*Biostatistics* (2013), **0**, 0, *pp.* 1–2
doi:10.1093/biostatistics/biosts-13132-File003

# Supplementary material to: High-Dimensional, Massive Sample-Size Cox Proportional Hazards Regression for Survival Analysis*

SUSHIL MITTAL[1], DAVID MADIGAN[1], RANDALL S. BURD[2], MARC A. SUCHARD[3,4,5]

[1]*Department of Statistics, Columbia University, New York, NY, USA*
[2]*Division of Trauma and Burn Surgery, Joseph E. Robert Jr. Center for Surgical Care, Children's National Medical Center, Washington, DC, USA*
[3]*Department of Biomathematics and* [4]*Department of Human Genetics, David Geffen School of Medicine at UCLA, and*
[5]*Department of Biostatistics, UCLA Fielding School of Public Health, University of California, Los Angeles, CA, USA*

mittal@stat.columbia.edu

## 1. Additional Experiment on Pediatric Trauma Data

To provide further insight into the predictive performance of Cox models, we group the subjects into low, medium and high risk groups by sorting their relative risk scores in increasing order and using threshold values at the 33rd and 66th percentiles (Raftery *and others*, 1996). For each group, we count the number of events (number of non-censored observations). Table 1 reports the results for the Gaussian and Laplacian priors. The data show that in both cases, the subjects assigned to the high risk group by the high-dimensional models have more events than the ones assigned high-risk by the low-dimensional models. Although both kinds of models assign similar number of subjects to the low risk group, the mean observed survival time (MST) of the subjects having events is larger for the high-dimensional models than for the corresponding

*Research reported in this publication was supported by the National Institute for General Medical Sciences of the National Institutes of Health under award number R01GM087600 and by the National Science Foundation under award number IIS-1251151

low-dimensional ones. These findings also establish that the high-dimensional models are better at indicating which subjects are at higher risk.

Table 1. *Comparison of number of events in low-, medium- and high-risk groups for various low and high-dimensional models using Gaussian and Laplacian priors for pediatric trauma data. MST stands for mean observed survival time in days.*

| | | | Low-dim | | High-dim | |
|---|---|---|---|---|---|---|
| Prior | Risk Group | # Subjects | # Events | MST | # Events | MST |
| | Low | 18,860 | 7 | 2.00 | 8 | 2.75 |
| *Gaussian* | Medium | 18,860 | 28 | 3.68 | 14 | 3.00 |
| | High | 19,433 | 790 | 3.68 | 803 | 3.68 |
| | Low | 18,860 | 7 | 2.00 | 5 | 3.20 |
| *Laplacian* | Medium | 18,860 | 29 | 3.63 | 20 | 3.30 |
| | High | 19,433 | 789 | 3.68 | 800 | 3.68 |

## References

Raftery, A.E., Madigan, D. and Volinsky, C.T. (1996). Accounting for model uncertainty in survival analysis improves predictice performance. In: *Bayesian Statistics V*. Oxford University Press. pp. 323–350.