

Uses and Applications of Paradata in Web and Mixed-Mode Surveys

Fedcasic

April 2019



100

STATISTICS CANADA

ONE HUNDRED YEARS AND COUNTING



Statistics
Canada

Statistique
Canada

Canada

Outline

- Paradata Sources and Auxiliary Information
- Uses and Applications of Paradata
 - Before, During and After Data Collection
 - Some Recent Examples
- Current and Upcoming Research Projects



100



Paradata Sources

Main Paradata Sources Available

- Transaction files
 - CATI: call / contact information
 - CAPI: visit / call / contact information
 - Web: type of actions & activities (e.g., login, submit, bounce back)
 - Multi-mode: call / actions & activities
 - Key strokes
 - CATI and CAPI : Interviewer key strokes (audit trail)
 - Web: Interviewer (I-EQ) & Respondents (R-EQ) key strokes (IIS-log)
 - Interviewer payroll information
-
- *Automatically recorded for all StatCan surveys*
 - *Different content by collection mode*
 - *Timely available - Updated on daily basis*



100

Auxiliary information

Analysis is very often enhanced in using

- Planning assumptions
 - Target response rate, expected hit rate and EQ completion rate
- Planned budget
 - Number of hours and money(\$)
- Production plan
 - Plan that describes how collection hours are spent throughout collection period including collection milestones
- Sample design (e.g., stratum, geography, sample unit characteristics)
- Paradata from previous cycle(s) and /or Paradata from paradata (e.g., call delivery system parameters, case priority flag)



100



Examples of use Before collection

- Realistic assumptions and budget costing
- Info sheet and production plans
 - Important input to the Resource Allocation System during collection
 - Process all surveys simultaneously
 - Use to allocate and assign automatically interviewers
- Planning the overall collection strategy (e.g., EQ reminder strategy)
- Responsive Collection Design (RCD)
 - RCD strategy and propensity models
- Plan and set-up active managements tools
 - General and survey specific tools and milestones



100





Info sheet

- Planning and target assumptions used to create the production plan
- Assess if the budget makes sense according to the assumptions

Production Plan Info Sheet - 1 per cycle

Survey Info		
Item	Value	Comments
Survey Name	National Cannabis Survey	WAVE 4
Project Code	PE: 1697	
Cycle #	TBD	
Collection Start Date	13/11/2018	
Collection End Date	10/12/2018	
Interview time (minutes)	15.5	Time it takes to go through the roster and questionnaire, once somebody has
Total Sample	12,000	
Hit Rate	100%	1 - % voids

CATI operation #1* (i.e. 'Cases that start in I-EQ')		
Item	Value	Comments
CATI Start Date	December-02-18	
CATI End Date	December-10-18	
Sample Size	8,620	
CATI operation #2* (i.e. 'Cases that start in R-EQ')		
Item	Value	Comments
CATI Start Date		
CATI End Date		
Sample Size		

* If there are more than 2 CATI operations, please fill in the above information for all of them.

RO Info	Budgeted System Hours *	Collection Hours **	CATI Completes
Total	1841	2454	2,964
SHE			
STU	1233	1644	1986
HAL			
WIN			
EDM	608	810	978

EQ portion (if applicable)		
Item	Value	Comments
EQ start date	13/11/2018	
EQ end date	10/12/2018	
EQ Sample	12,000	# of EQ invites
EQ Completes	3,380	25% uptake

Production Plan

- Plan that describes how collection hours are spent throughout collection period including collection milestones



Cannabis Survey - Production Plan												
Date	Weekday										TOTAL Responses	
		STU	EDM	NAT	STU	EDM	NAT	STU	EDM	NAT	NAT	
13/11/2018	Tue	0	0	0	0	0	0	0	0	0	89	
14/11/2018	Wed	0	0	0	0	0	0	0	0	0	255	
15/11/2018	Thu	0	0	0	0	0	0	0	0	0	488	
16/11/2018	Fri	0	0	0	0	0	0	0	0	0	677	
17/11/2018	Sat	0	0	0	0	0	0	0	0	0	802	
18/11/2018	Sun	0	0	0	0	0	0	0	0	0	919	
19/11/2018	Mon	0	0	0	0	0	0	0	0	0	1,048	
20/11/2018	Tue	0	0	0	0	0	0	0	0	0	1,149	
21/11/2018	Wed	0	0	0	0	0	0	0	0	0	1,295	
22/11/2018	Thu	0	0	0	0	0	0	0	0	0	1,489	
23/11/2018	Fri	0	0	0	0	0	0	0	0	0	1,733	
24/11/2018	Sat	0	0	0	0	0	0	0	0	0	1,895	
25/11/2018	Sun	0	0	0	0	0	0	0	0	0	2,004	
26/11/2018	Mon	0	0	0	0	0	0	0	0	0	2,105	
27/11/2018	Tue	0	0	0	0	0	0	0	0	0	2,217	
28/11/2018	Wed	0	0	0	0	0	0	0	0	0	2,304	
29/11/2018	Thu	0	0	0	0	0	0	0	0	0	2,424	
30/11/2018	Fri	0	0	0	0	0	0	0	0	0	2,577	
01/12/2018	Sat	0	0	0	0	0	0	0	0	0	2,767	
02/12/2018	Sun	203	100	303	271	133	404	390	192	582	3,473	
03/12/2018	Mon	178	88	266	238	117	355	714	352	1,066	4,042	
04/12/2018	Tue	158	78	237	211	104	315	986	486	1,472	4,528	
05/12/2018	Wed	143	70	213	190	94	284	1,218	600	1,818	4,960	
06/12/2018	Thu	130	64	194	173	85	259	1,416	698	2,114	5,326	
07/12/2018	Fri	120	59	179	160	79	238	1,589	783	2,371	5,642	
08/12/2018	Sat	111	55	166	148	73	221	1,740	857	2,597	5,916	
09/12/2018	Sun	90	44	135	120	59	180	1,860	917	2,777	6,140	
10/12/2018	Mon	98	48	147	131	65	196	1,986	978	2,964	6,344	
Total		1,232	607	1,839	1,642	809	2,452	1,986	978	2,964	6,344	

LEGEND:

- weekend
- LFS
- weekend + LFS

Work on Weekends:
33%

Examples of use During collection

- Resource allocation and case prioritization for CMP
- Active Management Activities and reports
 - Monitor effort and progress by survey and Regional Office
 - Monitor interviewer performance
 - And more...
- Power BI
- Geo visualization
- Ad hoc requests



100



Resource allocation and case prioritization



100

In the context of Collection Management Portal (CMP) and national sample

- How the collection resources (e.g. proportion of hours) will be allocated by Regional Office (RO), survey, cycle and group of cases?
 - The model optimizes the allocation of the number (or %) of collection hours by RO, survey, cycle, group of cases using : Production plan (PP), interviewer claims, budget, number of available and responding cases.
 - Take also into account many operational constraints e.g., collective agreement, capacity, budget constraint, absenteeism rate
- How the interviewers be assigned by Survey, group for a given RO and Day?
 - Interviewer Field Assignment for a session (automatic)
- How cases will be delivered to the interviewers?
 - Call scheduler and prioritization
- *Work still in-progress*

9



Active Management

- Survey Progress Monitoring Report
 - Consolidate all paradata sources into a single report including budget and planning assumptions
 - Highest level report for a given survey



Summary			
	Expected	Observed	Final
Sample	33,248	0	
Hit Rate	95.0%	98.2%	
EQ Completion	12.2%	23.9%	15.1%
EQ Responses	3,847	7,805	4,766
CATI Responses	2,524	1,834	11,135
Response Rate	20.2%	29.5%	50.3%

Survey	Cycle	First	Last	Current	% of Collection Period
CIUS	201811	16/11/18	11/03/19	06/01/19	44.3%

RO	Production Plan/ (expected)			Observed Results		
	% System Hours	% Collection Hours	% CATI Resp	% System Hours	% Collection Hours	% CATI Resp
EDM	15.5%	15.5%	15.5%	18.9%	18.2%	14.0%
HFX	24.2%	24.2%	24.2%	24.9%	24.8%	25.9%
SHER	21.7%	21.7%	21.7%	17.9%	19.5%	20.8%
STURG	23.1%	23.2%	23.1%	21.6%	21.5%	23.0%
WIN	15.4%	15.4%	15.5%	16.7%	16.1%	16.4%

Examples of use After collection

- Post mortem
 - Often to plan and improve next survey cycles
- Historical analysis
 - Why response rates are decreasing? (Question most frequently asked...)
- Develop new concepts, measures or metrics
 - EQ metrics, metrics for national sample
 - Framework to Assess the Maximum Expected Response Rate
- Specific analysis - Often asked by senior management
 - Impact of survey length on CATI and EQ dropping rates
 - Evaluate field tests, e.g., Impact of invitation wording letter
 - Impact of the voluntary/mandatory nature of the survey

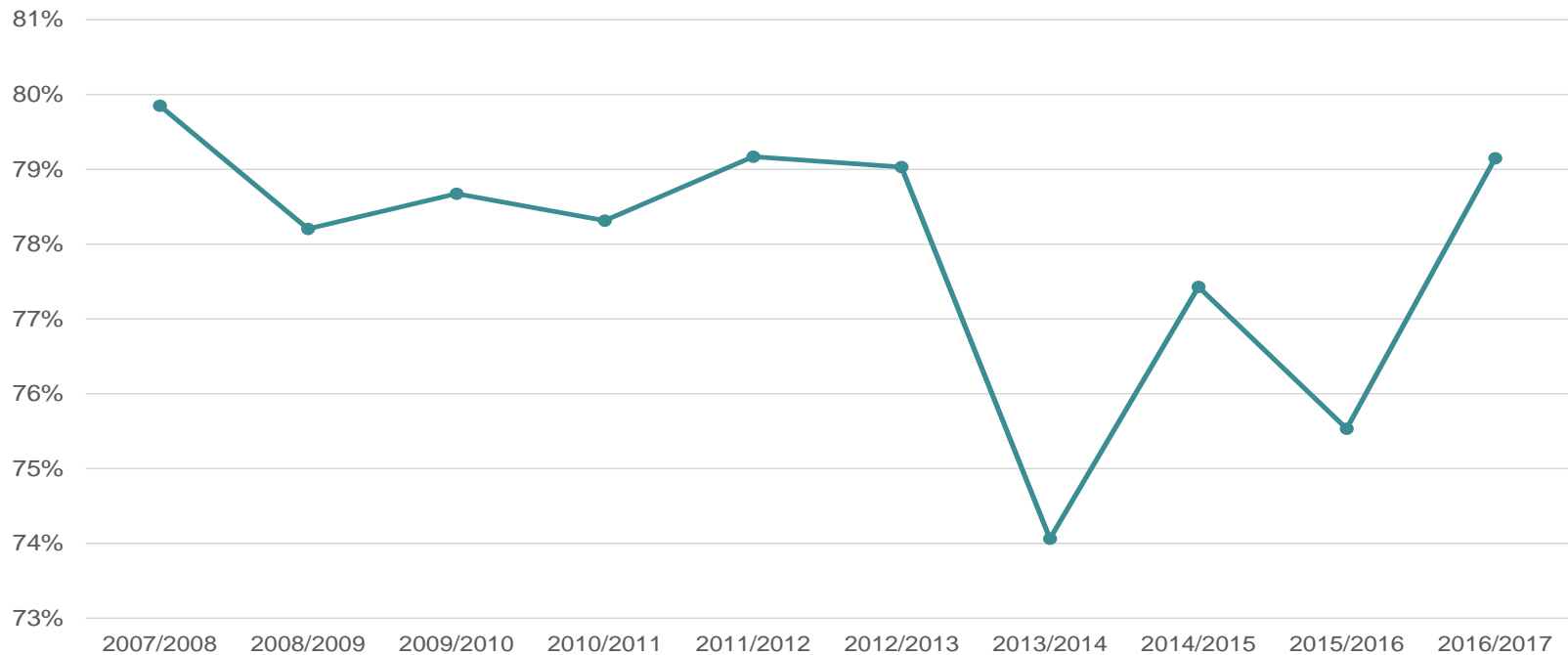


Historical Analysis



- CRRI for Social CATI and CAPI surveys 2007-2016
- CRRI for all the years for which it has been computed 2007/08 to 2016/17. The data comprises 1,100 survey cycles and 76 million calls/visits

CRRI for Social CATI and CAPI surveys 2007-2016



LFS response rates 201701-201811

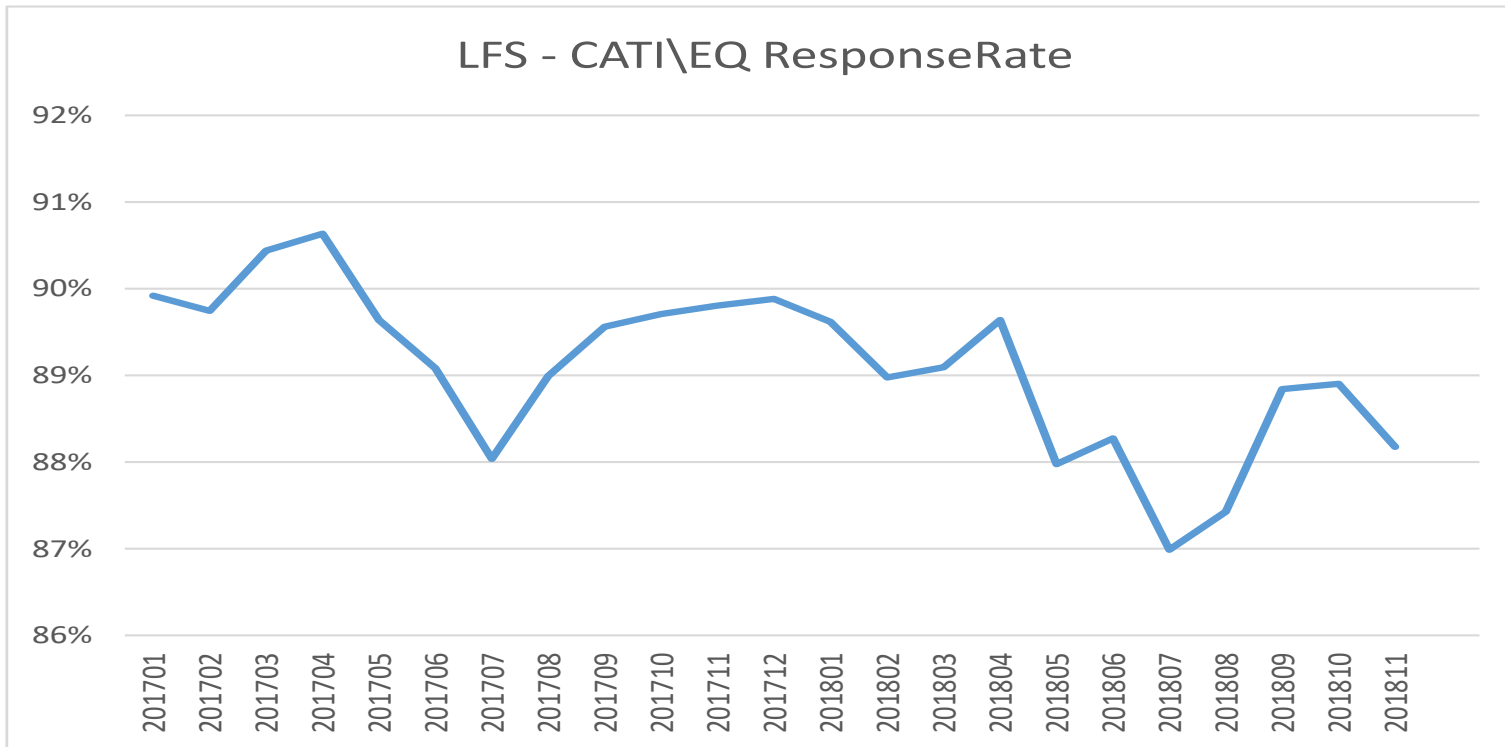
Decline - especially in 201801 - 201807 period

Why CATI response rate decreased in 2018?

- More refusals and less refusal conversion
- More no contact
- Less EQ responses especially in Québec and Ontario. Why?
- Interviewer workload composition?

100

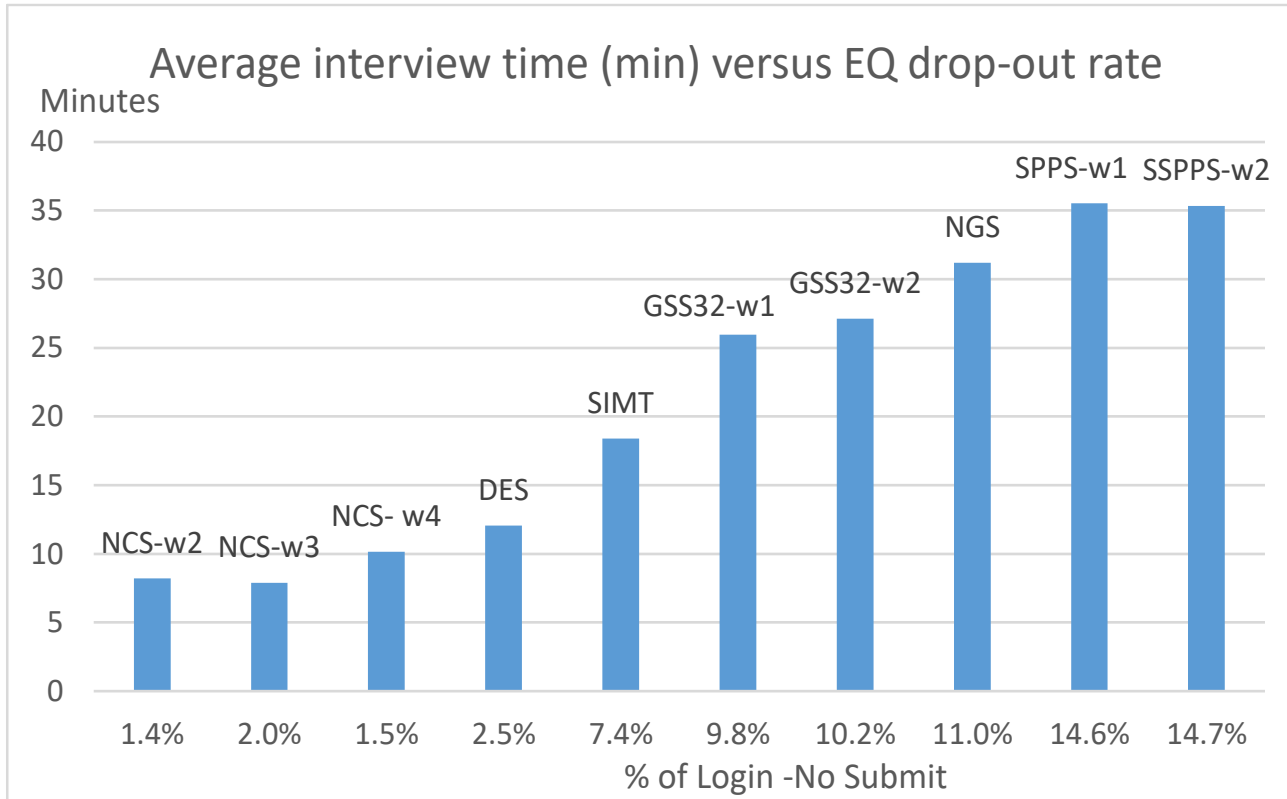
LFS - CATI\EQ ResponseRate



Impact of survey length on EQ drop-out rate



A thousand words are worth a picture



Innovation Research Projects

In addition to the current research projects and activities

- Testing the use of emails
- Testing the use of text (SMS)
- Researching and Testing the use of mobile applications - Time-use survey
- Cognitive Interactive Voice Response (IVR) system
- Crowdsourcing
- Research on the use of AI in collection
- Blockchain
 - Homeless project



100



100

For more information, please contact
Pour plus d'information, veuillez contacter

François Laflamme
francois.laflamme@canada.ca

Sylvie Bonhomme
Sylvie.bonhomme@canada.ca



Statistics
Canada

Statistique
Canada

www.statcan.gc.ca

Canada