



Minimizing Interviewer Effects: Innovations in Quality Assurance and Control in Multicultural, Multinational, and Multiregional Surveys

OECD Seminar on Managing the Quality of Data Collection in Largescale
Assessments

May 12th, 2017

Kristen Cibelli Hibben
Beth-Ellen Pennell
Lesli Scott
University of Michigan



Overview

- Role of the interviewer
- Interviewer effects on survey error
- Interviewer effects in the context of 3MC surveys
- Minimizing interviewer effects
- Trends and innovations in QA/QC
- Recommendations for PIAAC



Role of the Interviewer

Interviewers play a number of essential roles in survey execution:

- Building sampling frames by listing addresses
- Selecting respondents within selected units
- Gaining cooperation from sampled respondents
- Clarifying the respondent's role in the interview process
- Managing the question-and-answer process
- Recording respondent answers
- Editing answers for correctness and transmitting data to the survey organization

Groves et al. (2009)



Interviewer Effects on Survey Error

- Interviewers can affect many aspects of the survey process, both positively and negatively (West and Blom, 2016)
- Interviewers can affect four out of seven of sources of error in total survey error framework: (West and Blom, 2016)
 - Coverage; nonresponse; measurement; processing
- Error in survey estimates consist of:
 - Systematic deviations from a target value (bias) and
 - Variances of estimates (reflecting estimate instability over conceptual replications)



Research on Interviewer Related Bias

- Decreased reporting of socially undesirable responses and increased reporting of socially desirable responses (e.g., see Tourangeau and Yan (2007), Sakshaug, Tourangeau and Yan (2010), Tourangeau and Smith (1996))
- Altered reporting when interviewers' observable characteristics are related to the topic (e.g., race, gender) (see Davis et al., 2010 and West & Blom, 2016 for reviews)
- Differences in reporting related to interviewer experience (Hughes, Chromy, Giacoletti, & Odom, 2002), expectations (Fowler & Mangione, 1990), and attitudes (e.g. see Durrant et al. (2010) and Mneimneh et al. (2017))



Research on Interviewer Variance

- The sampling variance of a cluster design (used in most area surveys) is confounded with the non-sampling variance (i.e., that from interviewers)
- “Interpenetrated sample assignments” are needed to measure interviewer variance but are rare due to operational difficulties and costs
- O’Muircheartaigh and Campanelli (1998) :
 - The median interviewer effect was an 80% increase in the variance across the 820 study variables
 - Estimates suggest that the effect of the sample design and the interviewers was almost the same



Interviewer Effects in 3MC

- European Social Survey (ESS) - levels of interviewer variation vary substantially among countries and high ICCs despite detailed specifications (Beullens and Loosveldt, 2016)
- Interviewer error is probably the most serious error source in 3MC surveys (Mneimneh et al, forthcoming)



Minimizing Interviewer Effects

- Self-administered modes for sensitive questions
- Minimize questions that require non-standard interviewer behavior
- Minimize interviewer workload; cost trade-off
- Recruitment criteria (e.g., amount of experience, quality measures on previous studies, etc.)
- General Interviewing Techniques (GIT), standardization, non-directive probing, recording answers exactly as given, etc.
 - Train-the-trainer
 - New methods; streaming video, etc. to control costs
 - Training re: administering assessment; managing environment
- Supervision, quality assurance and quality control



Minimizing Interviewer Effects

- Implementing rigorous QA/QC standards and procedures at all stages of the survey lifecycle can help minimize interviewer effects
- Technological and methodological advances allow for more rigorous approaches to interviewer monitoring



Trends/Innovations in QA/QC

- Use of real-time interviewer-level indicators to guide QC interventions is limited in 3MC surveys
- However, examples can be drawn from innovative international studies and some exemplar 3MC studies:
 - Saudi National Mental Health Survey
 - China Mental Health Study
 - China Finance Study
 - Consumer Pyramids Survey, India
 - Survey of Health, Ageing, and Retirement in Europe (SHARE)



The Kingdom of Saudi Arabia: Saudi National Mental Health Survey

- Target population: non-institutionalized Saudi Nationals 15-65 years old
- CAPI with ACASI
- Sample design:
 - Multistage area probability
 - Stratified by 13 administrative areas
 - Sample Size:
 - 6,500 interviews in 3225 households
 - Two randomly selected respondents per household: female and male
 - Gender matching





Locating a Household in Riyadh

In collaboration with:

**King Faisal Specialist
Hospital and Research
Center, Prince Salman Center
for Disability Research, King
Saud University, King
Abdulaziz Center for Science
and Technology, the Ministry
of Health, and the Ministry of
Economy and Planning**





Analytical Reporting Process Chart

Raw Data

Data Warehouse

Reporting

Key
Stroke



Sample
Management
System



Call
Record



Evaluation



Verification



Survey Data



**Extract,
Transform,
& Load
(ETL)**



- Data Harmonization Across Projects

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

Dashboard

Predefined Report

Field Progress

OLAP Cube

Ad Hoc Report



Source - Raw Data	Data Warehouse	Analysis/Reporting
<ul style="list-style-type: none">➤ Key Stroke➤ Survey Management System (SMS)➤ Call records with time stamps and dispositions➤ QC (Verification & Evaluation)➤ Survey data	<ul style="list-style-type: none">➤ Point to SQL server➤ Based on multidimensional database➤ Data pre-aggregated at regular intervals➤ Cross-section vs. Panel data➤ Filter	<ul style="list-style-type: none">➤ Format can be Excel/SAS/PDF/Word/Web-based➤ Predefined reports – Standard (FPRs)➤ Predefined report – Customized (Dashboard)➤ Ad hoc reports (OLAP reports)



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



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 verification
 - Short Interview Length
 - Long pauses
 - Household roster delete
 - Number of completed interviews per day
 - Time between households



Key Stroke



Verification



Level 2 Indicators

- Lower level percentages or averages
- Rank by interviewer and then flag “worst”
- Type of Indicators
 - Other verifications (e.g., unable to verify, other outcomes (noninterviews))
 - % 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





Level 3 Indicators

- Lower level percentages or averages
- Rank by interviewer and then flag “worst”
- Type of Indicators
 - Average attempts per completed
 - Listed HH members by gender
 - Eligibility by gender
 - Response rates



Survey
Data



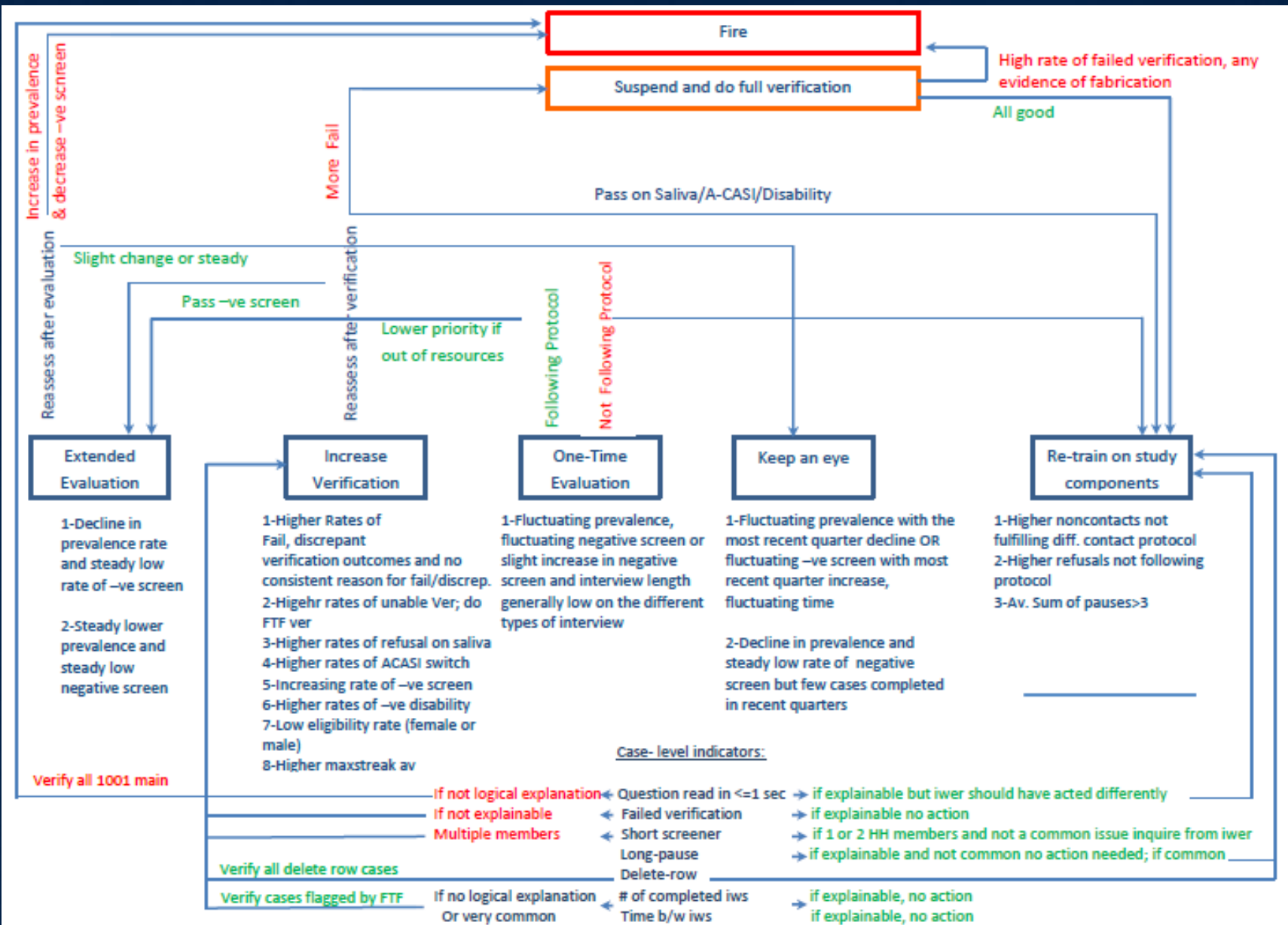
SMS



Call
Records



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



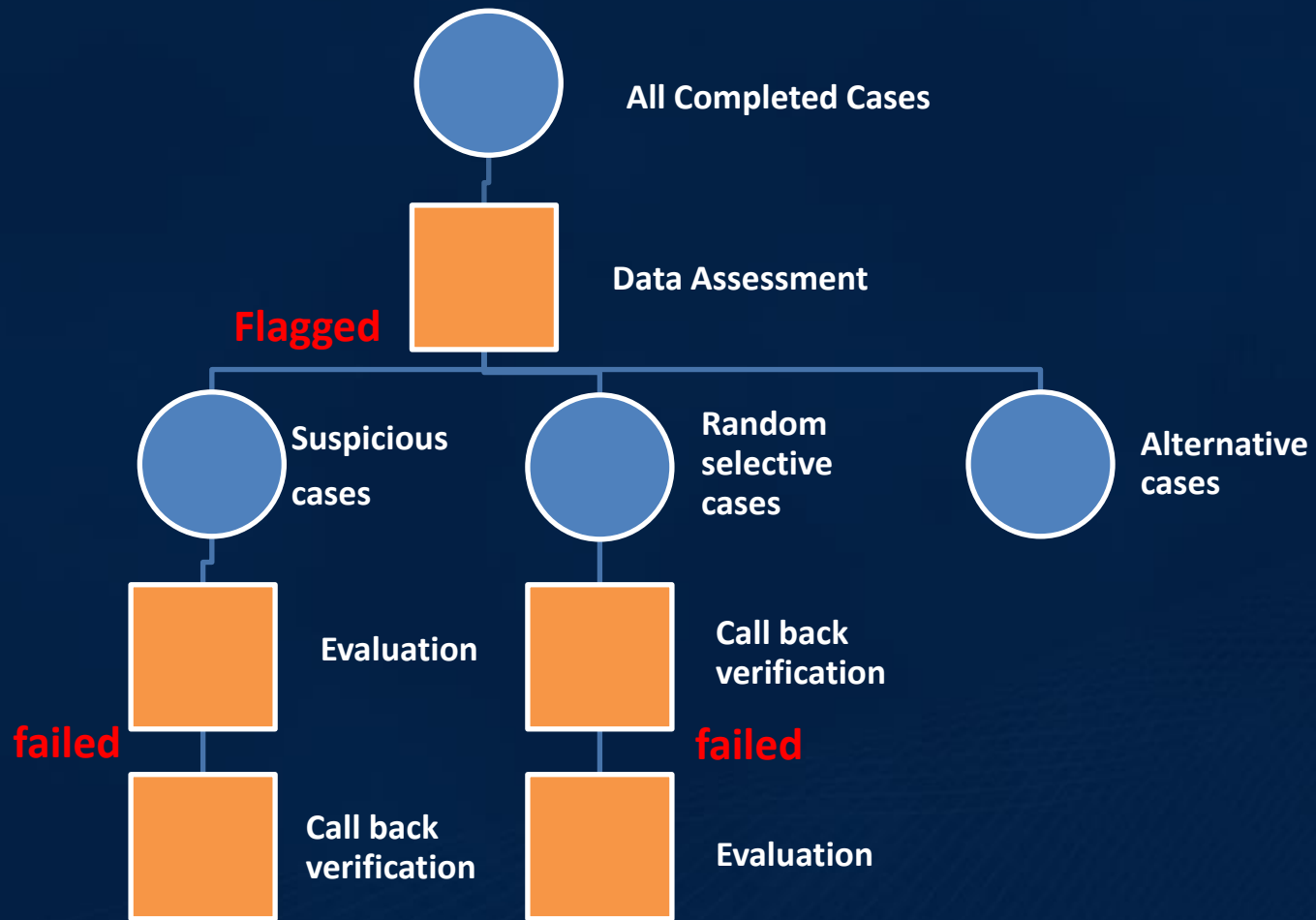


China National Mental Health Study

- Institute for Social Science Survey at Peking University
- National probability sample
 - 30,000 interviews
- Interviews are audio-recorded
- Paradata used to stratify quality control – optimize resources
- Stratified cases on interview length, missing data, and time of interview, etc.



Quality Control Optimization





Results

- Yielded higher rates of falsification and other interviewer errors, including problems with question reading than did those cases selected purely at random



India: Consumer Pyramid Panel

- Conducted by the Centre for Monitoring the Indian Economy (CMIE)
- Nationwide survey
- ~158,000 households, 3 times a year
- Sample released to enable **daily** estimates
- Paperless, mobile technology
- Real time validation
- Innovative use of GPS and Google Earth





Mumbai Slums





Mobile Data Collection





Household Selection Screen

16:20

Date Confirmation

Today's Date: 11-11-2014
Is it Correct?

Yes No

16:20

Home (Version : 16.4)

Select Household

Download or Upload Data

Exit

Status	No. of Household
TOTAL	: 27
PENDING	: 24
COMMENCED	: 1
COMPLETED	: 2
FORM_DOWNLOADED	: -
FORM_UPLOADED	: -
DATA_UPLOADED	: -
RECHECK	: -
REVISIT	: -
REJECT	: -
ACCEPTED	: -

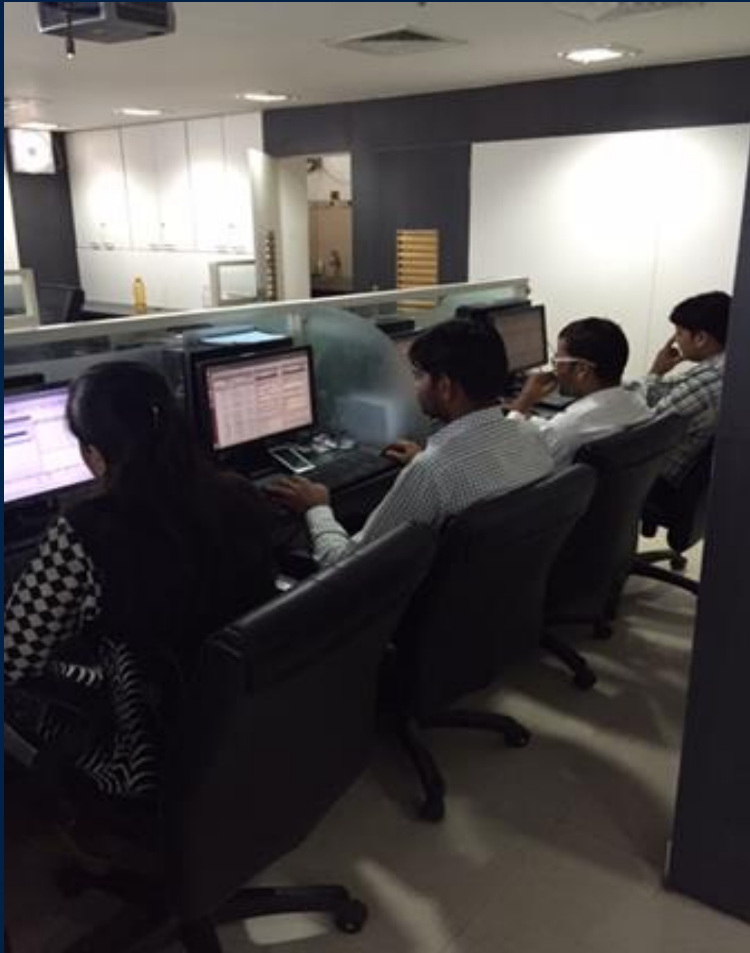
16:20

Select Household

Status	Town/Village	Ward	Ceb	HH No.
COMMENCED	Amode	0	0	1
PENDING	Amode	0	0	2
PENDING	Amode	0	0	3
PENDING	Amode	0	0	4
PENDING	Amode	0	0	5
PENDING	Amode	0	0	6
PENDING	Amode	0	0	7
PENDING	Amode	0	0	8
PENDING	Amode	0	0	9
PENDING	Amode	0	0	10
PENDING	Amode	0	0	11
PENDING	Amode	0	0	12
PENDING	Amode	0	0	13
PENDING	Amode	0	0	14



Quality Monitoring





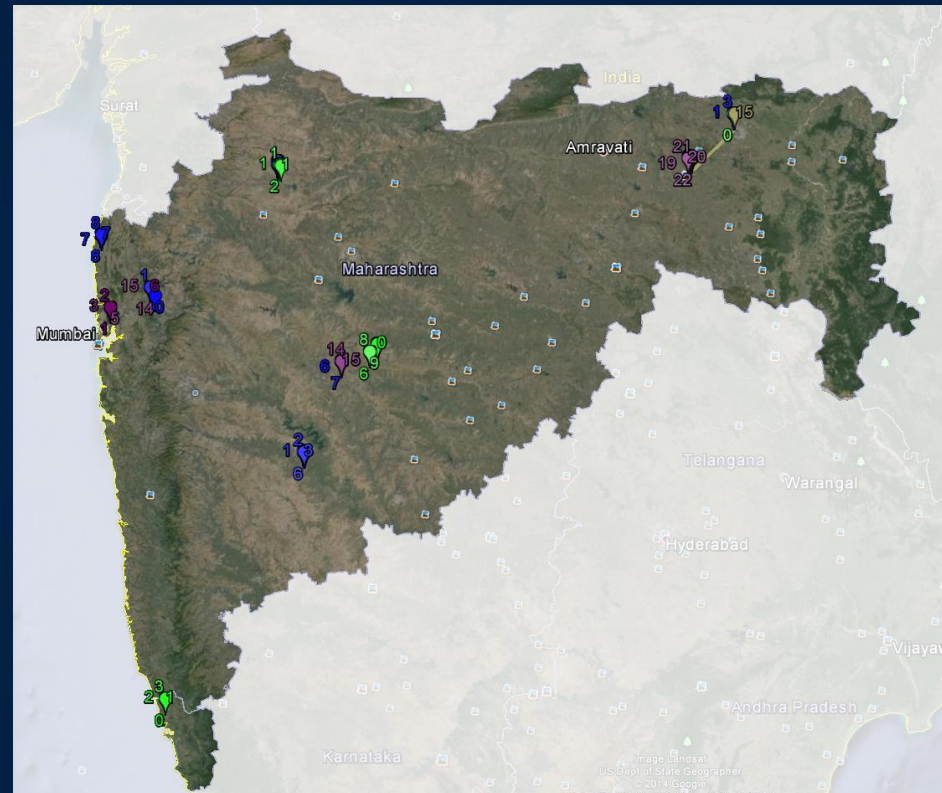
In-Field Validation Checks

1. Cross-check GPS location captured by mobile device against movement seen on Google Earth
2. Data checks
 - a. Matching of member details compared to past surveys
 - b. Matching of ownership of assets compared to past surveys
 - c. Completeness of data entered
 - d. Internal consistency of data entered
 - e. Consistency of values with past entries
3. Field Team Members are rotated from one round to the next
4. Avoid complete reliance on auto-checks of values



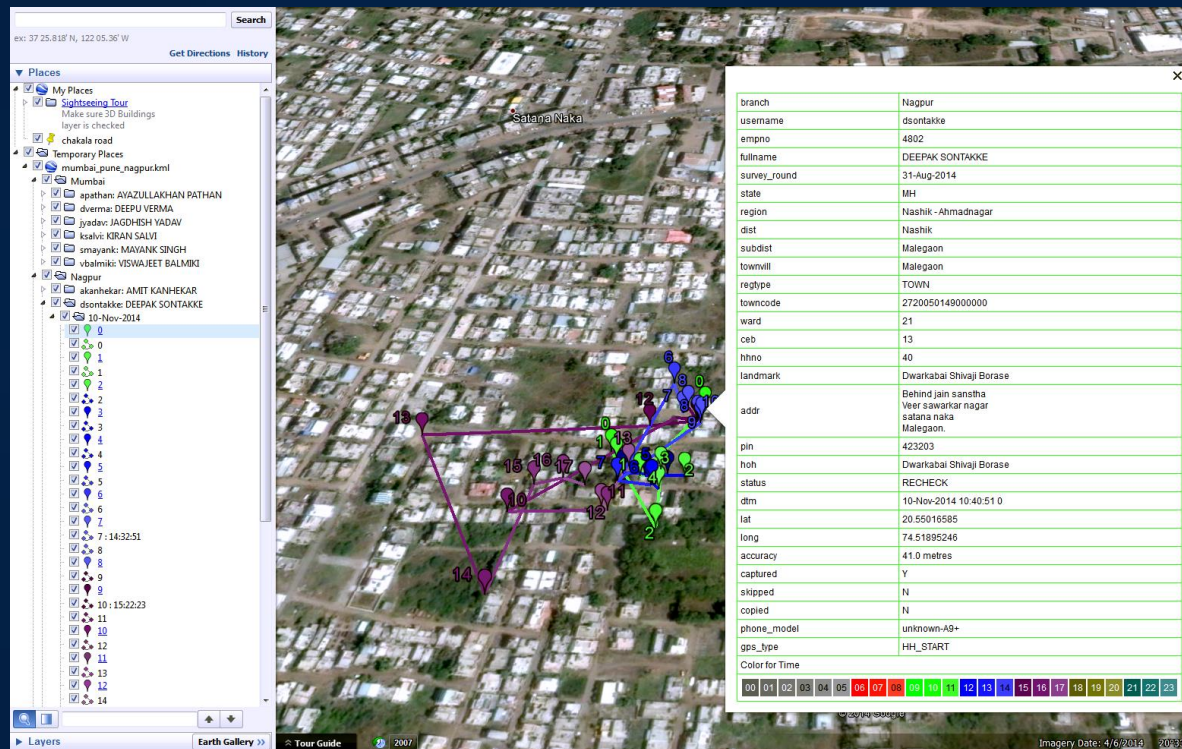


Maharashtra Example





Nagpur Example





China Household Finance Study

- Southwest University of Finance and Economics
- National survey about household income, expenses, assets, liabilities, insurance and securities, etc.
- In June-September, 2015, SWUFE sent 1600 students to interview 28,000 households in 1048 communities of 260 counties in 29 provinces.



INSTITUTE FOR SOCIAL RESEARCH
SURVEY RESEARCH CENTER
UNIVERSITY OF MICHIGAN



中国家庭金融调查
China Household Finance Survey

二零一三中国家庭金融调查访员暨录音核查员合影留念

2013年8月3日





让中国了解自己 让世界认识中国

Let China know herself Let the world know China

—甘翠

输入访员姓名查看轨迹及位置 搜索



中国家庭金融调查
China Household Finance Survey

今日 @CHFS

总成功量

0

0



中国小微企业调查
China Micro and Small Enterprise Survey

今日 @CMES

总成功量

0

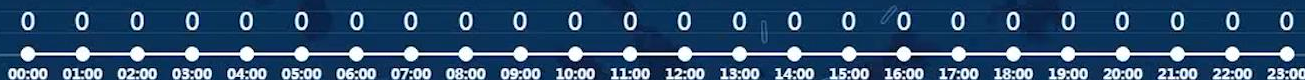
0

成功趋势

替换趋势

接触趋势

100
80
60
40
20
0



停止轮播

小组名称

目标数 完成数 完成比(%)

1	CHFS第153组	507	507	100↑
2	CHFS第125组	326	326	100↑
3	CHFS第69组	323	323	100↑
4	房产问卷第3组	317	317	100↑
5	CHFS第122组	304	304	100↑
6	CHFS第121组	299	299	100↑
7	CHFS第84组	294	294	100↑
8	CHFS第83组	290	290	100↑
9	房产问卷第1组	290	290	100↑
10	CHFS第140组	289	289	100↑
11	CHFS第118组	288	288	100↑
12	CHFS第123组	288	288	100↑
13	CHFS第117组	285	285	100↑
14	CHFS第150组	282	282	100↑
15	CHFS第151组	280	280	100↑
16	CHFS第156组	278	278	100↑
17	CHFS第68组	276	276	100↑
18	CHFS第十六组	271	271	100↑
19	CHFS第129组	268	268	100↑
20	CHFS第66组	260	260	100↑
21	CHFS第161组	260	260	100↑
22	CHFS第162组	259	259	100↑



访员分布



样本分布



Survey of Health, Retirement, and Aging in Europe



- Cross-national panel study that assesses the health, socioeconomic status and social and family networks among individuals 50 years+
- 20 European countries and Israel
- Since 2004, every two years



Survey of Health, Retirement, and Ageing in Europe



- Since Wave 1, common instrumentation and sample management system used in all countries across all data collection firms
- Quality indicators: the rate of attempted households, the rate of “reached” households, interview length, and length of long introductory items in the questionnaire, number of interviews per interviewer
- Measures delivered every two weeks
- Create ‘**Compliance Profiles**’



Recommendations for PIAAC

- PIAAC has established detailed and well-structured standards and guidelines
- Move further to data-driven assessment of interviewer behavior and adherence to protocol at all stages
 - Expand training on managing assessment environment and respondent motivation (a la Groves and McGonagle (2006)
- Allow for real-time or near real-time monitoring and interventions
- Delivery of audit files with interview data linked to interviewer ID
- Limit number of interviews per interviewer



Recommendations for PIAAC

- Technical infrastructure:
 - Common electronic interview system
 - Common sample management system
 - Audio recording built into the electronic interview system (also monitor assessment environment and interactions)
 - Interviewer observations (separate application?)
 - GPS-enabled devices for tracking interviewer locations
- Levels of indicators
 - Reporting of key indicators
 - Detailed follow -up protocol
 - Regular, scheduled reporting outcomes of follow-up



Recommendations for PIAAC

- But...requires expanded role and sufficient human and technical resources for central coordination
- Cost/quality tradeoff
- Follow-up critical
- Local buy in: capacity building



Recommendations for PIAAC

- Opportunities for future research:
 - Magnitude and potential sources of interviewer effects across countries
 - Interviewer variance in interview privacy
 - Recording interactions throughout PIAAC (background and assessment)
- Advance knowledge for PIAAC and 3MC studies more broadly



INSTITUTE FOR SOCIAL RESEARCH
SURVEY RESEARCH CENTER
UNIVERSITY OF MICHIGAN

Merci!