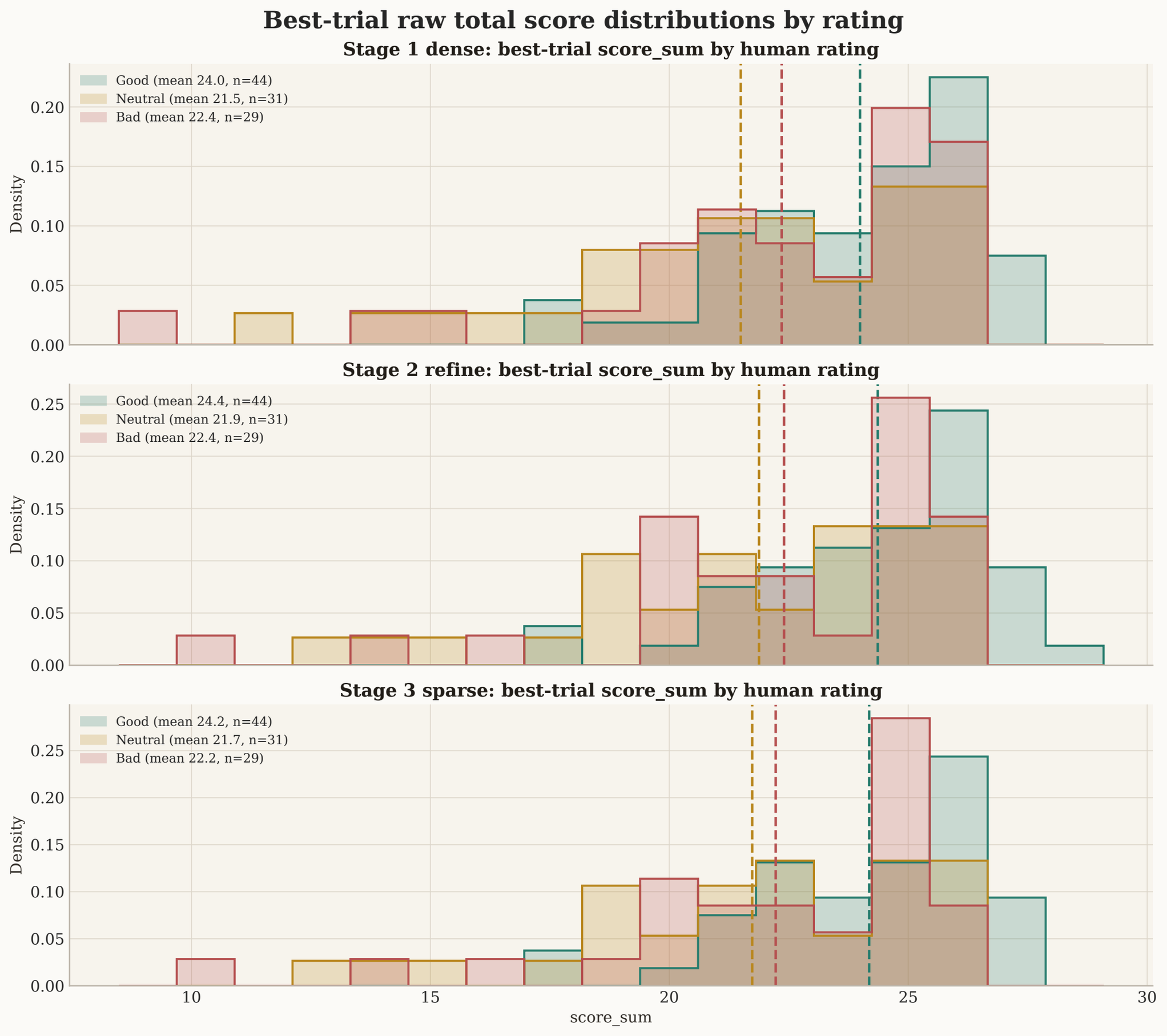
Best-trial raw total score distributions by rating
Stage 1 dense: best-trial score_sum by human rating
Good (mean 24.0, n=44)
Neutral (mean 21.5, n=31)
Bad (mean 22.4, n=29)
Density
Stage 2 refine: best-trial score_sum by human rating
Good (mean 24.4, n=44)
Neutral (mean 21.9, n=31)
Bad (mean 22.4, n=29)
Density
Stage 3 sparse: best-trial score_sum by human rating
Good (mean 24.2, n=44)
Neutral (mean 21.7, n=31)
Bad (mean 22.2, n=29)
Density
score_sum