SurveyCTO

Allowing Statistical Checks to Guide Survey Team Attention Toward Problems in Collected Data

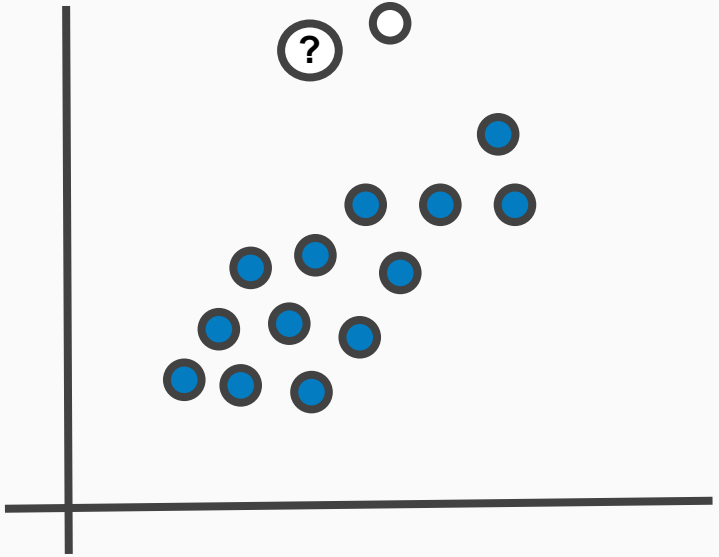
Amrik Cooper
Dobility, Inc.

Outline

1. Summary
2. Background
3. Problem Statement and Questions
4. Experiment
5. Survey
6. Conclusion

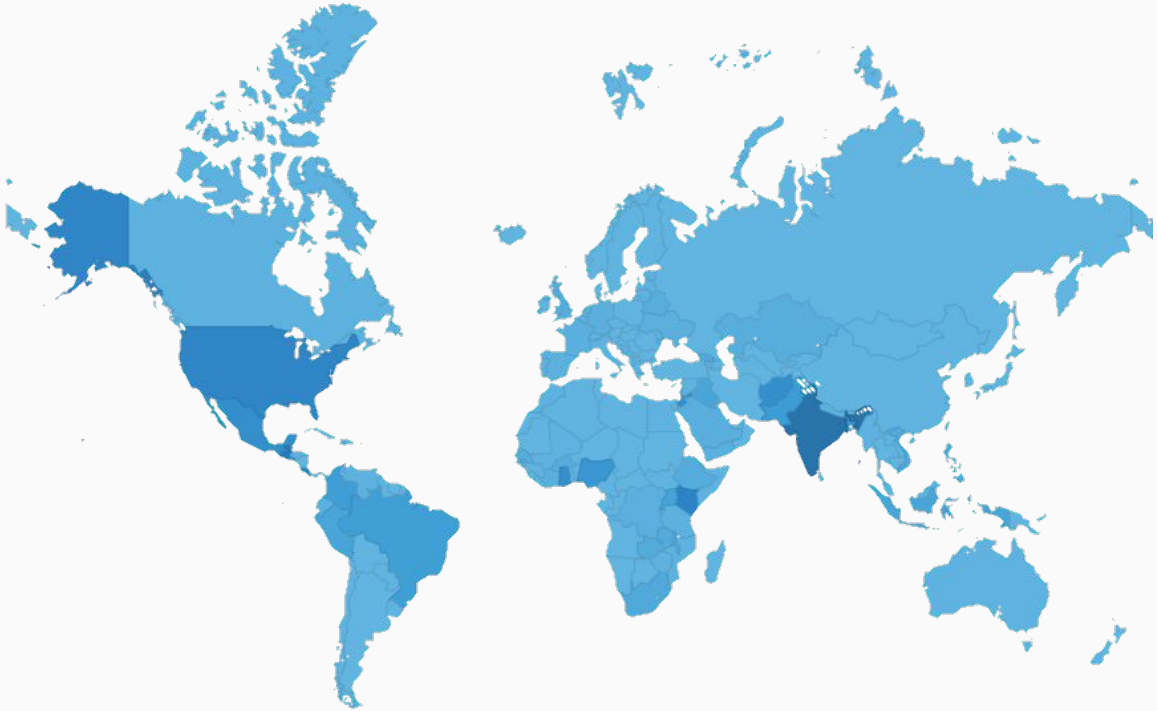
1. Summary

- Statistical checks for survey data quality assurance
- **Theory**: busy researchers sometimes don't configure quality checks in time
- **Survey results**: 40% don't use statistical checks
- **Experiment:** modest success with user nudges



2. Background

- SurveyCTO platform launched in 2013
- CAPI and CATI surveys
- Users in 165+ countries from different sectors

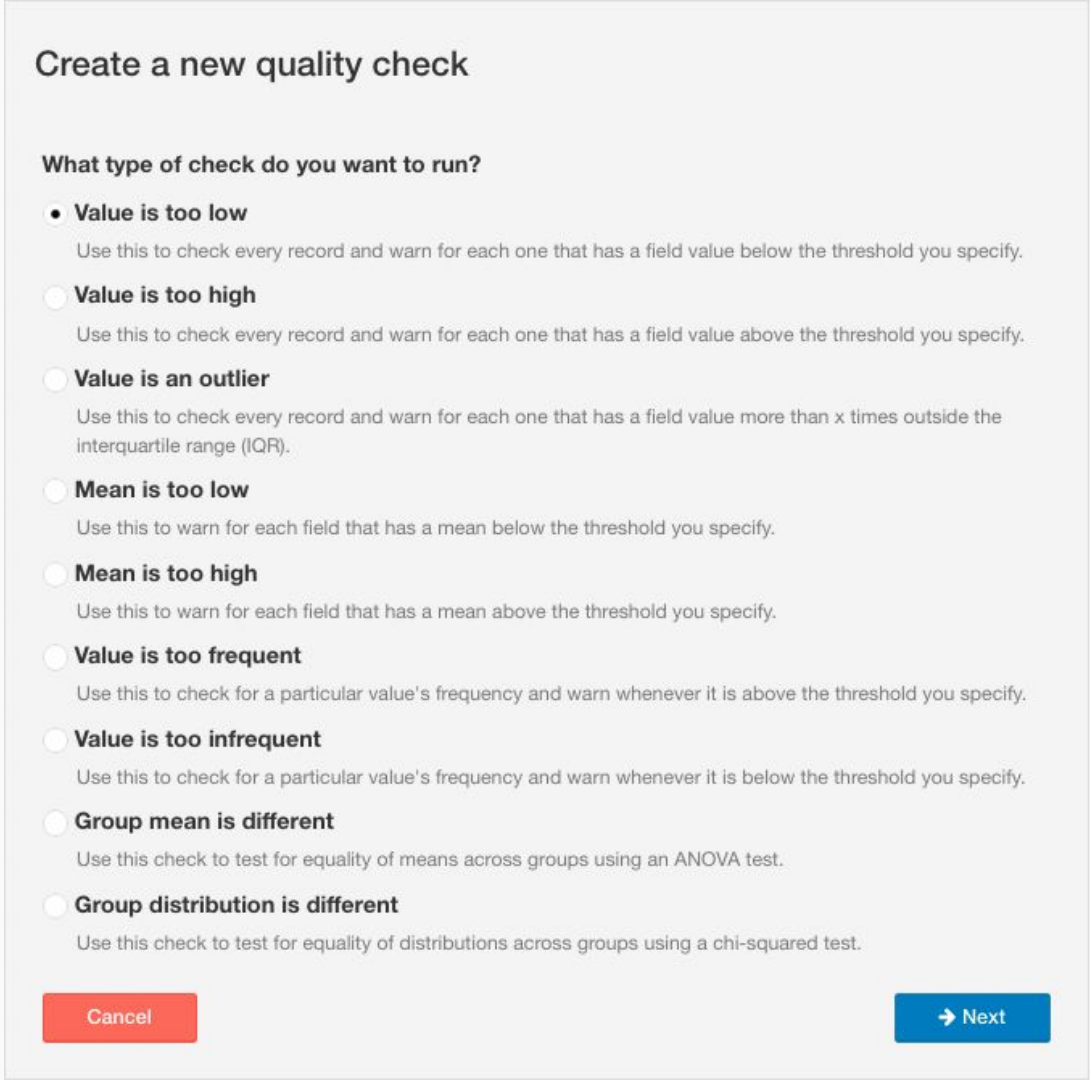


SurveyCTO

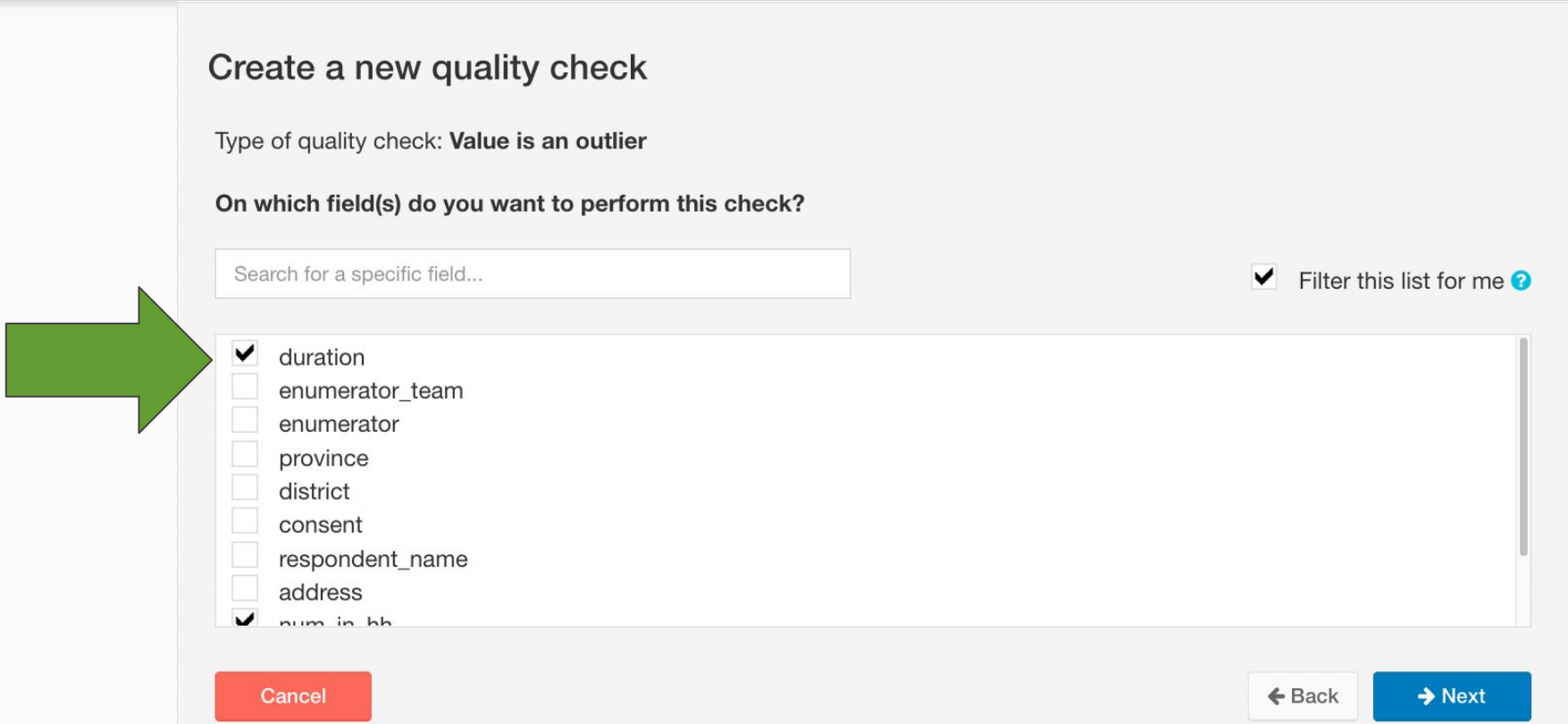
2. Background: Automated Quality Checks

- Statistical checks with user-defined thresholds
- Outliers, enumerator effects, high values
- Complement to field validation
- Creation process

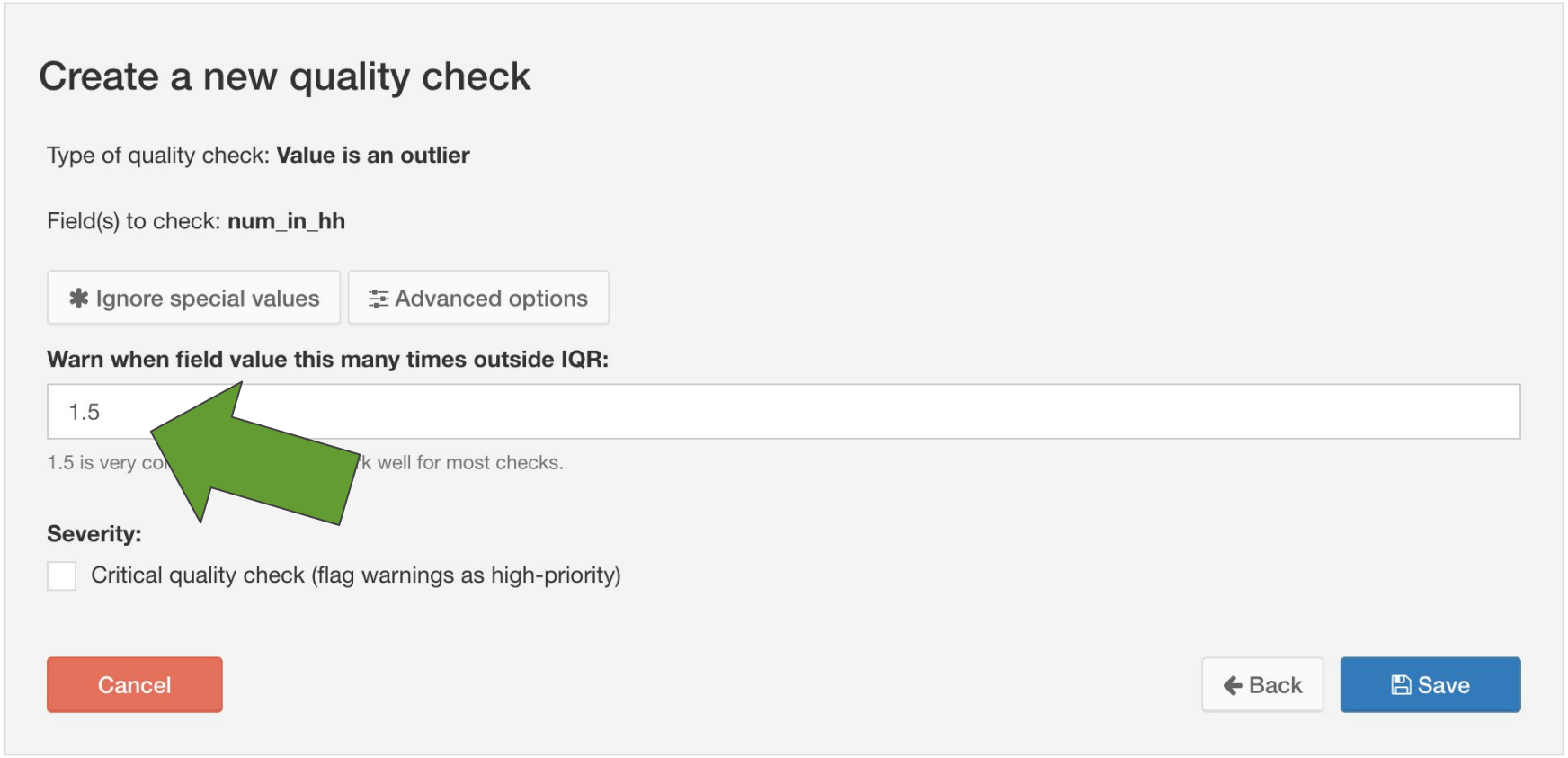
SurveyCTO



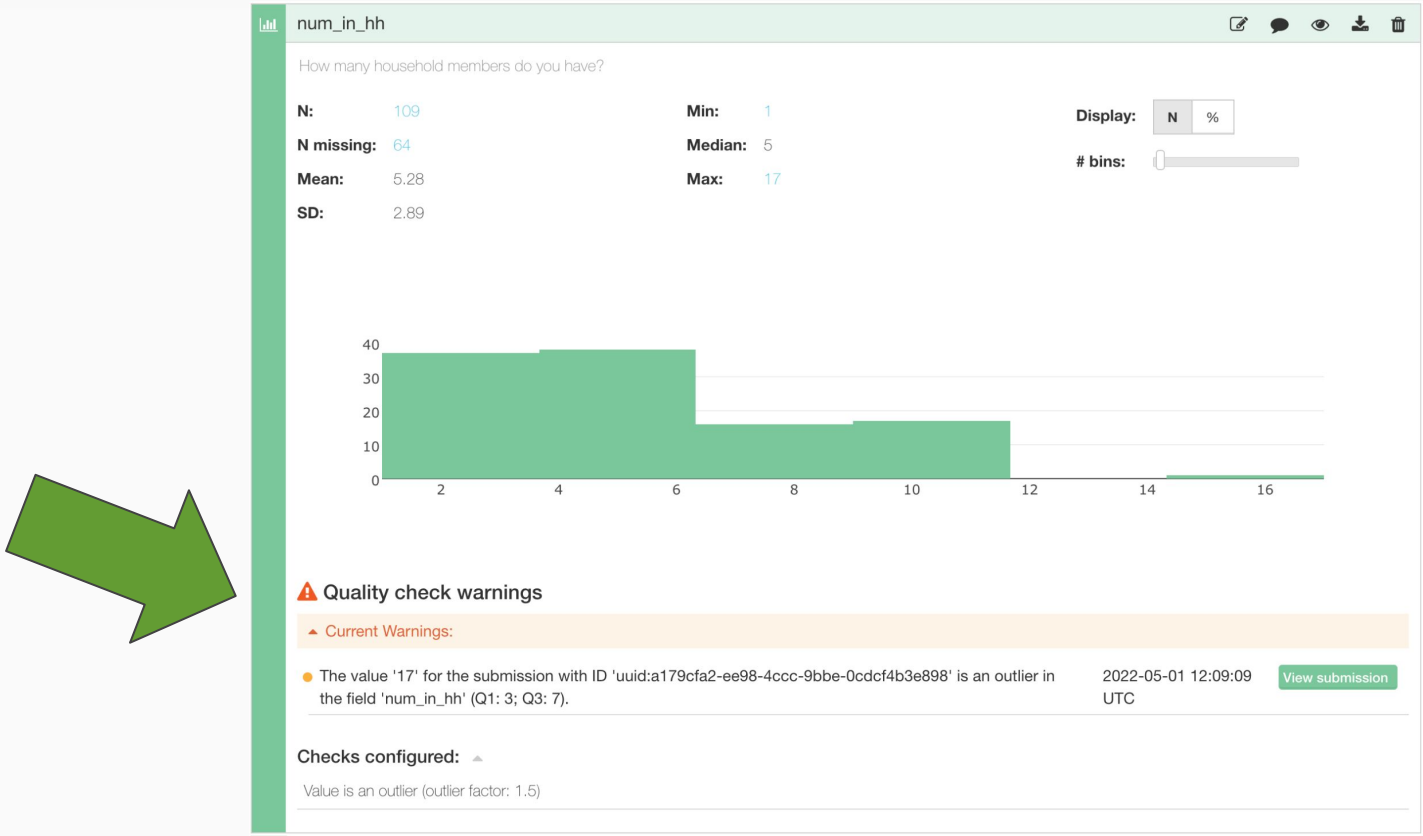
2. Background: Automated Quality Checks



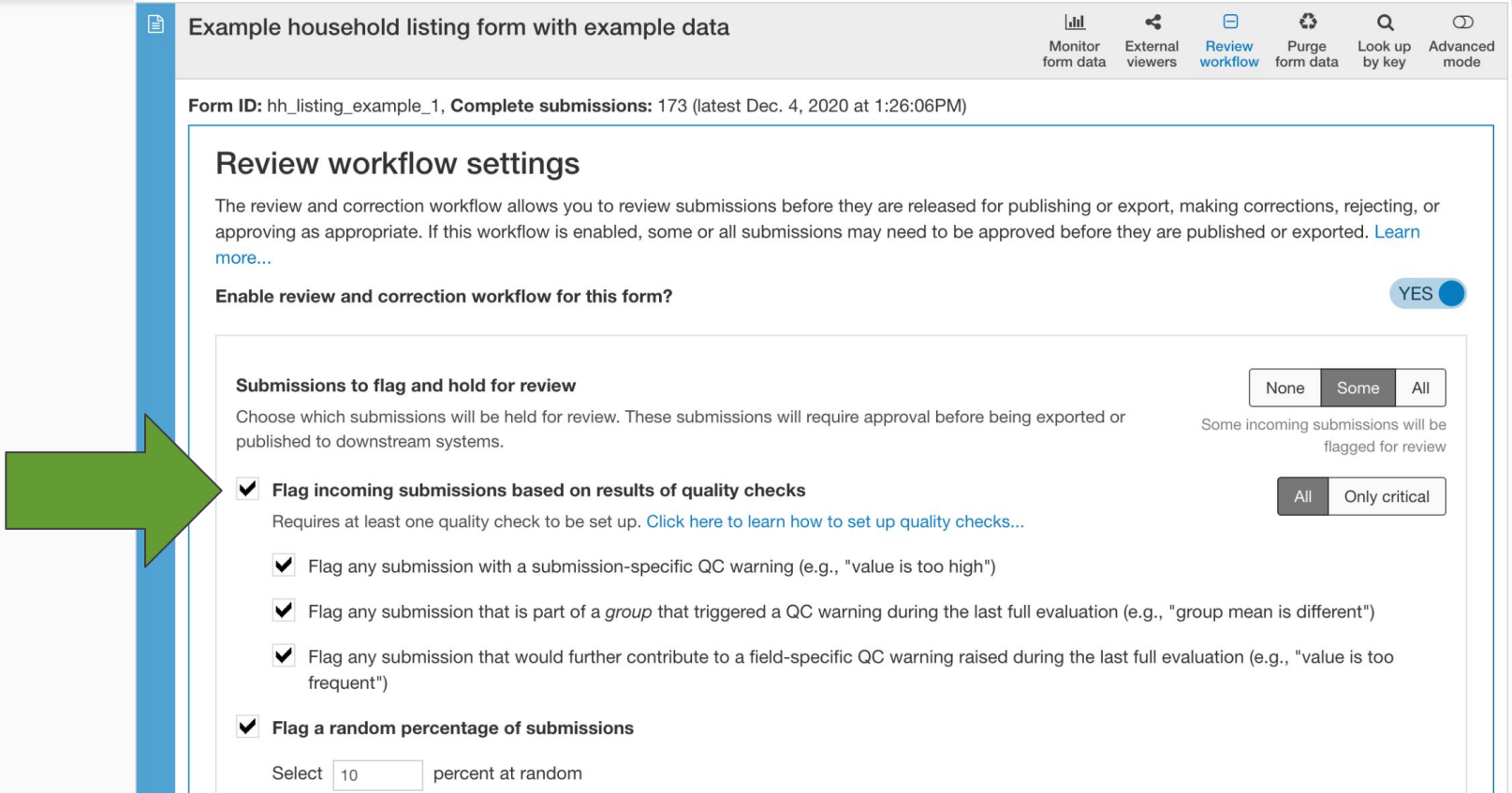
2. Background: Automated Quality Checks



2. Background: Automated Quality Checks



2. Background: Automated Quality Checks

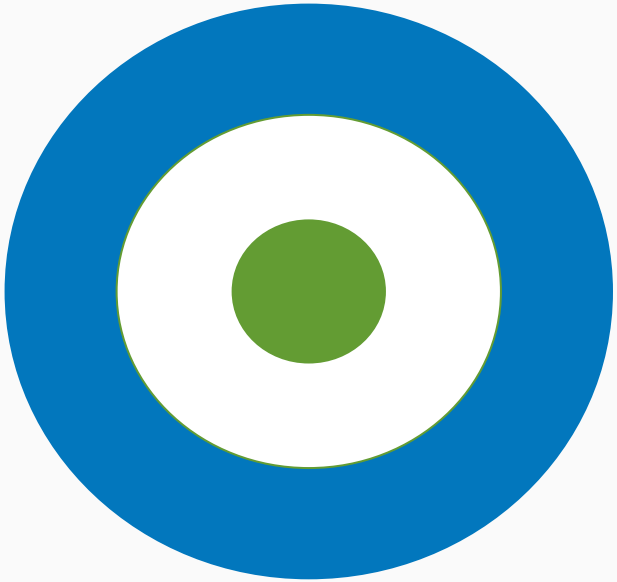


3. Motivation & Problem Statement

Problem: Low levels of statistical data quality check deployment.

Question 1: Why do relatively few users seem to use quality checks?

Question 2: Can we nudge users toward best practices and have them configure quality checks?



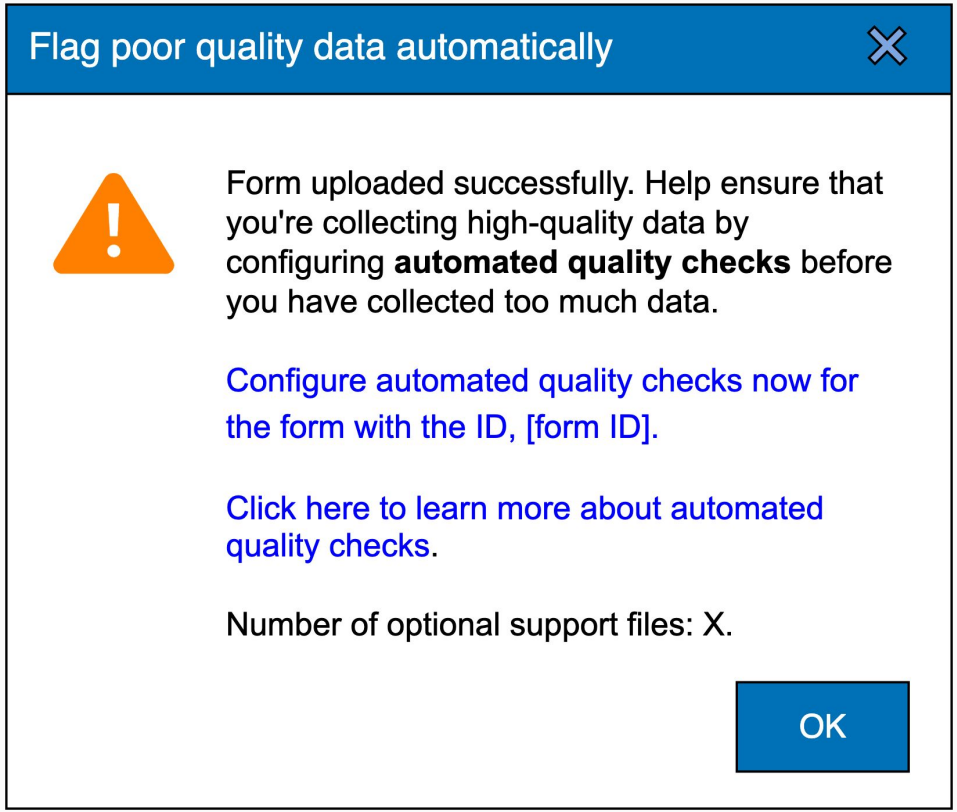
SurveyCTO

SurveyCTO

4. Experiment

4. Experiment Method

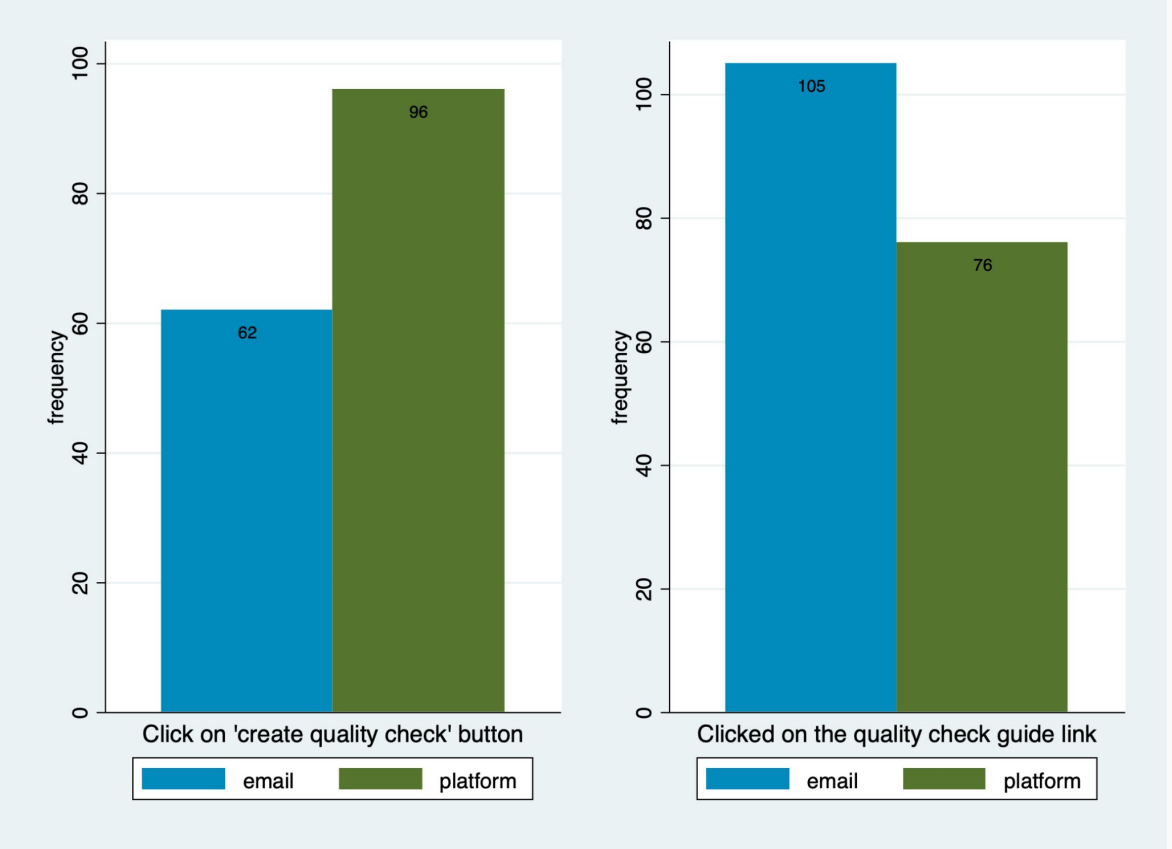
- Updated messaging to "nudge" users toward quality check creation
- Randomized experiment with three groups:
  - Control group
  - In-platform treatment
  - Email treatment
- Data collection with Google Analytics and a browser cookie



4. Experiment Interaction

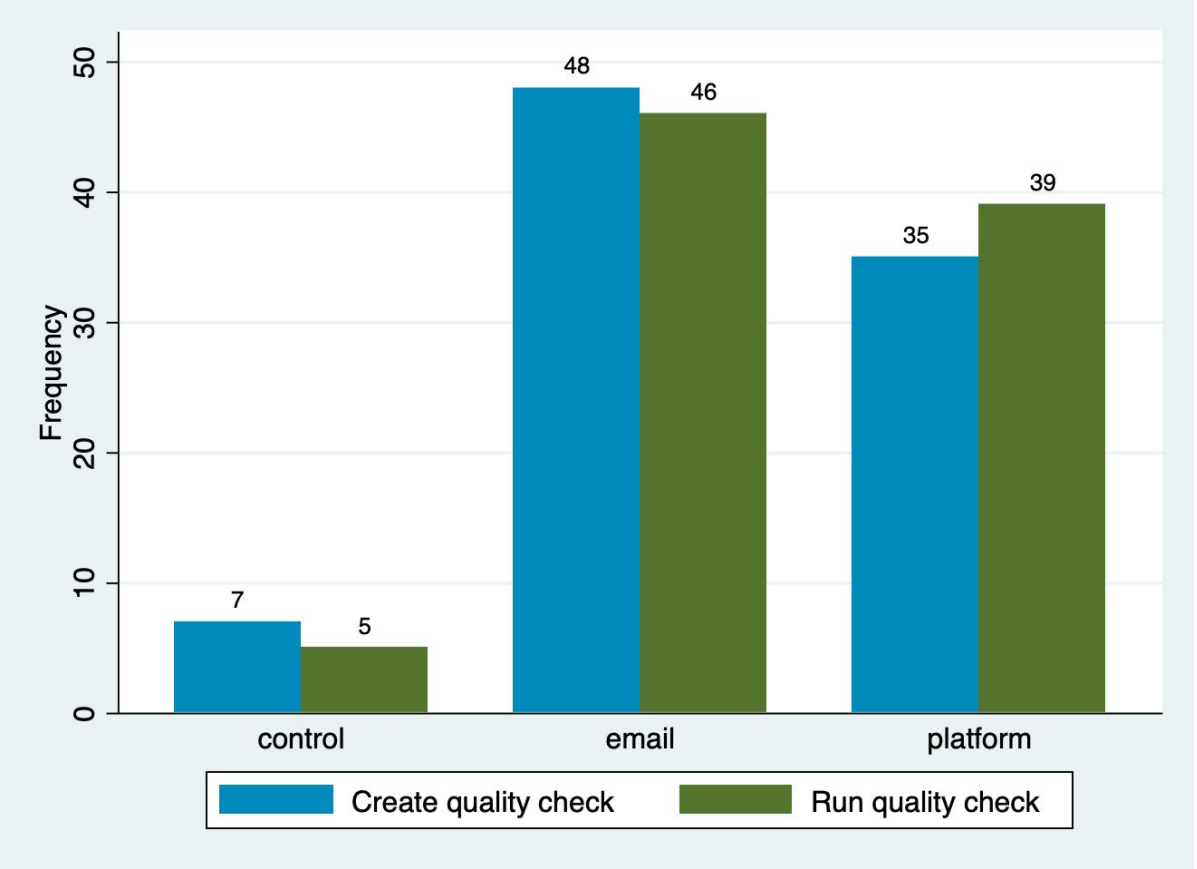
- Users in experiment: 1,181.
- Quality check experiment engagement:
  - In-platform: 96 engagements (22.8%*)
  - Email: 62 engagements (15.7%*)

* Percentages are response rates to treatments.



4. Experiment Results

- Both treatments outperformed control.
- Less significant difference between treatments, with email slightly ahead.
- Both event distributions significant at $p < 0.0001$.



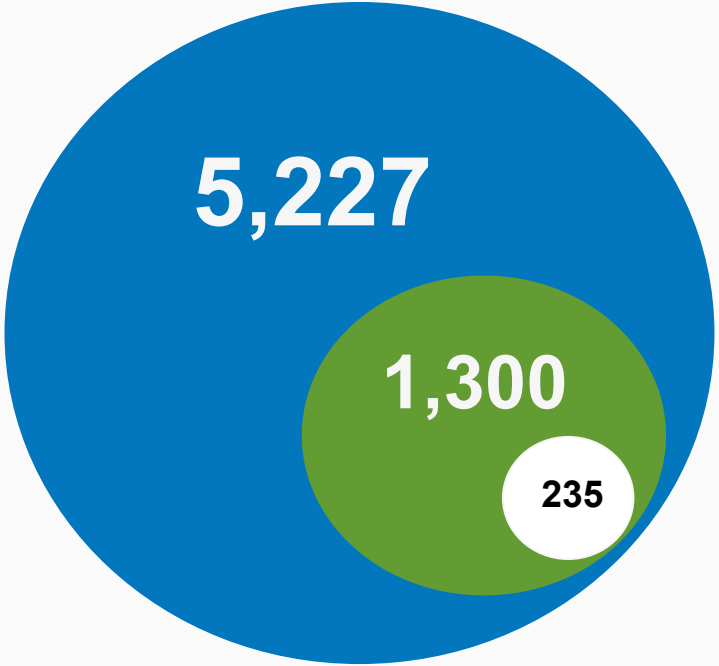
SurveyCTO

SurveyCTO

5. Survey

5. Survey Method

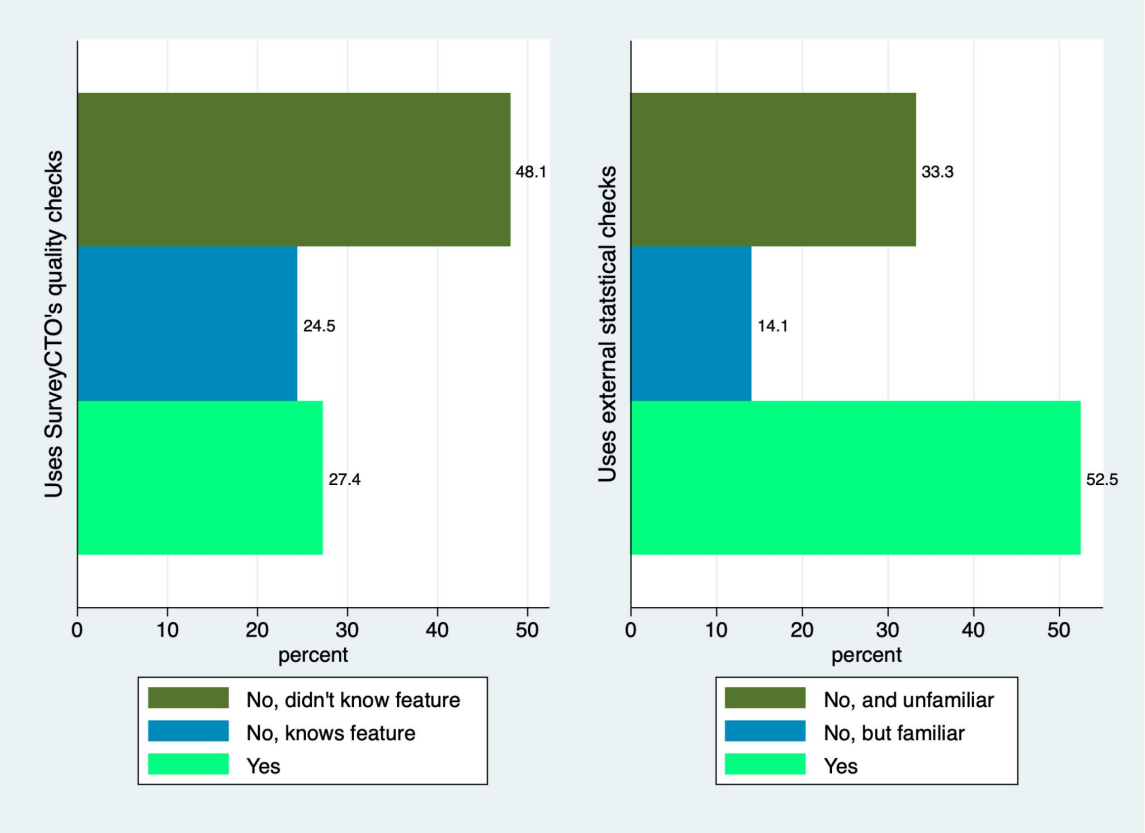
- **Population**: 5,227 paid administrative users logging in during a two-week period.
- **Sampled**: 1,300 users (~25%).
- **Responded**: 235 (~18%).



5. Survey Results

- 48.1% weren't aware of automated quality checks [41.3%, 54.9%].
- 52.5% used third party tools [45.5%, 59.5%].

Confidence intervals at 95%

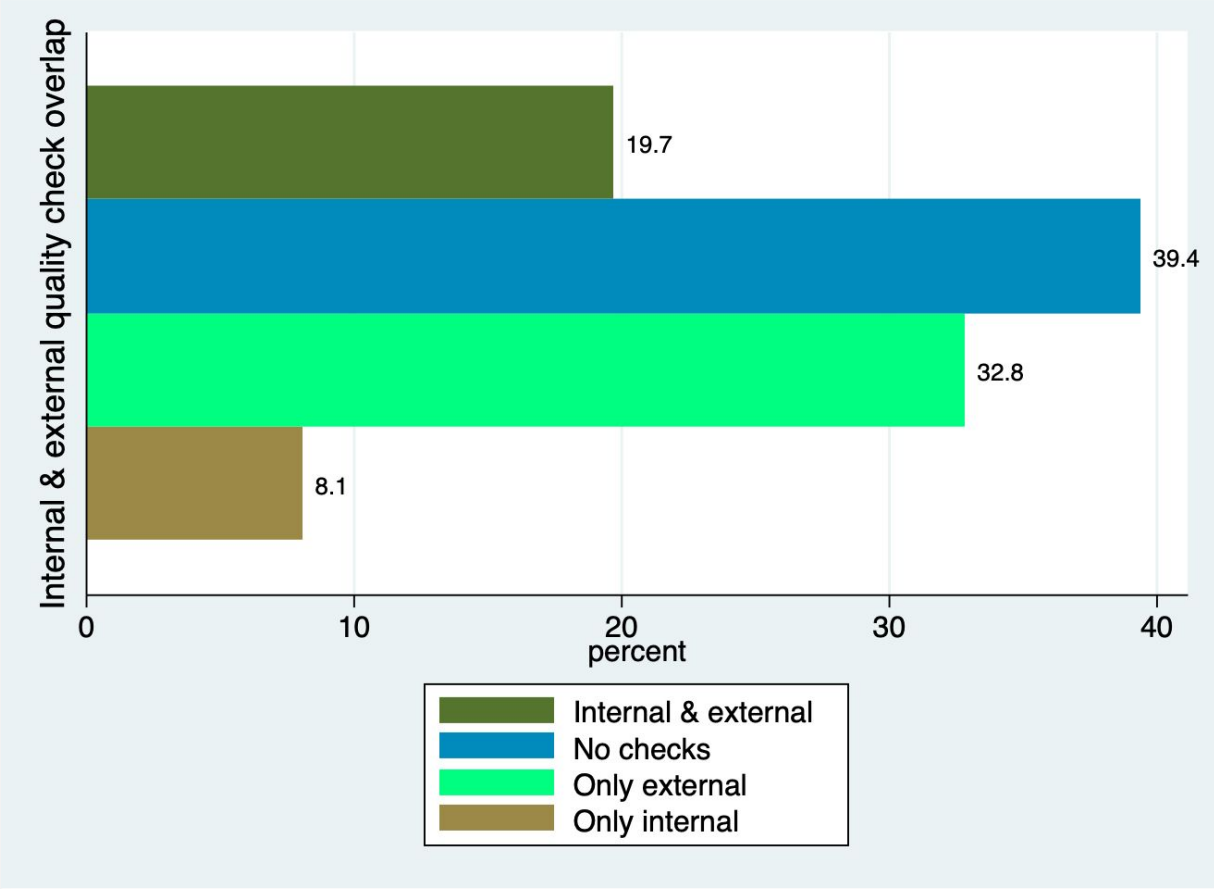


SurveyCTO

5. Survey Results

- Almost 20% use both [14.1%, 25.28%]
- 8.1% exclusive internal checks [4.25%, 11.9%]
- 39.4% use no checks [32.5%, 46.2%]

Confidence intervals at 95%

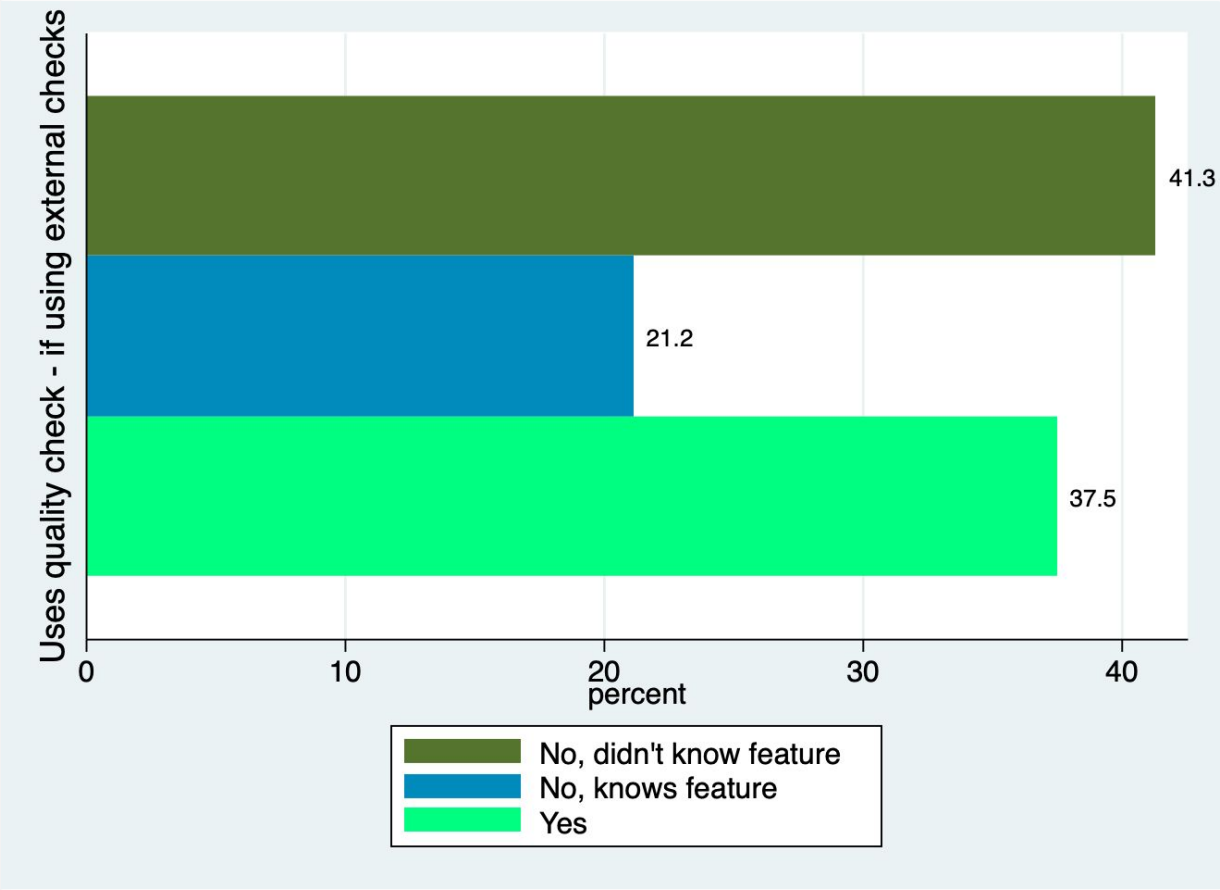


SurveyCTO

5. Survey Results

- 37.5% of external check users used internal checks too [28%, 46.9%]

Confidence intervals at 95%



SurveyCTO

6. Conclusions

- Statistically significant experiment results
- Treatment effectiveness rate: 9.27% vs 19% engagement rate
- Absolute effect: 71 net surplus quality checks
- 6 month history: 58.83 avg, std dev 51.84.
- Improved design in future, with better targeting?

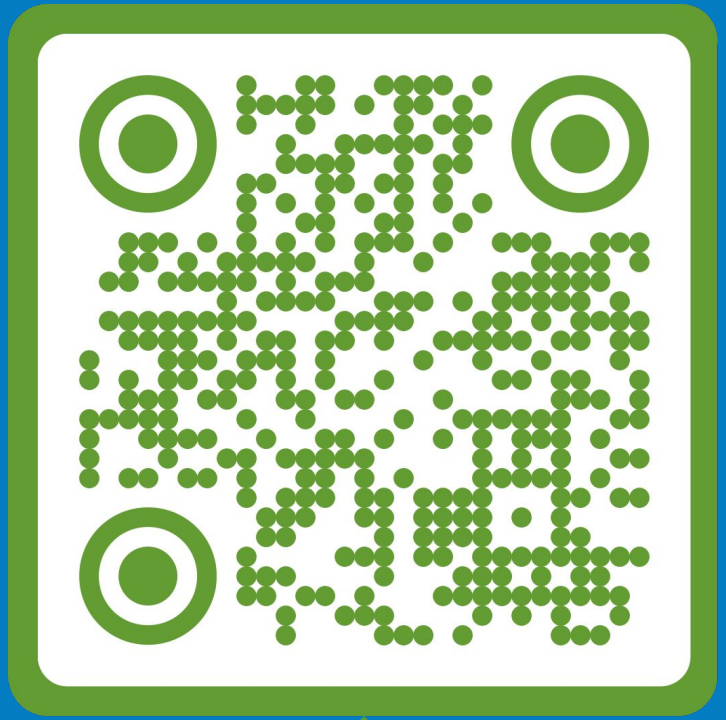
SurveyCTO

- 40% of users don't use checks [32.5%, 46.2%]
- Internal quality check usage overlap 70.9% with external quality check usage [58.51%, 83.3%]

Confidence intervals at 95%.

SurveyCTO

Thank you!

Amrik Cooper

amrik@surveycto.com