

reliability diagram (equal number of subpop. scores per bin)

