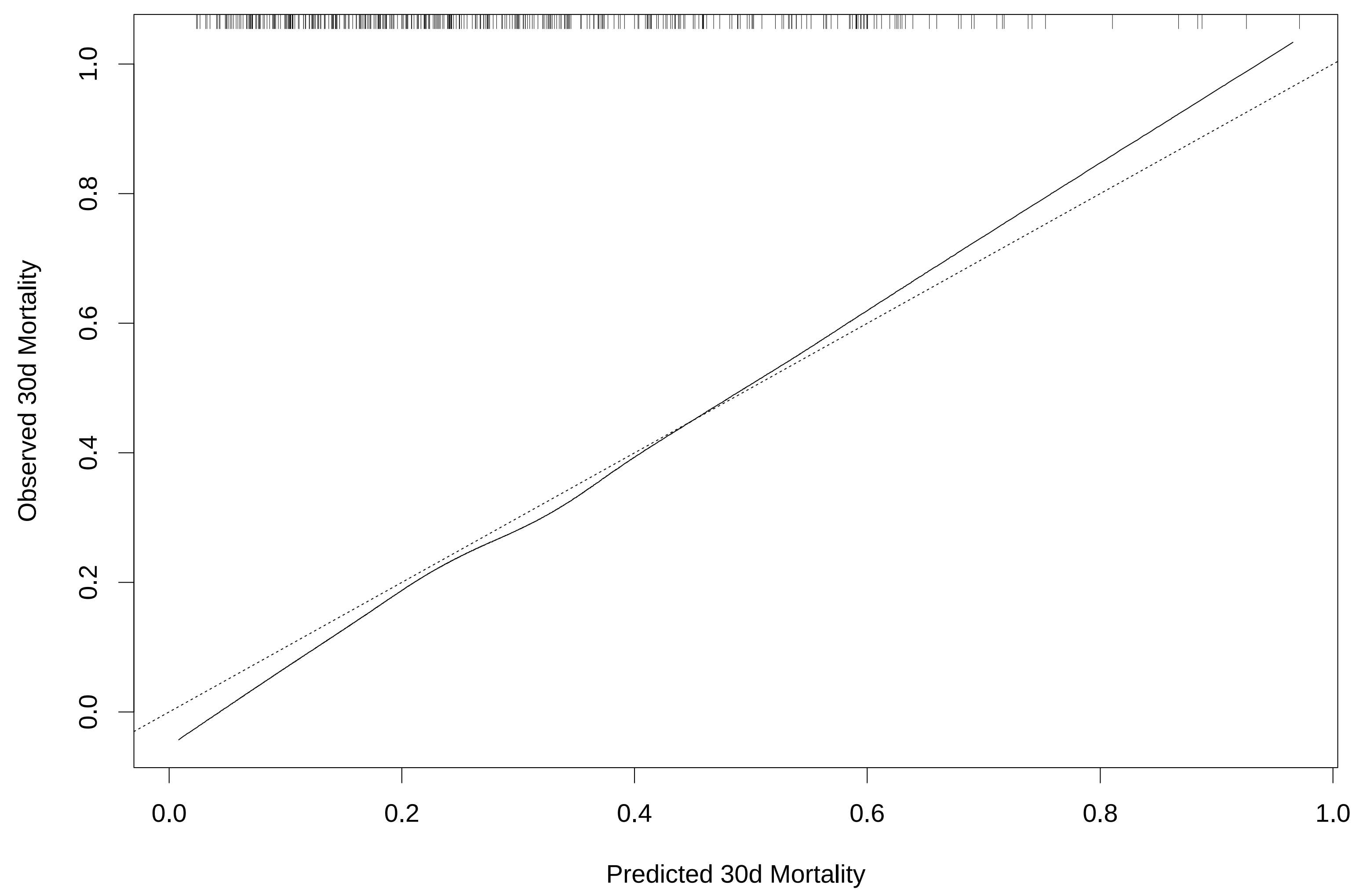
- Automatic interaction detection
- Recursive partitioning (& CART)
- Price of strictly empirical procedure not driven by science can be conservatism
  - Often must prune trees back until R^2 is low
- If additivity assumption is 0.6 correct, a flexible additive model can outperform recursive partitioning for commonly used N

Use Better Modeling Strategies

- Be unafraid of complex models; graph for non-statisticians
- Use a strategy you can program; can study properties by simulation
- If there is model uncertainty it is better to average models than to select a single “winner”
- Data reduction and shrinkage have advantages

- Pure discrimination: C-index; concordance between predicted and observed outcomes
- Calibration: absolute accuracy
  - Use a high-resolution assessment (easy for binary Y using lowess)

Validation of Published Model on New Sample



Judging Usefulness of Model for Individual Patients

- Folklore: $R^2 = 0.02$ unacceptable, $C > 0.8$ acceptable
- Better: once model is shown to be well calibrated, examine distribution of predicted values

Validation Methods

- Unless $N > 30,000$, holding back test data from model development is problematic
- Mean squared error of accuracy estimate is higher with data splitting or cross-validation than with bootstrap
- Resampling techniques preferred
  - Most validations claimed to be external are internal anyway
- Must consider **all** aspects of model uncertainty
- Need to start adding precision estimates to validated accuracy indexes