

# Potential uses of individual-level AR data & PIK coverage in the 2012 Survey of Business Owners

Adela Luque  
Renuka Bhaskar  
Sharon Ennis  
James Noon

U.S. Census Bureau

Federal Committee on Statistical Methodology

March 9, 2017

Acknowledgements: We wish to thank Elizabeth Goldberg (EWD), Richard Moore (ESMD), James Hunt (ESMD), Patrice Norman (EWD), Naomi Blackman (EWD), Joseph Barth (ESMD), Jim Liu (ESMD) for their time and help.

Disclaimer: Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

# Overview

- Background
- Objectives
- Data
- Findings:
  - Protected Identification Key (PIK) coverage
  - Veteran status
  - Nativity or citizenship at birth
  - Race & ethnicity
- Conclusion

# Background

- Survey of Business Owners
  - Only comprehensive source of info about employer & nonemployer businesses by gender, race, ethnicity and veteran status of the business owner
  - Survey of nonfarm businesses w/ revenues of \$1000 or more
  - Every 5 years in years ending in '2' and '7'
  - Includes employer and non-employer businesses, and different types of firms (i.e., sole-proprietorships, partnerships, private and public corporations)
  - Gender, race, Hispanic origin and veteran status are 'core demographics' in the SBO => they're imputed if missing

# Objectives

- Analyze 2012 SBO data on veteran status, race & ethnicity, and nativity in comparison to other data sources (i.e., administrative records and/or previous census records data)
- Assess the suitability of those data sources for imputation/direct replacement of veteran status, race & ethnicity and nativity in the SBO/business surveys
- Help inform business surveys' existing imputation methods – reduce respondent burden & operational costs while maintaining data quality

# Objectives

- PIK – anonymized unique individual identifier
- PIK coverage in 2012 SBO & potential expansion
- Why is this of interest?
  - PIKs - linking key across data sources
  - PIKs can help reduce respondent burden and operational costs
  - Core demographics (race, ethnicity, veteran status & gender) of *non-employer sole-proprietorship* owners in 2012 SBO imputed/replaced from ACS/decennial via PIK

# Data

- 2012 SBO firms with classifiable owner characteristics (e.g., no public corps) that have at least one alive/operating establishment
- Unweighted data => Percentages/figures not meant to be representative of underlying population
- By design, 2012 SBO aims for 100% PIK coverage of owners of ***non-employer sole proprietorships*** (but not for other types of firms)
  - Core demographics via PIK instead of survey response
  - Studies focus on sole proprietorships – since PIKs needed to link SBO business owners to other data

# PIK coverage of sole-proprietorship owners in 2012 SBO

	All Sole-props	PIKs	
	Number (1)	Row % (2a)	Col % (2b)
All	8,216,610	94.6%	100.0%
Non Employer	7,754,864	99.4%	99.2%
Employer	461,746	12.8%	0.8%
Non-employer, Form sent	641,057	93.0%	7.7%
Non-employer, Form not sent	7,113,807	100.0%	91.6%
Employer, Form sent	461,746	12.8%	0.8%
Employer, Form not sent	0	n/a	n/a

Source: 2012 SBO.

Can PIK coverage potentially be expanded?

- Yes, with tax data (& IRS approval)

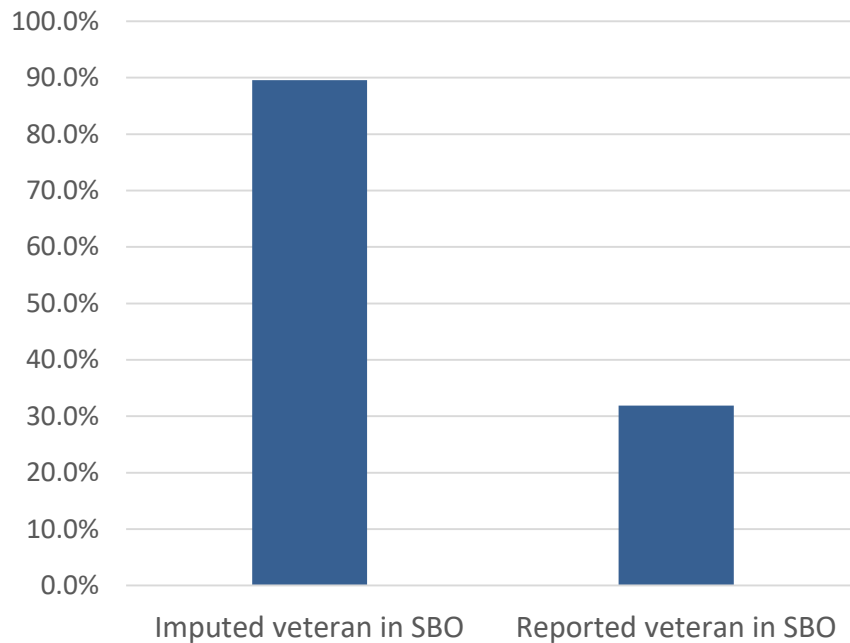
# 2012 SBO Veterans Study

- Veteran status is a 'core' demographic in SBO => imputed if missing
- Data used:
  - 2013 Veterans Administration (VA) data

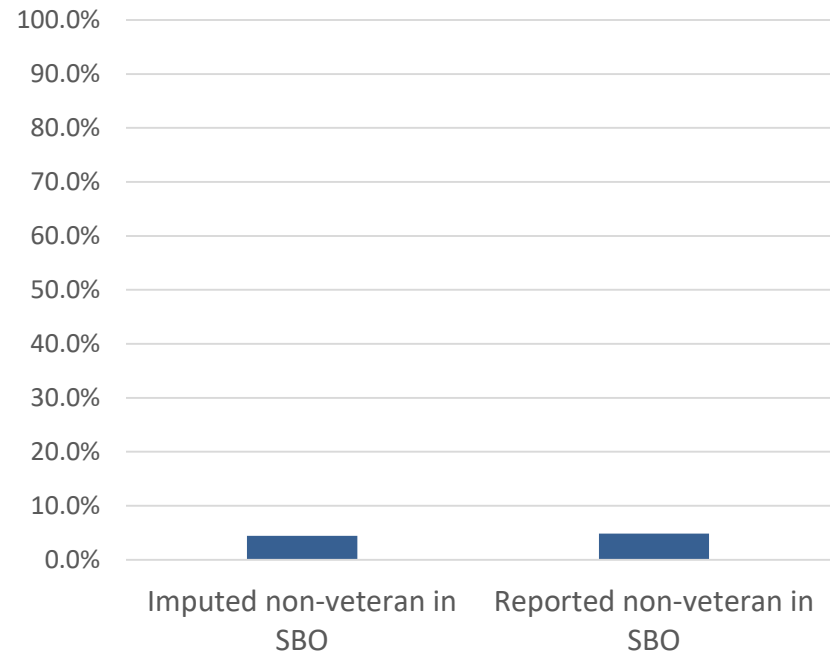


# Preliminary results - Disagreement rates between 2012 SBO & VA data

Percent of business owners coded as veterans in the 2012 SBO that are not veterans according to VA data



Percent of business owners coded as non-veterans in the 2012 SBO that are veterans according to VA data



Source: 2012 Survey of Business Owners (SBO) & 2013 Veteran Administration data.

# 2012 SBO Nativity Study

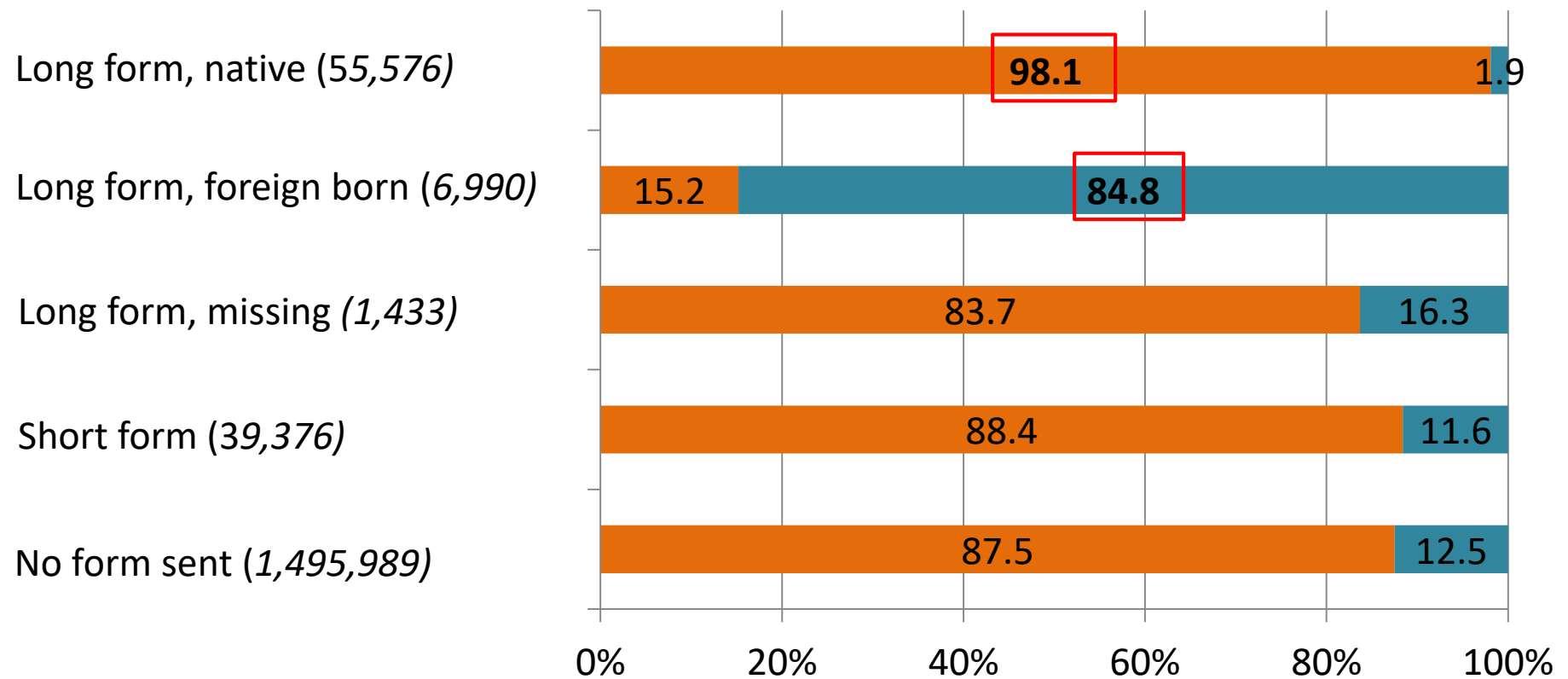
- Background:
  - Nativity (i.e, are you born a citizen of the U.S.?) is not a 'core' demographic in the SBO => not imputed if missing
  - Nativity only asked on the long form
- Data used:
  - Previous Census Responses (Census 2000 long form & American Community Survey 2001-2011)
  - 2011 Social Security Administration's Numerical Identification File (Numident)

# Agreement: SBO & Previous Census Records

## SBO Form Type and Nativity

### Previous Census Records Nativity

Native Foreign Born



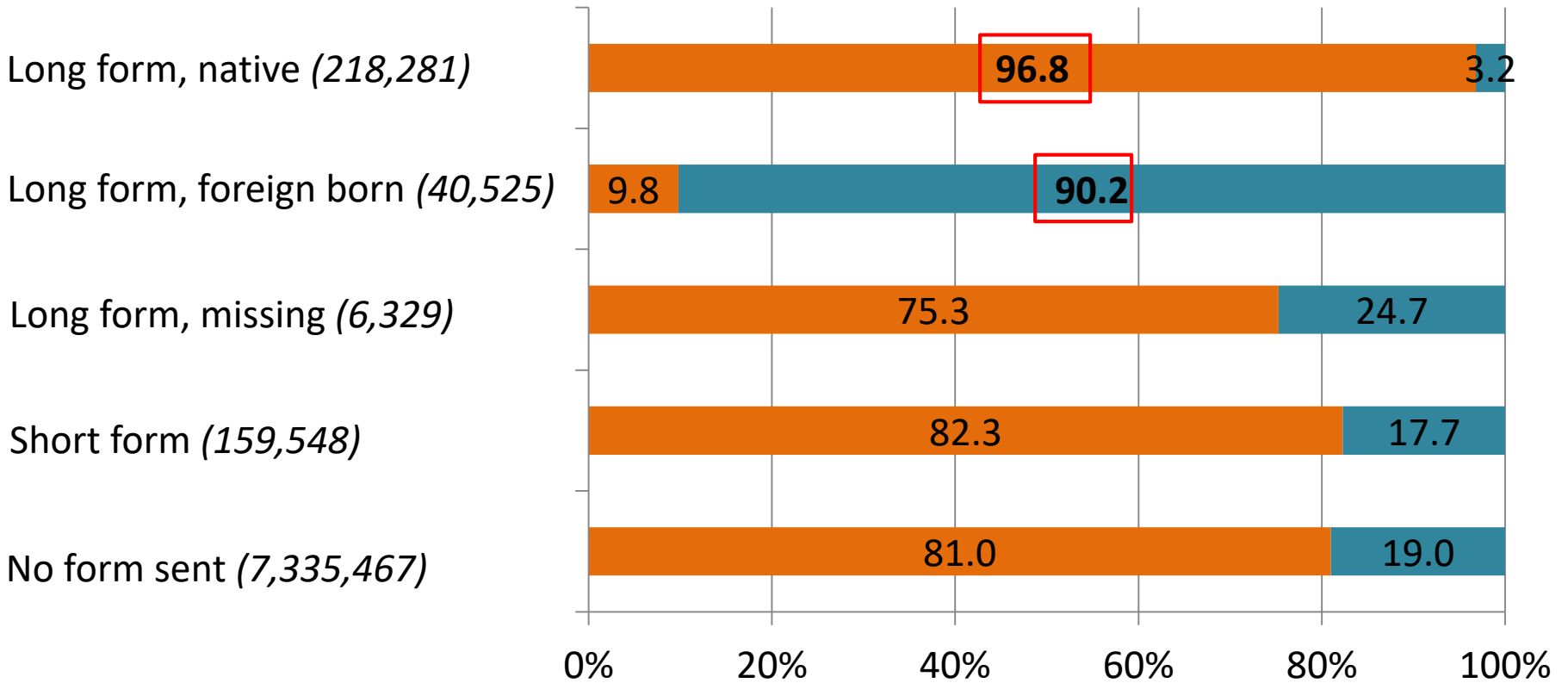
Source: 2012 Survey of Business Owners, Census 2000, and ACS 2001-2011.

# Agreement: SBO and Numident

## SBO Form Type and Nativity

### Numident Place of Birth Foreign Indicator

■ Born in the U.S. ■ Born outside the U.S.



Source: 2012 Survey of Business Owners and 2011 Numident.

# 2012 SBO Race & Ethnicity Study

- Race and ethnicity are 'core' demographics in the SBO
- 2012 SBO already used previous census records (2001-2010 ACS & 2000, 2010 decennials) to direct replace race & Hispanic origin for non-employer sole-props
  - Confirmation of high agreement rate (near 100%) between previous census records & SBO cases imputed using decennial or ACS data
- Study explores use of AR race and Hispanic origin composite
  - AR composite provides additional coverage for about 30,000 SBO owners that were mailed a survey form

# Conclusions

- Possible to increase PIK coverage to further reduce costs & maintain data quality - using tax data w/ IRS approval
- Business surveys can benefit from using VA data to assign/impute veteran status to business owners w/ PIKs – once quality of VA data is validated
- Nativity data are missing for most owners
  - Previous census records &/or Numident can be used to impute nativity information for business owners w/ PIKs
- In addition to ACS & decennial census data, business surveys could make use of the AR composite to impute/direct replace race and Hispanic origin
- Keep in mind: data agreements, potential bias in AR & definitional differences

# Thank you!

[adela.luque@census.gov](mailto:adela.luque@census.gov)