Using Text Messages to Increase Interviewer Compliance in the Survey of Income and Program Participation

Kevin Tolliver

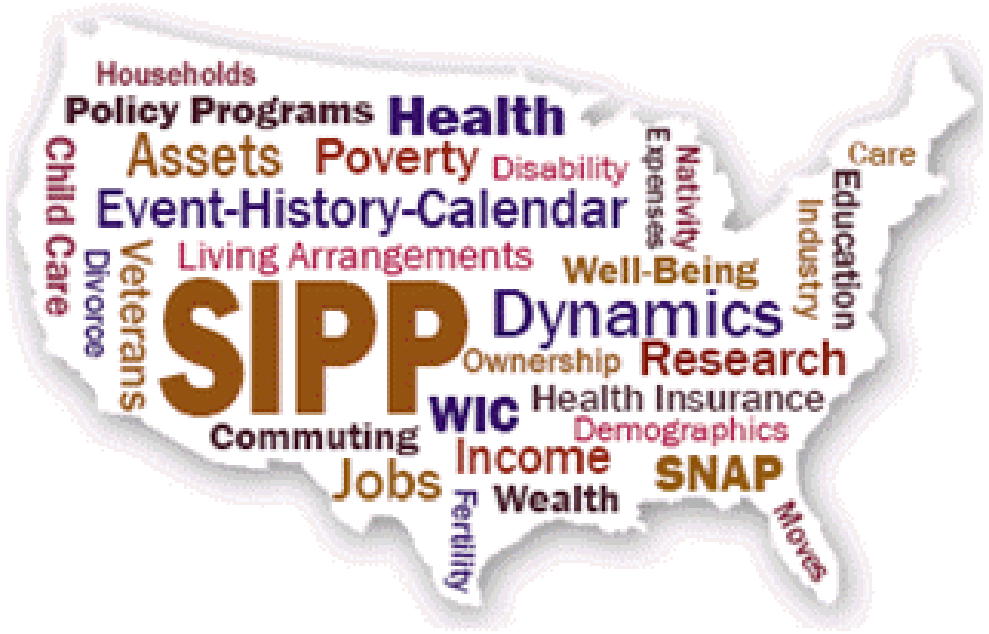
This presentation is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Survey of Income and Program Participation (SIPP)

- National longitudinal panel survey
  - 53,000 households each year
  - Conducted primarily face-to-face
  - multi-year panel

- Key Program Outcomes
  - Temporary Assistance for Needy Families (TANF)
  - Women, Infants, and Children (WIC)
  - Supplemental Security Income (SSI)
  - General Assistance (GA)
  - Supplemental Nutrition Assistance Program (SNAP)



- Challenges
  - Panel attrition
    (movers impact attrition)
  - Budgetary constraints

Motivation

Goal: To stimulate more effort on high priority cases and less effort on low priority cases with a text message

- In 2016 and 2017, SIPP experimentally prioritized select cases in order to improve data quality in the final product
  - Cases were assigned a H, M, L priority via transmissions to the laptop
  - Experiments resulted in modest improvements in data quality (2% point increases among targeted cases) with minor adverse effects to cost of data collection
- Priority protocol did not factor into performance evaluations. Research is still required to determine a fair assessment of priority compliance and an appropriate method for enforcing compliance.
- Not all interviewers showed signs of following the priority protocols evidenced by a post-hoc analysis.

Why Text Messages?

- Text messages are more flexible than phone reminders
- We believe that an interviewer is more likely to see the text in a timely manner than an email or a phone reminder
- Some interviewer indicated that they wanted a better way of knowing when priority changes occurred

Research Questions

- Does the number of text messages have an impact on interviewer behavior?
- What is the impact of a text message on interviewer behavior the following period?
- Do we plan on texting in the future? If so, how do we plan on texting?

Text Messages – Treatment Randomizations

- Dimension 1: Content
  - No Message
  - Reminder – Reminder to follow future prioritizations
  - Feedback – Feedback about prior prioritization
    - Positive or constructive feedback
    - Deterministic based on prior period behavior – discrimination criterion
- Dimension 2: Time
  - 10:00am, 12:00pm, 5:00pm, 7:00pm (Eastern Daylight Time)
  - Interviewers in Eastern/Central Time zone - eligible for all four times
  - Interviewers in Mountain/Pacific Time zone - not eligible for 10:00am texts

Text Messages – Content

- **Reminder Message:** This is a message from SIPP. Your case priorities have been updated. Please transmit and work accordingly. Contact your supervisor with any concerns.

- **Positive Feedback Message:** This is a message from SIPP. You worked your cases according to priorities last week. Thanks for your hard work! Contact your supervisor with any concerns.

- **Constructive Feedback Message :** This is a message from SIPP. To ensure data quality, please work your cases according to priorities. Contact your supervisor with any concerns.

Discrimination Criterion (DISC) as the weekly randomization condition

- DISC is an intermediate criterion which
  - Identifies *strong evidence* of working high priority cases
  - Determines positive or constructive feedback messages and the likelihood of receiving same content and timing
- Based on interviewer behavior in a prior time period
- *Strong Evidence* during the previous trial period (*t-1)* if either condition is met:
  - Condition 1: $\frac{\#\ Checked{-}in\ H\ cases}{\#\ Assigned\ H\ cases} > \frac{\#\ Checked{-}in\ M\ cases + \#\ Checked{-}in\ L\ cases}{\#Assigned\ M\ cases + \#Assigned\ L\ cases}$
  - Condition 2: $\frac{\#\ Attempts\ on\ H\ cases}{\#\ Assigned\ H\ cases} > \frac{\#\ Attempts\ on\ M\ cases + \#Attempts\ on\ L}{\#\ Assigned\ M\ cases + \#\ Assigned\ L\ cases}$

Design: Sequential Randomized Trials

- SRT was chosen to analyze the text effectiveness of a sequence of various events over multiple contact times.
- There were 13 different trials sent over 13 different weeks, in which 1250 interviewers were re-randomized, over the course of the 20 week data collections.
- The content and timing of the message were re-randomized every trial.
    - If the DISC condition was met, interviewers would receive: 1) positive feedback and 2) ~50 percent chance of receiving the exact same content and timing.
    - Otherwise, the content and timing were uniformly random within constraints, like timezones.

Evaluation: Behavioral Metric (BM) Considerations for Assessing an Interviewer's Contact Effort

- Following BMs will measure each interviewer's contact efforts assessed during a period of time *t*.

- BM 1: The number of contact attempts ($A_{c,t}$) made on case ***c*** between period ***t-1*** and ***t***.

  - The final number of contact Attempts A_c is

$$A_c = A_{c,1} + A_{c,2} + \cdots + A_{c,13}$$

- BM 2: The Evenness-of-Finding Attempts ($EFA_{c,t}$) on case ***c*** is the EFA by time period ***t***.

Evenness of Finding Attempts (EFA)

- Used for diversity indices from ecology and economics
- Coombs and Walsh (2014) adapted this idea for survey research methods when analyzing contact attempts
- Computation

$$EFA = \frac{1}{8} \times \frac{1}{\sum_{w=1}^{8} (proportion\ of\ attempts\ in\ window\ w)^2}$$

- Windows
  - Weekend 8am-12pm
  - Weekend 12pm-3pm
  - Weekend 3pm-6pm
  - Weekend 6pm-11pm
  - Weekday 8am-12pm
  - Weekday 12pm-3pm
  - Weekday 3pm-6pm
  - Weekday 6pm-11pm

Overall Behavioral Metric(s)

- There was no statistical difference in the final number of contact attempts made per case on high priority vs. medium priority cases.
- There were 25% fewer contact attempts per case made on low priority cases compared to high priority and medium priority.
- Mean EFA on high priority cases > Mean EFA on medium priority cases > Mean EFA on low priority cases.

Estimating the Effects of the Number of Text Messages

- We aim to test if more text message increases BM among high priority cases and decreases BM among low priority cases

- Consider
  - X_i : the number of text messages sent to interviewer *i*
  - BM_c: $A_c \times EFA_c$
  - $\overline{BM_i}[Priority] = \sum_{c=1}^{C} \frac{BM_c \times I(c \in Priority)}{I(c \in Priority)}, Priority = \{H, M, L\}$

- Simple Linear Regression model of $\Delta = \overline{BM_i}[H(or\ L)] - \overline{BM_i}[M]$ on *X*:

- The estimated β coefficients in the regression measures the effects the text message.

Preliminary Model Outcomes

- More text messages had positive effect on Δ, High Priority vs. Medium Priority, (28 percent increase, p-value = 0.020)

- More text message did not have a significant effect Δ, Low Priority vs. Medium Priority, (3 percent increase, p-value = 0.142)

Estimating the Effects of the Text Message with Repeated Measures

- We aim to test if a text message increases BM among high priority cases and decreases BM among low priority cases

- Generalized estimating equation (GEE) models of $\Delta = \overline{BM_i}[H(or\ L)] - \overline{BM_i}[M]$ at time t on X_t:
  - $BM_{c,t}$: $A_{c,t}, EFA_{c,t}\ or\ A_{c,t}\ \times EFA_{c,t}$
  - X_t : a collection of time-varying predictors at time t: e.g., Text Msg Sent, DISC, Same Content as prior period, Same Timing as prior period

- The estimated β coefficients in GEE measures the effects the text message.

Discussion

The following are notes about the data that impact model selection and conclusions:

- $A_{c,t}$
  - ~30% of observations being zero (no contacts made).
  - Data is right skewed.

- $EFA_{c,t}$
  - ~80% of observations being zero (no diversity).
  - Data is right skewed.

- By design, many other X_t are related to X_{t-1} when $DISC_{t-1}$ criteria is met.

Tailoring Messages for Future Data Collections

Once model has been selected and the experimental effects are estimated, we plan to:

- Categorize interviewers into groups based on similar traits
- Determine which timing and content worked best for each group
- Determine which change should be made if we are not seeing desired results

Thank You!

kevin.p.tolliver@census.gov