Center for Adaptive Design Update

Presented to the Census Scientific Advisory Committee
March 30, 2017

Michael Thieme, Ben Reist
Center for Adaptive Design
Research and Methodology Directorate
U.S. Census Bureau

Overview

- Improved survey representativeness without affecting response rates when compared to experimental control groups
- Enabled case prioritization that helps survey directors make informed cost/quality trade-off decisions during data collection operations
- Strengthened the Census Bureau Research and Methodology Directorate as an essential innovation partner with internal program areas and with our external partners

Overview

- Fielded Adaptive Design methodology experiments:
  - National Survey of College Graduates (NSCG) between March and August 2015, next experiment in the field April 2017
  - National Health Interview Survey (NHIS) between July and September 2016
  - Survey of Income and Program Participation (SIPP) between April and July 2016 and currently in the field

Overview

Achieved substantial IT system development in support of adaptive capabilities

- Operational control prototype used for the first time in the 2016 Census Test
- Established both lab and production instances of the Concurrent Analysis and Estimation System (CAES)
- Led the development to modernize four existing Census Bureau data collections systems (MCS, ROSCO, LCM, and UTS) enabling the adaptive case prioritization functionality used in this year's field experiments

National Survey of College Graduates (NSCG)

Metrics - Monitoring

Metrics to Evaluate Goal Achievement	Data Collection Goals			
	Sample Balancing	Controlling Costs	Increasing Timeliness	Reducing Burden
Response propensities		✓		
Response propensities by mode			✓	✓
Partial R-indicators	✓	✓		
Stability of estimates		✓	✓	✓
Interviewing paradata (e.g., locating attempts)		✓	✓	✓

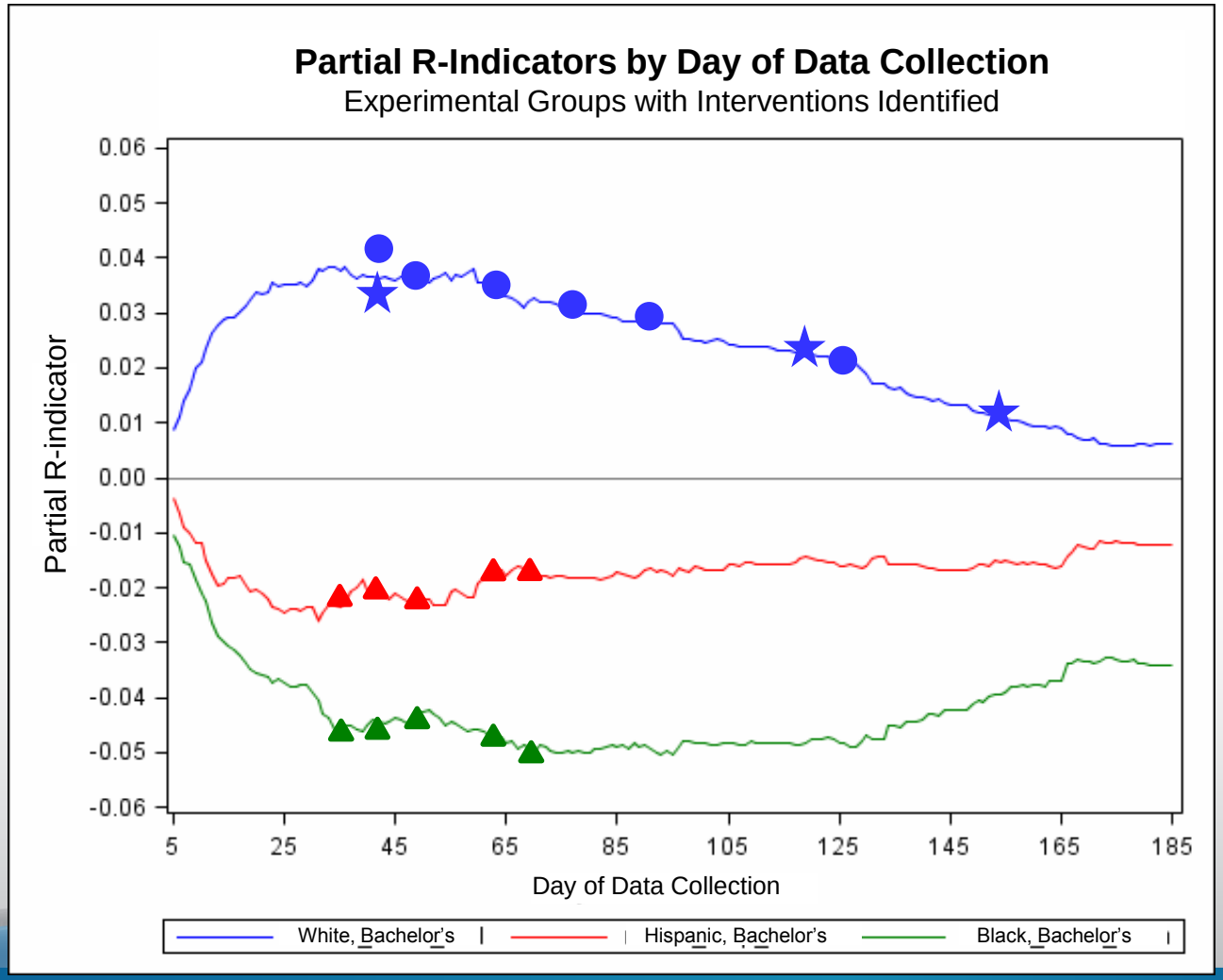
R-Indicators

- Sample R-Indicators $$R(\hat{\rho}) = 1 - 2\left(\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\frac{s_i}{\pi_i}(\hat{\rho}_i - \hat{\bar{\rho}})^2}\right)$$
  - Evaluate representativeness of respondent population as compared to the sample population, given a set of balancing variables
- Unconditional Partial R-Indicators
  - Variable-Level $$R_u(\text{var}, \hat{\rho}) = \sum_{k=1}^{K}\frac{N_k}{N}(\hat{\bar{\rho}}_{x,k} - \hat{\bar{\rho}}_x)^2$$
    - Evaluate which variables are driving the variation in propensities
  - Category-Level $$R_u(\text{var}, k, \rho) = \sqrt{\frac{N_k}{N}}(\bar{\rho}_{x,k} - \bar{\rho}_x)$$
    - Evaluate which subgroups of a variable or a cross of variables are over- or under-represented

2013 NSCG Experiment

- Research question: Is it possible to use adaptive design techniques in a federal survey?
- Treatment group sample size: 4,000 new sample cases
- Monitoring: Partial unconditional R-indicators
- Interventions
  - Overrepresented groups: sent web invites rather than paper questionnaires and put “on hold” in CATI
  - Underrepresented groups: moved cases to CATI

Interventions



Overrepresentation Interventions

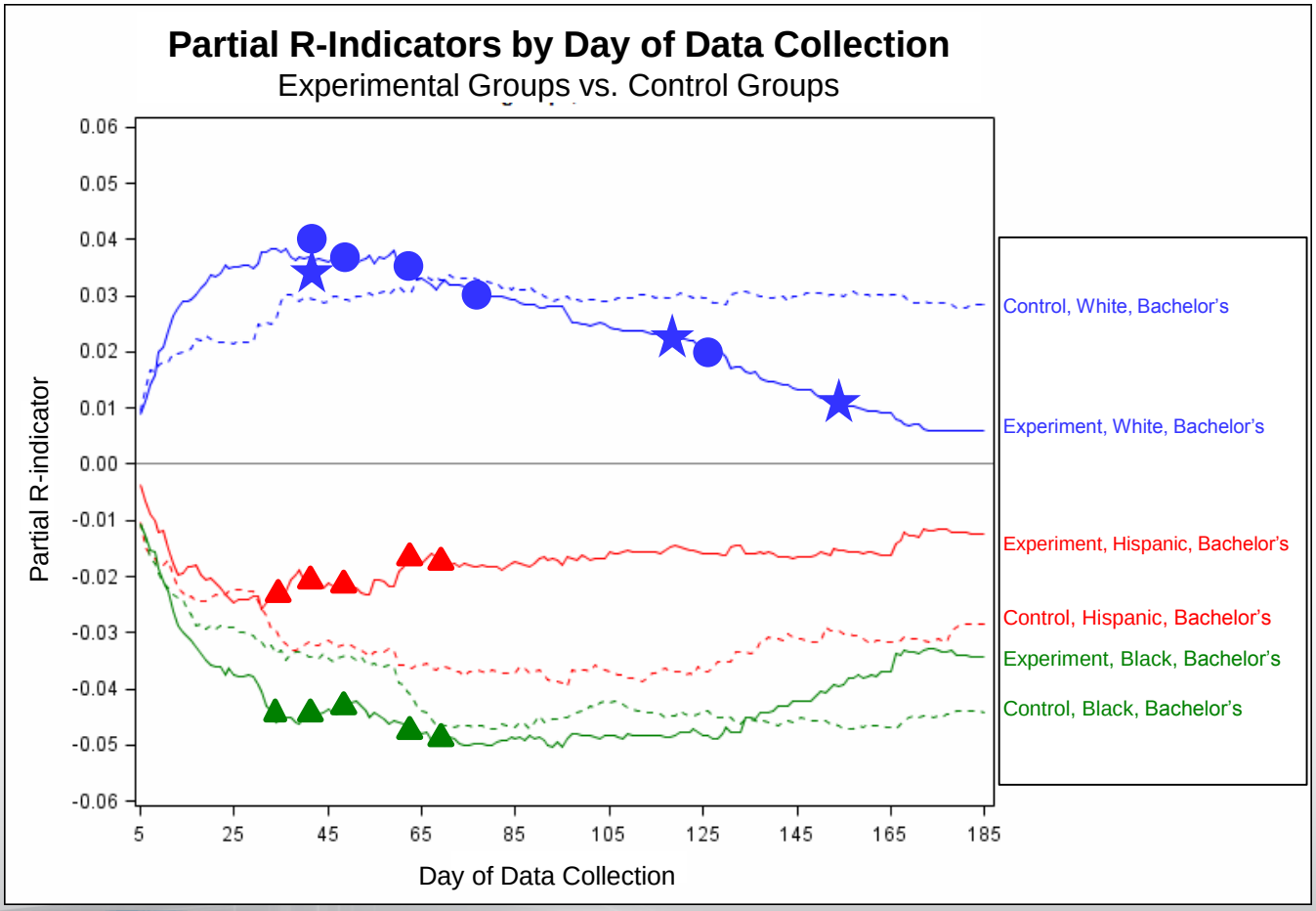
● Cases put on hold in CATI

★ Cases sent web invites rather than paper questionnaires

Underrepresentation Interventions

▲ Cases moved to CATI

Interventions vs. Control



<u>Overrepresentation Interventions</u>

● Cases put on hold in CATI

★ Cases sent web invites rather than paper questionnaires

<u>Underrepresentation Interventions</u>

▲ Cases moved to CATI

Conclusions – Successes 2013

- Operational success
  - Infrastructure worked
  - Can send frequent interventions to modes
  - Need to keep in mind mailing delays

- Interventions led to statistically significant changes trends of partial unconditional R-indicators

- Saved ~16% per case in the treatment group vs. control group

Conclusions – Limitations 2013

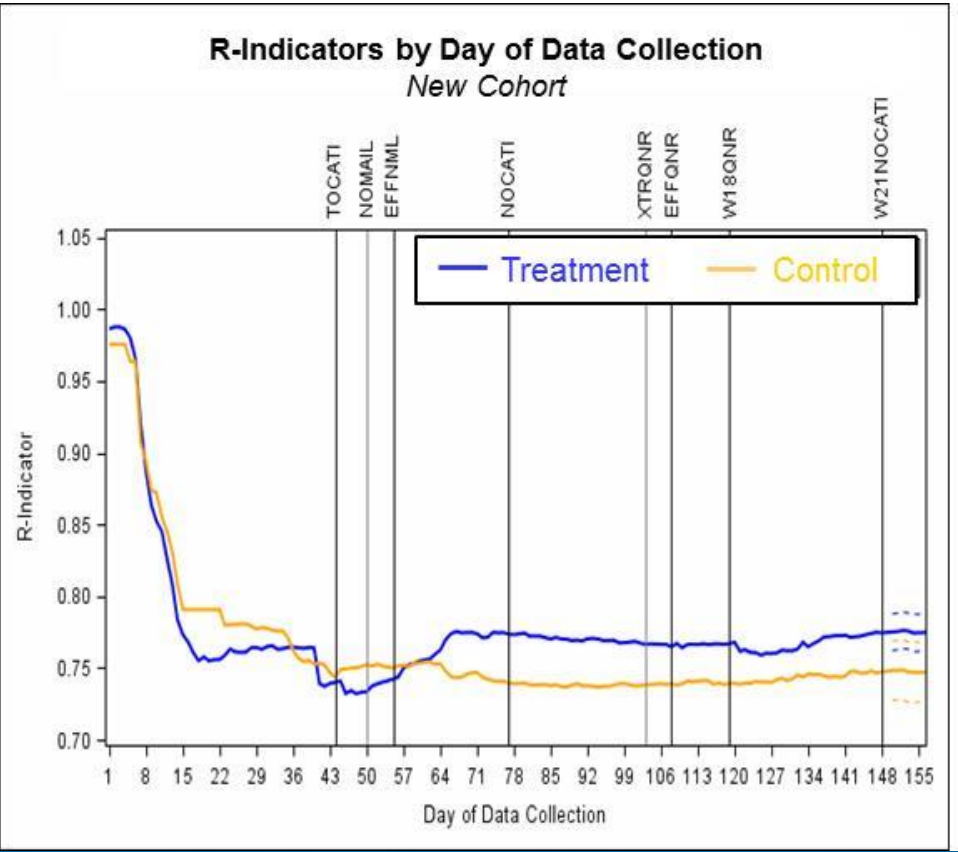
- Limitations
  - Small sample sizes (some groups <100)
  - Large standard errors on evaluation metrics
  - Made many small interventions – difficult to determine effectiveness of any particular one
- Most effect came from reducing contact on over-represented cases
  - Led to statistically significant reduction in response rate
  - "Gaming the R-indicator"?

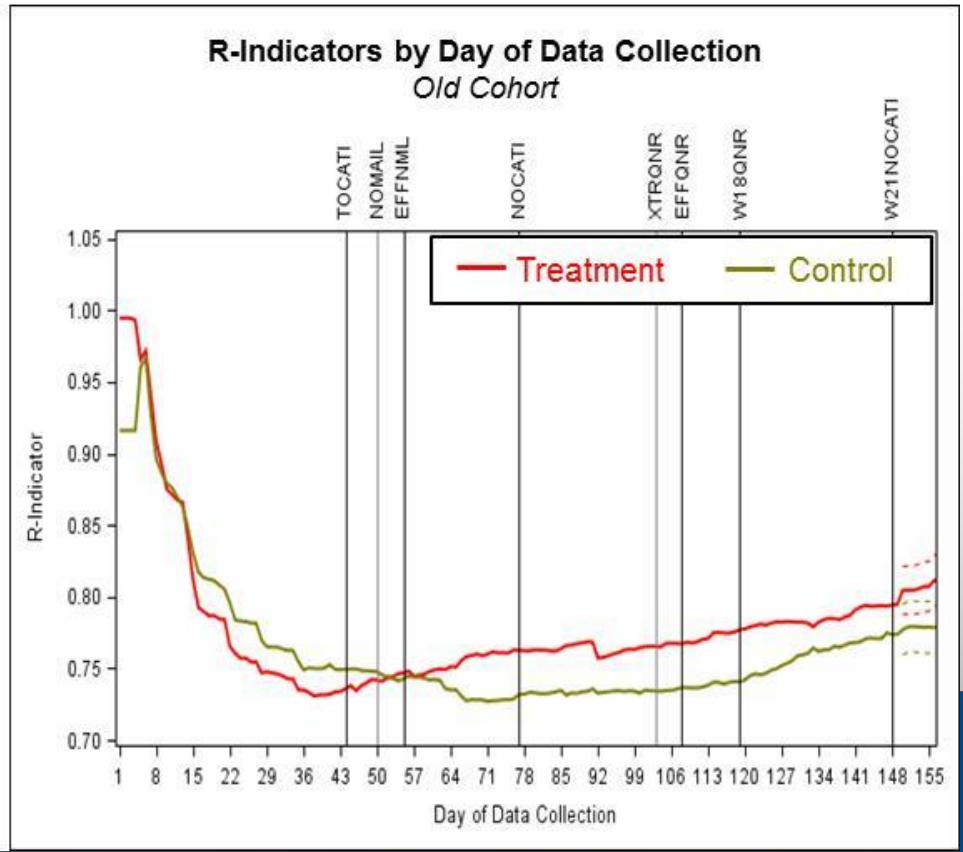
2015 NSCG Experiment

- Larger sample sizes
  - 8,000 cases in the new cohort (same in cntl)
  - 10,000 cases in the old cohort (same in cntl)
- Goal to evaluate specific interventions
  - Fewer, larger interventions
- Additional interventions
  - Send off-path questionnaires
- Targeted incentive in production for all cases
  - Will incentive remove need for mode switching?
- OMB involvement

2015 NSCG AD Test: Finding #1

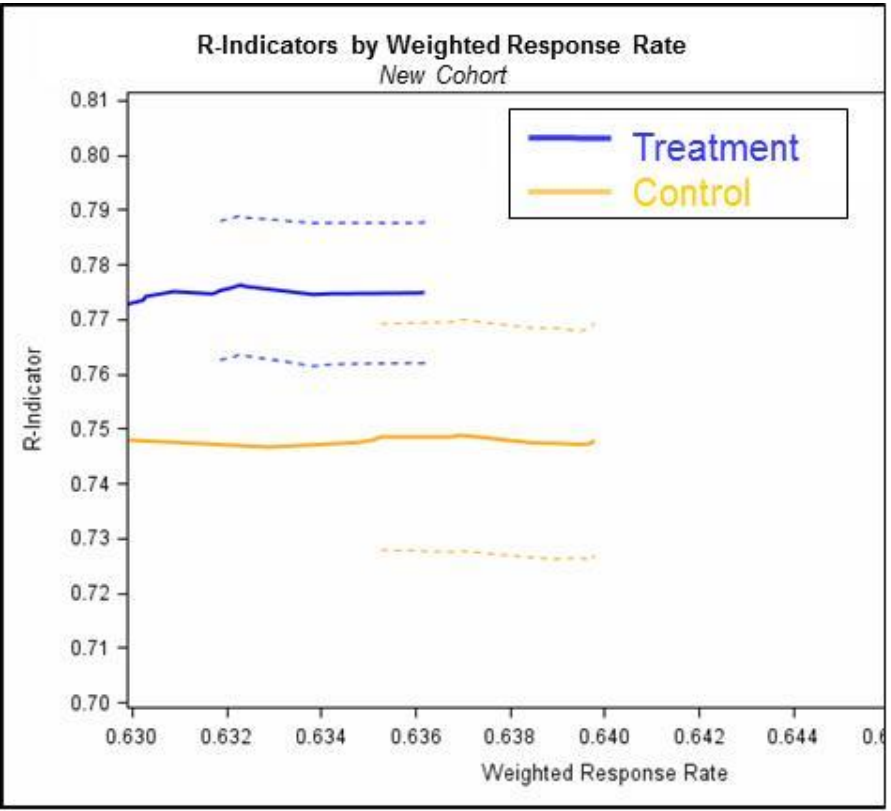
In both cohorts, once we started making interventions, the *R-indicator point estimate* for the treatment groups became <u>and stayed</u> more representative

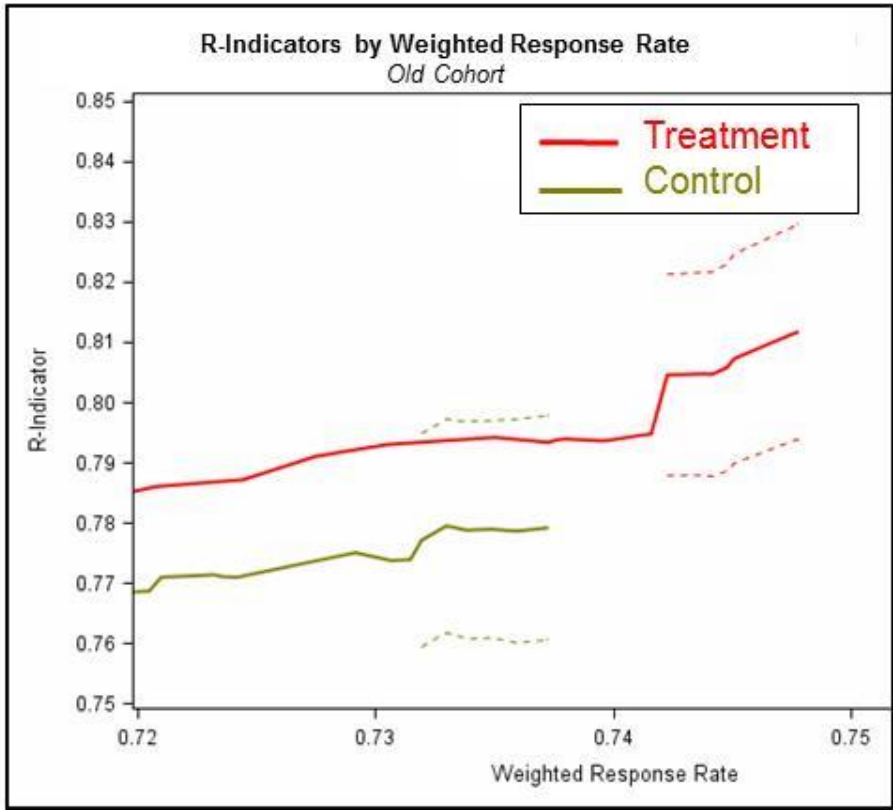




2015 NSCG AD Test: Finding #2

In both cohorts, we were able to reallocate resources through interventions without any adverse impact on response rate.

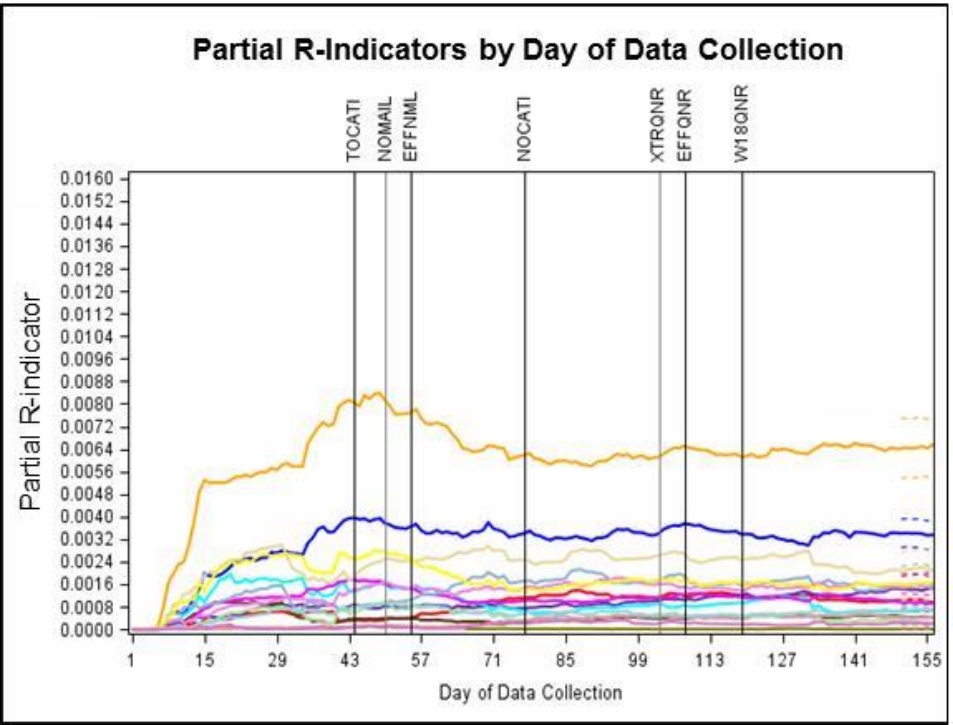




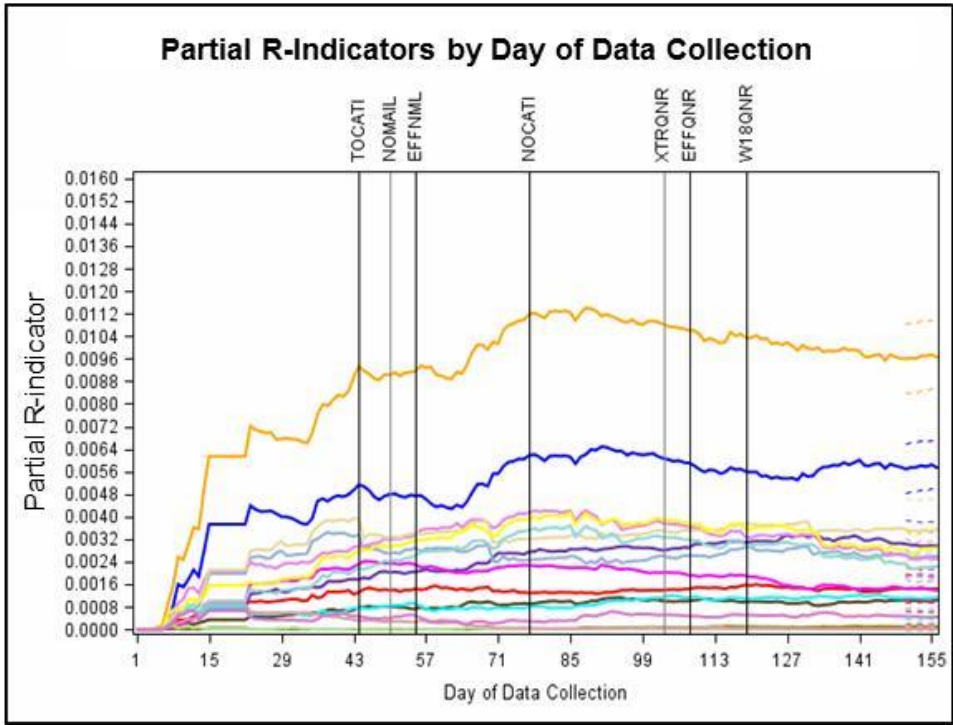
2015 NSCG AD Test: Finding #3

For the treatment groups in both cohorts, the improvement in variable-level representation is statistically significant for the variables on which we intervened (**age x income**, **age**).

Treatment Group



Control Group



Conclusions & Successes 2015

- Dynamic adaptive design requires a lot of planning, effort, and oversight
- All of the interventions were designed to improve sample balance concerns
- Interventions improved representation (sample balance) without any significant reduction in response
  - Likely due to the combination of the targeted incentive and adaptive design

Case Prioritization in the NHIS: Q3 2016

Motivation

- Case prioritization
  - Help reach data collection goals subject to constraints
    - Increase representativeness for a given cost
    - Maintain response rate for a lower cost
  - Idea is to alter the effort an interviewer spends on a case
    - Need to define data collection goals
    - Assign relative values to cases
  - Can be applied adaptively
    - Take advantage of information acquired during data collection
      - Contact History Instrument
      - Neighborhood Observations (uncommon at Census)
      - Cross-sectional survey – cannot leverage past responses

Conceptual Overview

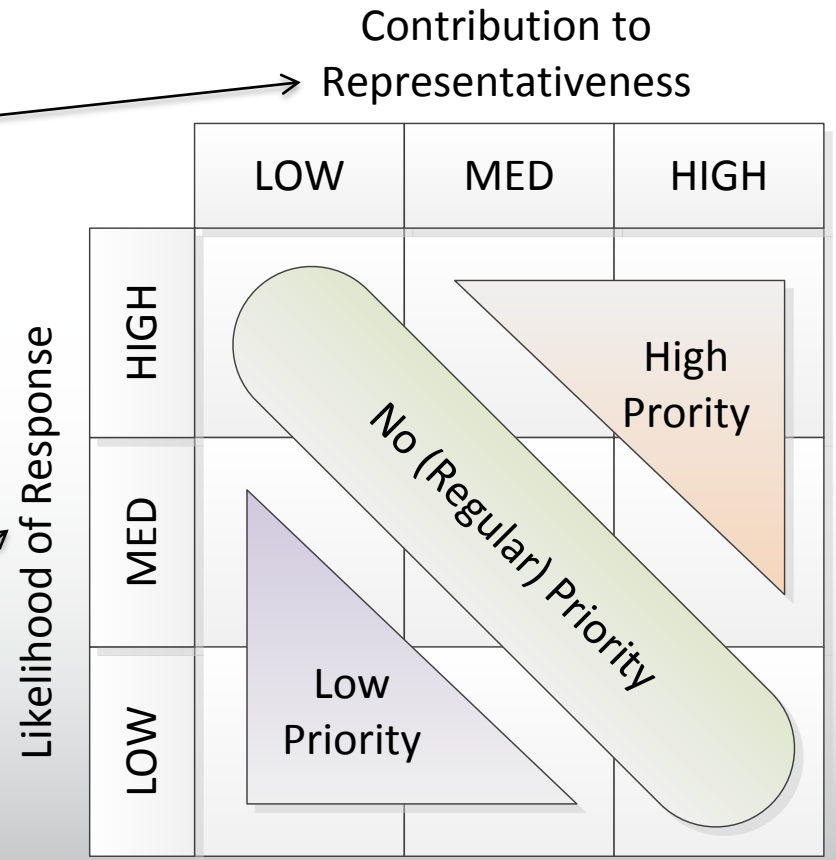
- Existing structure: Four-week data collection
  - First visit attempt: Neighborhood Observation Instrument (NOI)
    - Some items correlated with NHIS outcomes of interest
  - All contact attempts: Contact History Instrument (CHI)
    - Some items correlated with response propensity / effort
- Additional information
  - Sample information: weights, geocodes/location
  - Planning Database (PDB): block-group level aggregate data
    - Some items correlated with NHIS outcomes or effort
  - Have this information before data collection starts
- Leverage all of this information: Assign values and prioritize

Prioritization

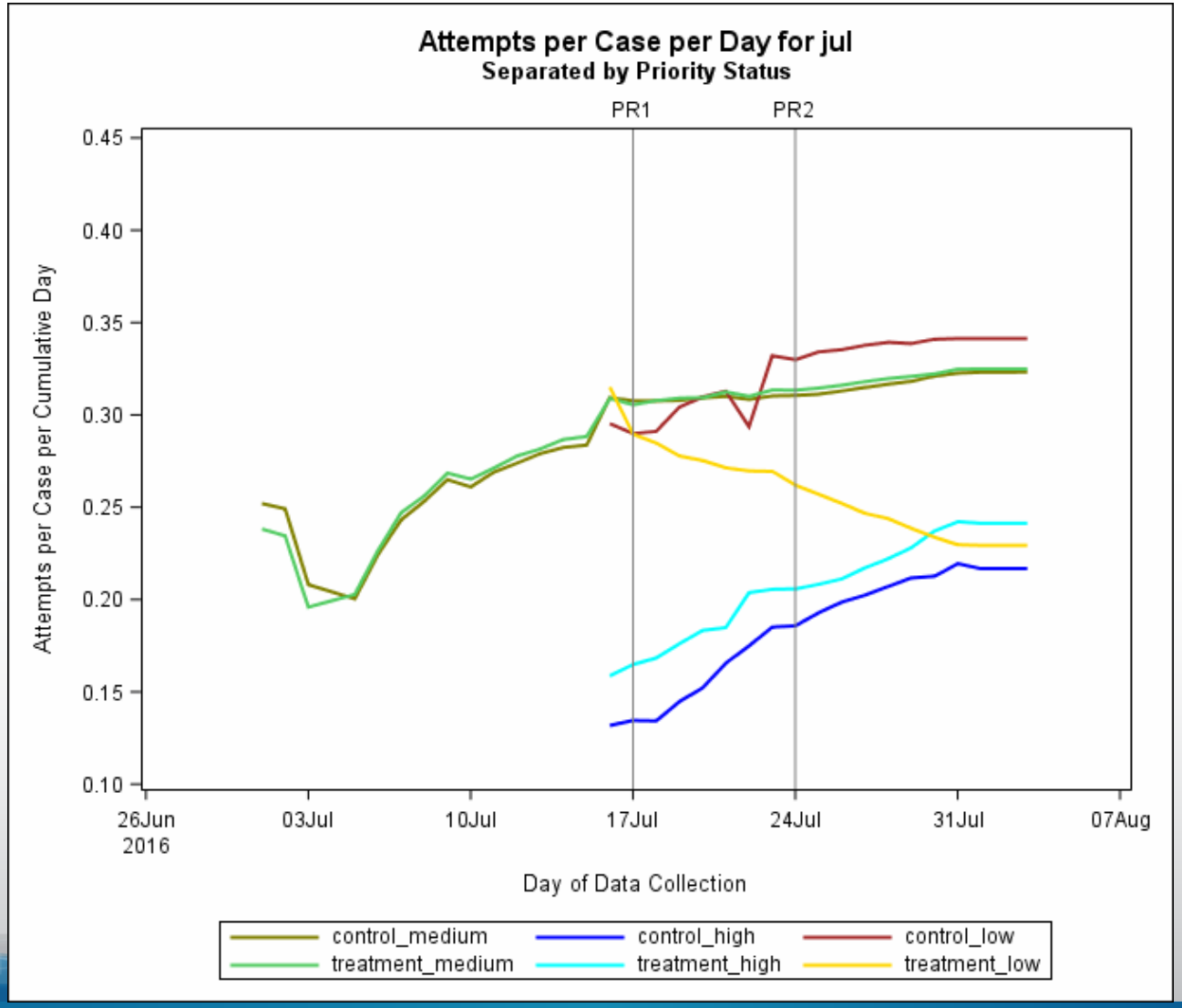
Build a response propensity model that *only* focuses on predicting estimates of interest (does not consider response)

How common is a type of case in the respondent population versus how likely is it for a specific case to respond?

Build a response propensity model that *only* focuses on predicting response (does not consider estimates of interest)



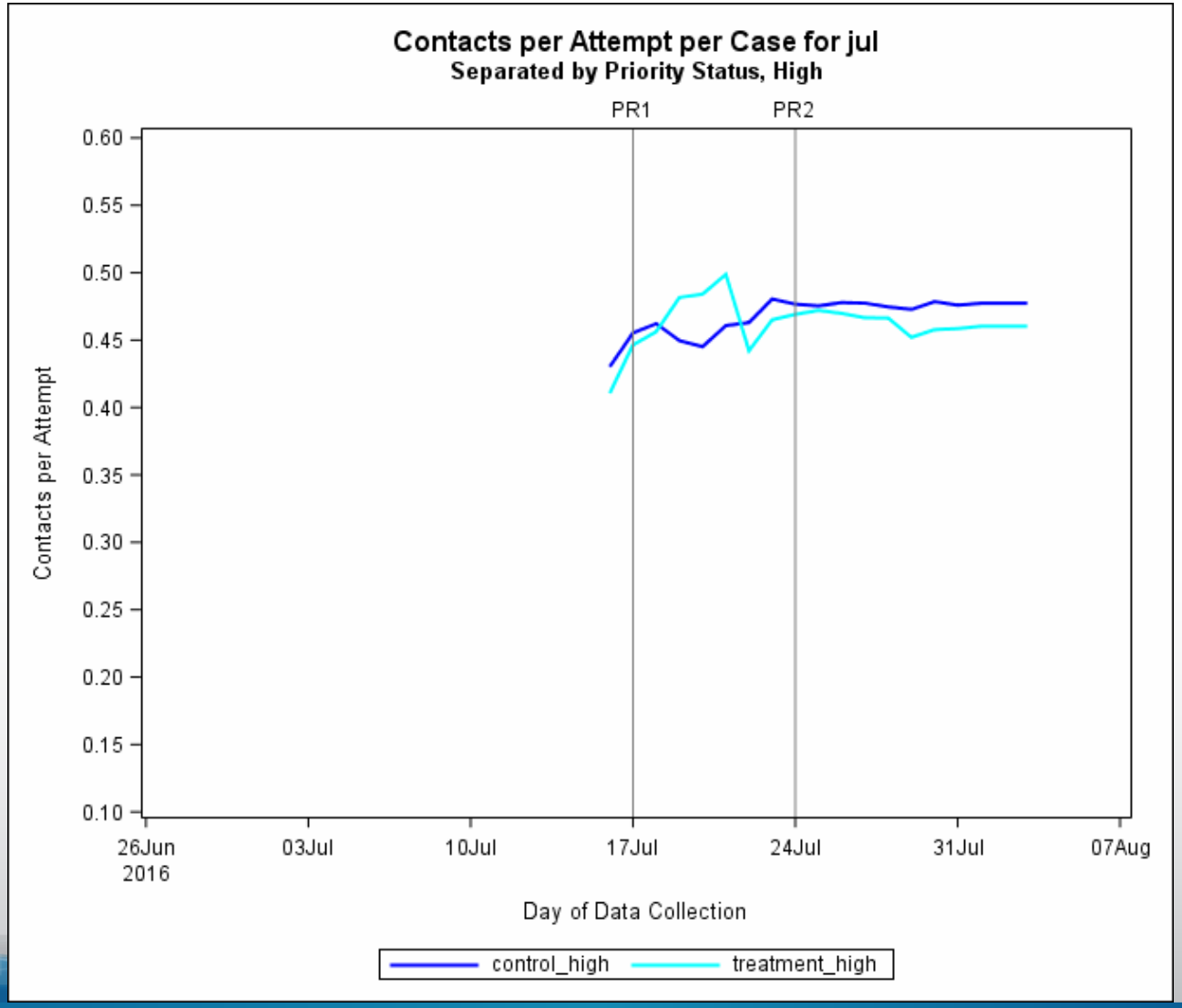
Average Attempts per Case per Day



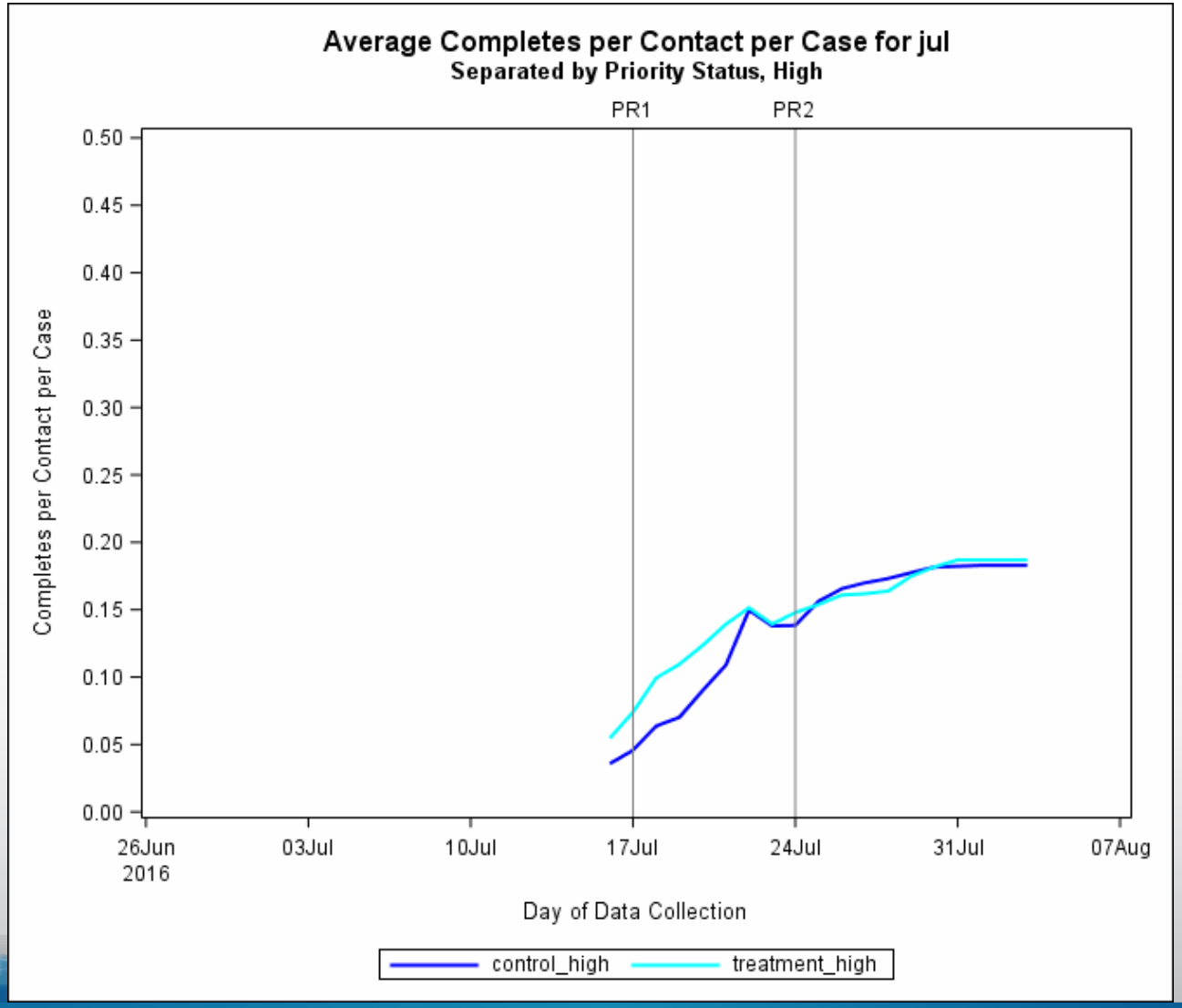
Does Prioritization Improve Outcomes?

- In July: Higher attempt rates in High Priority cases
- Did this result in:
  - Higher Contact Rates (Contact per Attempt Rate)? ✓
  - Higher Completion Rate (Completions per Contact)? ✓
- Interviewers shifted effort in an effective way
- Did not translate to top-line outcomes:
  - Small amount of prioritization (20% of open T cases)
  - Translates to about 7% of total workload

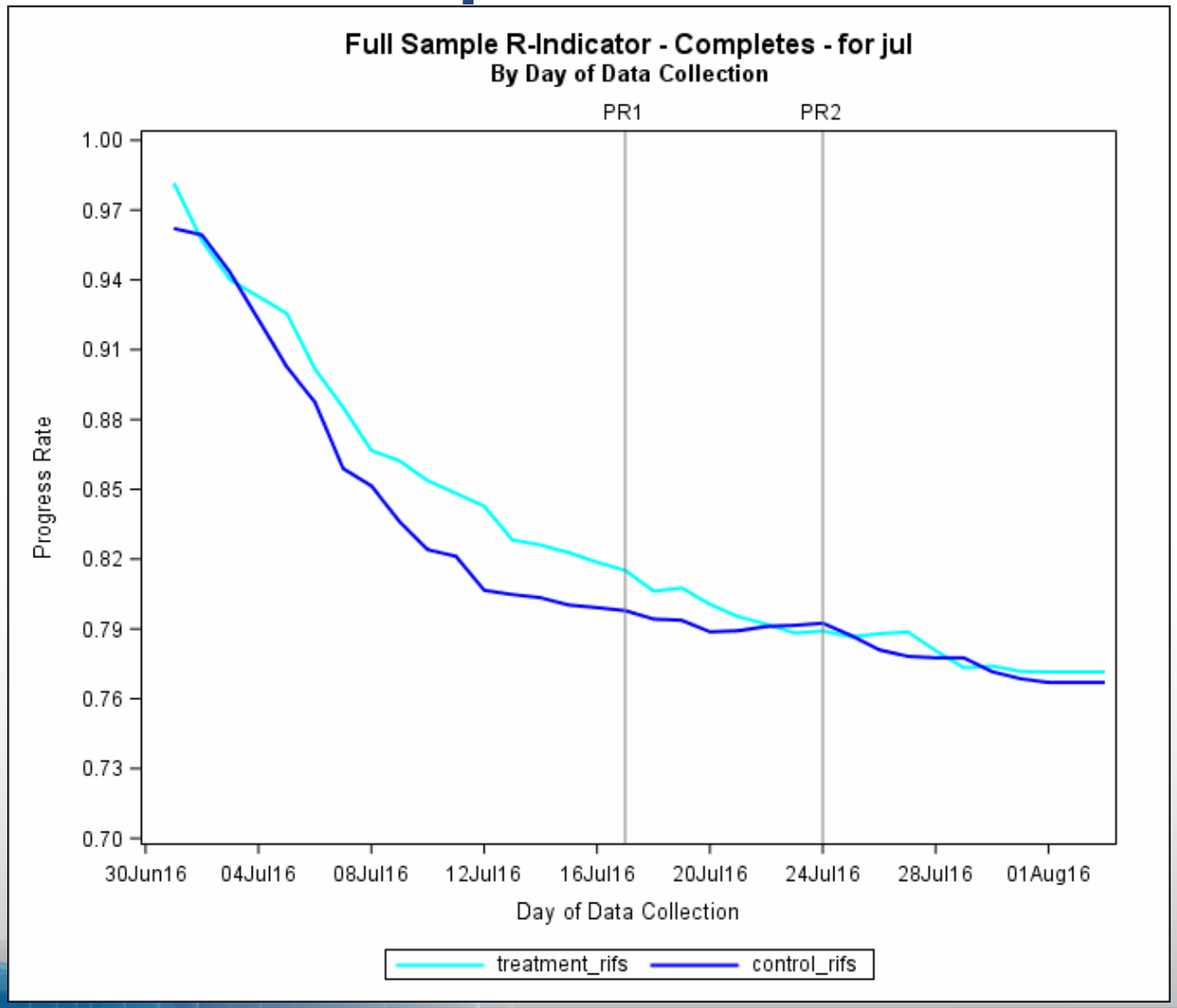
Does Prioritization Improve Outcomes?



Does Prioritization Improve Outcomes?



Full Sample R-Indicators



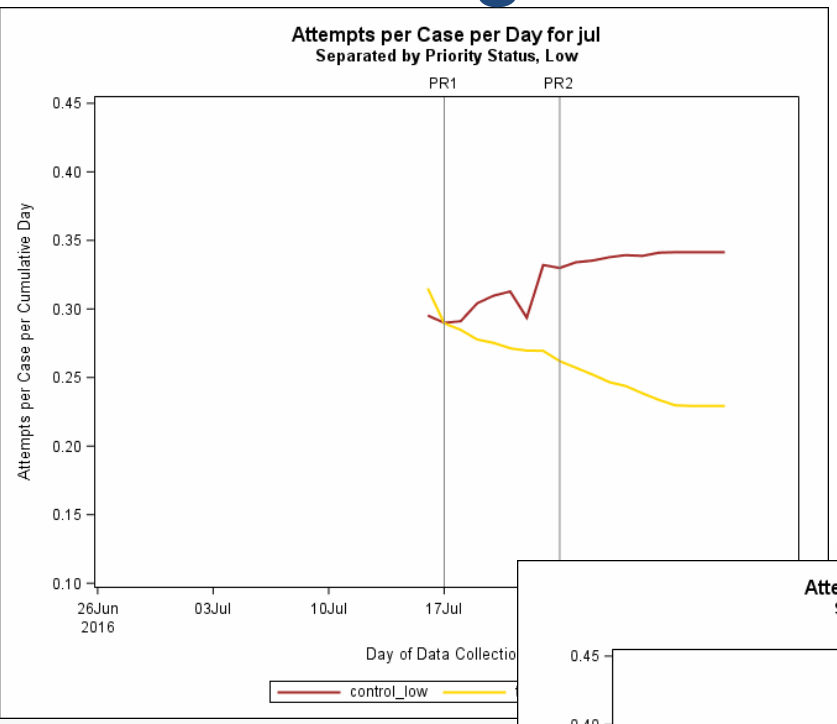
Summary

- Our treatment protocol was able to shift effort to more valuable cases during Week 3 of July.
- Shifted effort resulted in more attempts, a greater success rate of contacting respondents.
- Greater contact success in treatment led to a temporary bump in completions per attempt.
- No effect on overall completed interview number.
- No effect on full sample R-indicator for completed cases.
- Results were not replicated in Aug or Sept
- We have experimental data for further simulation

Field Implementation Issues

- Design: Case Sharing
  - Cases sometimes assigned to multiple FRs or both T/C FRs
  - Causes difficulty in implementation and evaluation
- Experiment: No Metrics Tied to Protocol
  - Interviewers followed protocol in July
  - Seemed to fall into past behavior for Aug, Sept
  - Need to provide clear metrics for continued compliance

Average Attempts per Case per Day



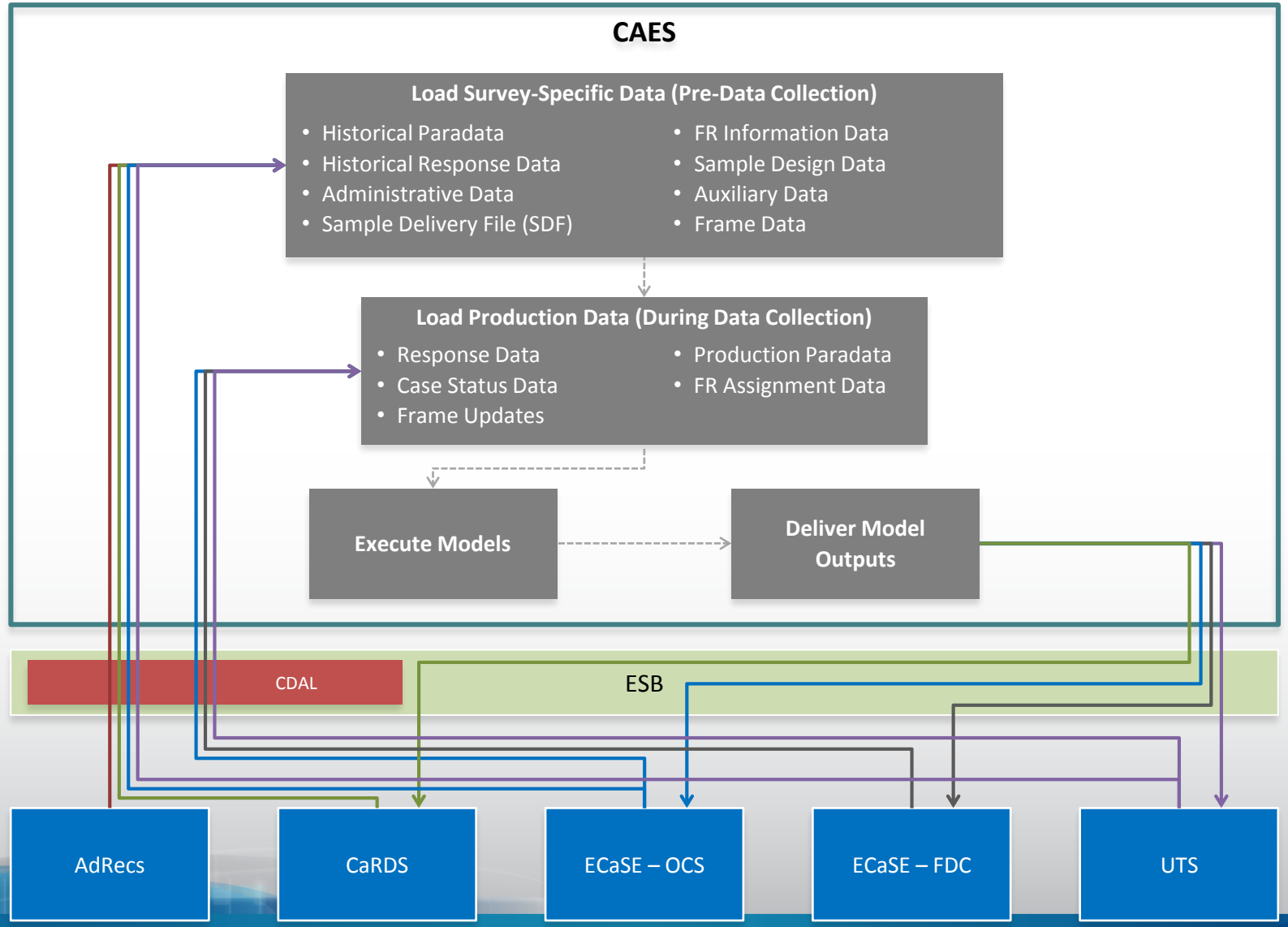
Concurrent Analysis and Estimation System (CAES)

CAES Overview

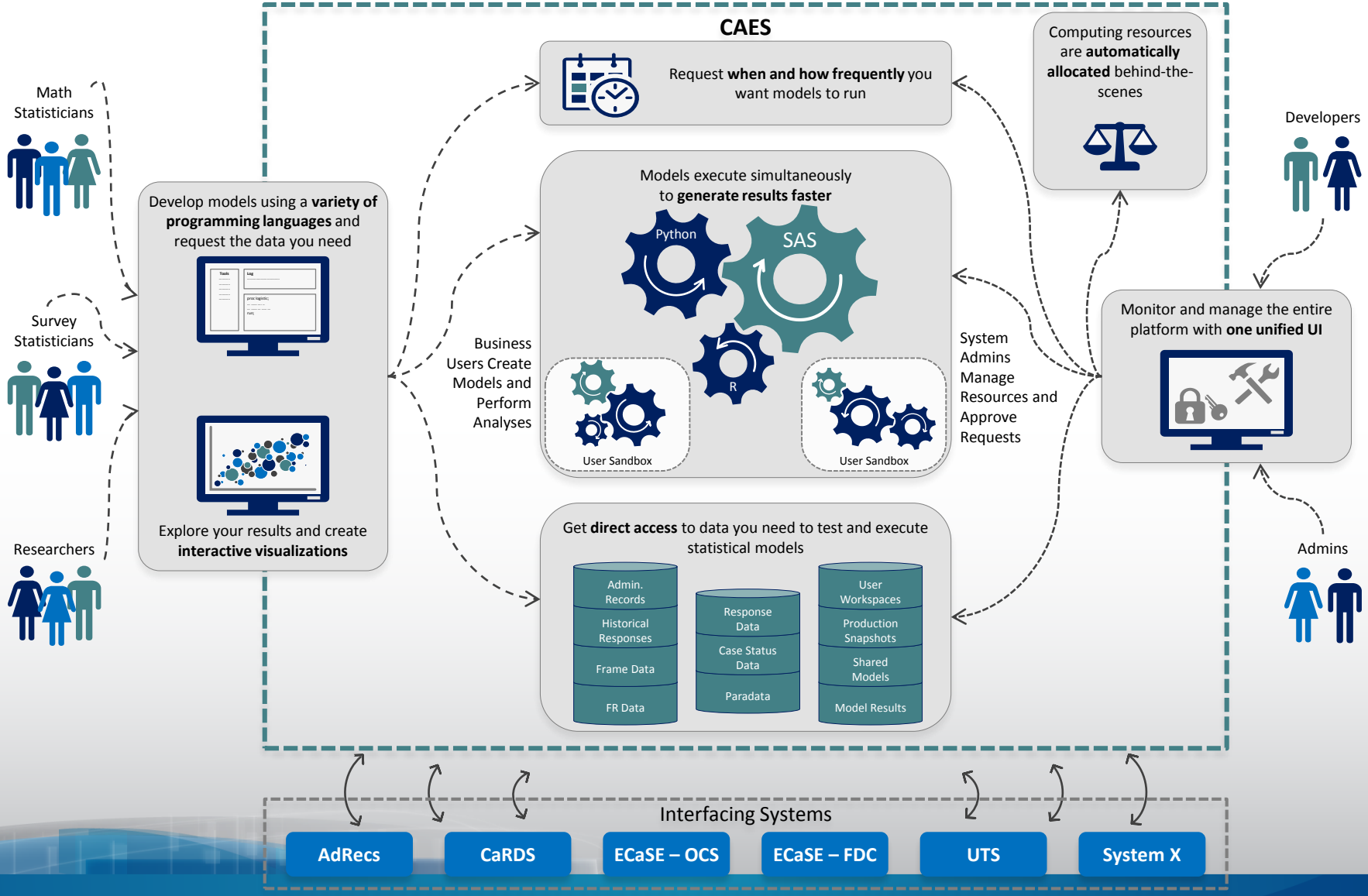
CAES is:

- A secure production hardware environment
- About three things:
  1. Analytical processing
  2. Flexible software choices
  3. **SPEED**

CAES Information Architecture



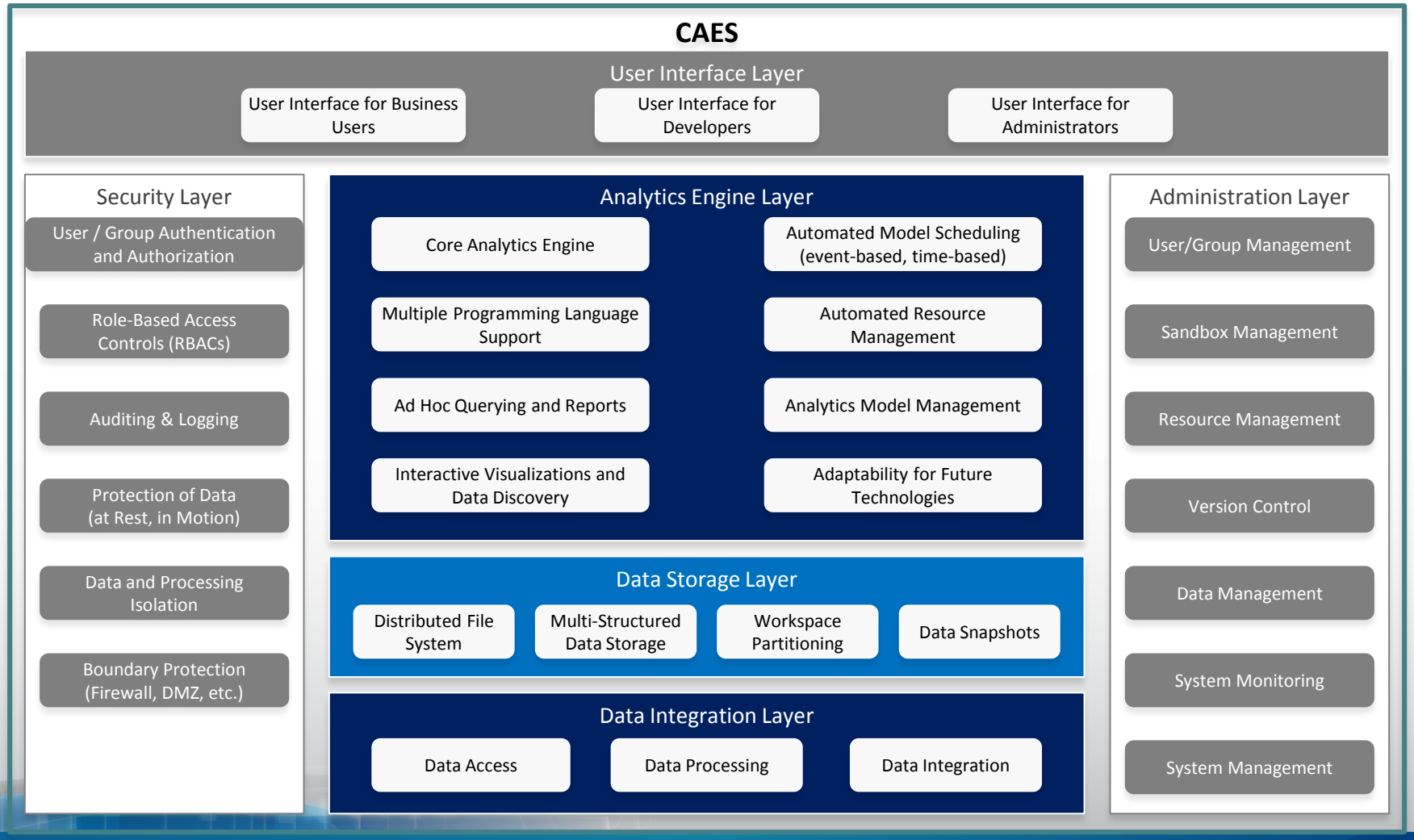
CAES User View



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

Administrator Actions — Business User Actions — Transfer of Data

CAES Logical Application Architecture



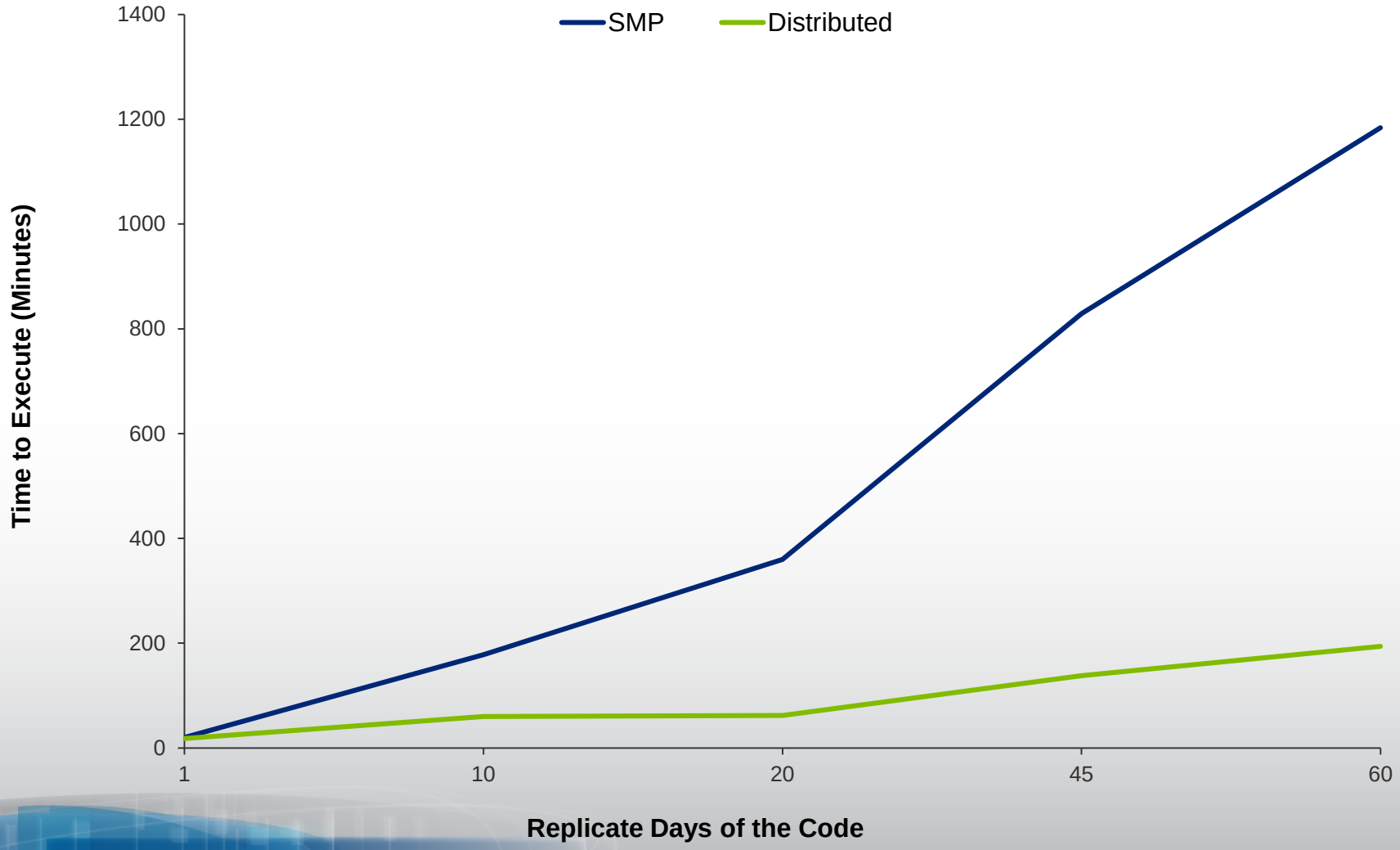
Initial Proof of Concept

Balancing Propensity and R-Indicator Model from the National Survey of College Graduates (NSCG)

- Single Server mode – approximately 4.5 days to run
- In-Memory Distributed mode (4 nodes) – approximately 1.5 days to run
- In Memory Distributed at SAS in Cary, NC (20 nodes) – approximately 3.5 hours to run
- The Big Data Platform (with HDFS) enables the storage of a large volume, velocity, and variety of data

  - Level of Effort to convert model for distributed processing was moderate
    - Meat of the code remains the same (e.g., structure of macros, specific calculations, etc.)
    - New processing ideas (e.g., partitioning data for parallel processing, using proc fcmp, etc.)
    - New modeling structure (e.g., creating temp tables instead of datasets, creating empty tables to fill with outputs, etc.)

Findings


SMP
Distributed
1400
1200
1000
800
600
400
200
0
Time to Execute (Minutes)
1
10
20
45
60
Replicate Days of the Code

CAES High Level Schedule

High Level Milestone	Date
Hadoop Cluster with initial Hortonworks and analytical SW running in lab	9/30/16
Hadoop Cluster with initial Hortonworks and analytical SW installed in Bowie	12/1/16
Receive CAES ATO	3/1/17
Run National Survey of College Graduates (NSCG) R-Indicator Model in CAES in parallel with production	4/15/17
Start testing in CAES of 2018 Decennial E2E Test Models	9/30/17

Discussion