

# Technology and Strategies for Improving Survey Data Quality and Fieldwork Efficiency in the Global South

**J. Daniel Montalvo, Ph.D.**

Director of Survey Research Operations  
Center for Global Democracy and LAPOP Lab  
Vanderbilt University

**Guy Solomon**

Dooblo CEO

April 3 , 2025

# LAPOP Lab

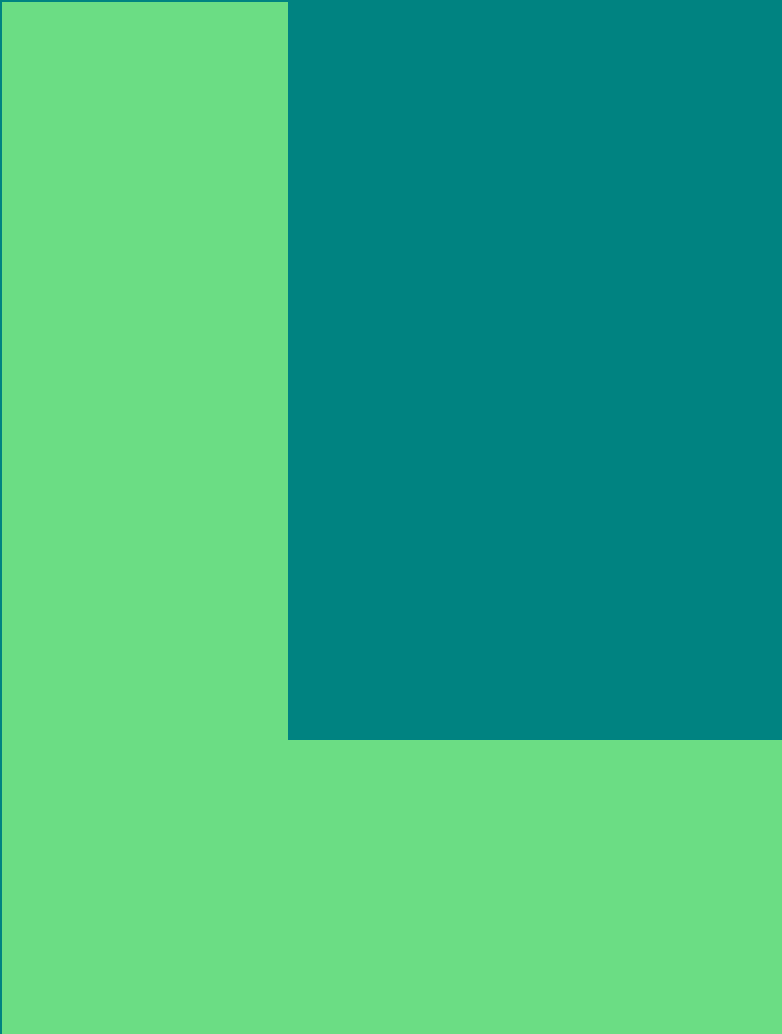
- ▶ **Leader in Public Opinion Research**  
Premier academic institution in international survey research with 30+ years of experience. Cornerstone of Vanderbilt's CGD
- ▶ **Gold-Standard Survey Methods**  
Unparalleled quality control in national surveys, impact evaluation studies, and reports on attitudes and experiences
- ▶ **The AmericasBarometer**  
Largest, most scientifically rigorous comparative survey of democratic values and behaviors covering 34 countries across the Americas and the Caribbean
- ▶ **Influence and Impact**  
Publishes high-quality academic studies and policy-relevant research. Extended network of academic institutions and NGOs across the Americas

# Dooblo

- ▶ **Leading Survey Software Provider**  
Specializes in multimode data collection, including offline capabilities, and fieldwork management solutions
- ▶ **Global Reach and Impact**  
Over 1,000 clients, 200 million completed interviews, and operations in 200+ countries
- ▶ **SurveyToGo Platform**  
Enhances data accuracy and efficiency with predictive quality control, real-time field management, and flexible data collection modes
- ▶ **Trusted Across Sectors**  
Used in market research, academic studies, and opinion research for reliable, real-time data collection

# Contents

1. Challenges to Survey Research in the Global South
2. Flagging Data Collection Errors: *FALCON*
3. Addressing Data Collection Challenges: *Quality Assessments*
4. Measuring the Impact of Data Quality Controls
5. How can Technology Help?



# 1

## Challenges to Survey Research in the Global South

---

# Best practices to reduce measurement errors

## LAPOP's approach to questionnaire design:

- ✓ Start with a set of **research topics** ahead of each round of the AmericasBarometer. Groups of experts design an initial set of questions that are intensively workshopped
- ✓ Formulation of a **draft core** questionnaire
- ✓ Each question in the draft is extensively pretested through iterative **cognitive interviewing** in 3-5 countries in the Americas
- ✓ The final core questionnaire is **country-customized**; inclusion of country-specific questions; additional cognitive interviews in each country
- ✓ Interviewer training, piloting of the electronic questionnaire, final edits

# Best practices to reduce representation errors

## During sampling, LAPOP:

- ✓ Collects the most recent **sampling frames**
- ✓ Draws standardized, **national probability samples**
  - **Stratified** by region, city size, and urban/rural location
  - Five selection **stages**, PPeS (cities or municipalities, census segments or electoral districts, blocks or rural areas, households, and respondents
  - Samples are **clustered** at the segment level
- ✓ Gathers maps and plan fieldwork logistics
- ✓ Trains interviewers and supervisors, and assign work areas

Following a meticulous design phase, the implementation phase is carried out in a terrain resembling this:

**Crime and insecurity:**

The LAC region has the highest murder rates in the World



**Political instability:**

Coups d'état, protests, riots, looting, and government surveillance are widespread in the region





And this:

**Environmental and infrastructure issues:**  
Difficult terrain, lack of resources



**Economic insecurity:** Trust issues, underpaid interviewers



# Many other fieldwork challenges can undermine data quality

- Dogs
- Illness
- Spouses
- Intimidation
- Distractions
- Escaped convict raids
- Low response rates, particularly differential non-response

# All these factors influence interviewer behavior

## ■ Representation errors:

- Conducting interviews in incorrect locations
- Not following the household selection protocol
- Tracking interview attempts improperly
- Conducting interviews with ineligible respondents
- Misreporting respondent characteristics (e.g., sex and age)
- Conducting the wrong number of interviews in a selected area

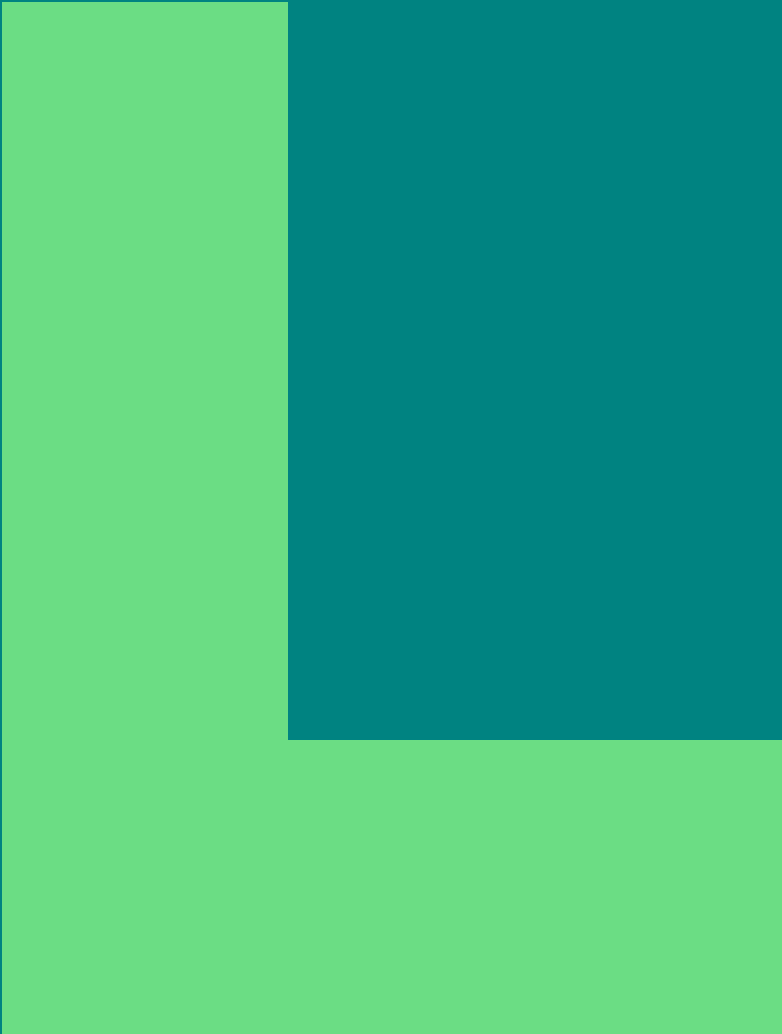
## ■ Measurement errors:

- Speeding through interviews
- Altering question wording
- Skipping or leading questions
- Fabricating responses
- Coding errors
- Outsourcing work to non-certified individuals
- Defining concepts
- Straightlining

# Case Study – LAPOP's Venezuela Survey (2016/17)

- Context of **acute civil/political unrest, economic deterioration, and insecurity**
  - Food scarcity
  - Efforts to recall the president
  - Frequent protests and looting
- **High levels of interviewer fraud and errors** (Castorena et al. 2023)
  - 650+ interviews out of 1,500 were canceled and replaced due to quality concerns
- Demonstrates how challenging conditions impact data collection reliability





# 2

## Flagging Data Collection Errors

*FALCON*

---

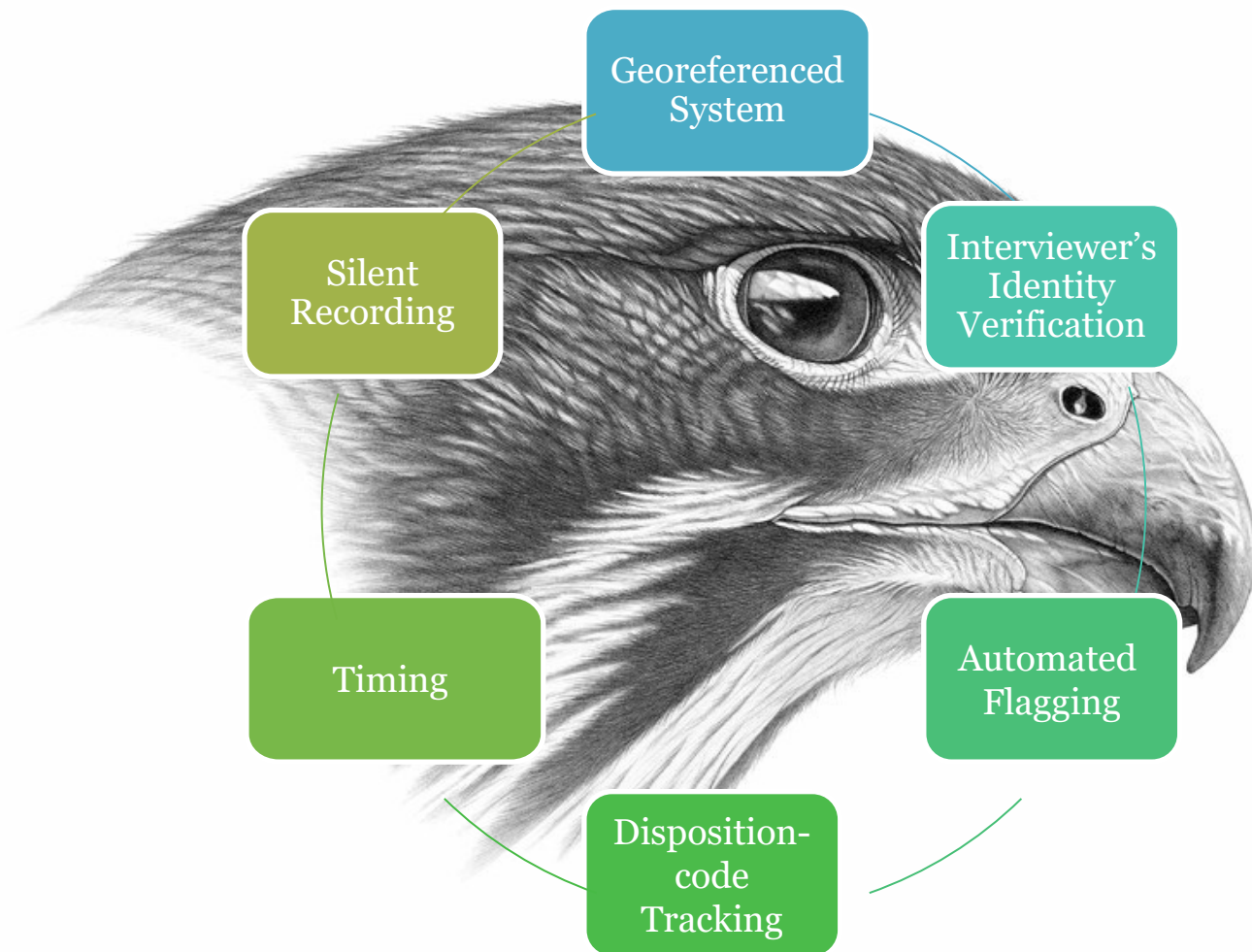
# LAPOP Lab addresses these challenges with three strategies:

1. Prevention
2. Detection
3. Correction

# Prevention through interviewer training

- **Training aims to minimize survey errors by:**
  - Explaining quality standards, ethical concerns, and project scope
  - Teaching best practices for interviews and technology use
  - Practicing questionnaire administration
  - Reviewing the selection of sampling points
- **Prevention strategies:**
  - Monitoring mechanisms (without revealing specific strategies)
  - Communicating penalties for non-compliance
- **Example:**
  - Interviewers are informed that some questions will be recorded, but not told which

# Detection of errors with LAPOP's *Fieldwork Algorithm for LAPOP Control Over Survey Operations and Norms* (FALCON)





# LAPOP's FALCON provides real-time oversight of interviews

- **FALCON** identifies quality control issues in data collection while fieldwork is in progress
- **It complements:**
  - Direct supervision (limited due to survey scale)
  - Post-hoc rejection of interviews (improves real-time monitoring)
- **Requirements:** Mobile devices, trained personnel, real-time internet connectivity, and more importantly, SurveyToGo
- Let's dive in...

# Georeferenced System

# Detecting sample distortions with FALCON

## Georeferenced System:

- Flags interviews conducted in incorrect locations
- Comprises 1) Geofencing, 2) Distance Audit Module, and 3) Route Tracking

1. **Geofencing:** Ensures interviews occur in the correct locations
2. **Distance Audit Module (DAM):** Computes the distance between the interview location and the geofence
3. **Route Tracking:** Tracks interviewer movement patterns



# Detecting sample distortions with FALCON

## Georeferenced System:

- Flags interviews conducted in incorrect locations
- Comprises 1) Geofencing, 2) Distance Audit Module, and 3) Route Tracking

1. **Geofencing:** Ensures interviews occur in the correct locations

2. **Distance Audit Module (DAM):** Computes the distance between the interview location and the geofence

3. **Route Tracking:** Tracks interviewer movement patterns

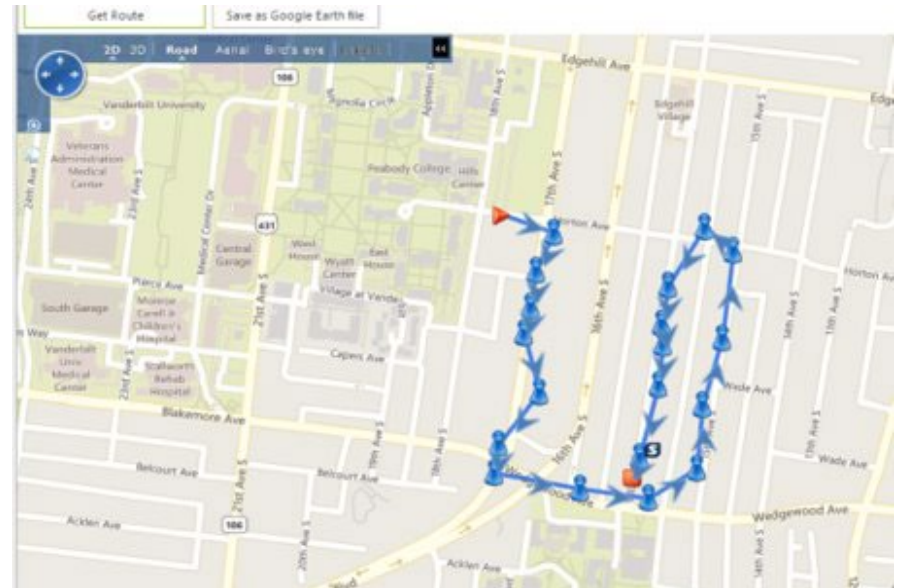
The screenshot displays the FALCON system interface. At the top, a red banner indicates a 'Flagged by Quality Control' status with the comment: 'By Script, Comment: La entrevista se hizo fuera del límite de la geo-cerca'. Below this, the 'Data' tab is active, showing various survey details such as Subject ID (41980984), Start/End times, and Surveyor (ECU Mariela Reibel). A 'Response to Question 115 for subject 41980984' dialog is open, showing a 'G DAM' entry with a value of 91.4267294293148, which is circled in red. The dialog also shows a list of questions and answers, and a 'Timestamp' of 11/26/2016 12:35:39 PM.

# Detecting sample distortions with FALCON

## Georeferenced System:

- Flags interviews conducted in incorrect locations
- Comprises 1) Geofencing, 2) Distance Audit Module, and 3) Route Tracking

1. **Geofencing:** Ensures interviews occur in the correct locations
2. **Distance Audit Module (DAM):** Computes the distance between the interview location and the geofence
3. **Route Tracking:** Tracks interviewer movement patterns



# Disposition- code Tracking

# Detecting disposition-code misreporting with FALCON

## ■ ART (Automated Response Tracker):

- Disposition codes are programmed at the outset of the electronic questionnaire
- Interviewers are instructed to register every attempt
- LAPOP uses this information to compute response rates



Attempt. Result of attempt:

- ☐ 1. Nobody at home
- ☐ 2. Unoccupied, abandoned, destroyed, or no house in this location. Business, organization or institution on site
- ☐ 3. Occupied, but rejects interview
- ☐ 4. Inappropriate respondent, such as deaf or blind, ill, intoxicated, threatening, etc.
- ☐ 5. Doesn't speak the language
- ☐ 6. Any other reason
- ☐ 7. Willing to proceed

9 of 304

LBDAY. Good morning/afternoon/evening, my name is {0} and I work for (Insert company name), an institution dedicated to knowing the opinion of people like you on different topics. We are currently conducting a survey on behalf of (Insert name of academic partner) and Vanderbilt University.

[DO NOT READ]: IF YOU HAVE COMPLETED A QUOTA, EDIT THE TEXT BELOW AS APPROPRIATE:

For this survey, I would like to speak with one person, who could be you, who is 18 years of age or older, who lives in this house, and who is present at this time.

Can I talk to that person?

[DO NOT READ]: IF THERE ARE TWO OR MORE ELIGIBLE PERSONS; THAT IS, THEY MEET THE SAME RESIDENCE, AGE AND GENDER CHARACTERISTICS, ASK:

Can I speak to the person who has had the most recent birthday?

- ☐ Yes
- ☐ No, decline the interview
- ☐ No, it does not meet the residency requirement
- ☐ No, it does not meet the quota

12 of 304

# Detecting disposition-code misreporting with FALCON

- ART (Automated Response Tracker):
  1. **Delayed reporting**, in which some estimate of the attempts before the successful one, are reported in one quick burst just as a successful interview is about to begin
  2. **Underreporting** (the main problem), in which only one attempt is reported (100% success rate), when in fact many had been made



# Silent Recording and Timing

# Detecting measurement errors with FALCON

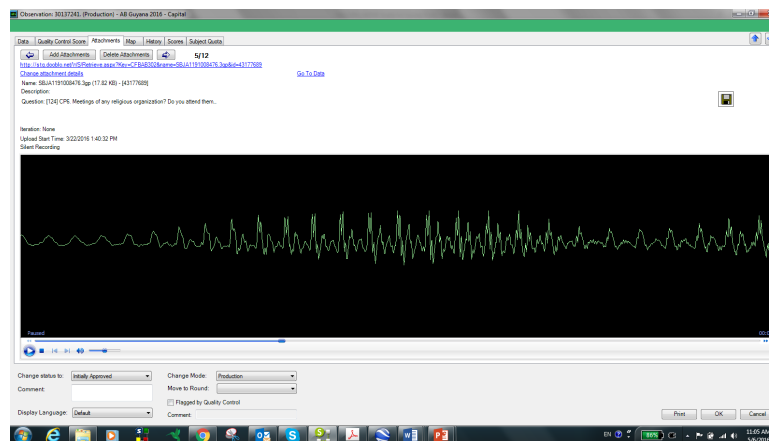
- **Speeding, wording changes, or data fabrication:**
  - **Silent Recording and Listening Feature:** Audits interview interactions
  - **Timing Feature:** Flags abnormally fast responses

Silent photo capture on entrance:

☐ Back camera

☐ Front camera

☒ Enable silent audio recording



# Interviewer's Identity Verification

# Detecting measurement errors with FALCON

- **Unauthorized interview outsourcing:**
  - **Silent photo capturing**, confirms certified interviewer presence
  - **Sworn statement**, formalizes interviewer identity

Iteration: None  
Upload Start Time: 12/9/2015 12:43:58 PM  
Silent Photo Capture  
Picture taken at: 2015-12-09 11:07:03  
Latitude: 10° 19' 49.95" Longitude: -83° 53' 9.96" [View in Google Maps...](#) [Copy coordinates \(decimal\)](#)  
☒ Fit to window

Silent photo capture on entrance:

☐ Back camera  
☒ Front camera  
☐ Enable silent audio recording

Change status to: Initially Approved Change Mode: Production  
Comment:  Move to Round:   
Display Language: Default ☐ Flagged by Quality Control  
Comment:



# Automated Flagging

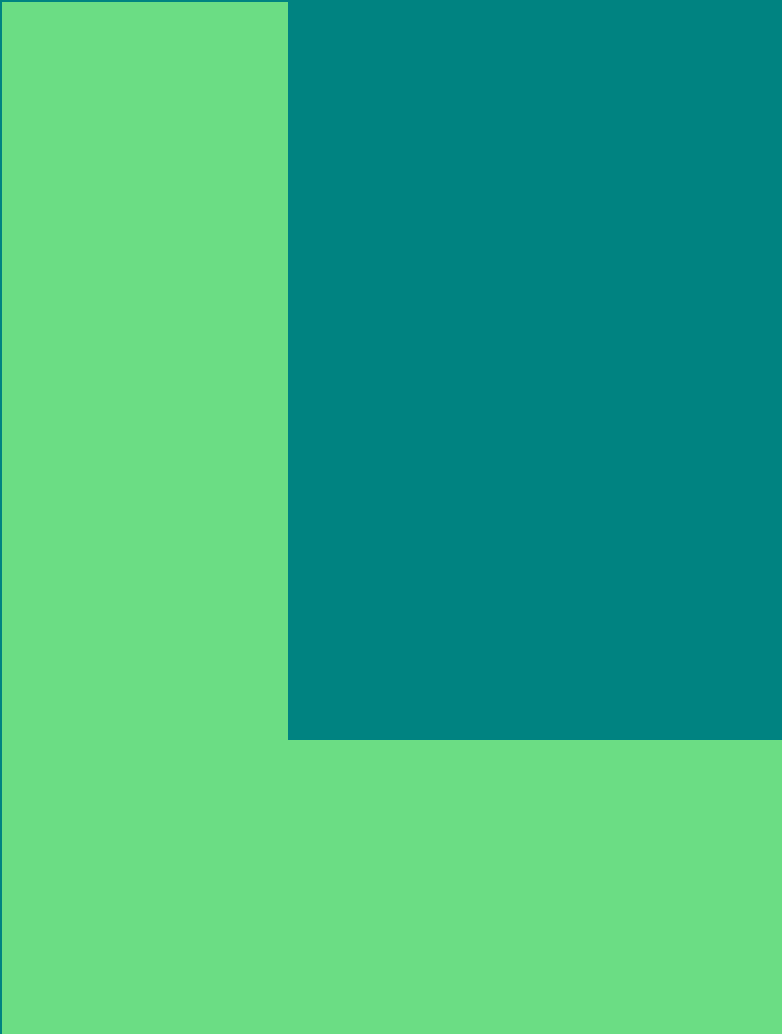
# Detecting measurement errors with FALCON

- **Automated interview flagging (system):**

- Duration flag
- Time-of-day flag
- (Absence of) GPS flag
- Fake GPS flag
- Straightlining flag
- Clock tampering flag
- Backtracking activity flag
- Attachment flags

# Detecting measurement errors with FALCON

- **Automated interview flagging (LAPOP)**
  - ✓ QDAM > 3 km OR no value in QDAM
  - ✓ Interview breakoffs
  - ✓ Timing of key questions
- **Total number of FALCON flags: 141**
  - ✓ However, 30 flags account for 93% of canceled interviews (see Cohen and Warner, 2020)



# 3

Addressing Data  
Collection Challenges

*Quality Assessments*

---



# Dedicated auditing teams study the flags of each interview

- **3-Tier Quality Control System**

## **LEVEL 1:**

Survey firm audits 100% of the interviews

## **LEVEL 2:**

Contracted firm and/or LAPOP audit 33-40% of audited interviews in L1

## **LEVEL 3:**

LAPOP audits 10% of audited interviews in L2, if L2 is outsourced



# Quality control protocol

## Auditors review the following:

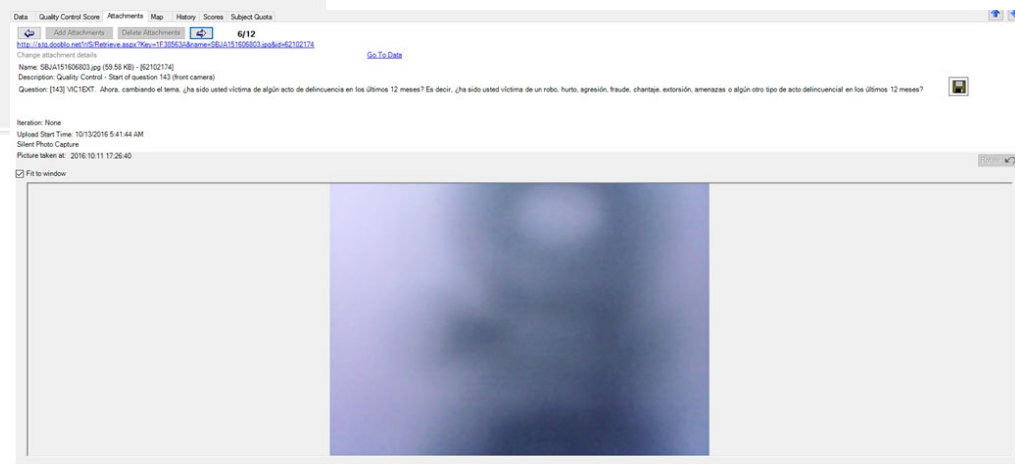
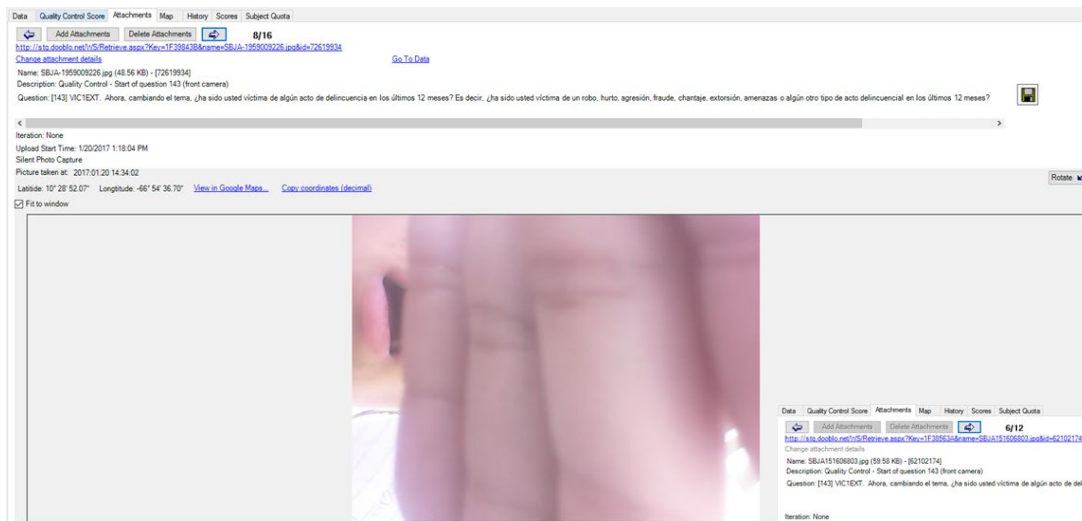
- Net duration of each interview
- Electronic questionnaire version
- Geotags, satellite imagery, and interview location
- Audio recordings
- Silent photo captures
- Automated flags

## Exhibit A: Outsourcing



# Exhibit B: Outsourcing

## Outsourced interviewers covering cameras to avoid being identified

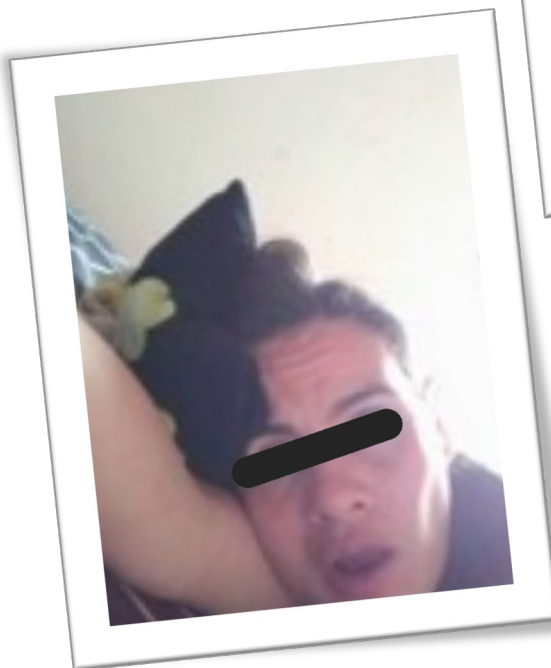


## Exhibit C: Data fabrication



## Exhibit D: Data fabrication

Silent Photo Capturing: Verifies in-person presence in designated areas



# Exhibit E: Data fabrication

## Interviewers rushing through the survey

Flag interviews that took:

☒ Less than 40 minutes to complete

☒ More than 120 minutes to complete

☐ Flag interviews if no GPS was captured

☐ Flag if a fake location was discovered during the interview



NET INTERVIEW  
DURATION



**Flagged by Quality Control:**

By Duration, Comment: Interview duration too short. Required: 30 min, Actual: 7:10 min

Read Only

Data Quality Control Score Attachments Map History Scores Subject Quota

|                              |                                |                         |                                                     |
|------------------------------|--------------------------------|-------------------------|-----------------------------------------------------|
| Status: Canceled             | Start: 2/17/2016 9:49:34 AM    | Scenario                | Total Duration: 00:00:07:12                         |
| Subject ID: 28948122         | End: 2/17/2016 9:56:46 AM      | Surveyor: Desma DaSilva | Net Duration: 00:00:07:10                           |
| Round:                       | Uploaded: 2/17/2016 9:56:19 AM | Task ID:                |                                                     |
| Subject UTC: (UTC-04:00)     | External:                      | Filtered on: N/A        | Collected using survey version: <a href="#">125</a> |
| Index: 966878-2              | Device Index: D966878-2        | Data:                   | Latest survey version: 127                          |
| Version: 1.32.218.2 AN       | Name: Region 2/ Anna Regina    |                         | Device ID: 613                                      |
| Device IMEI: 863448020034218 | Location:                      |                         |                                                     |

# Auditors report issues in the Quality Assurance Chapter (QuAC)

## ■ QuAC:

- The system computes an automatic penalty score
- When the penalty is *20 points or greater*, the interview gets canceled
- Interviews with a score lower than 20 are approved to be part of the dataset

## ■ Type of offenses:

- Setting the device to airplane mode is a serious offense, so the interview is automatically penalized with 100 points
- Reading questions too fast is a minor offense, however, interviewers receive a warning

Response to Question 683 for subject 42753546

|                                                                                                                                               |                                                       |
|-----------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------|
| Type                                                                                                                                          | Multiple Selection                                    |
| Question:                                                                                                                                     | QAC. La entrevista presenta los siguientes problemas: |
| <input type="checkbox"/> FRAUDE. EVIDENCIA DE FRAUDE                                                                                          |                                                       |
| <input type="checkbox"/> El encuestador se hace la entrevista a sí mismo                                                                      |                                                       |
| <input type="checkbox"/> Existen adjuntos de audio, pero no se oye a nadie hablar o sólo se oye al entrevistador                              |                                                       |
| <input type="checkbox"/> El entrevistador pone el dispositivo en "modo avión"                                                                 |                                                       |
| <input type="checkbox"/> El entrevistador apaga el GPS del dispositivo                                                                        |                                                       |
| <input type="checkbox"/> El entrevistador tapa la cámara o la deshabilita para evitar las capturas de fotos                                   |                                                       |
| <input type="checkbox"/> El encuestador entrevista a otro encuestador                                                                         |                                                       |
| <input type="checkbox"/> El encuestador entrevista a alguien que conoce                                                                       |                                                       |
| <input type="checkbox"/> IID. PROBLEMAS CON LA IDENTIDAD DEL ENCUESTADOR                                                                      |                                                       |
| <input type="checkbox"/> Las fotografías no corresponden con las del encuestador o hay inconsistencias en las fotografías                     |                                                       |
| <input type="checkbox"/> La voz de los audios no corresponden a la voz del encuestador                                                        |                                                       |
| <input type="checkbox"/> INC. ENTREVISTA INCOMPLETA                                                                                           |                                                       |
| <input type="checkbox"/> Se agotaron los intentos                                                                                             |                                                       |
| <input type="checkbox"/> El encuestado no permite terminar la entrevista y la abandona                                                        |                                                       |
| <input type="checkbox"/> El encuestador decide terminar la entrevista por cualquier otra razón                                                |                                                       |
| <input type="checkbox"/> NETGEO. PROBLEMAS DE UBICACIÓN Y/O DE DURACIÓN DE LA ENTREVISTA                                                      |                                                       |
| <input type="checkbox"/> La entrevista se llevó a cabo en un lugar equivocado (Centros comerciales, parques, gasolineras, universidades, etc) |                                                       |
| <input type="checkbox"/> La entrevista inició y terminó en lugares diferentes                                                                 |                                                       |
| <input type="checkbox"/> La duración neta de la entrevista fue de menos de 25 minutos o de más de 2 horas                                     |                                                       |
| <input type="checkbox"/> MISREAD1. PROBLEMAS EN LA LECTURA DE LA HOJA DE INFORMACIÓN DEL ESTUDIO (EX-CARTA DE CONSENTIMIENTO INFORMADO)       |                                                       |
| <input type="checkbox"/> No lee nada de la hoja de información del estudio                                                                    |                                                       |
| <input type="checkbox"/> Lee solamente partes de la hoja de información del estudio                                                           |                                                       |
| <input type="checkbox"/> Cambia palabras de la hoja de información del estudio                                                                |                                                       |
| <input type="checkbox"/> Cambia el tiempo de duración aproximada de la entrevista                                                             |                                                       |
| <input type="checkbox"/> Sigue con la entrevista, a pesar de que el informante dice que no desea participar                                   |                                                       |
| <input type="checkbox"/> MISREAD2. LEE PREGUNTAS DEL CUESTIONARIO DE MANERA INCOMPLETA O INCORRECTA                                           |                                                       |
| Answer                                                                                                                                        |                                                       |
| Other                                                                                                                                         |                                                       |
| Timestamp:                                                                                                                                    |                                                       |
| <input type="checkbox"/> Null Response <input type="checkbox"/> Do Not Score <input type="checkbox"/> Do Not report comment                   |                                                       |
| <input type="checkbox"/> Hide Other Spec from reports                                                                                         |                                                       |



# Correcting errors

- **Real-time feedback for better fieldwork:**
  - Allows for **course correction during data collection** rather than post-hoc rejections
  - Improves survey accuracy, efficiency, and integrity
  - **Strengthens the reliability** of LAPOP Lab's research across the Americas

----- Mensaje original -----

De:

Fecha: 08/17/2016 19:28 (GMT-05:00)

A:

CC:

Asunto: Problemas con la lectura del consentimiento informado y B0

[REDACTED]

Nuestro equipo de auditoría de nivel 3 sigue encontrando problemas recurrentes con algunos encuestadores a la hora de leer la carta de consentimiento informado, y el texto introductorio a la pregunta B0.

Por ejemplo, [REDACTED] no está leyendo varias frases de la carta de consentimiento informado. En la entrevista 35440677 no menciona nada sobre la grabación de la entrevista, o sobre dejar la carta con el encuestado. También se salta la parte que dice "desea participar". En la encuesta 35506741 no dice nada con respecto a que las respuestas se mantendrán confidenciales y anónimas. Así mismo, nuestro equipo ha identificado que la entrevistadora comete varios errores de lectura en ambas entrevistas.

Por otra parte, la entrevistadora [REDACTED] no permite a los encuestados responder a la pregunta B0 en las siguientes entrevistas:

35594373

35559248

35521687

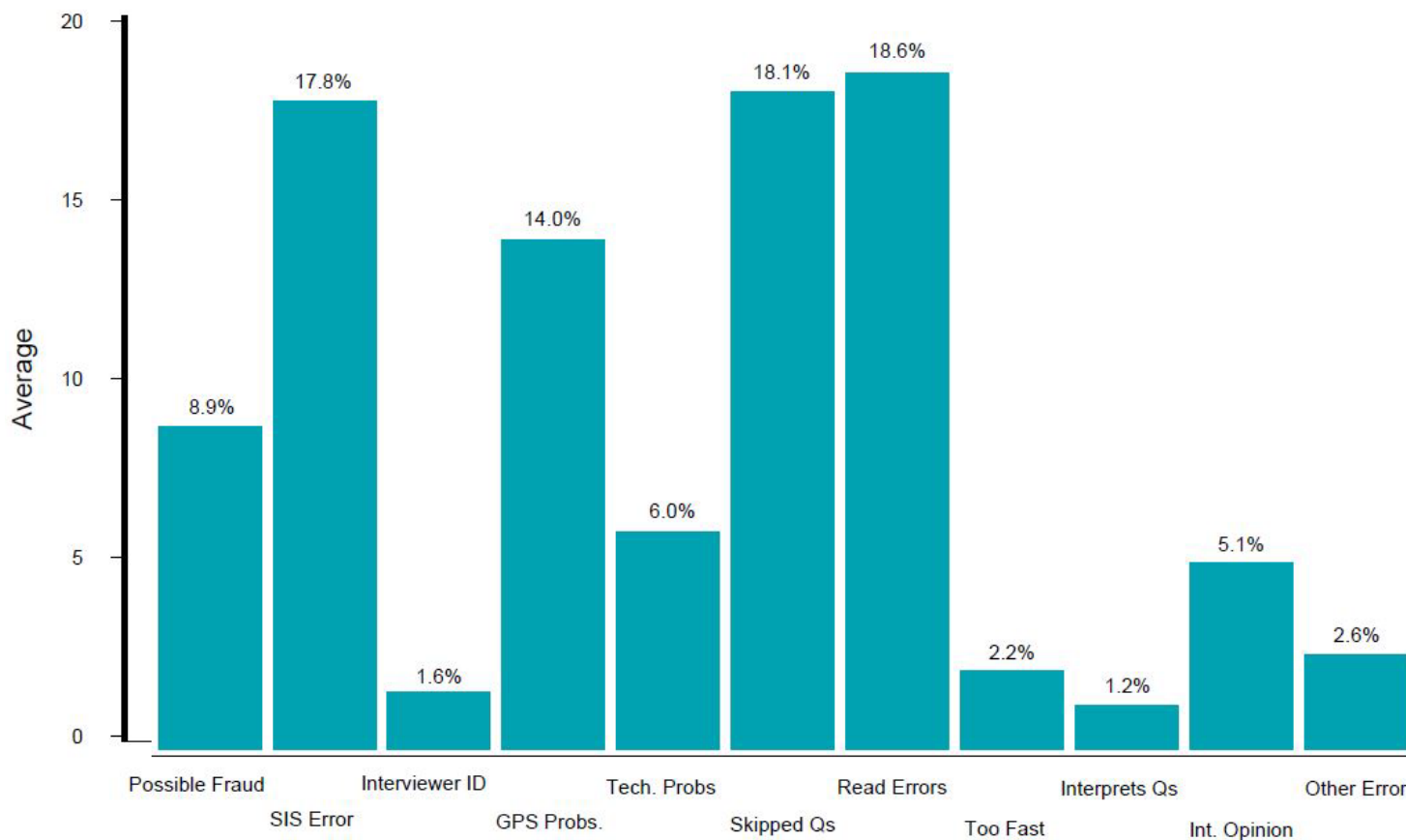


# 4

## Measuring the Impact of Data Quality Controls

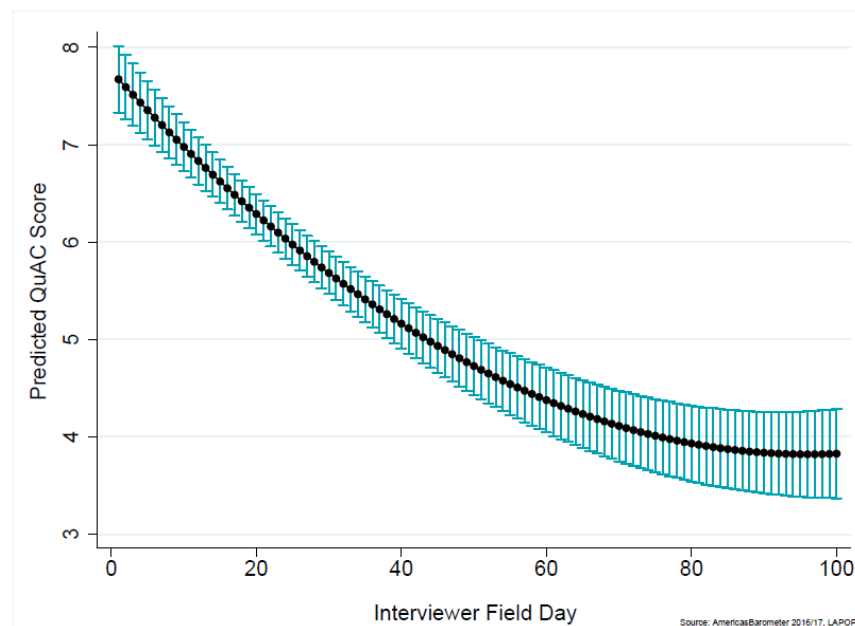
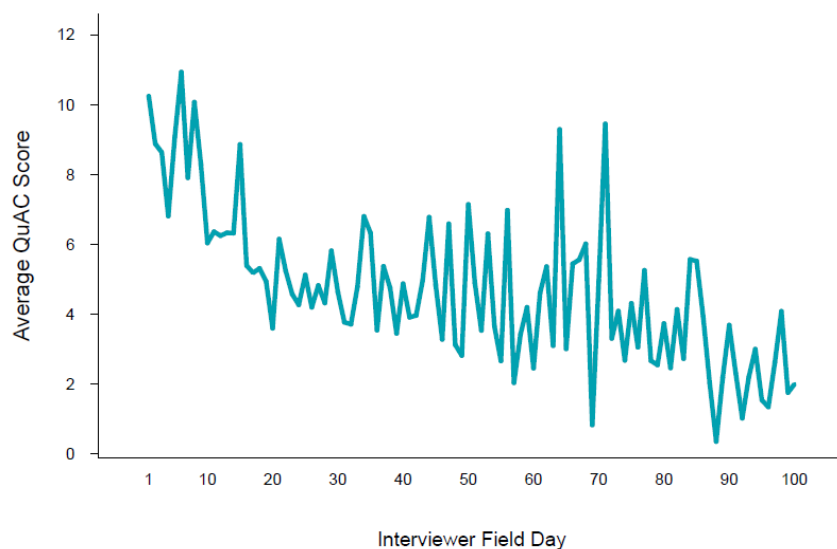
---

# Most errors among cancelled interviews pertain to reading errors

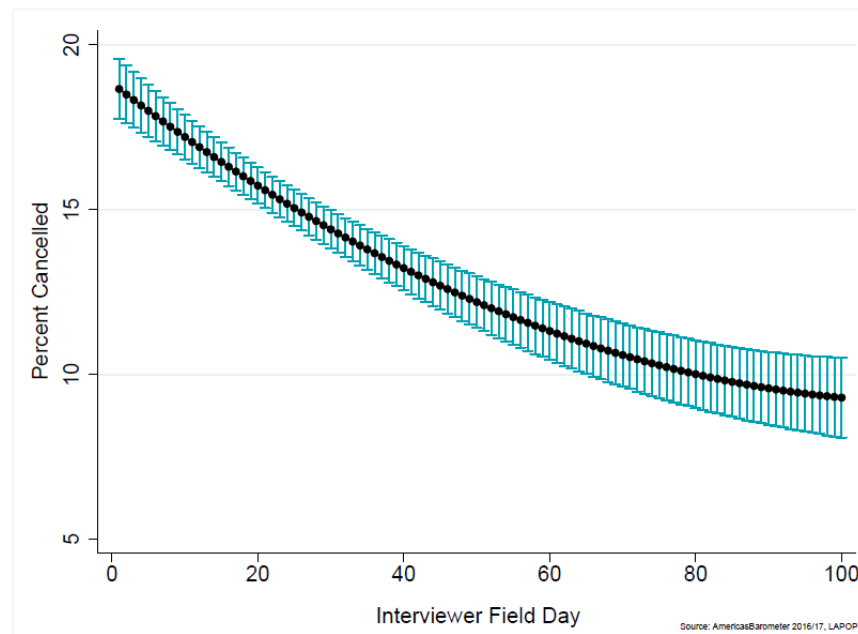
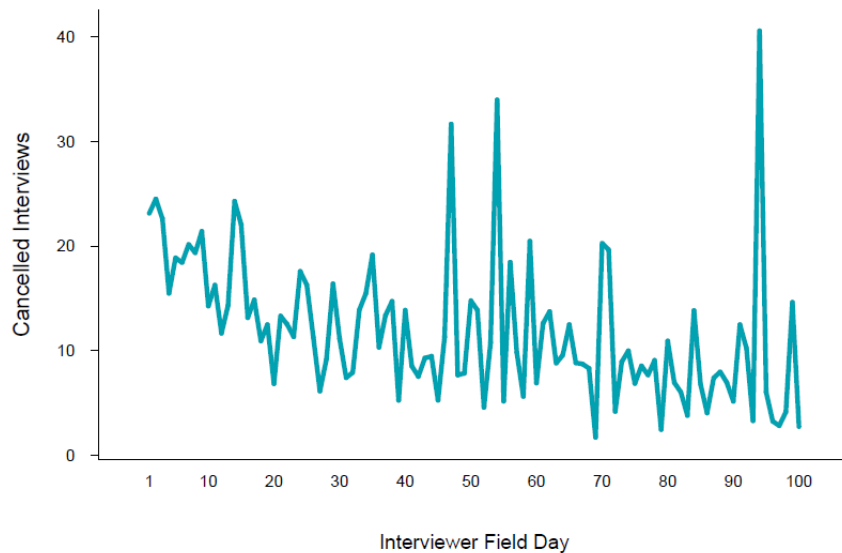


Source: AmericasBarometer 2016/17

# Penalty scores decrease as fieldwork progresses



# Errors consistent with fraud decrease with time





# 5

How Can Technology  
Help?

---

# More Realtime Visibility into Quality and Productivity

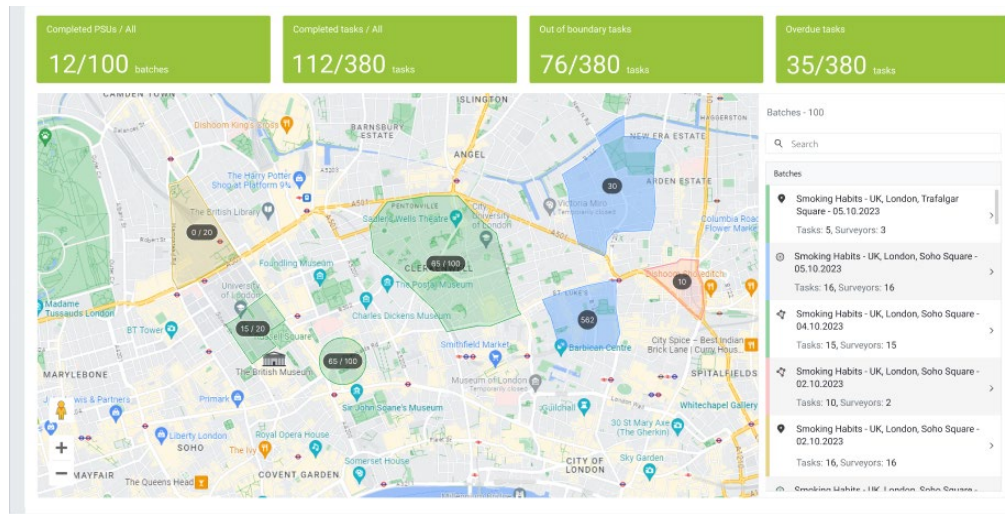
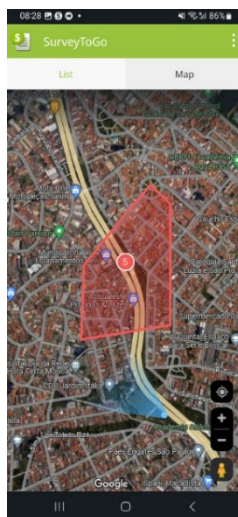


Realtime quality monitoring

- Daily
- Per interviewer

Location based

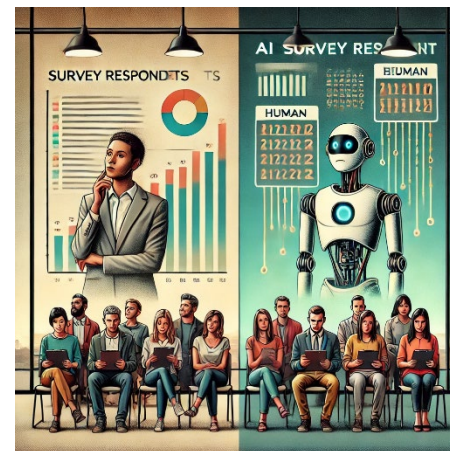
- Tasks
- Quality
- Quotas



# How Can AI Help Increasing Quality and Productivity?

**What should we be thinking of when engaging with AI?**

- Do I trust AI to have a leading role in my project?
- Am I compromising data quality?



**Until fully trusted, AI can help on both quality and productivity  
but must be closely monitored to prevent quality leaks**

## **Productivity**

- Transcribing
- Auto code open ended questions
- OCR
- Image analysis

## **Quality**

- Validate question asked correctly
- Consistency checks
- Quality score
- Offer follow up questions



# Key Lessons from the Webinar

## ✓ Survey Data Quality Faces Significant Challenges

- Sociopolitical and economic conditions can distort survey research through interviewer errors and fraud
- Common issues include sample distortion, measurement errors, and data fabrication

## ✓ Leveraging SurveyToGo Data Collection, FALCON Provides a Proactive Solution

- Combines real-time monitoring with post-hoc auditing to enhance data reliability
- Uses advanced tools like geo-fencing, silent recording, and response tracking to detect and prevent errors

## ✓ Data Quality Improves Over Time with Effective Monitoring

- Evidence shows a reduction in canceled interviews and errors as fieldwork progresses
- Quality control mechanisms work consistently across different countries and time periods

## ✓ Quality Control Measures Are Scalable and Adaptable

- The FALCON framework provides a model for improving survey data integrity globally
- Lessons from this approach can be applied to other research contexts facing similar challenges

# Technology and Strategies for Improving Survey Data Quality and Fieldwork Efficiency in the Global South

**J. Daniel Montalvo, Ph.D.**

Director of Survey Research Operations  
Center for Global Democracy and LAPOP Lab  
Vanderbilt University

**Guy Solomon**

Dooblo CEO

April 3 , 2025