Consumer sleep-tracker validation against PSG: one dot per device in each study (primary analysis rows only). All panels are 2-stage sleep/wake except the two STAGING panels (four-class wake/light/deep/REM).

Dashed green = candidate good zone (TST bias ±15 min; LoA ±60 min; sensitivity ≥92.5%; specificity ≥47.5%; accuracy ≥85%; κ ≥0.425; PABAK ≥0.725; F1 ≥0.90; staging accuracy ≥70%; staging κ ≥0.60). Whiskers = 95% CI where reported. LoA panel: dot = bias, bar = 95% LoA; dotted bar = LoA derived here as bias ± 1.96·SD (paper reported SD only). F1 (sleep) is derived, not reported: exact from epoch counts or wake-class F1 where published, otherwise estimated from sensitivity, specificity and PSG sleep prevalence (hollow markers). Sleep/wake κ is hollow where the paper reports none and it is reconstructed the same way. Sensitivity/specificity swapped to sleep-positive convention for Stucky 2021, Lee 2023 and Labie 2025. Diamonds = manufacturer white papers, not peer reviewed. '(median)' marks values that are medians of per-night values rather than means or pooled epochs (Apple 2025). Gray ring = sample mean age ≥ 50 y. n = participants analysed for that device.

