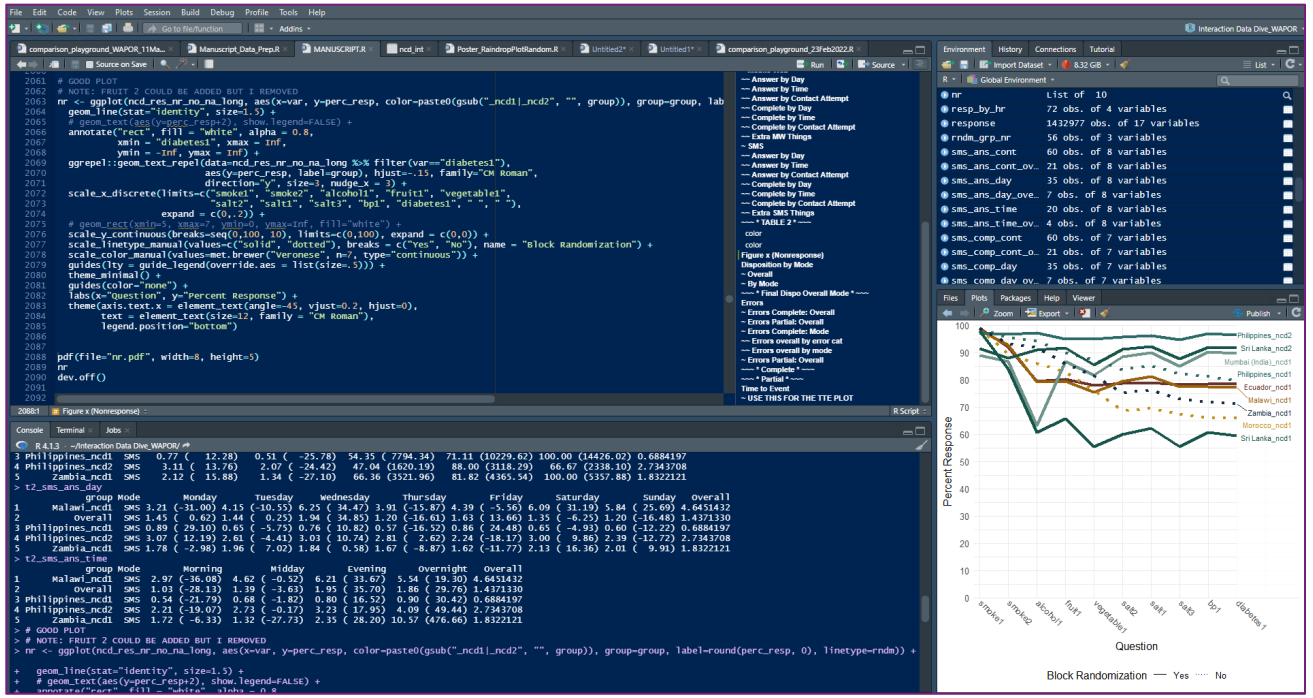
Using Mobile Phone Survey Paradata to Evaluate and Improve Processes: Lessons Learned and Best Practices from 9 Surveys in 7 Sites

Casey Siesel, MPH, Health Scientist
csiesel@cdc.gov

U.S. Centers for Disease Control and Prevention, Center for Global Health, Division of Global Health Prevention, Office of Global Noncommunicable Diseases



75th Annual World Association for Public Opinion Research Conference
November 13th, 2022

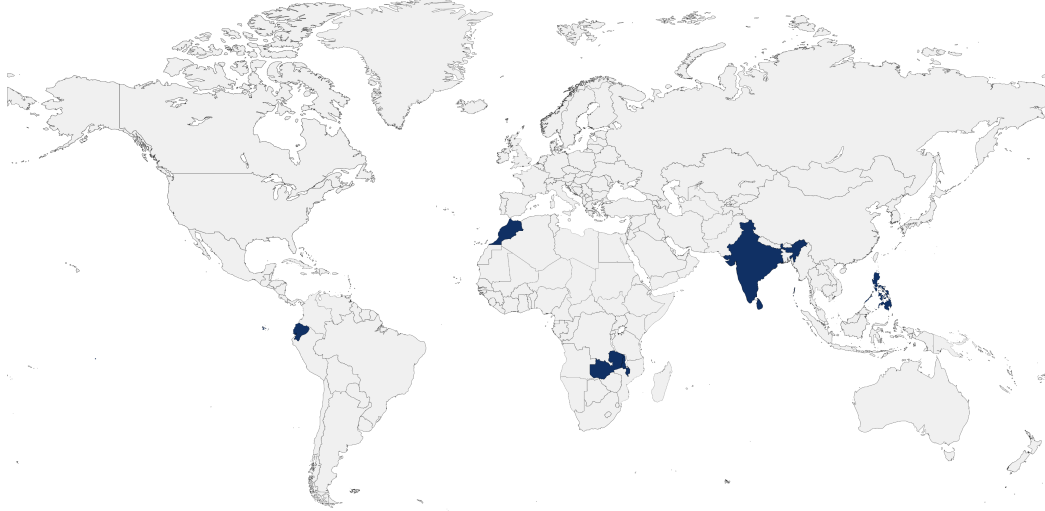
The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Bloomberg Philanthropies Data for Health Initiative

Mobile Phone Survey Background

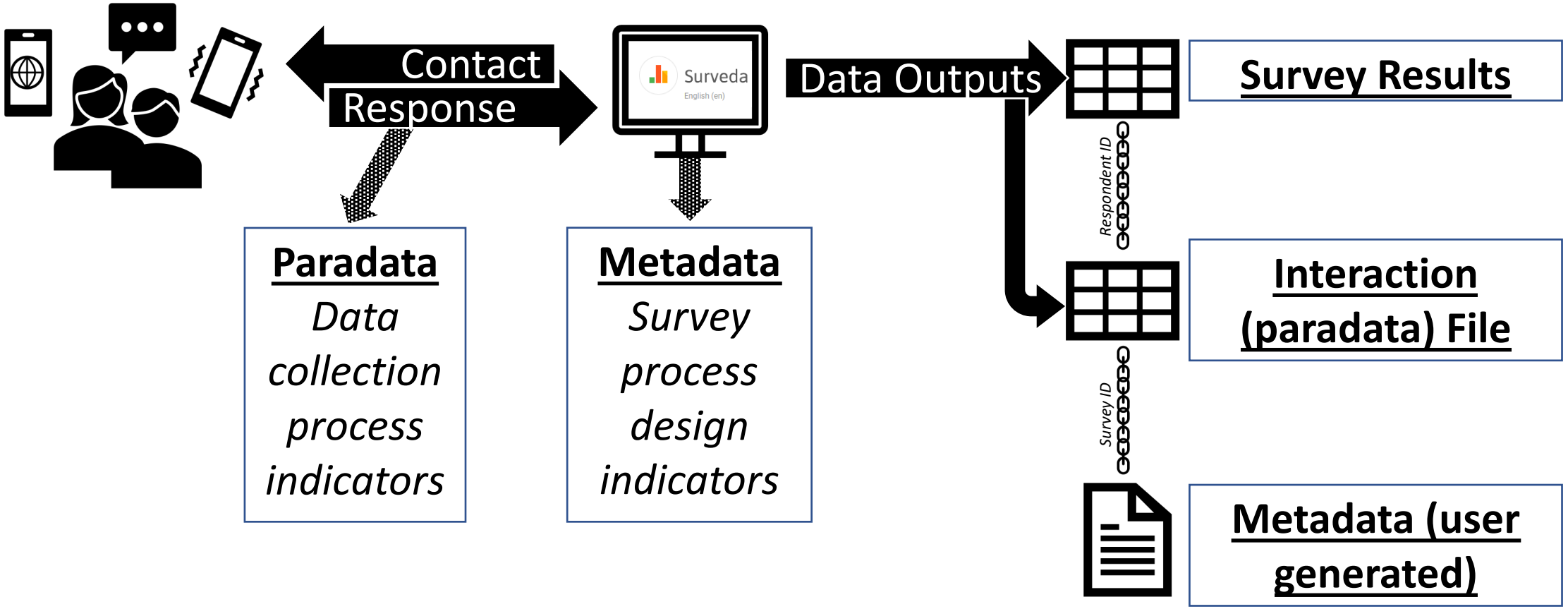
NCD MPS Overview

- First noncommunicable disease (NCD) mobile phone survey (MPS) launched in 2017
  - To-date, over *3.2 million phone numbers* have been contacted which has contributed to *37,750 completed interviews* representing over *243 million people*
- Fully automated, self-administered survey on behaviors and NCD risk factors for adults using random digit dialing
- Data is managed by Surveda, an open source, platform agnostic survey tool that can administer surveys via multiple modes (IVR, SMS, Mobile Web)



Site	Completed Interviews (Complete and partial)	Paradata Records
Ecuador	3104	658,184
Malawi	5816	1,944,633
Morocco	3528	2,625,628
Mumbai (India)	3432	5,082,487
Philippines x 2	3674 / 3103	7,346,148 / 1,683,509
Sri Lanka x 2	4656 / 4373	1,657,400 / 1,937,264
Zambia	6064	2,996,505
Overall	**37,750**	**25,931,758**

NCD MPS Data Flow



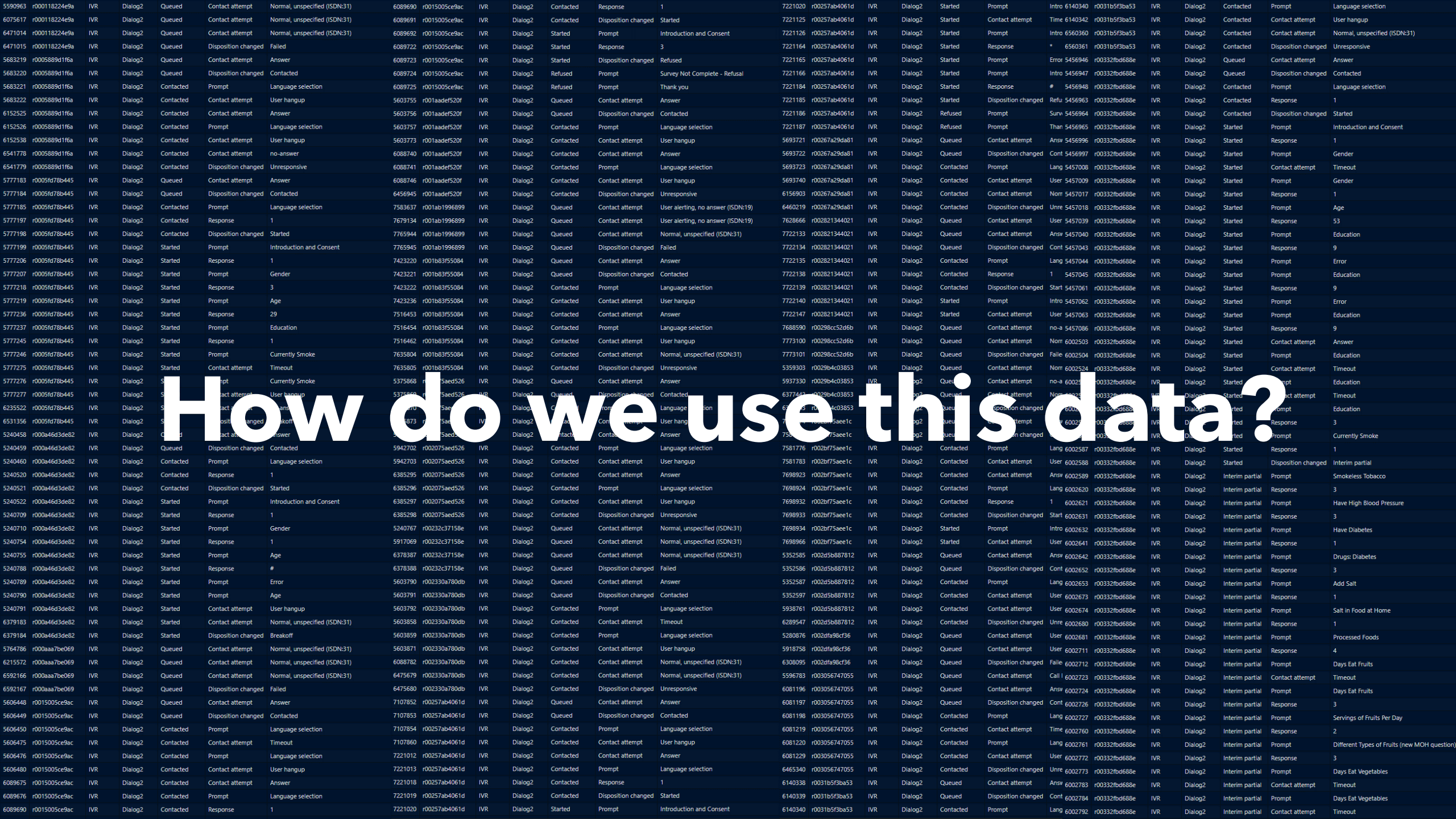
Paradata Description

Interaction File Output

“The Interactions audit file is the complete record of all actions for each phone number. It includes everything from the disposition history file, but with additional metadata. This file includes metadata, or data describing the response, of each response and corresponding contact attempt and mode.”

Interaction File for *re412ab0748f3*

	ID	Respondent ID	Mode	Channel	Disposition	Action Type	Action Data	Timestamp
1	11162380	re412ab0748f3	IVR	Dialog2	Queued	Contact attempt	Enqueueing call	2021-12-13 15:14:25 GMT+6
2	11228940	re412ab0748f3	IVR	Dialog2	Queued	Contact attempt	Answer	2021-12-15 09:18:43 GMT+6
3	11228941	re412ab0748f3	IVR	Dialog2	Queued	Disposition changed	Contacted	2021-12-15 09:18:43 GMT+6
4	11228942	re412ab0748f3	IVR	Dialog2	Contacted	Prompt	Language selection	2021-12-15 09:18:43 GMT+6
5	11228968	re412ab0748f3	IVR	Dialog2	Contacted	Contact attempt	Timeout	2021-12-15 09:19:19 GMT+6
6	11228969	re412ab0748f3	IVR	Dialog2	Contacted	Prompt	Language selection	2021-12-15 09:19:19 GMT+6
7	11228986	re412ab0748f3	IVR	Dialog2	Contacted	Contact attempt	Use	
8	11375314	re412ab0748f3	IVR	Dialog2	Contacted	Contact attempt	Ans	
9	11375315	re412ab0748f3	IVR	Dialog2	Contacted	Prompt	Lang	
10	11375359	re412ab0748f3	IVR	Dialog2	Contacted	Contact attempt	Tim	
11	11375360	re412ab0748f3	IVR	Dialog2	Contacted	Prompt	Lang	
12	11375378	re412ab0748f3	IVR	Dialog2	Contacted	Response	1	
13	11375379	re412ab0748f3	IVR	Dialog2	Contacted	Disposition changed	Star	
14	11375380	re412ab0748f3	IVR	Dialog2	Started	Prompt	Intro	
15	11375427	re412ab0748f3	IVR	Dialog2	Started	Response	1	
16	11375428	re412ab0748f3	IVR	Dialog2	Started	Prompt	Gen	
17	11375432	re412ab0748f3	IVR	Dialog2	Started	Response	2	
18	11375433	re412ab0748f3	IVR	Dialog2	Started	Prompt	Erro	
19	11375434	re412ab0748f3	IVR	Dialog2	Started	Prompt	Gen	
20	11375450	re412ab0748f3	IVR	Dialog2	Started	Response	3	
21	11375451	re412ab0748f3	IVR	Dialog2	Started	Prompt	Age	
22	11375494	re412ab0748f3	IVR	Dialog2	Started	Response	46	
23	11375495	re412ab0748f3	IVR	Dialog2	Started	Prompt	Edu	
24	11375530	re412ab0748f3	IVR	Dialog2	Started	Response	6	
25	11375531	re412ab0748f3	IVR	Dialog2	Started	Prompt	Add	
26	11375600	re412ab0748f3	IVR	Dialog2	Started	Response	5	
27	11375601	re412ab0748f3	IVR	Dialog2	Started	Prompt	Salt	
28	11375626	re412ab0748f3	IVR	Dialog2	Started	Response	3	
29	11375627	re412ab0748f3	IVR	Dialog2	Started	Prompt	Processed Foods	2021-12-19 15:13:11 GMT+6
30	11375670	re412ab0748f3	IVR	Dialog2	Started	Response	4	2021-12-19 15:13:58 GMT+6
31	11375671	re412ab0748f3	IVR	Dialog2	Started	Prompt	Have High Blood Pre...	2021-12-19 15:13:58 GMT+6
32	11375692	re412ab0748f3	IVR	Dialog2	Started			
33	11375693	re412ab0748f3	IVR	Dialog2	Started			
34	11375716	re412ab0748f3	IVR	Dialog2	Started			
35	11375717	re412ab0748f3	IVR	Dialog2	Started			
36	11375718	re412ab0748f3	IVR	Dialog2	Interim partial			
37	11375729	re412ab0748f3	IVR	Dialog2	Interim partial			
38	11375730	re412ab0748f3	IVR	Dialog2	Interim partial			
39	11375768	re412ab0748f3	IVR	Dialog2	Interim partial			
40	11375769	re412ab0748f3	IVR	Dialog2	Interim partial			
41	11375770	re412ab0748f3	IVR	Dialog2	Interim partial			
42	11375771	re412ab0748f3	IVR	Dialog2	Interim partial			
43	11375830	re412ab0748f3	IVR	Dialog2	Interim partial			
44	11375831	re412ab0748f3	IVR	Dialog2	Interim partial			
45	11375832	re412ab0748f3	IVR	Dialog2	Interim partial			
46	11375833	re412ab0748f3	IVR	Dialog2	Interim partial			
47	11375855	re412ab0748f3	IVR	Dialog2	Interim partial			
48	11375856	re412ab0748f3	IVR	Dialog2	Interim partial			
49	11375867	re412ab0748f3	IVR	Dialog2	Interim partial			
50	11375868	re412ab0748f3	IVR	Dialog2	Interim partial			
51	11375902	re412ab0748f3	IVR	Dialog2	Interim partial			
52	11375903	re412ab0748f3	IVR	Dialog2	Interim partial			
53	11375919	re412ab0748f3	IVR	Dialog2	Interim partial			
54	11375920	re412ab0748f3	IVR	Dialog2	Interim partial			
55	11375955	re412ab0748f3	IVR	Dialog2	Interim partial			
56	11375956	re412ab0748f3	IVR	Dialog2	Interim partial			
64	11376065	re412ab0748f3	IVR	Dialog2	Interim partial	Prompt	Different Types of Fru...	2021-12-19 15:19:18 GMT+6
65	11376070	re412ab0748f3	IVR	Dialog2	Interim partial	Response	1	2021-12-19 15:19:24 GMT+6
66	11376071	re412ab0748f3	IVR	Dialog2	Interim partial	Prompt	Days Eat Vegetables	2021-12-19 15:19:24 GMT+6
67	11376089	re412ab0748f3	IVR	Dialog2	Interim partial	Contact attempt	Timeout	2021-12-19 15:19:43 GMT+6
68	11376090	re412ab0748f3	IVR	Dialog2	Interim partial	Prompt	Days Eat Vegetables	2021-12-19 15:19:43 GMT+6
69	11376091	re412ab0748f3	IVR	Dialog2	Interim partial	Response	7	2021-12-19 15:19:44 GMT+6
70	11376092	re412ab0748f3	IVR	Dialog2	Interim partial	Prompt	Servings of Vegetabl...	2021-12-19 15:19:44 GMT+6
71	11376137	re412ab0748f3	IVR	Dialog2	Interim partial	Contact attempt	Timeout	2021-12-19 15:20:16 GMT+6
72	11376138	re412ab0748f3	IVR	Dialog2	Interim partial	Prompt	Servings of Vegetabl...	2021-12-19 15:20:16 GMT+6
73	11376149	re412ab0748f3	IVR	Dialog2	Interim partial	Response	2	2021-12-19 15:20:28 GMT+6
74	11376150	re412ab0748f3	IVR	Dialog2	Interim partial	Prompt	Different Types of Ve...	2021-12-19 15:20:28 GMT+6
75	11376166	re412ab0748f3	IVR	Dialog2	Interim partial	Contact attempt	Timeout	2021-12-19 15:20:44 GMT+6
76	11376167	re412ab0748f3	IVR	Dialog2	Interim partial	Prompt	Different Types of Ve...	2021-12-19 15:20:44 GMT+6
77	11376216	re412ab0748f3	IVR	Dialog2	Interim partial	Response	6	2021-12-19 15:21:03 GMT+6
78	11376217	re412ab0748f3	IVR	Dialog2	Interim partial	Prompt	Phone Numbers	2021-12-19 15:21:03 GMT+6
79	11376229	re412ab0748f3	IVR	Dialog2	Interim partial	Contact attempt	Timeout	2021-12-19 15:21:18 GMT+6
80	11376230	re412ab0748f3	IVR	Dialog2	Interim partial	Prompt	Phone Numbers	2021-12-19 15:21:18 GMT+6
81	11376248	re412ab0748f3	IVR	Dialog2	Interim partial	Contact attempt	Timeout	2021-12-19 15:21:33 GMT+6
82	11376249	re412ab0748f3	IVR	Dialog2	Interim partial	Prompt	Phone Numbers	2021-12-19 15:21:33 GMT+6
83	11376264	re412ab0748f3	IVR	Dialog2	Interim partial	Response	1	2021-12-19 15:21:43 GMT+6
84	11376265	re412ab0748f3	IVR	Dialog2	Interim partial	Prompt	Share Phone Number	2021-12-19 15:21:43 GMT+6
85	11376287	re412ab0748f3	IVR	Dialog2	Interim partial	Contact attempt	Timeout	2021-12-19 15:22:03 GMT+6
86	11376288	re412ab0748f3	IVR	Dialog2	Interim partial	Prompt	Share Phone Number	2021-12-19 15:22:03 GMT+6
87	11376308	re412ab0748f3	IVR	Dialog2	Interim partial	Response	1	2021-12-19 15:22:27 GMT+6
88	11376309	re412ab0748f3	IVR	Dialog2	Interim partial	Disposition changed	Completed	2021-12-19 15:22:27 GMT+6
89	11376310	re412ab0748f3	IVR	Dialog2	Completed	Prompt	Complete Message	2021-12-19 15:22:27 GMT+6
90	11376311	re412ab0748f3	IVR	Dialog2	Completed	Prompt	Thank you	2021-12-19 15:22:27 GMT+6


How do we use this data?

Metrics and Key Performance Indicators

- Evaluate performance of survey and Surveda settings (modes, recontact attempts, timing)
  - Surveda level: Response rate, time per interview, data entry errors, time of responses
  - Survey level: contact attempts, proportions of dispositions, progress towards quotas, item nonresponse

id	respondent_id	mode	channel	disposition	action_type	action_data	timestamp	mno	idx	cntry	srvy

Timing

Question: Are there days, times, and contact attempts that people are more likely to engage with MPS?

respondent_id → filter → action_type → filter → action_data → sort by → id → group by → timestamp

Stratify by: disposition, cntry, srvy, mode

Mode Effect

Question: How does survey mode impact engagement?

respondent_id → group by → disposition → group by → mode

Stratify by: cntry, srvy

Question Order

Question: What is the impact of question block randomization on nonresponse?

respondent_id → filter → disposition → filter → action_type → group by → action_data → filter → action_type → group by → action_data

Stratify by: cntry, srvy, mode

Time to Event

Question: Are there thresholds for time of engagement for participants by disposition?

respondent_id → filter → action_type → select → timestamp → filter → action_type → select → timestamp → group by → disposition

Stratify by: cntry, srvy, mode

Errors

Question: Are MPS participants facing issues when responding to questions?

respondent_id → filter → disposition → filter → action_type → group by → action_data → lead → action_type → group by → action_data

Stratify by: disposition, cntry, srvy, mode

id	respondent_id	mode	channel	disposition	action_type	action_data	timestamp	mno	idx	cntry	srvy

Timing

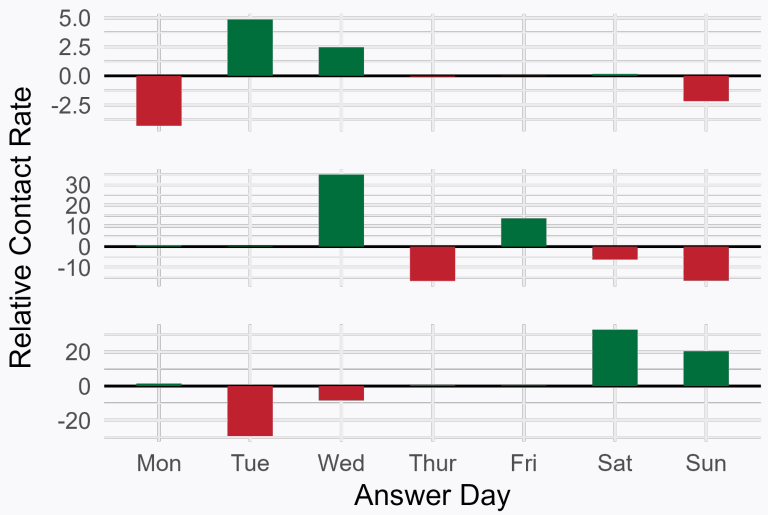
Question: Are there days, times, and contact attempts that people are more likely to engage with MPS?

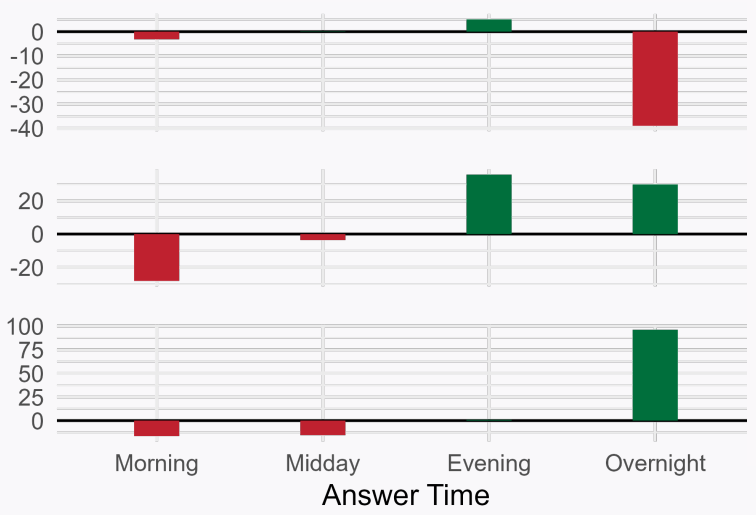
respondent_id → **filter** → action_type → **filter** → action_data → **sort by** → id → **group by** → timestamp

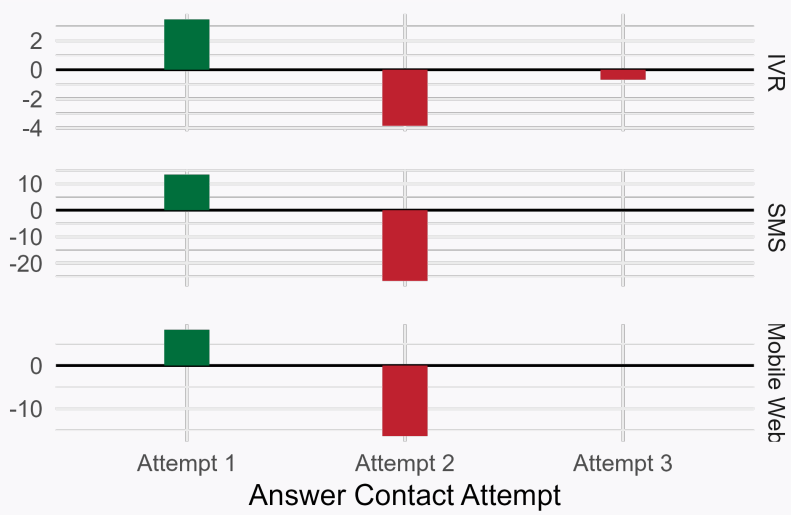
Stratify by: disposition, cntry, srvy, mode

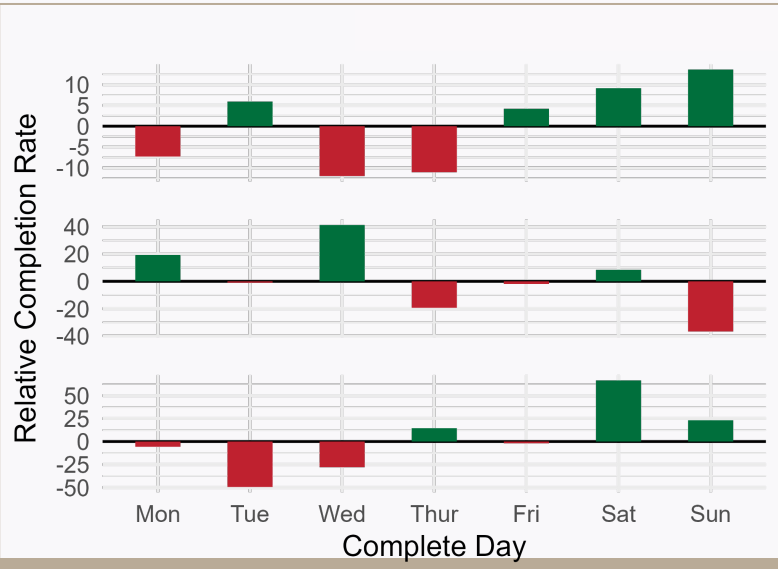
Timing

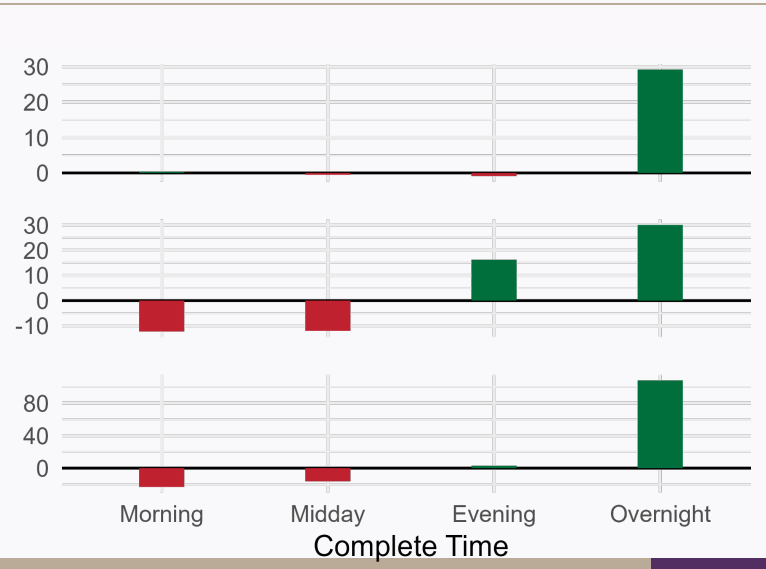
Contact and completion rates vary by mode, day and time. Contact and completion rates are highest on the initial contact attempt for all modes.

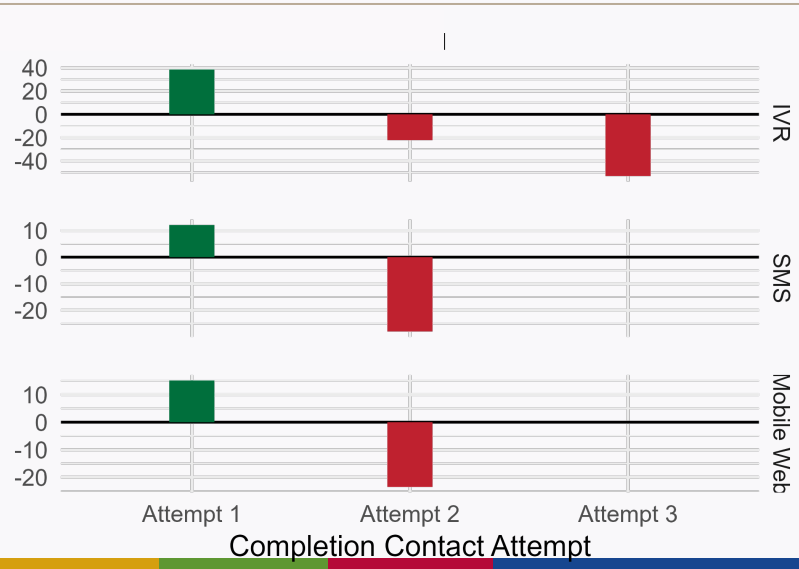












id	respondent_id	mode	channel	disposition	action_type	action_data	timestamp	mno	idx	cntry	srvy

Mode Effect

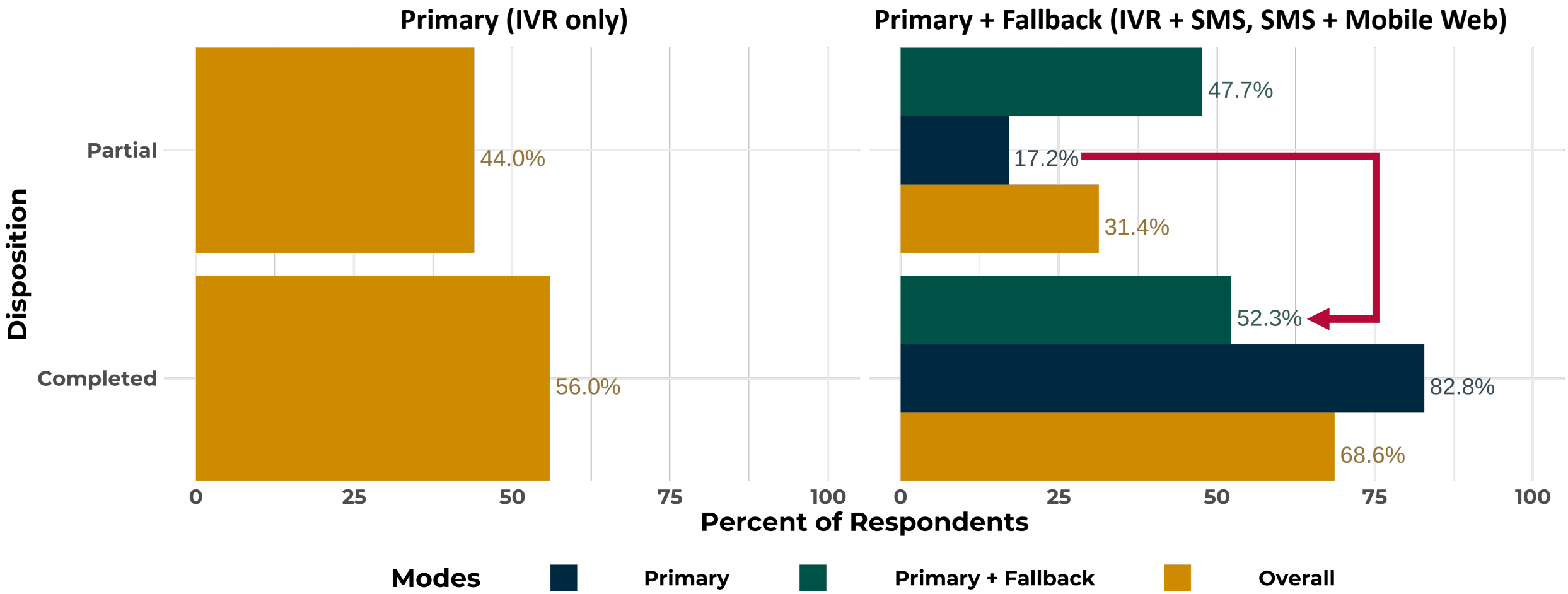
Question: How does survey mode impact engagement?

respondent_id → group by → disposition → group by → mode

Stratify by: cntry srvy

Mode Effect

MPS that include a fallback mode have a 72.4% higher ratio of completed to partially completed surveys.



id	respondent_id	mode	channel	disposition	action_type	action_data	timestamp	mno	idx	cntry	srvy

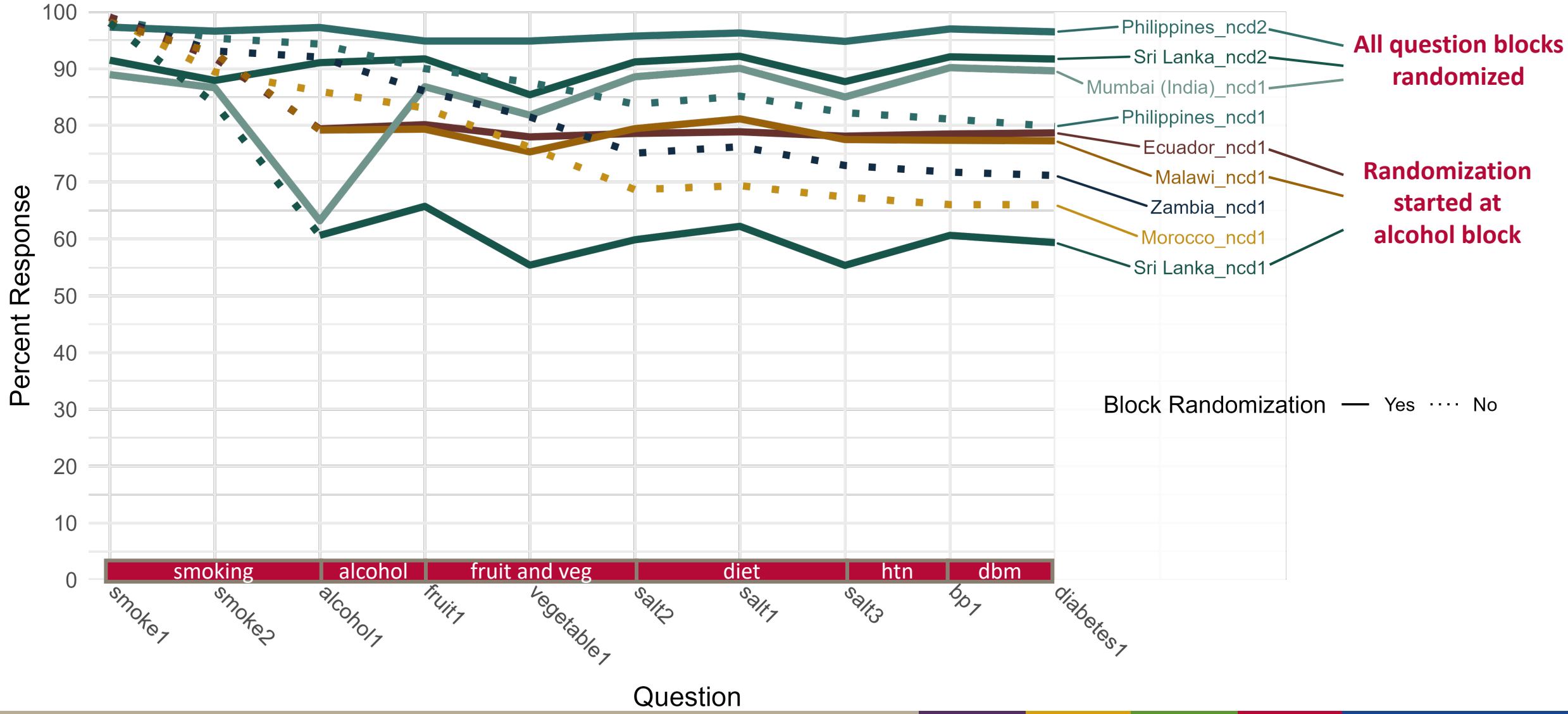
Question Order

Question: What is the impact of question block randomization on nonresponse?

Stratify by: cntry srvy mode

Question Order

MPS that utilized block randomization had 9.8% less nonresponse at the final question block (diabetes1) when compared to those without randomization.



id	respondent_id	mode	channel	disposition	action_type	action_data	timestamp	mno	idx	cntry	srvy

Time to Event

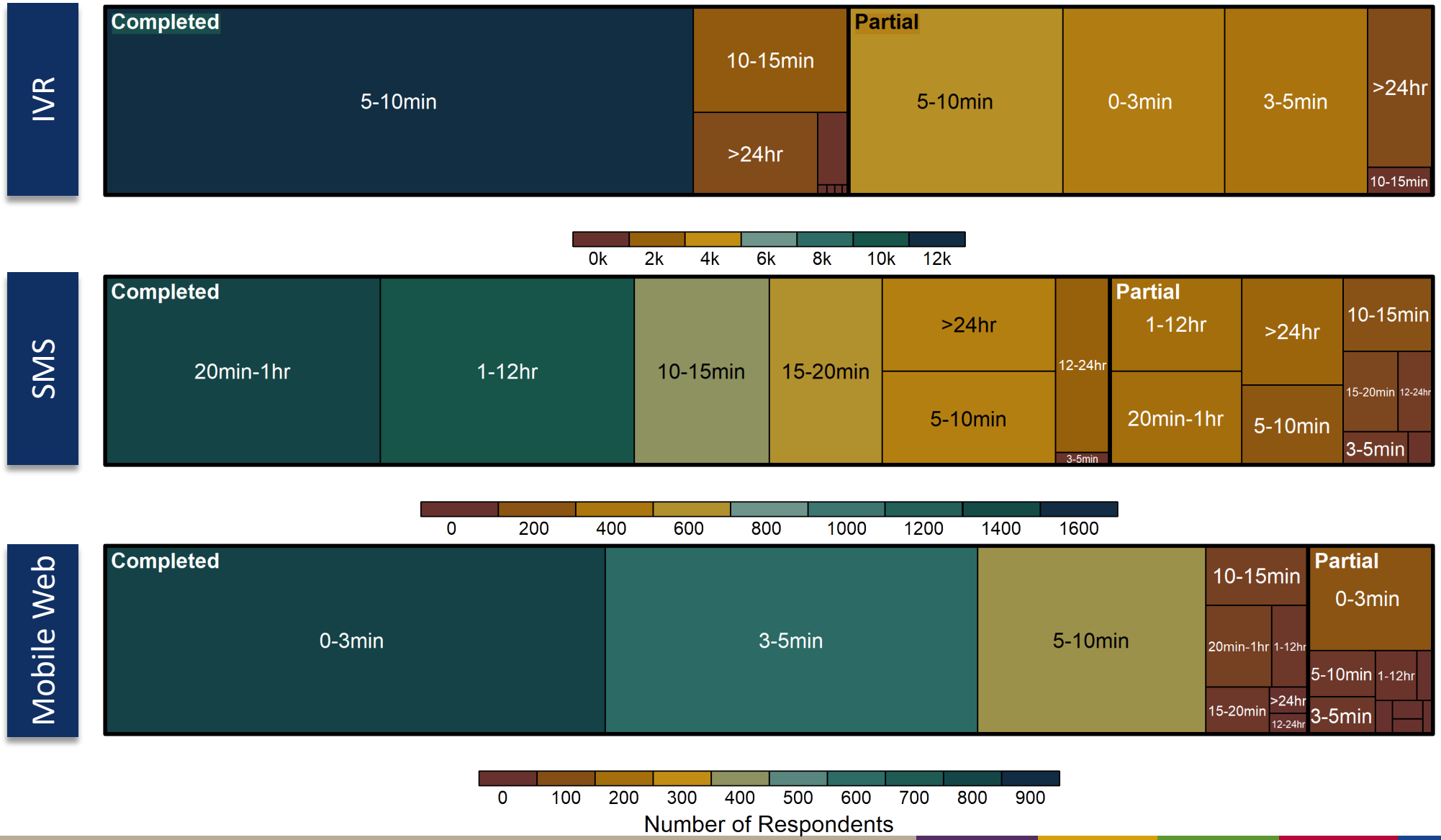
Question: Are there thresholds for time of engagement for participants by disposition?

respondent_id → **filter** → action_type → **select** → timestamp → **filter** → action_type → **select** → timestamp → **group by** → disposition

Stratify by: cntry srvy mode

Time to Event

Time to event varies by mode with greater variation observed in SMS and Mobile Web MPS



id	respondent_id	mode	channel	disposition	action_type	action_data	timestamp	mno	idx	cntry	srvy

Errors

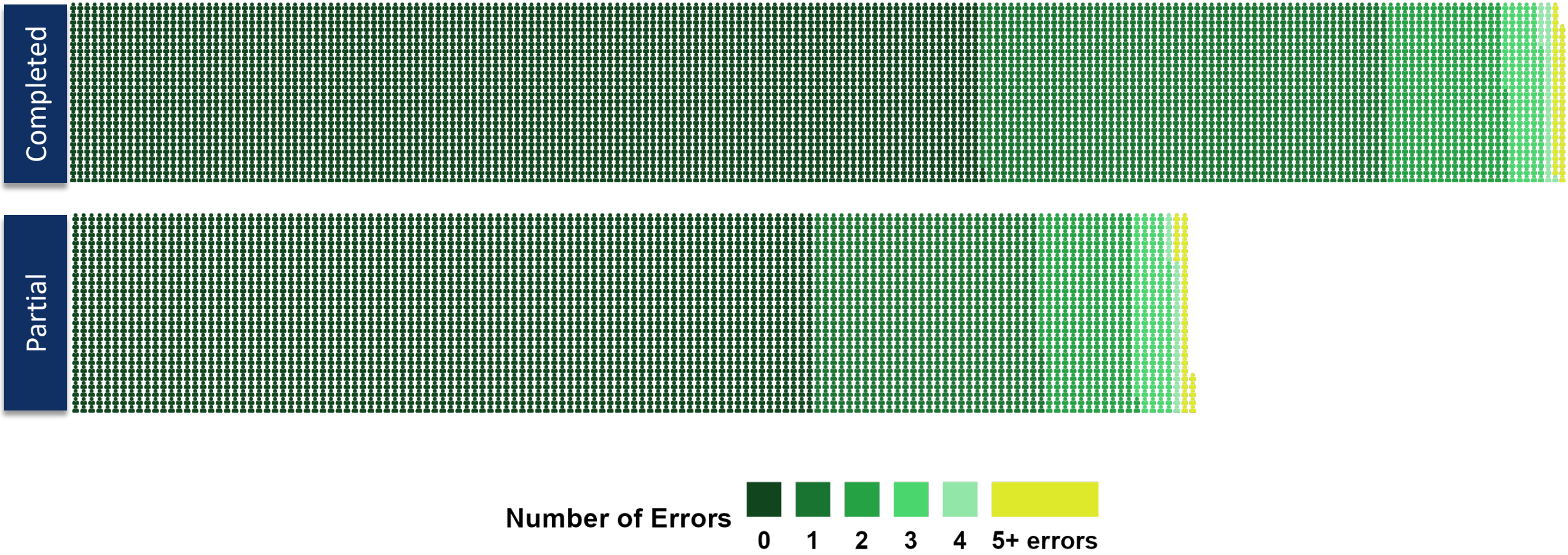
Question: Are MPS participants facing issues when responding to questions?

respondent_id → **filter** → disposition → **filter** → action_type → **group by** → action_data → **lead** → action_type → **group by** → action_data

Stratify by: disposition | cntry | srvy | mode

Errors

Over 60% and 66% of participants who completed or partially completed MPS did so with no errors in data entry



Lessons Learned & Best Practices

Lessons Learned & Best Practices

1. Paradata datasets offers big opportunities…if there is strong commitment

- Formalize paradata analyses in workplans
- Utilize exploratory visual data analyses to identify trends

Lessons Learned & Best Practices

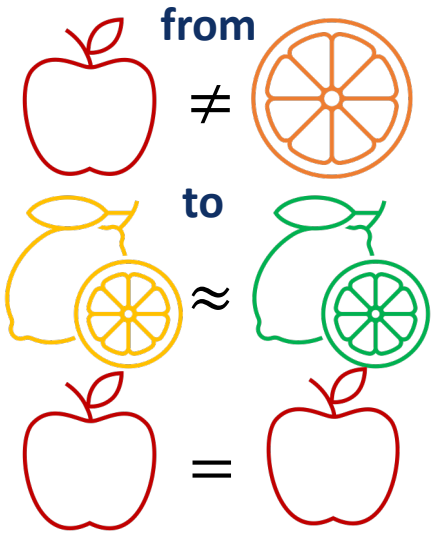
2. The front-end investment in paradata analyses is intensive (time, learning curve, computing power)…but pays dividends down the line

- Utilize a ‘building blocks’ approach to coding – distinct chunks for separate tasks which combine to tell a story

Lessons Learned & Best Practices

3. Small differences in survey design can introduce big challenges for data manipulation and comparisons

- Standardize approaches whenever and wherever possible (variable names, data formats, etc.)



Lessons Learned & Best Practices

4. Real-time paradata analysis is an invaluable tool for ongoing process evaluations and course-correcting during data collection

- Operationalize early and often
- Connect directly to data streams via APIs or web links
- Look at data through different lenses

Thank you!

For additional questions, contact me at csiesel@cdc.gov

For more information, contact CDC
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Funding for the Data for Health Initiative and NCD Mobile Phone Survey is provided by the Bloomberg Philanthropies and the Government of Australia, Foreign Affairs.