Decision Rules for Administrative Data Usage: In-Office Enumeration

Stephanie Coffey, Jennifer Bernard

U.S. Census Bureau, Center for Economic Studies

Joint Statistical Meetings

August 3, 2026

Disclaimer

Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau.

No data is disclosed in this presentation.

Motivation

- Nonresponse Follow-Up (NRFU) is a complex, costly operation
  - Recruit, hire and manage 100,000+ of enumerators
  - Knock on doors for non-self-responding addresses (~33% of addresses in 2020)
- In-Office Enumeration (InOE) to reduce costs and improve data quality
  - Use high quality administrative data (AD) *where possible*
  - Allow enumerators to focus on addresses with lower quality/missing AD
- How do we determine "*where possible*"?
- Builds on work from 2020 Census
  - In 2020, AD was used to limit fieldwork (focus on persons and adult/child status)
  - This work extends to more granular age groups, race/ethnicity characteristics

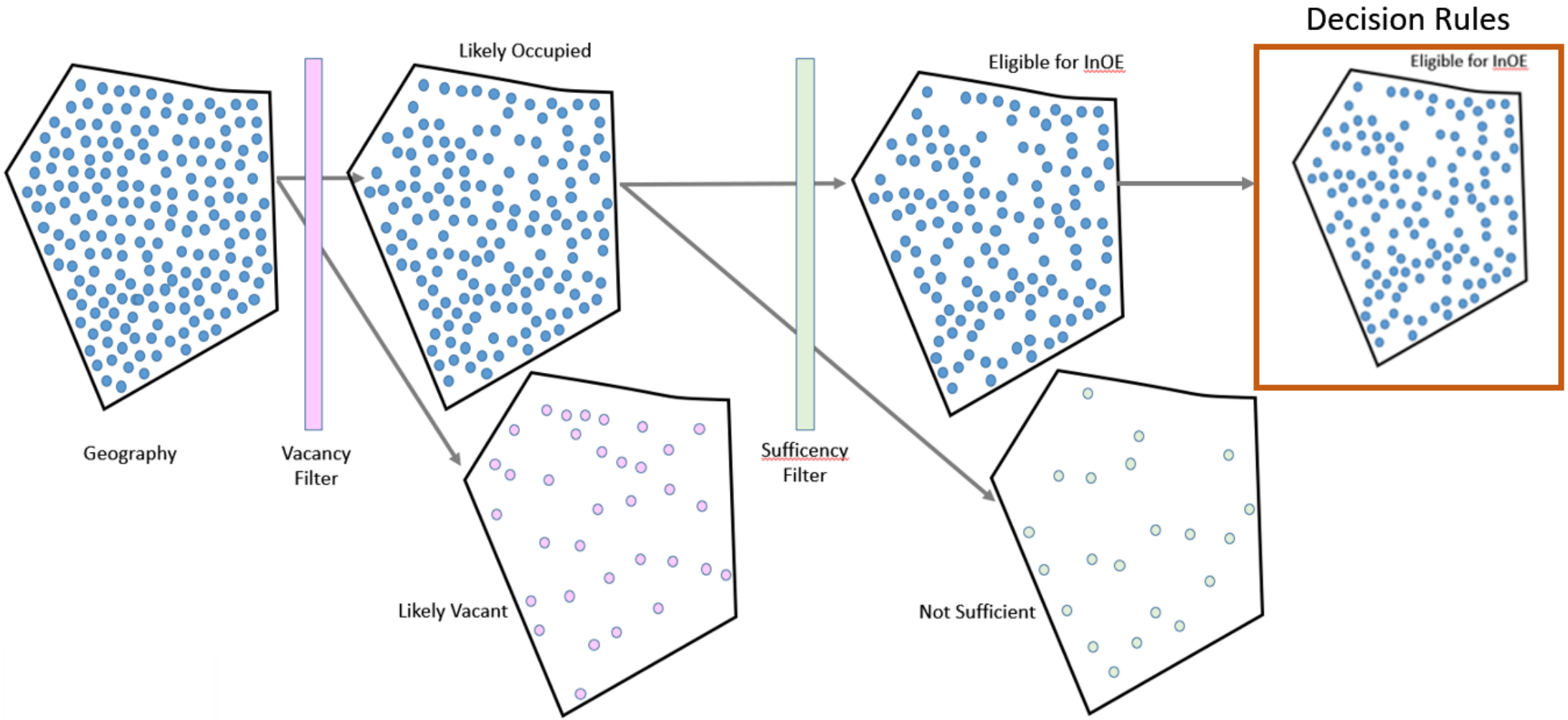
Roadmap

- Conceptual Overview of In-Office Enumeration
- Supporting Models
- Sample Decision Rules
- Wrap-Up

In-Office Enumeration: Conceptual Overview

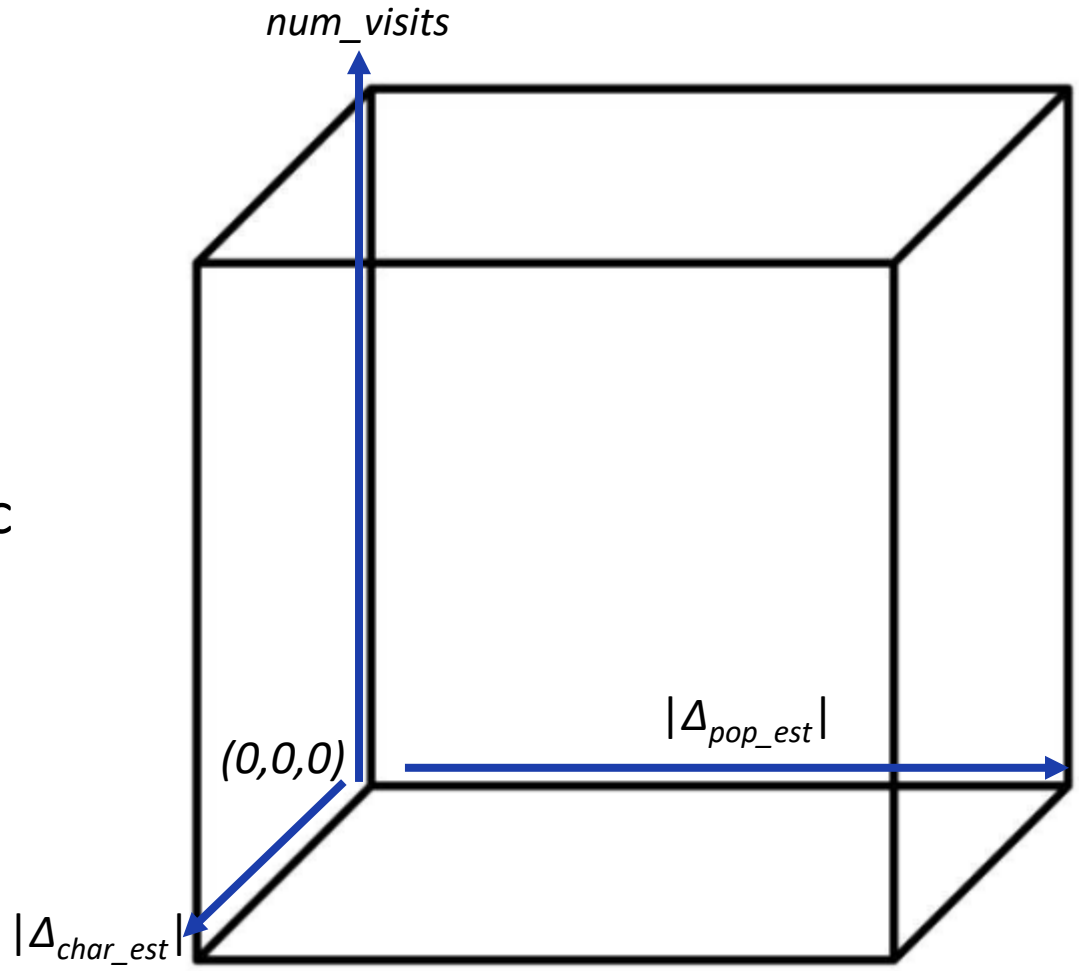
- Which addresses can be enumerated with AD rather than fieldwork?
- Multi-step decision process to determine eligibility and selectability
  - Eligible Cases:
    - Open and unresolved at time *t*
    - Not identified as likely vacant/delete *(vacancy modeling process)*
    - Have sufficient AR data *(sufficiency criteria)*
  - Selectable Cases:
    - Have an expected AR popcount “close” to theoretical field enumeration
    - Have expected demographic characteristics “close” to theoretical field enumeration
    - Require relatively “more” NRFU effort than other cases

In-Office Enumeration: Eligibility



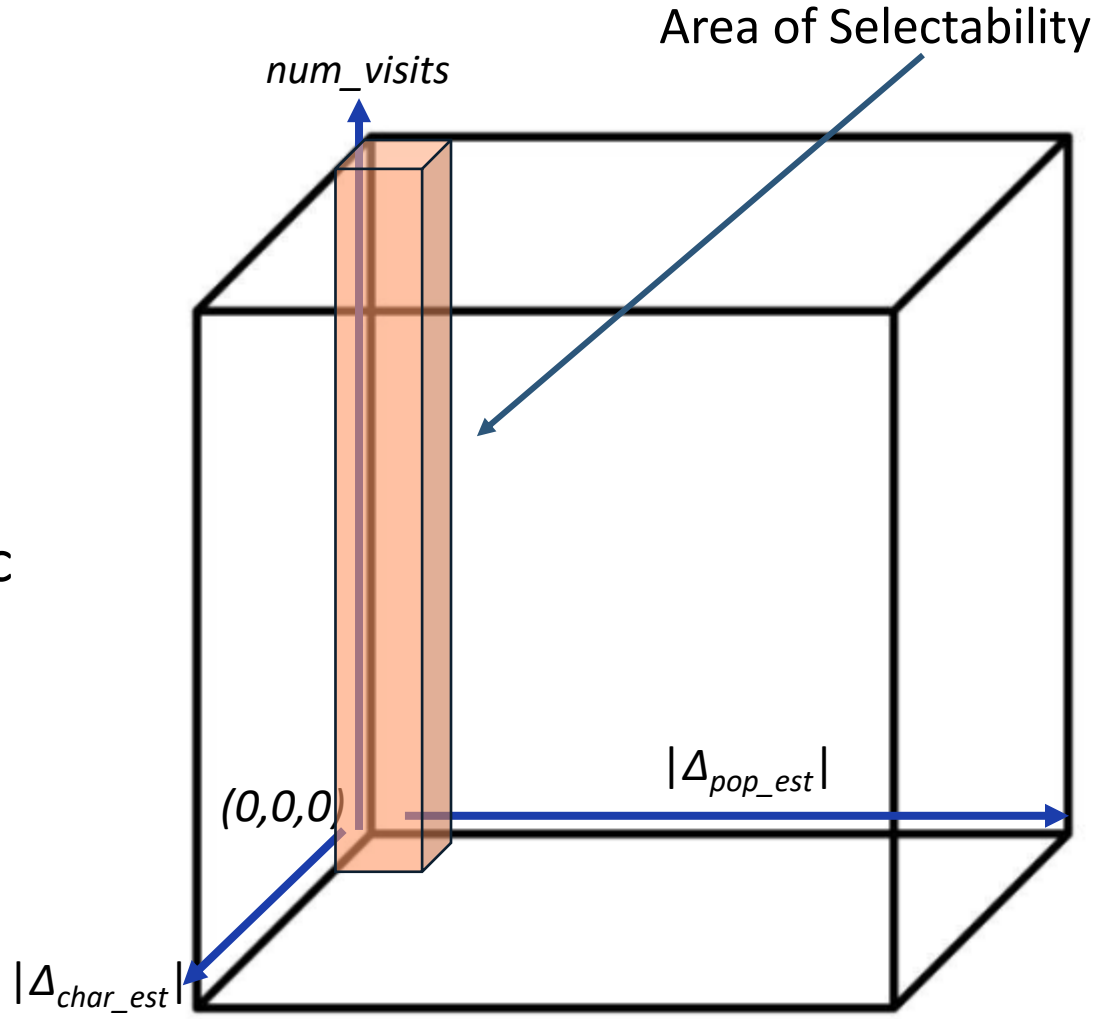
In-Office Enumeration: Selectability

- Three Criteria
  - $|\Delta_{pop\_est}|$: difference between number of people at an address in AD vs field enumeration
  - $|\Delta_{char\_est}|$: difference between expected count of people in different characteristic categories (sex, age, race, ethnicity) of people at an address in AD vs field enumeration
  - *num_visits*: expected number of NRFU attempts needed to enumerate the address



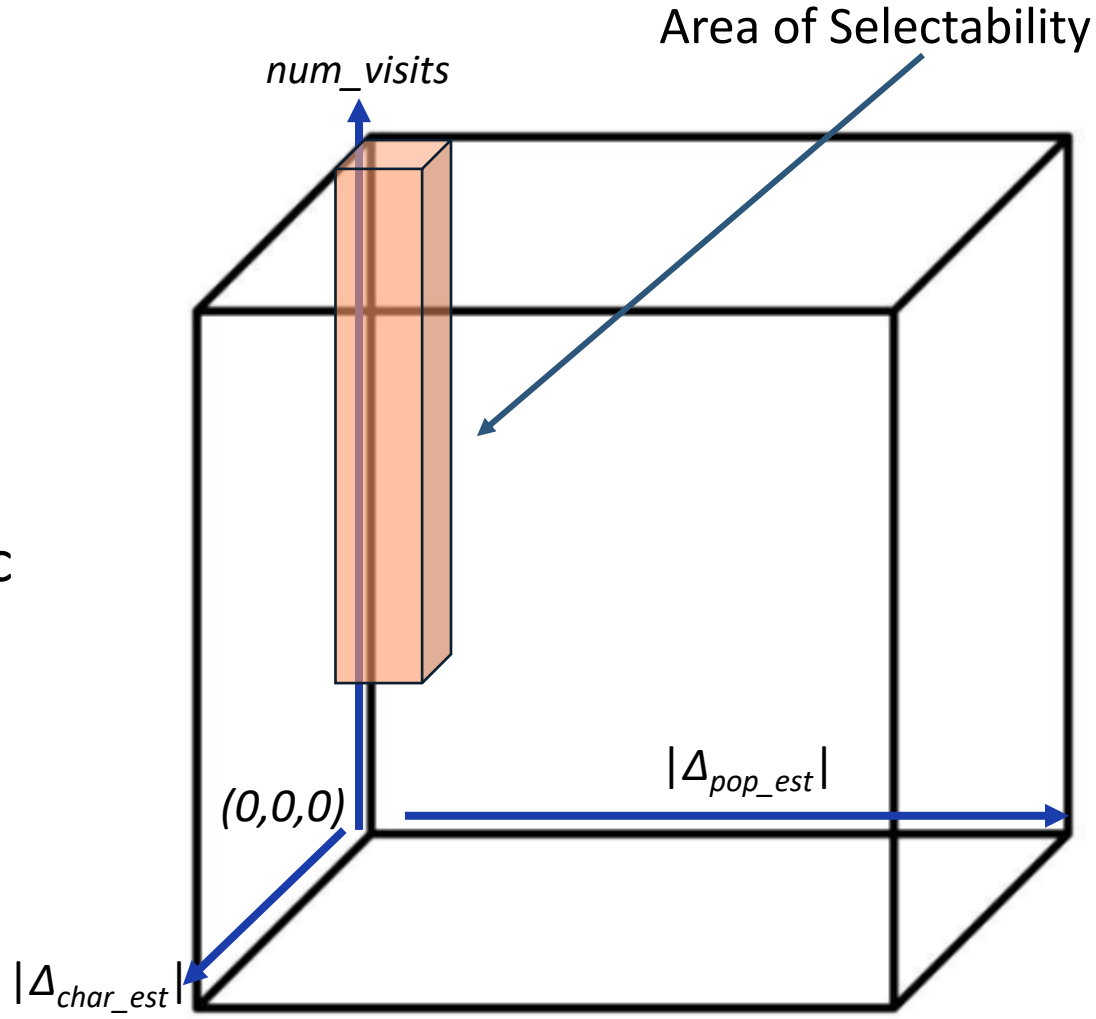
In-Office Enumeration: Selectability

- Three Criteria
  - $|\Delta_{pop\_est}|$: difference between number of people at an address in AD vs field enumeration
  - $|\Delta_{char\_est}|$: difference between expected count of people in different characteristic categories (sex, age, race, ethnicity) of people at an address in AD vs field enumeration
  - *num_visits*: expected number of NRFU attempts needed to enumerate the address



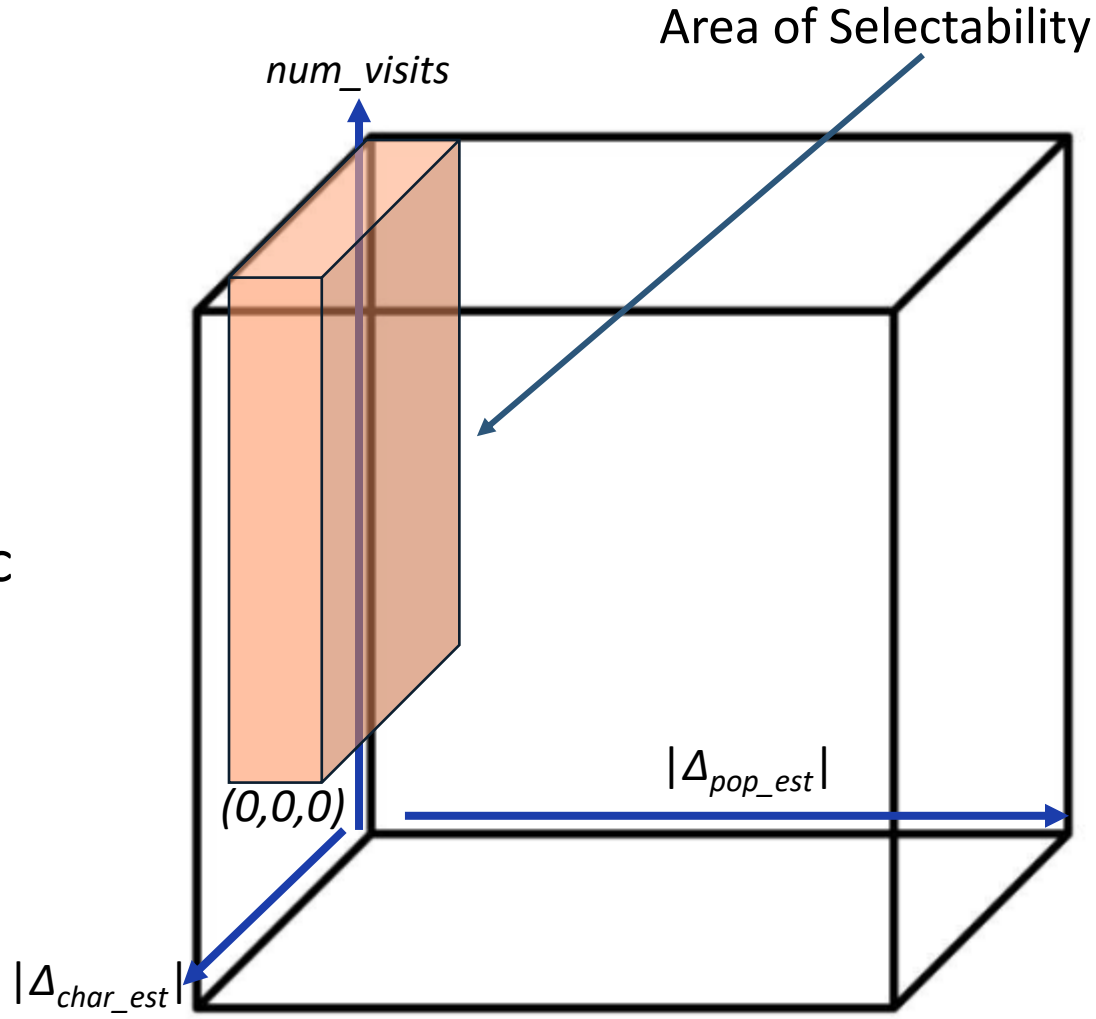
In-Office Enumeration: Selectability

- Three Criteria
  - $|\Delta_{pop\_est}|$: difference between number of people at an address in AD vs field enumeration
  - $|\Delta_{char\_est}|$: difference between expected count of people in different characteristic categories (sex, age, race, ethnicity) of people at an address in AD vs field enumeration
  - *num_visits*: expected number of NRFU attempts needed to enumerate the address



In-Office Enumeration: Selectability

- Three Criteria
  - $|\Delta_{pop\_est}|$: difference between number of people at an address in AD vs field enumeration
  - $|\Delta_{char\_est}|$: difference between expected count of people in different characteristic categories (sex, age, race, ethnicity) of people at an address in AD vs field enumeration
  - *num_visits*: expected number of NRFU attempts needed to enumerate the address



In-Office Enumeration: Modeling

- Focused on three concepts that we cannot know *a priori*
- Employ three key models
  Currently these models are built using a 50% split panel of data from the 2020 Census to estimate coefficients, but all cases are scored
  - Population count disagreement model $\approx |\Delta_{pop\_est}|$
  - Characteristic disagreement model $\approx |\Delta_{char\_est}|$
  - Enumeration effort model $\approx num\_visits$
- Once we have predictions from these models
  - Apply business rules
  - Select cases meeting business rule thresholds

In-Office Enumeration: Data Sources for Modeling

Source	Type of Information	Data Availability
Decennial Census Universe	Address-level geographic and housing unit information known prior to operations	Prior
Person Characteristic Frame (PCF)	Person Identification Key – Master Address File Identification (PIK-MAFID)-level file that includes person-level characteristic information pulled from AD sources	Prior; Updated During
PCF Enhancements	Additional model-based information useful for predicting accuracy of an AD enumeration (mobility/stability likelihoods, consistency of demographic characteristics across AD sources, etc.)	Prior; Updated During
Planning Database	Publicly available low-level aggregate socio-demographic information drawn from 5-year American Community Survey (ACS) Estimates	Prior
Decennial Census Paradata	Information about the number, type, and outcome of various contact attempts (strategies, respondent interactions, etc.)	Accumulated During
Decennial Census Response Data	Address-level response data from enumeration operations	Accumulated During

Modeling: Population Count (PopCount) Disagreement Model

- PopCount Disagreement Model
  - $\hat{\delta}_h^{pop} = (\hat{n}^{AD} - n^{EN}) = f(V, X, Z)$
    - $\hat{\delta}_h^{pop}$: Estimated disagreement for hth housing unit
    - $\hat{n}^{AD}$: Count of PIKs at a MAFID in the PCF
    - n^{EN}: Count of enumerated persons in the Decennial Response Data
    - X: Administrative data known prior to the start of data collection
    - Z: Decennial geographic data known prior to the start of data collection
    - V: Decennial operational data learned during census activities

Covariates *Expected* to Be Related to Larger Disagreement
Incomplete address information on the MAF
Increased count of adults, counts of children on AD input file
Existence and recency of information in AD input file
The more mobile individuals are predicted to be (year-over-year)
More turnover at the housing unit (year-over-year)

Modeling: Characteristic (Char) Disagreement Model

- Characteristic Disagreement Model (sex, age, race, ethnicity)
  - $\hat{\delta}_h^{char} = \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{k}\left(\hat{n}_{jk}{}^{AR} - n_{jk}^{EN}\right) = f(V, X, Z)$
    - δ_h^{char}: Estimated disagreement count for h^{th} housing unit across all four characteristics
    - $\hat{n}_{jk}^{AR}$: Count of PIKs on PCF having value *k* for the j^{th} char (e.g., how many males at the MAFID)
    - n_{jk}^{EN}: Count of individuals in response data having value *k* for the j^{th} char
    - X: Administrative data known prior to the start of data collection
    - Z: Decennial geographic data known prior to the start of data collection
    - V: Decennial operational data learned during census activities

Covariates *Expected* to Be Related to Larger Disagreement
Incomplete address on the MAF
Increased count of adults, counts of children on AD input file
Absence of information from various AD sources
Variability in characteristics across various AD sources
The more mobile individuals are predicted to be (year-over-year)

Modeling: Enumeration Effort (Effort) Model

- Enumeration Effort Model
  - Random Survival Forest Model (Ishwaran et al. 2008)
    - Similar to Random Forests
      - Nonparametric decision trees
      - Averages across many trees to identify best partitions
    - Accounts for censored data (~7%) – in 2020, about 7% of cases were not completed in 6 visits
    - Provides similar concepts as survival modeling
      - Conditional hazard rate $h(t|X) = (S(t-1) - S(t))/(S(t-1)$
      - Expected survival time $E[v] = [T|X] = \sum_{t=1}^{t} S(t-1)$

Covariates *Expected* to Be Important for Prediction
Indicators for varied contact attempt times
Likelihood housing unit is vacant or a delete
Housing unit self-response response propensity
Rural vs urban status
Having refused/expressed lack of interest in earlier attempt

Decision Rules (Models + Constraints)

- Models tell us:
  - PopCount Disagreement Model, $\hat{\delta}_h^{pop}$
    - How similar is the address-level popcount expected to be?
  - Characteristic Disagreement Model, $\hat{\delta}_h^{char}$
    - How similar are the address-level demographics expected to be?
  - Enumeration Effort Model, $E[T|X]$
    - How many visits would enumeration require?
- Constraints tell us:
  - How much error are we willing to absorb? *OR*
  - How much effort reduction do we need?

Sample Decision Rules

Use AD to enumerate cases where…

Rule	Criteria	Comments
1	Predicted PopCount disagreement = 0 $\hat{\delta}_h^{pop} = 0$	Controls error, focusing on only one dimension of error
2	Predicted PopCount disagreement = 0 *and* Predicted Char disagreement is < 1 char/person $\hat{\delta}_h^{pop} = 0 \quad and \quad \hat{\delta}_h^{char} / \hat{\delta}_h^{pop} = 0.25$	Controls error, focusing on two dimensions of error
3	Reduce expected effort by 30% by stopping "most effortful cases" (i.e., predicted to require most visits) $\{h \in (v_{(1)}, v_{(2)} \dots, v_{(n)}) \mid E[V]' = 0.70E[V]\}$	Controls effort/resources • Prioritizes cases expected to take the most visits • In effect this stops the fewest cases to reach 30% visit reduction
4	Reduce expected effort by 30% and minimize PopCount Disagreement $\{h \in (p_{(1)}, p_{(2)} \dots, p_{(n)}) \mid E[V]' = 0.70E[V]\}$	Controls effort/resources *and* error • Stops enough cases to meet resource constraints • Prioritizes cases we predict to have more accurate PopCount

Evaluation Criteria

- Initial factors for evaluation
  - Change in statewide workloads
    - How many cases are removed from NRFU vs 2020?
  - Change in enumeration attempts
    - How many visits are eliminated vs 2020?
  - Change in statewide population count
    - How different are population counts vs 2020?
  - Change in statewide demographics
    - How different are aggregate demographics vs 2020?

Ongoing Work

- Continued model improvement (new/updated datasets)
- Apply a variety of decision rules during simulation
  - Restrict/relax maximum level of error
  - Restrict/relax budget for NRFU visits
- Examine
  - Overall state-level results (like prior slides)
  - Lower-level geographic results (down to the tract level)
  - Key estimation/reporting domains like hard-to-count and historically undercounted populations
- Evaluate consistency of decision rule performance across geographies and domains
- Identify best candidate decision rules

Thank you!

We welcome comments on this ongoing research agenda.

Stephanie Coffey
stephanie.coffey@census.gov