NHANES Sample Design 2019-2022

Jennifer D. Parker, Ph.D., Senior Statistician
Division of Health and Nutrition Examination Surveys, NCHS

NCHS Board of Scientific Counselor Meeting, May 19, 2016

Planning for 2019-2022 NHANES Sample

- Planning is starting now

- Evaluations conducted by Westat and NCHS for proposed changes should begin summer 2016

- Design options needed by September 2016

- Final decisions on design parameters by June 2017

- Sample selection November 2017

- Planning for implementation begins 2017

Objective

- To present options for the 2019-2022 NHANES sample for discussion

    - Current design features

    - Current design considerations

    - Specific design questions

- Obtain input on design, particularly on areas where research or evaluation could be useful prior to final decisions

Current Design Features

- Complex sample survey, including stratification and clustering

  - Priority domains (e.g. groups by race/ethnicity and age) determine probabilities of selection of locations and individuals

- Implementation over 4-years, data released in 2-year cycles

- Three Mobile Examination Centers (MECs), with 2 operating at any given time. Each location (or stand) open for about 8 weeks, with 5 weeks of MEC operation.

- 15 locations per year, with 300-450 examined persons per location, for a target of 5,000 examined persons per year

Current Design (2015-2018) Features

- Target sample sizes for race/ethnicity, income and age domains

  - Hispanics (target 25%)

  - non-Hispanic black (target 25%)

  - non-Hispanic and non-black Asians (target 13%)

  - low income non-Hispanic white/others (target 12%)

- 40% of the sample under 20 years of age

NHANES Sampling Domains 2015-2018

- Stratification and sampling rates determined by need for reliable estimates for specific domains based on age, race/ethnicity, gender, income

  - 15 age-sex subdomains (Combined ages M & F : < 1, 1-2, 3-5; Separate M & F: 6-11, 12-19, 20-39, 40-49, 50-59,60+ years)
    - Hispanics
    - Non-Hispanic black
    - Non-Hispanic, non-black Asian

  - 18 age-sex subdomains (Combined ages < 1, 1-2, 3-5; Separate M & F : 6-11, 12-19, 20-29,30-39, 40-49, 50-59,60-69, 70-79, 80+ years)
    - White/other
      - Low-income
      - Non low-income

Design Considerations

- Two main requirements were established for NHANES III (6 year sample) and have been considered for Continuous NHANES:

  - An estimated prevalence statistic for a specific domain of 10% should have a relative standard error of 30% or less

  - Estimated (absolute) differences between domains of at least 10% should be detectable with a Type I error rate(α)of $\leq$ 0.05 and a Type II error rate (β)of $\leq$ 0.10.

- Assuming design effect of 1.5, a sample size of about 150 is necessary for first condition and 420 for second

- Coarsening over survey years or by combining subdomains is needed to obtain target sample sizes

  - Fasting and environmental subsamples will require more coarsening

# Specific design questions

- Population subgroups that have been suggested

  - Adolescents

  - 0-24 months

  - Pregnant women

  - Asian
    - **Chinese subgroup?**

  - Older adults

- Targets can be reached within current design or by a separately funded parallel study aligned with current design (using another trailer like NHANES National Youth Fitness Survey)

Specific design questions

- Asian oversample has contributed to lower response rates
  - Asians less likely to respond
  - Others in high-Asian areas less likely to respond
- Given that Asians are an important population group, should we:
  - Continue to meet current Asian oversample targets?
  - Consider something else?
    - Oversample at lower rate?
      - Explore model-based estimation?
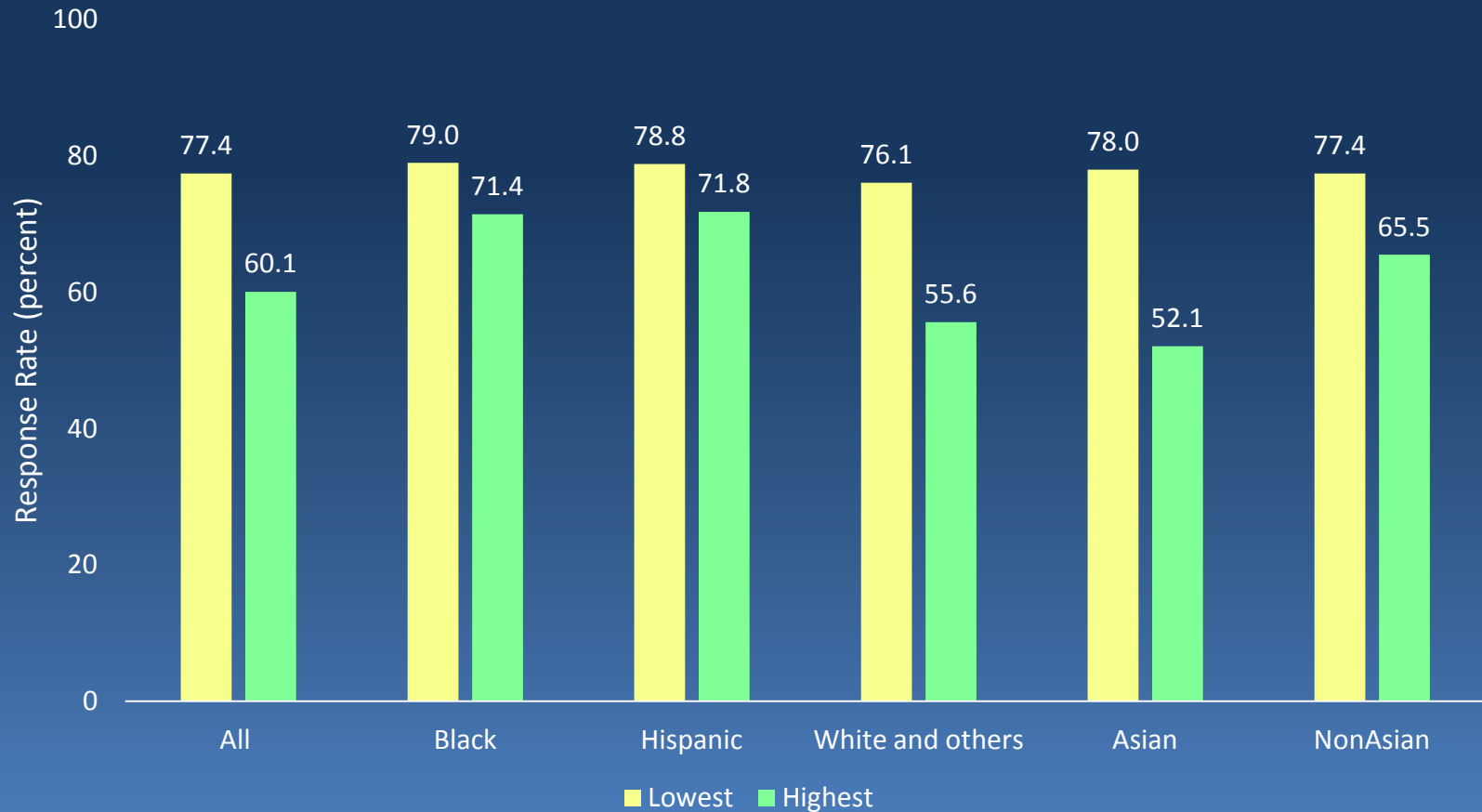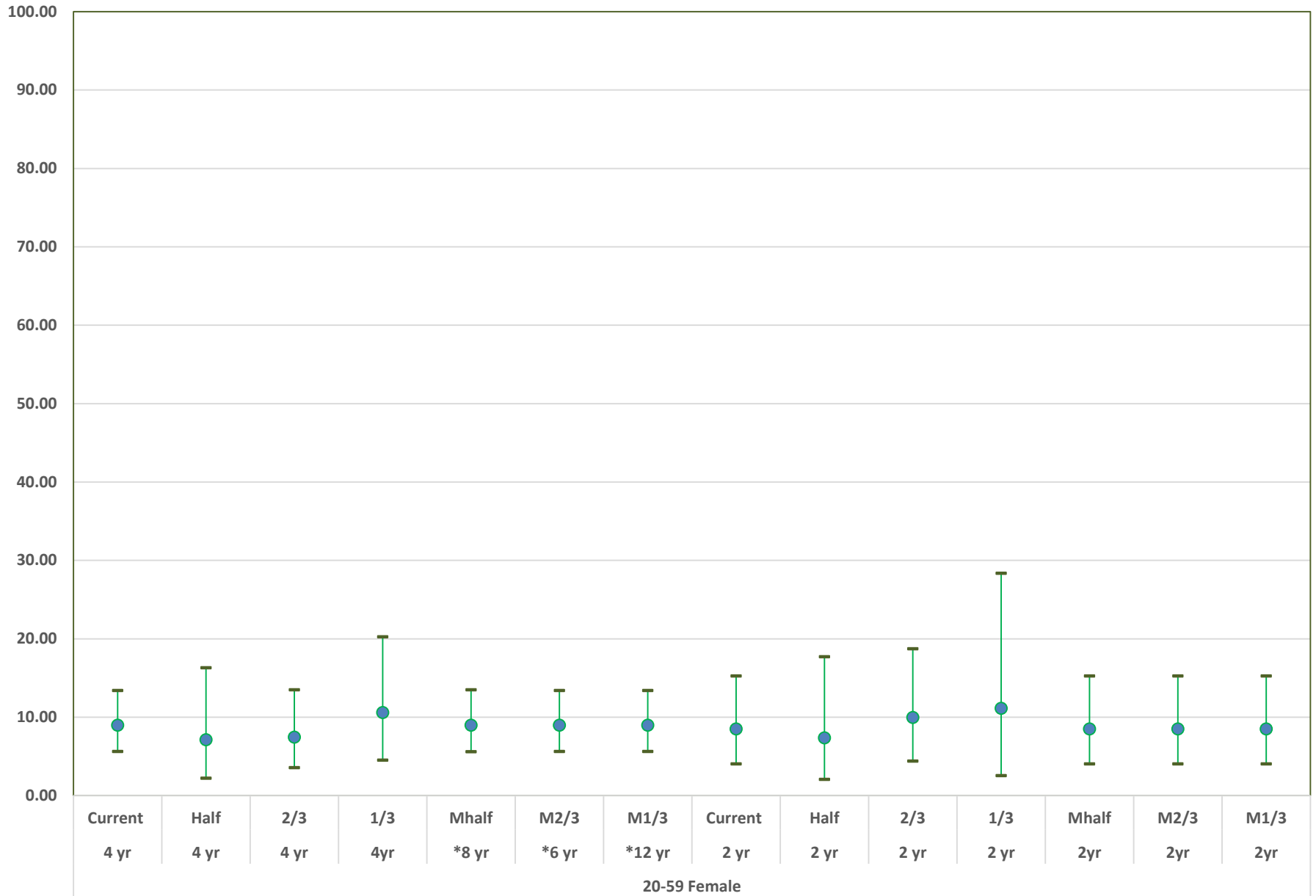   - Sample separately?

Response Rates by Race/Ethnicity 2007-2015
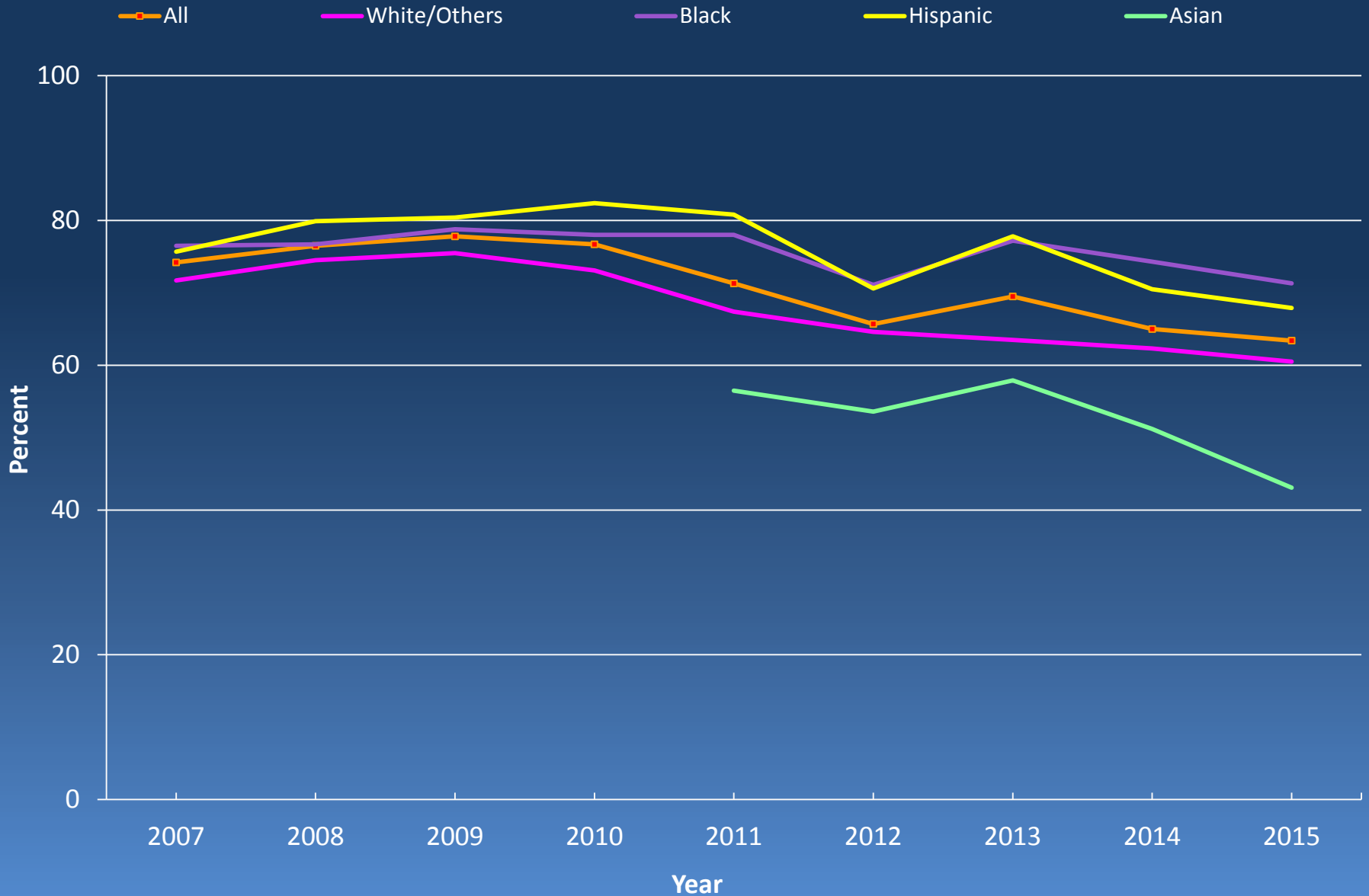
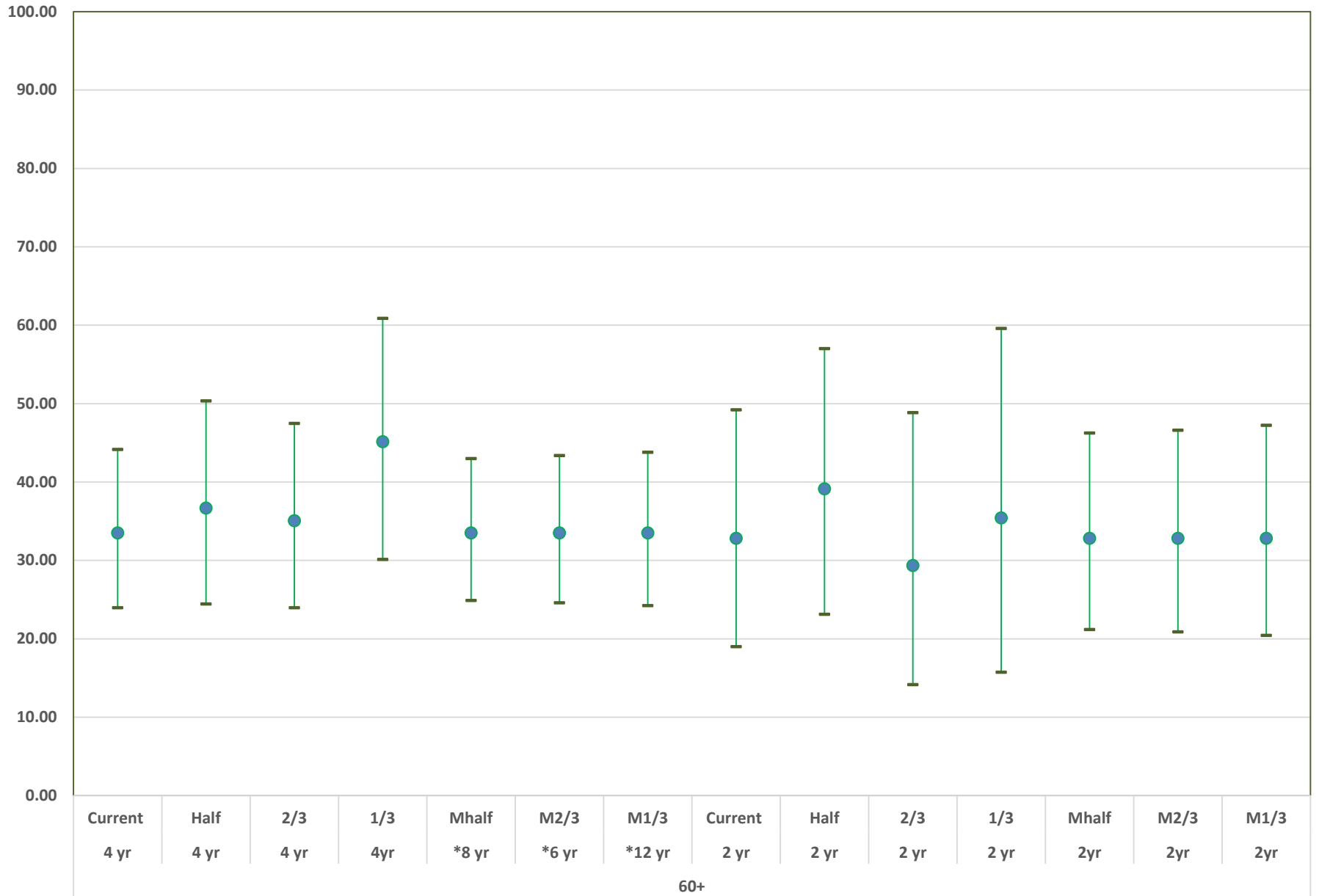Response Rates by Quintiles of Percent Asian Screened in Stands
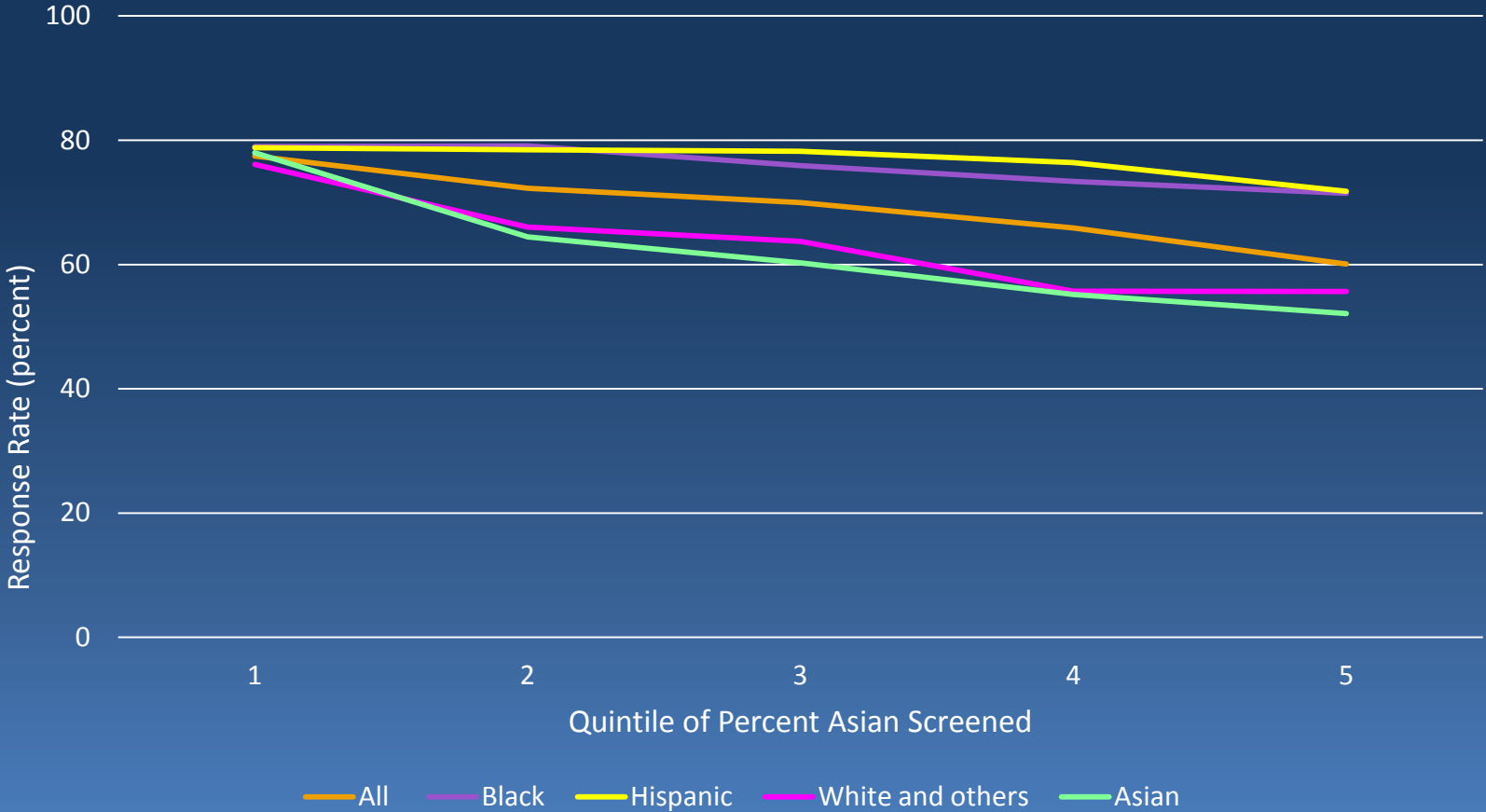
Mean stand response rates by quintile of Asian screened in stand, lowest and highest quintile, NHANES 2011-2014

Diabetes prevalence for Asian females ages 20-59 years

Diabetes prevalence for Asians ages 60 and over

Summary of issues to consider

- Should we consider response rates in determining priorities?

- Should we develop new criteria for precision?

  - For which domains? Over how many years?

  - Fasting and environmental subsamples?

- Should we consider a six-year design?

- Are there other groups we want to consider oversampling?