**Supplementary Figure 1: Stylized example of how a random forest consisting of three trees is used to make a prediction.**



**Figure legend: In this example, there are four covariates: duration of surgery (in minutes), blood loss (in mL), surgical specialty (one of Urology, Cardiac or Neurosurgery) and length of stay (in days). The red paths correspond to the prediction made by each tree for a hypothetical patient in the neurosurgery service line, whose surgery lasted 200 minutes, with an estimated blood loss of 350 mL and a hospital stay of 6 days. Trees 1 and 2 predict no readmission, while Tree 3 predicts readmission. The overall probability of readmission prediction of the random forest is therefore 1/3 = 0.333. Supplementary Figure 2: Calibration curves for a single random split of the UCLA Ronald Reagan Medical Center data.**



**Figure legend: L1 regularized logistic regression, random forest and gradient boosted tree model were trained using *general* and *labs* features, computed using the test set for a single random split of the data. Each curve is obtained by binning the test set observations by deciles of predicted readmission probability. Each point of each curve corresponds to a decile/bin; the *x* coordinate is the average predicted readmission probability within that decile, while the *y* coordinate is the actual/observed readmission probability within the decile (calculated as number of patients readmitted / total number of patients). The diagonal dotted line indicates perfect calibration. For comparison, the calibration curves for HOSPITAL and LACE are also shown.**

**Supplementary Figure 3: Random forest variable importance plot for a single random split of the UCLA Ronald Reagan Medical Center data.**

****

**Figure legend: Random forest model shown was trained with the *general* and *labs* features. Variable importance was computed using the permutation importance metric, which measures the increase in classification error when the chosen variable is randomly permuted.**

**Supplementary Figure 4: Gradient boosted tree model variable importance plot for a single random split of the UCLA Ronald Reagan Medical Center data.**

****

**Figure legend: Gradient boosted tree model shown was trained using *general* and *labs* features. Variable importance was computed using the “Gain” metric in *xgboost*.**

**Supplementary Figure 5: Plot of standardized coefficients for L1 regularized logistic regression for a single random split of the UCLA Ronald Reagan Medical Center data.**

****

**Figure legend: L1 regularized logistic regression model shown was trained using *general* and *labs* features.**

**Supplementary Figure 6: UCLA Ronald Reagan Medical Center test set area under the curve values for pre-discharge prediction.**

****

**Figure legend: AUC: area under the curve.** **The x-axis indicates how many days after surgery predictions are made.** **Values are calculated by computing the test set area under the curve for each of the ten random splits and averaging; error bars correspond to one standard deviation.**

**Supplementary Figure 7: Inclusion/exclusion criteria for overall data set and emergency department readmissions at the UCLA Santa Monica Hospital (SMH).**

****

**Supplementary Figure 8: UCLA Santa Monica Hospital test set area under the curve values for pre-discharge prediction.**

****

**Figure legend: AUC: area under the curve. The x-axis corresponds to the cutoff parameter *T,* which indicates how many days after surgery predictions are made.** **Values are calculated by computing the test set area under the curve for each of the ten random splits and averaging; error bars correspond to one standard deviation.**