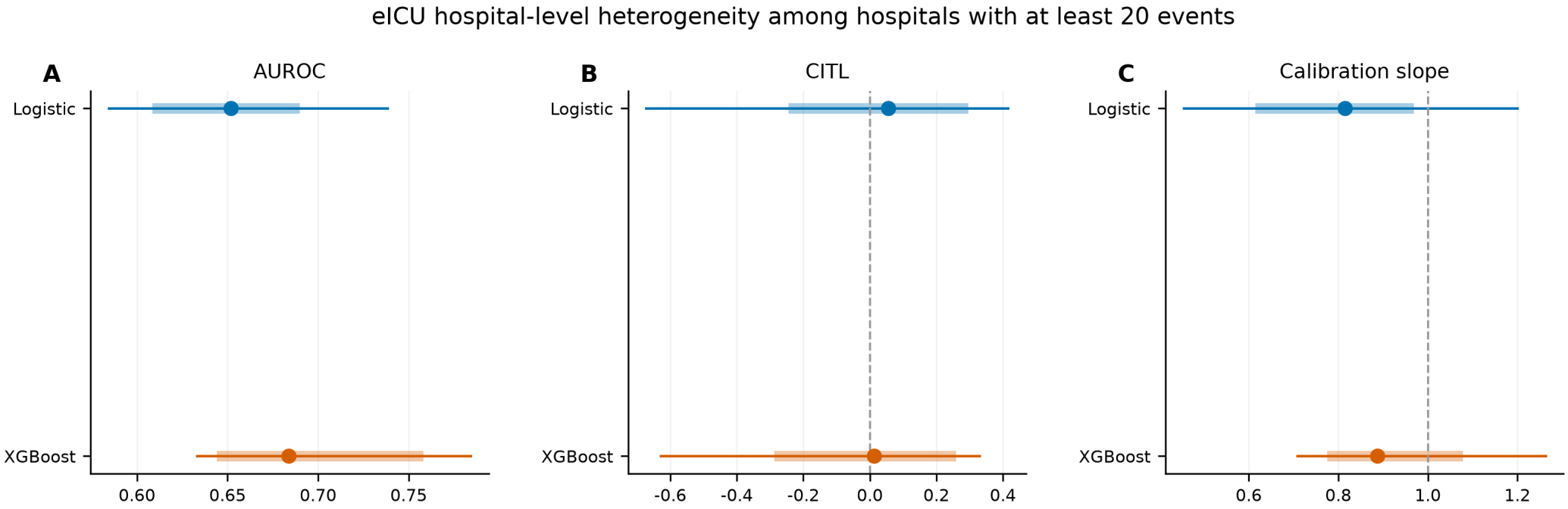
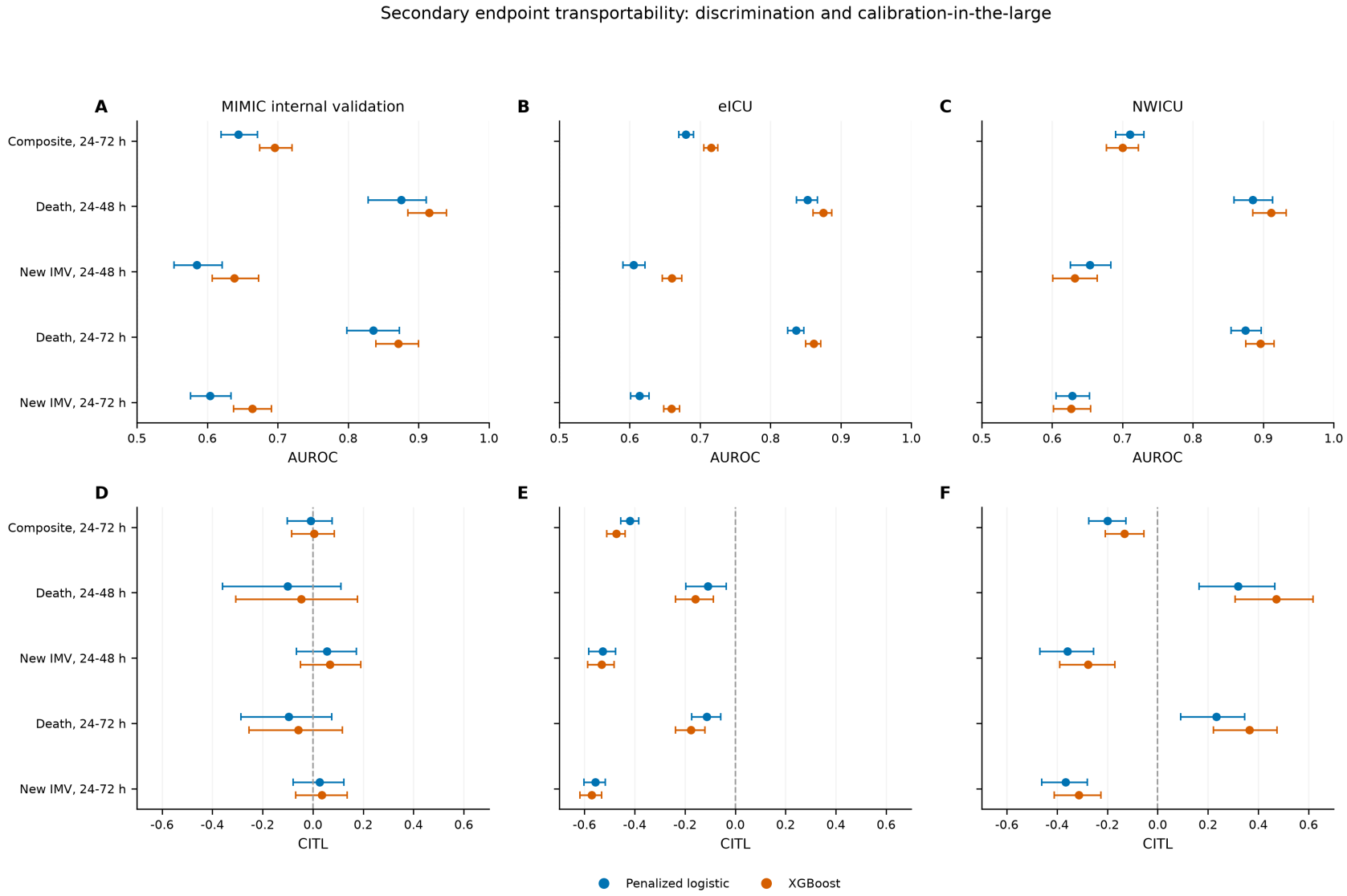
Figure S1. eICU hospital-level deployment heterogeneity



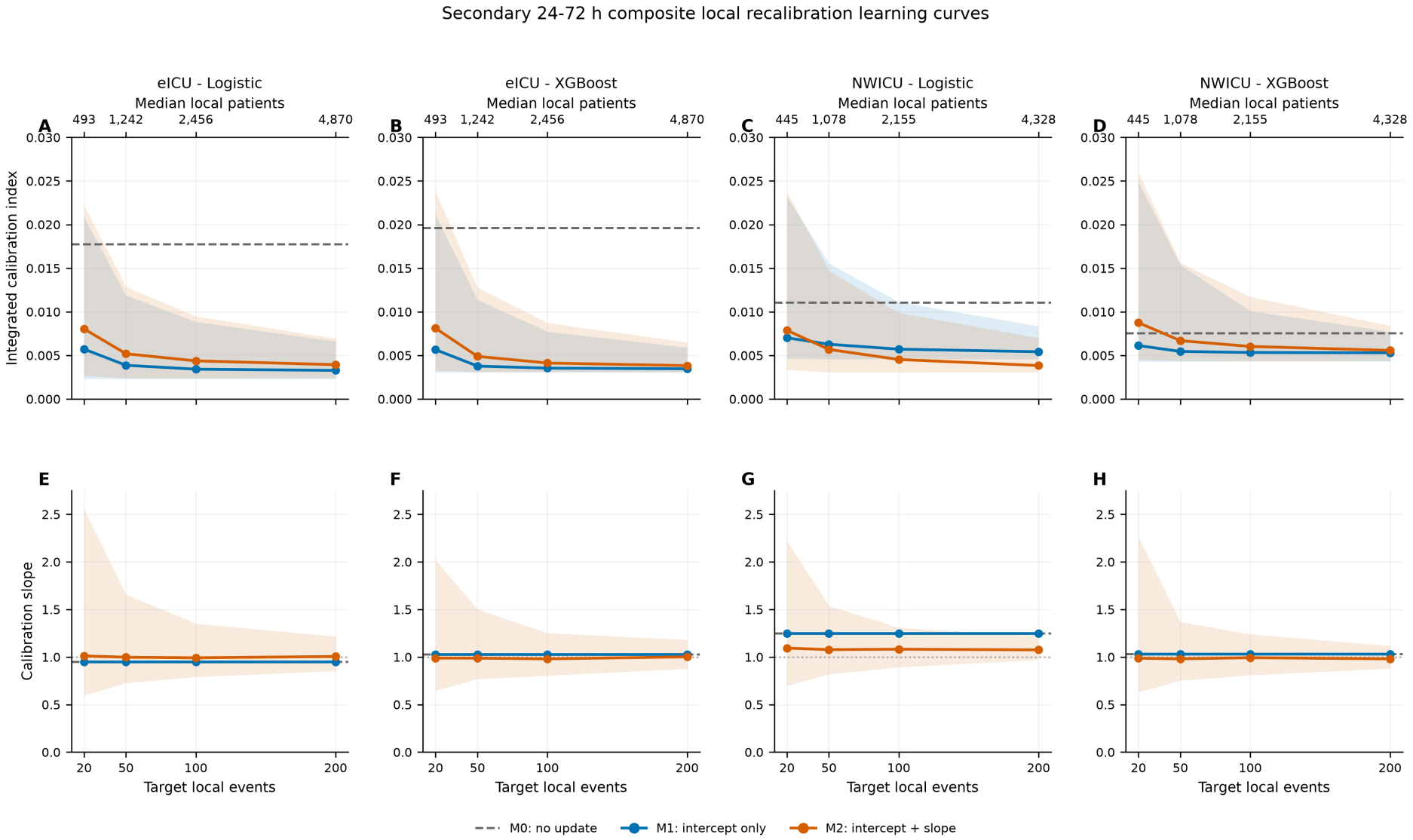
eICU hospital-level deployment heterogeneity among hospitals with at least 20 primary outcome events. Points show medians, thick segments show interquartile ranges, and thin segments show 10th-90th percentiles across hospitals.

Figure S2. Secondary endpoint transportability



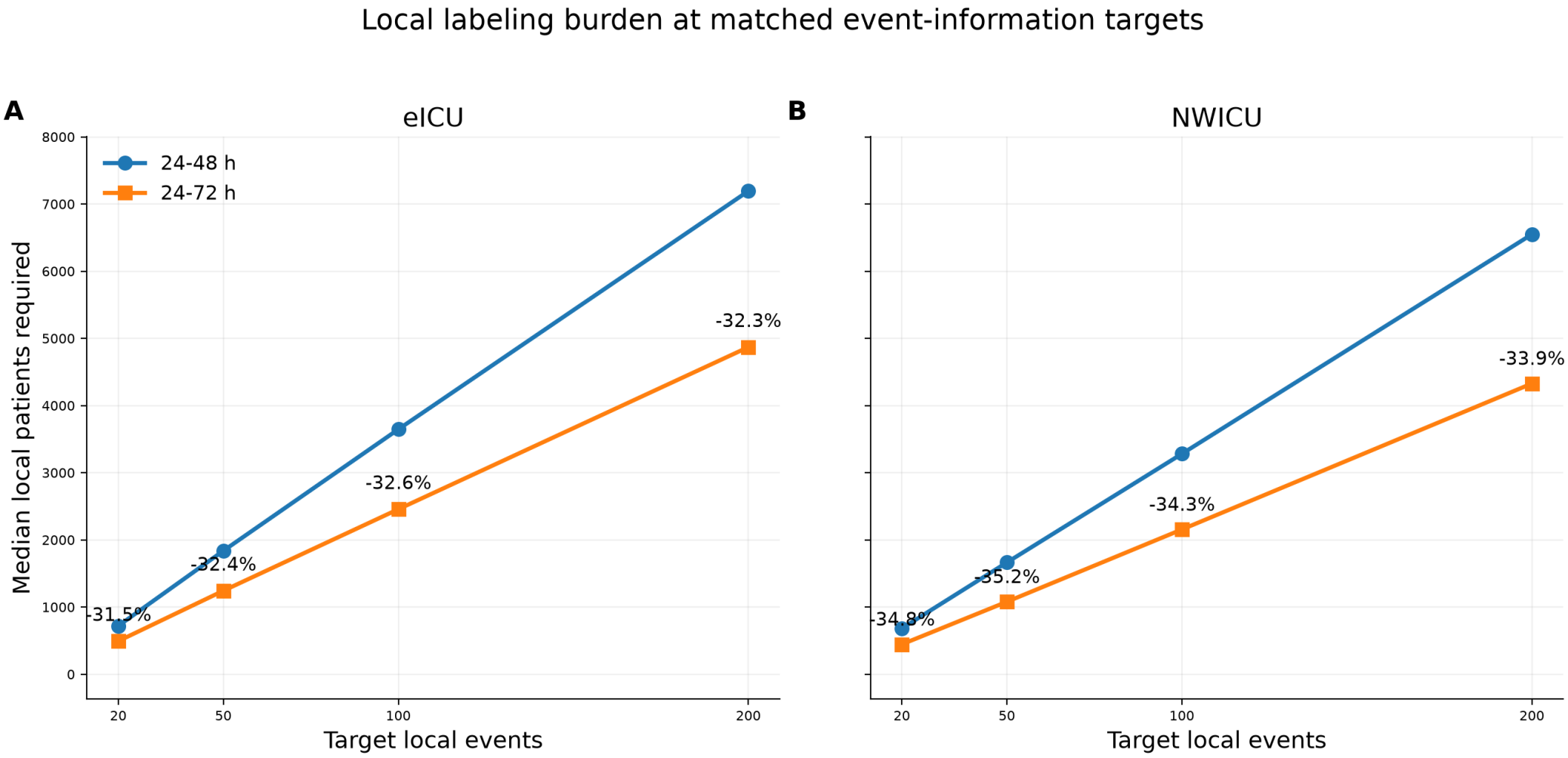
Transportability of fixed secondary outcome models. Patient-clustered 95% bootstrap percentile intervals are shown for AUROC and calibration-in-the-large across MIMIC internal validation, eICU, and NWICU for the 24-72 h composite and the death and new invasive mechanical ventilation components.

Figure S3. Secondary 24-72 h composite recalibration learning curves



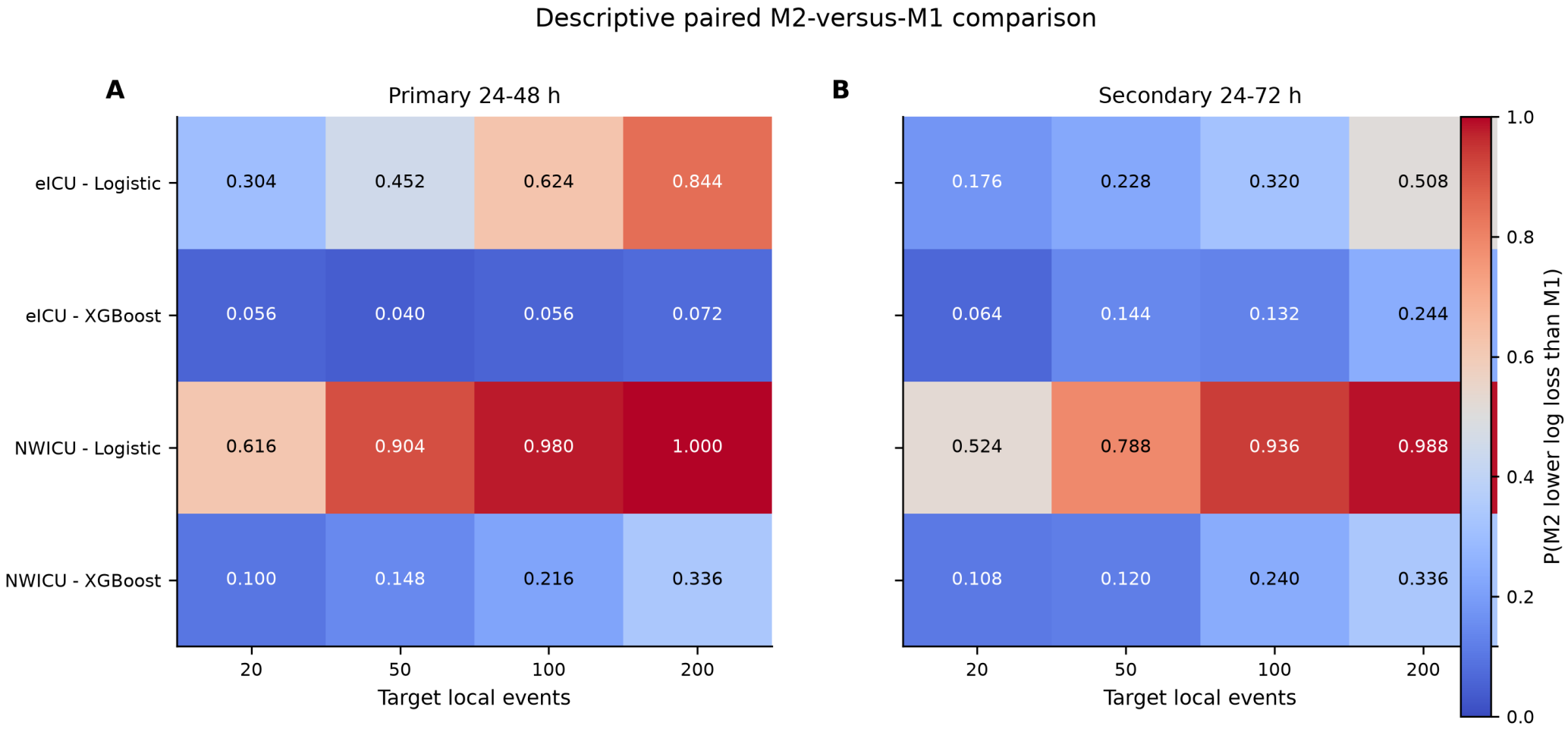
Secondary 24-72 h composite local recalibration learning curves. The plotting conventions match Figure 3. Shaded regions are empirical adaptation-sampling quantiles rather than population confidence intervals.

Figure S4. Cross-horizon local labeling burden



Local labeling burden at matched event-information targets. Median patient counts required to accrue 20, 50, 100, or 200 events are compared between the 24-48 h and 24-72 h composite outcomes using the same fixed adaptation pools and repeated sampling design.

Figure S5. Descriptive M2-versus-M1 paired comparison



This comparison is descriptive only; the prespecified estimands remain M1 versus M0 and M2 versus M0.

Descriptive paired comparison of M2 versus M1. Cells show the fraction of paired repeated adaptation samples in which M2 had lower log loss than M1. This comparison is descriptive only; the prespecified estimands remain M1 versus fixed M0 and M2 versus fixed M0.