

Conducting Targeted Nonresponse Follow-Up in the Economic Census in an Adaptive Design Framework

Stephen J. Kaputa (U.S. Census Bureau) and Yukiko Ellis (U.S. Census Bureau)

Goal

- Target establishments in the Economic Census using different collection protocols with the intention of reducing cost
- Select subsamples in low responding domains to receive a certified letter as opposed to a standard letter

Research Projects

Adaptive Design Simulation Study (2014)

- Explored optimized allocation methods for nonrespondent subsampling
- Found that subsampling nonrespondents without changing the contact strategy may have minimal benefits

Field Test (2015)

- Tested three different contact strategies to be paired with nonrespondent subsampling
- Recommended certified letter for contacting subsampled establishments
- Problem: Certified letters became a standard nonresponse follow-up method (NRFU)

Field Test (2016)

- Compared uniform certified NRFU to targeted certified NRFU and nonrespondent subsampling with certified NRFU
- **Recommended targeted certified NRFU to be implemented in the 2017 Economic Census**

Economic Census

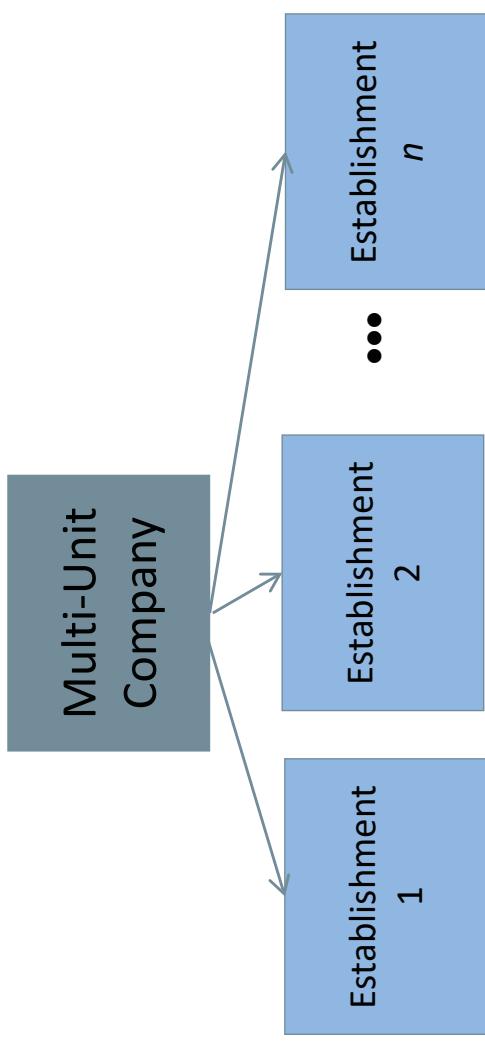
- Conducted every five years (2017, collected in 2018)
- Covers eighteen non-farm sectors
- Surveys over 4 million establishments
- Produces national and sub national benchmarks
 - Industry Totals
 - Products Totals
- Provides data for sampling frames
- Mail-out contact with online response

Business Organization Structures

Single Unit (SU)



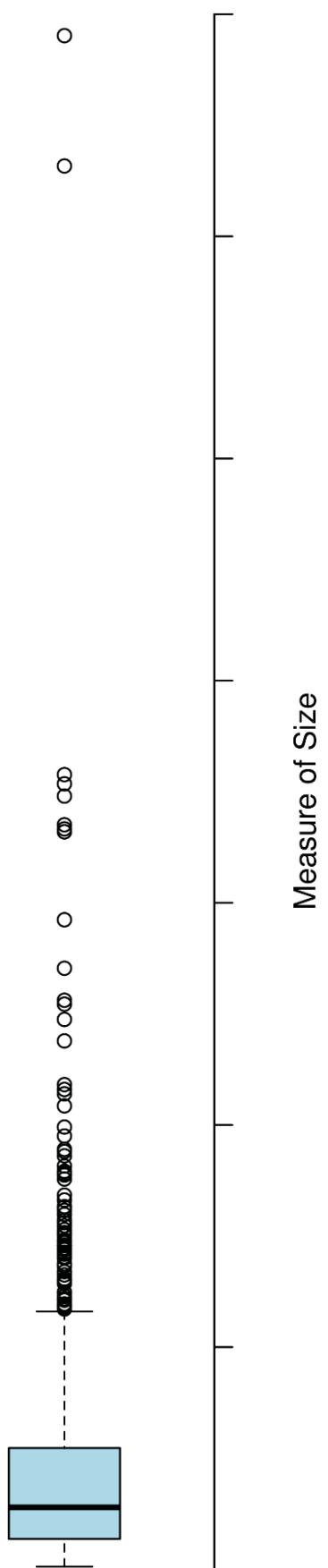
Multi Unit (MU)



One primary industry

Can operate in more than one industry

Highly Skewed Distributions

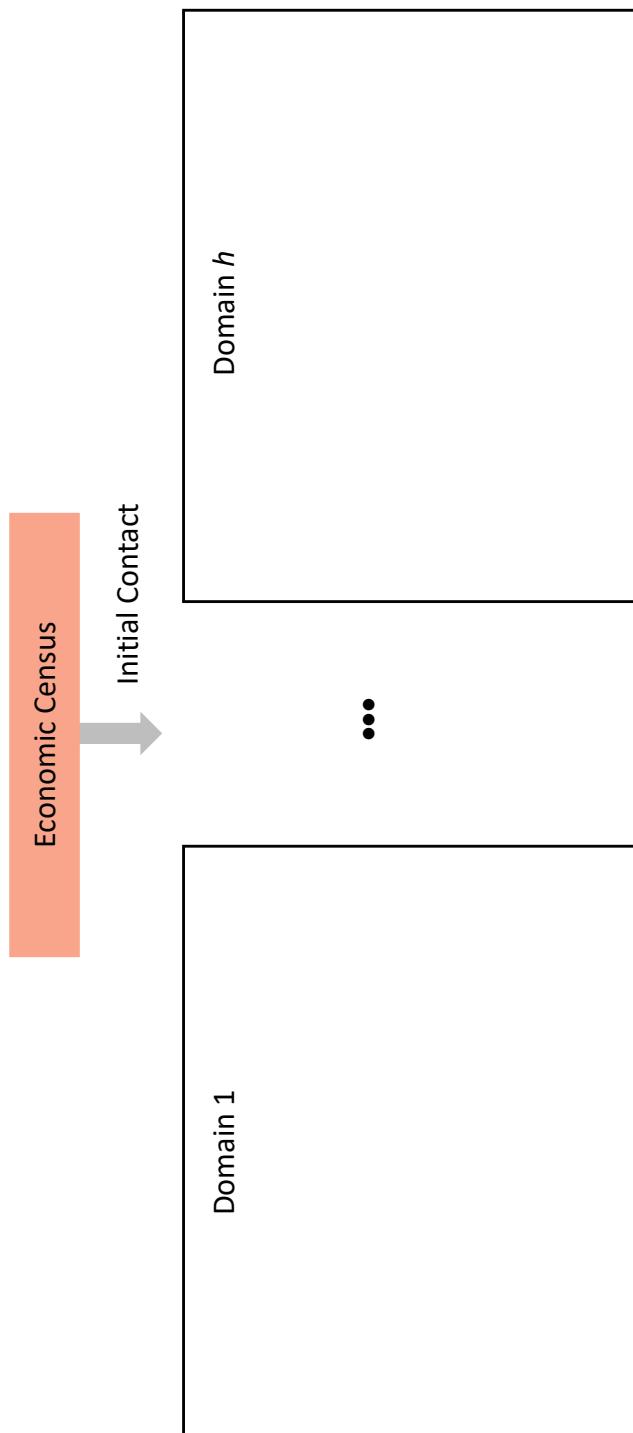


- Large establishments or companies can contribute substantially to an industry total

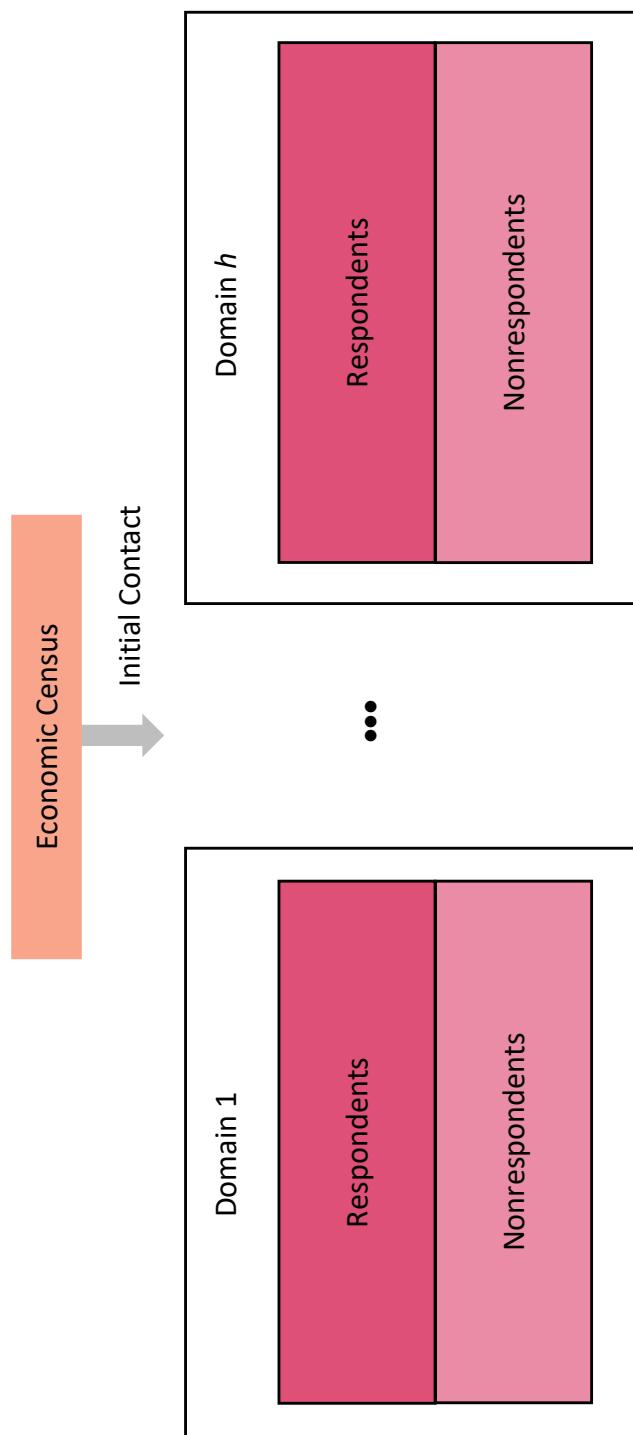
Econ Census Mail-Out Schedule

Contact Round	Method	Approximate Date
Initial Contact	Non-certified letter	May 1-3, 2018
Due Date Reminder	Non-certified letter in pressure-sealed envelope	May 29, 2018
Due Date	NA	June 19, 2018
1st Follow-up	Non-certified letter in pressure-sealed envelope	July 6-20, 2018
2nd Follow-up	Non-certified letter in pressure-sealed envelope	August 20-31, 2018
3rd Follow-up (Targeted)	Certified first class letter or non-certified (standard) letter	October 11-22, 2018
4th Final Follow-up	Non-certified Office of General Counsel letter	November 13-28, 2018
Final Closeout	NA	April 30, 2019

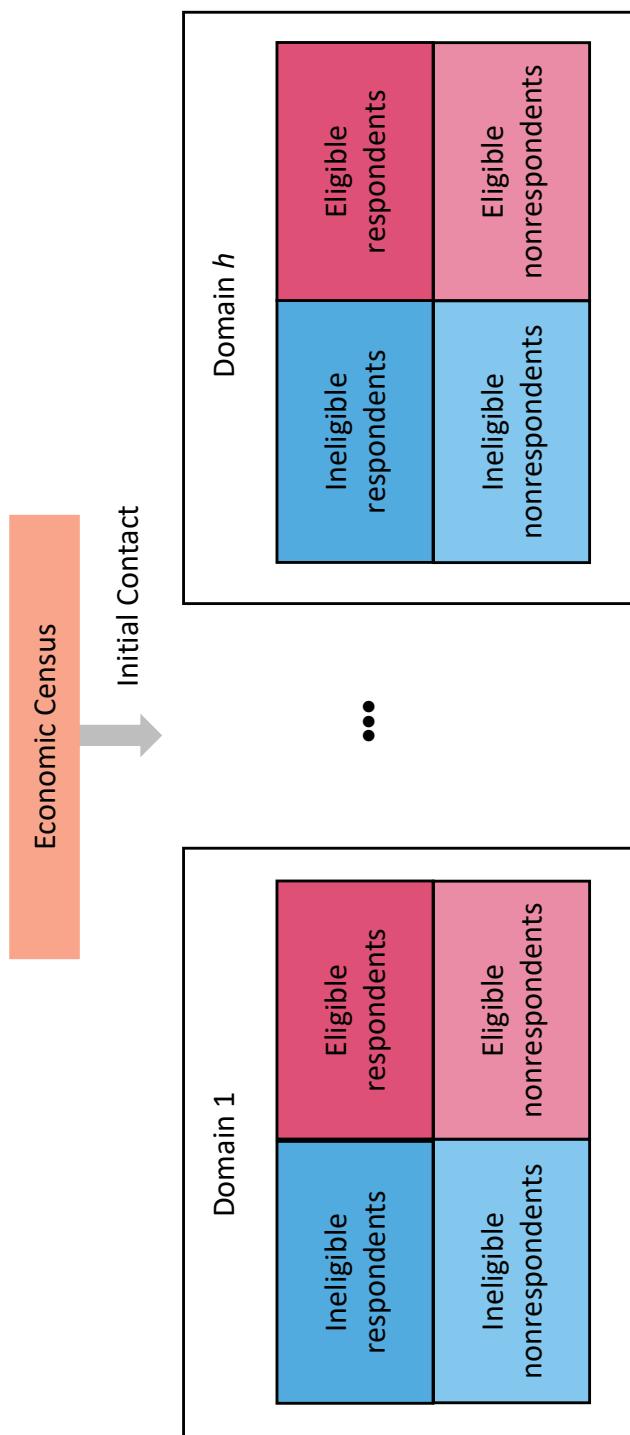
Targeted Follow-up Framework



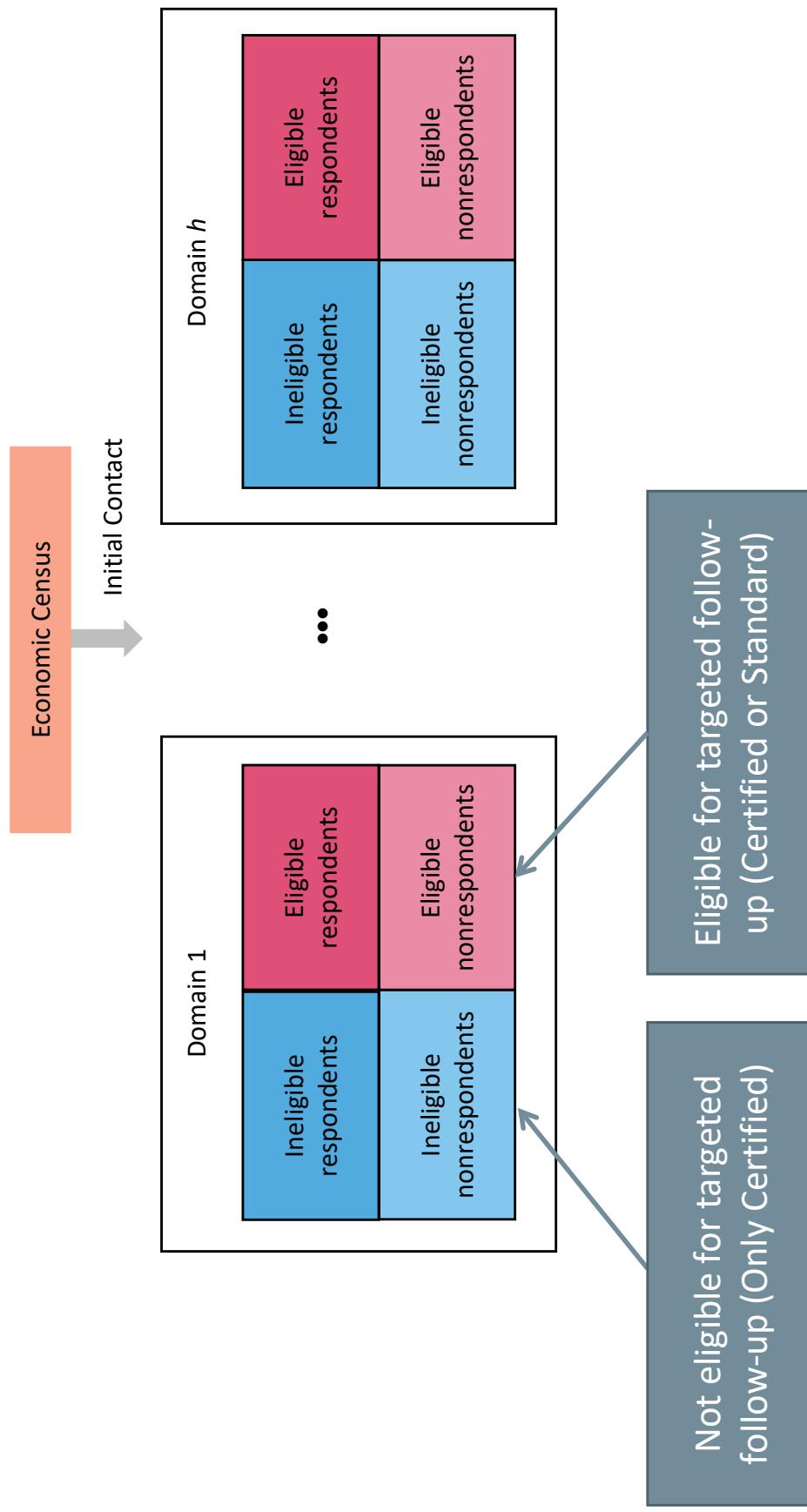
Targeted Follow-up Framework



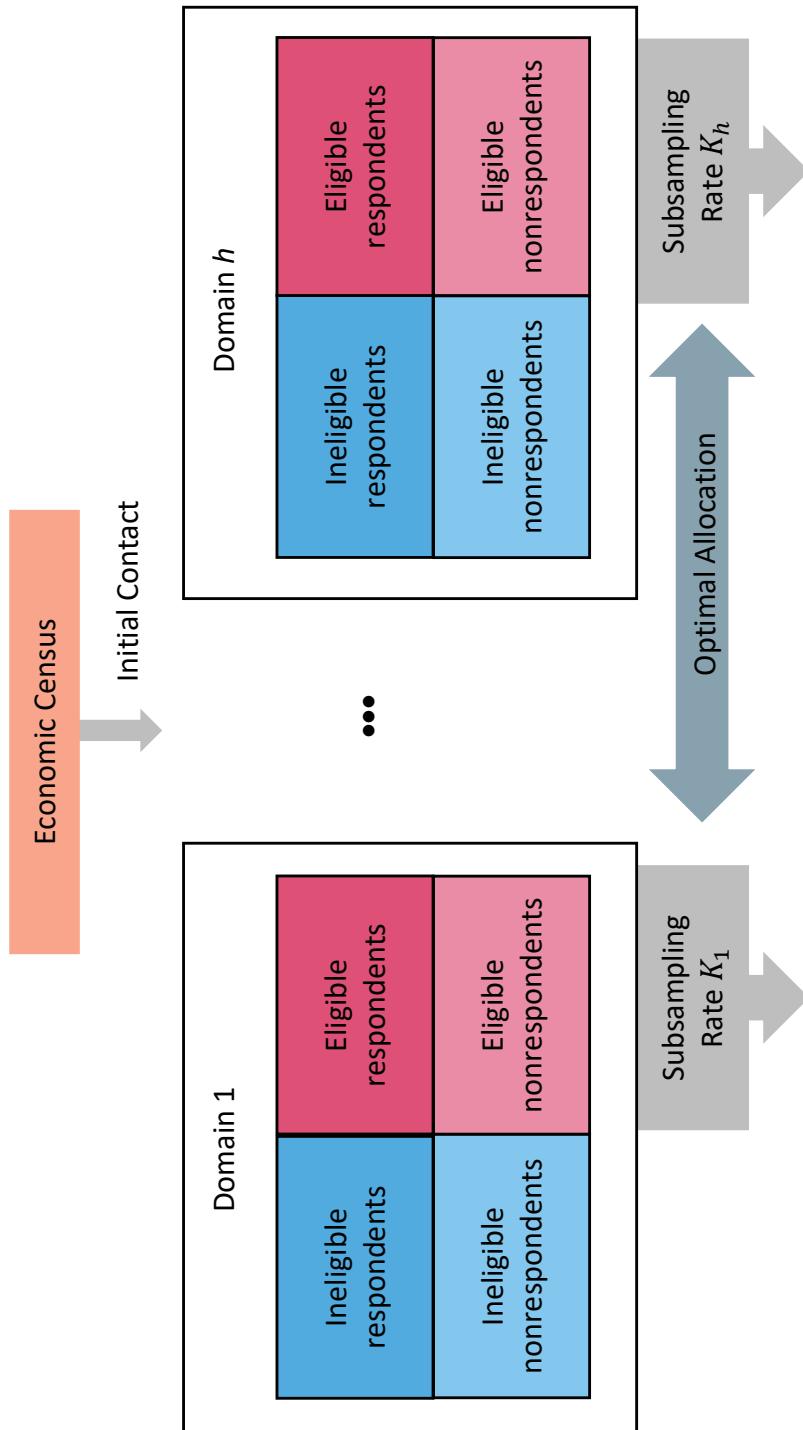
Targeted Follow-up Framework



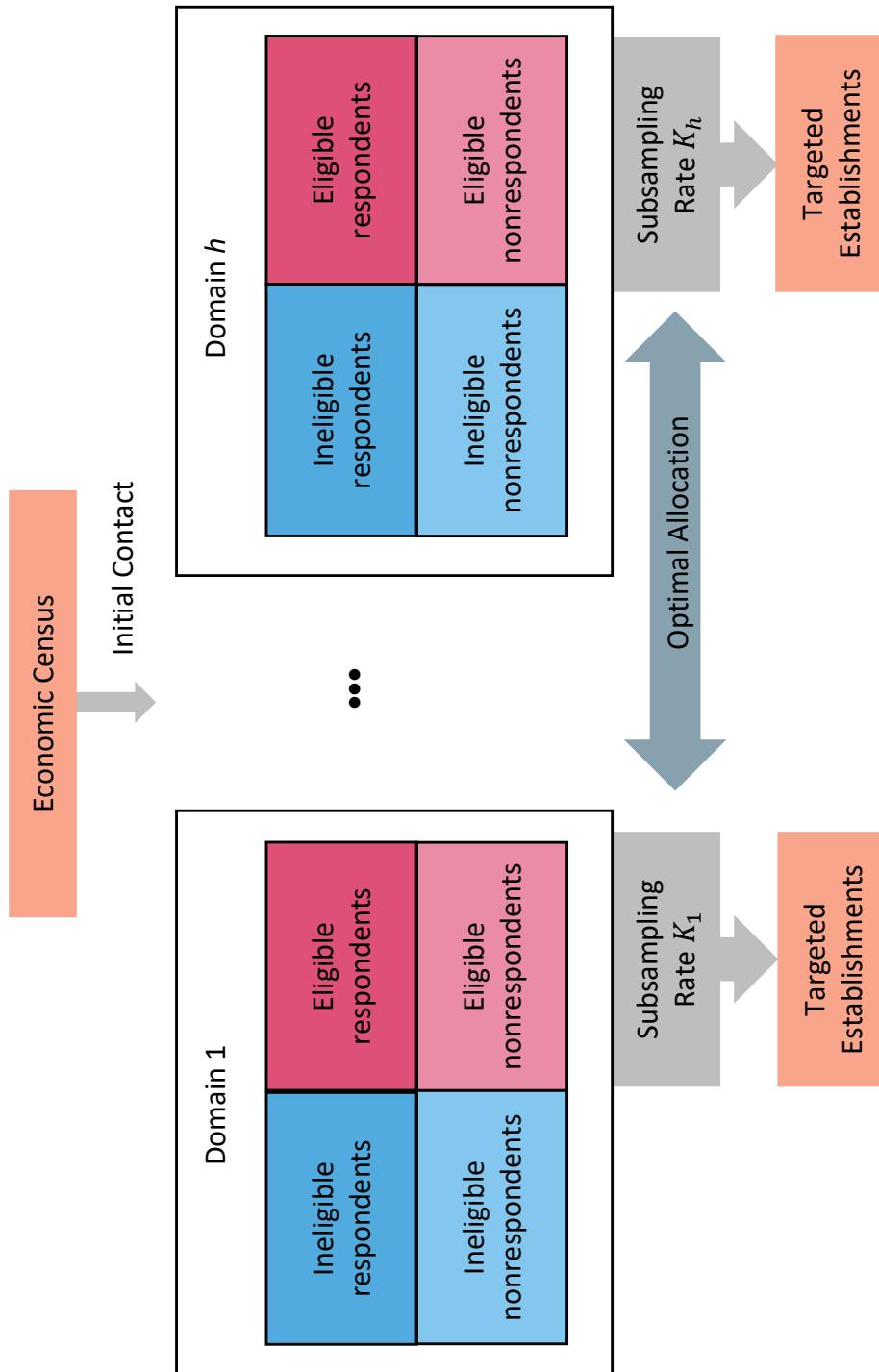
Targeted Follow-up Framework



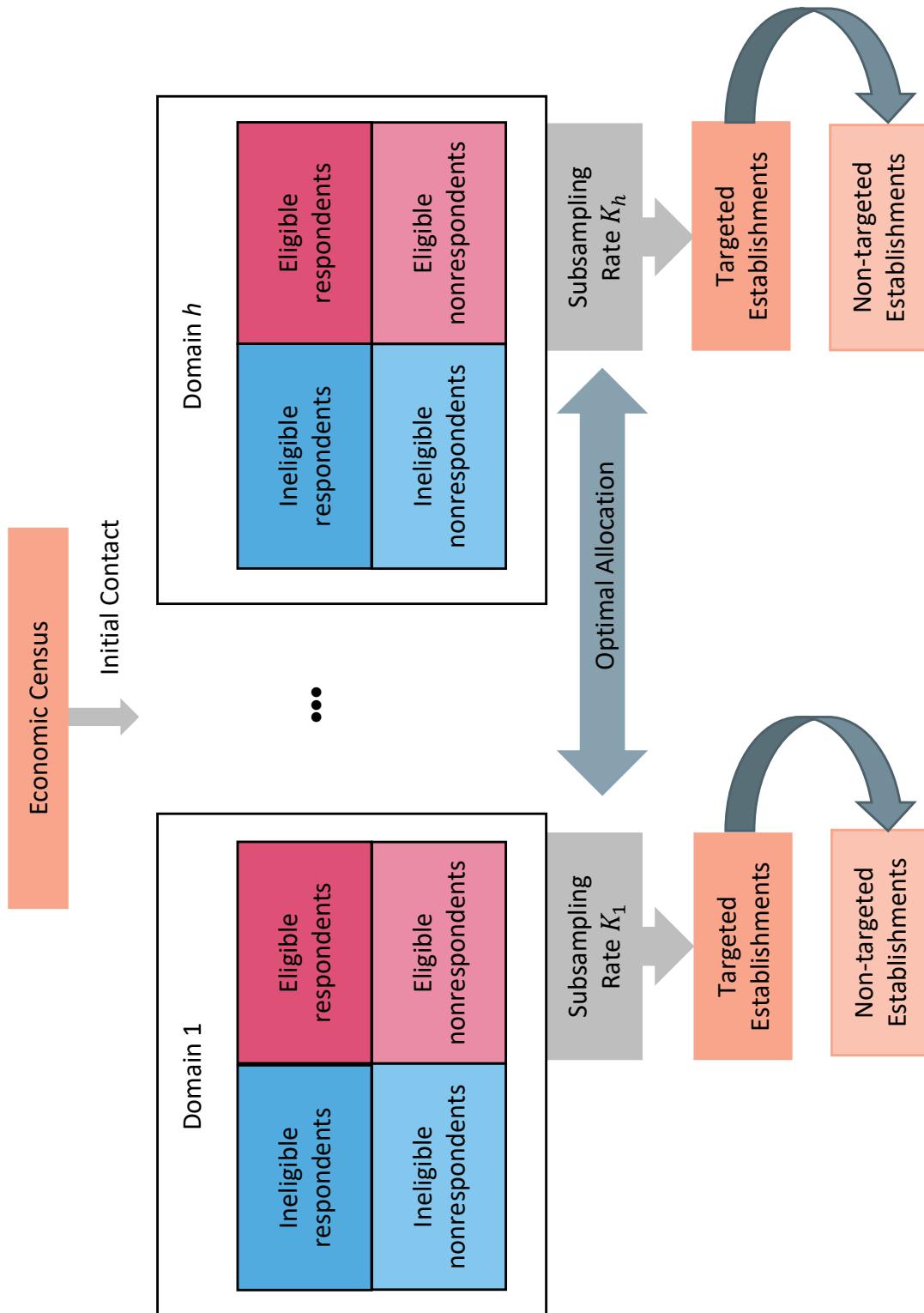
Targeted Follow-up Framework



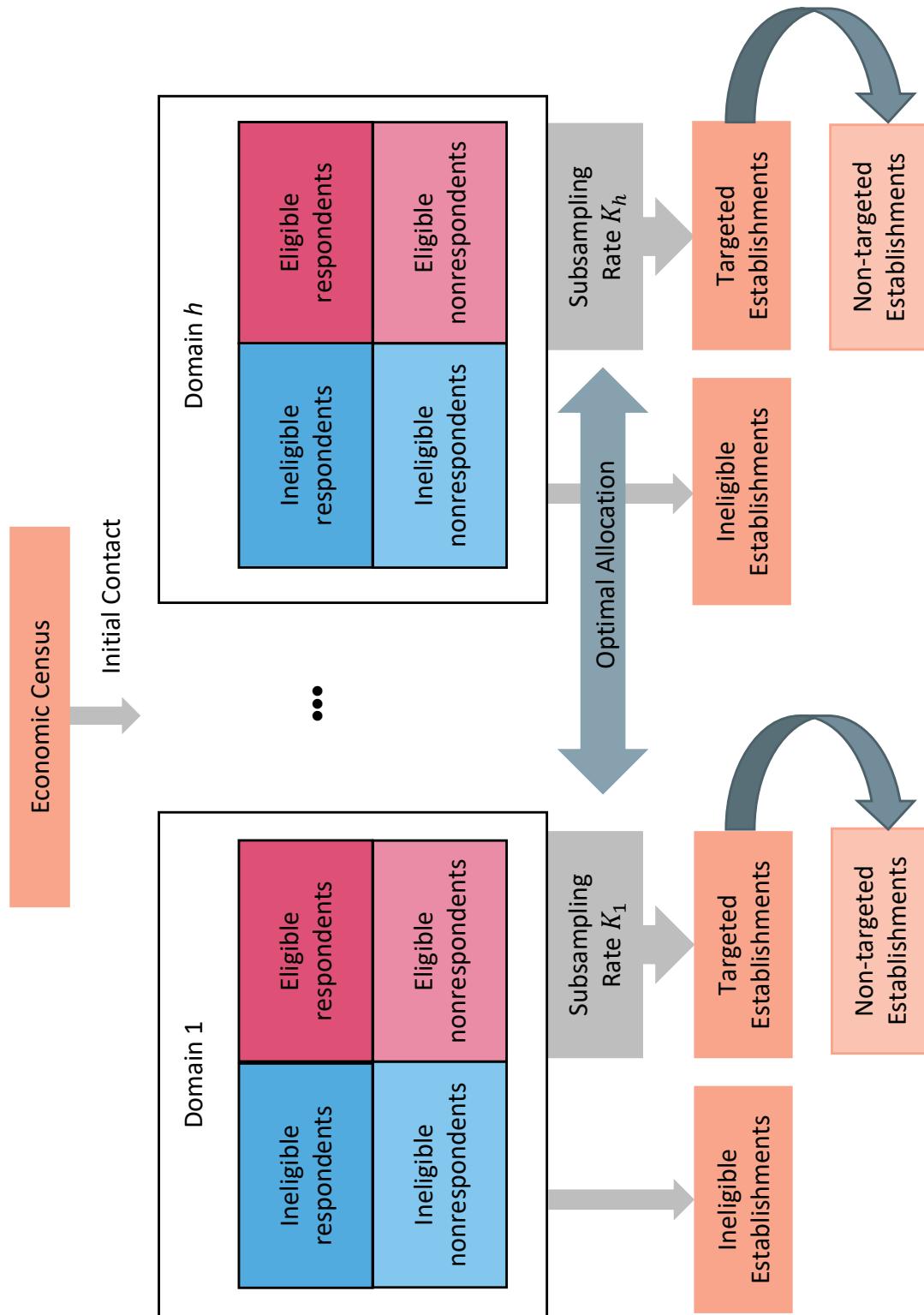
Targeted Follow-up Framework



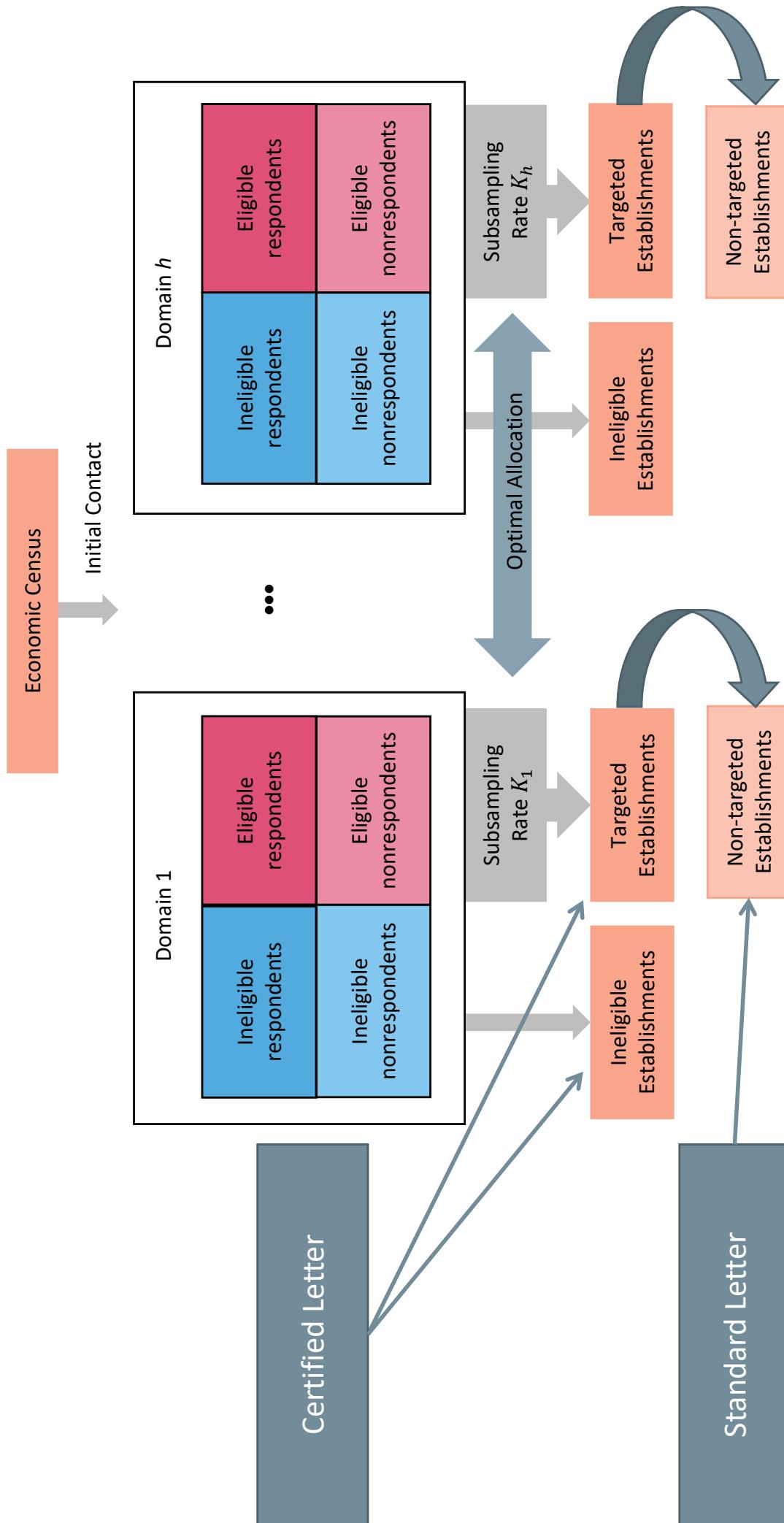
Targeted Follow-up Framework



Targeted Follow-up Framework



Targeted Follow-up Framework



Allocation Method

- Min-URR: Minimizes the squared deviation between domain response rates
- Allocation program formulated as a quadratic program with linear constraints
 - Can only subsample overall rate of K
 - Cannot subsample more establishments than present in domain
 - Cannot take a negative subsample

Min-URR

Minimize (squared) deviation between predicted domain response rate and overall target response rate

Objective function

$$\min \sum_h (URR_h^P - URR^T)^2$$

Predicted domain
response rate

$$URR_h^P = \frac{(r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h)}{n_h}$$

Target response rate

$$URR^T = \frac{\sum_h ((r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h))}{\sum_h n_h}$$

Min-URR

Minimize (squared) deviation between predicted domain response rate and overall target response rate

Objective function
Predicted domain response rate

$$\min \sum_h (URR_h^P - URR^T)^2$$

Target calculated with overall subsampling rate

Target response rate

$$URR_h^P = \frac{(r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h)}{n_h}$$

$$URR^T = \frac{\sum_h ((r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h))}{\sum_h n_h}$$

Min-URR

Minimize (squared) deviation between predicted domain response rate and overall target response rate

Objective function

$$\min \sum_h (URR_h^P - URR^T)^2$$

Predicted domain response rate

$$URR_h^P = \frac{(r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h)}{n_h}$$

Target response rate

$$URR^T = \frac{\sum_h ((r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h))}{\sum_h n_h}$$

$$URR^T = \frac{\sum_h ((r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h))}{\sum_h n_h}$$

Min-URR

Respondents in domain h
before intervention

Minimize (squared) deviation between predicted domain
response rate and overall target response rate

Objective function

$$\min \sum_h (URR_h^P - URR^T)^2$$

Predicted domain
response rate

$$URR_h^P = \frac{(r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h)}{n_h}$$

Target response rate

$$URR^T = \frac{\sum_h ((r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h))}{\sum_h n_h}$$

Min-URR

Minimize (squared) deviation between predicted domain response rate and overall target response rate

Objective function

$$\min \sum_h (URR_h^P - URR^T)^2$$

Predicted domain
response rate

$$URR_h^P = \frac{(r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h)}{n_h}$$

Target response rate

$$URR^T = \frac{\sum_h ((r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h))}{\sum_h n_h}$$

Ineligible nonrespondents in
domain h before intervention

Eligible nonrespondents in
domain h before intervention

Min-URR

Minimize (squared) deviation between predicted domain response rate and overall target response rate

Objective function

$$\min \sum_h (URR_h^P - URR^T)^2$$

Predicted domain response rate

$$URR_h^P = \frac{(r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h)}{n_h}$$

Target response rate

$$URR^T = \frac{\sum_h ((r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h))}{\sum_h n_h}$$

Estimated nonresponse conversion rate for Ineligible units in domain h

Estimated nonresponse conversion rate for Eligible units in domain h

Min-URR

Minimize (squared) deviation between predicted domain response rate and overall target response rate

Objective function

$$\min \sum_h (URR_h^P - URR^T)^2$$

Predicted domain response rate

$$URR_h^P = \frac{(r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h)}{n_h}$$

Target response rate

$$URR^T = \frac{\sum_h ((r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h))}{\sum_h n_h}$$

Predicted Ineligible Respondents
in domain h

Predicted Eligible Respondents
in domain h

Production Run

- Define Response Status
- Define Eligibility for Targeted Allocation
- Estimate Conversion Rates
- Establish Allocation Domain

Response Status

- Official response status cannot be assigned till after data collection and analyst review
- Check-in status was used as a proxy for response status

Eligibility

- Ineligible
 - All multi-unit establishments

Eligibility

- Ineligible
- All multi-unit establishments
- High priority single units

Eligibility

- Ineligible
 - All multi-unit establishments
 - High priority single units
 - Small domains (< 5 establishments)

Eligibility

- Ineligible
 - All multi-unit establishments
 - High priority single units
 - Small domains (< 5 establishments)
 - Establishments with incomplete industry codes

Eligibility

- Ineligible
 - All multi-unit establishments
 - High priority single units
 - Small domains (< 5 establishments)
 - Establishments with incomplete industry codes
 - Two high priority industry sectors

Eligibility

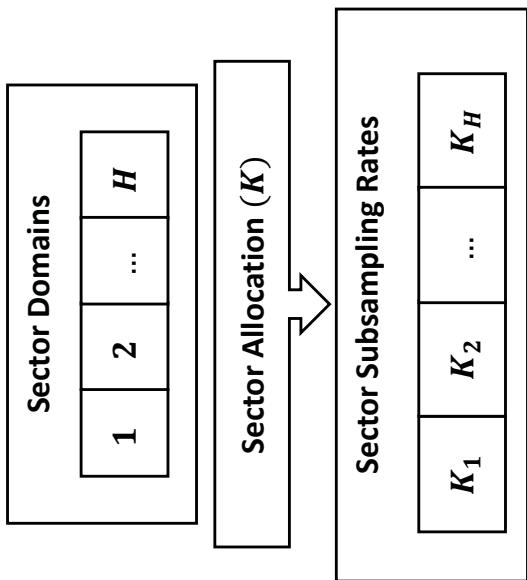
- Ineligible
 - All multi-unit establishments
 - High priority single units
 - Small domains (< 5 establishments)
 - Establishments with incomplete industry codes
 - Two high priority industry sectors
- Eligible
 - Remaining single units

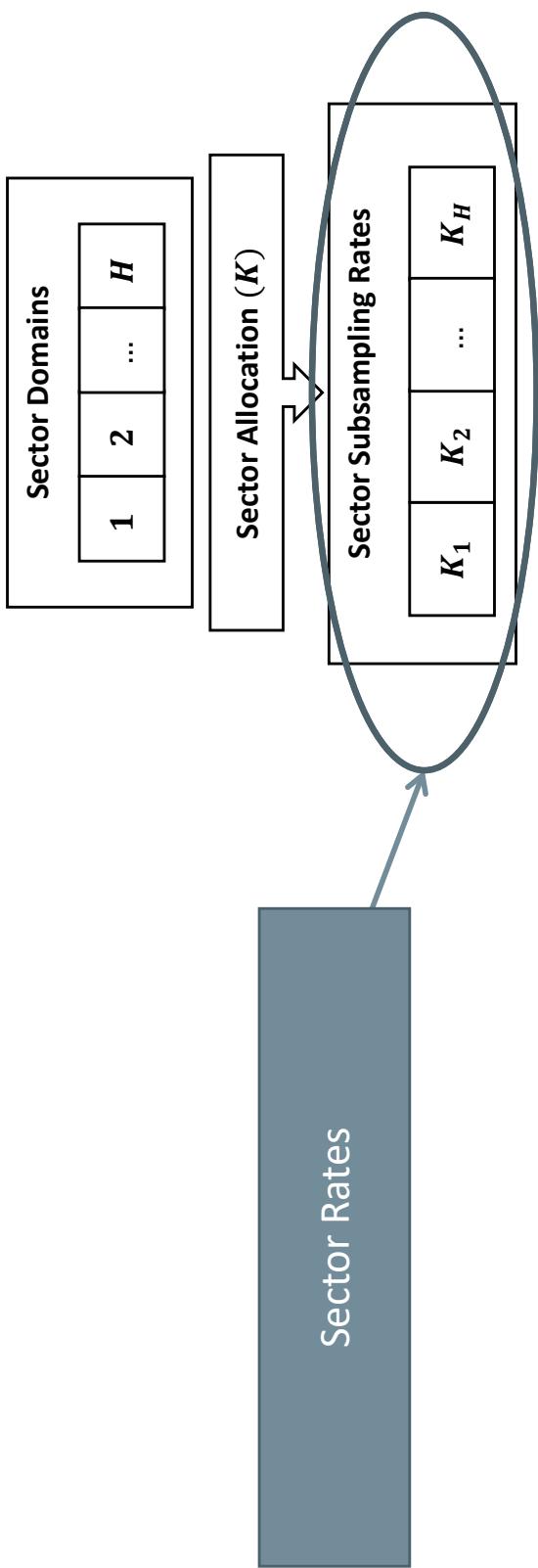
Conversion Rates (q)

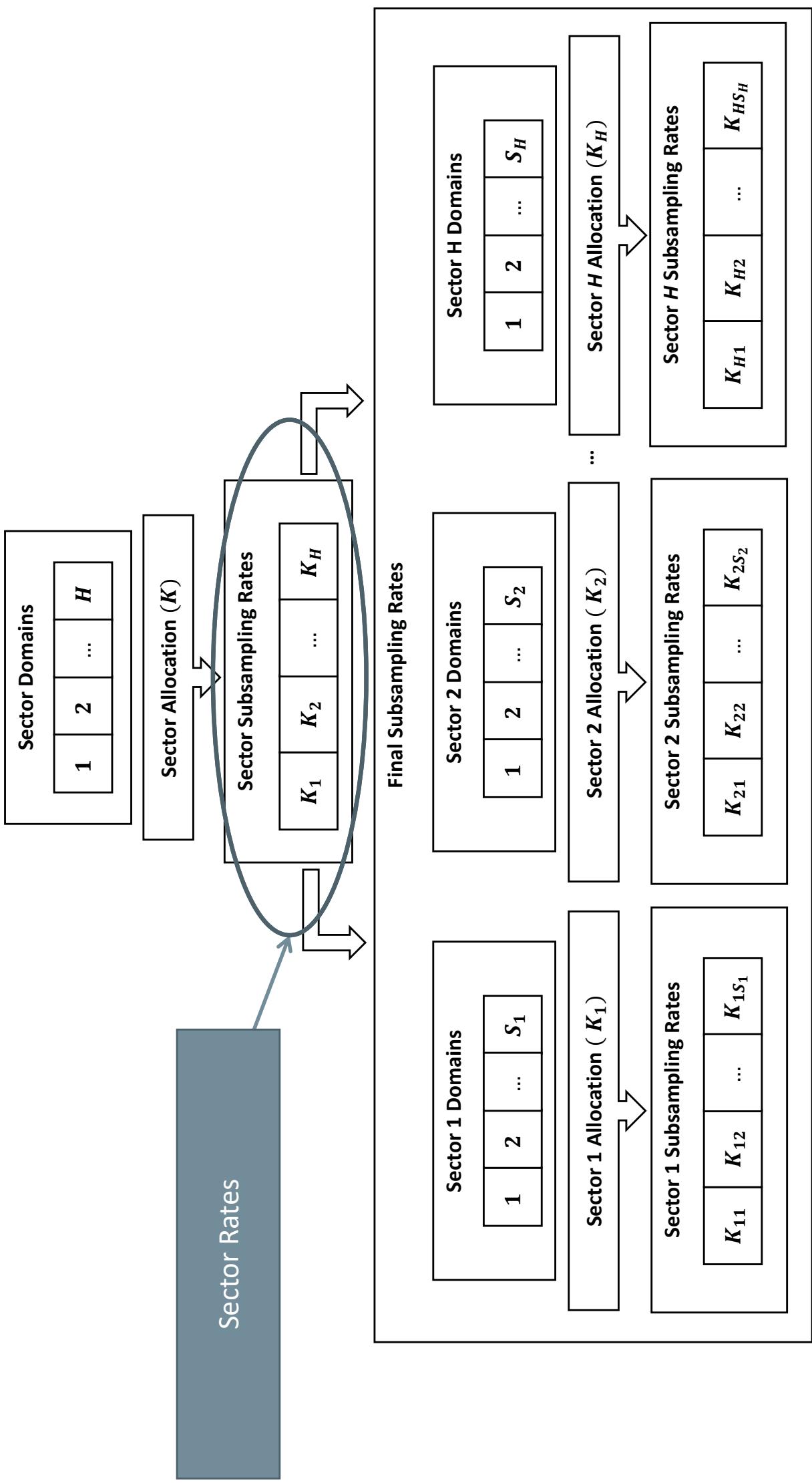
- Can be difficult to estimate due to new collection protocols
- Ineligible = 0.65
- Eligible = 0.50

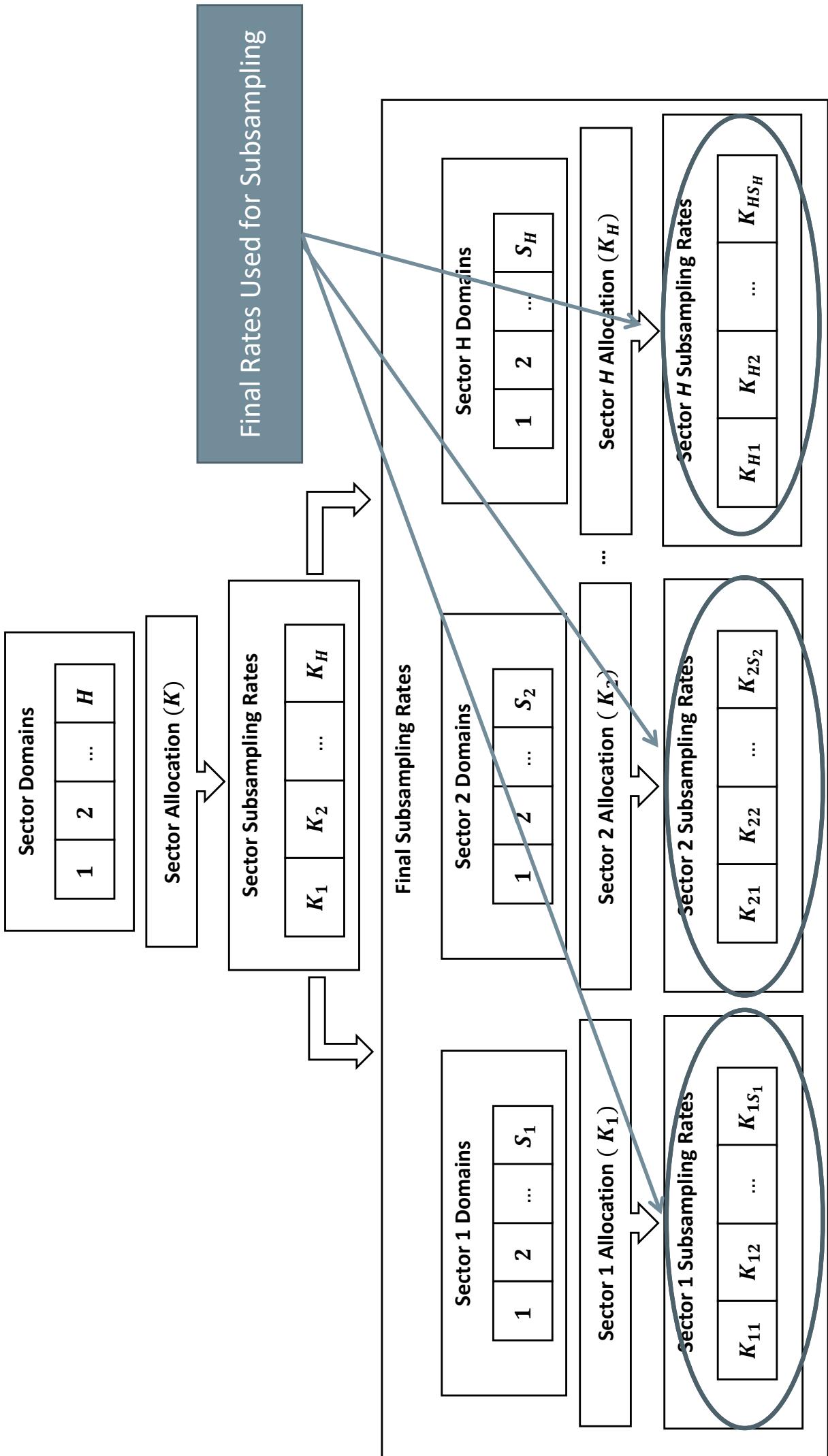
Allocation Domains

- Initially domains were set at 6-Digit NAICS crossed with state
 - Created over 15,000 domains
 - Long run time
 - Uninterpretable results
- Two stage Allocation
 - First Stage: Sector
 - Second Stage: 6-Digit NAICS crossed with state





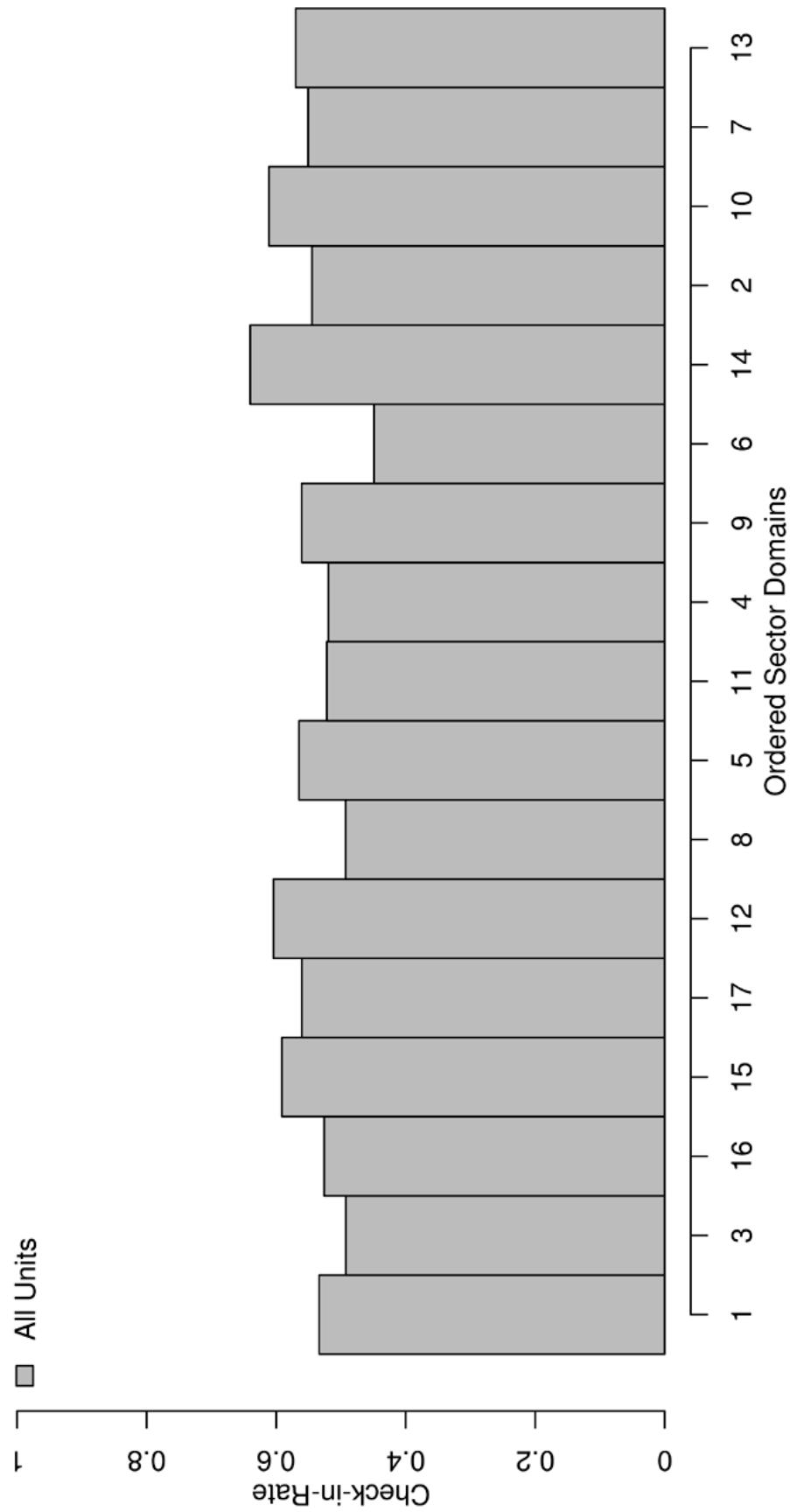




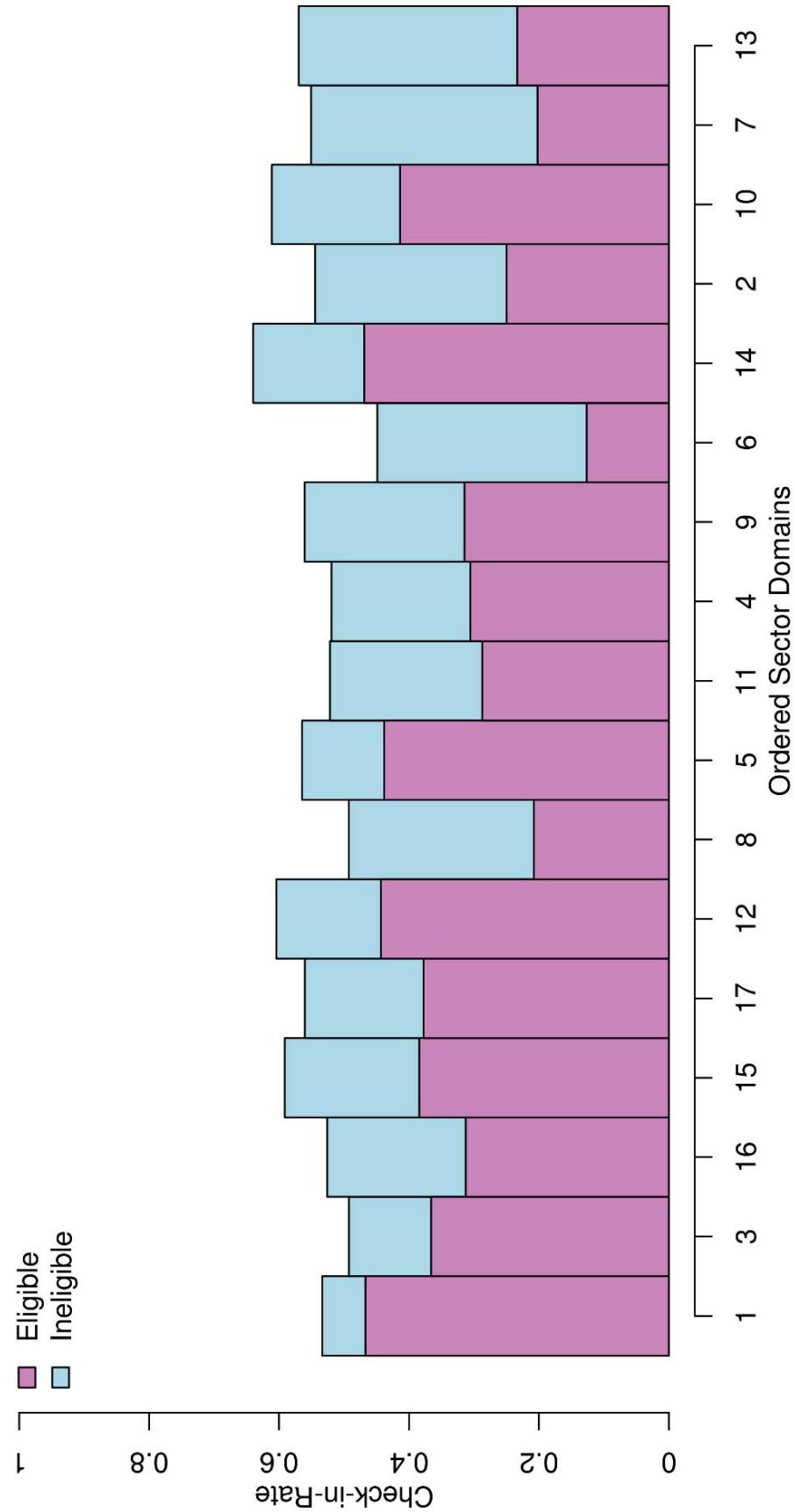
Results From Production Run

- Intervention check-in rates
- Predicted check-in rates
- Observed check-in rates
- Observed conversion rates

Intervention Check-in Rates



Intervention Check-in Rates



Allocation Check-in Rate

Predicted Domain
Check-in Rate

$$URR_h^P = \frac{(r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h)}{n_h}$$

“Fixed” Domain
Allocation Rate

$$URR_h^A = \frac{r_h + (m_h^i * q_h^i)}{n_h}$$

Allocation Check-in Rate

$$URR_h^P = \frac{(r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h)}{n_h}$$
$$URR_h^A = \frac{r_h + (m_h^i * q_h^i)}{n_h}$$

Predicted Domain Check-in Rate

“Fixed” Domain Allocation Rate

Allocation Check-in Rate

Predicted Domain
Check-in Rate

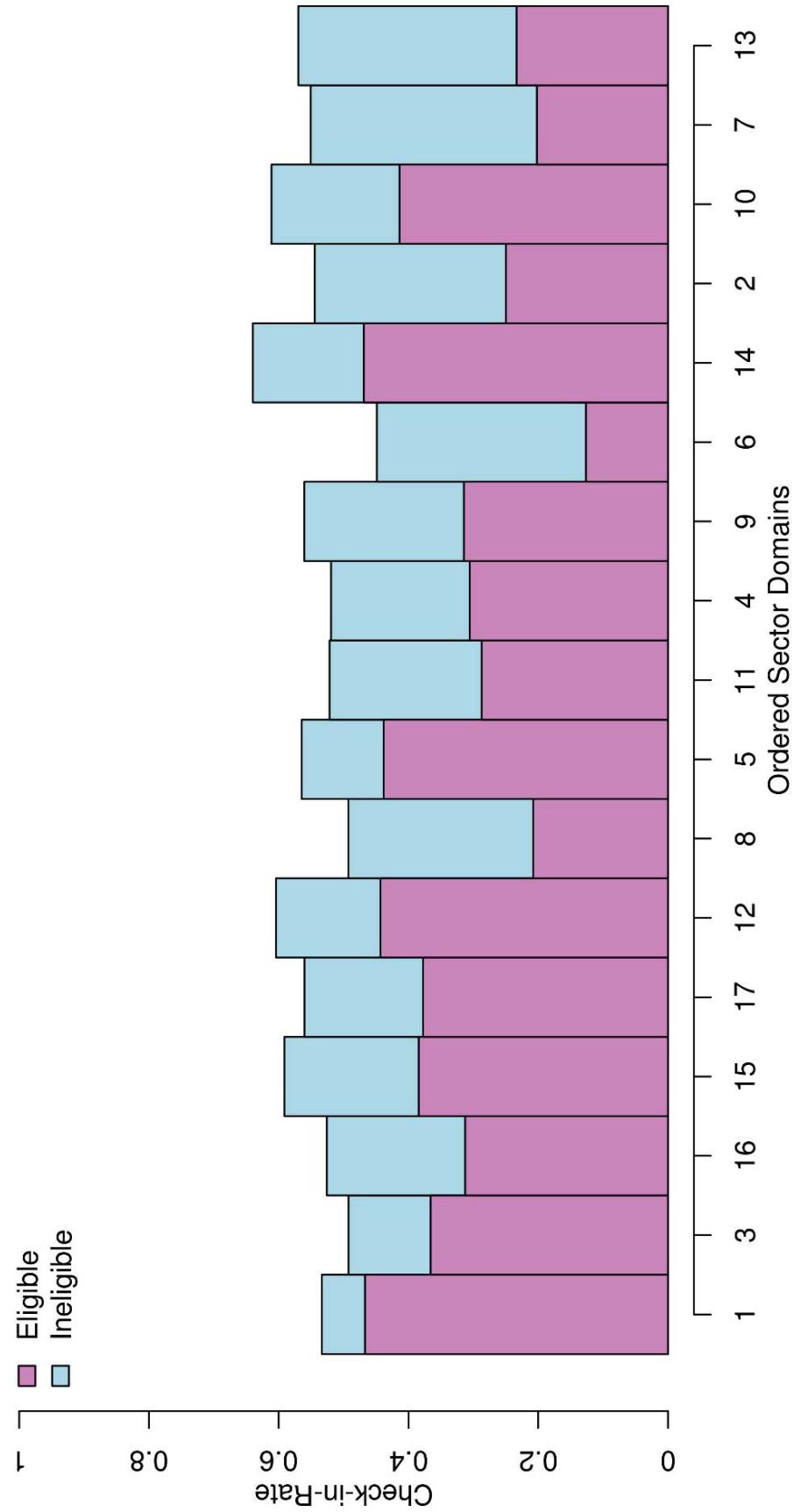
$$URR_h^P = \frac{(r_h + (m_h^i * q_h^i)) + (m_h^e * q_h^e * K_h)}{n_h}$$

“Fixed” Domain
Allocation Rate

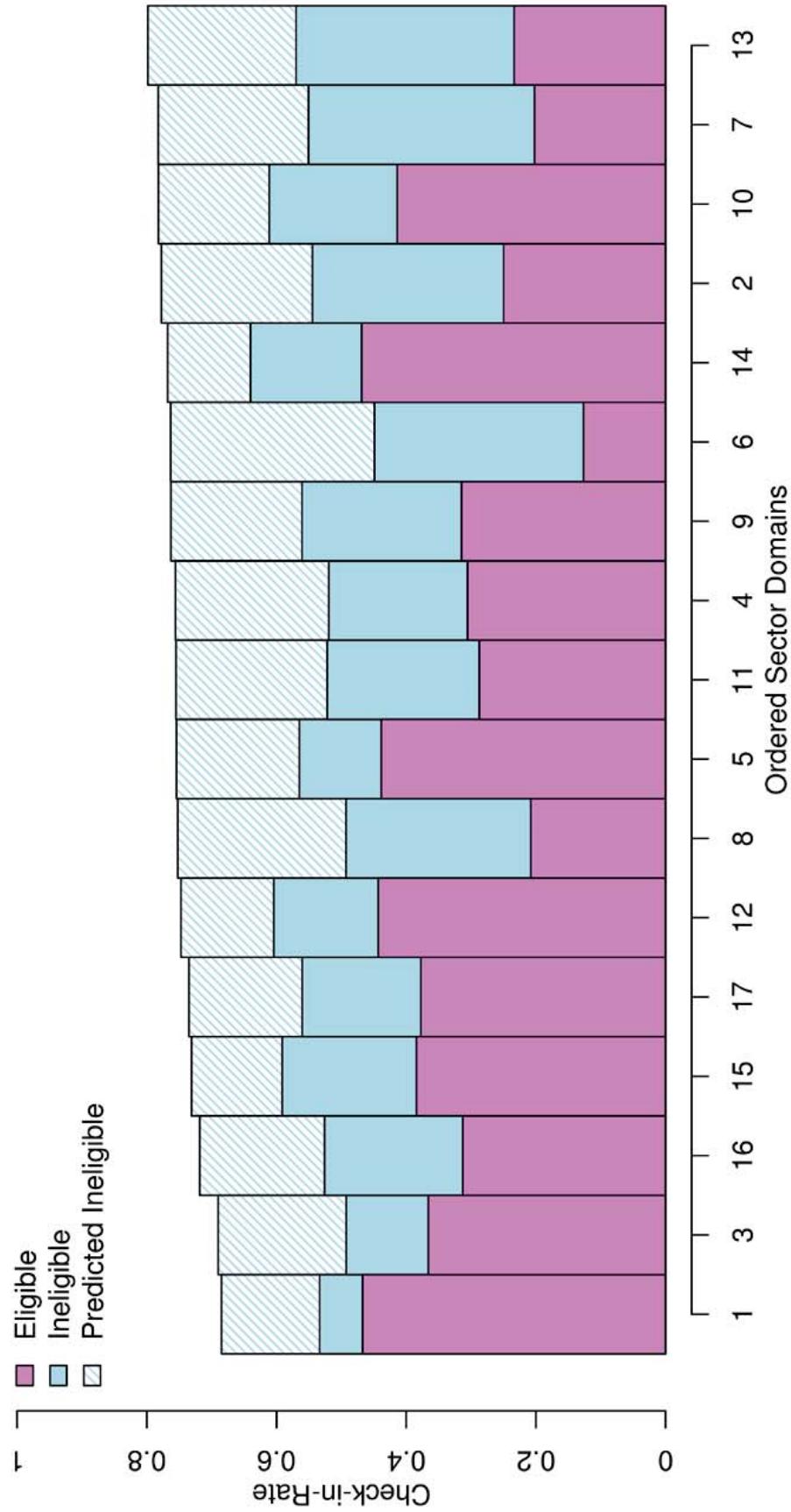
$$URR_h^A = \frac{r_h + (m_h^i * q_h^i)}{n_h}$$

Starting point for allocations

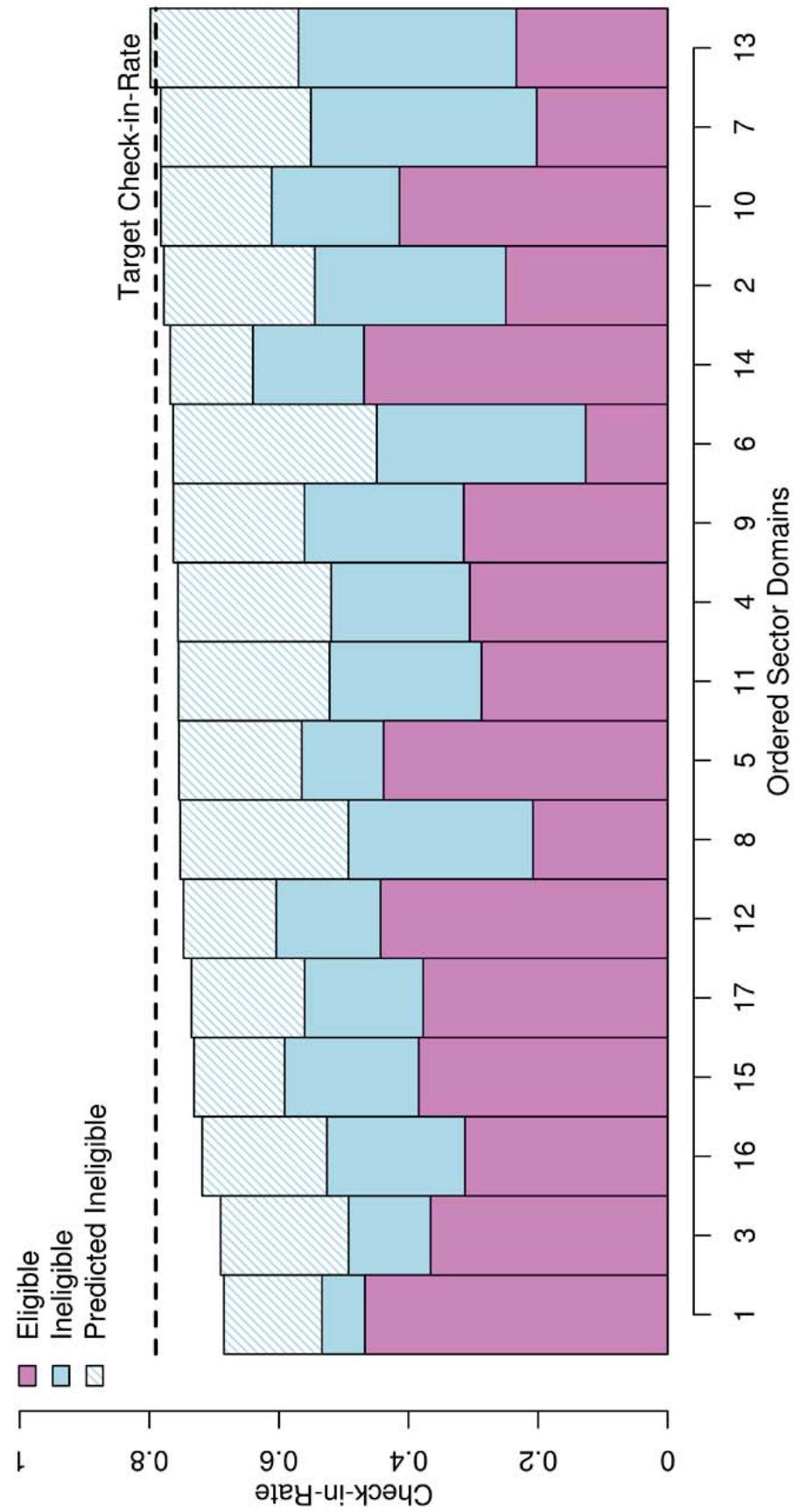
Intervention Check-in Rates



Intervention Allocation Rate



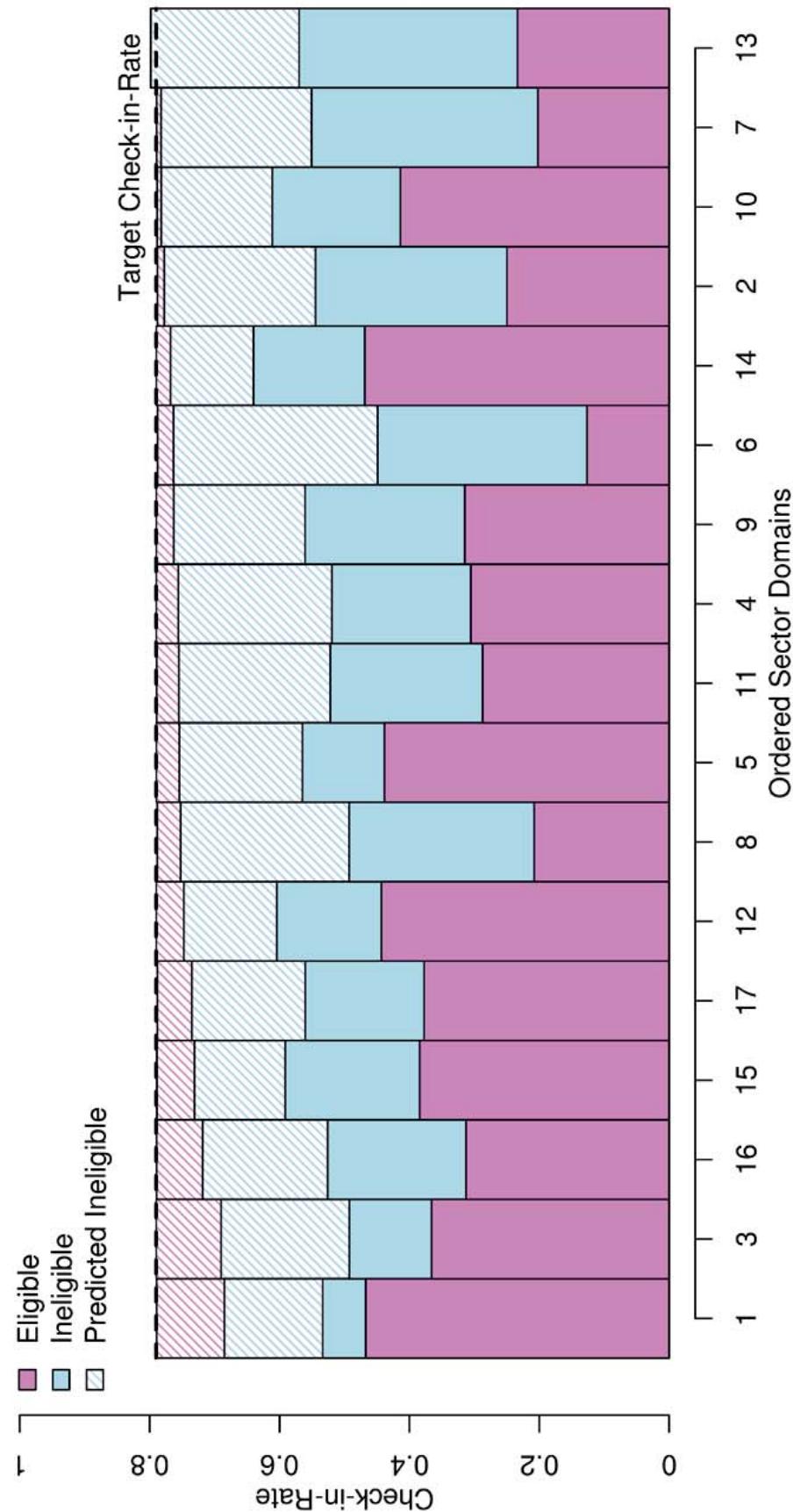
Target Check-in Rate



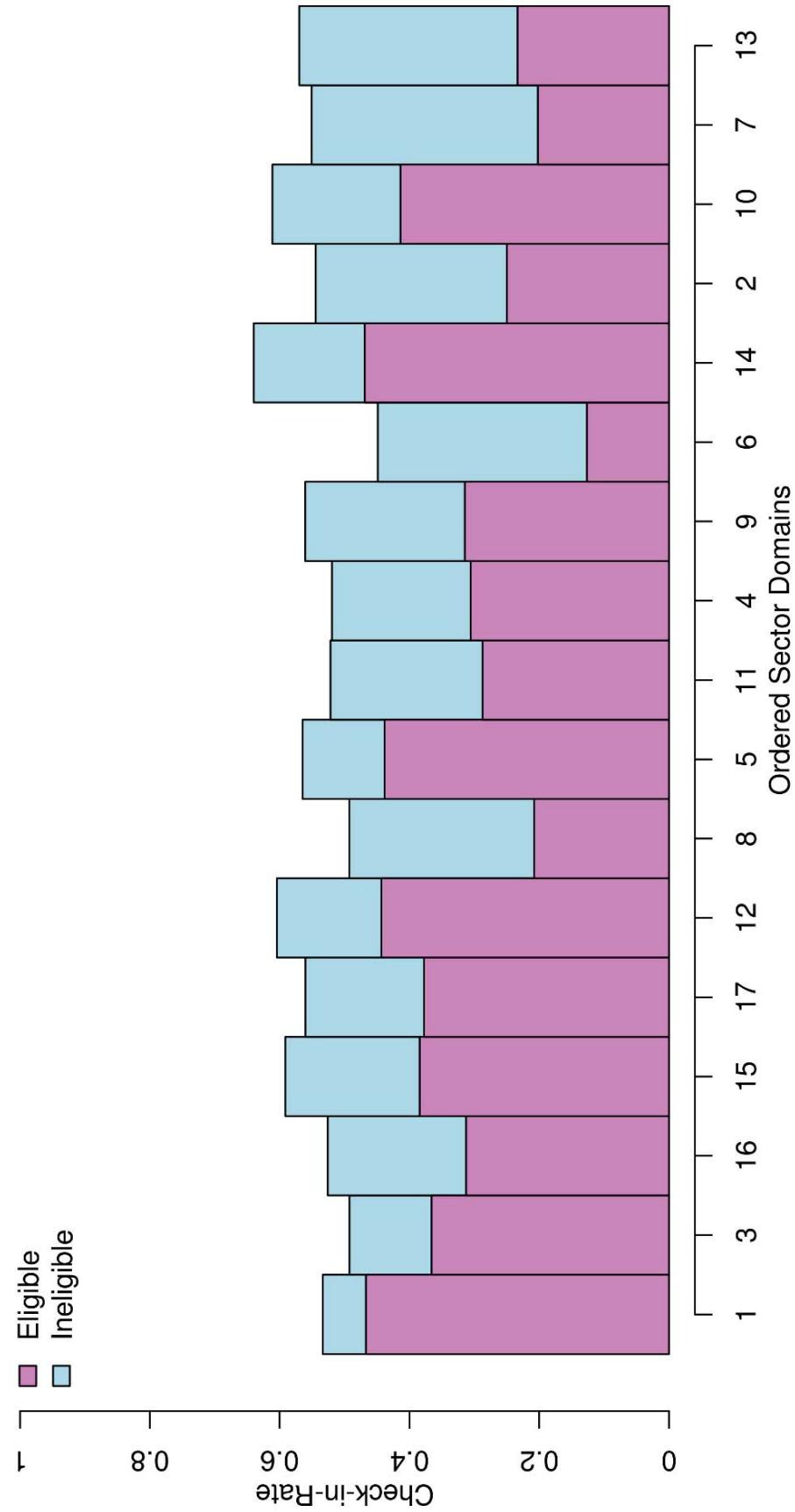
Target Check-in Rate



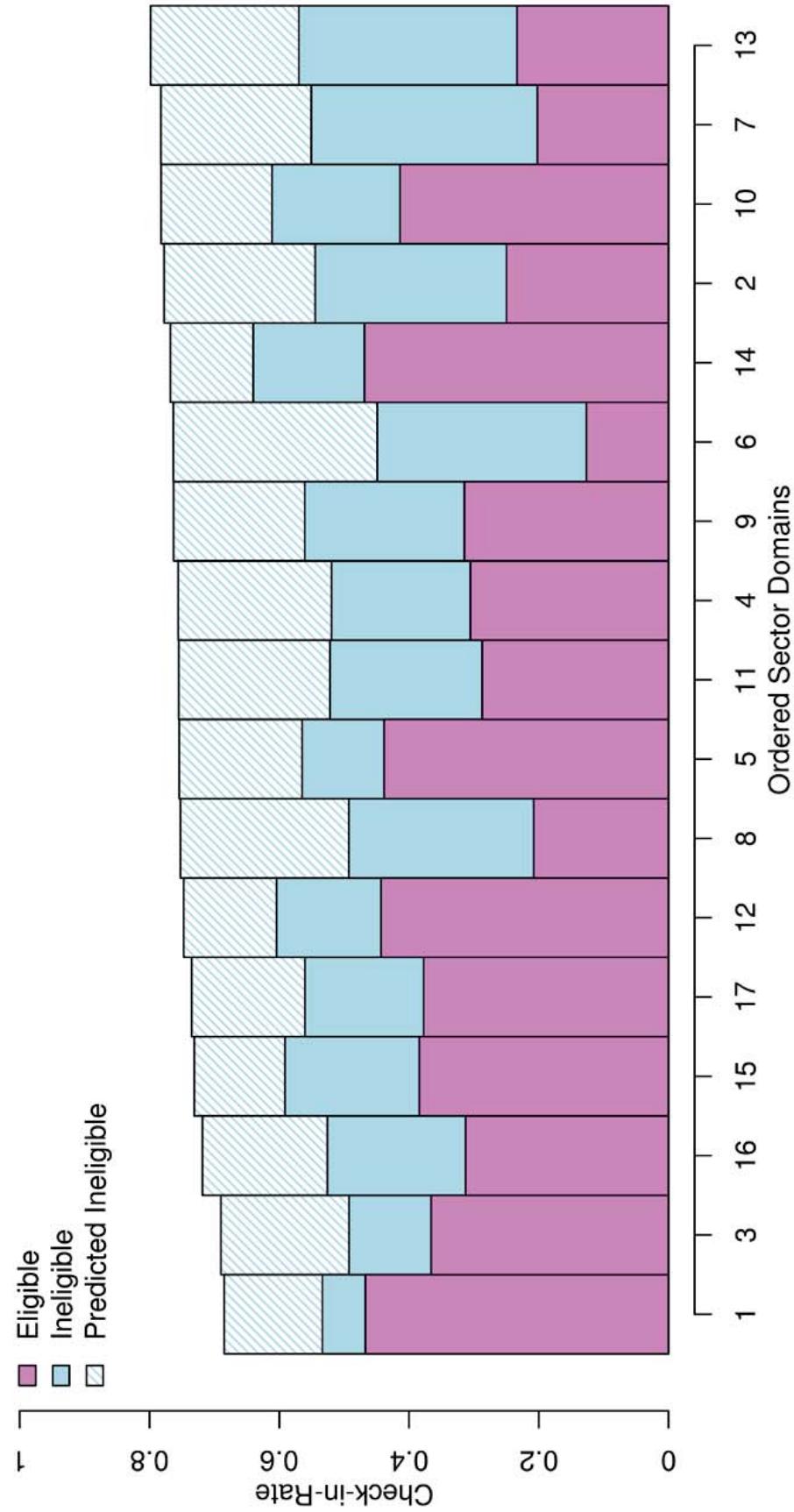
Predicted Check-in Rate after Intervention



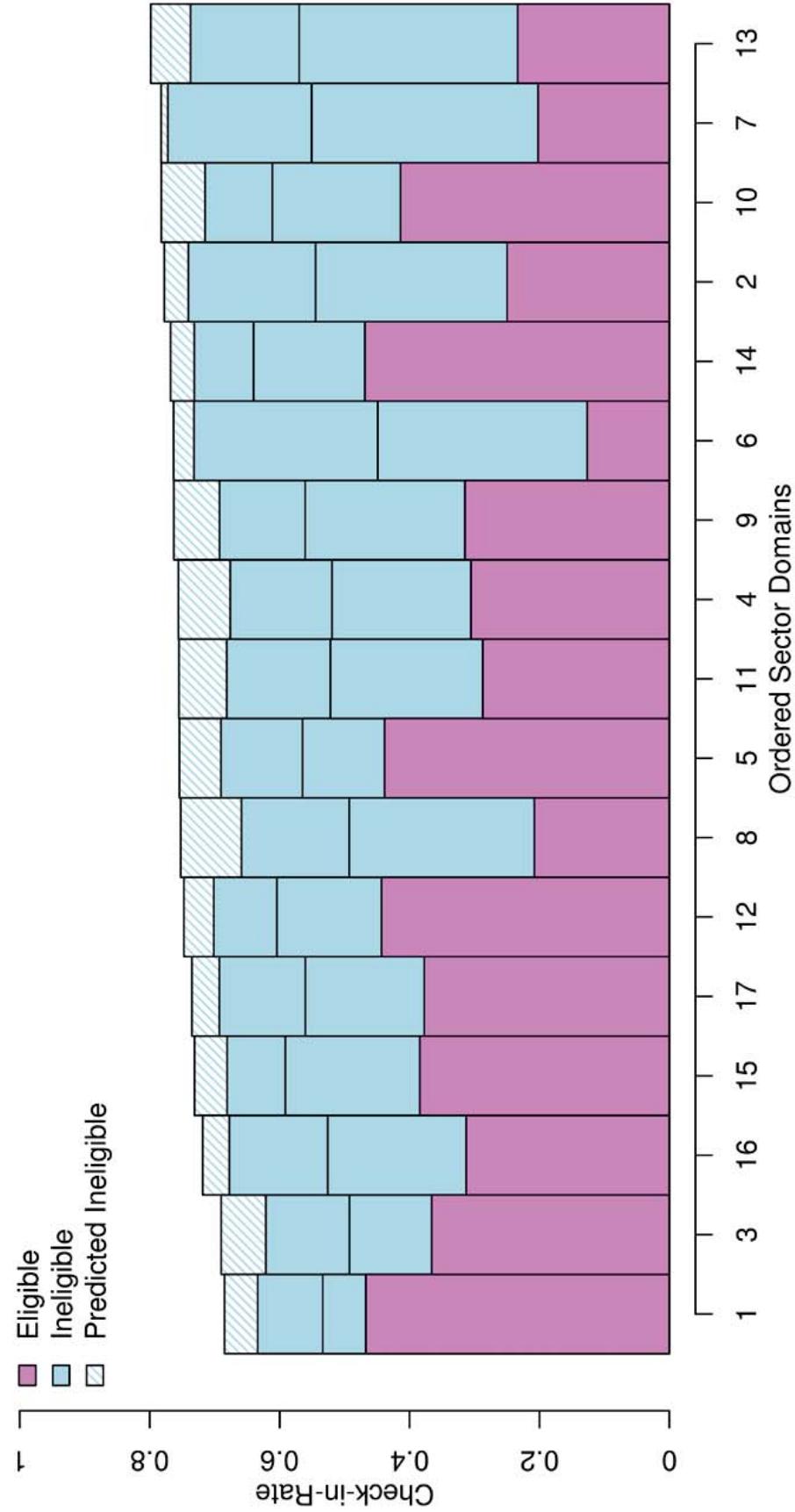
Results



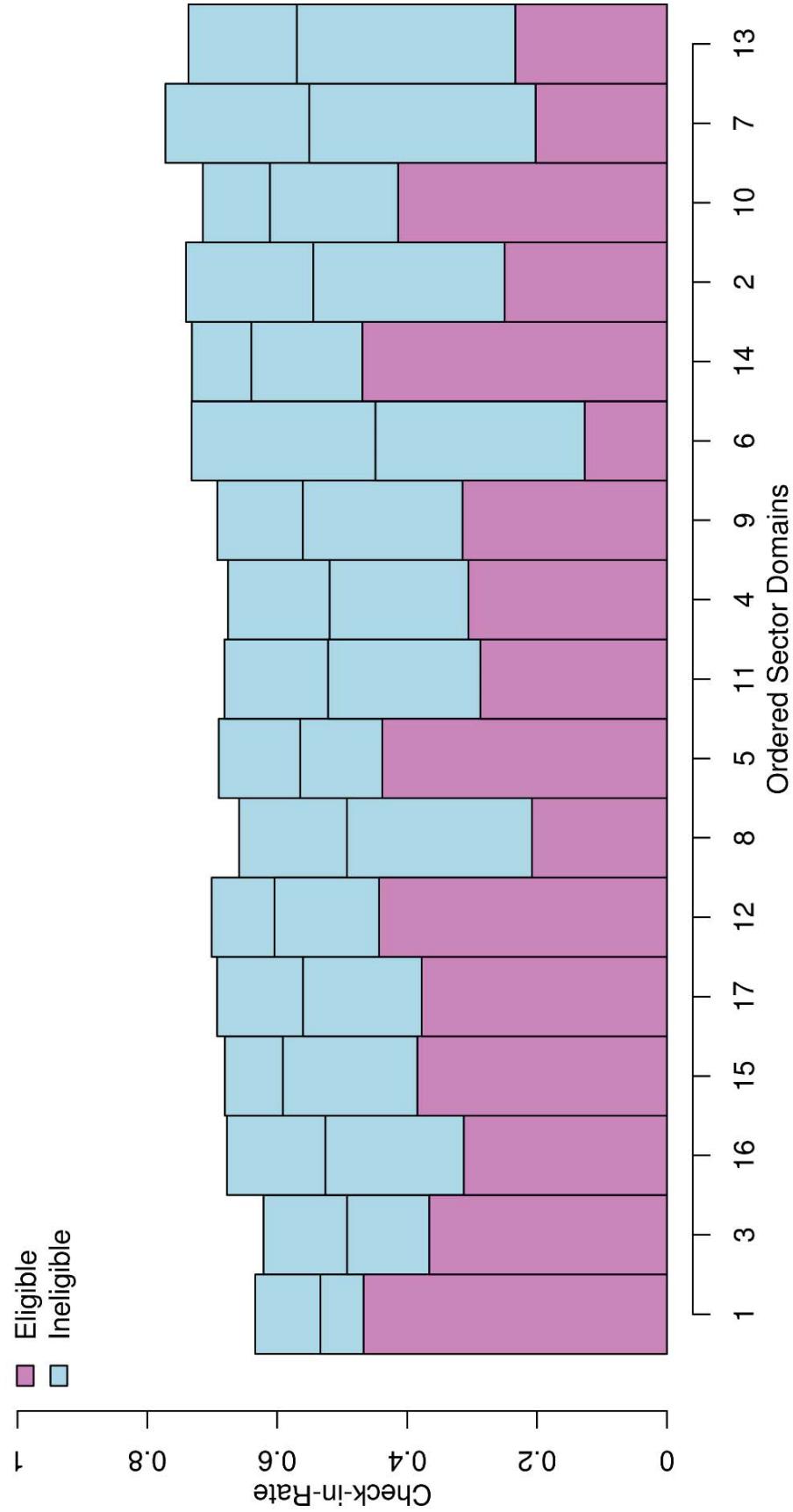
Intervention Allocation Rate



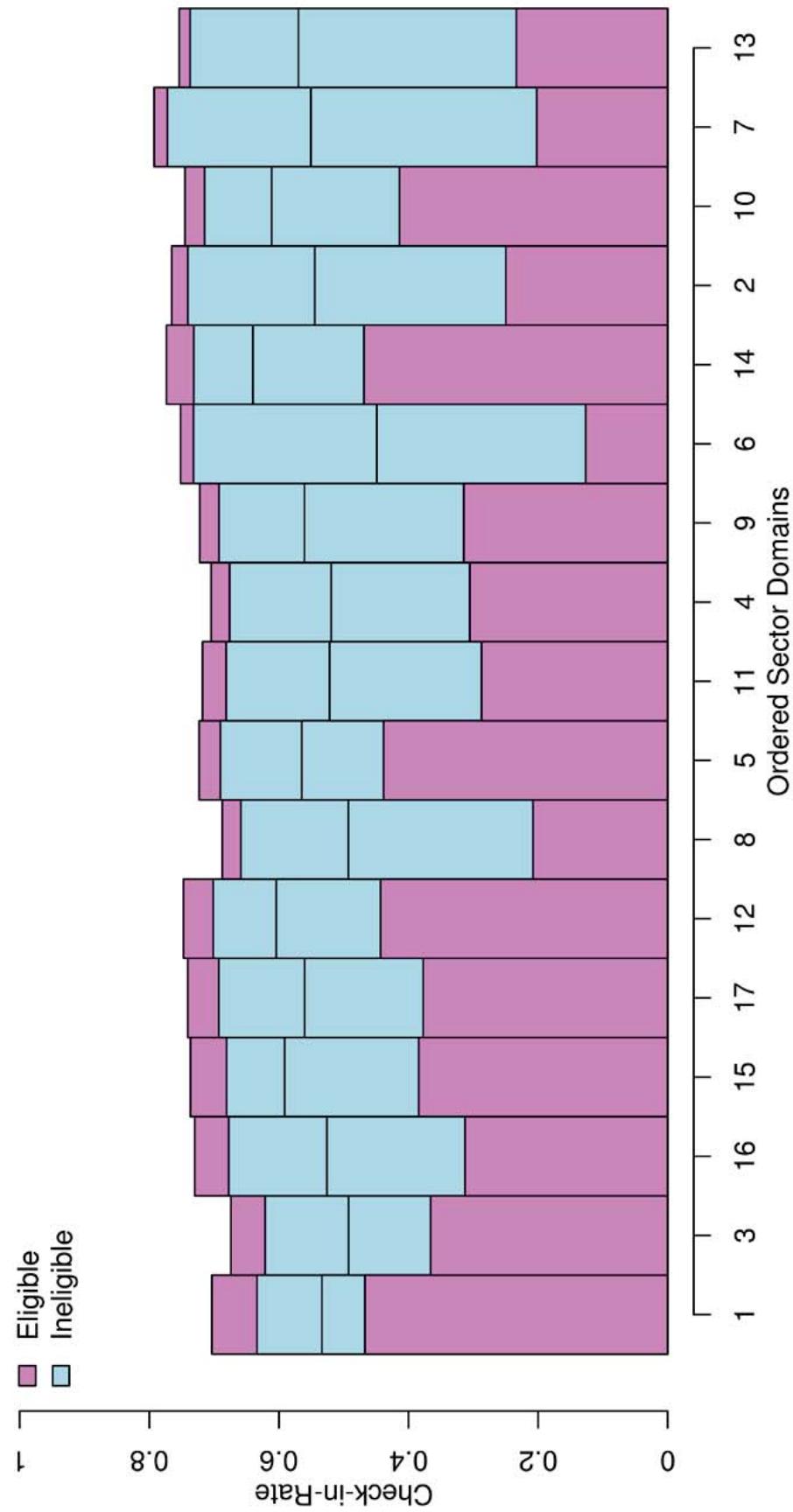
Observed Allocation Rate



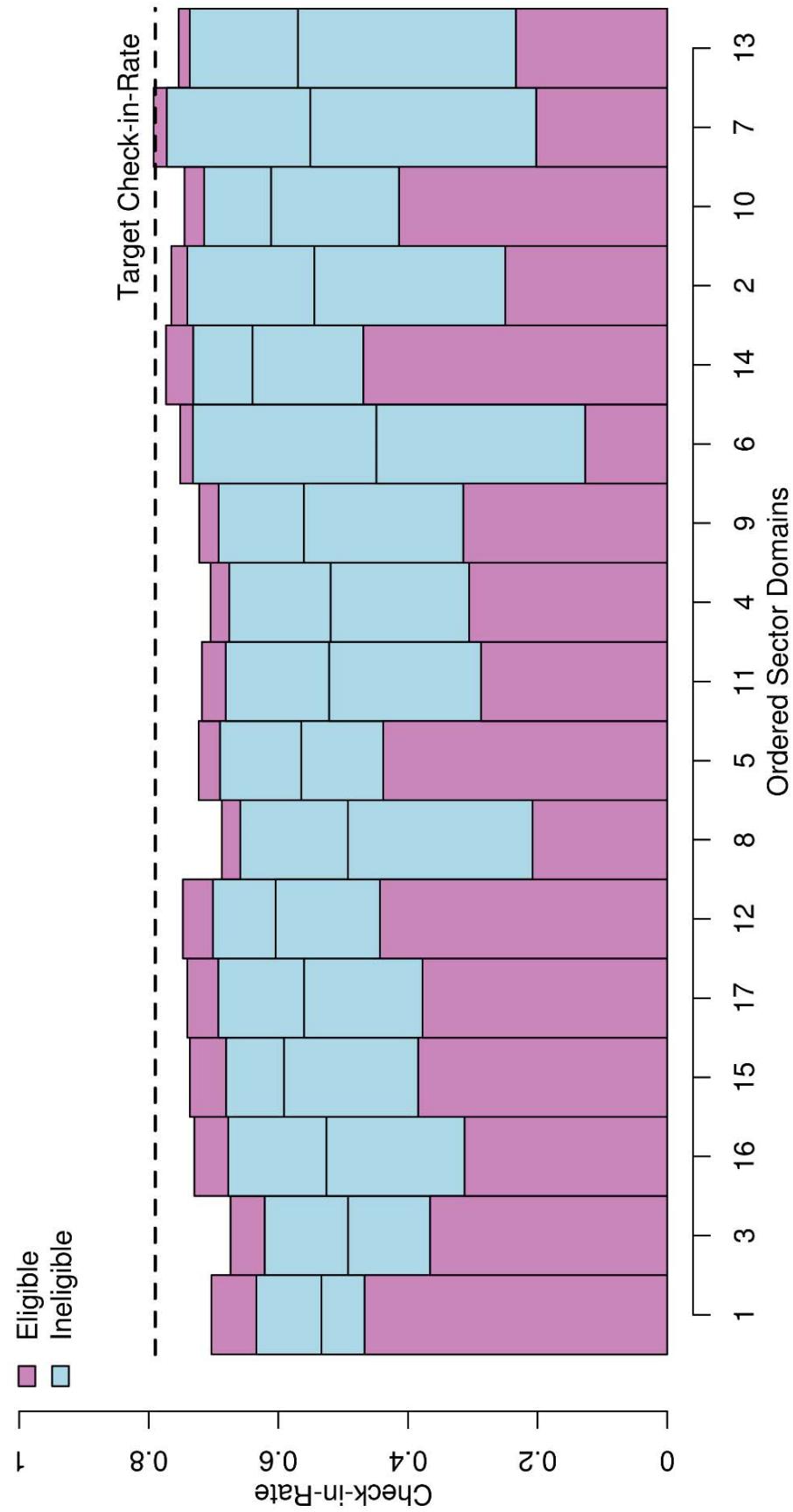
Observed Allocation Rate



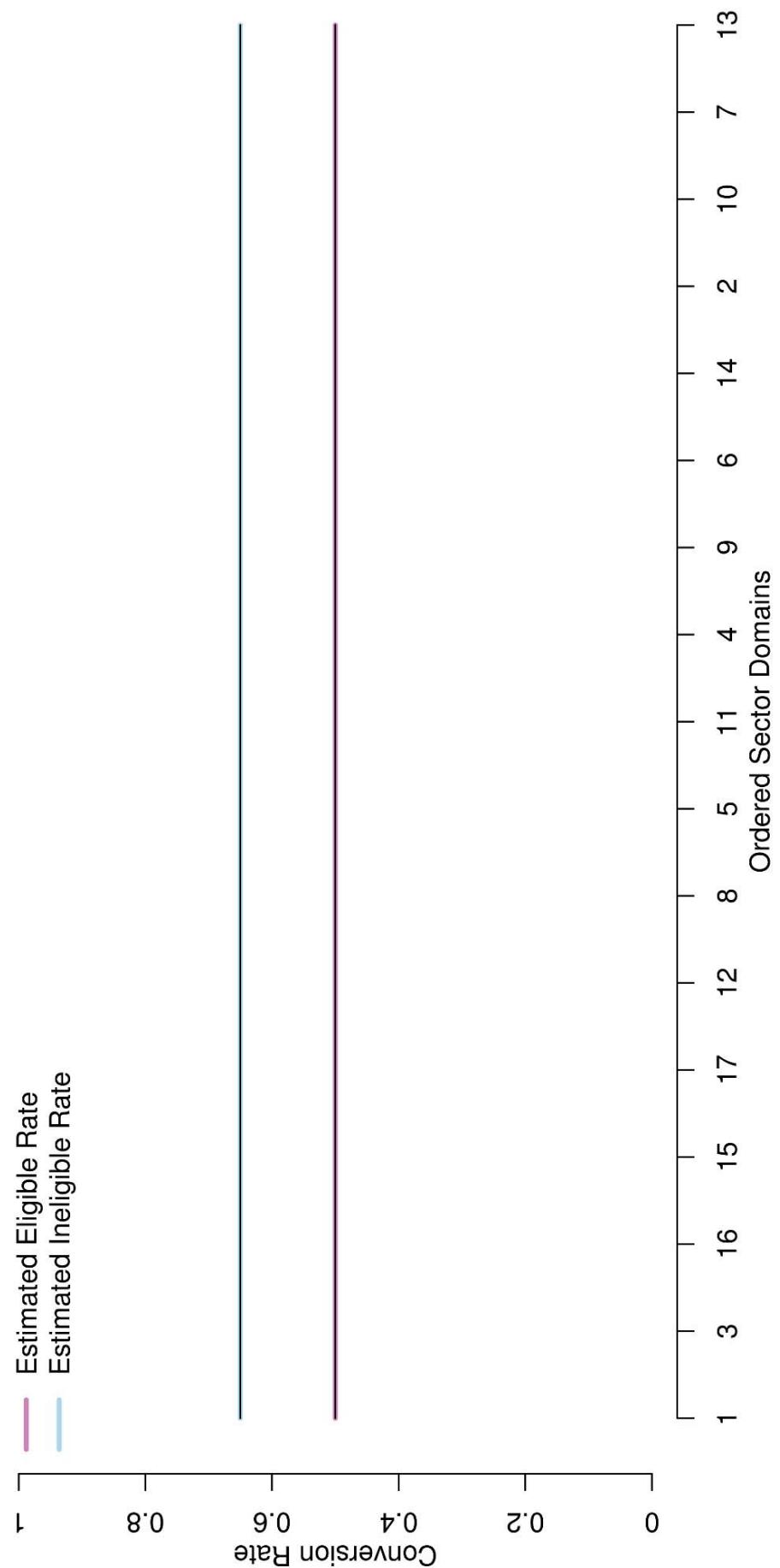
Observed Check-in Rate



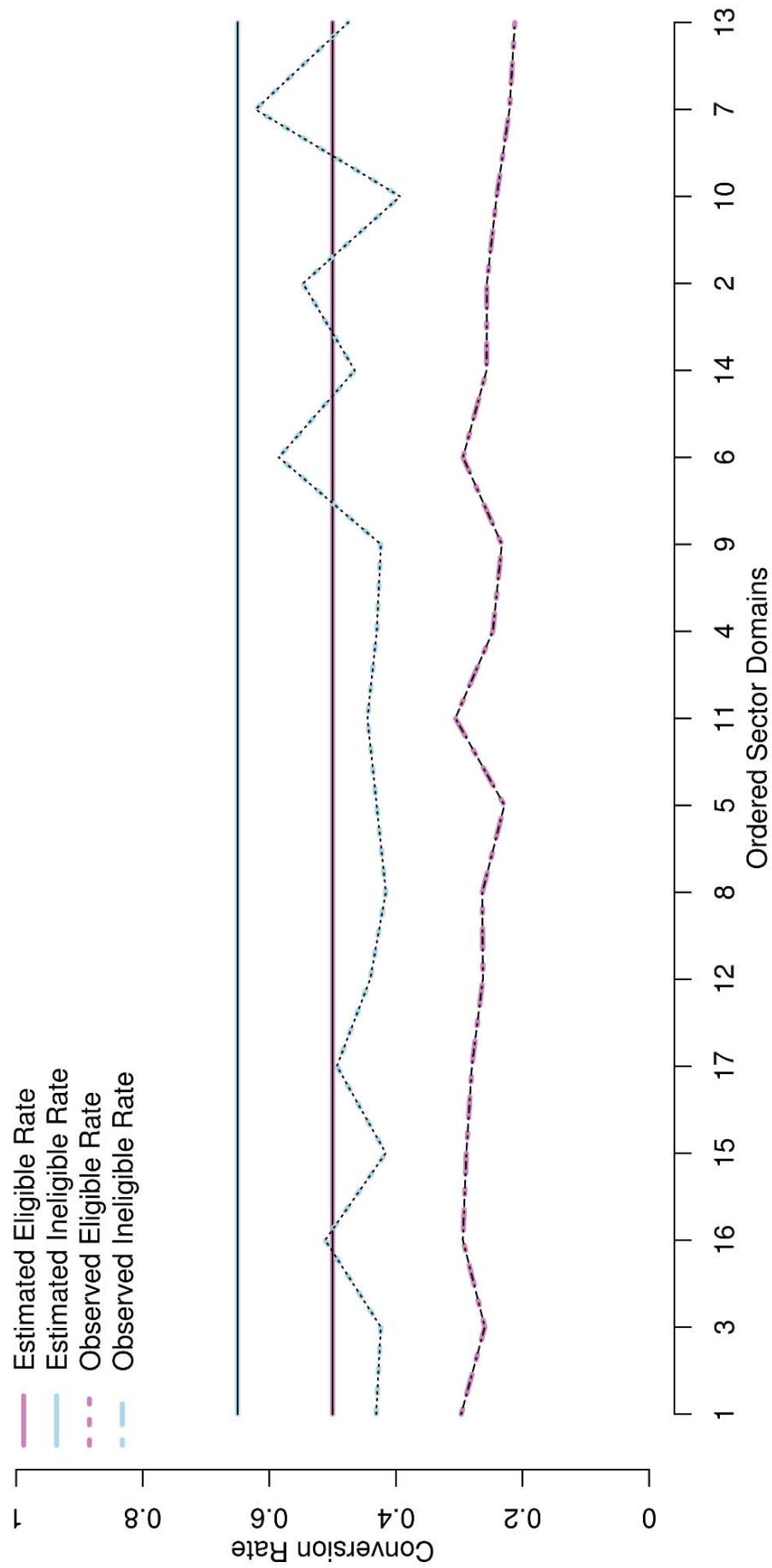
Observed Check-in Rate



Estimated Conversion Rates



Observed Conversion Rates



Conclusion

- Hard to judge quality without a baseline or control group
- Did successfully implement adaptive design methods in a large scale data collection effort
- Took coordination between program managers, methodologists, and programmers across different trade areas

E-mail:

stephen.kaputa@census.gov



Questions