Still measuring

PR cycle time changed by -9.7 hours in the rollout group relative to the comparison group (95% range -11.3 to -8.1 hours). The rollout group started at 45.7 hours.

- The plan is 10 measurement weeks. 4 are on file, so this is a reading so far.

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

PR cycle time

Change	**-9.7 hours**
95% range	**-11.3 to -8.1 hours**

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.

- PRs merged: +0.37 PRs per person per week, range +0.3 to +0.45 PRs per person per week, agreed
- direction
- Change failure rate: +0.3 pp, range -0.2 pp to +0.9 pp, range includes no change
- Time to first review: -3 hours, range -4 to -2.1 hours, agreed direction

What it cost

Loaded pay is an assumption you can edit: $4,500 per person per week. Cost per merged PR 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	**$2,880**
Token and API spend	**$676**
Agent spend	**$0.00**
Other spend	**$0.00**
In this window	**$3,556**

In a typical measurement week, 70% of the 80 seats in the rollout group were used.

- People, before: $2,833 per merged PR. Loaded pay in the rollout group, divided by weekly volume
- before the rollout.
- People plus the tool, after: $2,279 per merged PR. 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.

75% of job titles were recognized without an edit.

- Light: 27 people, 2.10 sessions each, breadth 22%, outcome change -1.5
- Selective: 31 people, 7.50 sessions each, breadth 48%, outcome change -4
- Embedded: 16 people, 14.00 sessions each, breadth 74%, outcome change -10

Baseline response 62% of 124 people invited.

Reading

The planned weeks are not all on file. This is a reading so far.

Who was in it

- Platform: First wave, 36 people
- Product: First wave, 28 people
- Data: First wave, 16 people
- Mobile: Hold back, 18 people
- Security: Hold back, 12 people
- Reliability: Hold back, 14 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 1.3 hours. You said 4.6 hours would matter, and this study can detect a change of that size.

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.