Predicting the Effect of Adding a Citizenship Question to the 2020 Census

J. David Brown
Misty L. Heggeness
Suzanne M. Dorinski
Lawrence Warren
Moises Yi

April 11, 2019

The analysis, thoughts, opinions, and any errors presented here are solely those of the authors and do not reflect any official positions of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. The Disclosure Review Board release number is DRB-B0035-CED-20190322.

United States™ Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Citizen Voting Age Population (CVAP) Statistics

- Produced by Census Bureau annually at block group level
- Source: 5-year ACS
- Population and persons age 18 and over who are U.S. citizens, by race/ethnicity
- CVAP used by Dept. of Justice for Voting Rights Act enforcement
- 2011 CVAP used 2005-2009 ACS, released near same time as 2010 Census PL94 redistricting data (April 1, 2011)
- On Dec. 12, 2017 Dept. of Justice requested citizenship question be added to 2020 Census so CVAP could be produced at block level

Why Household Self-Response is Important

- If household doesn't self-respond
  - Enumerators attempt contact on up to 6 days
  - Seek proxy response from neighbor
  - Whole-household imputation

- Cost increases by estimated $55 million for every percentage point increase in Nonresponse Followup (NRFU)

- Quality declines
  - In 2010, 97.3% correct enumeration rate for self-responses, 93.4% for household interviews, and 70.2% for proxy responses
  - 96.7% linkage rate to administrative records for self-responses, 33.8% for proxy responses

Literature (1 of 2)

- Dillman, Sinclair, and Clark (1993)
  - Randomized Controlled Trial (RCT) shows that asking for SSN decreases decennial response by 3.4 percentage points overall, and by 6.2 percentage points in areas with low mail response rates
- Guarino, Hill, and Woltman (2001)
  - 2000 Census RCT shows 2.1 ppt lower self-response rate in high-response areas, 2.7 ppt lower rate in low-response areas with questionnaires containing SSN request
- Singer, Mathiowetz, and Cooper (1993)
  - Households with confidentiality concerns were less likely to self-respond to the 1990 Census
- Singer, Van Hoewyk, and Neugebauer (2003)
  - Belief that census may be misused for law enforcement purposes was significant negative predictor of self-response in 2000 Census

Literature (2 of 2)

- O'Hare (2018)
  - Citizenship question has higher item allocation rate in ACS than other variables that will be in 2020 Census
  - Increasing over time
  - Higher for racial and ethnic minorities, foreign born, and self-responders
- McGeeney et al. (2019)
  - In 2020 Census Barriers, Attitudes, and Motivators Study (CBAMS), 32.5% of foreign-born respondents "extremely concerned" or "very concerned" that Census Bureau will share answers with other govt. agencies, vs. 24.0% among others
  - 34.0% of foreign-born "extremely concerned" or "very concerned" that answers will be used against them, vs. 22.0% among others
- Escudero & Becerra (2018)
  - In survey in Providence, Rhode Island (site of 2018 End-To-End Census Test), 75% of men and 83% of women agreed with statement "many people in Providence County will be afraid to participate in the 2020 Census because it will ask whether each person in the household is a citizen."

Measuring Effect of Citizenship Question on Self-Response Rate

- Natural experiment: random sample of 1,418,000 households receiving both ACS (with citizenship question) and Census (without) in 2010
- Households may be less willing to respond to one survey than the other for reasons other than citizenship question
- Divide households into ones likely more vs. less sensitive to citizenship question
  - Less sensitive: everyone in household is citizen in ACS and admin. data
  - More sensitive: all other households
- Difference between self-response rate across surveys for less sensitive group represents general difference in propensity to self-respond across surveys
- Difference-in-differences can isolate citizenship question effect

Measuring Effect of Citizenship Question on Self-Response Rate

- $G \in (S, U)$, *S* is potentially sensitive to a citizenship question, while *U* group is not
- $R_{G_i ACS_t}$ and $R_{G_i Census_t}$ = 1 if household *i* in group *G* self-responds in year *t* to the ACS and Census, respectively, and zero otherwise
- Difference between the survey responses is

$$\Delta R_{G_i t} = R_{G_i ACS_t} - R_{G_i Census_t}$$

- Difference-in-differences in expected self-response rates across the two surveys for the two groups *S* and *U* in year *t* is

$$\Delta\Delta R_{SU_t} = E\left(\Delta R_{S_t}\right) - E\left(\Delta R_{U_t}\right)$$

Data Sources

- American Community Survey (ACS) in 2010, 2017
- 2010 Census
- 2010, 2017 Social Security Administration (SSA) Numident
  - Misses persons without Social Security Numbers (SSNs)
  - Not all naturalized persons report their status change to SSA, or they do so with delay
- Individual Tax Identification Numbers (ITINs)
  - Persons who need to pay taxes, but do not have work authorization

Comparison of 2010 ACS to 2010 Census Self-Response Rates

	Self-Response Rate (%)		Difference
	2010 ACS	2010 Census	
All other households	42.0	62.7	-20.7
AR & ACS all-citizen households	65.6	74.4	-8.9
Difference-in-differences			-11.9

Blinder-Oaxaca Decomposition

- Households potentially containing noncitizens could have a greater difference between their Census and ACS self-response propensity for reasons other than citizenship question
  - Those containing noncitizens may be more likely to be linguistically isolated
  - Linguistically isolated households may find a longer questionnaire particularly burdensome
- Blinder-Oaxaca decomposition can control for systematic observable differences between groups like linguistic isolation

Blinder-Oaxaca Decomposition

- We estimate OLS models for each household group:

- $\Delta R_{S_{it}} = X'_{S_{it}}\beta_{S_t} + \varepsilon_{S_{it}}$

- $\Delta R_{U_{it}} = X'_{U_{it}}\beta_{U_t} + \varepsilon_{U_{it}}$

- $\Delta\Delta R_{SU_t} = E\left(\Delta R_{S_t}\right) - E\left(\Delta R_{U_t}\right)$

- $\Delta\Delta R_{SU_t} = \left[E\left(X_{S_t}\right) - E\left(X_{U_t}\right)\right]'\beta_{U_t} + \left[E\left(X_{S_t}\right)'\left(\beta_{S_t} - \beta_{U_t}\right)\right]$

Blinder-Oaxaca Decomposition

- Explanatory variables (*X*'s) include
  - log household size and its square
  - owned vs. rented
  - housing structure type
  - household income
  - presence of related and unrelated children, unrelated adults, only working adults
  - householder sex crossed with marital status
  - householder age, race/ethnicity, education, recently moved here
  - linguistic isolation
  - shares of housing units in block group with at least one noncitizen, under poverty line, vacant
  - tract population density

Blinder-Oaxaca Decomposition of Comparison of Predicted 2010 ACS to 2010 Census to Self-Response Rates by All-Citizen vs. All Other Households

	2010 ACS – 2010 Census
All other households	-20.7
AR & ACS all-citizen households	-8.9
Difference-in-differences	-11.9
Explained	-3.1
Unexplained	-8.8

Blinder-Oaxaca Unexplained Component Using 2017 ACS Characteristics

$$UV_{2017} = E\left(X_{S_{2017}}\right)' \beta_{S_{2010}} - E\left(X_{S_{2017}}\right)' \beta_{U_{2010}}$$

	2017 ACS – 2010 Census
All other household model ($\beta_{U_{2010}}$)	-19.9
AR & ACS all-citizen household model ($\beta_{S_{2010}}$)	-11.9
Difference-in-differences	-8.0

N=755,000 households

Blinder-Oaxaca Decomposition: Robustness

- Try 227 variables from entire ACS, in addition to 39 in base specification, to estimate the all-citizen household model
- 3 versions of Least Absolute Shrinkage and Selection Operator (lasso) procedure
  - EBIC information criterion (149 variables selected)
  - cross-validation method (157 variables selected)
  - AIC information criterion (157 variables selected)
- Principal Components Analysis (PCA) using top 20, 50, and 100 factors
- Run Blinder-Oaxaca Decomposition with the selected variables in 2010
- 6.3-6.4 ppts unexplained with lasso, 7.0-7.2 unexplained with PCA

Effect on Overall Self-Response Rate

- Apply 8.0 ppt drop to 28.1% of housing units potentially having at least one noncitizen (estimated in 2017 ACS)
- Results in 2.2 ppt drop in housing unit self-response
- At a cost of $55 million per ppt, this would mean an increase in NRFU fieldwork costs of $121 million

Caveats

- Assumes self-response rate of all-citizen households will be unaffected by citizenship question
- Some households in group potentially containing at least one noncitizen likely contain only citizens, which may understate the citizenship question effect on households actually containing at least one noncitizen
- Does not capture change in degree of sensitivity to citizenship question since 2010

Conclusions

- Households potentially containing at least one noncitizen have a 11.9 ppt larger drop-off in self-response to the 2010 ACS vs. the 2010 Census compared to all-citizen households
- 6.3-8.8 ppt of the difference-in-differences is unexplained, which we attribute to sensitivity to the ACS citizenship question
- We estimate a 2.2 ppt overall drop in self-response, increasing NRFU cost by $121 million and lowering quality

Ideas for Future Research

- Randomized Control Trials
  - Measure effect of citizenship question on all-citizen household unit self-response rate
  - Effect of citizenship question on net undercount

- Comparisons of citizenship information across multiple administrative sources

- How to combine data sources to produce “best” citizenship variable

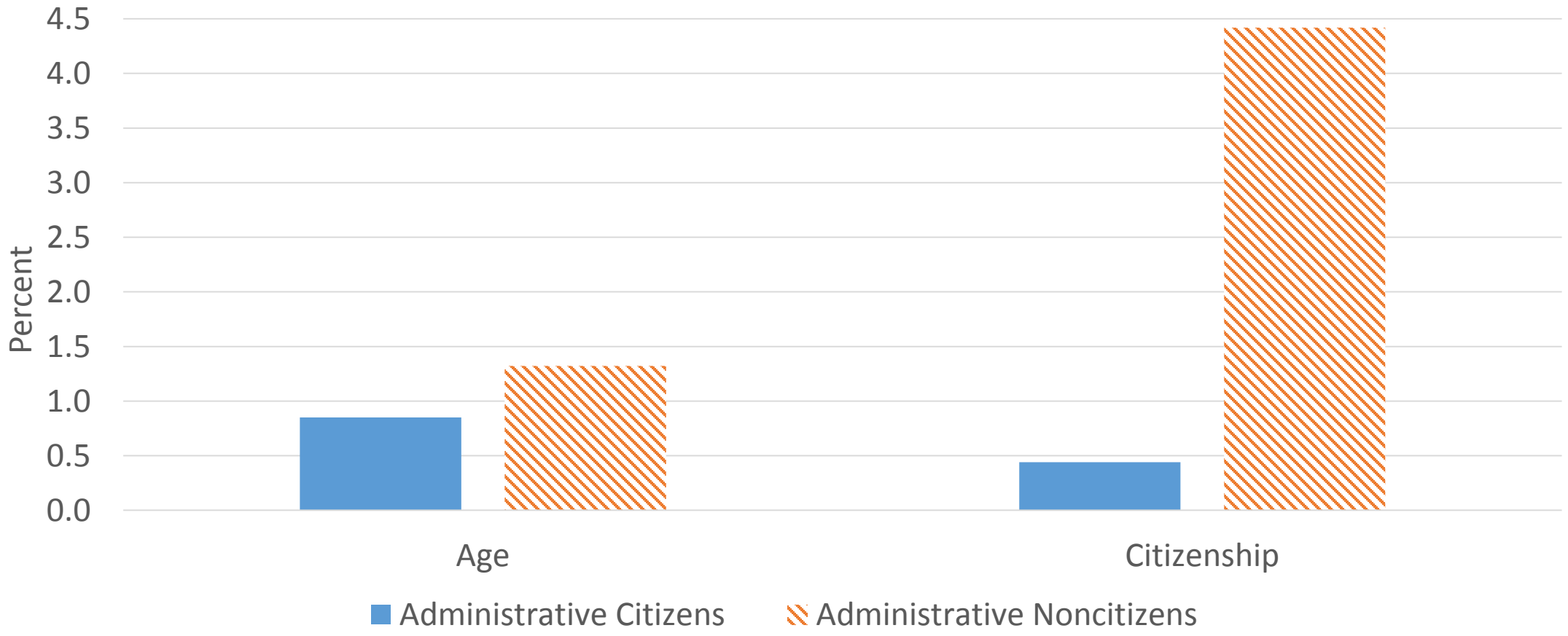
Misty L. Heggeness
U.S. Census Bureau
Research and Methodology Directorate

e-mail: misty.l.heggeness@census.gov

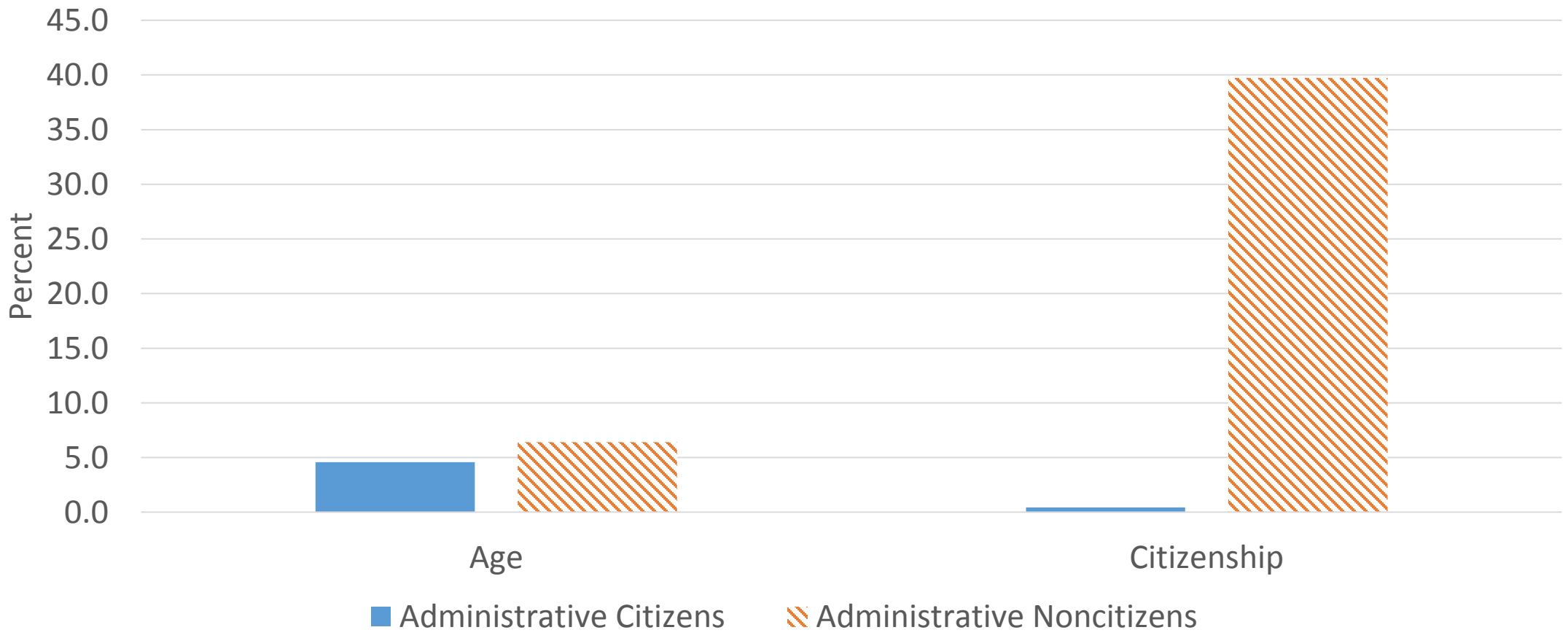
On  : @m_heggeness

Website: https://sites.google.com/view/misty-l-heggeness/home

2017 ACS Item Nonresponse: Administrative Record Citizens and Noncitizens



2017 ACS-Administrative Record Disagreement: Administrative Record Citizens and Noncitizens



Explaining Administrative Record Noncitizen Item Nonresponse and Discrepant Response

- Respondent misunderstands the question
  - more discrepancies when linguistically isolated, in self-response
- Respondent doesn't know person's status
  - more nonresponse and discrepancies with nonrelatives, little difference between noncitizens and citizens
- Respondent has privacy concerns
  - more nonresponse and discrepancies among noncitizens relative to citizens
- Incorrect linkage to administrative records
  - more discrepancies with lower-quality linkage, little difference between noncitizens and citizens
- Administrative data are incorrect (missing naturalizations)
  - more discrepancies when reporting about self, mode shouldn't matter

Item Nonresponse Regressions

- $Item_{Cj} = X'_{Cj}\beta_{Cj} + \varepsilon_{Cj}$
- $Item_{NCj} = X'_{NCj}\beta_{NCj} + \varepsilon_{NCj}$
- *Item j* = age, and citizenship in 2017 ACS
- *X* includes relationship to householder, race/ethnicity, working or search for a job, linguistic isolation, linkage quality, self-response vs. fieldwork, education, household income, share of households in block group with at least one noncitizen, share of households in block group below poverty line
- Sample size:
  - 4,108,000 for administrative record Citizens
  - 253,000 for administrative record noncitizens

Age and Citizenship Status Disagreement Regressions

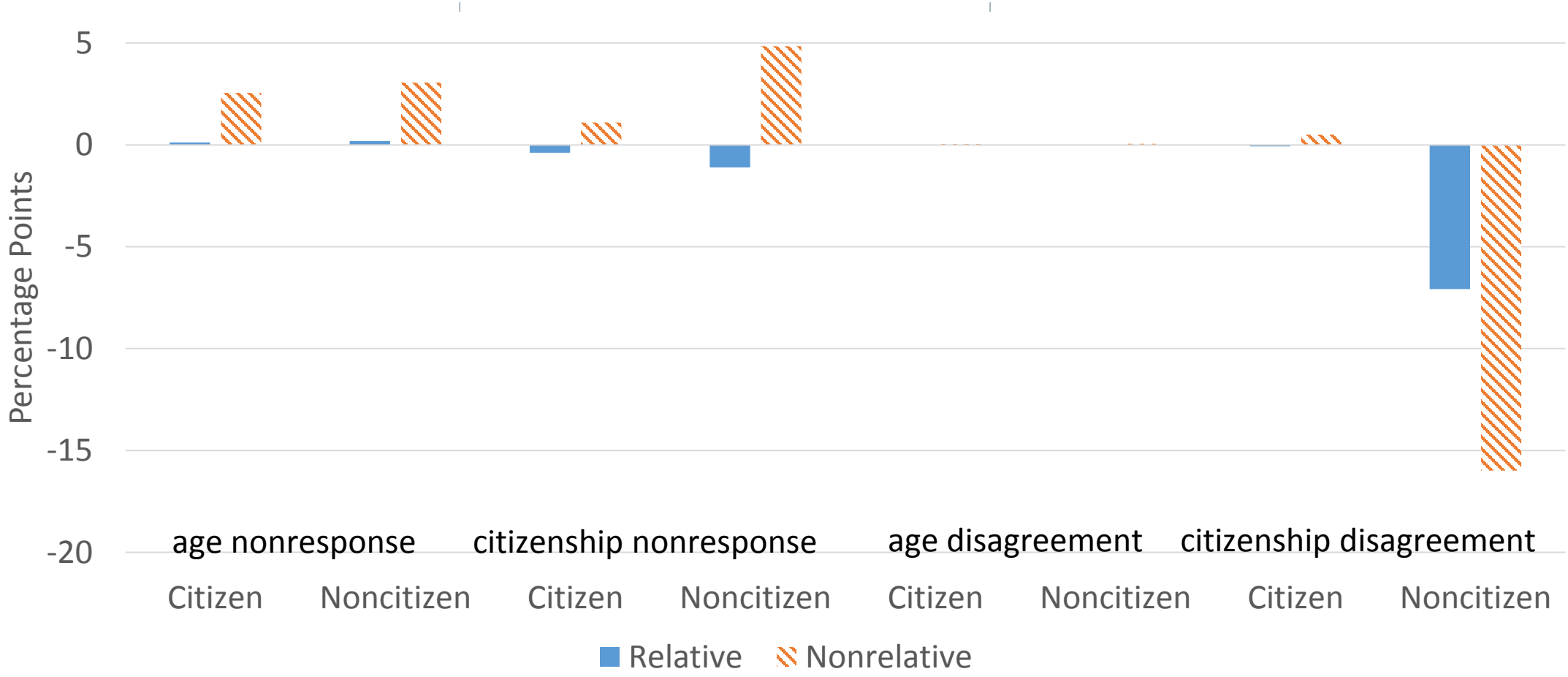
- $Disagree_k = X_k'\beta_k + \varepsilon_k$

- *k* = admin. citizen-ACS noncitizen, admin. noncitizen-ACS citizen

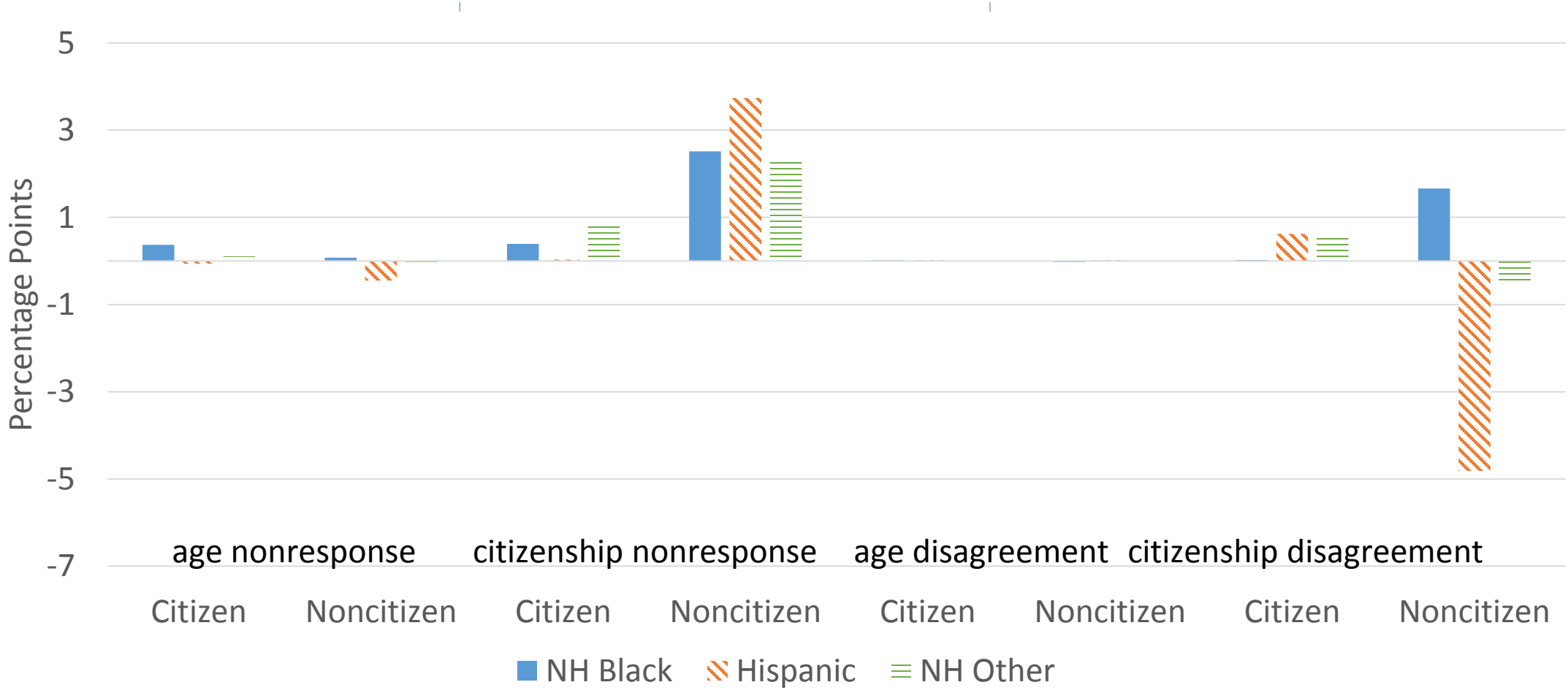
- *X* includes relationship to householder, race/ethnicity, working or search for a job, linguistic isolation, linkage quality, self-response vs. fieldwork, education, household income, share of households in block group with at least one noncitizen, share of households in block group below poverty line

- Sample size:
  - 4,060,000 for administrative record citizen age disagreement regression
  - 249,000 for administrative record noncitizen age disagreement regression
  - 3,872,000 for administrative record citizen – ACS noncitizen regression
  - 229,000 for administrative record noncitizen – ACS citizen regression

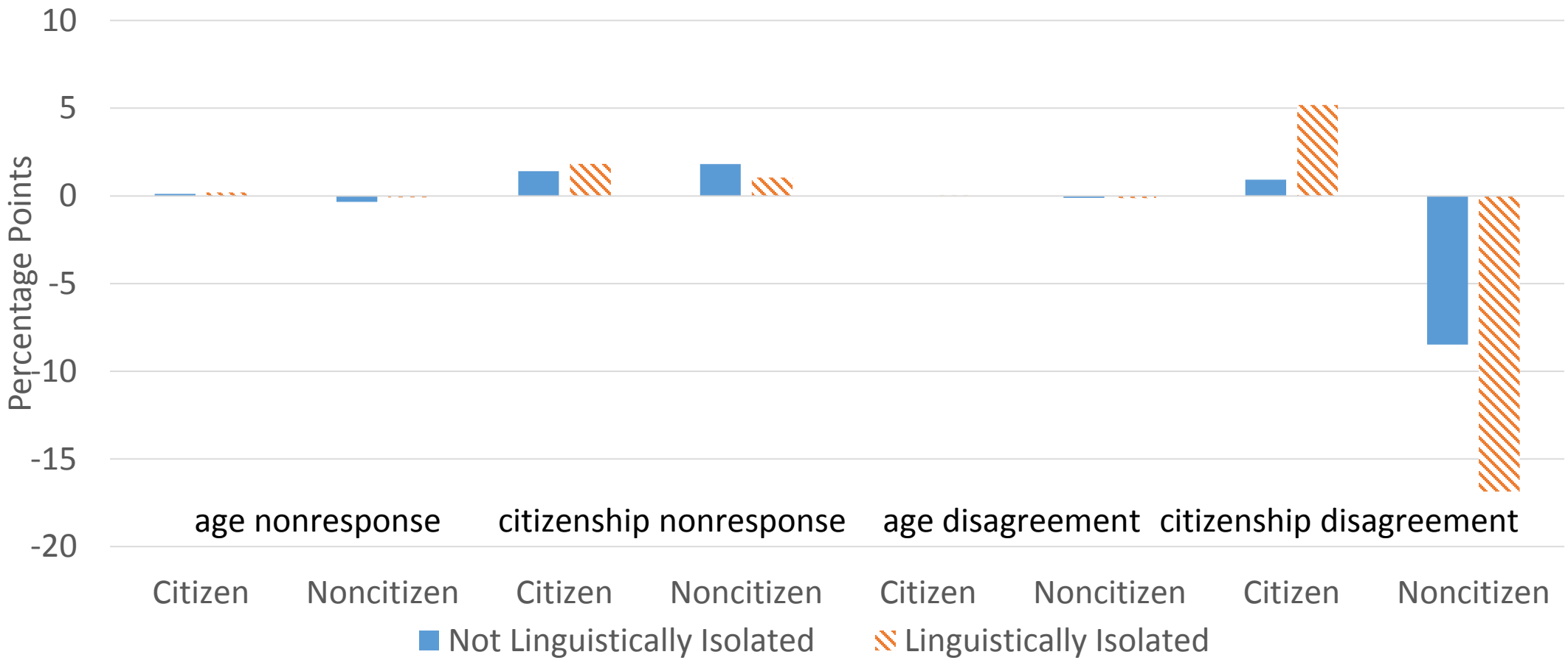
Relatives and Nonrelatives vs. Respondent



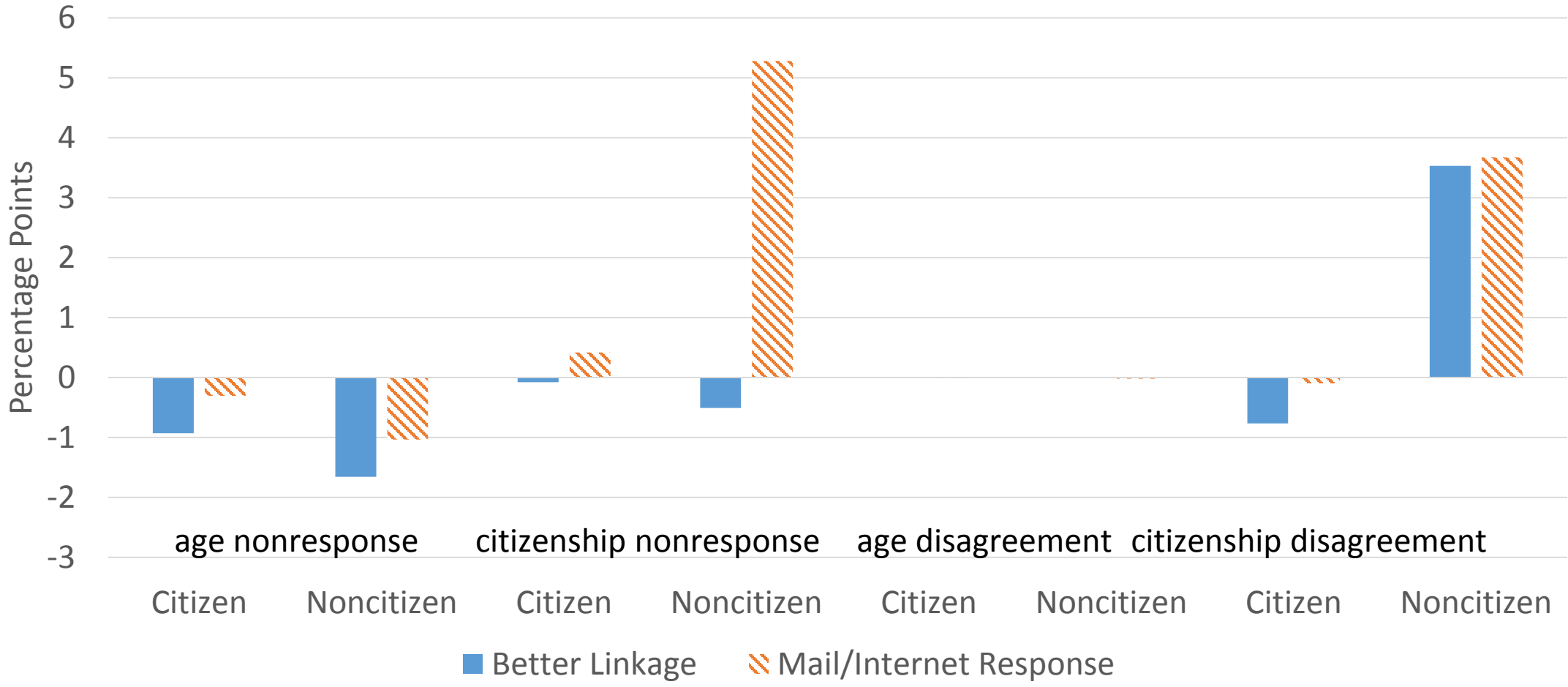
Race/Ethnicity vs. Non-Hispanic White



Linguistic Isolation



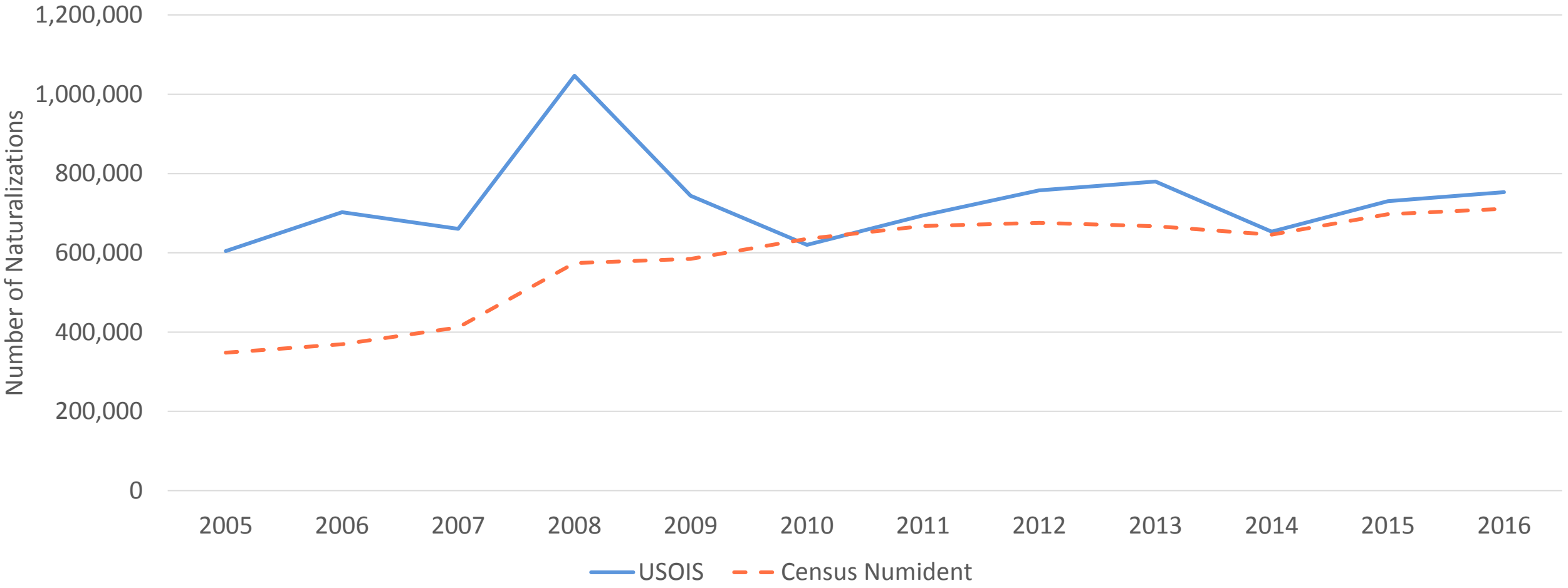
Better Linkage, Mail/Internet Response



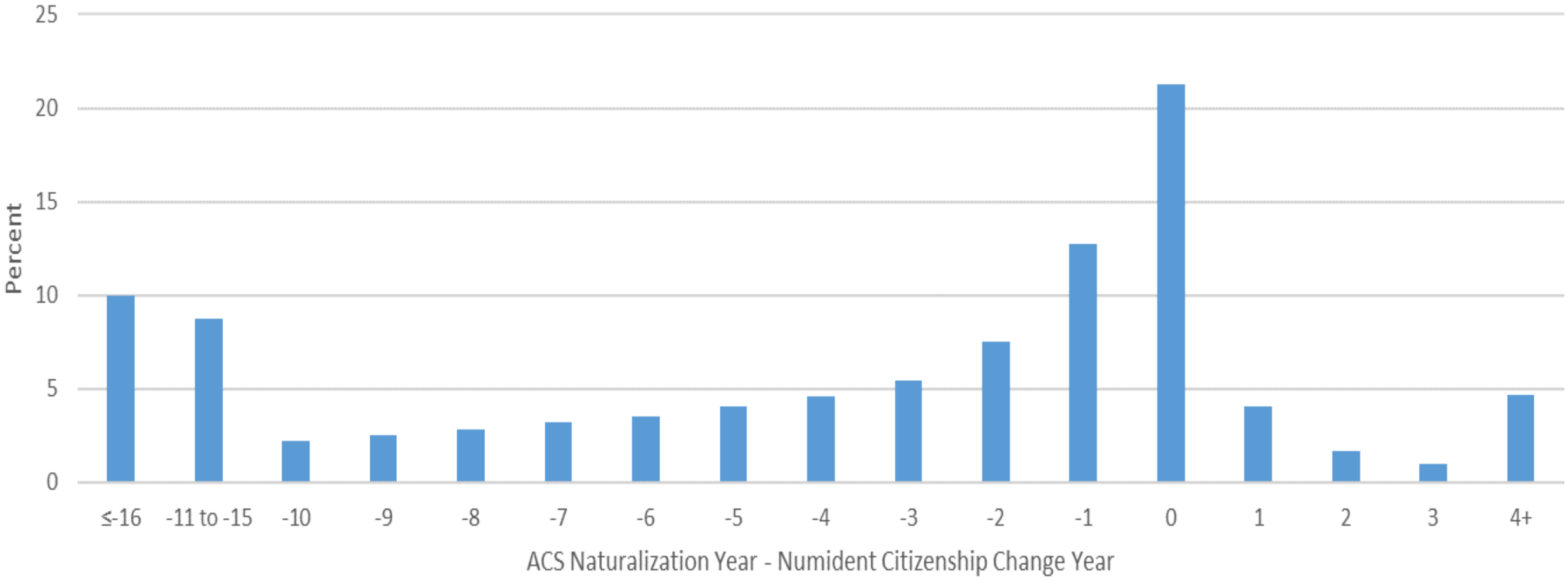
Blinder-Oaxaca Decomposition of Differences in Problematic Response to Citizenship and Age Questions by Administrative Record Citizenship Status

	Problematic Response Rate (%)		Difference
	Citizenship	Age	
AR Noncitizens	44.6	8.0	36.6
	(0.15)	(0.07)	(0.17)
AR Citizens	5.9	5.8	0.1
	(0.03)	(0.02)	(0.04)
Difference-in-differences			36.5
			(0.08)
Explained			-1.0
			(0.04)
Unexplained			37.4
			(0.09)

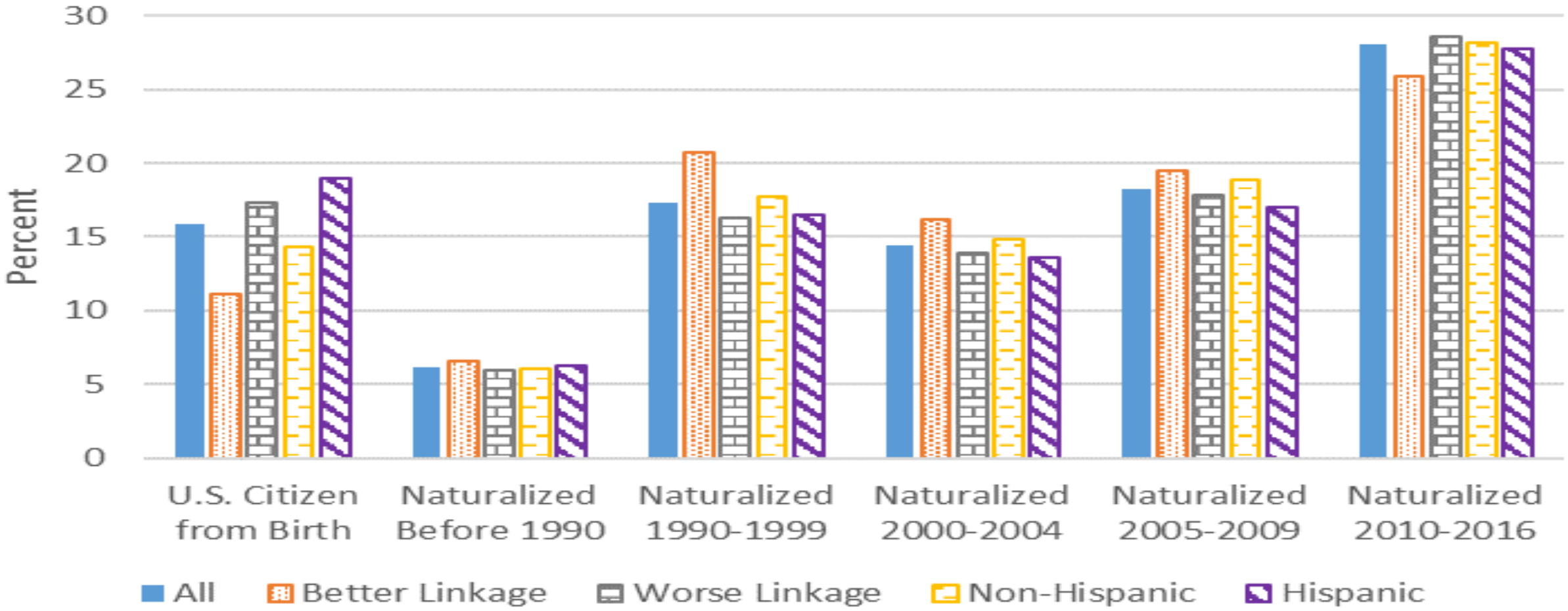
Estimated Annual Naturalizations in 2017 Numident vs. USOIS Statistics



Difference between 2016 ACS Naturalization and Numident Citizenship Change Years



Distribution of 2016 ACS Citizenship Receipt Timing for Administrative Record Noncitizen-ACS Citizens by Linkage Quality and Ethnicity



Enumeration Quality in Mailout/Mailback and Nonresponse Follow-up (NRFU) Proxy Responses

	Mailout/Mailback Response	NRFU Proxy
Correct Enumerations	97.3	70.2
Erroneous Enumerations	2.5	6.7
Whole-Person Census Imputations	0.3	23.1
Person Linkage Rate	96.7	33.8

$55 million estimated fieldwork cost for each percentage point drop in self-response rate