



2024 H2

RideCheck Holdout Group

Safety Effectiveness Measurement

Uber

EXHIBIT 1068

Deponent: Ross 30(b)(6)
Date: 6/12/2025

Rebecca Romano CSR No. 12546

Trial Exhibit No.
P-00320

Meeting goals

- 01 Align on Problem Definition
- 02 Refine Proposed Solutions
- 03 Kick-off Final Solution Design



Project Overview

PROBLEM STATEMENT

We don't know RideCheck's current (baseline) impact on incident rate, and want to track that for future changes.

PROPOSED SOLUTION

A per-trip, always-on holdout group design allowing us to set a baseline and measure product evolution.

KEY LINKS

[PRD](#) | [ERD \[tbd\]](#) | [DS Doc](#)

- [RideCheck SOT \(2024\)](#)
- [RideCheck SOT \(2021\)](#)

RESPONSIBLE

Data [@issamu.okada](#)

Engineering [@harish.shanker](#)

Product [@bruno.grimaldi](#)

Product Ops [@vidya.daryanani](#)

APPROVERS

Data Leadership [@sunny.wong](#)

Product Leadership [@mike.akamine](#)

REVIEWERS

Legal [@dkolta](#)

Product [@mariana.esteves](#)

Safety Ops [@alex.madsen](#)



01 Problem Definition

RideCheck is evolving



Problem definition

RideCheck today

simple heuristics → passive interventions

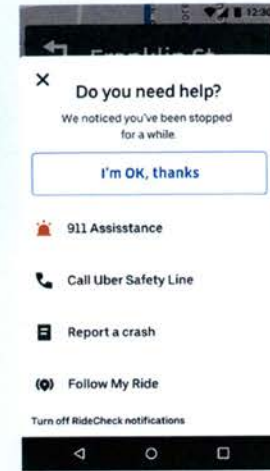
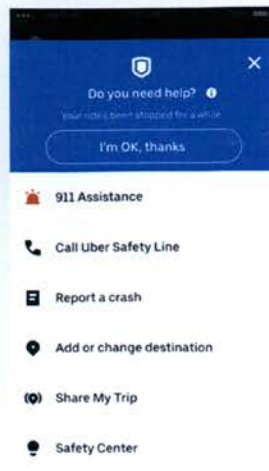
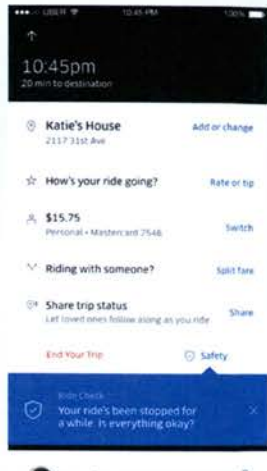
- Car stopped for 5+ minutes on trip
- Trip ends earlier than expected
- Route is too different from our estimate
- Push notification
- In-app callout with Safety Tools

Problem definition

At a glance

Current Heuristics

Heuristic	Description	Trigger Rate	Precision	Recall
Long Stop	Car has been stopped on-trip for a significant amount of time	1.14%	0.36%	7.90%
Midway Drop-Off	Trip completed at significant distance from destination.	0.76%	0.95%	13.70%
Route Deviation	On-trip at significant distance from closest point in "reference route".	0.08%	0.34%	0.50%
Crash Detection	Significant impulse (gyro + accelerometer) followed by long stop.	<0.01%	50.88%	12.25%



At a glance Current Interventions

RideCheck's interventions acknowledge that the trip is not going according to plan, and surface useful Safety Tools for the Rider and Driver in those moments. At their core they are:

- Low-intrusion
- Educational
- Quality of life
- Empathetic

Problem definition

Ops has questions re: RideCheck's effectiveness...

In Serious Incident investigations, we are often asked why RideCheck didn't do more:

- Why it didn't trigger earlier
- Why it didn't escalate actions
- Why it chose to not act at all

...because our
understanding of it
is changing

On the one hand, RideCheck is a low-intrusion, educational, quality-of-life feature.

On the other hand, it is "the" one platform in our portfolio that is watching trips in real-time and is capable of acting on them.

Problem definition

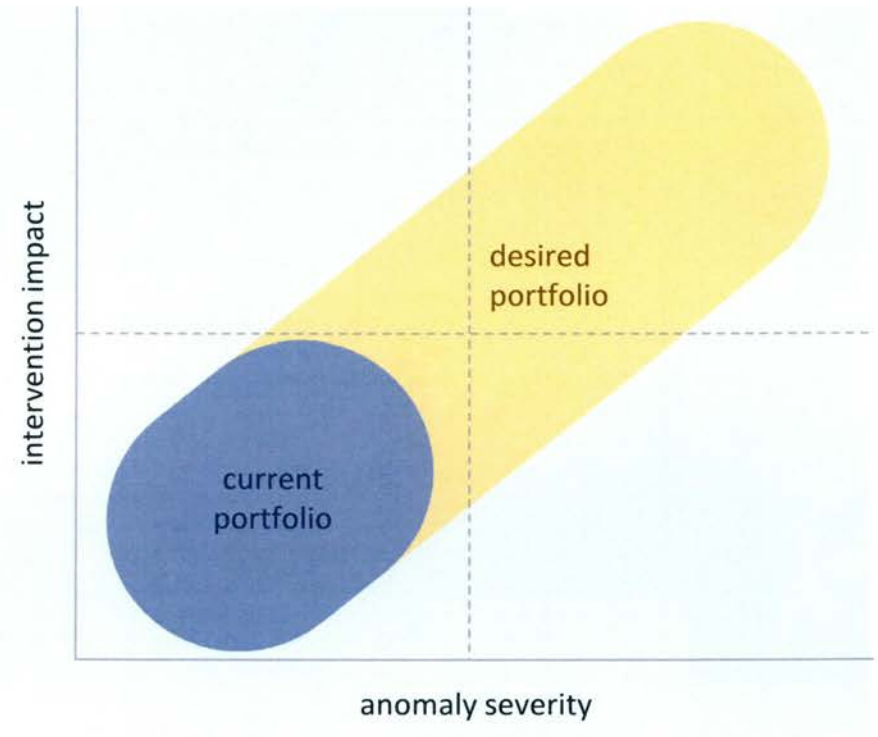
We want to extend RideCheck's portfolio

From quality of life check-ins

The original intent behind RideCheck is not abandoned. We intend to perfect current quality-of-life check-ins, evolving both their heuristics and the check-in experience.

To on-trip Safety policies

Simultaneously, we want to explore how we can use the RideCheck engine to better identify and react to sensitive moments for our users.



Problem definition

To support

simple heuristics → passive interventions

but also

sophisticated heuristics → proactive interventions

Anomaly Progressions

- Allows multiple signals to be considered
- First step towards a live on trip safety model

Destination Linger

- Suite of anomalies for a post-trip period
- Extends the coverage of RideCheck

Call the User

- Robocalls
- Live Agent-initiated calls*

Escalate

- Internally (eg. CommOps IIT*)
- Externally (eg. Guardian, U4B Operator*)

Problem definition

A "Live Trip Monitoring" mini-program

Baseline Anomaly Improvements

- Crash detection ML model
- Route Deviation "reference route"
- Multiple anomaly Long Stop

High-Severity Anomalies

- Anomaly Progressions
- Destination Linger





Issamu (Data Science):
But how will we measure the success of this mini-program? We've only ever measured RideCheck as a low-intrusion quality-of-life feature.

Problem definition

We don't know RideCheck's impact on Safety Incidents

On a montly basis, we only check how well RideCheck heuristics model incidents and how often we're sending checks.

In 2021 we did a treatment/control experiment and survey, tracking IPC deterrence potential and a few Safety Sentiment metrics.

Objective	Measured
Trigger Rate Are we sending too many checks?	Monthly
Precision % anomalies that became incidents	Monthly
Recall % incidents that had anomalies	Monthly
Committed to Safety Using our Brand Tracking question	Last in 2021 survey / experiment
Overall IPC Rate Is RideCheck is an IPC deterrent?	Last in 2021 survey / experiment



Issamu (Data Science):

To measure the impact of upcoming product development, we need to set a baseline of how RideCheck performs as an "on-trip safety policy" platform today.

02 Proposed Solution

RideCheck Holdout Group

Proposed solution

Holdout Description

1

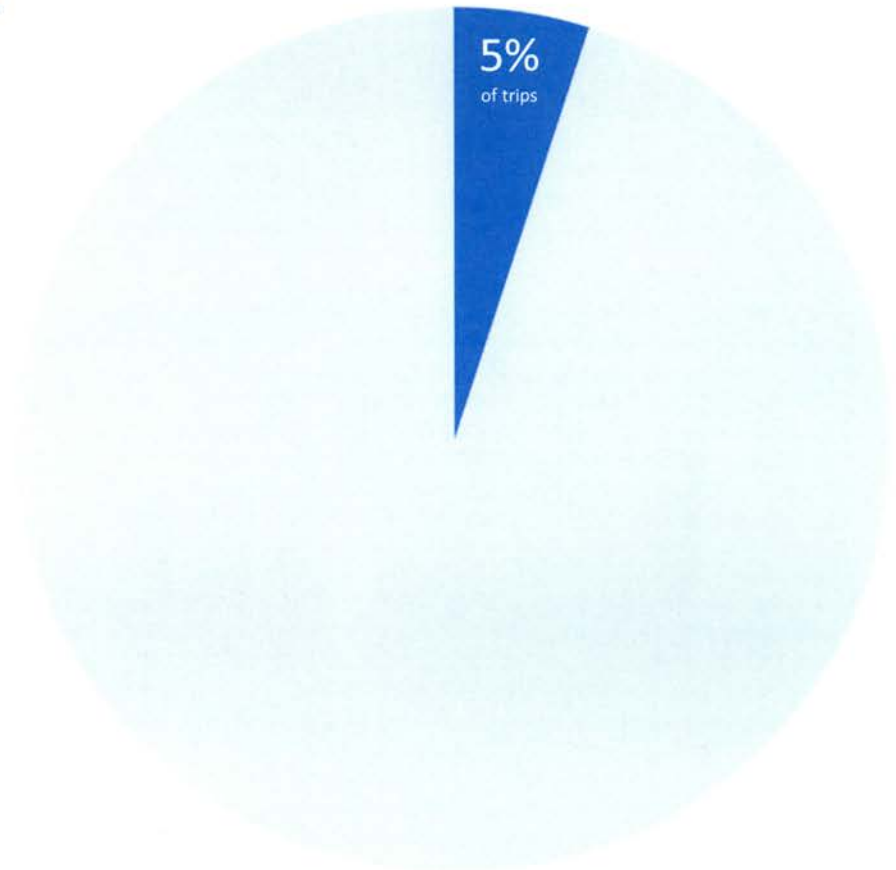
Methodology

At request time, a trip is assigned treatment or control. If it is in the control group, it will receive no RideChecks included in the scope of the Holdout Group. Treatment trips receive all eligible RideChecks as normal. On the backend, we keep track of all anomaly triggers as normal.

→ Per trip = All users benefit from RideCheck most of the time

In Scope

All Active Safety checks for Mobility trips. ([See what's out of scope here.](#)) Exclusions on exception basis only, such as regulatory requirements — and always accompanied by a DS analysis on exclusion impact to the Holdout.



Slide 16

1 @bruno.grimaldi@uber.com Is it 5% of all trips? or 5% of the triggered anomalies?
Reassigned to bruno.grimaldi@uber.com
Mariana Esteves, 8/16/2024

1 of all trips
Bruno Grimaldi, 8/16/2024

Proposed solution

Out of scope



✗ Delivery Workflows

Courier deliveries look very different from mobility trips, which would confound our analysis. They are also covered by a single RideCheck as of 2024 H2.



✗ Crash Detection

Road Safety's Crash Detection RideCheck is optimized for its own set of goals. It already has a more complex heuristic and higher-level interventions compared to other Mobility RideChecks.

Proposed solution

Known risks

Scale / Impact

An effective Holdout Group will exclude somewhere between 5–10% of trips from receiving any RideCheck intervention.

This is a significant volume of trips that will have one less Safety product covering them, which is compounded by our ambition to make this product more important to the overall on-trip Safety journey. There is a chance that an incident will occur that we did not have interventions for.

Inconsistent Experience

Because our design is per-trip, users (esp. Drivers) may realize their trip-to-trip experience is inconsistent, and attribute that to faulty or random behavior, eroding some of their trust in our Safety product portfolio.



Discussion

What metric are we
measuring here?

Proposed solution

Safety Incident Rate ← Our main candidate

Pros

A no-brainer

As "the" top-line Safety KPI, it requires no explanation. It is everywhere, allowing for comparison between RideCheck and other Safety features, as well as more systematic analyses.

Alignment

This metric corresponds with our intent to extend the RideCheck portfolio into higher-impact anomalies and interventions.

Cons

Confusion

A new RideCheck intervention may at once both decrease real-life incidents and also increase the rate of reporting. Because this metric encompasses the two effects, it becomes hard to interpret.

Whether this number goes up, down or stays the same is neither good nor bad if we can't separate those two effects.

Proposed solution

Safety Incident Rate Holdout Setup

2 5% holdout group (1 in 20 trips) is sufficient to measure effects as small as 3% in Safety Incident Rate in 1.5 months, and effects as small as 1% within a year.

	5% holdout	10%	15%	20%	25%
1% effect	52 wks	26	17	13	10
2%	13	6	4	3	3
3%	6	3	2	1	1
4%	3	2	1	1	1
5%	2	1	1	1	0
6%	1	1	<1	<1	<1
7%	1	1	<1	<1	<1
8%	1	<1	<1	<1	<1
9%	1	<1	<1	<1	<1
10%	<1	<1	<1	<1	<1

Uber | RideCheck Holdout Group Check Point | [Source](#)

Slide 21

2

Is the proposal to run this for at least 1yr and then revisit if we need to have this as an evergreen approach?

Mariana Esteves, 8/16/2024

2

The proposal is for an evergreen H/O. The 1 year figure is how long it would take to measure an effect as minute as 1%, given a 5% holdout.

Bruno Grimaldi, 8/16/2024

Proposed solution

L3/L4 Incident Rate ← An alternative

Pros

De-Escalation

Sophisticated heuristics with progressive interventions should help us de-escalate and prevent Serious Incidents.

While the overall Safety Incident Rate may go up by educating users on Safety tools, Serious Incidents should hopefully always trend down.

Cons

Exceptionality

Uber is already an exceptionally safe platform. L3/L4 incidents are rare enough that measuring effects in this rate is a finicky process subject to long lead times and susceptible to random variations.

Proposed solution

L3/L4 Incident Rate Holdout Setup

A 10% holdout group could measure medium effects (5%) on L3/L4 Incident Rate within a year, and large effects (10%+) within 3 months.

	5% holdout	10%	15%	20%	25%
1% effect	2,316 wks	1,158	772	579	463
2%	576	288	192	144	115
3%	255	127	85	64	51
4%	143	71	48	36	29
5%	91	45	30	23	18
6%	63	31	21	16	13
7%	46	23	15	11	9
8%	35	17	12	9	7
9%	27	14	9	7	5
10%	22	11	7	6	4

Uber | RideCheck Holdout Group Check Point | [Source](#)

Proposed solution

Discussion: Should we have a metric?

Diff-in-Diff Enabler

Having historic holdout data allows us to monitor the evolution of RideCheck's performance at an aggregate level (all heuristics and all interventions considered). This can help us steer efforts towards incremental gains, and give insight into how to balance user annoyance with being there when we're needed.

Research and Brand Tracking

If we opt for a per-user holdout design, we can use the historical data to support Brand Tracking and Research on RideCheck's evolving impact on Safety Sentiment, Safety Feature Awareness and other user behavior.

Discussion: Alternative metrics

Velocity metrics

Take a page from CommOps' "Average Handling Time" and measure the speed with which a Safety Incident is handled.

Velocity metrics include:

- Time to first report (how long a user takes to report an Incident to Uber)
- Time to normalization (how long it takes between an anomaly being detected and it being closed)

Tracking metrics*

Work with UXR and Marketing Research teams to track users confidence and self-reported behavior.

Tracking metrics include:

- CTS Score (how much the user believes Uber is Committed to Safety)
- Feature awareness (how many users know about Uber's Safety Tools, and what tools they know specifically)

Operational metrics†

Metrics that express the performance of RideCheck in reducing Safety Incidents frequency or severity.

Operational metrics include:

- Serious-by-Overall Incident Ratio (of all Incidents reported via RideCheck, what % are serious incidents)
- % Incidents Addressed (genuine incidents first reported via RideCheck)

Thank You | Questions?

03 Final Solution Design

General-Purpose RideCheck Holdout

Design

Detailed solution review with final Designs

Recap problems to be solved and constraints

Recap project goals

Show an overview of the chosen solution from a design perspective

Detail each part of the user journey, including happy paths and exceptions

! If your project does not include user-facing designs, please add the detailed business rules implemented and how users will be impacted

Take note of stakeholder comments and requests

04 XP / Launch Plan

TBD

Plan details

How are you building and launching

How are you measuring impact? (XP, survey, etc)

Detail: Primary Metrics, Secondary and Guardrails, with targets and goals.

How are you launching: XP, gradual?

Which markets are impacted?

Are there CommOps / Support needs?

Eng timelines

Take note of stakeholder comments and requests

05 XP Results / Rollout Plan
TBD

XP Results

XP Readout + Launch plans

How did it perform against success metrics?

Add quantitative and qualitative insights

Are there improvements to be made?

What is the final recommendation (launch/not launch)

Add global launch plan and timelines

Take note of stakeholder comments and requests