

Administrative Data Limitations and the Need for Continued Improvement of Face-to-Face Interviewing for Non-English Speakers in National Surveys

Patricia Goerman, U.S. Census Bureau
Alisú Schoua-Glusberg, Research Support Services
Leticia Fernandez, U.S. Census Bureau

*Presentation for the Federal Committee on Statistical
Methodology (FCSM) 2021 Research and Policy Conference:
November 2-4, 2021
Virtual conference*

Outline of Presentation

- Background on use of Administrative Records (AR) in surveys
- Hard to Count (H2C) populations and AR linkage
- Qualitative insights about survey participation of H2C populations from a 2020 Census evaluation
- Recommendations about improving coverage of under-represented groups

Background

- Administrative Records are data collected by federal and state governments in the course of providing services to program participants
- The Census Bureau has been using administrative records (AR) for decades to update and supplement survey frames, inform edit and imputation routines, and construct new estimates (Mule 2021).
- In the current environment of budget constraints and declining response rates, AR can be a cost-effective way to supplement census data but there are some challenges:
 - Not collected for statistical purposes
 - May not include subpopulations that do not interact with agencies/institutions

Hard-to-Count Populations

- Research has consistently reported that socially and economically marginalized subpopulations are most at risk of census omission (O'Hare 2019).
- Community stakeholders and researchers have voiced concerns regarding how well AR cover historically HTC populations (McClure et al. 2017).
 - homeless, recent immigrants, young children, squatters, formerly incarcerated people, runaway or trafficked youth, etc. (Cohn 2016).
 - Implications for federal/state/local funding for community programs
 - Indicators of well-being could be biased

Hard-to-Count Populations and Administrative Records Linkage

In order to use Administrative Records to fill in survey gaps, it is necessary to link individuals across databases

- Individuals are linked across surveys and AR using a unique **Protected Identification Key (PIK)** based on personal identifiers using probability record linkage techniques.
- PIKs cannot be assigned to individuals with insufficient personal information (SSN, DOB, name, address). PIK assignment is not random (Bond et al. 2014).
- One way to assess coverage of HTC in AR is to match (by PIK) to other sources that may include them (e.g., survey of prison inmates). But there is no gold standard as surveys may also miss them.

Foreign Born in Administrative Records

- Example: study to matched foreign born individuals in the ACS 2006-2010 to Administrative Records (Bhaskar et al. 2018).
 - Of the 2.2 million foreign-born in the ACS sample, 79.5 percent were assigned a unique identifier (PIK) to be able to link them to Admin Records
 - For those assigned a PIK, 97.6 percent could be matched to Admin records
 - Low PIK rate for foreign born means that only 77.6 percent of foreign born in the ACS sample could be matched to Admin Records
 - Foreign born *NOT* in Admin Records had Hard-to-Count characteristics:
 - Recent immigrants, more likely to speak English ‘not well’ or ‘not at all,’ not naturalized citizens, lower levels of education, more likely to report not working last year, more likely to live in tracts with higher presence of foreign born, living in larger households

H2C Populations missing from censuses, surveys and Administrative Records

- H2C populations such as immigrants and non-English speakers maybe missing from censuses, surveys and Administrative Records
- The best chance to get them included in a census or other survey maybe through interviewer administered modes
- Qualitative methods can be extremely useful in finding out how to better reach and include these populations in our data collections

Qualitative Insights from a 2020 Census Doorstep Messaging Evaluation

- Insights about respondent concerns that are missing from administrative records
- Project methodology:
- Initial plan to observe bilingual interviewers and do 5-minute respondent debriefings
- Bilingual enumerator debriefing focus groups over the telephone
- 28 focus groups with up to 6 enumerators per group from various regions across the country.
- Plan: 4 focus groups per language group (English only, and bilingual enumerators who spoke the other 6 languages)

Arabic, Chinese (Mandarin/Cantonese), Korean, Russian, Spanish, Vietnamese

Interviewer Focus Group Composition

Group Language	Group 1	Group 2	Group 3	Group 4	Group 5	Total per language
Monolingual English	5	4	4	N/A	N/A	13
Bilingual Spanish	4	3	3	2	5	17
Bilingual Chinese	5	3	4	5	N/A	17
Bilingual Vietnamese	4	4	4	6	N/A	18
Bilingual Korean	3	5	4	2	N/A	14
Bilingual Russian	1	6	2	4	N/A	13
Bilingual Arabic	4	5	6	4	N/A	19
Total enumerators who participated						111

Insights Gained about Participation of Hard-to-Count Respondents

1. Doorstep messages for encouraging reluctant respondent participation
2. Challenges faced by interviewers
3. Undercount of children

1. Doorstep Messages for encouraging reluctant respondent participation

- Interviewers across languages reported having told respondents about how filling out the census would help their communities in terms of:
 - Schools
 - Community centers
 - Hospitals
 - Emergency services
- Choice of messages based on respondent characteristics
- Confidentiality messages, particularly helpful with non-English speakers
- Enumerators reported that telling the respondent that completing the interview would prevent a follow up visit from another enumerator was helpful

2. Challenges Faced by Interviewers

- Immigrant concerns
 - For Spanish, Arabic, Korean and Chinese speakers, concerns about sharing information
“I just think anybody who’s an immigrant right now might be afraid with the current administration and not trusting what’s going on and think they might use it against them.” – Arabic-speaking enumerator
 - For Spanish speakers, mentions of the government appeared to be a deterrent for participation:
“For me, I think sometimes when I mentioned that it was a government thing, they’d immediately say, ‘Oh, no, no, no, no. I don’t want to do it no more. Never mind.’”
- Sharing a common language builds trust (Finding across languages)
“You know, one thing I sort of noticed I guess is that it felt like having a Korean person go to a Korean household, for me it seemed like I was able to close the case or to complete the interview whereas in the case notes that I read before I knocked on the door, it often said, like, language barrier or refusal. So, it seemed like when I got to their door and immediately started speaking Korean to them, they were still annoyed but more open to I guess completing the interview.”

3. Undercount of Children and Others

- Undercount of children age 0-4 is a well documented and recurrent problem across multiple censuses (Jensen et al. 2018)
- It has gotten worse in this age group, while coverage of other age groups seems to have improved (O'Hare 2015)
- Children do not always get a PIK (for lack of some or all data needed), and this is more pronounced for Hispanic and racial minorities (Fernandez et al. 2018)
- Interviewers felt that omissions of children easily corrected in instrument
 - Observation and respondent debriefing would likely be a better way to study this
- Incomplete information about children seemed more common to them than complete omissions
 - Reluctance about sharing details on children in the household, especially date of birth
- Apartment buildings where more people living than allowed
 - Interviewers sometimes suspected omissions from respondents and apartment managers

Recommendations for Improving Coverage of Under-represented Groups

- While Administrative Records can fill in many gaps due to decreasing response rates, they are not equally useful and accurate across all populations
- We recommend continued multi-mode research and improvements targeting inclusion of Hard-to-Count populations in national surveys
- In our example, the interviewer perspective provided many rich and interesting details. May not be the same as how respondents perceive things and/or what an observer would see in the field
- Recommend future observational field studies that include incorporating respondent voices into the process

Administrative Data Limitations and the Need for Continued Improvement of Face-to-Face Interviewing for Non-English Speakers in National Surveys

Thank you!

For more information:

Patricia.L.Goerman@census.gov

alisu@researchsupportservices.com

leticia.esther.fernandez@census.gov

This presentation is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not those of the U.S. Census Bureau. This talk has been reviewed for disclosure avoidance and approved under clearance number: CBDRB-FY21-CBSM002-015

References

- Bhaskar, Renuka, Leticia Fernandez and Sonya R. Porter. 2018. Assimilation and coverage of the foreign-born population in administrative records. *Statistical Journal of IAOS*, 34(2): 191-201.
- Bond, Brittany, J. David Brown, Adela Luque, Amy O'Hara. 2014. The Nature of the Bias When Studying Only Linkable Person Records: Evidence from the American Community Survey. CARRA Working Paper Series #2014-08. Washington, DC: U.S. Census Bureau.
- Cohn, D'Vera. 2016. Census Bureau hopes to use data from other government agencies in 2020. Washington, DC: Pew Research Center.
- Fernandez, Leticia, Rachel Shattuck, and James Noon. 2018. The Use of Administrative Records and the American Community Survey to Study the Characteristics of Undercounted Young Children in the 2010 Census. CARRA Working Paper Series #2018 – 05. Washington, DC: U.S. Census Bureau.
- McClure, Dave, Robert Santos and Shiva Kooragayala. 2017. Administrative Records in the 2020 US Census, Civil Rights Considerations and Opportunities. Washington, DC: Urban Institute.
- O'Hare, William. 2019. Differential Undercounts in the U.S. Census Who is Missed? Springer Open.