



National Center for Health Statistics

Data Linkage

Record and Geographic Linkages to Inform Health Disparities

Jennifer Parker and Lauren Rossen
Office of Analysis and Epidemiology

National Center for Health Statistics
Office of Analysis and Epidemiology



Outline

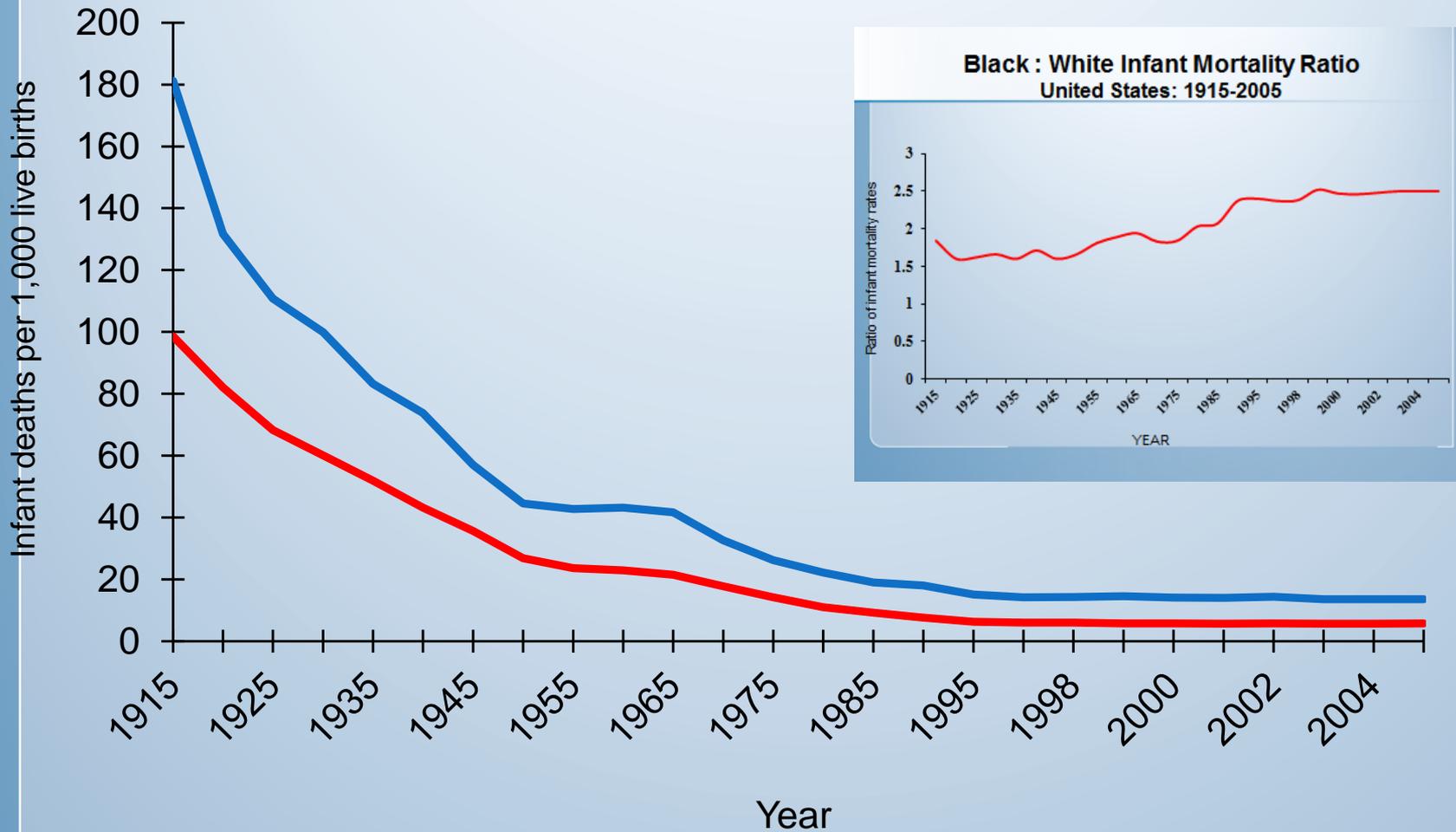
- Why combine data?
- Linked birth and infant death files
- NCHS record linkage program
 - Linked mortality files
 - Linked Medicaid and Medicare data
- Geographically linked data
- Access to NCHS linked data
- Summary

Why link data for disparities?

- Possibly better ascertainment of race and ethnicity information
- Additional information on
 - socioeconomic related variables
 - contextual exposures
 - urban-rural status
 - outcomes or intermediate endpoints related to outcomes

Infant Mortality Rates by Race United States, 1915-2005

— Black — White



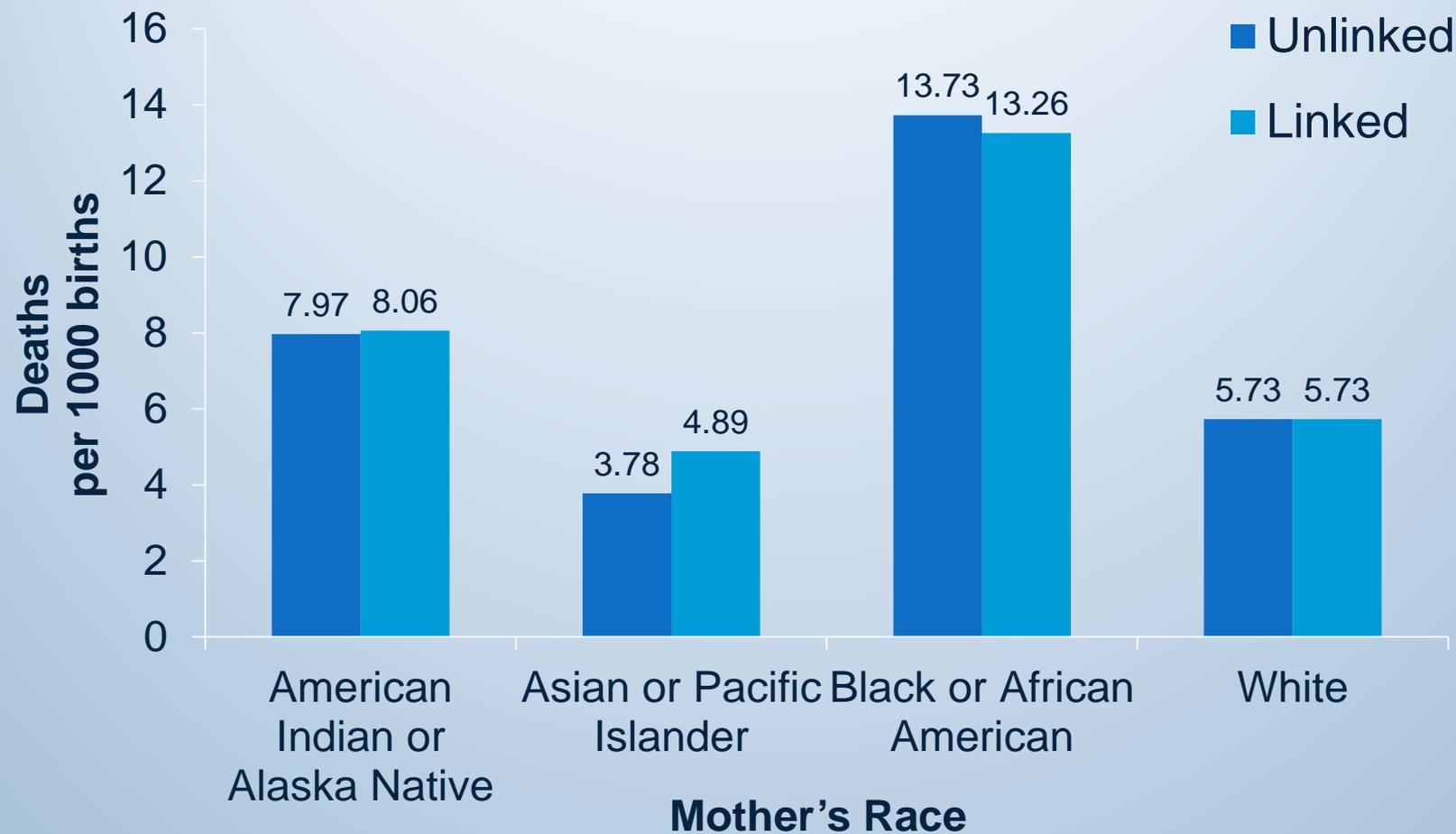
NCHS Linked Birth and Infant Death files

Why link infant death records to infant birth records?

Birth records add:

- Race of mother versus race of infant
 - Better information for smaller race and ethnicity groups
- Mother's demographic information (e.g. education, marital status)
- Infant and maternal health information (e.g. birth weight, gestational diabetes)

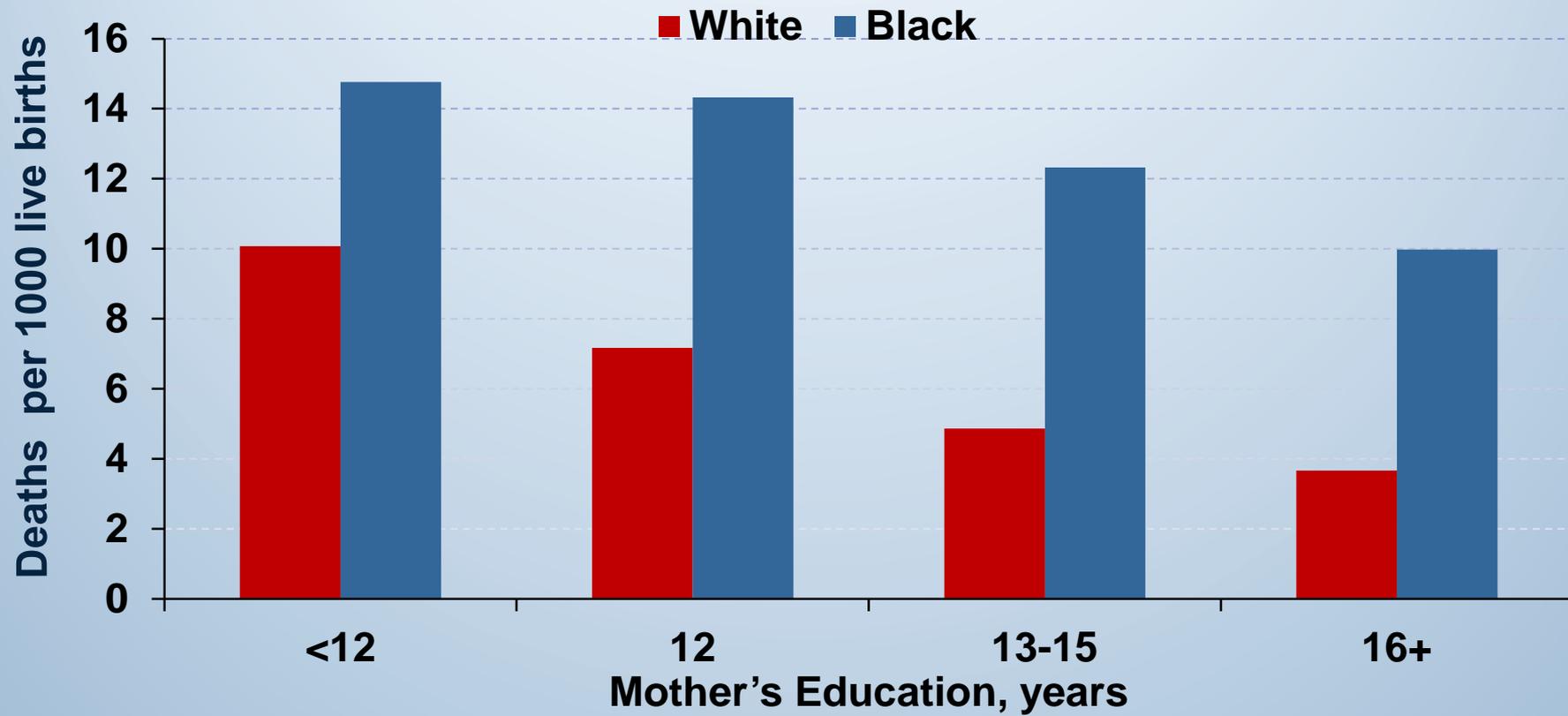
2005 Infant Mortality Rates by Mother's Race: Unlinked and Linked Births



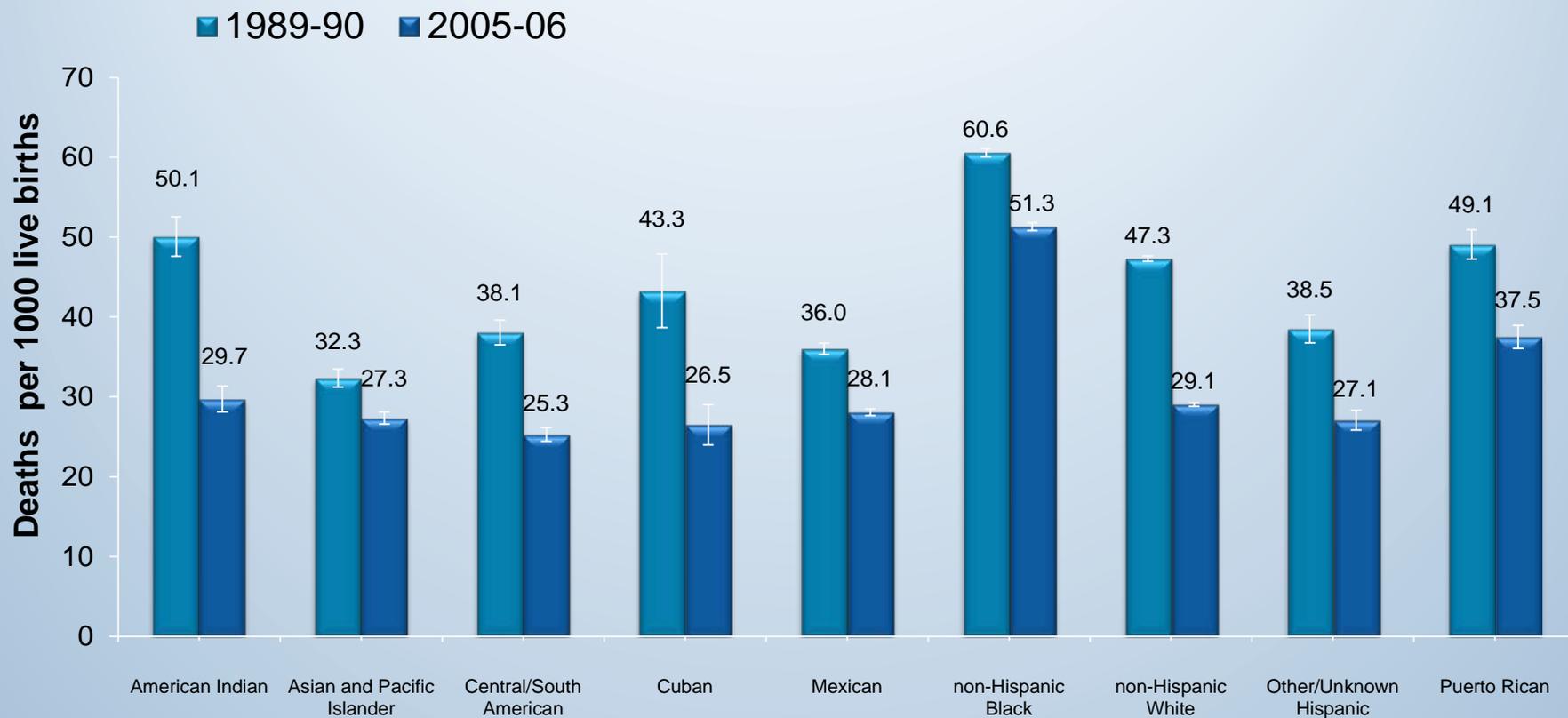
SOURCE: CDC Wonder

Infant mortality by maternal education and race

United States, 2005



Pre-term (<37 weeks) Infant Mortality rates, 1989-2006



NCHS Record Linkages

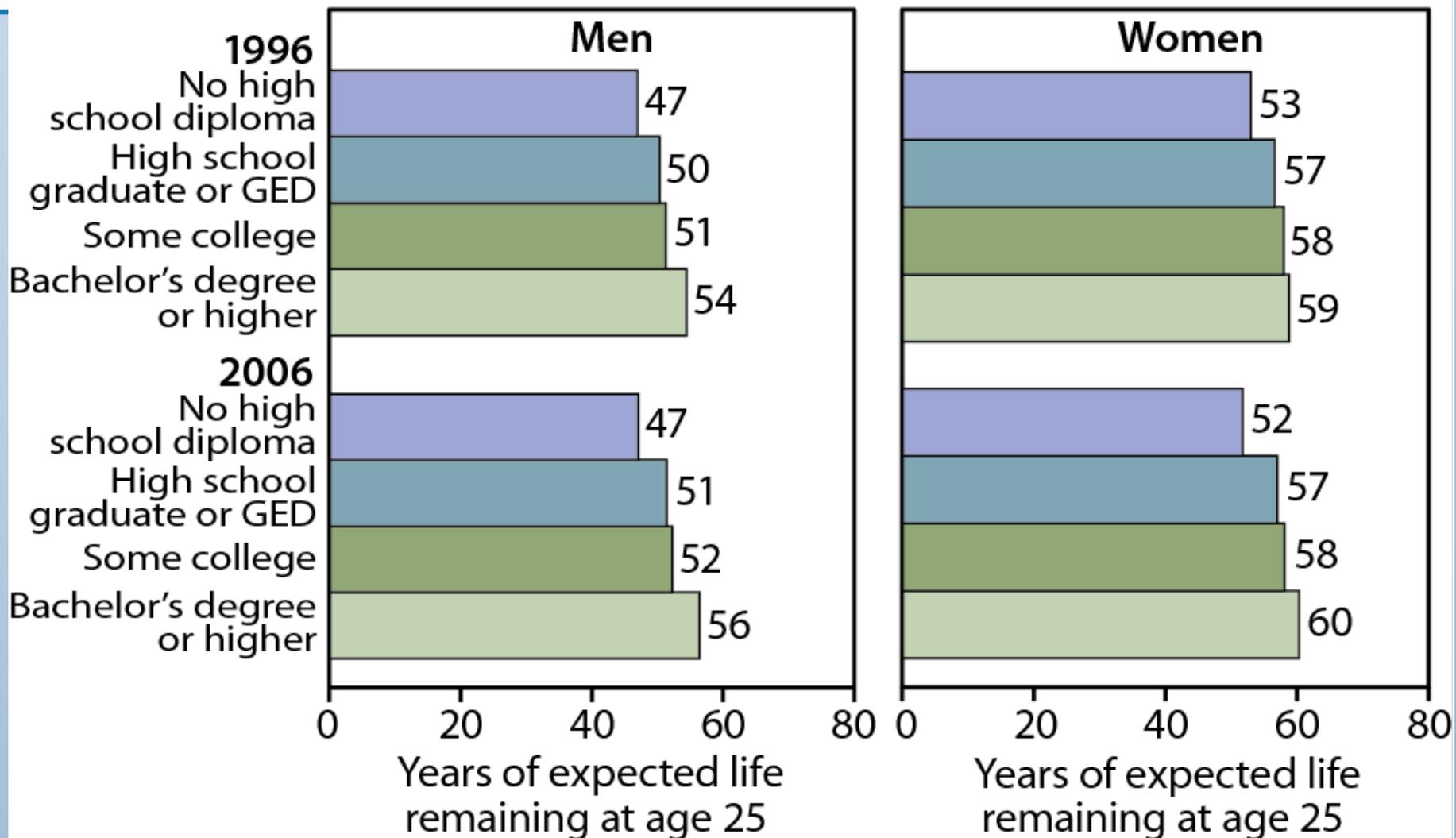
- NCHS record linkage program links survey data to administrative records using confidential personal identifying information (e.g. names, Social Security Numbers, dates)
- Administrative records
 - Mortality
 - Centers for Medicare and Medicaid Services (CMS)
 - Social Security Administration
 - Pilot projects (e.g. Florida Cancer Data System, Texas Supplemental Food and Nutrition Program)

NCHS Record Linkages

Why link survey data to administrative records?

- Survey data
 - better detail on race, ethnicity, socioeconomic indices (education, income), baseline health status, self-reported program participation
- Administrative records
 - program participation (e.g. Medicaid, SNAP)
 - costs of medical care, benefits
 - longitudinal data adds ability examining health prior to or after the survey
 - mortality and cause of death

Life expectancy at age 25

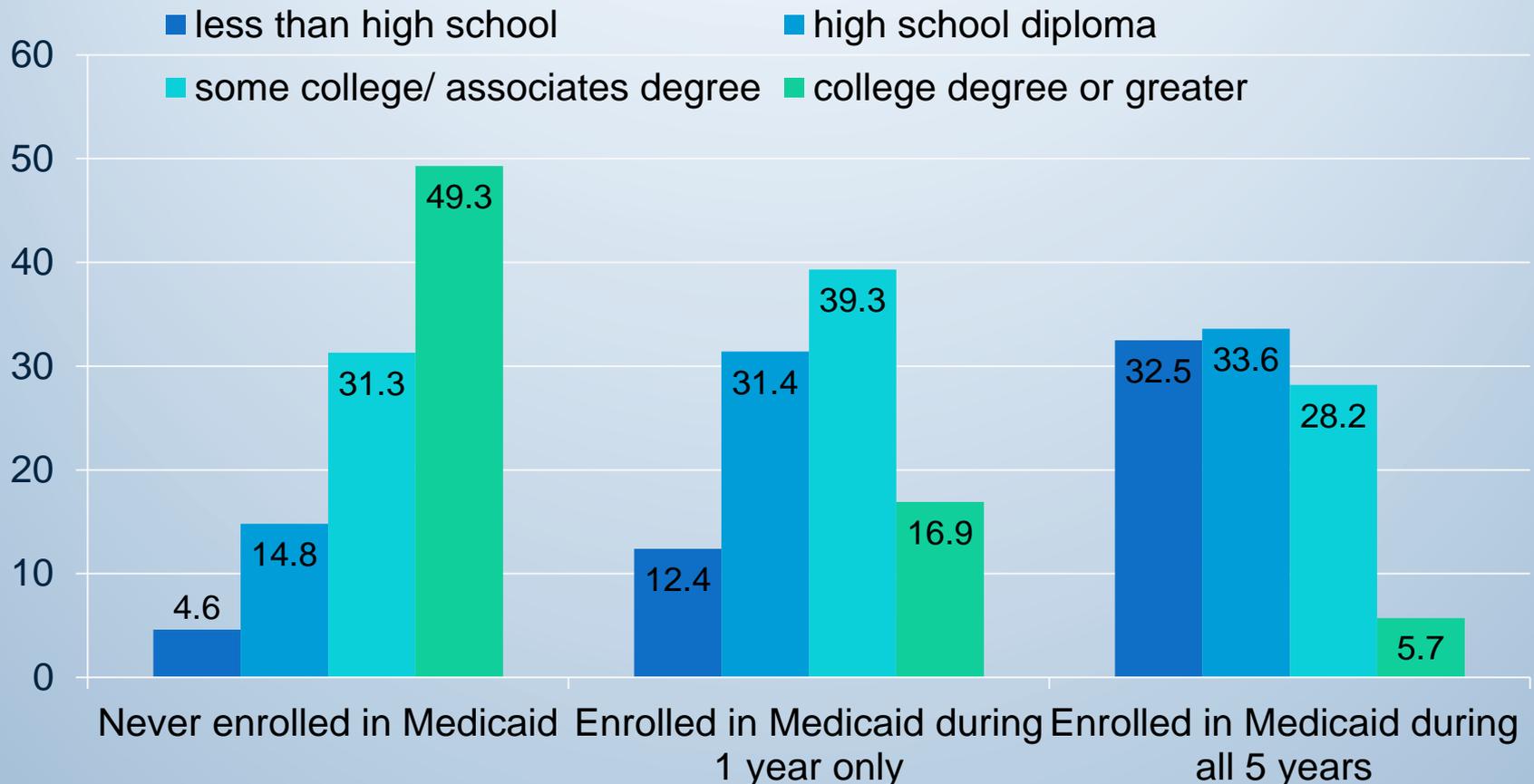


NOTE: GED is General Educational Development high school equivalency diploma.

SOURCE: CDC/NCHS, *Health, United States, 2011*, Figure 32. Data from the National Health Interview Survey Linked Mortality File.

Percent distribution of parental education by children's enrollment in Medicaid during a 5 year period

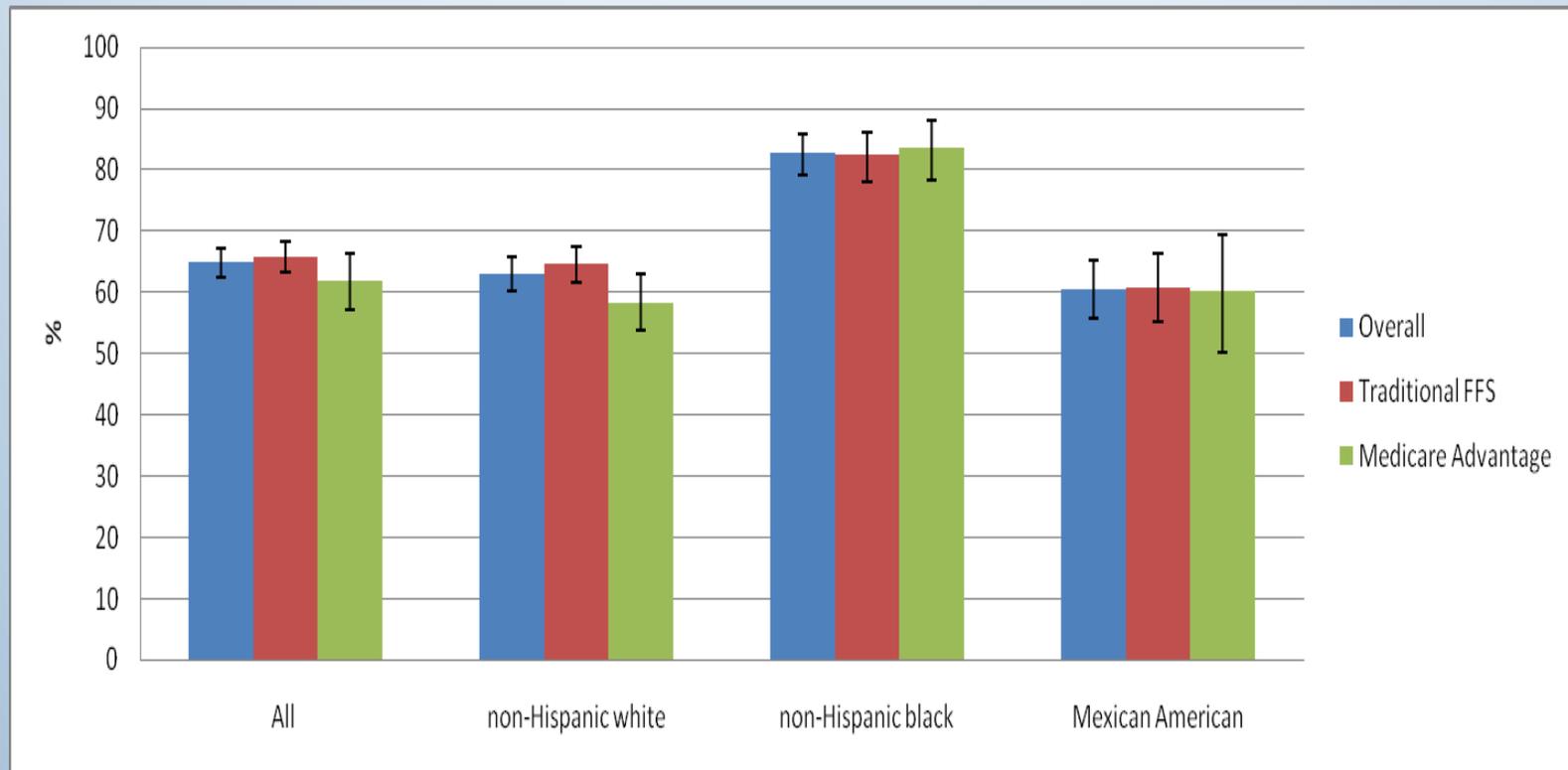
2004 NHIS linked to 2004-2008 Medicaid



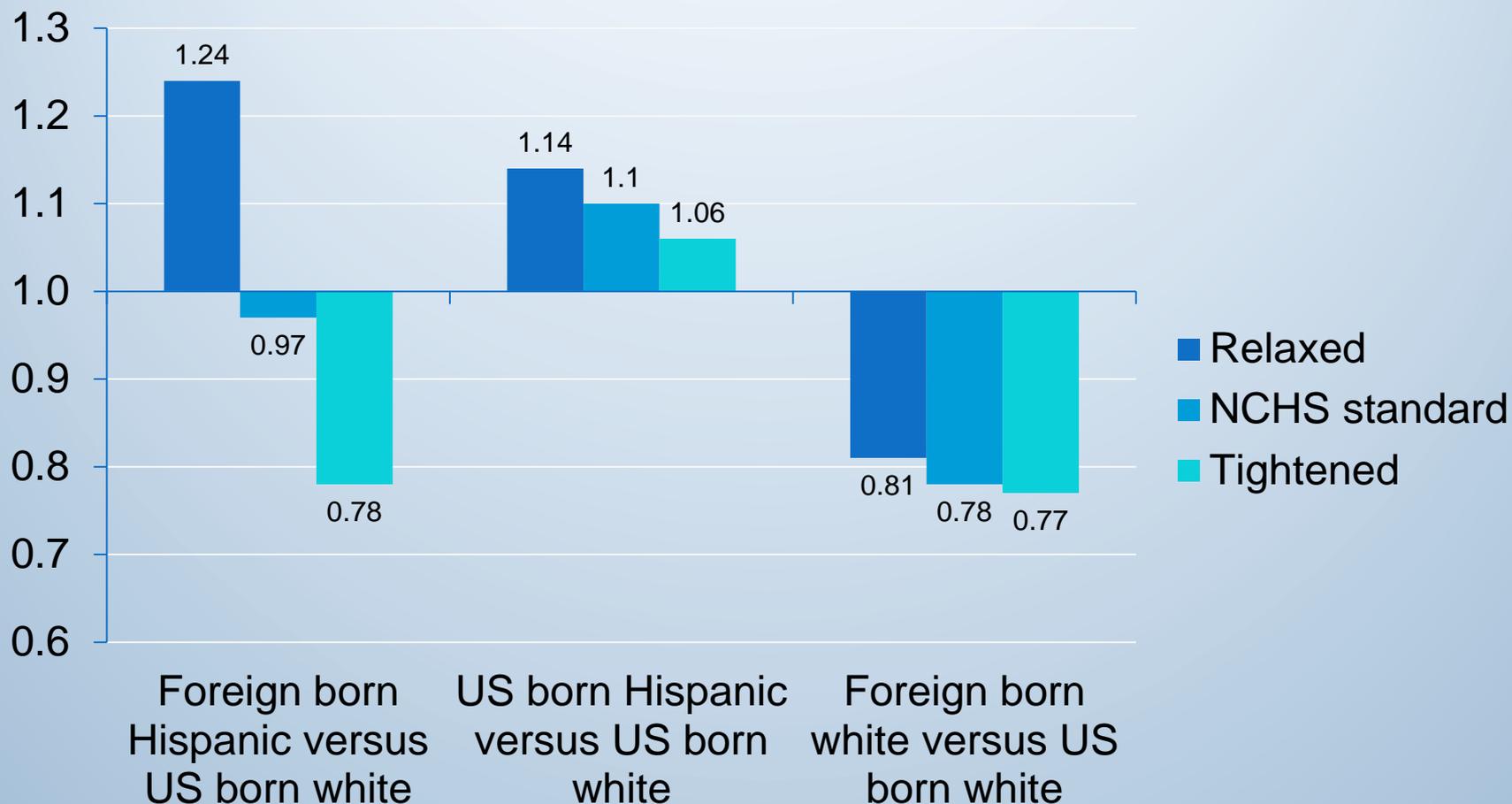
SOURCE: Simon et al, preliminary results

Percent hypertension by Medicare enrollment (fee for service or Medicare Advantage) by race/ethnicity

1999-2004 NHANES linked to 2007 Medicare



Association between ethnicity and mortality using 3 linkage criteria

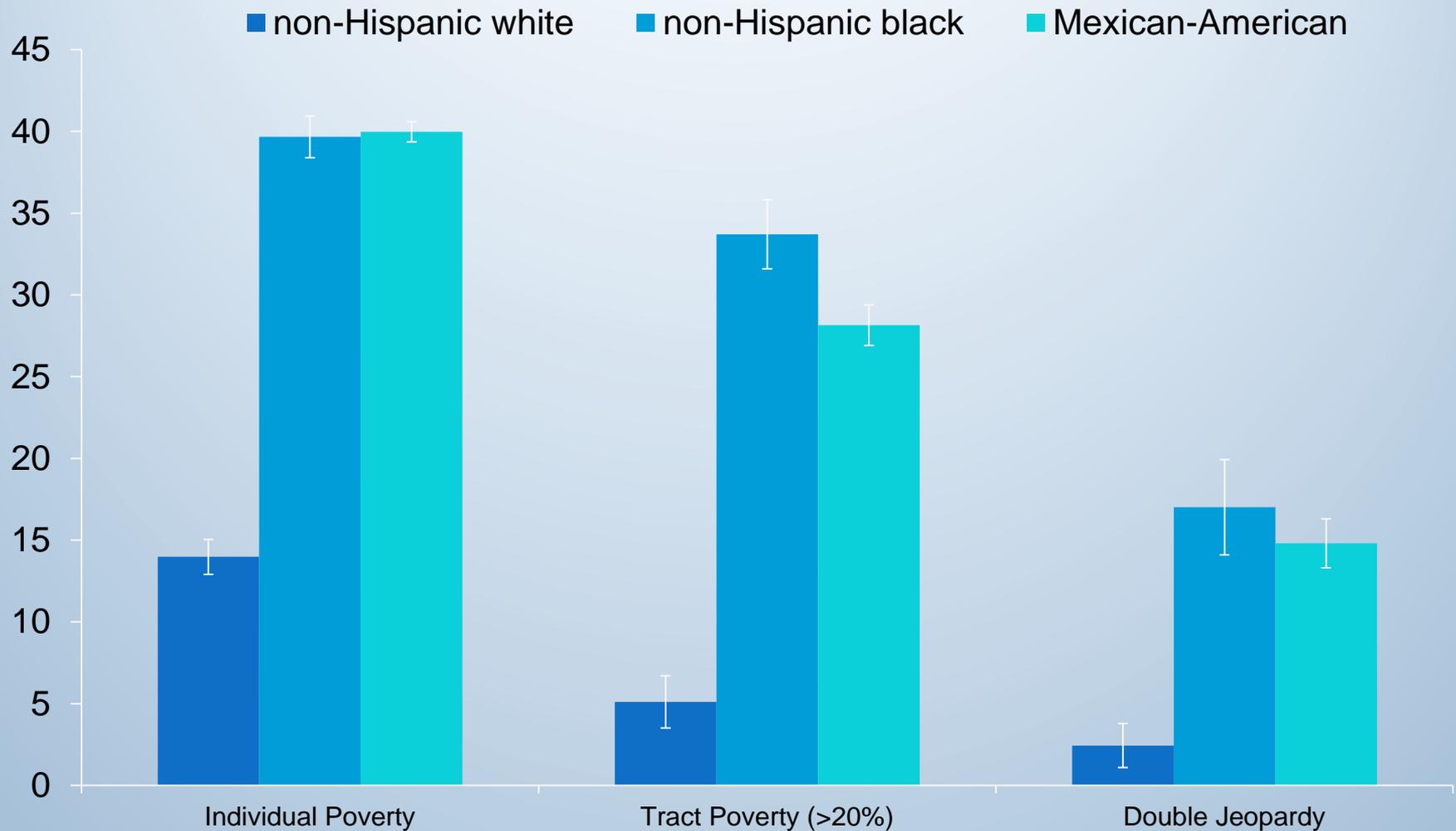


Geographic linkages

Why link survey data to geographic data?

- Measures of
 - contextual SES indices (median income, % poverty)
 - exposures (e.g. pollution, liquor stores)
 - access (e.g. Community Health Centers)
- Very limited geographic detail on public use files
 - External data can be merged by administrative units or using GIS methods (Research Data Center)

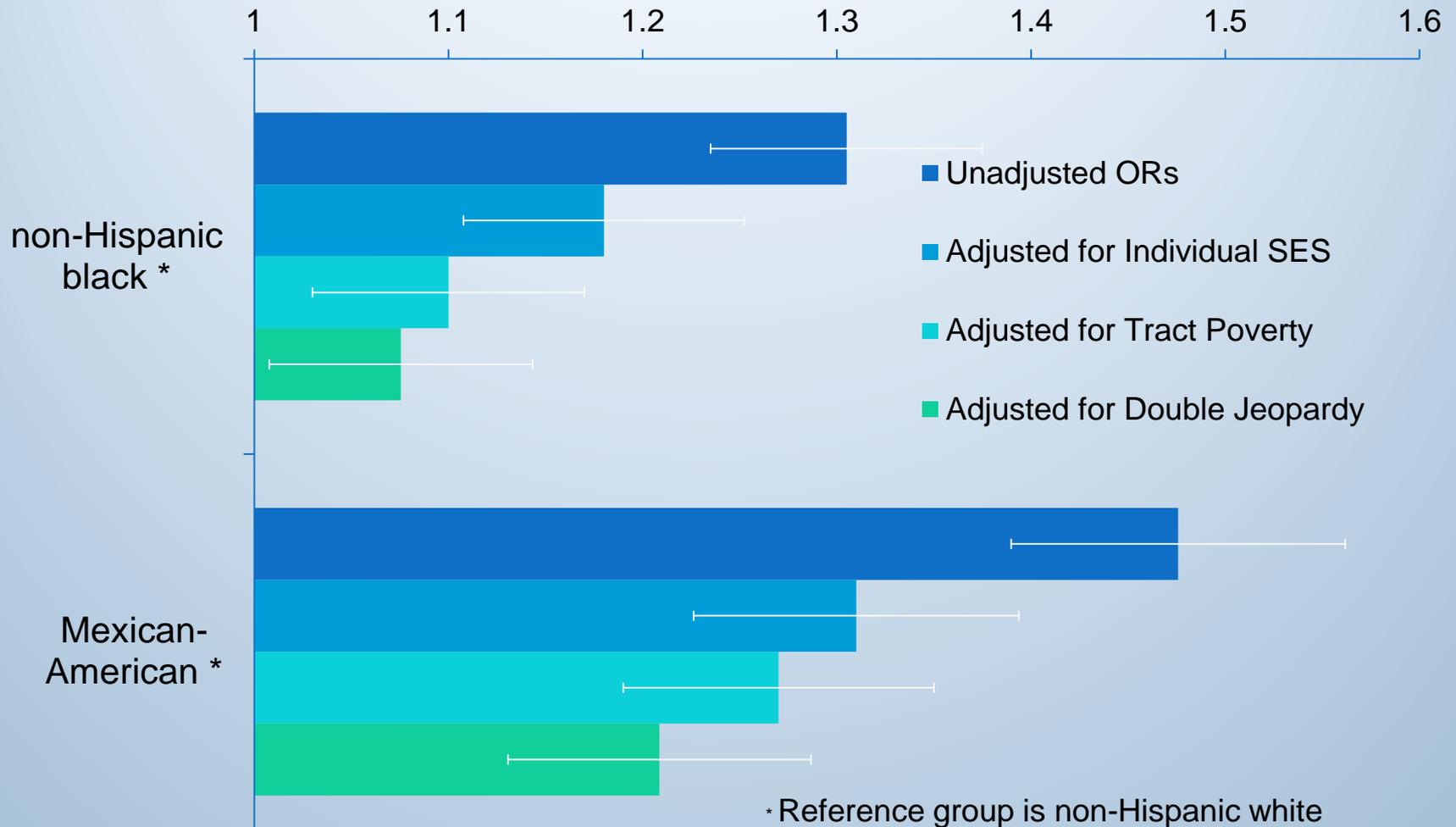
Percent of Children (2-18 yr) in Poverty (Family and Tract level) and “Double Jeopardy”, NHANES 2001-2010 linked to 2000 Census



SOURCE: Rossen et al, preliminary results

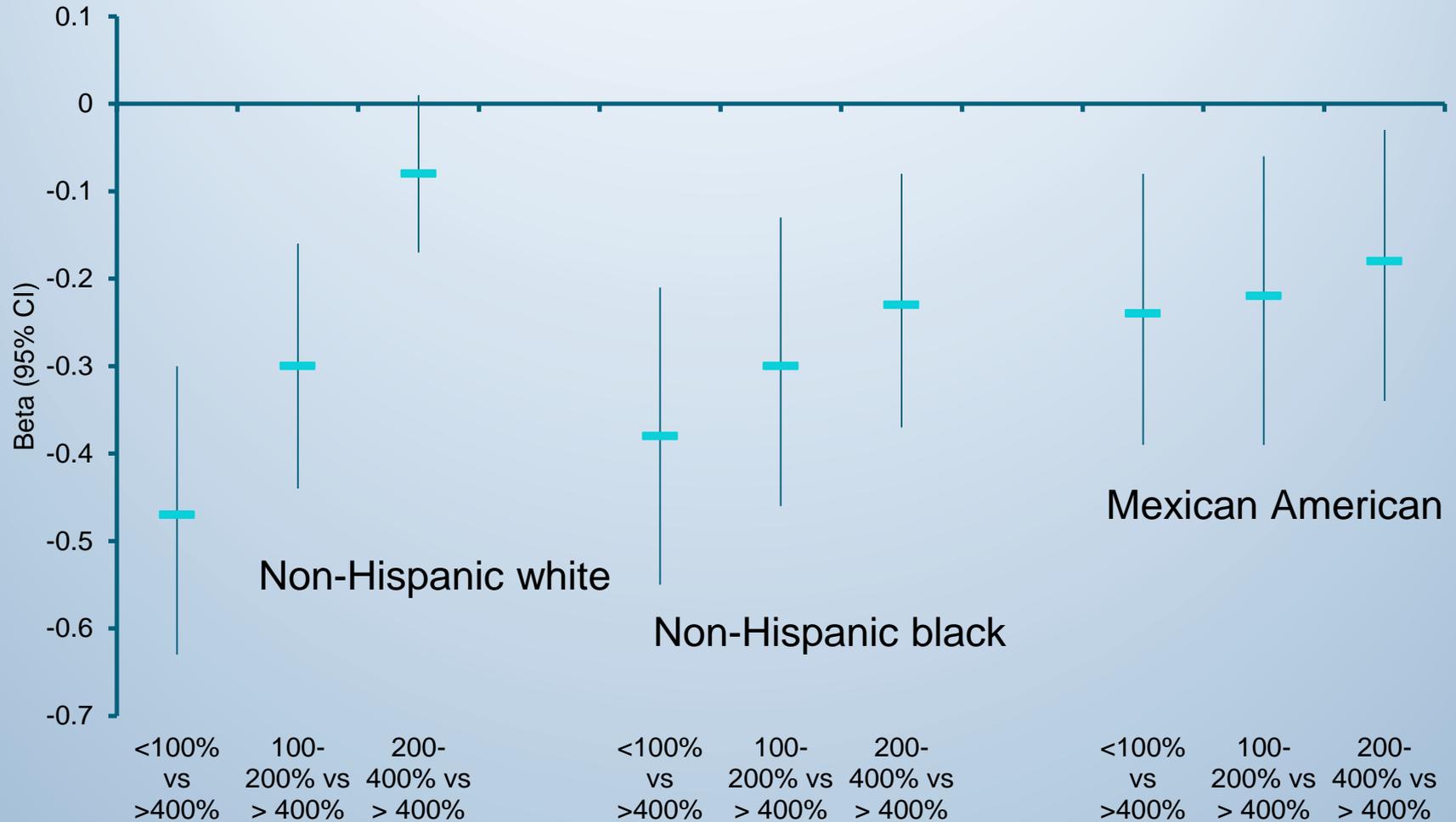
Racial/Ethnic Disparities in Childhood Overweight: The Effect of Adjusting for Poverty and Double Jeopardy.

NHANES 2001-2010 linked with Census 2000 tract poverty



SOURCE: Rossen et al, preliminary results

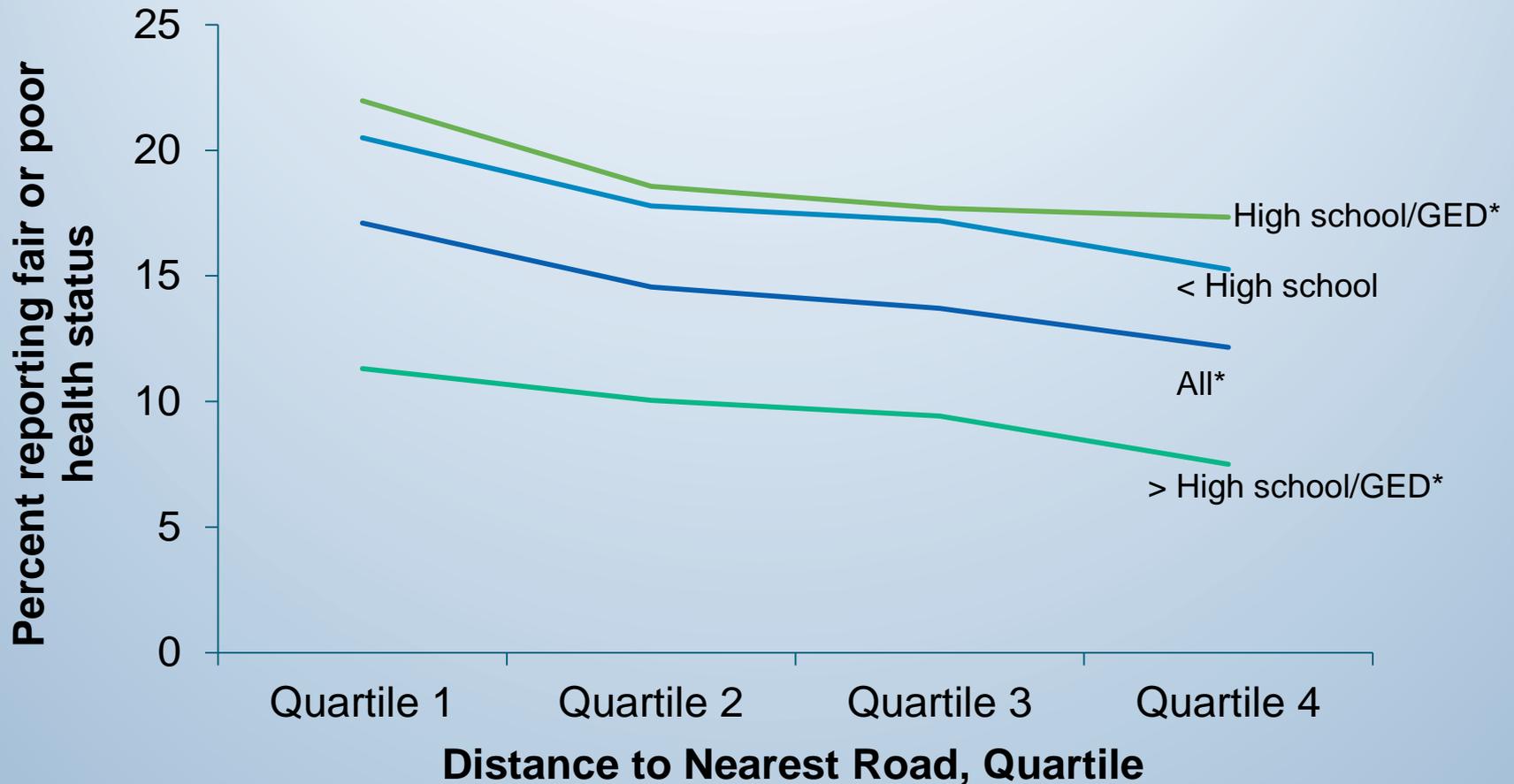
Distance to nearest road (log) by poverty status & race/ethnicity NHANES 1999-2008 linked to 2005 National Highway Planning Network



SOURCE: Berko et al unpublished results

Percent fair/poor health by distance to nearest road

NHANES 1999-2008 linked to 2005 National Highway Planning Network



SOURCE: Parker et al 2012

Considerations

- Temporal relationship between data sources
 - Prospective and retrospective analyses
 - Need to keep track of data collection years for each source
- Linkage bias
 - All survey records may not be linked
 - Linkage quality depends on accurate identifiers
 - All geographic areas in survey may not have contextual data
- Record linked and geographic linked data pose risks of disclosure
 - Access through the NCHS Research Data Center

Summary

- So, why link data for disparities?
 - Survey data
 - better detail on race, ethnicity, socioeconomic indices (education, income), baseline health status, self-reported program participation
 - Administrative records for individuals
 - program participation (Medicaid, SNAP)
 - costs of medical care, benefits
 - longitudinal data adds ability examining health prior to or after the survey
 - mortality and cause of death
 - Geographic data
 - contextual socioeconomic status
 - measures of exposure (pollution, locations of exposures)
 - measures of access (locations of services)
- However,
 - Most analyses of NCHS linked and geographic data must be done in the RDC
 - Care must taken to understand temporal and geographic relationships

Acknowledgements

- Nataliya Kravets
- Jeff Berko
- Ken Schoendorf
- Alan Simon
- Lisa Mirel
- Kim Lochner



National Center for Health Statistics

Data Linkage

More information

Jennifer D. Parker

(301) 458-4419, jdparker@cdc.gov

Disparities in Exposure to Air Pollution during Pregnancy

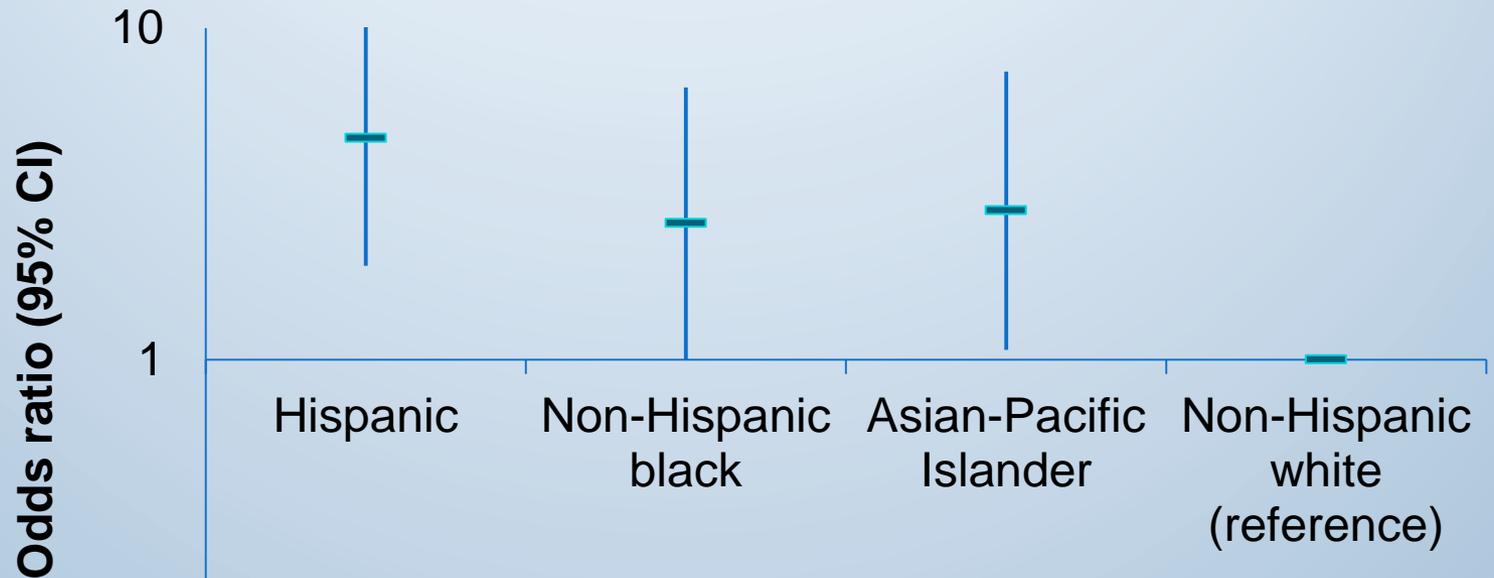
Tracey J. Woodruff,¹ Jennifer D. Parker,² Amy D. Kyle,³ and Kenneth C. Schoendorf²

¹National Center for Environmental Economics, U.S. Environmental Protection Agency, San Francisco, California, USA; ²Infant and Child Health Studies Branch, National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland, USA;

³School of Public Health, University of California, Berkeley, California, USA

Previous research shows poorer birth outcomes for racial and ethnic minorities and for persons with low socioeconomic status (SES). We evaluated whether mothers in groups at higher risk for poor birth outcomes live in areas of higher air pollution and whether higher exposure to air pollution contributes to poor birth outcomes. An index representing long-term exposure to criteria air pollutants

defined as living near particular sources. This approach leaves out other sources of air pollution, such as that from motor vehicles, that could be important. Using air quality monitoring data is a useful way to evaluate the relationship



Relationship between high air pollution index and maternal race. US Natality 98-99

Infant Mortality Rates Among Term Infants (≥ 37 weeks), 1989-2006

