

# Least Angle Regression

Tim Hesterberg, Chris Fraley

Research Department  
Insightful Corp., Seattle

Supported by NIH SBIR Phase II 2R44GM074313-02

timh@insightful.com, fraley@insightful.com

## Summary

Least Angle Regression is a promising new technique for variable selection applications, offering an attractive alternative to stepwise regression in terms of stability and prediction accuracy. It provides a unified framework that explains the similar behavior of Lasso (L1-penalized regression) and forward stagewise regression, leading to a fast implementation of both. We'll demonstrate a prototype open-source S-PLUS/R package "glars" for generalized least angle regression, extending the "lars" package of Efron and Hastie and "glmplath" of Park and Hastie. We invite outside collaboration.

[www.insightful.com/Hesterberg/glars](http://www.insightful.com/Hesterberg/glars)

## LARS and related methods

### Why is LARS important?

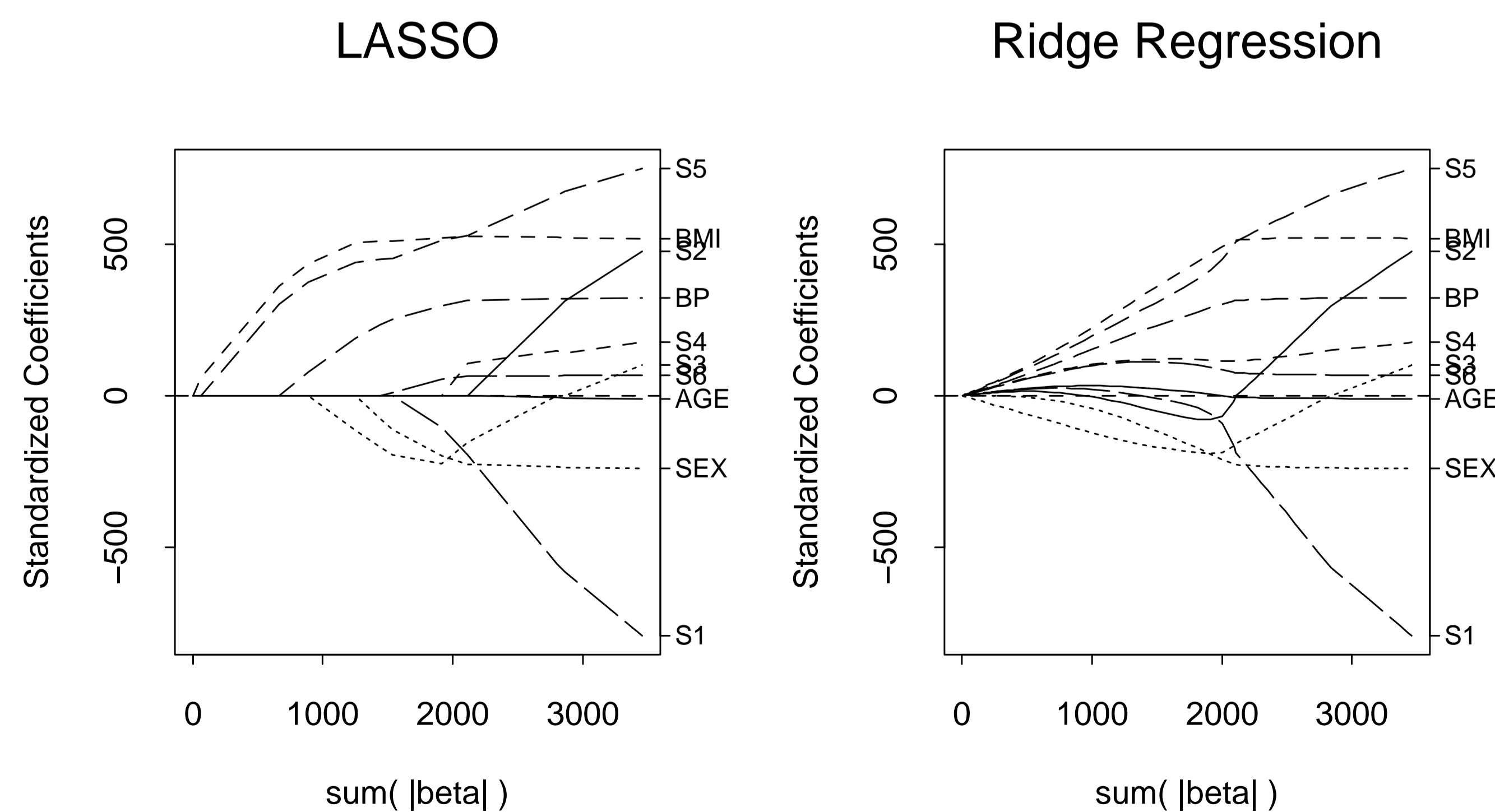
Variable selection in regression is important, especially with developments like microarrays, with more variables/predictors than observations. Approaches include stepwise, all-subsets, ridge, stagewise, boosting, LASSO, regularization, .... There are interesting connections between the methods. Least Angle Regression (Efron et al 2004 Annals) unifies Lasso and Forward Stagewise, provides a fast implementation, and a way to choose the number of terms to include.

### Ridge Regression and LASSO

**Ridge regression:** minimize  $\sum (Y_i - \hat{Y}_i)^2 + \lambda \sum \hat{\beta}_j^2$

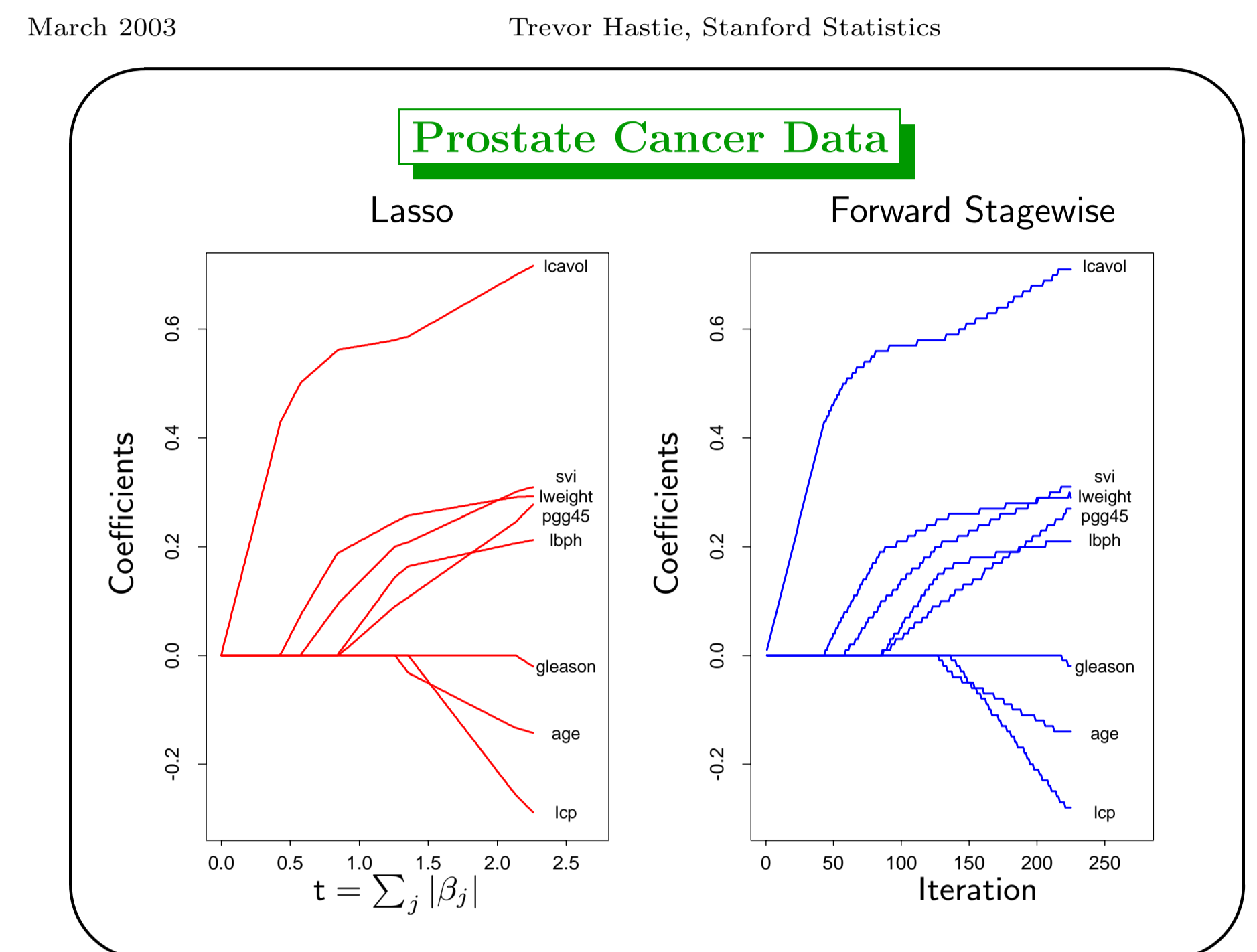
**LASSO:** minimize  $\sum (Y_i - \hat{Y}_i)^2 + \lambda \sum |\hat{\beta}_j|$

The LASSO penalty forces small coefficients to 0, giving simpler models. It penalizes large coefficients less; this gives better predictions than ridge regression, because important terms are penalized less. However, the implementation is more complicated and slower.



## Stepwise, Stagewise, and Least Angle

Lasso and infinitesimal forward stagewise are identical for orthogonal predictors, otherwise similar.

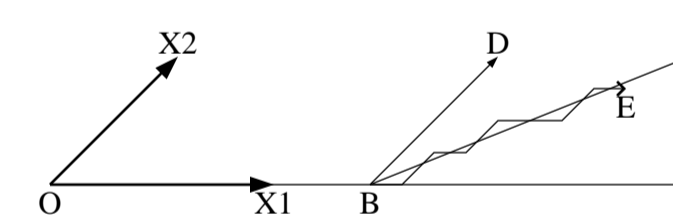


LARS provides the geometrical explanation for the similarity. It is also interesting to contrast LARS and stepwise regression; LARS is more conservative, taking smaller steps.

**Stepwise regression:** Pick predictor most correlated with  $y$ , full LS solution ( $A$ , then  $C$ ).

**Forward stagewise:** Pick predictor most correlated with  $y$ , increment coefficient for predictor (*tiny steps*,  $O$  to  $B$ , to  $E$ , to  $C$ )

**Least Angle Regression:** Pick predictor most correlated with  $y$ , bring predictor into model only to extent it is better than others, move in least-squares direction until another variable is as correlated (*jump directly*,  $B$ , then  $C$ )



$C$  = projection of  $y$  onto space spanned by  $X_1$  and  $X_2$ ; least-squares solution.

$B$  = first step for least-angle regression

$E$  = point on stagewise path

## Stopping Rules

Degrees of Freedom - use for  $F$ -test

- variable selection like stepwise regression increases effective DF. This tends to be over-optimistic, and too many variables are included if  $F$  tests are used for stopping.
- penalty methods like ridge regression: decrease effective DF
- LARS: these effects cancel, DF = number of nonzero terms. Hence can use  $F$  test for stopping.

Cross-validation is an alternative, but is slower.

## Software

### Existing Software in R and/or S-PLUS

**lasso** : Tibshirani (1995)

linear models with  $L_1$  constraint

**lasso2** : Lokhorst, Turlach, Venables (1999)

linear and generalized linear models with  $L_1$  constraint

**brdgrun** : Fu (2000)

shrinkage estimator for LMs with bridge penalty  $\lambda \sum |\beta_j|^\gamma$

**lars** : Efron and Hastie (2004)

linear regression: least angle, lasso, forward stagewise

**elasticnet** : Zou and Hastie (2005)

LM with penalty  $\lambda_1 |\beta|_1 + \lambda_2 |\beta|_2$

**glasso** : Kim, Kim, Kim (2005), generalized lasso: multiple constraints; multiple linear predictors

**glmplath** : Park and Hastie (2006)

GLM and Cox proportional hazards

### S+GLARS

S-PLUS and R, open source, incorporates `lars`, `glmplath`. We plan to provide a framework to incorporate work by others, with a consistent user-friendly interface.

### Extensions:

- Numerically-accurate calculations
- Factors, splines, polynomials, interactions, ...
- Other models (robust regression, ...)
- Other penalties
- Missing data
- Massive data sets
- Diagnostics, tools for selecting tuning parameter

LARS is a field of active research. We are working with outside researchers on these and other extensions. If you are interested, please contact us.

A prototype is available, see:

[www.insightful.com/Hesterberg/glars](http://www.insightful.com/Hesterberg/glars)



## Insightful Research Department

We turn research into software for wide use. We have high standards for ease of use, robustness, testing, and documentation. Our projects are varied, and they usually involve collaboration with academics (I've worked in resampling, missing data, group sequential designs, simulation-based econometric software, functional data, stable distributions, proteomics, microarrays, frailty models, causal modeling, ...). *We're hiring, and we're always looking for good projects and collaborators.* Contact Tim Hesterberg [timh@insightful.com](mailto:timh@insightful.com) or Director of Research Jill Goldschneider [jrgold@insightful.com](mailto:jrgold@insightful.com)