Unmarked

Time to fill changed by +1.3 days in the rollout group relative to the comparison group (95% range -5.2 to +7.8 days). The rollout group started at 48.2 days.

- The smallest change these groups can reliably detect is about 13.1 days. You said 4.8 days would
- matter, and this study is unlikely to see a change that small. A flat result would be ambiguous.

These are sample numbers. Replace them with an admin export or CSV.

- Baseline pulse response is 33%, under the 40% mark. Treat the experience numbers as thin.

Time to fill

Change	**+1.3 days**
95% range	**-5.2 to +7.8 days**

Change in the rollout group minus change in the comparison group. Each week counts equally. Inside a week, each team counts by the number of observations on that row. The range is a 95% interval from week-to-week variation.

The other metrics

The agreed metric is the one in the reading above. These are the other metrics on file.

- 90-day retention: +0.1 pp, range -1.5 pp to +1.7 pp, range includes no change
- HR case cycle time: +0.83 hours, range -1.3 to +3 hours, range includes no change
- Requisitions filled: +0.01 requisitions per week, range -0.25 to +0.26 requisitions per week, range
- includes no change

What it cost

Loaded pay is an assumption you can edit: $2,800 per person per week. Cost per requisition filled uses the rollout group's before and after. It is not adjusted by the comparison group. The agent's own cost, when shown, leaves human pay out and is not comparable to the other rows.

Seat spend	**$1,530**
Token and API spend	**$360**
Agent spend	**$0.00**
Other spend	**$0.00**
In this window	**$1,890**

In a typical measurement week, 71% of the 17 seats in the rollout group were used.

- People, before: $7,329 per requisition filled. Loaded pay in the rollout group, divided by weekly
- volume before the rollout.
- People plus the tool, after: $6,872 per requisition filled. Loaded pay plus seat, token, and other tool
- spend, divided by weekly volume after.

Adoption

Who used it, at the team. This explains the rollout. It is not a second result.

100% of job titles were recognized without an edit.

Baseline response 33% of 30 people invited.

Reading

The smallest change these groups can see is larger than the gain that was set as mattering.

Who was in it

- Recruiting: First wave, 9 people
- Total rewards: First wave, 8 people
- People ops: Hold back, 7 people
- Learning: Hold back, 6 people

How this was calculated

The number is the first wave's change minus the held-back group's change over the same weeks. Each week counts equally. Inside a week, each team counts by the number of observations on that row. A row under the privacy floor is left out, and a week is left out when the included observations add up to less than the floor. The range is how much those weekly results bounced. It is not proof the tool caused the change.

The reading stays a description when the metric or the groups were changed after they were confirmed, when everyone got it at once, when the range still includes both no change and an improvement that would matter, or when the groups are too small to see that improvement. Groups under 5 are hidden before any of this runs.

The smallest change these groups can reliably detect is about 13.1 days. You said 4.8 days would matter, and this study is unlikely to see a change that small. A flat result would be ambiguous.

The calculations workbook repeats every row. On its Estimate sheet, column B is a formula and column C is the number in this report. They should match.