

# 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

FedCASIC  
April 11, 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

# Background

- Survey of Business Owners
  - Only comprehensive source of info about 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
- Help inform SBO's existing imputation methods – reduce 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 - instead of via survey response

# 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 (1a)	Col % (1b)	Number (2a)	Row % (2b)	Col % (2c)
<b>All</b>	8,216,610	100.0%	7,769,108	94.6%	100.0%
<b>Non Employer</b>	7,754,864	94.4%	7,710,185	99.4%	99.2%
<b>Employer</b>	461,746	5.6%	58,923	12.8%	0.8%
<b>Non-employer, Form sent</b>	641,057	7.8%	596,378	93.0%	7.7%
<b>Non-employer, Form not sent</b>	7,113,807	86.6%	7,113,807	100.0%	91.6%
<b>Employer, Form sent</b>	461,746	5.6%	58,923	12.8%	0.8%
<b>Employer, Form not sent</b>	0	0.0%	n/a	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



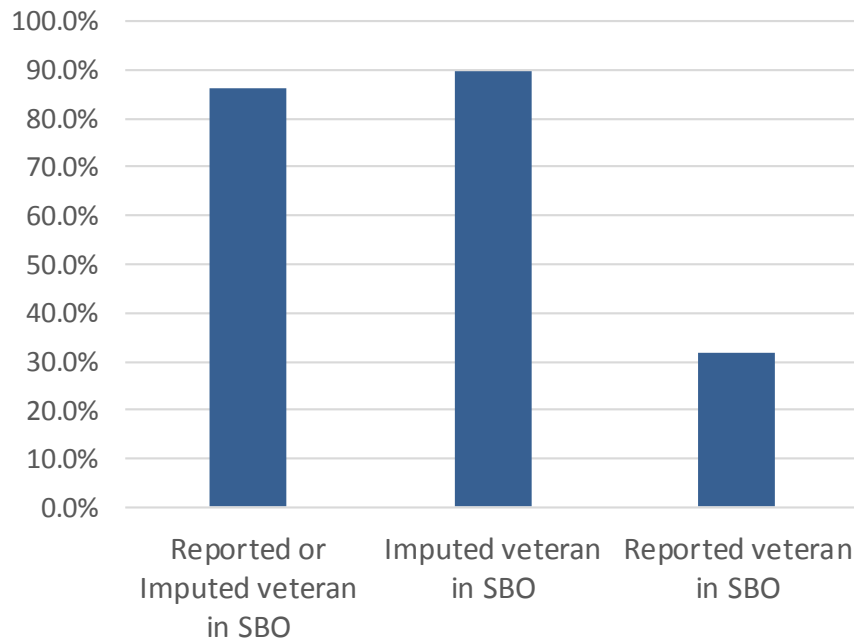
## Veteran Status Distribution of 2012 SBO Owners by Imputation Status – Sole Proprietorships

	Number	Column %	Column %
Total	8,216,590	100.0%	100.0%
Reported	725,680	8.8	
Veteran	80,251		1.0
Not Veteran	645,429		7.8
Imputed	7,490,910	91.2	
Veteran	741,892		9.0
Not a Veteran	6,749,018		82.2

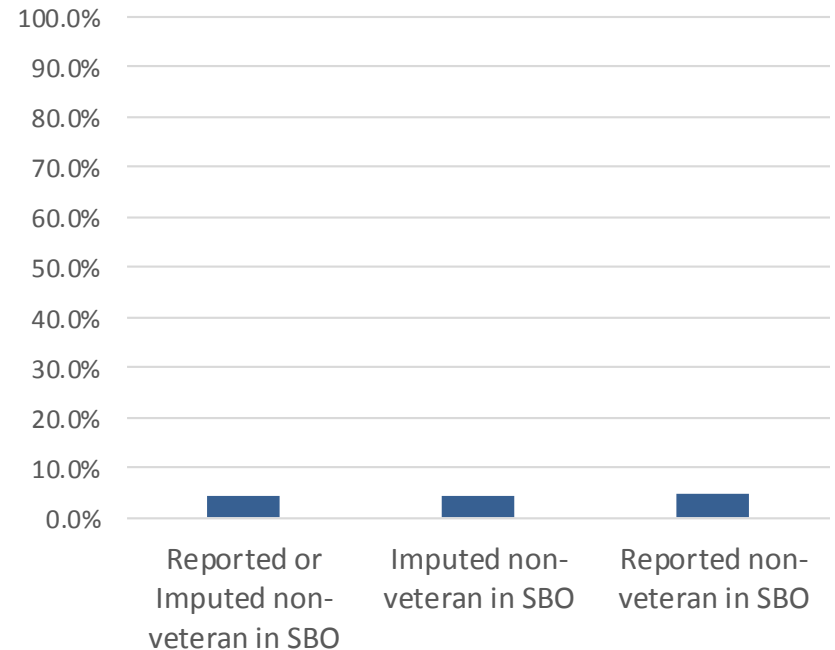
Source: 2012 Survey of Business Owners

# 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 only
  - As a result, nativity information is missing for most owners in 2012 SBO
- Data used:
  - Previous Census Responses (Census 2000 long form & American Community Survey 2001-2011)
  - 2011 Social Security Administration's Numerical Identification File (Numident)

# Form Type and Nativity, Sole Proprietorships

Form type and nativity	Number	Percent
Total	8,216,590	100.0
Long Form	422,538	5.1
Native	346,123	4.2
Foreign born	67,656	0.8
Missing value	8,759	0.1
Short form	321,735	3.9
No form	7,472,317	90.9

Source: 2012 Survey of Business Owners

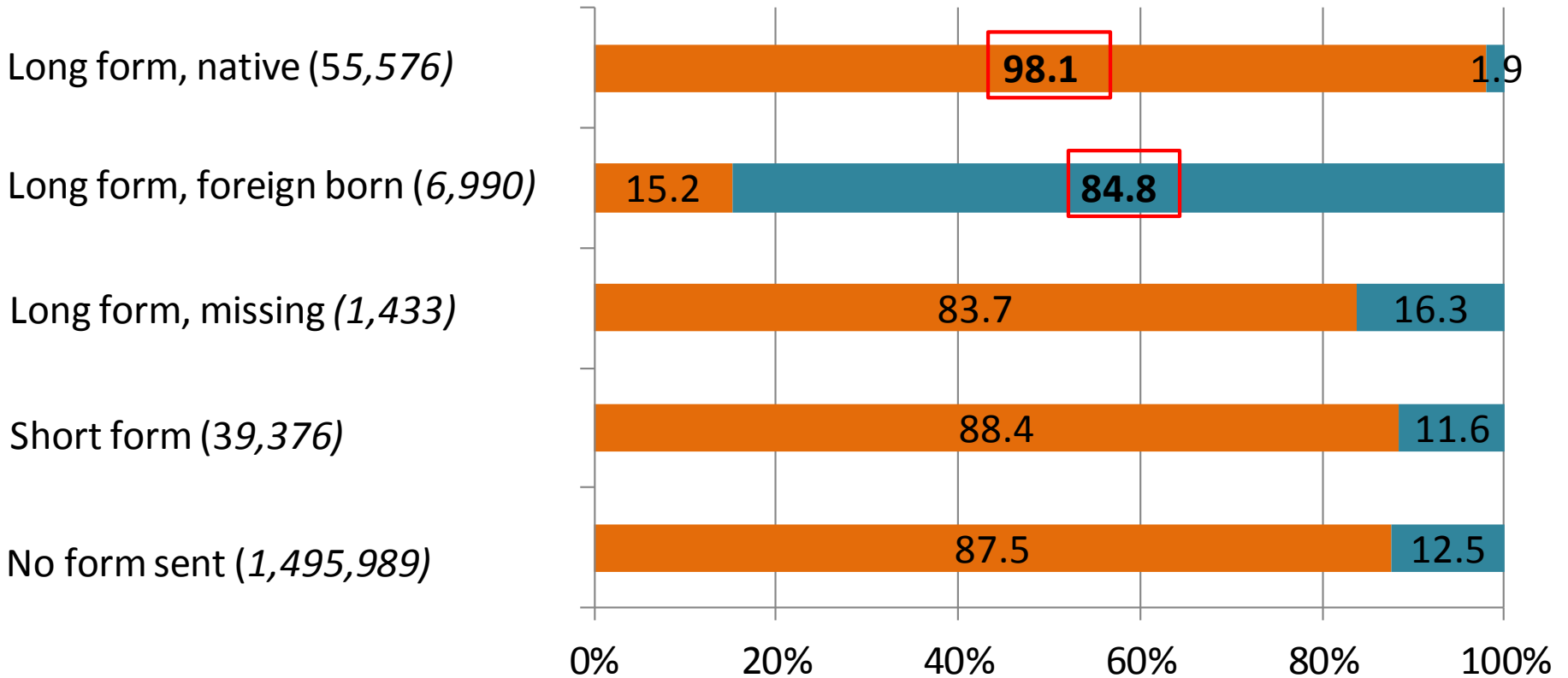
**Nativity information is available for only about 5 percent of SBO owners of sole proprietorships.**

# Agreement: SBO and Previous Census Records

## SBO Form Type and Nativity

### Previous Census Records Nativity

Native Foreign Born



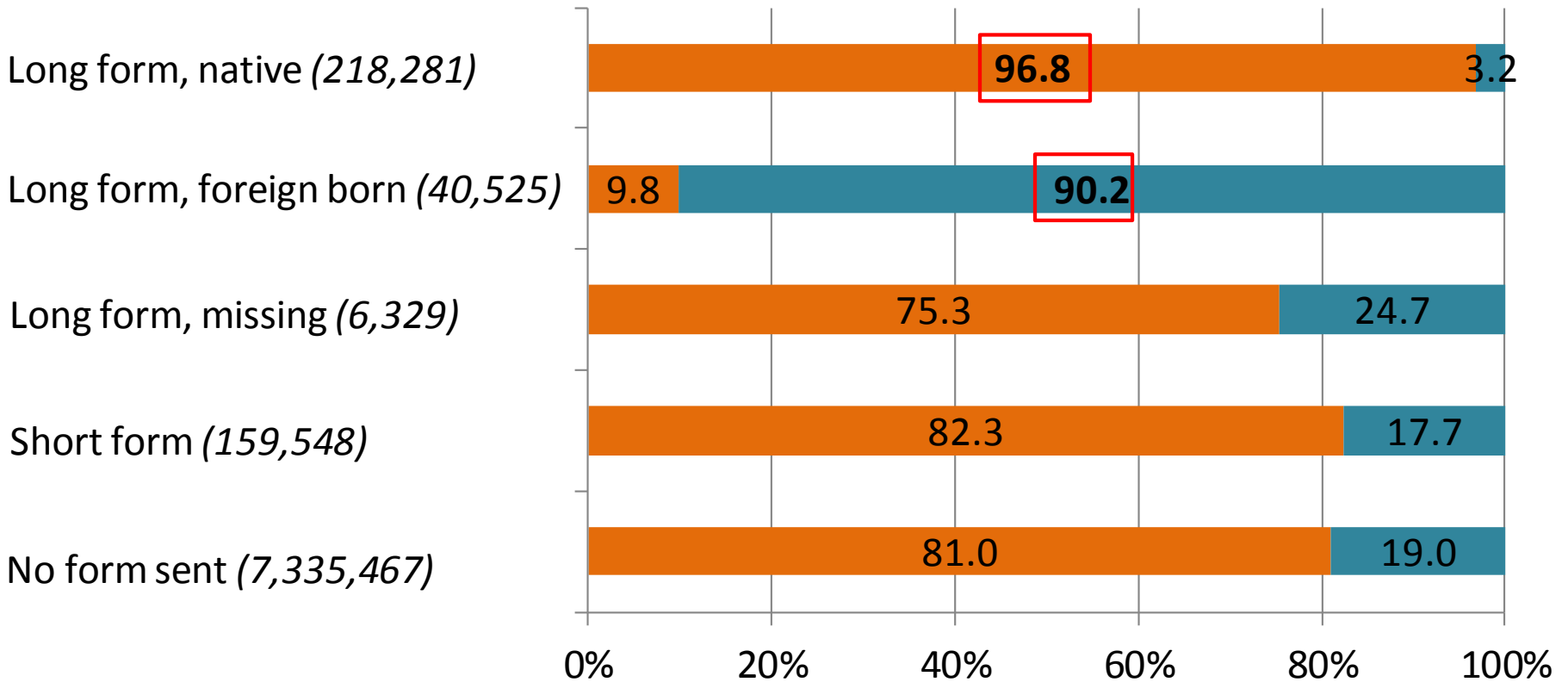
Source: 2012 Survey of Business Owners, Census 2000, and ACS 2001-2011.

# 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
- Study explores use of AR race and Hispanic origin composite

# 2012 SBO Race & Ethnicity Findings

- Study confirms high agreement rate (near 100%) between previous census records & SBO cases imputed using decennial or ACS data
- 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
- SBO can benefit from using VA data to assign veteran status to SBO owners w/ PIKs
- Nativity data are missing for most owners
  - Previous census records &/or Numident can be used to impute nativity information for SBO owners w/ PIKs
- In addition to ACS & decennial census data, SBO could make use of the AR composite to impute/direct replace race and Hispanic origin (for SBO owners w/ PIKs)
- Keeping in mind: data agreements, potential bias & any definitional differences

# Thank you!

[Adela.Luque@census.gov](mailto:Adela.Luque@census.gov)

# PIK coverage expansion?

- Potentially feasible to expand PIK coverage in future SBOs:
  - to low 90s% if person tax records of **employer sole-props** can be used to assign PIKs, and
  - to mid/high 90s% if K1 tax data for **partnerships and S-corps** can be used to assign PIKs
- Not currently feasible to get PIKs for owners of private corporations (non S-corps)

# PIK Coverage of Sole-proprietorship Owners in 2012 SBO by (core) Demographic Group

	All Sole-props		PIKs		
	Number (1)	Col % (2)	Number (3)	Row % (4)	Col % (5)
<b>All</b>	8,216,610	100.0%	7,769,108	94.6%	100.0%
<b>Race</b>					
White	5,565,949	67.7%	5,184,905	93.2%	66.7%
Black	1,532,625	18.7%	1,517,638	99.0%	19.5%
Asian	444,667	5.4%	404,536	91.0%	5.2%
AIAN	114,623	1.4%	111,233	97.0%	1.4%
NHPI	22,283	0.3%	21,561	96.8%	0.3%
SOR	536,463	6.5%	529,235	98.7%	6.8%
<b>Hispanic origin</b>					
Hispanic	1,314,963	16.0%	1,287,997	97.9%	16.6%
Non-Hispanic	6,901,647	84.0%	6,481,111	93.9%	83.4%
<b>Sex</b>					
Female	4,215,300	51.3%	4,107,871	97.5%	52.9%
Male	4,001,310	48.7%	3,661,237	91.5%	47.1%
<b>Veteran status</b>					
Veteran	822,145	10.0%	773,163	94.0%	10.0%
Non veteran	7,394,465	90.0%	6,995,945	94.6%	90.0%

Source: 2012 SBO.

# Non Sole-props - Employer Identification Number (EINs)\* & PIKs

Firms w/ multiple owners (non sole-props)	
572,344	
With EINs (cell 1)	W/out EINs (cell 2)
400,764	171,580
70.0%	30.0%

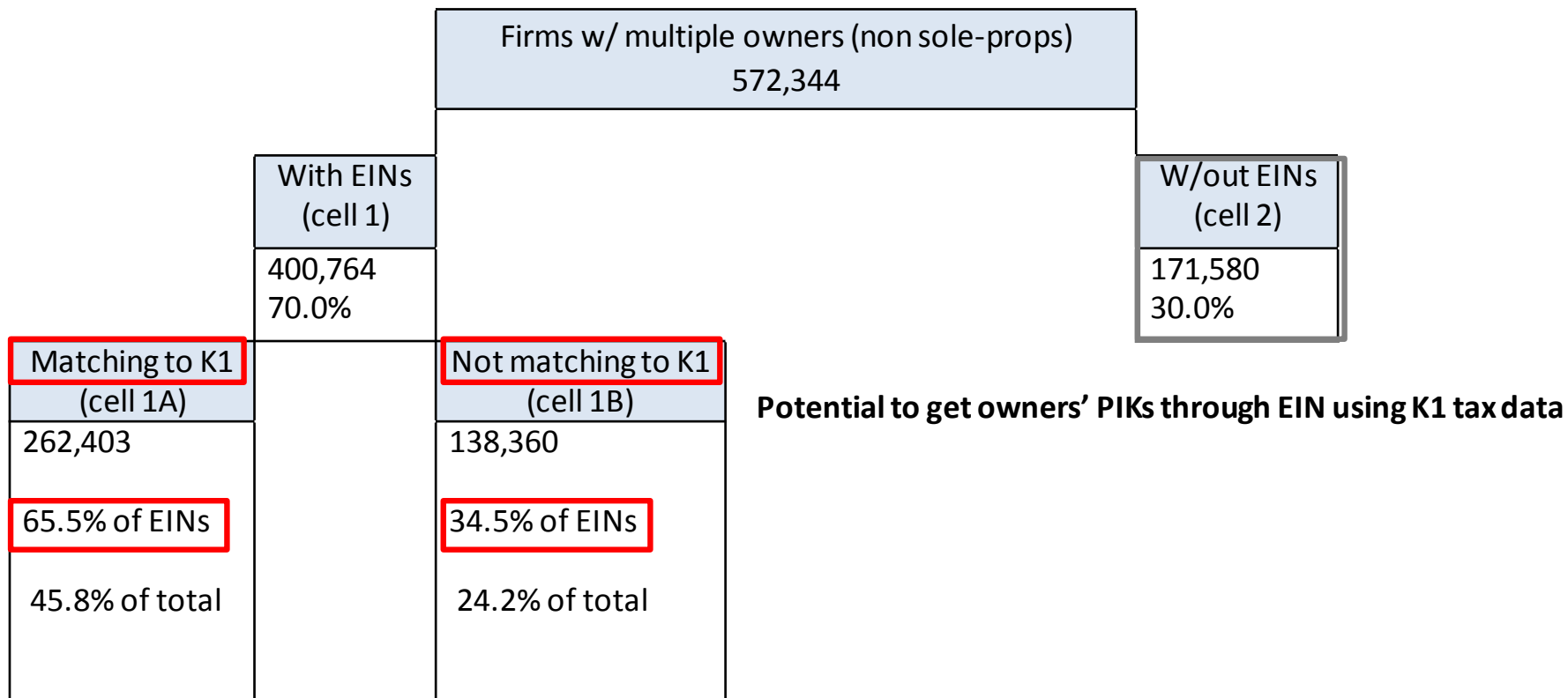
- 99% are employers
- 53% are male owned

- 100% are non-employers
- 53.2% are female-male equally owned

Sources: 2012 SBO & K1 tax data.

\*The source of EINs in the SBO is the Business Register.

# Non Sole-props - Employer Identification Number (EINs)\* & PIKs



Sources: 2012 SBO & K1 tax data.

\*The source of EINs in the SBO is the Business Register.

# Non sole-props - Employer Identification Number (EINs)\* & PIKs

Firms w/ multiple owners (non sole-props) 572,344			
With EINs (cell 1)			W/out EINs (cell 2)
400,764 70.0%			171,580 30.0%
	w/ PIK owner-spouse (cell 2A)	No PIK found (cell 2B)	
	89,359	82,221	
	52.1% of firms w/out EIN	47.9% of firms w/out EIN	
	15.6% of total	14.4% of total	

- Over half of non sole-props w/out an EIN have PIKs of owner & spouse
- Not surprisingly, 72.6% of them are equally owned by a man & a woman
- 100% sole-props accdg. to LFO field

Sources: 2012 SBO & K1 tax data.

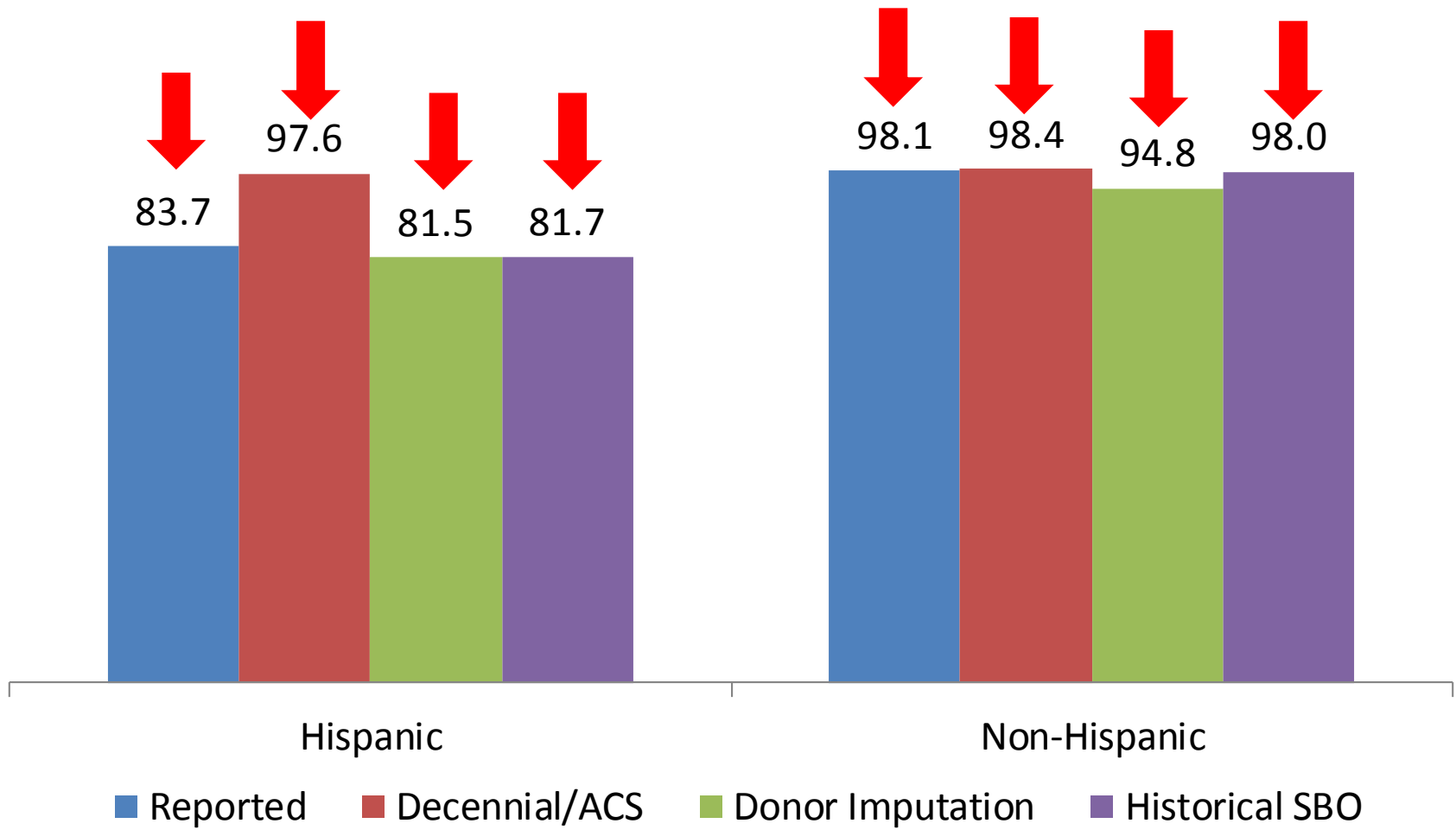
\*The source of EINs in the SBO is the Business Register.

# Summary of Findings

- Nativity data are missing for most owners
- Numident
  - provides high coverage of SBO records
  - place of birth data in high agreement with SBO nativity
- Previous census records
  - provides lower coverage of SBO records
  - nativity data in high agreement with SBO nativity

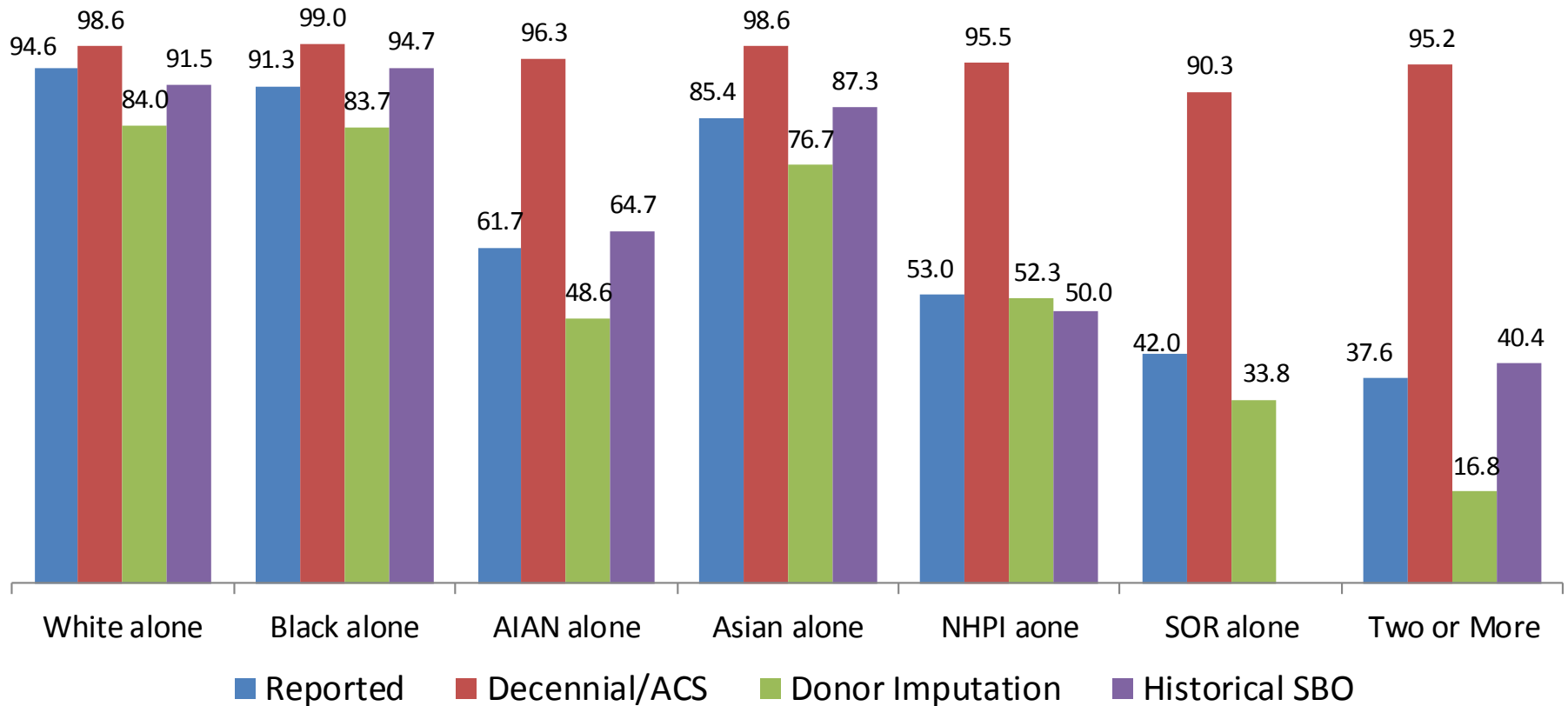


# SBO/Previous Census Records Hispanic Origin Response Agreement – Sole Proprietorships



Sources: 2012 SBO and Previous Census Records.

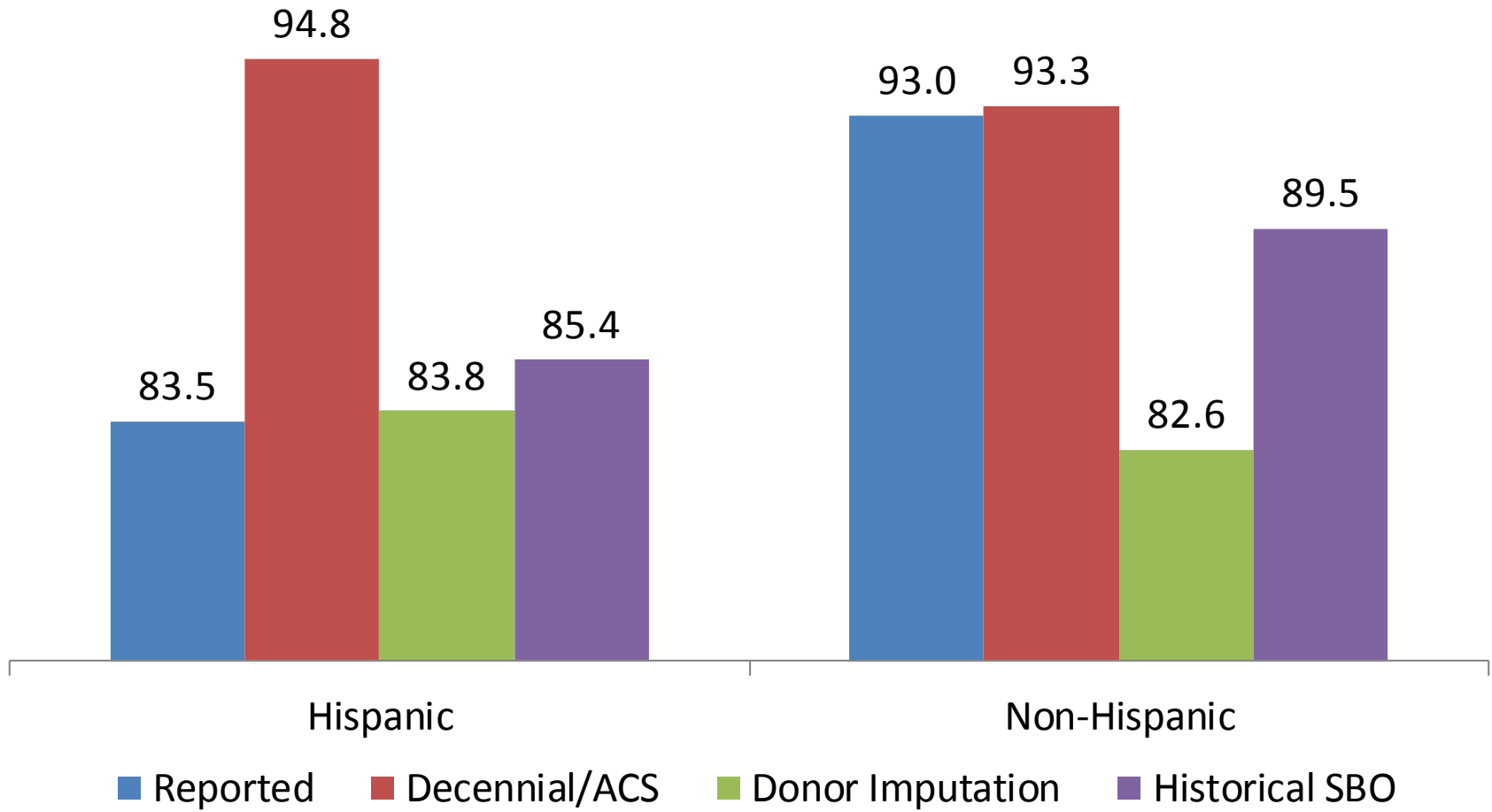
# SBO/Previous Census Records Race Response Agreement – Sole Proprietorships



Sources: 2012 SBO and Previous Census Records.

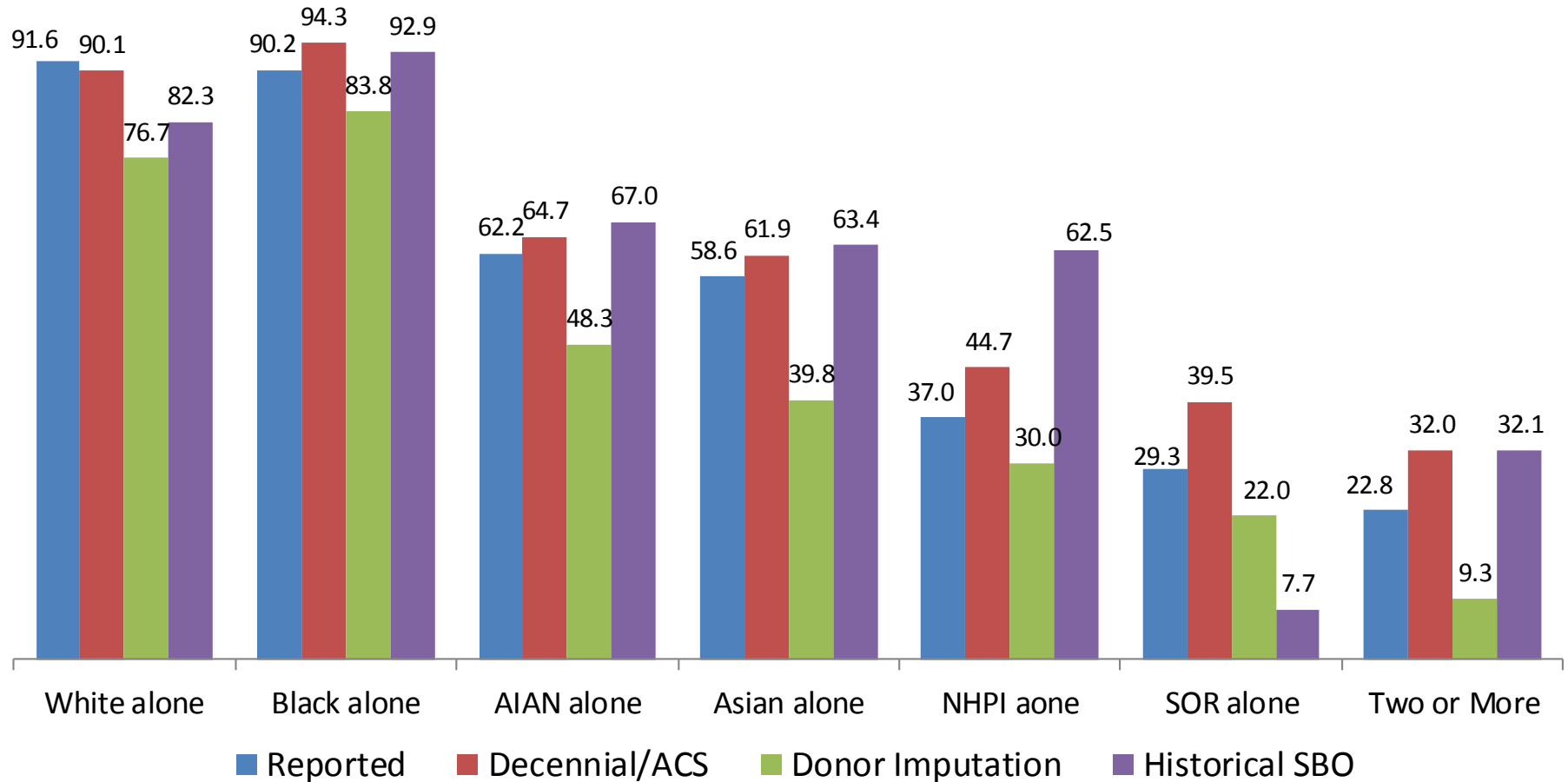
AIAN = American Indian or Alaska Native; NHPI = Native Hawaiian or Other Pacific Islander; SOR = Some Other Race

# SBO/Administrative Records Hispanic Origin Response Agreement – Sole Proprietorships



Sources: 2012 SBO and Administrative Records.

# SBO/Administrative Records Race Response Agreement – Sole Proprietorships



Sources: 2012 SBO and Administrative Records.

AIAN = American Indian or Alaska Native; NHPI = Native Hawaiian or Other Pacific Islander; SOR = Some Other Race