

Consistency of Race and Ethnicity Reporting in Two National Health Surveys

Jacqueline Wilson Lucas, M.P.H.

Health Statistician, Division of Health Interview Statistics

Jennifer H. Madans, Ph.D.

Associate Director for Science, National Center for Health Statistics



Background

- Women age 15-44 from households in the 1993 NHIS were selected for the 1995 NSFG (N=10,847 women)
- Opportunity to examine a variety of interesting research questions because of the breadth of information available in the combined data set
- **Methodological issues surrounding reporting of information in the two surveys also possible**



Research question

- How consistent are respondents to these two surveys in reporting their race and ethnicity?
 - *Are reports of Hispanic origin consistent across the two surveys?*
 - *Are reports of race – including multiple race reporting – consistent across the surveys?*
 - *If there are inconsistencies, what factors might help to explain them?*



Questionnaire comparison

1993 NHIS

- *“Other” offered as race response category*
- *Race information by proxy and self-report*
- *Detailed API and AIAN race groups subgroups on hand card (no specific order)*
- *Only 2 race mentions keyed in data file*
- *Hispanic origin in HH comp (front); race in demographic background (end of core)*
- *Hispanic origin asked before race*
- *Very little imputation done (crude for race)*
- *Race asked in two-part question*
- *General health content*

1995 NSFG

- *“Other” not offered as race response category*
- *Race information by self-report only*
- *Detailed subgroups not offered - only 4 OMB race groups (in alphabetical order)*
- *Up to 4 race mentions recorded*
- *Hispanic origin and race in demographic section (near end of ques)*
- *Hispanic origin asked before race*
- *Imputation done for all variables with missing data*
- *Race asked in two part question*
- *Focus on women’s reproductive health issues*



Methods

- Link between 1993 NHIS core in-house file and 1995 NSFG public use respondent file created (Lucas, Zhen)*
- Variables from NHIS and NSFG included in data set
- Bivariate and multivariate analyses run (logistic regression)
- NSFG weight used for SUDAAN

* Non-public use data; available through RDC



Variables used in analysis

- race (coded as single and multiple race mentions in both surveys)
- Hispanic origin
- marital status
- educational attainment
- age
- income
- region of residence
- MSA residence
- US/Foreign born status
- respondent reporting status (self report whole or partly used in analysis)
- residence in a mono-racial or multiracial household



Key definitions

- ***Self report:***
 - In the NHIS, defined using the respondent status variable (*self entirely or partially/not self-report*)
 - In the NSFG, **all information is self-reported**
- ***Level of agreement on race, ethnicity measures:***
 - Any match on single or multiple race mentions, DK or refused responses was considered **agreement** on race responses
 - Any response that did not match - including refusal, DK, “other” on NHIS - classified as **no agreement** on race responses
 - All persons identified as Hispanic on both surveys classified as agreement; otherwise no agreement



Comparison of race and ethnicity measures

	1993 NHIS	1995 NSFG
<i>Single race reporting</i>	97.9 (0.20)	98.8 (0.15)*
<i>Multiple race reporting</i>	1.6 (0.18)	1.0 (0.14)*
<i>Hispanic origin reporting</i>	10.7 (0.65)	11.2 (0.66)
<i>Any mention of White</i>	79.9 (0.72)	80.9 (0.75)
<i>Any mention of Black</i>	13.5 (0.65)	13.8 (0.64)
<i>Any mention of AIAN</i>	1.8 (0.20)	2.9 (0.26)*
<i>Any mention of API</i>	3.3 (0.40)	3.3 (0.38)
<i>Any mention of Other</i>	2.4 (0.65)	-- na --



How did NHIS respondents who reported their race as “Other” report race on the NSFG

<i>White</i>	68.1 (4.31)
<i>AIAN</i>	16.5 (3.22)
<i>API</i>	6.2 (1.84)
<i>Black</i>	5.7 (1.61)
<i>White/AIAN</i>	2.2 (1.04)



Summary measure of agreement

Race measure

92.8 (0.39)

7.2 (0.39)

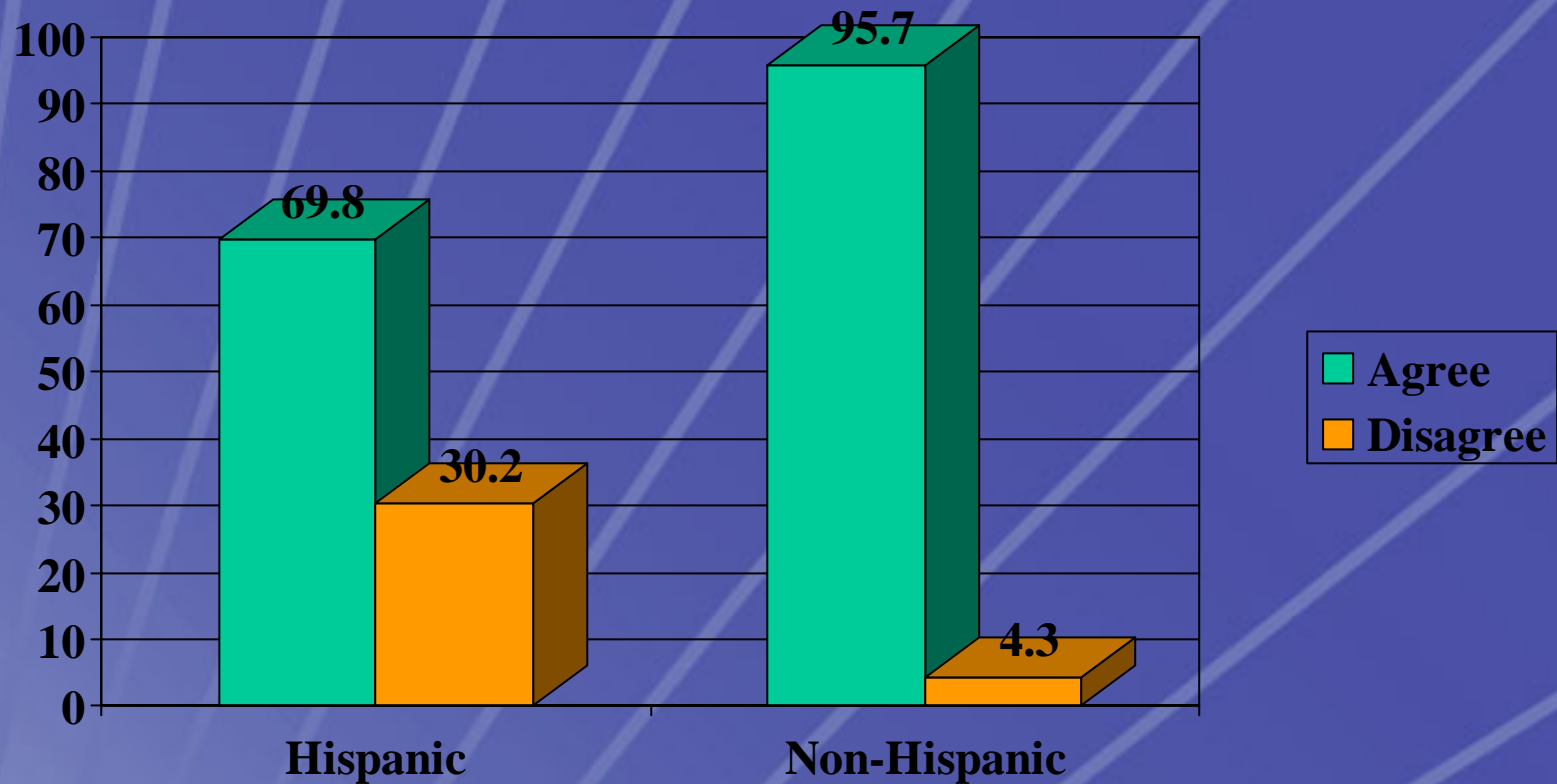
*Hispanic origin
measure*

98.7 (0.15)

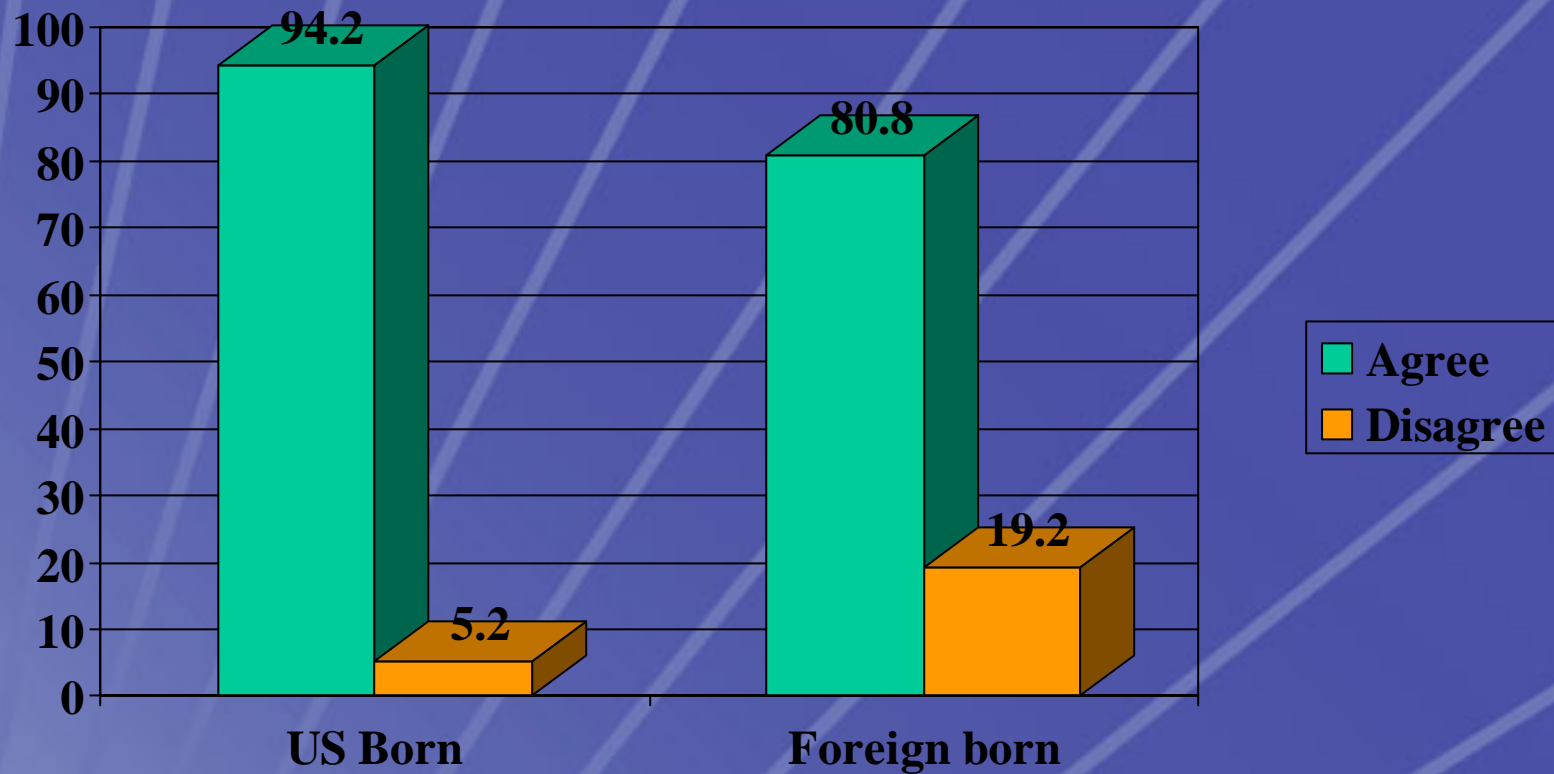
1.3 (0.15)



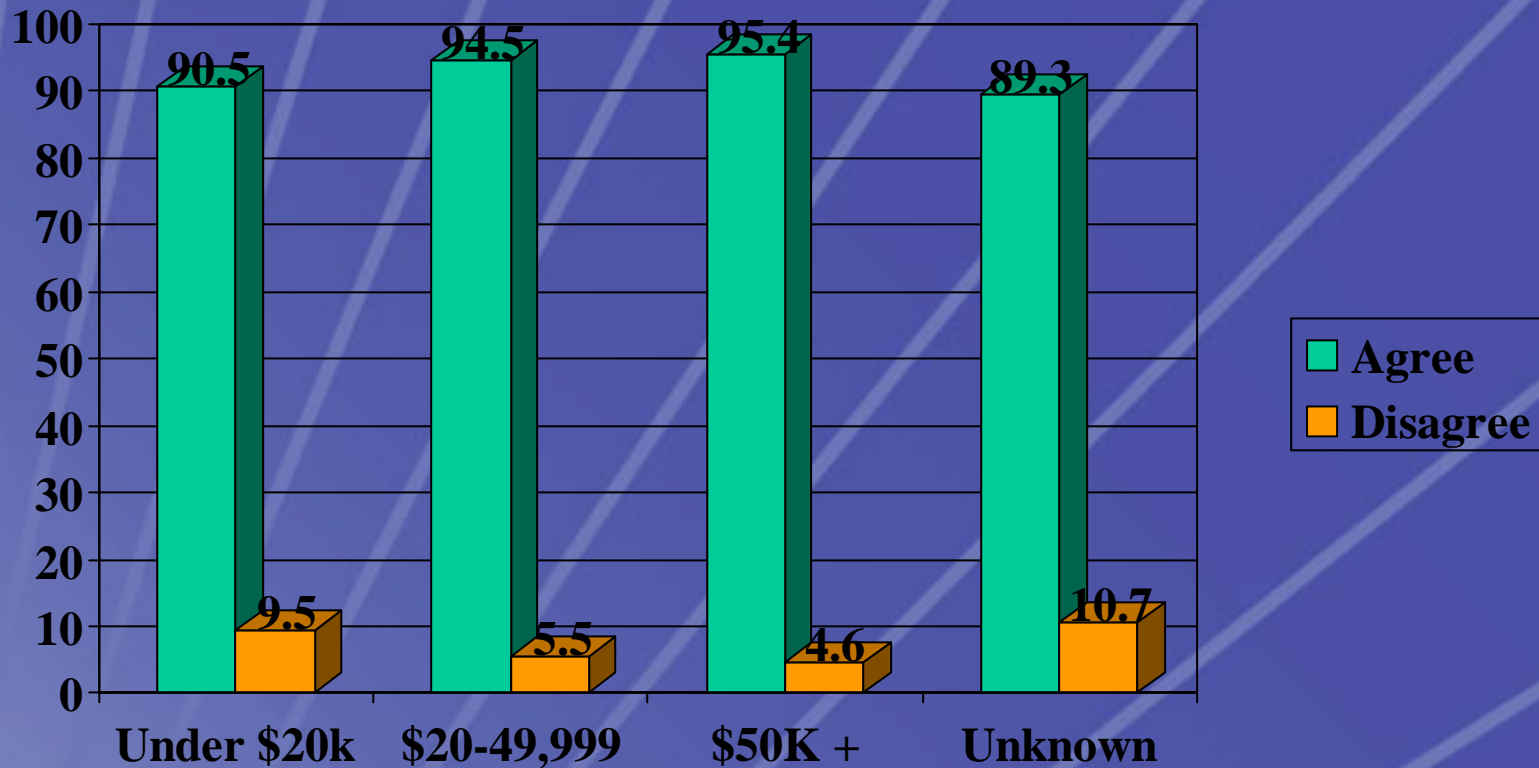
Agreement of race variables by Hispanic origin



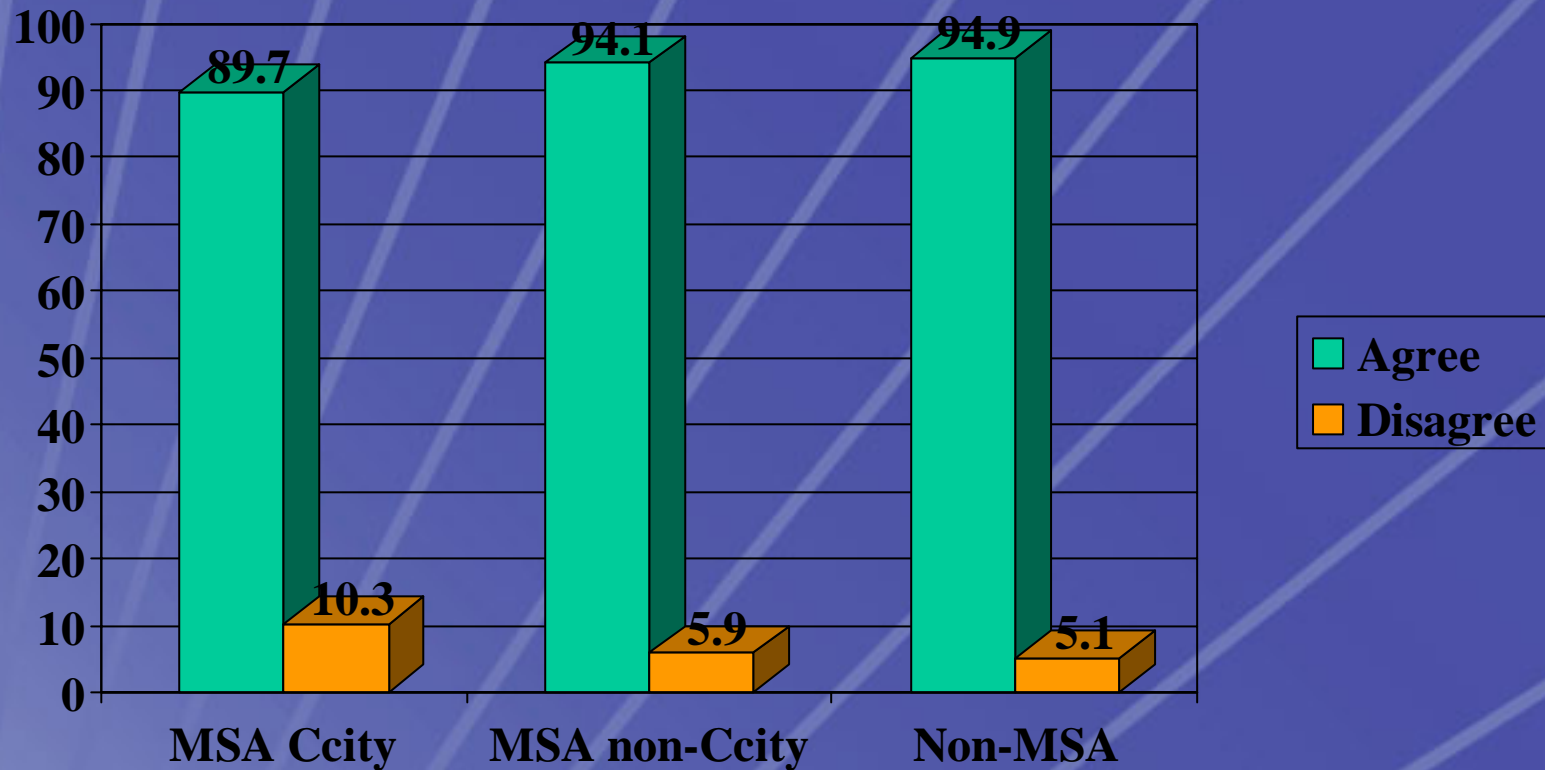
Agreement of race variables by nativity



Agreement of race variables by family income



Agreement of race variables by MSA



Model specifications

	<i>All women</i>	<i>Single race women</i>	<i>Multiracial women</i>
<i>Age</i>	YES	YES	YES
<i>Hispanic origin</i>	YES	YES	YES
<i>Nativity</i>	YES	YES	YES
<i>Marital status</i>	YES	YES	YES
<i>Any mention AIAN</i>	NO	NO	YES
<i>Respondent status</i>	YES	YES	YES
<i>Education</i>	YES	YES	YES
<i>Poverty level income</i>	YES	YES	YES
<i>Region of residence</i>	YES	YES	YES
<i>MSA residence</i>	YES	YES	YES
<i>Household race status</i>	NO	NO	YES



Odds of disagreement on race

	<i>All women</i>	<i>Single race women</i>	<i>Multiracial women</i>
<i>Age 30-34</i>	0.71**	0.72**	NS
<i>Hispanic</i>	6.74**	11.77***	SIG****
<i>Foreign born</i>	1.63**	1.60**	NS
<i>Never married</i>	1.25**	NS	NS
<i>Any mention AIAN</i>	NA	NA	SIG****
<i>Proxy report of race</i>	NS	1.32**	NS
<i>Less than HS education</i>	1.35**	1.48**	HSGRAD/SIG****
<i>Poverty level income</i>	NS	NS	NS
<i>South/Midwest</i>	0.66/0.75**	NE/1.39**	NS
<i>MSA Central City</i>	NS	1.83**	NS
<i>Multiracial household</i>	NA	5.03***	NS

** p=.05 *** p=.001 **** data not reliable



Summary of findings

- Overall proportions reporting Hispanic origin and single and multiple races were similar in the two surveys
- Agreement between survey responses for people was high for Hispanic origin, lower for race
- Adjusting for other factors, Hispanic origin, nativity , marital status, education, and region of residence all play a role in the consistency of race responses



Implications of the data

- The understanding of and ability to report race remains a problem for persons of Hispanic origin
- High level of agreement on Hispanic origin reporting and significantly lower level of agreement on race reporting - particularly multiple race reporting - raises some important questions about measuring multiple race groups



Limitations of the data

- Many aspects of the questionnaires are quite different, including content, categories offered, order of responses, etc. - this may have a huge impact on the level of agreement of responses
- Very small sample sizes limit ability to look at data in detail
- Analysis limited to women – are there sex differences in the consistency of race reporting?
- Other characteristics that might influence race reporting (interviewer characteristics, community level characteristics)– not included in these analyses (future work)

