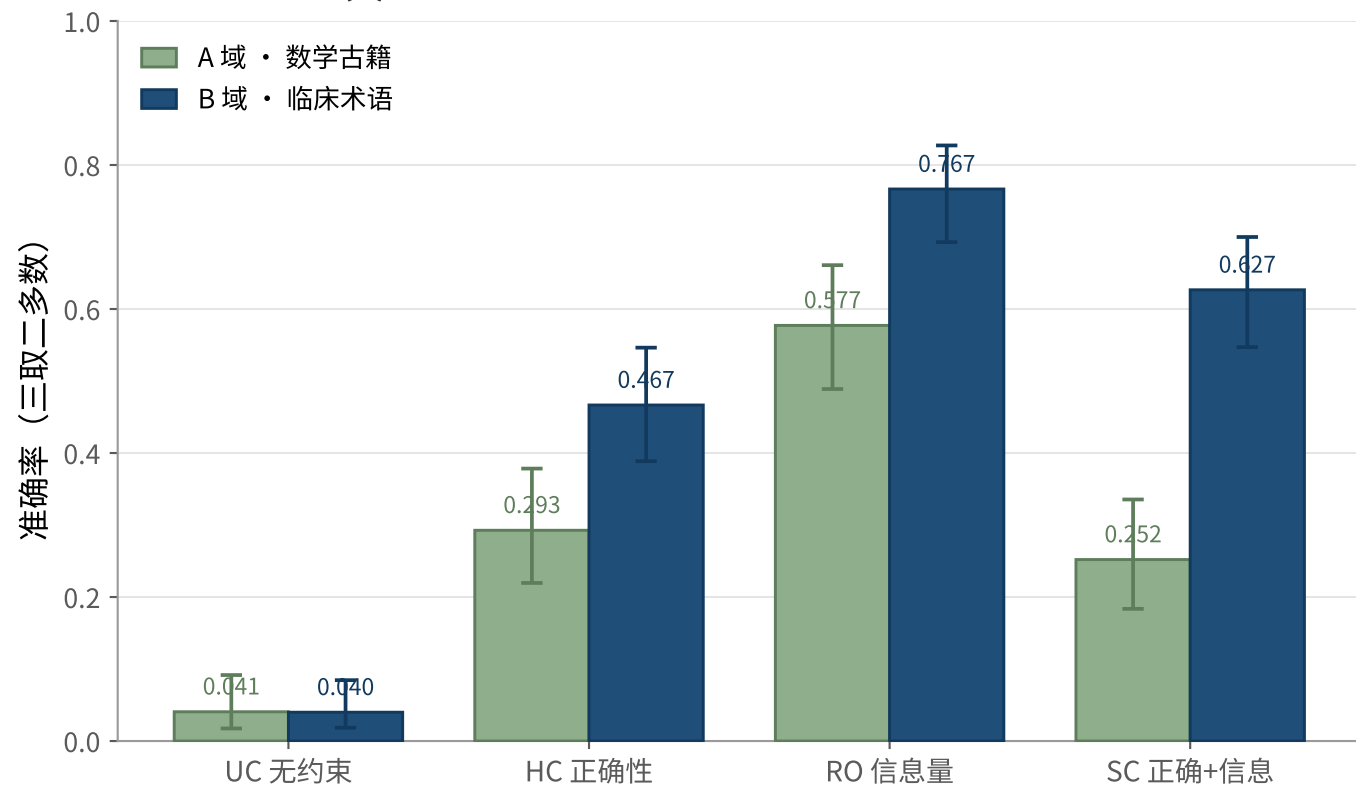


图 B 信息量 > 正确性：四条件行为与基线因果检验

(a) 四条件准确率（三取二多数）与 95% 置信区间



检索器 top-1 处处高于 HC，却始终不及 RO → 增益并非来自「含答案的候选」，而是来自信息量。

(b) 基线检索器 vs HC / RO 天花板

