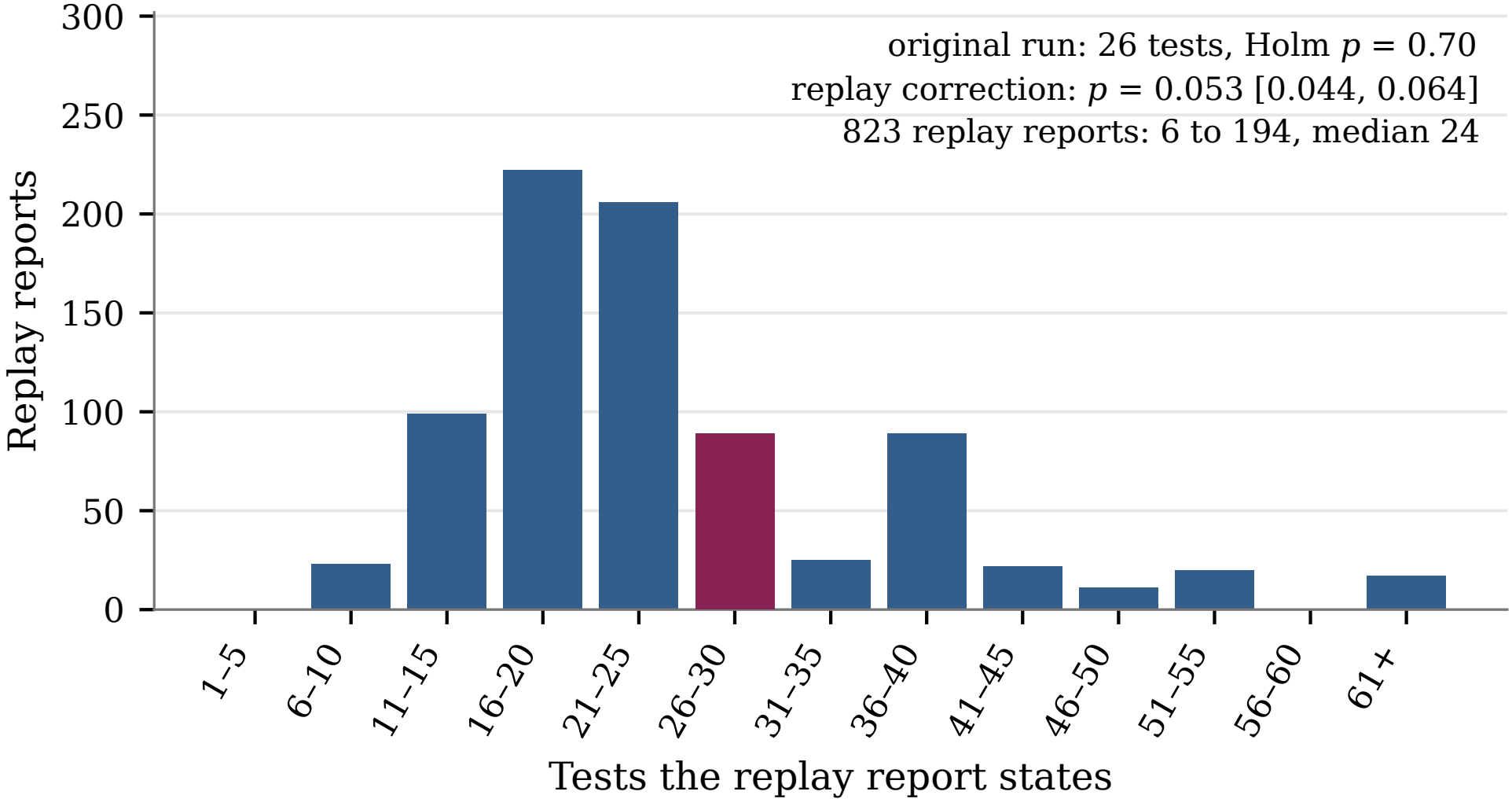

original run: 26 tests, Holm p = 0.70
replay correction: p = 0.053 [0.044, 0.064]
823 replay reports: 6 to 194, median 24
Replay reports
300
250
200
150
100
50
0
1–5
6–10
11–15
16–20
21–25
26–30
31–35
36–40
41–45
46–50
51–55
56–60
61+
Tests the replay report states