



Statistics Canada
www.statcan.gc.ca

Tools and Logistics Discussion

4th Workshop Advances in
Adaptive and Responsive Survey

Manchester, November 2015

Outline



□ Three presentations - Summary and discussion

◆ Peter Miller:

- The U.S. Census Adaptive Design Research Agenda

◆ Peter Lynn:

- Targeted designs in longitudinal surveys: prospects and definitions

◆ Michael Thieme:

- Intra-Application Distributed Processing: Computing Speed and Agility for Adaptive Design

□ Some general thoughts about

- ◆ Main factors affecting the choice of RCD strategy
- ◆ Tools and logistics for RCD surveys

Peter Miller



□ Previous work

- ◆ Census test - pilot of adaptive approaches to nonresponse follow-up
- ◆ National Survey of College Graduates - mode switching to improve sample balance, response rates and reduce costs.

□ Current work

- ◆ Indicators of quality (r-indicator and stability of estimates)
- ◆ Paradata and auxiliary data to improve data collection management
- ◆ Design features and interventions
 - Test different strategies to create workload assignment
 - Case prioritization for CAPI interviews
 - Improve CATI procedures to reduce the number of transferred cases to CAPI

Peter Miller



□ Current work (continued)

◆ Propensity models

- Priority to cases that have particular contribution to nonresponse bias
- Prediction of the likelihood of interview completion
- Propensity model for identifying vacant housing

◆ Tools and logistics

- Construction of computing capabilities needed to execute adaptive design interventions
- Development of a formal framework protocol for adaptive design
 - Standardization??

◆ Context of “Normal” survey practice

- Research studies (including field test evaluation) are very often impacted by confounding factors that are very difficult to control in the field and in the analysis (noise)

Peter Lynn



□ Target design - Definition

- ◆ Planned strategy that identify before the beginning of data collection the sub-groups to be treated and their specific treatment (no further adaptations are made during collection)
 - Generally used for short data collection surveys or rotating panel sample design surveys including longitudinal surveys

- ◆ My definition: Responsive Collection Design (RCD) generally refers to strategy that can be modified during collection while Adaptive Collection Design (ACD) takes advantage of the lessons learned from previous collection cycles to improve the next ones

- ◆ Should we care about definitions ? Response: NO
 - Lars Lyberg (paradata workshop (London 2009)) argued (and I agree) that we should not allow our quest for a definitive definition of “paradata”. Same thing for responsive, adaptive, targeted designs. The terms just will have to be clearly defined every time it is used.

Peter Lynn



- ❑ Relationship between cost and error
 - ◆ RD/AD objectives could be to reduce error (essentially non response error). For example, target low propensity sub-groups (e.g. population with traditionally low response rate)
 - But nothing prevent us to aim to reduce measurement error as well
- ❑ Discuss auxiliary information used to define the subgroups available prior to collection (e.g. socio-demographics, paradata)
- ❑ Discuss the treatments (Differential incentives, prioritising certain types of cases, differential communication strategies) – provide example
- ❑ Looking forward
 - ❑ Target feature should be the norm (Routine) not the exception (one time experiment) - exactly StatCan direction
 - ❑ Different field procedure for hard to reach population

Michael Thieme

- ❑ Computer application is generally developed in silo (one application by survey step) following in the survey lifecycle pattern (e.g. frame development and maintenance, sample selection, sample control, data collection, reporting, analysis, etc.)
 - ◆ Duplication of capabilities and effort
 - ◆ Limit opportunities to optimize applications to take advantage of the strengths different computing architectures
 - ◆ Less flexible

- ❑ From a system and computer application development perspective, they believe decomposition of the process must go beyond the survey lifecycle.
 - ◆ Providing them as single shared services for all applications (regardless of where they fall in the survey lifecycle). For example, the analysis system can be used at different survey steps.

Michael Thieme

- System that consolidates, integrates, timely processes many different data sources constitutes the heart of any RCD
 - ◆ Cornerstone of Active Management (AM) that aims to timely monitor, analyse and communicate data collection process during collection
 - ◆ But we need timely and accessible paradata and relevant auxiliary info.

- Standardization of RCD strategy and system (including testing process) is more than important
 - ◆ More efficient processing (as suggested by Michael)
 - ◆ Ensure consistency and coherence across all surveys
 - ◆ Facilitate the implementation of new RCD surveys
 - ◆ Reduce the time and risks of implementation
 - *RCD approach was initially developed with a research rather than a production perspective - Need significant adjustment*
 - *By default all StatCan CATI surveys are using RCD strategy*



Some General Thoughts

Main Factors Affecting the Choice of RCD Strategy



□ RCD objectives

- ◆ Improve response rate and quality
 - Reduce error - especially non response error (peter)
- ◆ Reduce cost
- ◆ Improve sample representativeness
- ◆ Any combination of these potential objectives

□ Mode of collection (e.g. CATI, CAPI, multi-mode)

□ Type and quality (including timeliness) of information available during collection

- ◆ Paradata and other data sources (e.g. cost)
- This factor not only impacts the feasibility and choice of RCD strategy but also the tools and logistics (e.g. ability to actively monitor the survey progress)

Main Factors Affecting the Choice of RCD Strategy (cont'd)

- ❑ The availability of the information prior to collection
 - ◆ Sampling frame and sample design information
 - ◆ Paradata from previous data collection survey or cycle
- ❑ Duration of collection period
- ❑ Practical considerations (e.g. technical limitations & communication channels)
 - ◆ System development including paradata processing and quality (e.g. Timeliness, accessibility, interpretability and accuracy) - Impact on tools and logistics
 - ◆ Which type of new data collection features can be added to the collection system? System flexible enough?
 - ◆ Which type of interventions can be done during collection?

Tools and Logistics (Active Management)

- Set of plans and tools to manage data collection while in progress
 - ◆ Tools and Logistics need to be developed prior to data collection

- In RCD context, Active Management (AM) is used:
 - ◆ to provide timely information on survey progress and performance using key indicators
 - ◆ to decide when is the right moment to change strategy
 - ◆ to determine if interventions are required
 - If so, determine which ones are the most appropriate to meet RCD objectives

 - Active Management (AM) is essential for any RCD
 - Without no RCD but ACD still possible