Optimizing Call Attempts Using Response Propensity and Marginal Yield Analysis

Rafael Sobrino and Rachid Boudiaf

Statistics Canada

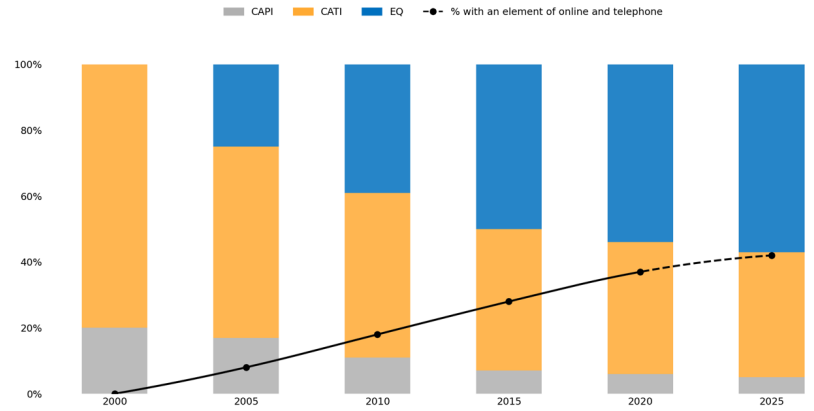
FedCASIC - April 21, 2026

Background

- Collection costs are rising across all major statistical agencies
- Response rates for household surveys have been declining for decades
- Static collection rules fail to account for heterogeneity
- A data-driven, adaptive approach is needed

Rising Costs and Declining Response Rates Demand New Strategies

Statistical agencies worldwide are grappling with a fundamental challenge: the cost of data collection is increasing while response rates are declining. Traditional "one-size-fits-all" collection rules are no longer efficient.



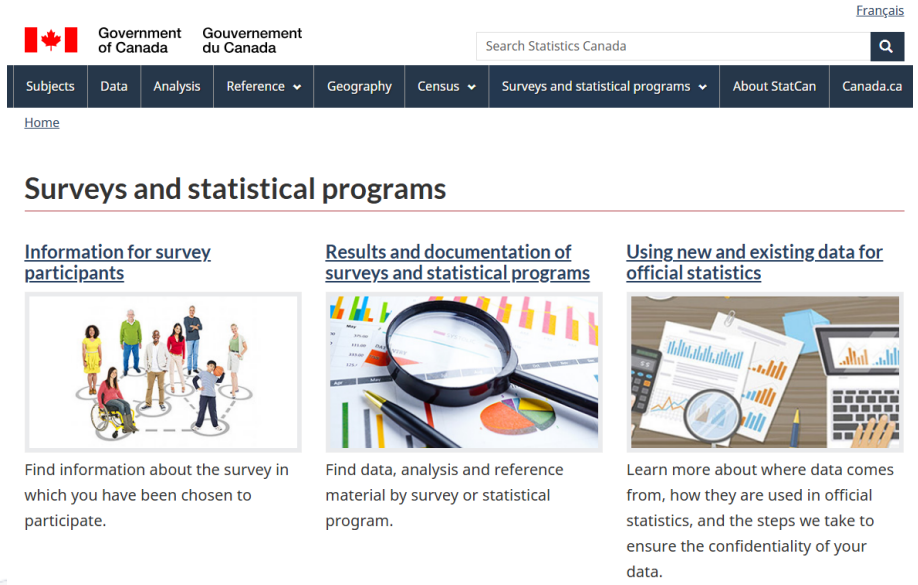
The shifting landscape of data collection modes requires adaptive strategies.
Source: CPRD, Statcan 2026

Statistics Canada Vision

- **Sustainable collection models**
  Centralized expertise and shared tools
- **Optimized reminder strategies**
  - Analysis of the impact of emails, SMS and Letter reminders
  - Targeted follow-ups using paradata
- **Budget optimization**
  - Limited number of attempts
  - Prioritized high-probability cases

Adaptive and Sustainable Collection Strategies

Statistics Canada is moving toward adaptive, data-driven collection approaches to address rising costs and declining response rates. Traditional collection strategies are being replaced by responsive and efficient strategies.



Our Goal

Identify When to Stop Calling to Maximize Efficiency

Moving beyond fixed collection rules by leveraging response propensity models to tailor collection efforts and optimize resource allocation.

Segment the Sample

Use a response propensity model to divide the sample into quintiles based on predicted behavior.

Analyze Marginal Yield

Evaluate the effectiveness of each additional CATI call attempt for every segment.

Identify Stopping Point

Determine the point of diminishing returns where further calls yield minimal results.

Provide Recommendations

Formulate actionable, evidence-based strategies for sustainable data collection.

A Three-Stage Analytical Framework

1 **Response Propensity Modeling**

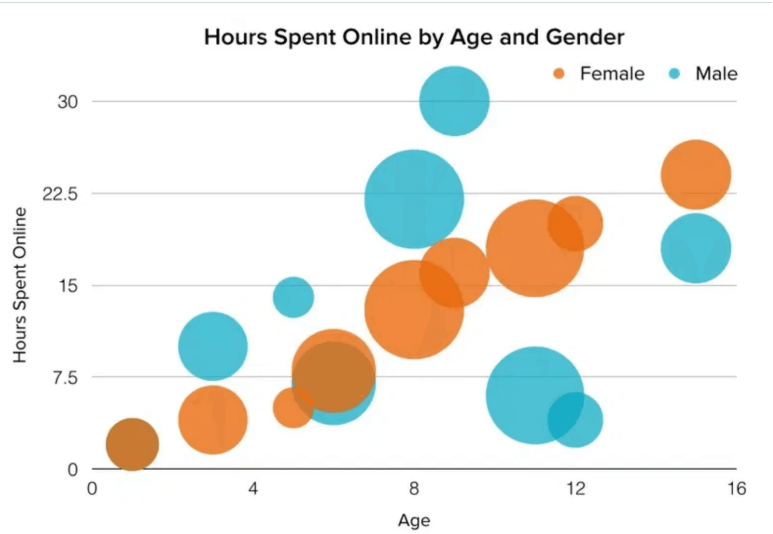
Using Elastic Net regression in R to predict response probability based on socio-demographic data.

2 **Sample Segmentation**

Dividing the sample into quintiles based on predicted propensities to create homogeneous groups.

3 **Cost-Benefit Analysis**

Evaluating call attempt efficiency using both dynamic thresholds and fixed operational rules.



Visualizing the analytical process from data to insights.
Source: CRPD, Statcan, 2026

Reference Survey

Canadian Social Survey (CSS)

The Canadian Social Survey (CSS) is a **voluntary** quarterly survey by Statistics Canada aimed at understanding various social issues affecting Canadians, covering topics like health, well-being, and quality of life.

Data Collection and Usage

The CSS employs an electronic questionnaire (EQ) with follow-up telephone interviews for non-response (CATI). Data collected is used by various levels of government and organizations to inform policy development and improve services for Canadians.

Future of CSS

Although CCS will no longer be part of the concurrent surveys for now, it is considered a good example because of its structure.

1 - Response Propensity Model

Predicting Response Probability Using Socio-Demographic Data

The response propensity model was built using **Elastic Net regression**, a technique that combines Ridge and Lasso regularization. This method is ideal for datasets with many potentially correlated explanatory variables, as it performs variable selection and regularization simultaneously.

As the survey frame relies on administrative sources, not all information is available for every case. Some variables are imputed and may reflect outdated data (e.g., last census).

Key Predictors

Socio-Demographics

Age, Gender, Household Size, Contact Information Availability, Household Income

Modelling and Optimization

Elastic Net Regression with Cross-Validation

Alpha = 0.5 (Balanced Ridge/Lasso)

$\min\beta(\|y-X\beta\|2+\lambda1\|\beta\|1+\lambda2\|\beta\|2)$

Results

Propensity Prediction

The propensity to respond was predicted for each case using the final model

Segmentation

Creation of Deciles and Quintiles

Segmentation into deciles and quintiles based on the distribution of predicted propensity

2 - Sample Segmentation
Socio-demographic Profiles (Quantiles)

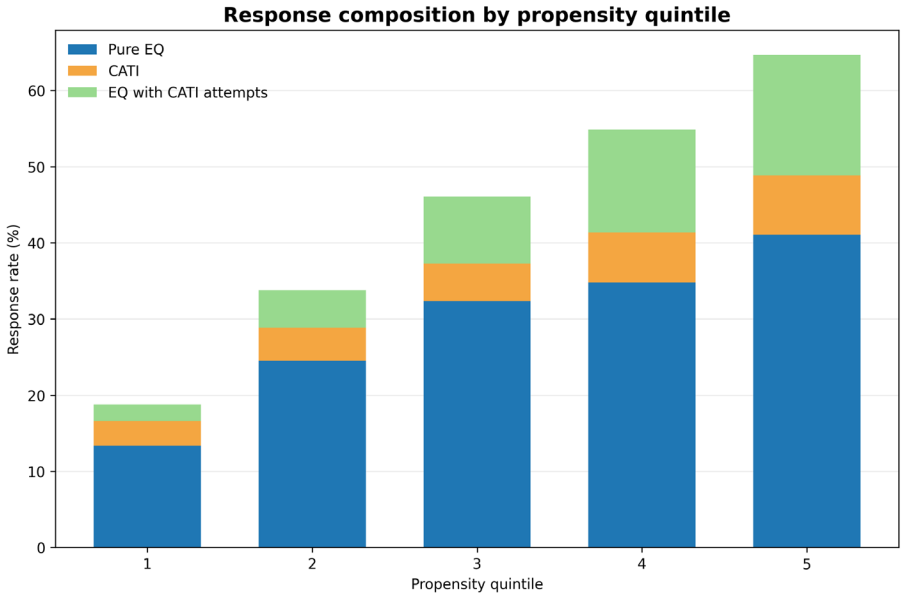
Q1 — **Very Low Propensity**
Precarious households, limited digital contact means, very low response rates.

Q2 — **Low Propensity**
Socially modest, uncertain participation, potential for targeted follow-up.

Q3 — **Moderate Propensity**
Standard family households, well-equipped digitally, relatively responsive.

Q4 — **Good Propensity**
Mostly homeowners, well-connected (93% phone, 96% email), easy responders.

Q5 — **High Propensity**
Affluent, small households, universal access, highly responsive.



CSS Sustainable Collection Results
Source: CRPD, Statcan, 2026

3 - Cost benefit Analysis

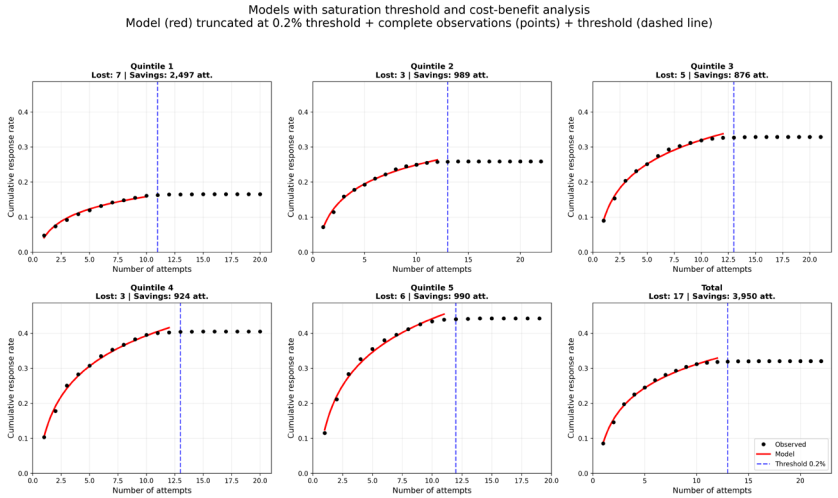
We evaluated two distinct approaches to define the optimal stopping point for call attempts. Both approaches converge, continuing call attempts beyond approximately 12 attempts is a low-yield strategy.

Dynamic Approach

0.2%

Marginal Gain Threshold

Identifies the point of diminishing returns empirically. Saturation is reached when the marginal gain falls below **0.2%**. Applied independently to each quintile for maximum flexibility.

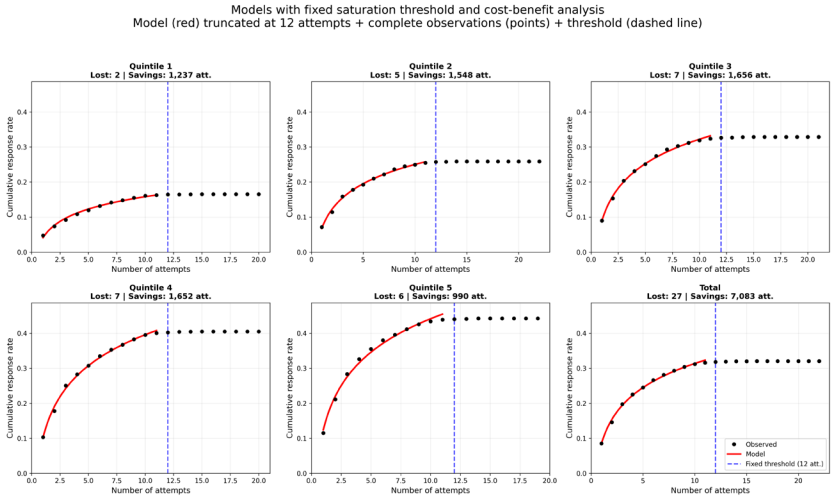


Operational Approach

12

Fixed Attempts

Simulates a simple, uniform management rule. Saturation is arbitrarily set at **12 attempts** for all cases. Prioritizes operational ease and standardization.



Analysis Results

We evaluated two distinct approaches to define the optimal stopping point for call attempts

Saturation point

Both Approaches Converge on a 12-Attempt Stopping Rule

Attempts Saved

The fixed rule is slightly more conservative, generating an additional 807 saved attempts.

Lost Responses CATI

The number of lost CATI responses is identical in both scenarios.

Lost Responses EQ

The difference in lost responses comes exclusively from EQ responses, which are less costly to obtain.

Marginal Efficiency

The efficiency of attempts beyond the threshold is identical, confirming that efforts beyond 12-13 attempts yield very low returns.

Conclusion: A 12-Attempt Rule Saves 7% of Budget with Minimal Response Loss

7.2%

Total Budget Savings (CATI Collection Hours)

Marginal Gain Is a Baseline – Not a Universal Stopping Rule

Three metrics, three denominators – three different questions

Each metric measures something real, at a different scale.

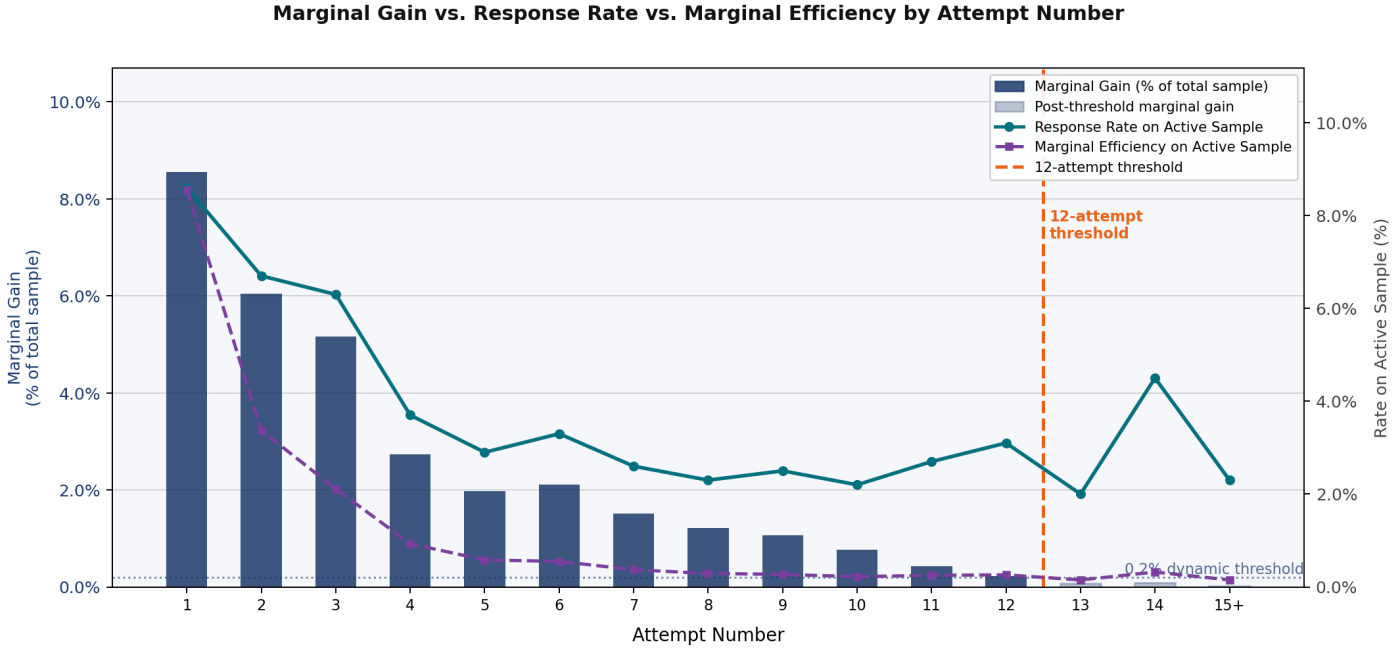
Marginal Gain asks: how much does this attempt contribute to the total sample?

Response Rate on Active Cases asks: how likely is a response among cases still reachable?

Marginal Efficiency asks: how productive are the accumulated call-attempts on those same cases?

The right threshold depends on collection goals

Groups underrepresented or critical for dissemination may warrant more attempts. Overrepresented groups may need fewer than 12. Marginal gain is the baseline – adaptive collection adjusts from there.



Navy bars: Marginal Gain (% of total sample). **Teal line:** Response rate on active cases.

Purple dashed line: Marginal Efficiency = responses / (active cases × attempt number)

Same data – three denominators.

7.2%	3.6%	<0.2%
Collection hours saved	sample active at attempt 13	marginal gain beyond threshold

Literature Review

External Research Validates the 12-15 Attempt Threshold

Our findings are strongly supported by the academic and professional literature on survey methodology. Multiple studies on CATI operations identify an optimal number of call attempts, often cited as being between 12 and 15, before efficiency plummets.

Adaptive Survey Design (ASD)

Agencies like the U.S. Census Bureau and Statistics Canada advocate for using paradata to dynamically manage collection strategies, reducing costs by nearly 10% without significant data quality loss.

Vicente et al. (2017)

Identifies 12-15 calls as the optimal range for CATI surveys.

Laflamme & Karaganis (2010)

Details Responsive Collection Design (RCD) implementation at Statistics Canada.

Coffey & Elliott (2023)

Demonstrates potential for ~10% cost reduction via data-driven interventions.

The Operational Challenge

Implementing Stopping Rules Requires Proactive Planning

While optimal for costs, stopping rules can lead to underutilization of assigned staff if production plans are rigid.

Proactive Budget Planning

Integrate saturation thresholds into initial budget and staffing allocations.

Interviewer Versatility

Cross-train staff on multiple surveys to allow for rapid reassignment.

Continuous Training

Utilize freed-up hours for skills development and support activities.

A Roadmap for Sustainable Collection

1. **Adopt a n-Attempts Stopping Rule**

   Implement the optimal threshold for CATI operations to maximize efficiency based of analysis for recurrent surveys.

2. **Reallocate Freed Budget**

   Invest savings into targeted email reminders, SMS, or incentives.

3. **Implement a Hybrid Strategy**

   Stop CATI early but extend EQ collection to capture remaining responses. This has been already implemented for other surveys.

4. **Integrate Saturation Thresholds**

   Incorporate data-driven limits into production planning and staffing.

5. **Research and Implementation of New Collection Features**

   Conduct Research and implementation of new features for the current collection systems

Opportunities for Further Optimization

1 **Validate Findings**

Confirm quintile segmentation and cost-benefit results for other surveys and subsequent collection cycles

2 **Optimize Call Timing**

Analyze and refine strategies based on day of week and time of day efficiency.

3 **Real-time Dashboard**

Develop paradata monitoring tools to track efficiency and trigger interventions live.

4 **Machine Learning**

Investigate advanced ML models to further refine response propensity predictions.

Key Takeaways

Data-Driven Strategies Enable Sustainable Survey Collection

This project demonstrates how response propensity modeling and marginal yield analysis can be operationalized to inform more efficient strategies.

By quantifying the point of diminishing returns, we provide actionable evidence to support targeted CATI strategies. An attempt cap could offer a simple, robust way to reduce costs significantly with minimal impact on response outcomes.

Optimal & Evidence-Based

An attempt cap stopping rule as a validated saturation point for efficiency.

High Impact Savings

Quantify savings in collection hours while sacrificing minimal responses.

Transferable Methodology

The approach can be adapted on other surveys to balance costs and respondent burden.

STAY CONNECTED

Eh Sayers podcast

StatsCAN Plus

The Daily

My StatCan

Website

Surveys and statistical programs

Data service centres

Questions?

Contact us: rafael.sobrino@statcan.gc.ca, rachid.boudiaf@statcan.gc.ca

STATISTICS CANADA

Delivering insight through data for a better Canada