

Supplemental material 3

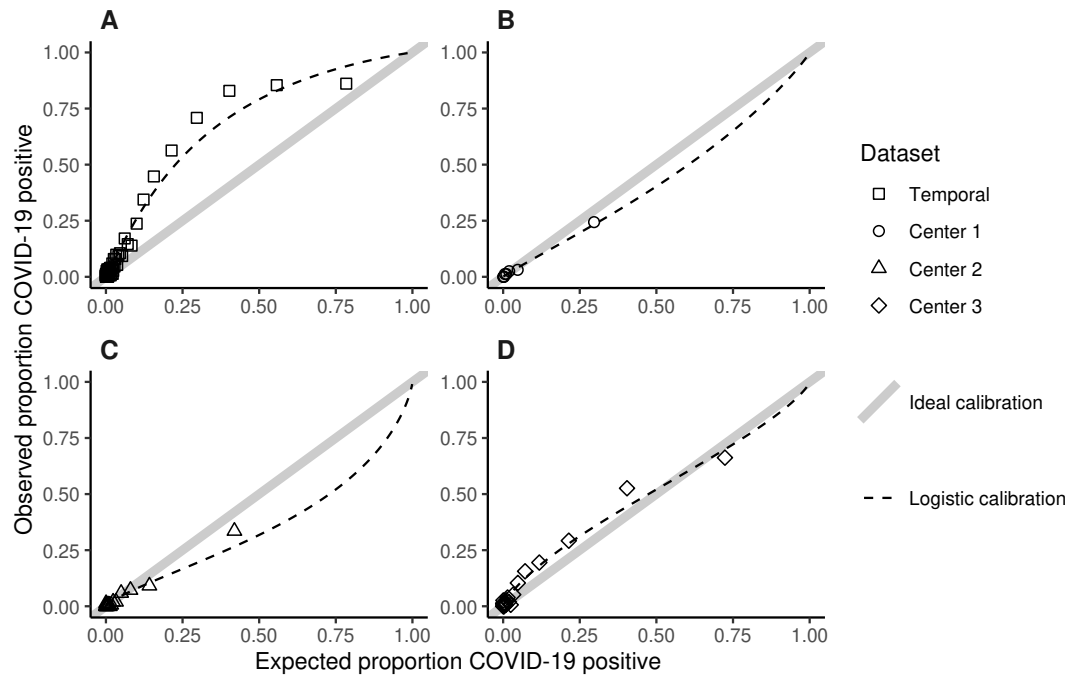


Figure 1: CoLab-score calibration plots of the temporal validation (A), external validation center 1 (B), external validation center 2 (C) and external validation center 3 (D).

In the calibration plots, the proportion of observed COVID-19 positives versus expected probabilities are plotted. Observations are grouped with an average of 150 observations per group. The expected probabilities follow from applying the inverse logit function to the CoLab-linear predictor calculated from Table 2. If the observed proportion in an external dataset is lower than the expected proportion, this means risks are over-estimated, if the observed fraction is higher, risks are under-estimated. Ideally, observed proportions are equal to expected proportions, this ideal-calibration-line is shown as a straight line through the origin with a slope of 1. The logistic calibration line is a logistic regression fit of the predicted probabilities. [Intercept, slope] for plots A-D: A [1.34, 1.08], B [-0.39, 0.92], C [-0.76, 0.77], D [0.08, 0.79]. Although no validation datasets show perfect calibration, this is the result of differences in COVID-19 prevalence in the temporal validation dataset (7.4% versus 2.2%) and differences in calibration of laboratory equipment in the three external centers.

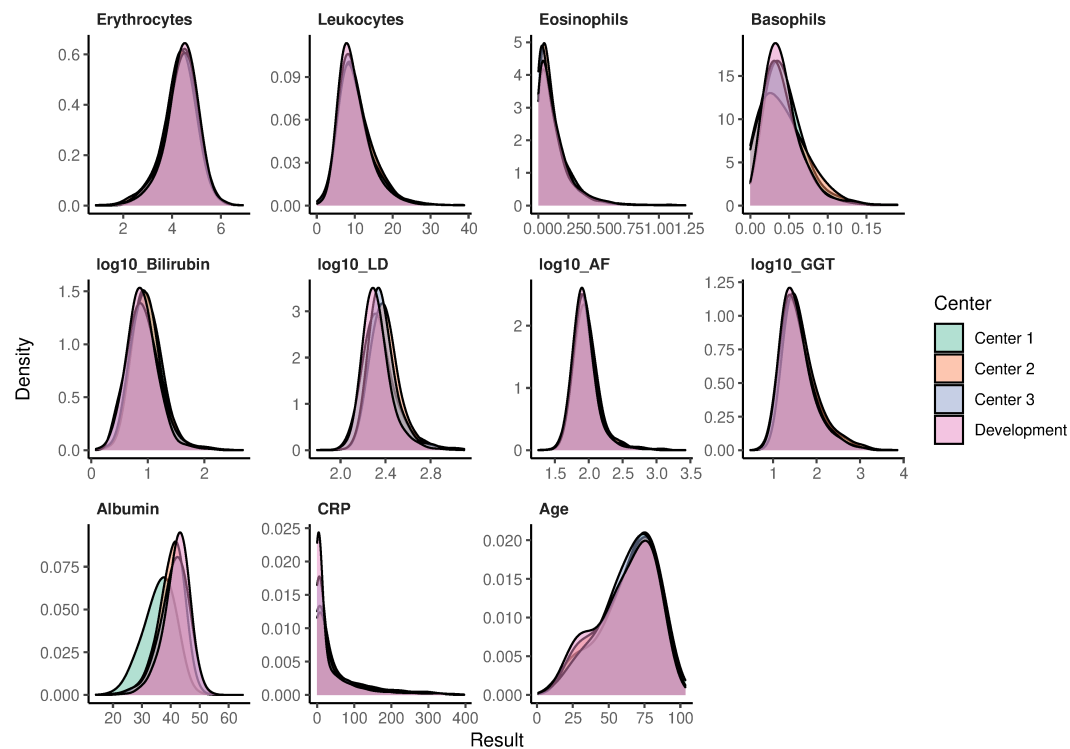


Figure 2: Probability density plots of laboratory parameters.

Probability density plots are shown for all control patients of the development dataset and the three external centers. Ideally all distributions should overlap since this implies that control patient populations are most likely similar in the development dataset to the external datasets.

When comparing the distribution of the CoLab variables for all control-patients across different external validation datasets, albumin and LD show the largest deviations.