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

Using Paradata to Monitor Interviewers' Behavior: A Case Study from a National Survey in the Kingdom of Saudi Arabia

Yu-chieh (Jay) Lin
Jennifer Kelley
Zeina N. Mneimneh
Beth-Ellen Pennell

March 28, 2014
The International Workshop on Comparative Survey Design and Implementation
Bethesda, Maryland

Outline

- World Mental Health Initiative
- Saudi National Mental Health Survey
- Quality Control
  - Analytical Reporting Process Chart
  - Measures
  - Dashboard + Cube = Dynamic Dashboard
- Demonstration
- Lessons Learned

World Mental Health Initiative

- A coordinated series of community psychiatric epidemiological studies carried out in countries throughout the world, many of them never having previous information about the prevalence, treatment, or societal burden of mental disorders.
    - Cross-sectional.
    - Retrospective lifetime assessment.
    - Fully-structured interview only validated in a small number of countries.
    - Limited to the household population.
- Implementation is facilitated by access to a WMH Data Collection Coordination Centre that provides key infrastructure support and consultation from experts in survey research.
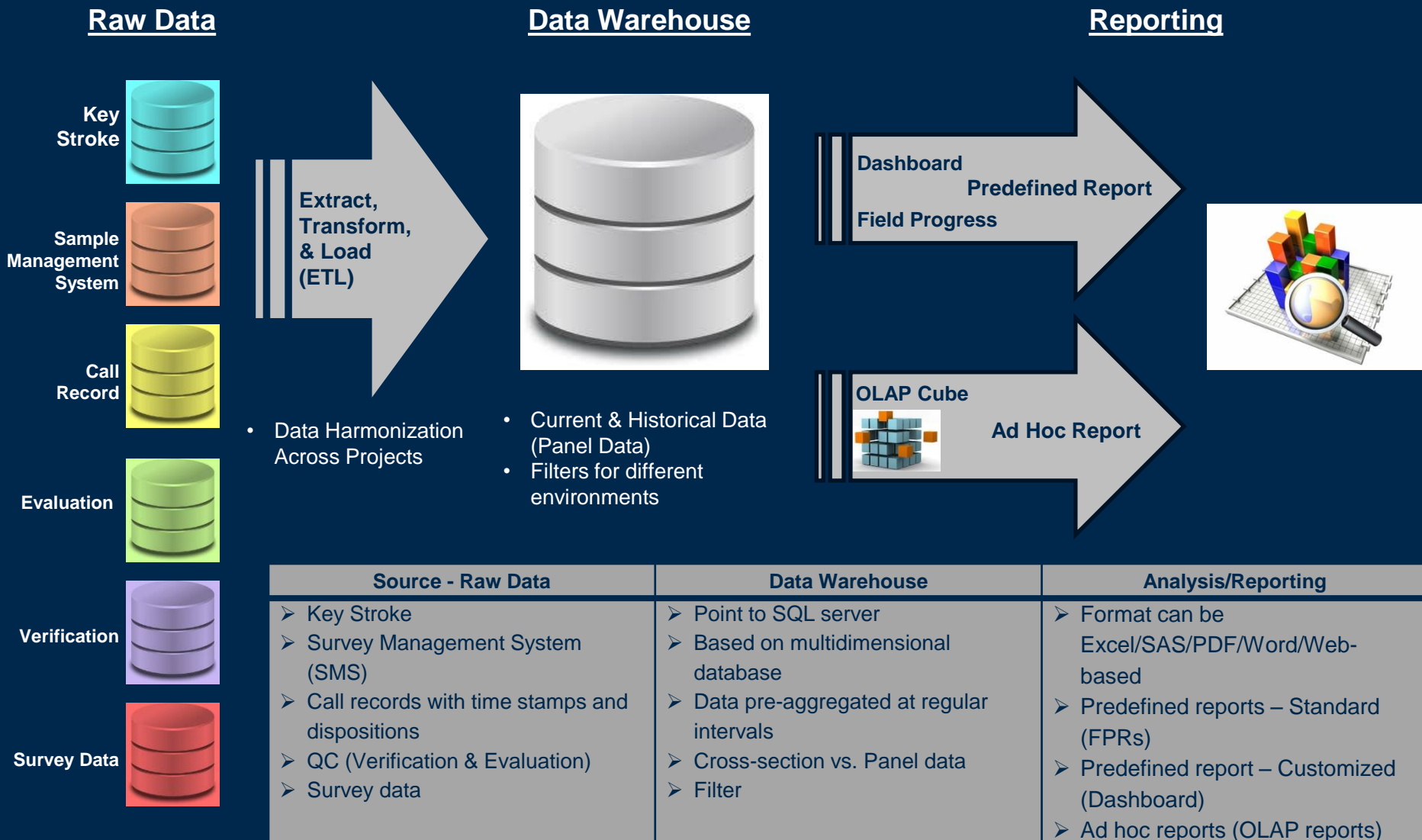
Map of Countries
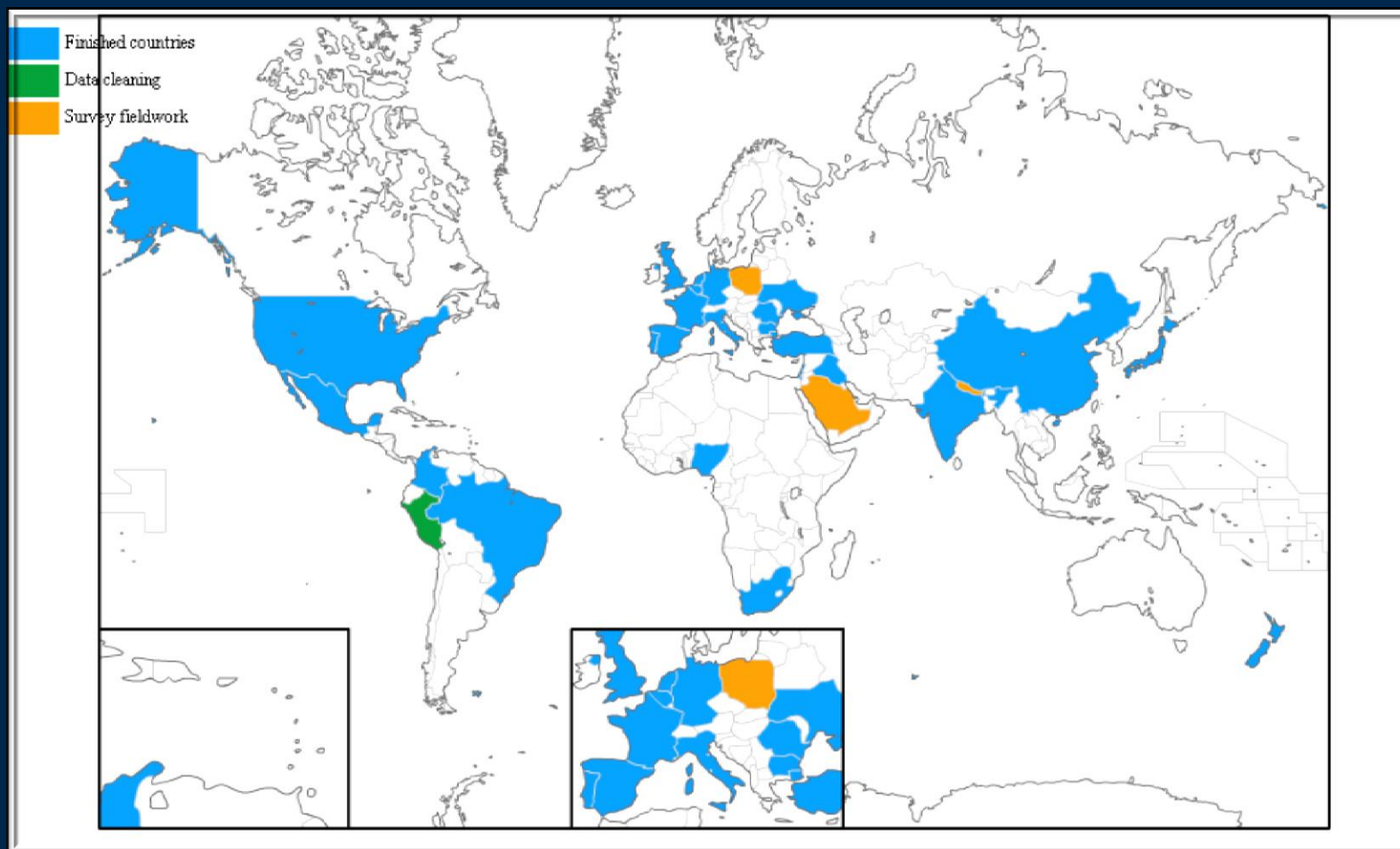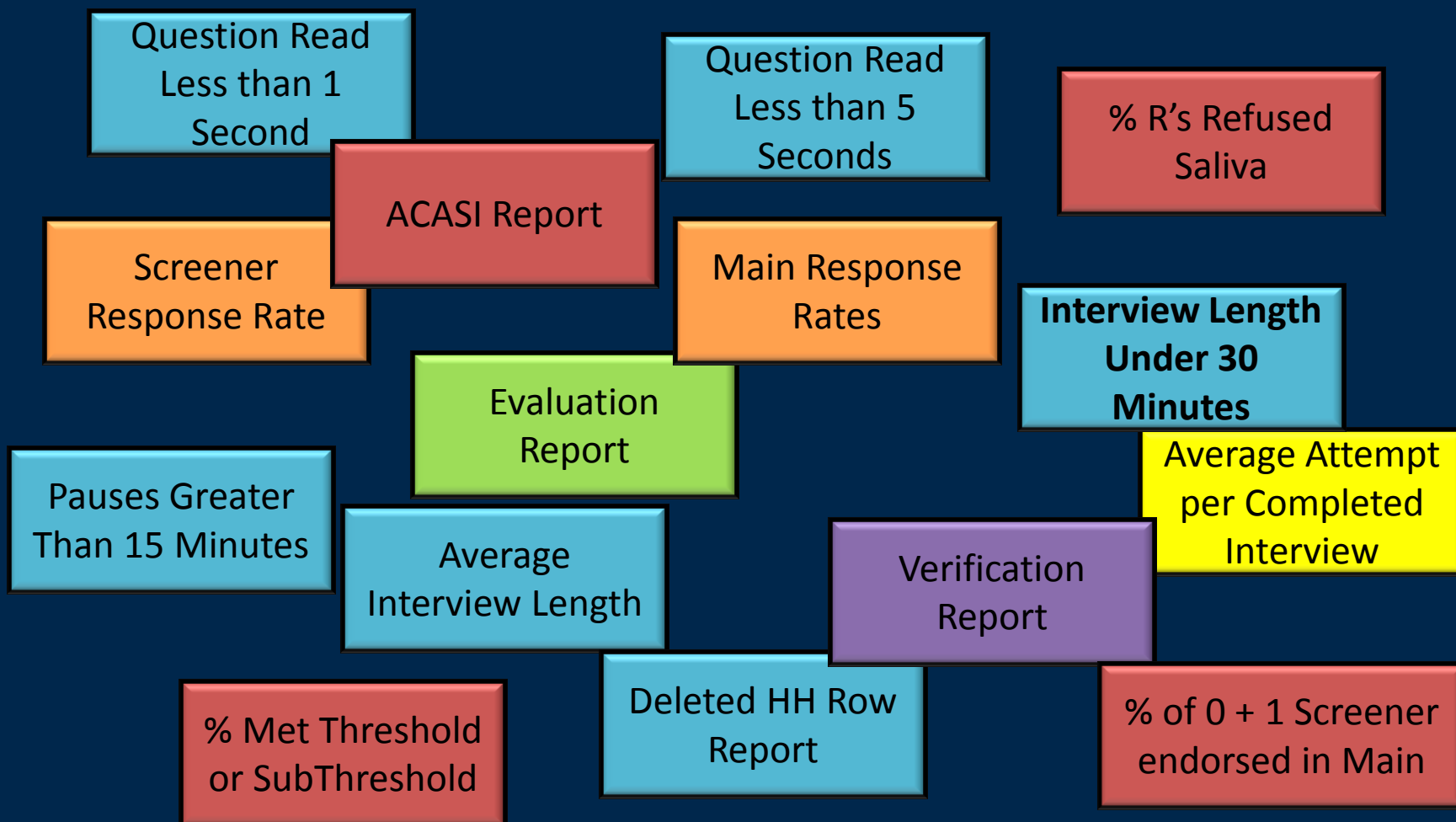
Saudi National Mental Health Survey

- National Probability Sample.
  - Two respondents selected randomly per household & 6,500 Adults from all areas of Kingdom of Saudi Arabia (KSA).
- Began field work in February 2014.
  - Interview lengths: median = 180 mins
  - Face-to Face.
  - Computerized Administered Personal Interviews (CAPI) using Blaise.
  - Audio-Computerized Administered Self Interview (A-CASI)
- Gender match interviewers with respondents.
- Interviews are NOT recorded.
- Saliva collected.

Analytical Reporting Process Chart

Raw Data

- Key Stroke
- Sample Management System
- Call Record
- Evaluation
- Verification
- Survey Data

Extract, Transform, & Load (ETL)

- Data Harmonization Across Projects

Data Warehouse

- Current & Historical Data (Panel Data)
- Filters for different environments

Reporting

Dashboard — **Predefined Report**

Field Progress

OLAP Cube — **Ad Hoc Report**

Source - Raw Data	Data Warehouse	Analysis/Reporting
➢ Key Stroke	➢ Point to SQL server	➢ Format can be Excel/SAS/PDF/Word/Web-based
➢ Survey Management System (SMS)	➢ Based on multidimensional database	
➢ Call records with time stamps and dispositions	➢ Data pre-aggregated at regular intervals	➢ Predefined reports – Standard (FPRs)
➢ QC (Verification & Evaluation)	➢ Cross-section vs. Panel data	➢ Predefined report – Customized (Dashboard)
➢ Survey data	➢ Filter	➢ Ad hoc reports (OLAP reports)

Quality Control Measures

Question Read Less than 1 Second

Question Read Less than 5 Seconds

% R's Refused Saliva

ACASI Report

Screener Response Rate

Main Response Rates

Interview Length Under 30 Minutes

Evaluation Report

Pauses Greater Than 15 Minutes

Average Interview Length

Verification Report

Average Attempt per Completed Interview

% Met Threshold or SubThreshold

Deleted HH Row Report

% of 0 + 1 Screener endorsed in Main

Dashboard + Cube = Dynamic Dashboard

- Dashboard displays data in an easy-to-read way, but contains static data
- Cube is linked directly to data and can "drill down" or aggregate, but lacks easy-to-read display

- Dynamic Dashboard combines dashboard and cube
  - Requires extra effort to set up initially
  - Data refreshed every time you open file
  - Easy-to-read displays
  - Drill down to case and question level
  - Allows data exploration

Dynamic Dashboard

KSA_Interviewer_QC_Summary_2014-03-18.xlsx - Microsoft Excel

| Explore Data | QC Summary | Interviewer Rankings | Denominators |

	Indicator Label	Description of QC Indicator
Level 1 Indicators		
Question field time – Under 1 sec (Main - Screening module)	Qtime_SC	Flags interviewers who have at least one main interview where any question in the Screener module is read in under 1 second; and displays count of occurrences across all interv
Question field time – Under 1 sec (Main - All other CIDI modules)	Qtime_Main	Flags interviewers who have at least one main interview where any question in the Main interview is read in under 1 second; and displays count of occurrences across all intervie
Failed Verification	Ver_Fail	Flags interviewers who have failed a verification; and displays a percentage of occurrences across verified cases.
Interview length (Screener)	Short_SCRN	Flags interviewers who have a screening interview length under 6 minutes, and displays a percent of occurrences across all interiwer's completed screeners.
Interview length (Main) for Short interviews	Short_Main_Sh	Flags interviewers who have a Main - Short interview length under 35 minutes; and displays a percent of occurrences across all of interviewer's completed short mains.
Interview length (Main) for Long Threshold	Short_Main_LT	Flags interviewers who have a Main - Long interview length under 121 minutes; and displays a percent of occurrences across all of interviewer's completed long threshold mains.
Interview length (Main) for Long Others	Short_Main_LO	Flags interviewers who have a Main - Long Others interview length under 62 minutes; and displays a percent of occurrences across all of interviewer's completed long other main
Long pauses (Main)	Long_Paus	Flag interviews who have at least one pause of more than 10 minutes in the main interview; and displays a percent of occurrences across all interviewer's completed main intervi
Household Roster Delete Report (Screener)	Delete_HH	Flags interviewers who deleted a household member from the roster in at least three screening interviews; and displays a count of occurrences across all of interviewer's complet
Number of completed interviews per day	#Iw_Day	Flags interviewers who completed three or more main interviews in one day.
Times between interviews (e.g. too short)	Short_Tim_Bw_Iw	Flags interviewers who moved between different households in under 10 minutes; and displays a count of occurrences across all of interviewer's sample.
Level 2 Indicators		
Other Verifications	Ver_Dis_Unabl	Ranks interviewers from high to low on percentage of verifications that have a discrepancy or unable-to-verify outcome; and flags the interviewers with the three highest percent
Percent of 0, 1, 2, or 3+ Screener endorsed in Main	Negative_SC	Ranks interviewers from high to low on percentage of screener module and screening questions in diagnostic sections in the long version (in the main interview) that have only 0
Concerning pattern of Screener endorsed in Main in quartiles	Quar_Neg_SC	Ranks interviewers from high to low on the difference in percentage of screeners flagged for endorsement across quartiles (difference between Q1-Q2; Q1-Q3, Q1-Q4;) and flags
Prevalence rates	Prevalence	Ranks interviewers from low to high on prevalence of mental health disorders; and flags the interviewers with the three lowest prevalence rates.
Patten of consecutive "No's" in Main - Screening module	Max_Streak	Ranks interviewers from high to low on average maximum number of consecutive Nos in the screener module in completed main interviews; and flags the interviewers with the th
Percent of HH18-HH25 endorsed (Screener)	HH18_HH25	Rank interviewer from high to low on percentage of screening interviewers with zero disability questions endorsed; and flags the interviewers with the three highest percentages.
Average interview length (Screener)	Av_Time_SCRN	Ranks interviewers from shortest to longest average interview length for screening interviews; and flags the interviewers with the three lowest average interview length.
Average interview length (Main)	Av_Time_Main	Ranks interviewers from shortest to longest average interview length for main interviews; and flags the interviewers with the three lowest average interview length.
Average Main interview length - Long Threshold	Av_Time_Main_LT	Ranks interviewers from shortest to longest average interview length for Main - Long; and flags the interviewers with the three lowest average interview length.
Average Main interview length - Long Others	Av_Time_Main_LO	Ranks interviewers from shortest to longest average interview length for Main - Long Others; and flags the interviewers with the three lowest average interview length.
Average Main interview length - Short	Av_Time_Main_S	Ranks interviewers from shortest to longest average interview length for Main - Short; and flags the interviewers with the three lowest average interview length.
Decline of average Main interview length	Quar_Av_Time_Main	Ranks interviewers from high to low on the difference in average Main length across quartiles (difference in average Main - Long interview length from Q1-Q2; Q1-Q3, Q1-Q4;); an
Decline of average Main interview length - Long Threshold	Qaur_Av_Time_LT	Ranks interviewers from high to low on the difference in average Main - Long interview length across quartiles (difference in average Main - Long interview length from Q1-Q2; Q
Decline of average Main interview length - Long Others	Quar_Av_Time_LO	Ranks interviewers from high to low on the difference in average Main - Long Others interview length across quartiles (difference in average Main - Long Others interview length
Decline of average Main interview length - Short	Quar_Av_Time_S	Ranks interviewers from high to low on the difference in average Main - Short interview length across quartiles (difference in average Main - Short interview length from Q1-Q2; Q
Sum of pauses >= 5 minutes	Sum_Pause	Ranks interviewers from high to low on average total length of summed 5 minute pauses within main interview across their completed mains; and flags the interviewers with the
Percent of saliva not given	Refus_Saliva	Ranks interviewers from high to low on percent of saliva not given; and flags the interviewers with the three highest percentage.
Percent of ACASI switching to CAPI	ACASI_sw_CAPI	Ranks interviewers from high to low on percent of ACASI switching to CAPI; and flags the interviewers with the three highest percentage.
Level 3 Indicators		
Average attempts for all SIDs with final result codes (Screener, excluding NS SIDs)	Av_Attemp_SCRN	Ranks interviewers from high to low on average attempts for finalized screenings lines (excludes non-sample SIDs); and flags the interviewers with the three highest averages.
Average attempts for all SIDs with final result codes (Main, excluding NS SIDs)	Av_Attemp_Main	Ranks interviewers from high to low on average attempts for finalized Main lines (excludes non-sample SIDs); and flags the interviewers with the three highest averages.
Listed Female Members	Listed_Female	Rank interviewers from high to low on percentage of screening interviews with no Female household member listed; and flags the interviewers with the three highest percentage
Listed Male Members	Listed_Male	Rank interviewers from high to low on percentage of screening interviews with no Male household member listed; and flags the interviewers with the three highest percentages.
Eligible Female	Elig_Female	Rank interviewers from high to low on percentage of screening interviews with no eligible Female household member listed; and flags the interviewers with the three highest percenta

Table_of_Contents | Explore_Data | QC_Summary | Iwer_Rankings | Denominators | Qtime_SC | Qtime_Main | Ver_Fail | Short_SCRN | Short_Main_Sh | Short_Main_LT | Short_Main_LO | Long_Paus | Delete_HH | #Iw_Day | Short_Tim_Bw_Iw | Bw_Iw_Details | Ver_Dis_Unab

Running OLAP query ... (Press Esc to cancel)

Level 1 Indicators

- Flag single case or instance
- Can start flagging on Day 1 of data collection
- Intervene immediately
- Type of Indicators
  - Question field time – Under 1 sec
  - Failed verifications
  - Short Interview Length
  - **Long pauses**
  - Household roster delete
  - Number of completed interviews per day
  - **Time between households**

Key Stroke

Verification

Level 2 Indicators

Key Stroke

- Iwer level percentages or averages
- Rank by interviewer and then flag 3 "worst"
- Type of Indicators
  - Other verifications (e.g. unable to verify)
  - % Gate questions endorsed
  - Prevalence rates
  - Pattern of consecutive no's
  - Average interview length
  - Decline of average interview length
  - **Sum of pauses**
  - % Saliva not given
  - **% ACASI switch to CAPI**

Verification

Survey Data

Level 3 Indicators

- Iwer level percentages or averages
- Rank by interviewer and then flag 3 "worst"
- Type of Indicators
  - Average attempts per completed
  - **Listed HH members by gender**
  - Eligibility by gender
  - Response rates

Survey Data

SMS

Call Records

Quality Control Summary

Back to T

LEVEL 3 INDICATORS

Row Labels	QC_Av_Attemp_SCRN	QC_Av_Attemp_MAIN	QC_Listed_Female	QC_Listed_Male	QC_Elig_Female	QC_Elig_Male	Sum_Level3
Iwer 1	0	0	1	0	1	1	3
Iwer 2	0	1	0	0	1	1	3
Iwer 3	1	1	0	0	0	0	2
Iwer 4	0	0	0	0	1	0	1
Iwer 5	0	0	1	0	0	0	1
Iwer 6	1	0	0	0	0	1	2

(Sum_Level3 continues: 2, 1, 0, 1, 0, 0, 0, 0, 0, 16)

Back to T

LEVEL 2 INDICATORS

Row Labels	QC_Ver_Dis_Unabl	QC_Negative_SC	QC_Max_Streak	QC_HH18_HH25	QC_Av_Time_SCRN	QC_Av_Time_Main_LT	QC_Av_Time_Main_LO	QC_Av_Time_Main_S	QC_Pause_Num	QC_Refus_Saliva	QC_ACASI_sw_CAPI	Sum_Level2
Iwer 1	0	1	1	1	1	1	1	1	0	1	0	8
Iwer 2	1	1	1	0	0	0	0	0	1	0	1	6
Iwer 3	1	1	0	1	0	1	0	0	0	0	0	4
Iwer 4	0	0	0	0	1	0	1	1	0	1	0	4
Iwer 5	0	0	0	1	0	0	0	0	0	0	1	2
Iwer 6	0	0	0	0	1	1	0	1	0	1	0	5

(Sum_Level2 continues: 2, 2, 2, 1, 0, 0, 0, 0, 0, 36)

Back to Table of Contents

LEVEL 1 INDICATORS

Row Labels	D_Scrn_1001_8010_ADT	D_Main_1001_1005_ADT	Sum_QC_Indicators	QC_Qtime_SC	QC_Qtime_Main	QC_Ver_Fail	QC_Short_SCRN	QC_Short_Main_Sh	QC_Short_Main_LT	QC_Short_Main_LO	QC_Long_Paus	QC_Delete_HH	QC_Iw_Day	QC_Short_Tim_Bw_Iw	Sum_Level1
Iwer 1	53	18	16	0	1	0	1	1	0	0	1	0	0	1	5
Iwer 2	11	9	11	0	0	1	0	0	0	0	1	0	0	0	2
Iwer 3	21	20	10	0	1	1	1	0	0	0	1	0	0	0	4
Iwer 4	26	8	9	0	1	0	1	0	0	0	1	0	0	1	4
Iwer 5	29	11	8	0	1	0	1	0	0	0	1	1	0	1	5
Iwer 6	10	8	8	0	0	1	0	0	0	0	0	0	0	0	1
Iwer 7	24	20	7	0	1	0	1	0	0	0	1	0	0	0	3
Iwer 8	28	9	6	0	0	0	1	0	0	0	1	0	0	1	3
Iwer 9	16	5	5	0	1	0	0	0	0	0	1	0	0	1	3
Iwer 10	12	8	3	0	1	0	0	0	0	0	0	0	0	0	1
Iwer 11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Iwer 12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Iwer 13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Iwer 14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Grand Total	230	116	83	0	7	3	6	1	0	0	8	1	0	5	31

	A	B	
1			
2	Explore Data	QC Summary	Interviewer Rankings	Deno
3			
4			
5	**Level 1 Indicators**	**Indicator Label**	**Description of QC**
6	Question field time – Under 1 sec (Main - Screening module)	Qtime_SC	Flags interviewers w
7	Question field time – Under 1 sec (Main - All other CIDI modules)	Qtime_Main	Flags interviewers w
8	Failed Verification	Ver_Fail	Flags interviewers w
9	Interview length (Screener)	Short_SCRN	Flags interviewers w
10	Interview length (Main) for short interviews	Short_Main_Sh	Flags interviewers w
11	Interview length (Main) for Long Threshold	Short_Main_LT	Flags interviewers w
12	Interview length (Main) for Long Others	Short_Main_LO	Flags interviewers w
13	Long pauses (Main)	Long_Paus	Flag interviews who
14	Household Roster Delete Report (Screener)	Delete_HH	Flags interviewers w
15	Number of completed interviews per day	#Iw_Day	Flags interviewers w
16	Times between interviews (e.g. too short)	Short_Tim_Bw_Iw	Flags interviewers w
17	**Level 2 Indicators**		
18	Other Verifications	Ver_Dis_Unabl	Ranks interviewers
19	Percent of 0, 1, 2, or 3+ Screener endorsed in Main	Negative_SC	Ranks interviewers
20	Concerning pattern of Screener endorsed in Main in quartiles	Quar_Neg_SC	Ranks interviewers
21	Prevalence rates	Prevalence	Ranks interviewers
22	Patten of consecutive "No's" in Main - Screening module	Max_Streak	Ranks interviewers
23	Percent of HH18-HH25 endorsed (Screener)	HH18_HH25	Rank interviewer fr
24	Average interview length (Screener)	Av_Time_SCRN	Ranks interviewers

Long Pauses

- # Aggregate

- # Drill down

- # Export

Time Dimension

- Aggregate

- Drill down

- Export

					Total D_ Main_1001_ADT	Total Av_Time_Main
Row Labels	Q1 Q2 Q3 Q4	Q1	Q2	Q3	Q4		
⊟ lwer 1	2 3 2 3	197.0	161.7	238.5	149.0	10	180.3
⊞ 2014-02-12	1	118.0				1	118.0
⊞ 2014-02-14	1	276.0				1	276.0
⊞ 2014-02-22	1		69.0			1	69.0
⊞ 2014-02-27	1		321.0			1	321.0
⊞ 2014-02-28	1		95.0			1	95.0
⊞ 2014-03-02	2			238.5		2	238.5
⊞ 2014-03-12	1				315.0	1	315.0
⊞ 2014-03-16	1				83.0	1	83.0
⊞ 2014-03-22	1				49.0	1	49.0
⊞ lwer 2	2 3 3 3	84.5	176.3	116.3	163.3	11	139.7
⊞ lwer 3	5 5 5 5	209.6	124.0	135.4	116.6	20	146.4
⊞ lwer 4	2 2 2 3	186.5	47.5	132.5	125.7	9	123.3
⊞ lwer 5	2 2 2 3	264.5	260.5	91.5	325.0	9	245.3
⊞ lwer 6	2 3 3 3	364.5	272.3	175.3	151.3	11	229.6
⊞ lwer 7	5 6 5 6	173.2	105.8	115.4	92.0	22	119.5
⊞ lwer 8	5 6 5 6	204.8	249.7	142.6	173.0	22	194.2
⊞ lwer 9	3 3 3 3	246.3	258.7	216.3	193.3	12	228.7
Grand Total	28 33 30 35	N/A	N/A	N/A	N/A	132	N/A

Lessons Learned

- When cases are cumulated for the long term monitoring
  - If the performance is NOT stable
    - Most likely it is <u>sample</u> effect
  - If the performance is stable
    - Most likely it is <u>interviewer</u> effect

- Interpretations of flagged interviewers/cases
  - Necessary adjustments of flagging protocols
    - Cross-cultural effect (long pause)
    - Respondent behavior (ACASI)
    - Sample characteristic (single Male HH)

Lessons Learned (cont.)

- Dynamic process

  - Define measures > use them > re-define them > modify flagging protocols or programming codes

- Create ranking vs. Focus on distribution of Iwers' performance

Acknowledgements

Special Thanks

to

Joel Devonshire

Brad Goodwin

Feedback & Questions?

yuchiehl@umich.edu