

Implementing Adaptive Design for the National Survey of College Graduates

Stephanie Coffey
U.S. Census Bureau
National Survey of College Graduates
March 20, 2013



Introduction

- National Survey of College Graduates
 - Sponsored by National Center for Science and Engineering Statistics (NCSES) at the National Science Foundation (NSF)
 - Part of the Science & Engineering Statistical Data System (SESTAT)
 - Person-level survey sampled from American Community Survey (ACS)
 - Target population is college graduates
 - Occurs every 2-3 Years



Introduction

- 2013 Data Collection
 - February 21 – August 25
 - Sample Size ~143,000 cases
 - 83,000 in New Cohort (2011 ACS)
 - 60,000 in Old Cohort (2009 ACS + 2010 NSRCG)
 - Data collection modes include: internet, mail, phone
 - Different costs and effort



Introduction

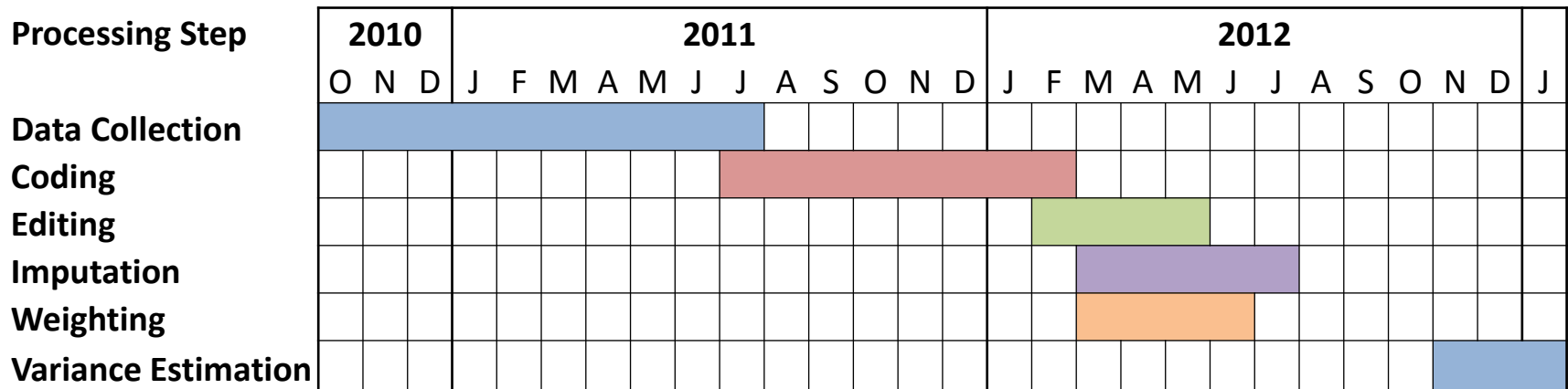
- Methodology Studies: What Strategies Work?
 - Incentive Timing
 - Priority Mail vs. First Class Mail
 - Mode Switching
 - Incentive Conditioning

Motivation for Adaptive Design

- General:
 - Allocate data collection resources efficiently
 - Avoid exhausting money and time
 - Move beyond response rate as the major metric of survey quality

Motivation for Adaptive Design

- NSCG Specific:
 - Reduce the time from start of data collection to delivery of finished product.



Needs to be done **without** sacrificing data quality!



Motivation for Adaptive Design

- Reduce Processing Time After Data Collection
 - Start processing earlier
 - Optimize processing
- Reduce Data Collection Period
 - Understand data quality better
 - Targeted interventions – mode switching
 - Frame data, contact data and paradata



Challenges to Implementation

- System:
 - Independent data collection systems
- Processing:
 - Move processing
 - Make assumptions
- Data Quality:
 - What measures do you use?
 - How do you use them in the decision-making process?

Adaptive Design Components Targeted for 2013 NSCG

Challenges Served as a Roadmap for 2013

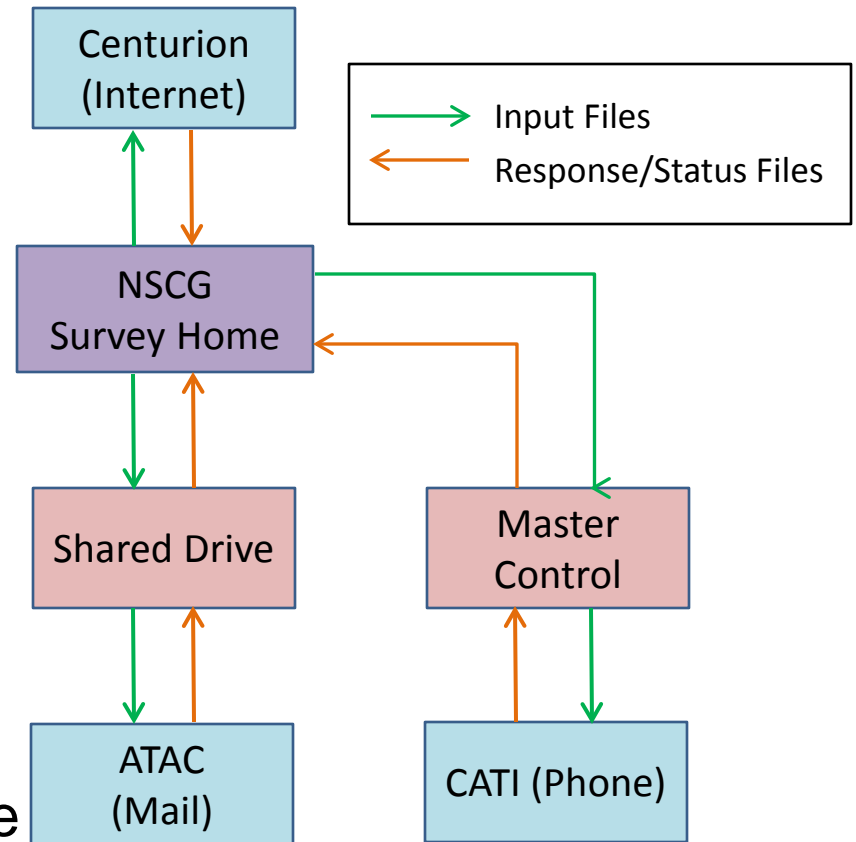
- Integrate Disparate Data Collection Systems
 - Integrated Systems
 - Integrated Reporting
- Institute Flow Processing
- Data Monitoring Methods
 - Implement Methods
 - Increase Access to Paradata
- Determine Possible Interventions

Integration of Systems

Integration of Systems

Baseline (2010)

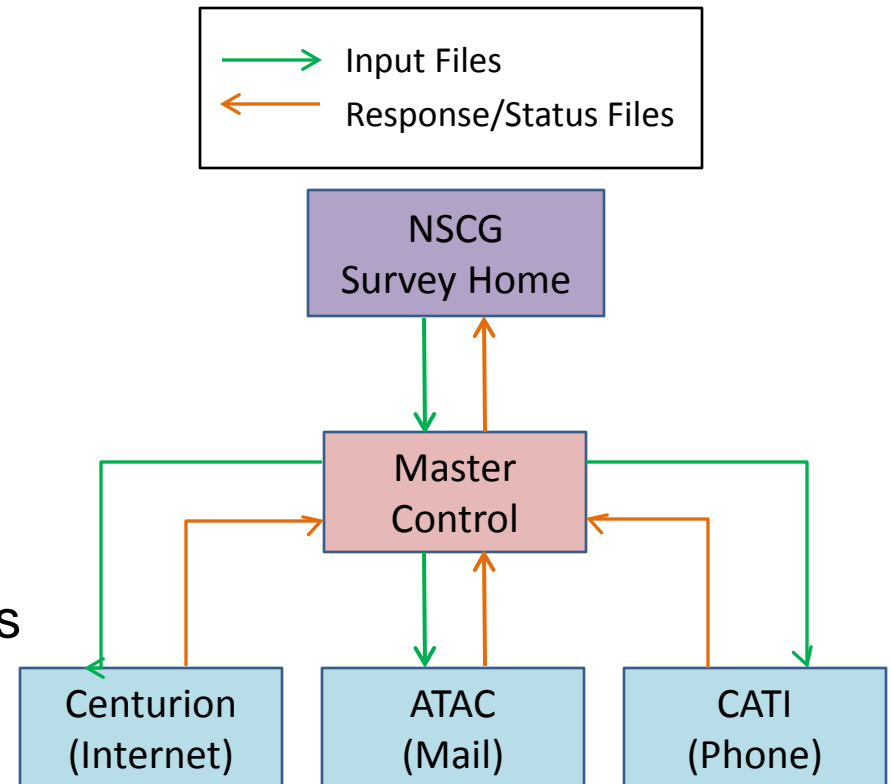
- Input files must be delivered to several different locations
- Many unrelated handoffs
- Separate intermediaries for mail and telephone
- Response files located in several different locations
- No mode-level interventions or communication without data flow to/from NSCG
- Different contact paths by mode



Integration of Systems

New Version (2013)

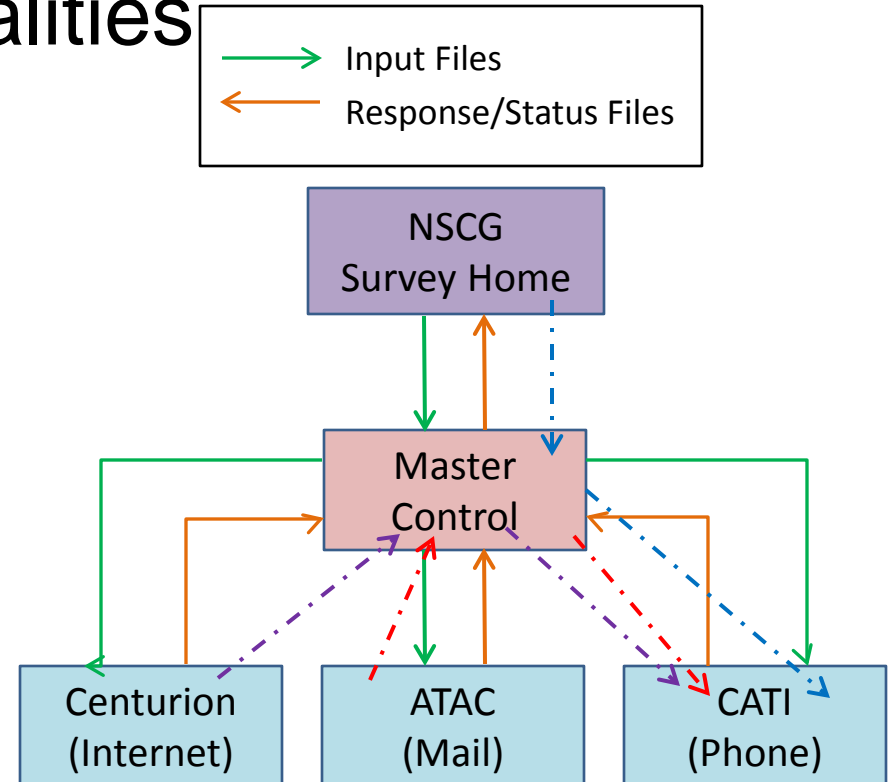
- Input files now delivered to one location
- Response files are now all in one location
- Single intermediary
 - Aware of all modes
 - Can pass info between modes
 - No need to wait for NSCG to affect action/interventions
- Single contact path for all modes



Integration of Systems

New System Functionalities

- CATI Holds from Internet
 - Every 2 Hours
- Mail Processing Holds
 - Daily
- Data Monitoring Holds
 - Weekly
- Integrated Reporting
 - Daily





Integrated Reporting

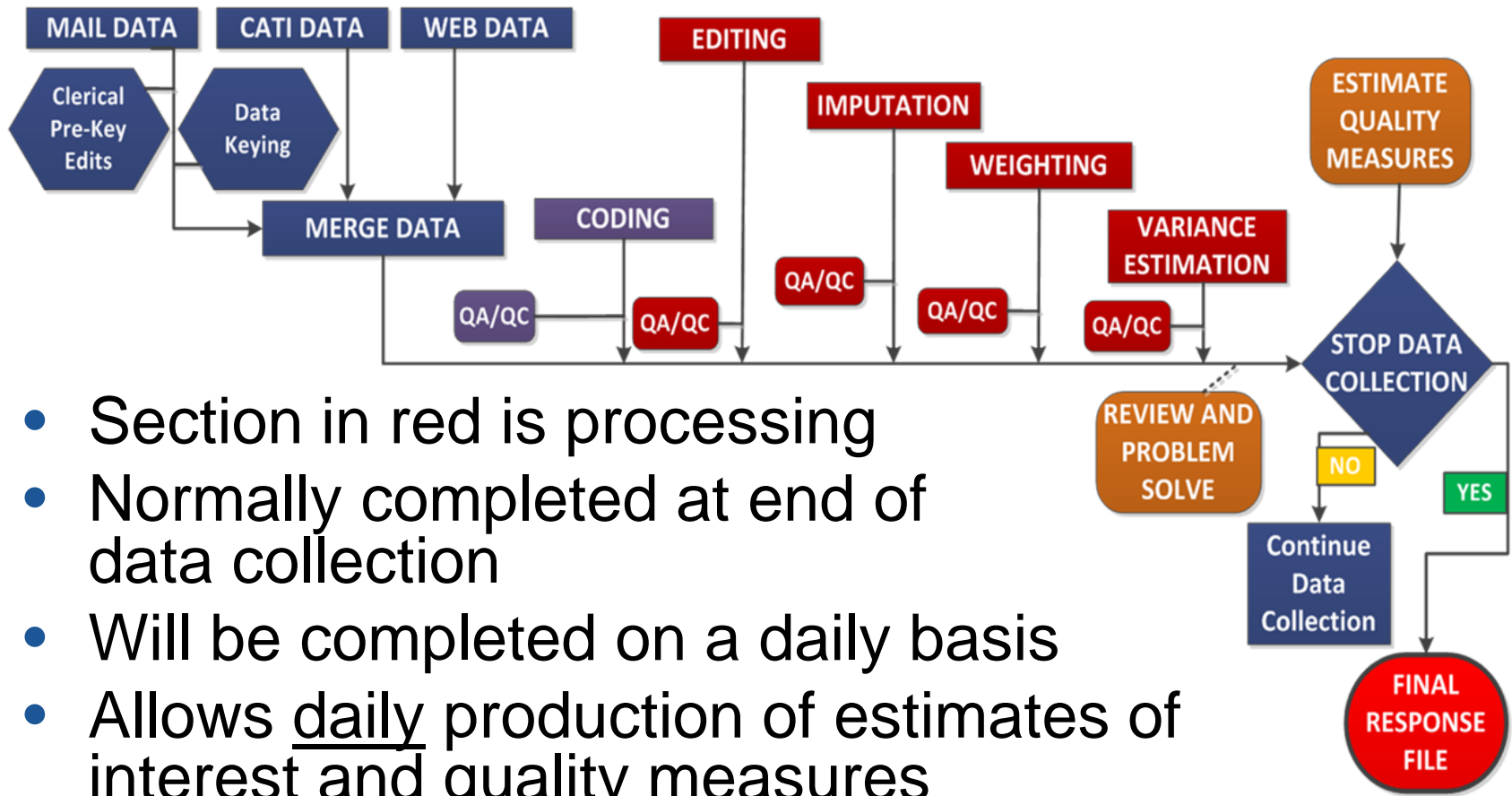
- Universal Tracking System (UTS)
 - Census Bureau enterprise-wide reporting system
 - Combines data streams from various systems
 - Met two major NSCG needs for adaptive design
 - Full Contact Path Report
 - Chronological report of all contacts for a sample person
 - Allowed us to respond to a specific sample person request
 - Contact Aggregation Report
 - Total contacts by category for a sample person
 - Include in data monitoring

Flow Processing

Flow Processing

- Complete most/all parts of processing
- NSCG has a goal of daily processing
- Make some assumptions
 - Less editing or less manual review
- Need coding, editing, imputation, weighting, and variance estimation

Flow Processing





Flow Processing Benefits

- Operational Benefits
 - Processing programs completed earlier
 - Real-world testing opportunities
- Data Benefits
 - See effects of changes in editing or imputation rules immediately in the data
 - Daily views of “final” data and data quality
 - This information is important for data monitoring

Data Monitoring

Data Monitoring

- Data-driven view of “What’s going on?”
- Stage 1: No flow processing required
 - Response Rates: Unweighted and Weighted
 - Exploring Paradata; Incorporate into Models
 - R-Indicators: Sample, Partial, Level-Based
 - Identify variables that drive variation in response propensity
 - Identify which levels of a variable are under/over-represented (group-level measure)



Data Monitoring

- Stage 1: No flow processing required (continued)
 - Response Propensities by Mode
 - What mode is a respondent likely to respond by?
 - Mahalanobis Distance (near-term future work)
 - Case-level distance measure
 - Identifies cases that are “different”
 - Dependent on Quality of Models
 - Carefully build propensity models
 - Need to validate models across years



Data Monitoring

- Stage 2: With results from flow processing
 - Benchmarking to Frame Totals
 - After non-response adjustment, how well do our estimates of frame totals represent the actual frame totals?
 - Stability of Estimates
 - Are new respondents providing new information?
 - Are estimates of interest or their variances changing?

Interventions

Interventions

- Data Monitoring provides information
 - Watch it or act on it?
 - 2013 NSCG includes mode-switching test
 - Monitoring methods help identify target cases
 - Move case to mode with the highest response propensity
 - Hold a case in web if it is a “low impact” case
 - Put a CATI case on hold (no contacts) if R-indicator indicates the group is over-represented
 - Need to identify more possibilities
 - Interventions are part of cost/quality tradeoff in adaptive design



Interventions

- Other types of interventions
 - Investigate and react to issues in data collection
 - Web server was extremely slow during first week of data collection
 - Used web paradata to identify time frame of slow service
 - Identified respondents affected by slow service
 - Mailed apology letter

Wrap-Up

Ongoing/Future Work

- Continue to explore new data monitoring methods
- Carefully document mode switching experiment
 - What did we do? Do we think it worked?
 - What additional variables or paradata might help make decisions?
 - Codify business rules and interventions that drive them
- Consider concerns about asymmetric data collection
- Refine models

Conclusion

- Work in progress
 - Integrated systems changes for adaptive design
 - Implementing initial monitoring methods
 - Enacting interventions and observing
- Adaptive Design Goals for NSCG:
 - Shorten data collection without sacrificing quality
 - Use fewer resources where it makes sense
 - Respond when necessary to information about the state of data collection

Acknowledgements

Steve Cohen (NCSES)

John Finamore (NCSES)

Ben Reist (Census)

Center for Adaptive Design (Census)

Contact:

Stephanie Coffey

stephanie.coffey@census.gov