

Supplementary Table S6. Estimator sensitivity and bootstrap stability of the two reference-standard-derived cutoff strategies

Final analytic cohort (n = 528). Both the extended Youden index and the ROC-surface (VUS) criterion were derived against the same manual three-tier interpretation as the reference standard. Cut points are expressed as p53 H-scores.

(A) Effect of the estimator on the derived cut points

Criterion	Estimator		Lower cutoff	Upper cutoff
Extended Youden index	Normal-distribution (as reported in the main text)		14.5	100.8
Extended Youden index	Empirical (distribution-free)		12.5	102.9
ROC surface (VUS)	Empirical (distribution-free; as reported in the main text)		12.4	103.8

Under a common empirical estimator the two criteria agree almost completely: they place their lower boundary in the same gap of the observed distribution (no core lies between 12.4 and 12.5) and assign 527 of the 528 cores identically. The two rows report the boundary under different conventions — the VUS cut point is the observed H-score at the boundary (103.8), whereas the empirical Youden cut point is the midpoint of the adjacent interval (102.9, i.e. between the observed values 102.0 and 103.8). Because these two conventions fall on either side of a single core (H-score 103.8), that core is assigned to overexpression by the empirical Youden cut points and to wild-type by the VUS cut points; every other core is assigned identically. The residual difference between the two criteria is therefore one of estimator and reporting convention rather than of the criterion itself.

(B) Bootstrap stability of the cut points (2,000 case-resampling replicates)

Cut point	Point estimate	Bootstrap mean	SD	95% CI
Extended Youden index, lower	14.5	14.4	1.30	11.8–16.8
Extended Youden index, upper	100.8	100.9	2.87	95.5–106.9
ROC surface (VUS), lower	12.4	12.4	0.88	9.3–13.4
ROC surface (VUS), upper	103.8	103.3	3.07	95.3–108.1

(C) Paired between-method difference in the cut points (extended Youden index minus VUS)

Cut point	Mean difference	95% CI
Lower	+2.04	−0.20 to +4.35
Upper	−2.44	−8.52 to +4.75

Both confidence intervals include zero, indicating that the cut points derived by the two criteria are not statistically distinguishable.

(D) Concordance of the resulting three-tier classifications

Comparison	Discordant cores	Cohen's kappa
Extended Youden index (14.5/100.8) vs. VUS (12.4/103.8)	14/528 (2.7%)	0.955

Eleven of the 14 discordant cores lie between the two lower cut points (H-score 12.4-14.5) and 3 lie between the two upper cut points (H-score 100.8-103.8).

CI, confidence interval; SD, standard deviation; VUS, volume under the ROC surface.

Bootstrap replicates were drawn by resampling cores with replacement; for each replicate, both criteria were re-estimated de novo. Replicates in which any reference tier was empty were discarded. In the bootstrap and in the empirical sensitivity analysis, candidate cut points were taken as the midpoints between adjacent observed H-scores.