



INSTITUTE FOR SOCIAL RESEARCH • SURVEY RESEARCH CENTER  
SURVEY RESEARCH OPERATIONS  
UNIVERSITY OF MICHIGAN

# Using Paradata to Investigate an Unexpected Production Outcome & Associated Interviewer Behaviors

Shonda Kruger Ndiaye

CSDI 2014



# Acknowledgements

- Joel Devonshire—ADT OLAP Cube Design
- Minako Edgars—Analysis support
  
- Major Funding: NSF
  
- Additional funding: NIA, NICHD, Office of the Assistant Secretary for Planning and Evaluation of the US DHHS, Economic Research Service of the United States Department of Agriculture, HUD, DOL, the Center on Philanthropy at the Indiana University-Purdue University



## Panel Study of Income Dynamics (PSID)

- A national, longitudinal household panel study
- Started in 1968 with 5,000 families
- 38 waves of data collection
- Extensive content: housing, earnings, income, employment, wealth, health, philanthropy, and more





# 2013 Wave

- 9,107 interviews (2013); 92% RR
- 43 wk data collection—13 wk period of Lab + Field work
- All experienced interviewers—Lab interviewers new to PSID
  - 103 Field (~60% PSID-experienced)
  - 12 Lab (0% PSID-experienced)
- Sample assignment to group not random, but not believed to differ systematically



# Interview Length on the PSID

- Closely monitored:
- Known to vary greatly by:
  - Interview type
  - Interviewer PSID experience
  - Also by:
    - Resistance
    - Use of Cell Phone
    - # of Suspend
    - Interview Consent to record status, etc.

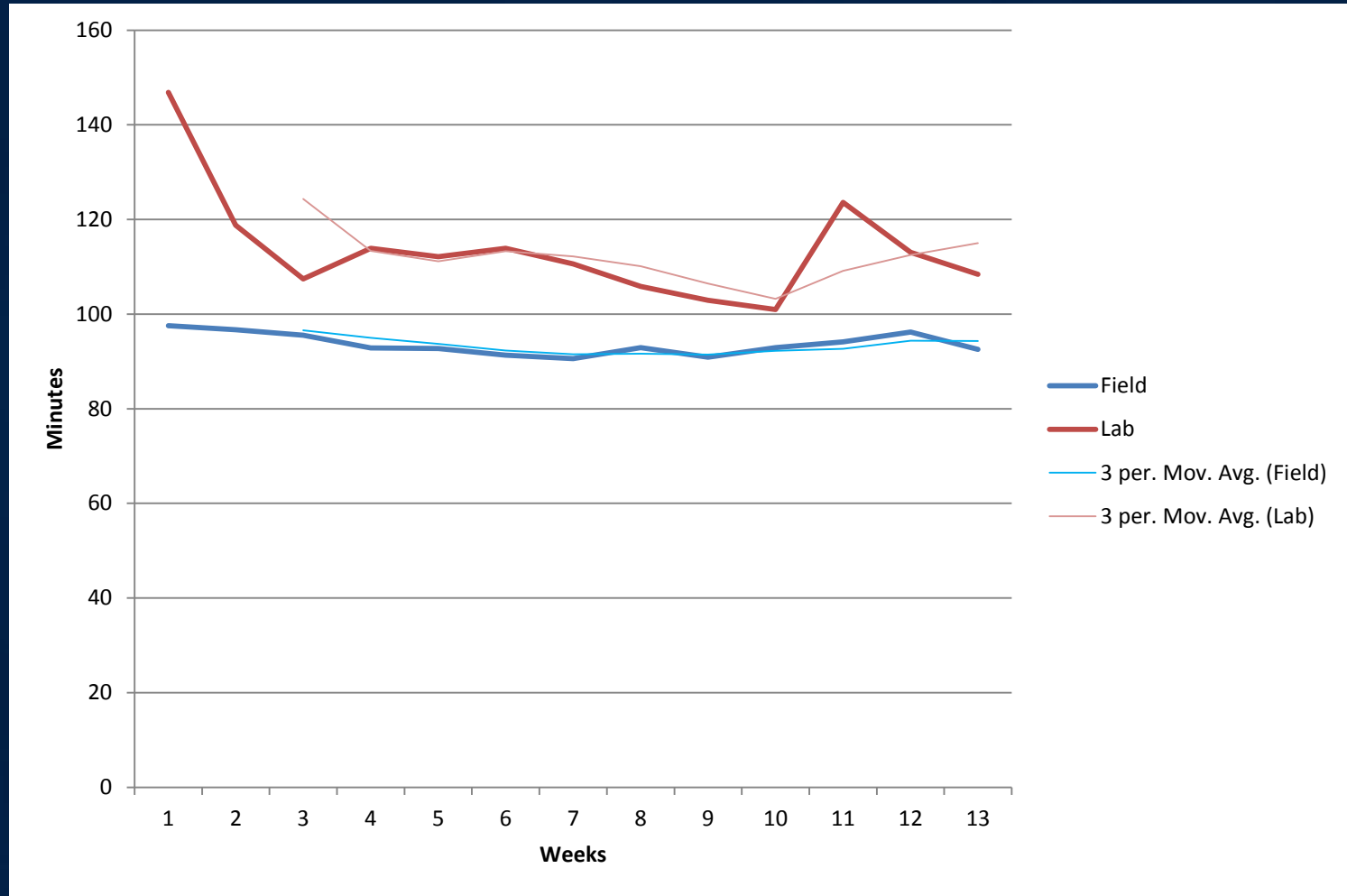


## Initial Observations on Length Differences

- Length differences apparent as soon as interviews came in.
- Wk 4
  - Field Avg = 83.10
  - Lab Avg = 97.78 (~18% longer)
- No real benchmark for “true avg”
  - Substantial content changes 2011 – 2013
  - 2013 Pretesting w/convenience sample
  - Major q’naire modifications post-pretest



# Difference persisted over time:





## Final Iw Length for 13 wks of Lab + Field

- Field Iws = 5872
- Lab Iws = 341
  
- Field Avg = 93.41 (StdDev: 31.71)
- Lab Avg = 111.16 (StdDev: 33.38)
  
- Lab Iws—19% longer





# Areas of Investigation

1. Inherent Differences in Sample/Interviews
2. Interviewer Behavioral Differences
3. Technical Differences



# Audit Trails/Blaise ADT Files

## Definition:

- Automatically generated files with records for each visit to a survey item or Web page:
  - # of visit to the item (page)
  - Start and end time
  - Keystrokes, function keys, mouse clicks
  - Final response/value

# Blaise Audit Trail with Keystrokes

## Timestamps

Hours:Minutes:Seconds:Thousands of a second

Case ID in Blaise database

```
"1/17/2012 9:00:06:304 AM", "Enter Form:1", "Key:3975053020 " ← Sample ID
"1/17/2012 9:00:06:304 AM", "Metafile name:C:\blproj\HRS2012\work\HRS12.bmi"
"1/17/2012 9:00:06:304 AM", "Metafile timestamp:Friday, January 06, 2012 1:08:04 PM"
"1/17/2012 9:00:06:304 AM", "WinUserName:14554015" ← Interviewer ID
"1/17/2012 9:00:06:304 AM", "DictionaryVersionInfo:0.0.0.0"
...
"1/17/2012 9:00:12:702 AM", "Enter Field:SecA.StartInterview.A007TRALive_A", "Status:Normal", "Value:"
"1/17/2012 9:00:13:965 AM", "(KEY:)1[ENTR]" ← Time of first keystroke
"1/17/2012 9:00:14:276 AM", "Action:Store Field Data", "Field:SecA.StartInterview.A007TRALive_A"
"1/17/2012 9:00:14:328 AM", "Leave Field:SecA.StartInterview.A007TRALive_A", "Cause:Next Field",
  "Status:Normal", "Value:1"
...
"1/17/2012 9:02:51:681 AM", "Enter Field:SecJ.WORKSTATUS.J005MCurrEmpStatus[1]", "Status:Normal", "Value:"
"1/17/2012 9:02:55:971 AM", "(KEY:)15[BACK][BACK]5[ENTR]"
"1/17/2012 9:03:03:209 AM", "Action:Store Field Data", "Field:SecJ.WORKSTATUS.J005MCurrEmpStatus[1]"
"1/17/2012 9:03:03:256 AM", "Leave Field:SecJ.WORKSTATUS.J005MCurrEmpStatus[1]", "Cause:Next
  Field", "Status:Normal", "Value:5"
...
"1/17/2012 9:13:24:923 AM", "Enter Field:IWComplete", "Status:Normal", "Value:"
"1/17/2012 9:13:28:480 AM", "(KEY:)1[ENTR]"
"1/17/2012 9:13:29:650 AM", "Action:Store Field Data", "Field:IWComplete"
"1/17/2012 9:13:29:728 AM", "Leave Field:IWComplete", "Cause:Next Field", "Status:Normal", "Value:1"
"1/17/2012 9:13:30:056 AM", "Leave Field:IWComplete", "Cause:Exit", "Status:Normal", "Value:1"
"1/17/2012 9:13:30:056 AM", "Leave Form:1", "Key:3975053020 "
```

Start IW  
Audit trail file  
information

Question

Question  
with changed  
answer

Complete IW



# Trace movement...

```
"12/17/2010 1:36:14 :391PM", "Enter Field: Section_A.A45a", "Status:Normal", "Value:"  
"12/17/2010 1:36:15 :796PM", "(KEY:)[SHFT]This is a test[ENTR]"  
"12/17/2010 1:36:22 :295PM", "Action: Store Field Data", "Field: Section_A.A45a"  
"12/17/2010 1:36:22 :530PM", "Leave Field:Section_A.A45a", "Cause:Next  
Field", "Status:Normal", "Value:This is a test"  
"12/17/2010 1:36:22 :540PM", "Enter Field:Section_A.A45b", "Status:Normal", "Value:"  
"12/17/2010 1:36:26 :578PM", "(KEY:)[UP]"  
"12/17/2010 1:36:26 :591PM", "Leave Field:Section_A.A45b", "Cause:Move  
Up", "Status:Normal", "Value:"  
"12/17/2010 1:36:26 :607PM", "Enter Field:Section_A.A45a", "Status:Normal", "Value:This is a test"  
"12/17/2010 1:36:31 :001PM", "(KEY:)this is another test, changing te[BACK]ext.[ENTR]"  
"12/17/2010 1:36:49 :948PM", "Action:Store Field Data", "Field:Section_A.A45a"  
"12/17/2010 1:36:49 :985PM", "Leave Field:Section_A.A45a", "Cause:Next  
Field", "Status:Normal", "Value:this is another test, changing text."
```



# Disadvantages of raw ADTs

- Running even simple queries takes a lot of time and system resources
- Requires query programming skills
- To be really useful, ADT data should be pre-joined with other paradata
- Thus, we need a fast and flexible query tool that empowers the analyst
  - Must be relatively easy to use, and allow both data exploration and reporting capability



# SRO's ADT OLAP Cube

- Online Analytical Processing (OLAP)
  - Based on multidimensional database
  - Data pre-aggregated at regular intervals
- Uses the ADT database as main source
  - But can supplement with any source data (e.g., sample management system data)
- Data are piped into a new interim database (data warehouse), then cube is built.



# 1. Inherent Differences in Sample/Interview – Lab vs. Field



## Interview Type

- Recontacts and Splitoffs, longer than Reinterviews
- Interview Length differences existed within groups:

Type	Overall	Field	Lab	% Longer Among Lab lws
Recontact	113.3	111.65	236.16	111.52%
Reinterview	96.84	92.73	110.51	19.17%
Split from Recon	114.99	111.31		
Splitoff	113.19	109.34	134.75	23.24%





# Interview Type (2)

- Interview type sample distribution:

Type	Field	Lab
Recontact	1.4%	0.9%
Reinterview	95.0%	98.0%
Splitoff	3.5%	1.1%

- Distribution favored shorter iws in the Lab



# Questions in Interviews

- Distinct Field Count (from ADT)
  - Field Avg/lw = 507
  - Lab Avg/lw = 515 (1.62% more)
  
- Increasing Field Avg lw Length 1.62%
  - Adjusted Field lw Length = 94.92
  - Lab Length still ~17% longer



## 2. Interviewer Behavioral Differences



# “Best Behavior” Hypothesis

- Same recording/evaluation protocol across groups, however . . .
  - Lab interviewers are subject to monitoring at any time
  - Field interviewers know which interviews are subject to recording/evaluation



## “Best Behavior” Hypothesis (2)

- Avg Iw Length of Field iws consented to record = 101.01
- Avg Iw Length of SSL iws = 111.16
- Lab Iws 10% longer



# Use of Notes/Remarks

- Question-level notes
- Some needed and systematically reviewed; others not
- Avg Remarks/lw:
  - Field—5.7
  - Lab—10.6
- Adjusted to equalize distinct field count—Lab entered ~83% more Notes

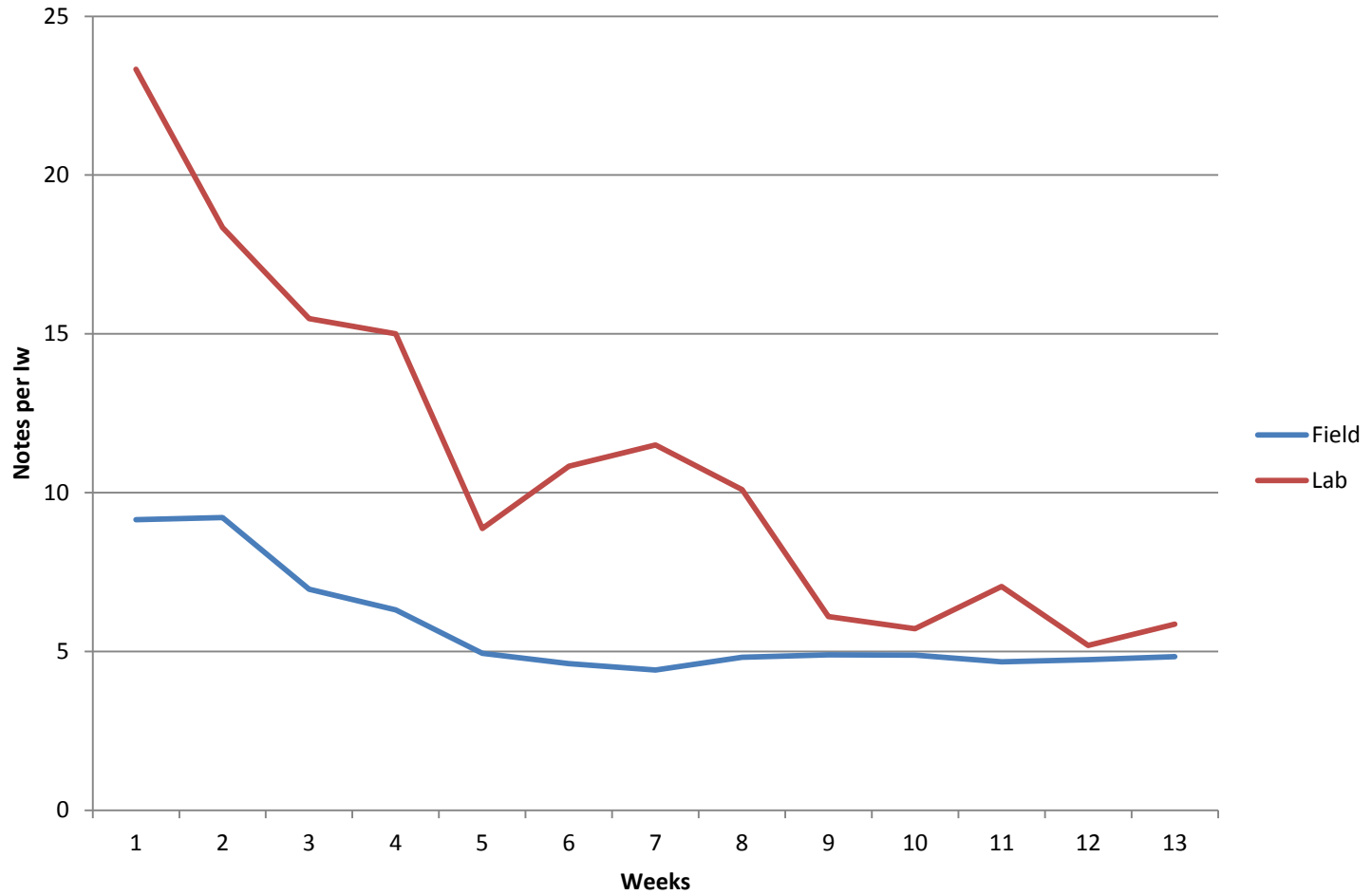


## Use of Notes/Remarks (2)

- Could calculate time spent typing notes, however, note-taking is also a function of R behavior
- Intervened mid-data collection to clarify use of Notes



## Notes per interview







# Use of Question-Level Help

- Question-by-Question Objectives (Q by Qs)
- Avg Times Q by Qs Used/lw
  - Field—2.1
  - Lab—4.9
- Adjusted to equalize distinct field count—Lab used Q by Qs 130%



# Backups

- Interviewers back up to view/correct previously-entered data
- Diligence or challenges with interview navigation?
- Avg backups/iw
  - Field—35.87
  - Lab—51.18
- Adjusted to equalize distinct field count—Lab backed up ~40% more frequently



# Effect of Experience

- 2011 interviewers (all Field):
  - New to PSID Iw Avg Length = ~4% longer than Experienced on PSID
- 2013 Interviewers New to PSID:
  - Field Iw Avg Length = 96.21
  - Lab Iw Avg Length = 111.16
  - Lab Iws ~16% longer



# Experience—Notes, Q by Qs, Backups

	Notes/lw	Q by Qs/lw	Backups/lw
<b>Field</b>	5.72	2.06	35.87
2011 Experience	5.06	1.97	32.43
Experience pre-2011	4.54	1.88	28.90
No PSID Experience	6.75	2.09	40.63
<b>Lab</b>	10.59	4.88	51.18
No PSID Experience	10.59	4.88	51.18
<b>Grand Total</b>	5.99	2.21	36.71



# Skipping or Shortening Q Reading

- Could have timed reading of shortest version of certain Qs
- Simpler: % of Questions administered in <2 seconds.
- Complicated due to Q format
  - E.g. date questions with mo/day/year on separate screens



## Skipping or Shortening Q Reading (2)

- % of Questions administered in <2 seconds:
  - Field—4.70%
  - Lab—3.12%
- ~33% more in the **Field**
- Due to greater variance in questions reached among Field iws? (e.g. more mm/dd/yyyy questions)
- Due to real differences in administration?



## Skipping or Shortening Q Reading (3)

- Intro to Philanthropy Section
  - Field 84.1 sec
  - Lab 60.2 sec
- Gateway into subsequent Philanthropy questions
  - Field 46.25
  - Lab 52.79
- No evidence so far of skipping or shortening?



# 3.

## Technical Differences





# Computer Processing Speed

- Lab used desktops; Field used laptops. Could there have been important differences?
- Between-field lengths from ADT
  - x Distinct field count
- Mins of iw length due to processing speed (w/distinct field count equalized)
  - Field—1.4
  - Lab—1.7
- Only  $\sim 1/3$  min attributable to processing speed differences



# Challenges

- Fully understanding meaning of audit trail variables (e.g. unit of measure).
- Properly aggregating audit trail data for each analysis.
- Reaching actionable conclusions in time to inform production.
- If actions driving length adhere to interviewers' training, what to do?



# Suggestions for Next Wave

- Pre-build more interview-length monitoring tools
  - Monitor by more factors up-front
  - Monitor at iwer, question level
- Consider refining interviewer training and/or questionnaire design
  - Find the questions where Lab length is significantly longer. Devote additional training?
  - Legitimize appropriate streamlining