Figure S2 – Reliability and calibration

Boosted trees usually do not produce well-calibrated posterior probabilities. They are biased towards the center and avoid the extremes. The reliability plot of the raw model output for pre- & post-op features (black curve) shows a systematic overestimation in 5x5 cross-validation. Fitting an isotonic regressor that maps the output of the classifier to a calibrated probability in [0, 1] (grey, interrupted curve) shows a clear improvement of predicted probabilities after calibration in this manner.