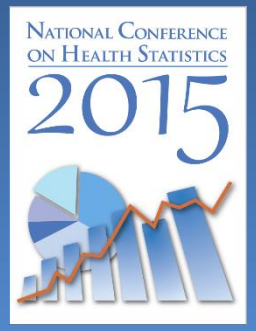
Differential Linkage by Race/Ethnicity and Availability of a Social Security Number in the Linkage with the National Death Index

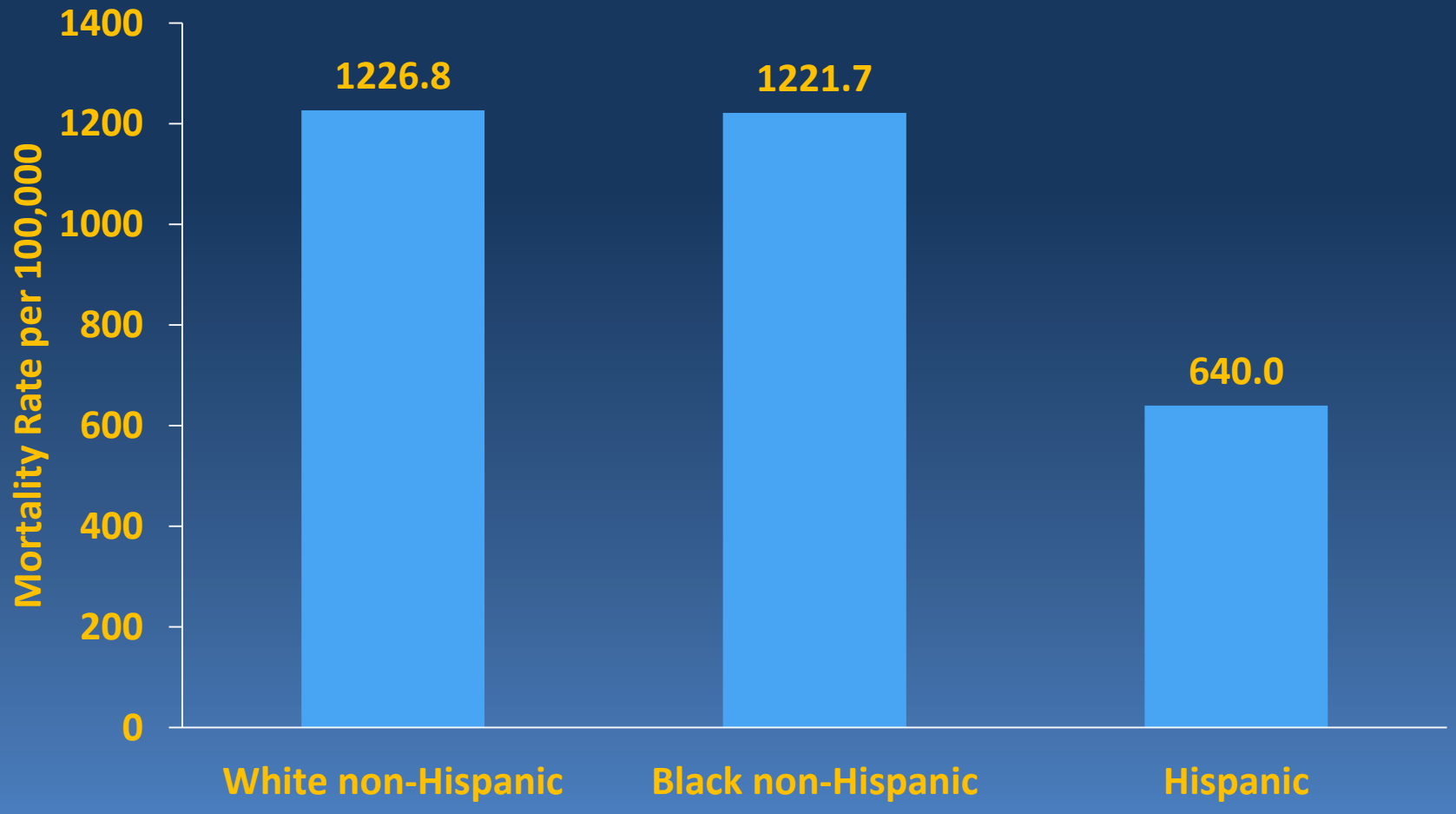
Eric A. Miller[1*,2], Frances McCarty[2], Jennifer D. Parker[2]

[1]ICF International
[2]National Center for Health Statistics

*Work conducted while at NCHS



Mortality Rate Among Adults Aged 25 Years and Over by Race/Ethnicity: NHIS Linked Mortality Files (1997-2004)



Based on data from Borrell et al. AJPH 2012; 102: 836-843

Hispanic Paradox*

Despite having a higher risk profile, Hispanics have been found to have lower mortality rates compared to non-Hispanic whites

If true, this has significant public health implications

- More resources devoted to studying the paradox
- Resources diverted to focus on other disparities / issues

*Markides and Coreil (1986). Public Health Reports; 101: 253-265

Potential Reasons for Paradox

- Health selective immigration
- Salmon bias (return migration)
- Advantageous health behaviors and social support

Potential Reasons for Paradox Cont'd

Health selective immigration

Salmon bias (return migration)

Advantageous health behaviors and social support

Data quality / Insufficient linkage

Reasons for Differential Linkage

Naming conventions for Hispanics differ from other US populations

- Use of mother's and father's surname
- May not have single middle name

Less likely to have Social Security Number

- Especially among older adults and foreign born
- Without a SSN, may be more difficult to link than white or black non-Hispanic participants

Importantly, this could apply to other race and ethnic groups in the US (e.g. Asian population)

SSN Availability in the National Health Interview Survey (NHIS)

There have been a number of changes to the collection of SSN in the NHIS since 2002

- No longer collected for everyone in the household (currently just sampled adults and children)

As a result, many more NHIS participants are being linked without a SSN

- Unclear how availability differs by race/ethnicty

Objectives

Compare the availability of a SSN for linkage by race/ethnicity

Estimate the percentage of links missed due to survey participants not having or providing a SSN (overall and by race/ethnicity)

Secondarily, if there are differences do they contribute to mortality disparities (results not presented)

Study Population

1992-2009 NHIS data linked with the National Death Index (NDI - follow-up through December 31, 2011)

Adults aged 65 years and over at the time of survey

- With sufficient identifiable information provided
- Did not refuse to have data linked
- Source of death was NDI

Analysis Methods

Analyses conducted in SUDAAN to account for complex survey design

Calculated weighted distributions by race/ethnicity

- Hispanic, Asian/Pacific Islander, black non-Hispanic, white non-Hispanic
- Chi square tests to examine differences

Estimated missed links using preventive fraction methods*

*Kleinbaum, DG, Sullivan KM, Barker ND. Active Epi Companion Textbook (2013)

Analysis Methods Continued

Estimating missed links with preventive fraction

Considered *not having* a SSN as protective against linking (i.e. death)

Modeled the survival of participants with and without a SSN adjusting for age, sex, education level, US nativity, respondent reported health status, and survey year

- The reduced risk of linking among those without a SSN can be translated into the percent of links missed (1-RR)

Models were run overall and stratified by race/ethnicity to compare missed links across groups

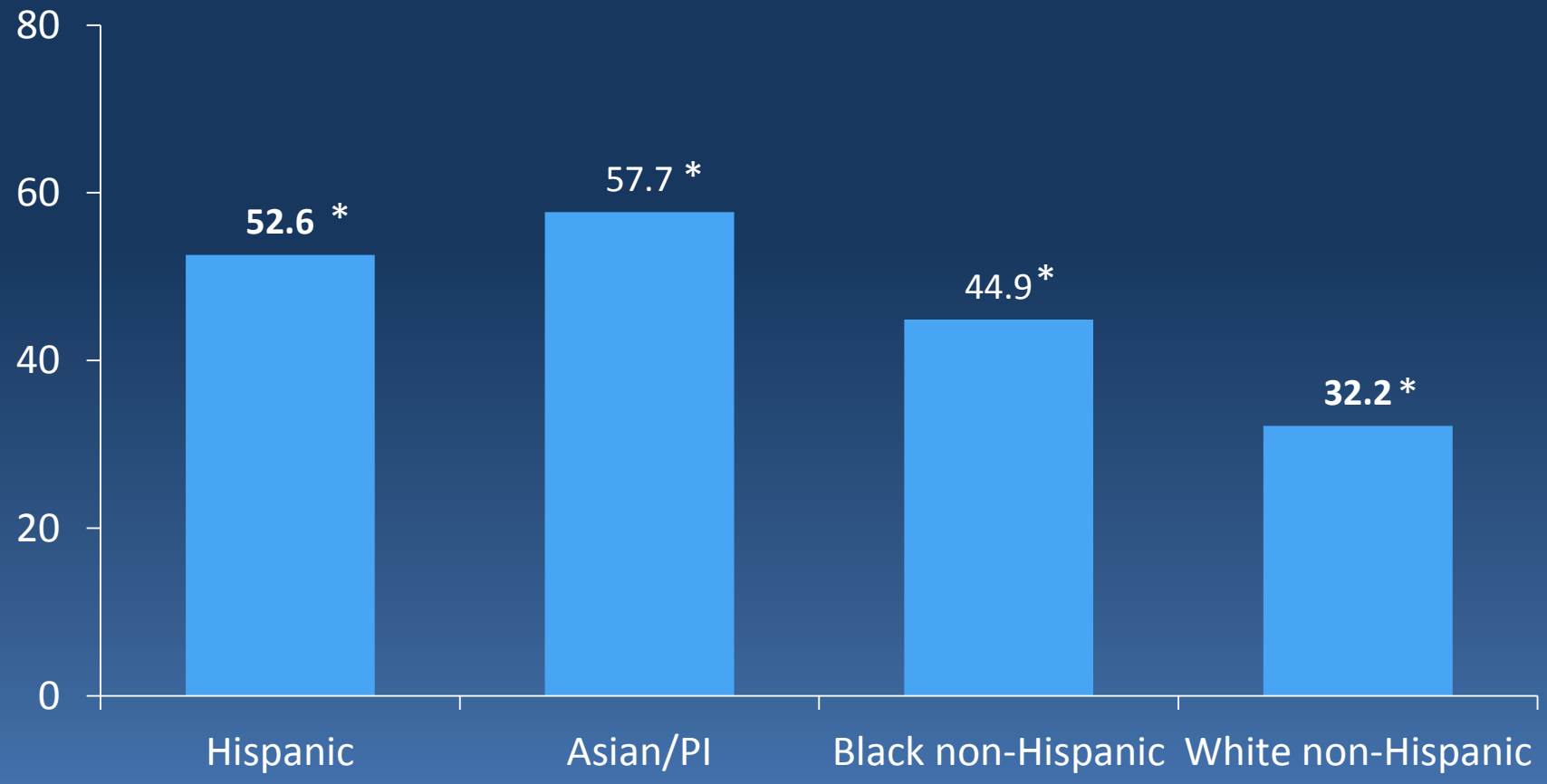
Study Population – 1992-2009 NHIS

Adults aged 65 years and over at the time of survey.

Race/Ethnicity	n (%)	Percent Linked
Hispanic	17,193 (5.2)	35.6
Asian/PI	4,590 (2.0)	27.7
Black non-Hispanic	24,529 (8.3)	49.2
White non-Hispanic	174,428 (84.5)	50.1

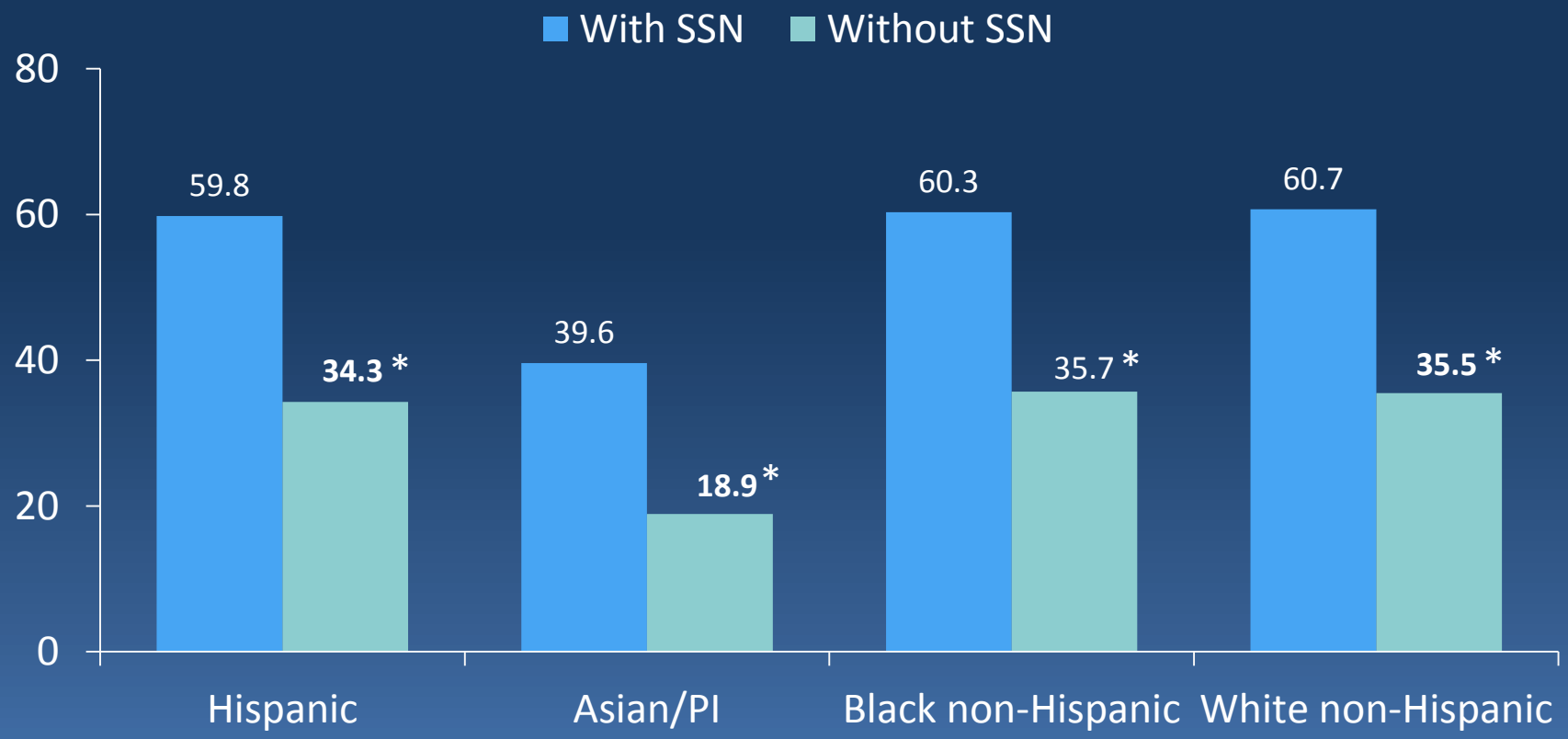
Significant differences by race/ethnicity for all demographic characteristics: age, sex, US nativity, education level. Also significant differences by respondent rated health and survey year (due to changes in oversampling)

Percent Without SSN



*Significantly different from white non-Hispanic ($p<0.05$)

Percent Linked by SSN



*Significantly different from available SSN ($p<0.05$)

Estimated Percentages of Missed Links

Race/Ethnicity	Estimated Percent of Links Missed Among Those Without a SSN* (95% CI)**	Estimated Percent of Links Missed in Group Overall*** (95% CI)**
Overall	24.8 (23.6 – 25.9)	10.7 (10.2 – 11.2)
Hispanic	34.4 (30.4 – 38.7)	18.1 (15.9 – 20.4)
Asian/PI	32.0 (21.9 – 41.0)	18.5 (12.7 – 23.9)
Black non-Hispanic	29.8 (26.5 – 33.0)	13.4 (11.9 – 14.8)
White non-Hispanic	23.5 (22.1 – 24.8)	9.9 (9.3 – 10.5)

*Calculated as 1-HR

**95% Confidence intervals (CI) calculated using bootstrap methods

***Calculated as the prevalence of participants without a SSN multiplied by (1-HR)

Summary

A non-trivial percentage of links are missed among participants without a SSN

A greater percentage are missed for Hispanic, Asian/PI and black non-Hispanic participants compared to white non-Hispanic participants

However, in a separate analysis, we found it had only a modest impact on mortality ratios compared to white non-Hispanic participants

Implications

These issues are not unique to this linkage

Researchers should consider data limitations when presenting disparities in mortality (especially with linked data)

Efforts need to be (are being) made to improve linkage in these populations

“Walk through the Valley”

Thanks!

Dean Judson
Jim Brittain
Keith Zevallos
Jesse Bassich
Cordell Golden
Eileen Call
Susan Wilson
Hannah Day
Judith Weissman
Patsy Lloyd
Deborah Ingram
Makram Talih

Contact Info:
Eric A. Miller eric.miller@icfi.com
or
datalinkage@cdc.gov

NDI Matching Algorithm

Social Security Number
First name
Middle initial
Last name
Month of birth
Day of birth
Year of birth
Sex
Father’s surname
State of birth
Race
State of residence
Marital Status

Conclusions

Other factors likely contributing to mortality disparity

- Names and naming conventions
- Unknown exact dates of birth

- Salmon Bias or longevity cannot be ruled out

Mortality Ratios After Correcting for Missed Links

Race/Ethnicity	Observed Percent Linked	Observed Ratio	Corrected Percent Linked	Corrected Ratio
Hispanic	35.6	0.71	42.1	0.77
Asian/PI	27.7	0.55	32.8	0.60
Black non-Hispanic	49.2	0.98	56.0	1.02
White non-Hispanic	50.1	1.0 (Ref)	54.7	1.0 (Ref)