

Module 6: Data quality monitoring

Video 6 of 6: Practical quality control

Sharan Sharma



THE WORLD BANK

MANNHEIM
BUSINESS SCHOOL

November 2020

Module 6 – Remote training on Phone Surveys

So much data ! What should I choose?

Paradata ?

Listening-in?

CARI ?

GPS?

Data indicators?

Re-interviews

Use “whatever we can lay our hands on” is probably not the right answer!

Each data source/method has its own strength/shortcoming...

Paradata

+

- Rich
- Detail-oriented
- Error-free
- No extra cost to collect (apart from initial set-up programming)
- Unobstrusive

-

- Unknown properties
- Need technical resource to parse
- Can be too much to handle
- Not confirmatory

CARI

+

- 'Proof'
- Can tell us about the interaction
- Qualitative inputs
- Unobstrusive

-

- Needs someone to hear the recordings
- Can be difficult to transmit in some geographies

Technology has aided QC enormously, but principles are the same e.g. Continuous Quality Improvement (CQI)

What are we really after?

Continuous Quality Improvement (CQI)

1. Prepare a workflow diagram of the process and identify key process variables.
2. Identify characteristics of the process that are Critical To Quality (CTQ)
3. **Develop real-time, reliable metrics for the cost and quality of each CTQ.**
4. Verify that the process is stable (i.e., in statistical control) and capable (i.e., can produce the desired results).
5. Continuously monitor costs and quality metrics during the process.
6. Intervene as necessary to ensure that quality and costs are within acceptable limits.

Source: Biemer, 2010

Illustrative example for “Question administration”

Process	CTQ	Metrics	
1. Read question	a. Read question as worded.	CARI Observations	<div style="border: 1px solid black; padding: 5px; margin-bottom: 10px;">Flag if time-to-data entry < 3 sec OR less than 75% of pre-test time-to-data</div>
	b. Adhere to interviewer instructions.	CARI Observations	
2. Probe	a. Probe when needed.	CARI Observations	
	b. Use only neutral probes.	CARI Observations	
	c. Make a note in case of any relevant observation.	Any F2 remark ?	
3. Clarify	a. Correct clarification given.	CARI Observations	
	b. Help file accessed if needed.	F1 accessed	
	c. Make a note in case of any relevant observation.	Any F2 remark ?	
4. Give feedback	No judgemental feedback given.	CARI Observations	<div style="border: 1px solid black; padding: 5px;">Flag if start of data entry to item exit > 5 seconds</div>
5. Enter data	a. Enter data verbatim.	CARI Observations	
	b. Make a note in case of any relevant observation.	Any changed entry (on 1st visit) ? Any F2 remark ?	
6. Reconcile and move forward	a. No pop-up errors.	Hard-error popped-up ? Soft-error popped-up ?	
	b. If hard errors induced, make correct changes.	Whether 'jumped to' and made changes.	
	c. Go to the next question in order.	If no error, does the interviewer still go back ?	
	d. Make a note in case of any relevant observation.	Any F2 remark ?	
7. Offline reconciliation	a. Do the data pass the common sense test? ...	How different from previous wave data (for panels) Percentile on a distribution within stratum	

QC indicators need to be 'fit for use'

- Productivity (is not only an administrative aspect!)
 - Call attempts per hour
 - Interviews per day
 - Ineligibles per hour
- Non-response
 - Response rate
 - Response rate by respondent demographic
- Measurement error
 - Length of interview
 - Number of comments

Some QC guiding principles

1. Check available resources: common to see a lot of data collected completely unused.
2. Appoint a QC champion: formal appointments better.
3. Keep it simple !
4. Do not need many indicators – a few thought-fully chosen ones will do. Common problem in QC is to get **OVERWHELMED**.

Some QC guiding principles

5. Make it action-oriented

- Lot of data collected are not actioned : wasted time and resource
- Keep asking: “How do I action these?”

Visual displays can facilitate pattern exploration

- Heat-maps
- Dashboards

Heat-maps

Interviewer ID	Response rate	Average Interview length (mins)	Numeric responses with rounding	Behavior coding % 'Fatal' error
1	100%	69	5%	5%
2	56%	40	58%	14%
3	75%	74	7%	2%
4	85%	78	12%	5%
5	52%	60	25%	6%
6	88%	80	32%	8%
7	67%	88	19%	2%
8	81%	76	33%	9%
9	73%	62	21%	2%
10	85%	72	5%	1%

Dashboards

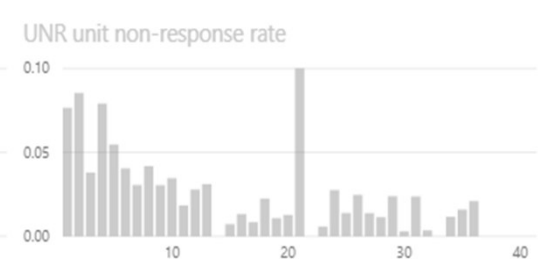
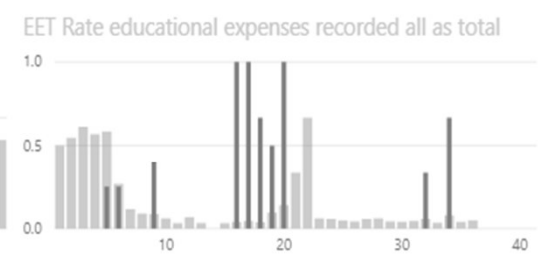
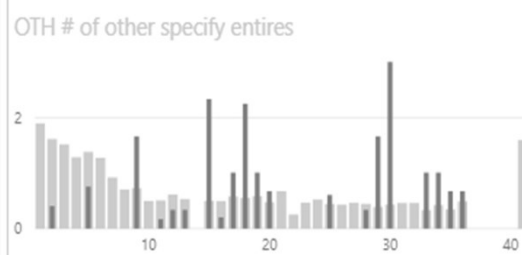
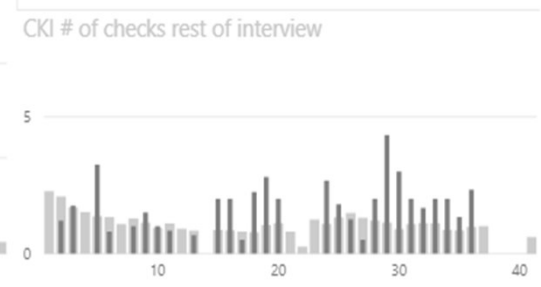
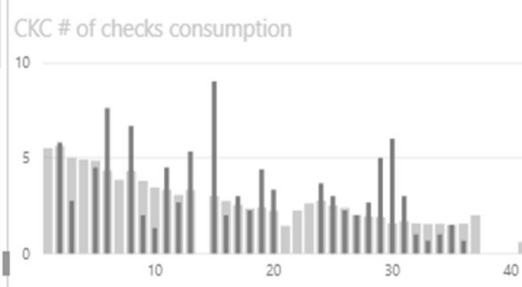
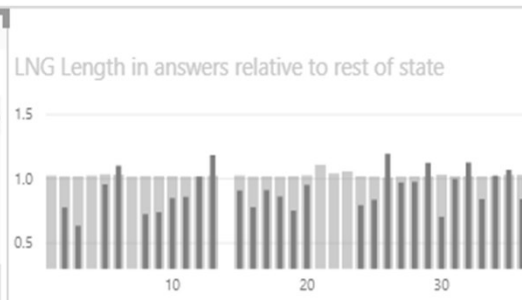
- Buzzword; nice for simultaneously looking at multiple indicators.
- Standardizes what different people in an organization are seeing.
- Can have different levels of access e.g. field supervisor versus survey manager.

Dashboards...

- Separate overall performance metrics and then drill-down by region/interviewer/supervisor etc.
 - Beware of a masking effect (one indicator 'low' + another 'high' -> overall average)
- Some principles to keep in mind:
 - Do not clutter!
 - Focus on at the most 3 metrics on one-screen.
 - Keep related metrics together.
 - Doesn't need to involve expensive systems. Software like Tableau or Microsoft Power BI ('desktop' version is free) or open-source programming languages like R (with its Shiny package for interactive web-based capabilities).

1. Interviewer Risk Index

state_id	IRX	N	LNG	FDD	CKC	CKI	OTH	EET	NOC	UNRA
BayeN1	0.32	89	0.14	0.37	0.06	0.26	0.40	0.36	0.57	0.51
BayeN2	0.36	116	0.12	0.56	0.34	0.01	0.47	0.18	0.72	0.47
BayeN3	0.27	117	0.47	0.18	0.61	0.05	0.74	0.00	0.00	0.00
BayeN3	0.54	117	0.81	0.26	0.97	0.09	0.43	0.55	0.77	0.41
BayeS	0.57	40	0.65	0.48	0.90	0.73	0.86	0.00	0.92	0.00
Benue										
	0.32	90	0.14	0.37	0.06	0.26	0.40	0.36	0.57	0.51
BenuN1	0.49	117	0.62	0.70	0.30	0.15	0.20	0.95	0.67	0.00
BenuN2	0.38	117	0.54	0.16	0.50	0.32	0.24	0.91	0.00	0.41
BenuN3	0.39	117	0.51	0.55	0.26	0.37	0.47	0.68	0.00	0.00
BenuS	0.33	39	0.06	0.23	0.87	0.93	0.50	0.38	0.00	0.00
Borno										
	0.32	51	0.14	0.37	0.06	0.26	0.40	0.36	0.57	0.51
BorN1	0.40	117	0.06	0.28	0.57	0.27	0.59	0.90	0.96	0.00
BorN2	0.69	117	0.77	0.78	0.66	0.81	0.54	0.83	0.98	0.00
BorN3	0.60	116	0.41	0.29	0.59	0.74	0.95	0.96	0.91	0.47
BorS	0.66	39	0.98	0.34	0.55	0.77	0.95	0.75	0.97	0.00
Cross River										
	0.32	90	0.14	0.37	0.06	0.26	0.40	0.36	0.57	0.51
CrossN1	0.29	118	0.08	0.61	0.60	0.53	0.38	0.00	0.00	0.00
CrossN2	0.51	117	0.67	0.55	0.49	0.45	0.39	0.19	0.59	0.52
CrossN3	0.33	115	0.53	0.40	0.68	0.39	0.25	0.17	0.00	0.00
CrossO1	0.44	47	0.74	0.49	0.43	0.10	0.82	0.00	0.00	0.59
CrossO2	0.66	46	0.77	0.72	0.87	0.35	0.99	0.00	0.76	0.63
CrossS	0.29	40	0.41	0.18	0.58	0.04	0.31	0.00	0.77	0.00
Delta										
	0.32	98	0.14	0.37	0.06	0.26	0.40	0.36	0.57	0.51
DeltN1	0.42	117	0.31	0.32	0.78	0.80	0.66	0.17	0.00	0.52
DeltN2	0.66	117	0.89	0.76	0.74	0.70	0.49	0.29	0.59	0.52
DeltN3	0.40	117	0.33	0.09	0.92	0.66	0.55	0.24	0.00	0.81

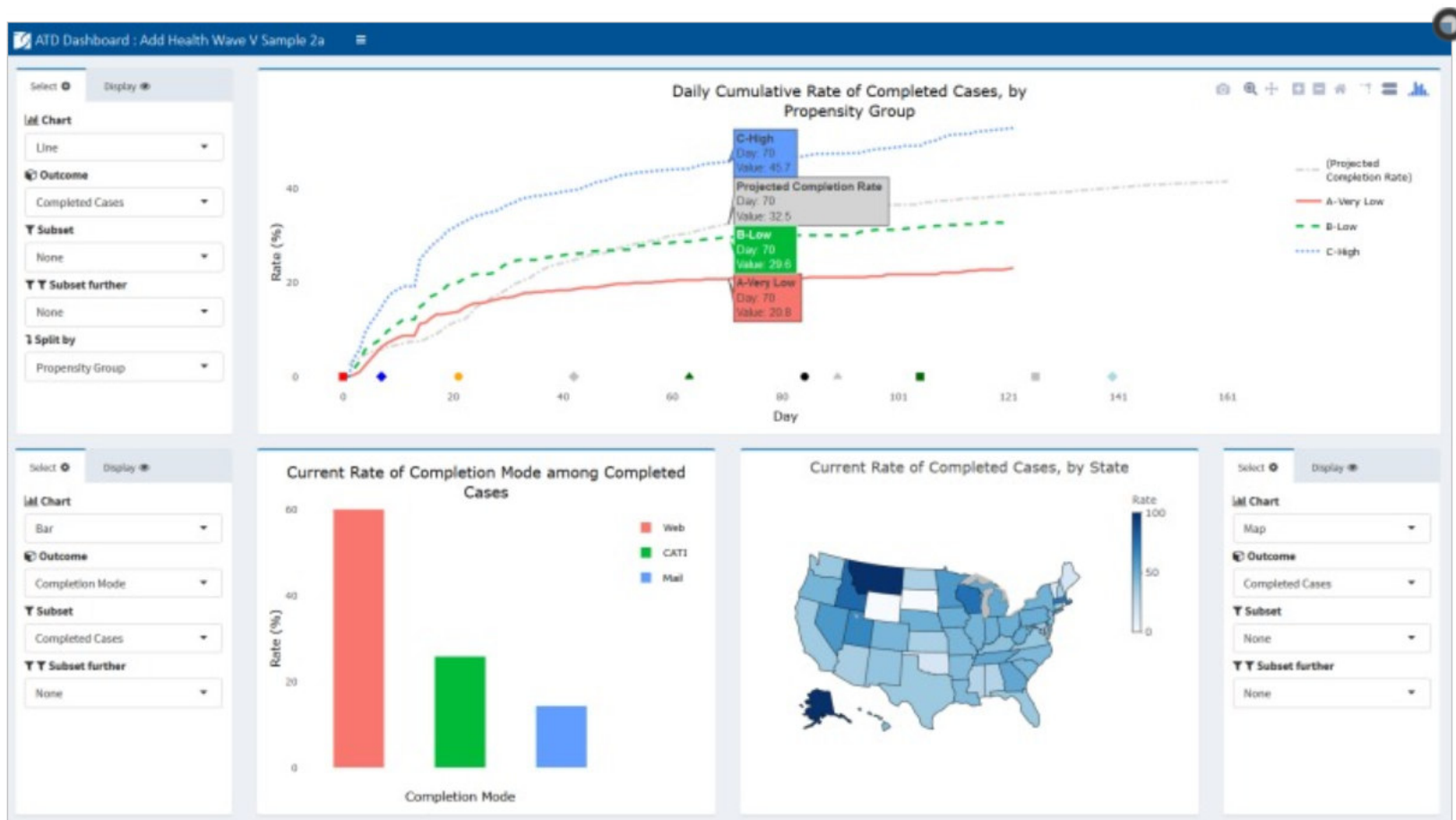


Survey: Nigeria General Household Survey (GHS) Wave 4

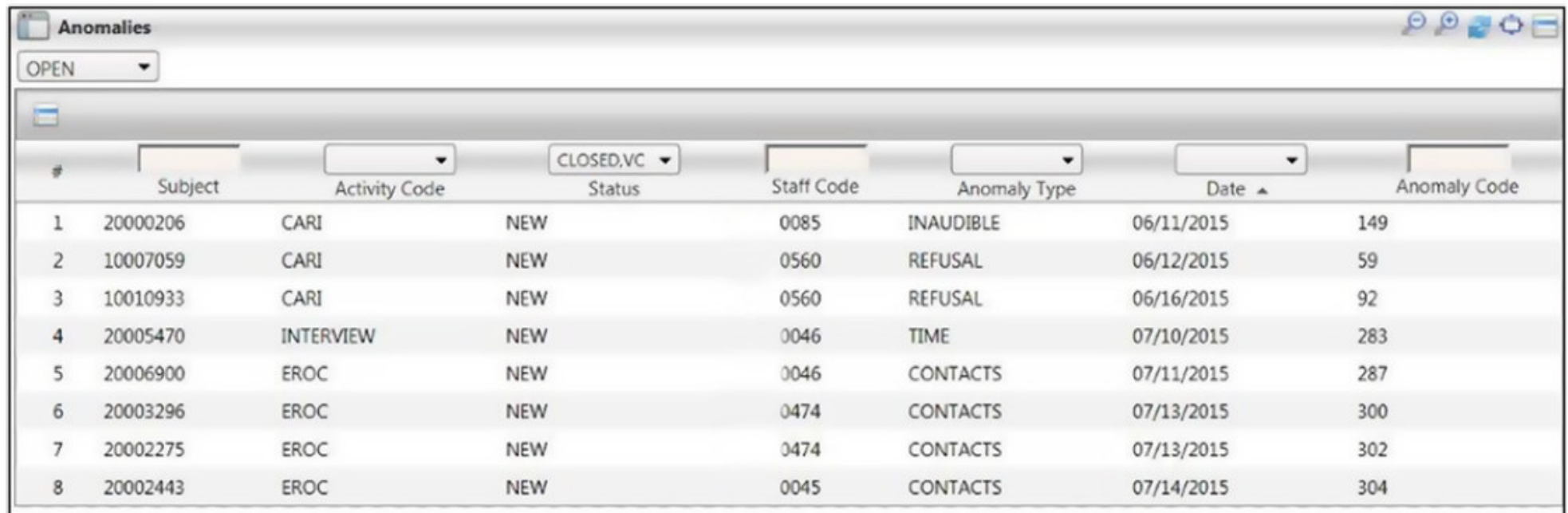
Dashboards/Reports

- Application that produces monitoring reports for high-frequency COVID-19 surveys (based on Survey Solutions):
https://github.com/arthur-shaw/survey_manager_covid#monitor
(in French: https://github.com/arthur-shaw/survey_manager_uemoa)

RTI's adaptive design dashboard using R/Shiny



WESTAT's dashboard alerts



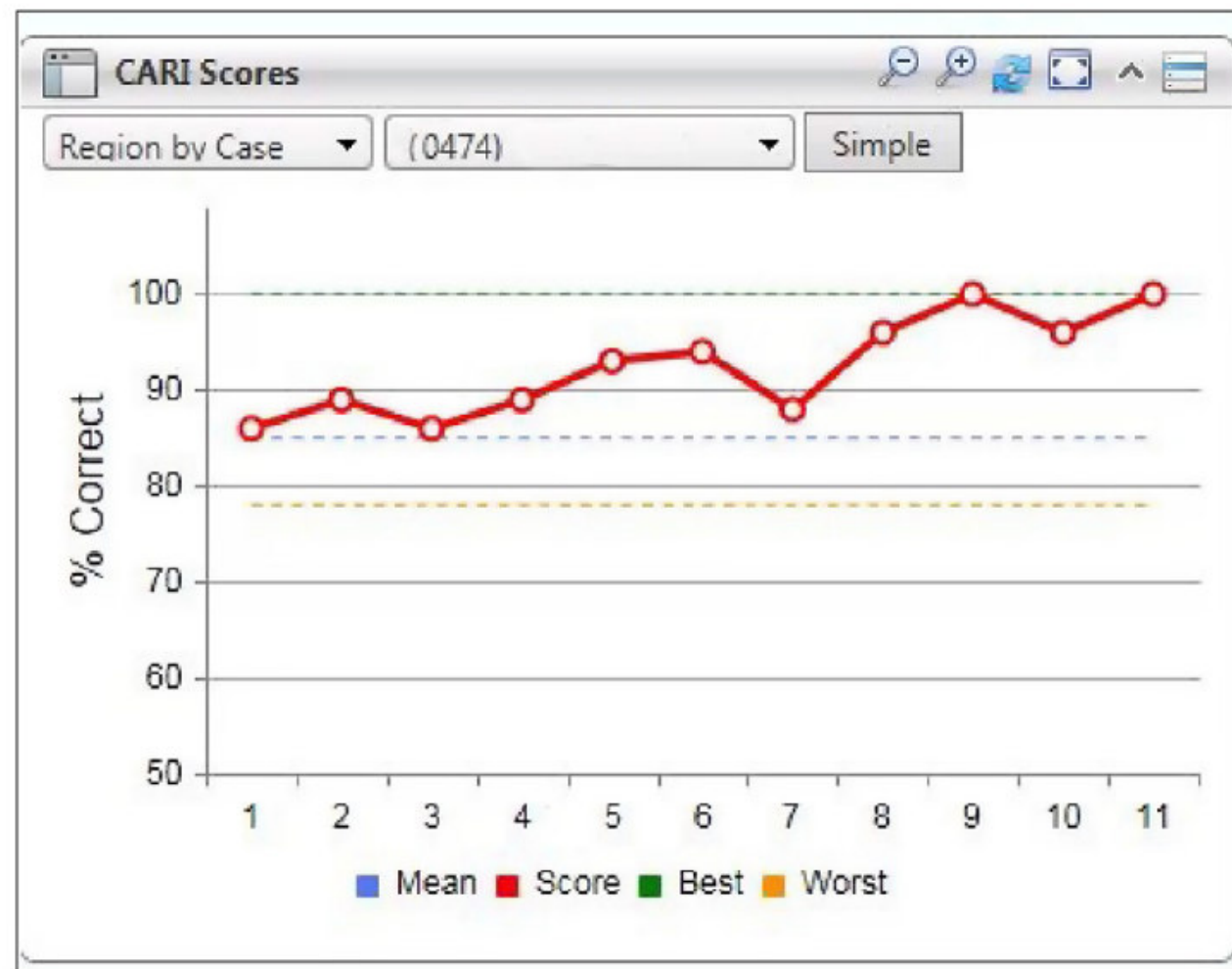
The screenshot shows a web application window titled "Anomalies". At the top left, there is a dropdown menu set to "OPEN". Below this is a table with the following columns: #, Subject, Activity Code, Status (with a "CLOSED,VC" dropdown), Staff Code, Anomaly Type, Date (with an upward arrow), and Anomaly Code. The table contains 8 rows of data.

#	Subject	Activity Code	Status	Staff Code	Anomaly Type	Date	Anomaly Code
1	20000206	CARI	NEW	0085	INAUDIBLE	06/11/2015	149
2	10007059	CARI	NEW	0560	REFUSAL	06/12/2015	59
3	10010933	CARI	NEW	0560	REFUSAL	06/16/2015	92
4	20005470	INTERVIEW	NEW	0046	TIME	07/10/2015	283
5	20006900	EROC	NEW	0046	CONTACTS	07/11/2015	287
6	20003296	EROC	NEW	0474	CONTACTS	07/13/2015	300
7	20002275	EROC	NEW	0474	CONTACTS	07/13/2015	302
8	20002443	EROC	NEW	0045	CONTACTS	07/14/2015	304

Source: *Edwards et al, 2017*

**After
drilldown ->**

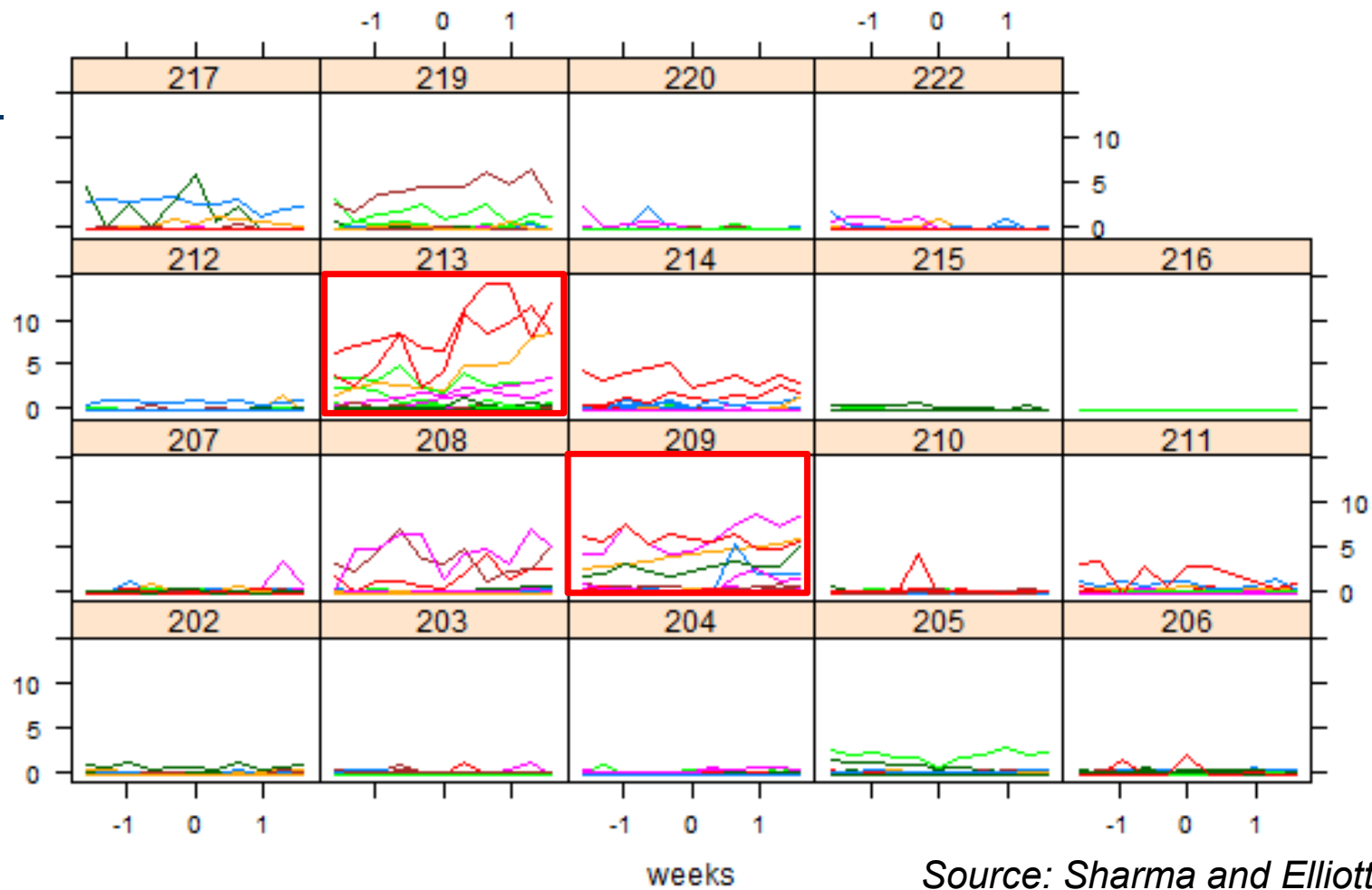
*Observe the Upper and
Lower limits*



Source: Edwards et al, 2017

Graphs also used to identify 'outliers'

- No standard definition of the term 'outlier'
 - Common-man language: "Very far from what we expect"
- Can be a good starting point for exploration
 - Standard usage: Classify all data that are more than 3 standard deviations away from the average as an outlier.
 - Can apply this to individual respondents or to compare interviewers.
 - Must not clean out outliers automatically.
 - Need to document any procedure undertaken.



Source: Sharma and Elliott (2019)

Each panel is an interviewer and each line is the trend for a respondent (data collected across weeks)

Post-identification activities

1. Distinguish between 'Common cause' and 'Special cause'.
 - Did response rates fall due to interviewer-reassignments? Or were there political tensions?
 - Get to the actual cause rather than only address the symptom. What is really causing response rates to fall for a particular interviewer ?
2. Note that errors interact, e.g., Efforts to reduce non-response can increase measurement error. Need judicious decision-making.
3. Specific and frequent verbal feedback tends to improve the quality of interviewers' administration of survey questions. (Edwards et al, 2017)
4. Detect bad habits early : Monitor initial interviews more closely.

Post-identification activities...

5. Distinguish a QC meeting from a meeting on other operational issues.

6. Clarify expectations clearly. Again, get to the cause: “You need to accurately record” rather than “There’s too many rounded numbers being reported”.

7. Mix individual and group feedback.

8. Written feedback helps maintain a record. Track feedback over time. Also, check if the feedback is working!

9. Don’t come across as negative. Also provide positive feedback !

References

- Beaman, L., & Dillon, A. (2012). Do household definitions matter in survey design? Results from a randomized survey experiment in Mali. *Journal of Development Economics*, 98(1), 124–135. doi:10.1016/j.jdeveco.2011.06.005
- Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to Survey Quality*. Wiley Series in Survey Methodology. doi:10.1002/0471458740
- Blasius, J., & Thiessen, V. (2012). Institutional quality control practices. In J. Blasius & V. Thiessen (Eds.), *Assessing the Quality of Survey Data* 57-80. London: Sage.
- Couper, M. P., Holland, L. & Groves, R. M. (1992) Developing systematic procedures for monitoring in a centralized telephone facility. *Journal of Official Statistics*, 8, 63-76
- DeMatteis, J.M., Young, L.Y. , Dahlhamer, J., Langley, R.E., Murphy, J., Olson, K., Sharma S. (2020) Falsification in Surveys, Task force report from the American Association of Public Opinion Research (AAPOR) and the American Statistical Association (ASA). Available at:
https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/AAPOR_Data_Falsification_Task_Force_Report.pdf
- Dupuis, G., Johnson, K.B., & Kaur, J. (2018) .Feed the Future Zone of Influence Surveys: Field Check Tables. Washington, DC: Bureau for Food Security, USAID. Available at: <https://www.agrilinks.org/post/feed-future-zoi-survey-methods>
- Edwards, B., Maitland, A., & Connor, S. (2017). Measurement Error in Survey Operations Management. *Total Survey Error in Practice*, 253–277. doi:10.1002/9781119041702.ch12
- Edwards, B., Schneider, S., & Brick, P. D. (2008). Visual Elements of Questionnaire Design: Experiments with a CATI Establishment Survey. *Advances in Telephone Survey Methodology*, 276–296. doi:10.1002/9780470173404.ch13

References

- Forsman, G., & Schreiner, I. (2004). The Design and Analysis of Reinterview: An Overview. Wiley Series in Probability and Statistics, 279–301. doi:10.1002/9781118150382.ch15
- Fowler, F., & Mangione, T. (1990). Standardized Survey Interviewing. doi:10.4135/9781412985925
- Groves, R.M. & Magilavy, L.J. (1986) “Measuring and Explaining Interviewer Effects in Centralized Telephone Surveys,” Public Opinion Quarterly, Vol. 50, pp. 251-266.
- Hicks, W. D., Edwards, B., Tourangeau, K., McBride, B., Harris-Kojetin, L. D., & Moss, A. J. (2010). Using Cari Tools To Understand Measurement Error. Public Opinion Quarterly, 74(5), 985–1003. doi:10.1093/poq/nfq063
- Holbrook, A. L., Green, M. C., Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. Public Opinion Quarterly, 67, 79–125.
- Hood, C.C. and Bushery, J.M. (1997), "Getting More Bang from the Reinterview Buck: Identifying 'At Risk' Interviewers," Proceedings from Section on Survey Research Methods, American Statistical Association, 1997, pp. 820-824.
- ICF Macro. 2009. Training Field Staff for DHS Surveys. Calverton, Maryland, U.S.A.: ICF Macro. Available at: https://dhsprogram.com/pubs/pdf/DHSM3/Training_Field_Staff_for_DHS_Surveys_Oct2009.pdf
- Kim, Y., Dykema, J., Stevenson, J., Black, P., & Moberg, D. P. (2018). Straightlining: Overview of Measurement, Comparison of Indicators, and Effects in Mail–Web Mixed-Mode Surveys. Social Science Computer Review, 37(2), 214–233. doi:10.1177/0894439317752406
- Kreuter, F. (Ed.). (2013). Improving Surveys with Paradata. doi:10.1002/9781118596869
- Kreuter, F. and C. Casas-Cordero (2010). Paradata. Working paper, German Council for Social and Economic Data.

References

- Loosveldt, G., Carton, A., & Billiet, J. (2004). Assessment of Survey Data Quality: A Pragmatic Approach Focused on Interviewer Tasks. *International Journal of Market Research*, 46(1), 65–82. doi:10.1177/147078530404600101
- Mneimneh, Z., Lyberg, L., Sharma, S., Vyas, M., Sathe, D. B., Malter, F., & Altwajiri, Y. (2018). Case Studies on Monitoring Interviewer Behavior in International and Multinational Surveys. *Advances in Comparative Survey Methods*, 731–770. doi:10.1002/9781118884997.ch35
- Morton, J.E., Mullin, P.A., Biemer, P.P (2008). Using Reinterview and Reconciliation Methods to Design and Evaluate Survey Questions. *Survey Research Methods*. Vol 2, No. 2. pp. 75-82. ISSN 1864-3361
- Mudryk, W., Burgess M. J., & Xiao, P. (1996). Quality control of CATI operations in statistics Canada. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 150–159.
- Murphy JJ, Duprey MA, Chew RF, et al. Interactive Visualization to Facilitate Monitoring Longitudinal Survey Data and Paradata (2019). Research Triangle Park (NC): RTI Press; Available from: <https://www.ncbi.nlm.nih.gov/books/NBK545492/> doi: 10.3768/rtipress.2019.op.0061.1905
- Murphy, J. J., Baxter, R. K., Eyerman, J. D., Cunningham, D., & Kennet, J. (2004). A system for detecting interviewer falsification. Presented at the American Association for Public Opinion Research 59th Annual Conference, Phoenix, AZ.
- Olson, K., & D. Smyth, J. (2020). What Do Interviewers Learn?: Changes in Interview Length and Interviewer Behaviors over the Field Period. *Interviewer Effects from a Total Survey Error Perspective*, 279–292. doi:10.1201/9781003020219-26
- Robbins, M. (2015, February). Preventing data falsification in survey research: Lessons from the Arab barometer. In *New Frontiers in Preventing, Detecting, and Remediating Fabrication in Survey Research*. Conferences hosted by NEAAPOR and Harvard Program on Survey Research, Cambridge, MA (Vol. 13).
- Schräpler, J.-P. (2010). "Benford's Law As an Instrument for Fraud Detection in Surveys Using the Data of the Socio-Economic Panel (SOEP)," *SOEPpapers on Multidisciplinary Panel Data Research* 273, DIW Berlin, The German Socio-Economic Panel (SOEP).

References

- Sharma, S., & Elliott, M. R. (2019). Detecting falsification in a television audience measurement panel survey. *International Journal of Market Research*, 62(4), 432–448. doi:10.1177/1470785319874688
- Sharma, S. (2019). *Paradata, Interviewing Quality, and Interviewer Effects*, Ph.D. dissertation, University of Michigan, Ann Arbor. Available at: <https://deepblue.lib.umich.edu/handle/2027.42/150047>
- Steve, K. W., Burks, A. T., Lavrakas, P. J., Brown, K. D., & Hoover, J. B. (2008). Monitoring Telephone Interviewer Performance. *Advances in Telephone Survey Methodology*, 401–422. doi:10.1002/9780470173404.ch19
- Swanson, D, Cho, M.J., and Eltinge, J. (2003) Detecting Possibly Fraudulent or Error-Prone Survey Data Using Benford's Law. *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Tarnai, J (2007). Monitoring CATI Interviewers. 62nd Annual Conference, American Association for Public Opinion Research, Anaheim, California. Available at: <https://opinion.wsu.edu/tarnai/paper/AAPOR07%20Monitoring%20CATI%20Interviewers%20Paper.pdf>
- Tourangeau, R., Kreuter, F., & Eckman, S. (2012). Motivated Underreporting in Screening Interviews. *Public Opinion Quarterly*, 76(3), 453–469. doi:10.1093/poq/nfs033
- West, Brady T., and Robert M. Groves (2013). "The PAIP Score: A Propensity-Adjusted Interviewer Performance Indicator." *Public Opinion Quarterly*, 77(1): 352-374.

END OF Module 6

Thank you !