

Census Scientific Advisory Committee

Administrative Records Working Group

**Barbara (babs) Buttenfield, Convener
University of Colorado – Boulder**

September 14 - 15, 2017

Context for Working Group Activities

- Reducing costs of Non-Response Follow-Up (NRFU) for 2020 Census without negative impacts on quality of data collection / reporting
- Use multiple sources of Administrative Records (ARs) from federal, state, private data to build and evaluate a roster of NRFU units, minimizing cost of follow-ups
- Strategies include ARs, statistical modeling and imputation, added field contacts (if needed)

Why is the Use of Administrative Records Important?

- Roughly 50 million NRFU addresses to check (Occupied, Vacant, Non-Existent)
- Use of ARs is an innovative strategy to significantly reduce number of field staff and repeated physical household visits
- Potential for \$1.4 billion reduction in follow-up data collection costs (Tom Mule)
- Time to 2020 Census is short - this needs to run smoothly and work effectively!

Current Sources for Administrative Records (examples)

- **Social Security, IRS 1040 and 1099, HUD data**
- **CMS Medicare and Medicaid**
- **Indian Health Service**
- **CARRA Best Race and Hispanic Origin data**
- **US Postal Service** (undeliverable list)
- State-level veterans, health and human service data (SNAP, **KidLink**)
- **MLS, tax, deed and parcel data** where available
- **2010 and ACS 5 year block group-level estimates**

Working Group Activities

- Review and assess current approaches including ongoing statistical testing
- Provide input on expanding and refining the use of administrative records to reduce respondent burden and to improve statistical analyses
- Offer recommendations on additional data sets to determine vacant, non-existent and occupied housing units

Working Group Activities

Offer recommendations on:

- Exploring ARs to determine characteristics of households and individuals and effects on differential undercounts
- Available state- and local-level and 3rd party resources
- Alternative statistical algorithms and potential for improving quality of estimates
- Methods to address possible regional variations

Working Group Members

CSAC

- Barbara Battenfield
- Allison Plyer
- Ken Simonson
- Jack Levis
- Krishna Rao
- Barbara Anderson
(ex officio)

Census Subject Matter Experts (SMEs)

- Tom Mule – Decennial
- Quentin Brummet –
CARRA
- Andy Keller -Decennial
- Other SMEs as
needed (Moises Yi,
Nicholas Jones)

Timeframe for Working Group Activities

- Very short!
 - By end of 2017, Census will make final decision on administrative records sources
 - On 1 April 2018, testing for 2020 Decennial collection takes place
 - Draft of final report by March 2018 CSAC meeting; final report by July 2018
- Working Group has been meeting monthly with Census SMEs since early August
- Thanks to SMEs for offering time and for answering many questions

Challenges

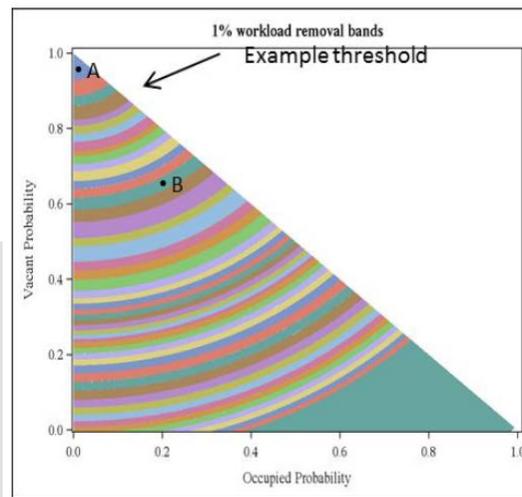
- Reduce risk of bias / error - balance costs of follow-up against improvements in reliability
- Time needed to evaluate and implement a solution is a critical path item

- How will ARs be used?
 - Statistical Imputation (ongoing)
 - How many models for the nation? (ongoing)
 - How to improve upon existing tests?

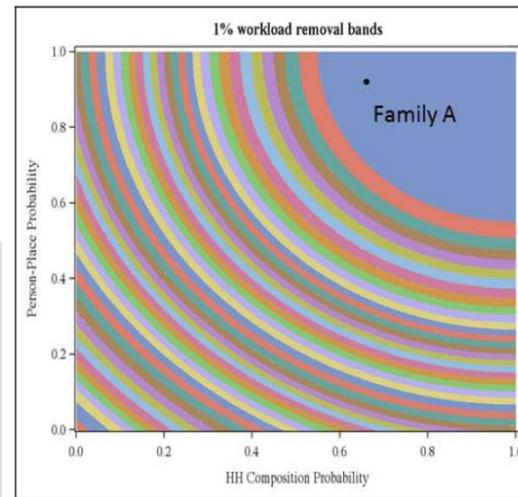
Census work discussed so far: Statistical Imputation (Vacant vs. occupied)

- Multinomial logistic regression creates workload cut-points
 - Predict probabilities for which units are vacant or non-existent, relative to occupied
 - Cut-points determine workload (Keller presentation) for estimating counts and family composition
 - Testing to introduce covariance criterion does not improve on baseline predictions
- Potential concern: imputation for other characteristics: e.g., race / ethnicity, income

Vacant Units



Occupied Units



Census Work Discussed So Far: Subnational Models

- Single model effective for the entire nation?
 - National Model is baseline
 - 3 binary: 'urban vs rural', renter / owner occupied, Hispanic ethnicity
 - 1 model using 4 Census geographic regions
 - Current results show insufficient differences from baseline to warrant subnational models
- Potential concern (approach driven by time constraints)
 - Single test study area (Maricopa County AZ) may not fully represent national range of demographic conditions
 - Subnational model groupings don't match national averages

Challenges to be Discussed with SMEs

- What characteristics can be reliably assigned?
 - Age, sex, household composition – very likely from ARs and imputation
 - Other characteristics – race / ethnicity, tenure, income (to be discussed)
- Assess quality
 - Compare imputed with reported data (2010 and ACS 5 year block group data)
 - Statistical simulation (Andrew Keller showed example for age imputation)
 - Other methods ?
- Potential concerns: reliability may vary by AR data source, by geographic region's demographic characteristics

Additional Topics to be Considered in Coming Months

- Assigning characteristics – how rosters are built (age, sex, race, Hispanic origin, tenure and relationship)
- Alternative sources of AR data (federal, state, 3rd party such as MLS)
- Addressing FRPA concerns about privacy, anonymity with state, local, 3rd party sources
- Additional suggestions of topics are welcome!