Overview of CATI Data Collection Research
Focussed on Developing Operational Strategies for Process Improvement

François Laflamme
Fedcasic, March 2009
Outline

- Introduction
  - Research Objectives
  - Statistics Canada’s Paradata Warehouse
- Initial Research
  - Strategic Opportunities for Improvement
- Overview of Ongoing Research
  - Productivity and Cost Analysis
- Current and Future Research Plan
- Conclusion
Data Collection Research Objectives

- Better understand data collection process
- Identify potential operational efficiencies
- Evaluate new initiatives: time slices, cap on calls
- Maintain and improve data quality
- Improve the way surveys are conducted and managed

Data collection is a key element of the survey process because it has a direct impact on the quality and the cost of many statistical programs
Data Collection Expenditures

- Data collection represents a large proportion of the total survey expenditures

<table>
<thead>
<tr>
<th>Survey / Program</th>
<th>Type</th>
<th>Percent of Data Collection Expenditures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labour Force Survey</td>
<td>CAPI/CATI</td>
<td>74.8%</td>
</tr>
<tr>
<td>Health Surveys</td>
<td>CAPI/CATI</td>
<td>47.7%</td>
</tr>
<tr>
<td>Survey of Labour and Income Dynamics</td>
<td>CATI</td>
<td>49.0%</td>
</tr>
<tr>
<td>General Social Survey</td>
<td>CATI</td>
<td>72.5%</td>
</tr>
<tr>
<td>Canadian Tobacco Use Monitoring Survey</td>
<td>CATI</td>
<td>69.4%</td>
</tr>
<tr>
<td>Survey of Household Spending</td>
<td>CAPI</td>
<td>63.7%</td>
</tr>
</tbody>
</table>
Statistics Canada’s Paradata Warehouse

Paradata Warehouse includes:

- Call and contact information for both telephone (CATI) and personal (CAPI) interview surveys
- Administrative and payroll information
- Historical information since 2003
- Updated on daily basis
Paradata Warehouse Advantages

- Statcan is responsible for data collection
  - Other organizations sub-contract
- All surveys represented
  - Availability of historical data
  - Comparison across survey cycles
  - Can compare different types of surveys to validate findings
- Paradata collection - no burden for interviewers
- Access to detailed and timely cost information
  - Permits survey cost analysis
Initial Research

- Objectives and scope
  - Better understand data collection process and practices and identify opportunities for improvement
  - Focus on CATI surveys: RDD, cross-sectional, longitudinal social and agricultural surveys (using call and contact information)

- Examples of analysis
  - Effort spent: attempts and system time
  - Reaching respondents: contact rate, sequence of calls
  - Relationship between production and cost
  - Same type of analysis across different types of surveys
  - Validate findings
Strategic Opportunities for Improvement

- Better use of pre-collection information
- Use information gathered during collection
- Need more emphasis on post-first contact research
- Assess the interaction between concurrent surveys
- Develop a responsive design framework
  - Active management and adaptive data collection strategy
    - Same strategy does not work effectively throughout collection period
- Continue to investigate the relationship between production and cost
  - Indicators to link effort and results during collection period
- Better predict collection resource requirements during collection period based on observed progress
  - Staffing impacts both survey productivity and cost
Overview of Ongoing Research Linked to Strategic Opportunities – Production versus Cost Data Analysis

Production and cost concepts

- **Production (system time)**
  - Complete Interview System Time: system time to complete interview
  - Total System Time: total system time includes all calls (e.g. appointments, tracing, answering machine, interview, etc...)

- **Cost (payroll hours)**
  - Direct Collection Payroll Hours: time charged (payroll hours) to conduct direct collection activities
  - Total Payroll Hours: total time charged
Relationship between Production and Cost throughout Survey Cycle

- Strong relationship between system time and payroll hours
  - System time is a good predictor for payroll hours charged
- These findings generated more focussed research projects
  - Survey productivity indicators
  - Survey cost analysis: Impact of cap on calls on survey costs

![Graph showing distribution of system time and payroll hours by collection day.](image)
Survey Productivity Indicators

- Daily Productivity Indicators – Example
  - Provide a link between effort and results during collection
  - Productivity ratios decrease during collection period
  - These ratios are affected by interview length and response rate

Note: Other types of productivity indicators can be derived
Survey Cost Analysis – Impact of Cap on Calls on Survey Costs

- Initial objectives of cap on calls
  - Reduce respondent burden
  - Manage interviewing effort more efficiently
  - No cost savings were explicitly planned at that time

- Cost analysis takes advantage of the relationship between production and cost during collection

- Maximum potential cost savings vary from 3.1% to 4.2% under the current cap on calls of 40 for longitudinal surveys

- Research limitations
  - Based on 2005\2006 data (before the introduction of the cap on calls)
  - Since then, work practices in ROs have been continuously improved
  - Survey planning would also have been different with a cap on calls
  - In practice, time spent over the cap on calls is not saved automatically – provides an estimate of the maximum potential savings (in theory)
    - When a case is capped on a given day, interviewers continue to work on the other available cases (always cases left to be worked on)
Current and Future Research Plan

- Responsive Design for CATI surveys
  - Adaptive collection
  - Multi-mode / multi-site environment
- Assessing and monitoring survey productivity and costs in CATI household surveys
- Sequence of calls (with the objective to increase likelihood of contact and interview)
- Planning and predicting the number of staffing hours based on survey progress
Current and Future Research Plan

- CAPI Surveys
  - Assess paradata quality and limitations
  - Develop productivity indicators
  - Include geographical characteristics
  - Evaluate and monitor CAPI sample coordination
Current and Future Research Plan

- Other Paradata Research Projects and Activities
  - Dashboard of key survey progress and productivity indicators based on paradata
  - Working groups:
    - RO CATI/CAPI research, cap on calls
  - Business surveys analysis
  - Audit trail projects: POINT and others
  - Ad hoc research

- Other Data Collection Research Projects
  - Mixed-mode collection
  - Cell phone
Conclusion

- Paradata has been the cornerstone of data collection research at Statistics Canada and continues to be extensively used.

- Research is based on objective and empirical measures automatically collected during collection:
  - Almost no collection cost and no interviewer burden.
  - Main cost is to create and maintain paradata database and to develop new analytical tools.

- Benefits of Paradata analysis can be important since data collection represents a large proportion of the overall survey cost.
For more information, please contact

François Laflamme
francois.laflamme@statcan.gc.ca