

Supplemental Table 1: Results of linear mixed effects model predicting AI Characteristics Transparency Reporting (ACTR) score

|                | Predictor                                | Estimate (95% C.I.) |
|----------------|------------------------------------------|---------------------|
| Fixed Effects  | Intercept                                | 2.44 (2.10, 2.78)   |
|                | Device approved post-2021 FDA guidelines | 0.88 (0.54, 1.23)   |
|                | Deep learning                            | 1.21 (0.84, 1.58)   |
|                | Predicate device uses AI                 | 0.36 (0.00, 0.71)   |
| Random Effects | Company (Intercept)                      | 1.72 (1.46, 1.99)   |
|                | Residual                                 | 2.33 (2.20, 2.45)   |

Supplemental Table 2: Association between input characteristics and AI Characteristics Transparency Reporting (ACTR) variability before/after 2021 Food and Drug Administration (FDA) reporting guidelines

| ACTR Score Input Characteristic                            | Chi Square Statistic | Degrees of Freedom | P-value |
|------------------------------------------------------------|----------------------|--------------------|---------|
| Dataset demographics                                       | 102.43               | 1                  | <0.001  |
| Test dataset size (patients or images)                     | 62.68                | 1                  | <0.001  |
| Train dataset size (patients or images)                    | 54.79                | 1                  | <0.001  |
| Train dataset source                                       | 40.44                | 2                  | <0.001  |
| Test dataset source                                        | 22.49                | 2                  | <0.001  |
| Other evaluation metric                                    | 8.68                 | 1                  | 0.003   |
| Accuracy                                                   | 8.41                 | 1                  | 0.004   |
| Clinical study sample size                                 | 7.53                 | 1                  | 0.006   |
| Sensitivity                                                | 6.85                 | 1                  | 0.009   |
| Specificity                                                | 6.23                 | 1                  | 0.013   |
| AUROC                                                      | 3.65                 | 1                  | 0.056   |
| PPV                                                        | 2.90                 | 1                  | 0.089   |
| Predetermined change control plan (PCCP)                   | 2.31                 | 1                  | 0.129   |
| Clinical study collection type (retrospective/prospective) | 1.86                 | 1                  | 0.172   |
| Model architecture                                         | 1.16                 | 1                  | 0.281   |
| NPV                                                        | 0.98                 | 1                  | 0.322   |
| Clinical study testing                                     | 0.62                 | 1                  | 0.429   |