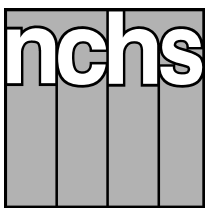
Exploring Alternative Methods for Bridging Multiple Race Data

Elizabeth Arias, Ph.D.,
Division of Vital Statistics

Jacqueline Wilson Lucas, M.P.H.,
Division of Health Interview Statistics

National Center for Health Statistics



CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION

Background

- Original project designed to explore how multiracial persons were classified on death certificates (**Pastor, et.al**.), and develop a mortality profile of multiracial persons
- Need to explore various options for bridging multiple race persons back to single race groups to maintain trends in data (applied to vital statistics, record-based surveys)
- One of at least two projects currently underway at NCHS to explore bridging methodologies for vital statistics data – **Schenker, Parker, and Ingram** working on an **official** bridging methodology for vital statistics data, using different statistical approach

What Is Bridging?

- The process of ***re-categorizing*** data collected under the new standards to make them comparable to data collected under the old standards - multiple race to single race group
- Designed to allow for maintenance of ***trends*** in various kinds of data collected by race
- ***Methods*** developed and evaluated by a committee of statisticians and other consultants looking at ways to tabulate multiple race data

Bridging Methods Recommended by OMB

- **Whole allocation**
  - Smallest group
  - Largest group other than white
  - Largest group
  - ***NHIS largest group***
- **Fractional allocation**
  - Equal fractions
  - ***NHIS fractions***

Alternative Bridging Techniques

- **Model-based approach to bridging**
  - More statistically defensible
  - Use characteristics of respondents to determine how to allocate multiple race persons
  - Use variables in models common to most surveys (individual and aggregate level characteristics)

Study Questions

- **What are the factors associated with selection of a ‘primary’ race among multiple-race individuals?**
- **How do the various multiple race groups differ from each other in their selection of a ‘primary’ race?**
- **How can we translate observed findings into a bridging algorithm that will re-categorize multiple race individuals into single race groups?**

Data sources

- **Merged National Health Interview Survey data 1986-1994 (used in original match to the NDI):**
  - NHIS has allowed respondents to report more than one race since 1976 (adults 1976-1981, all household members since 1982)
  - NHIS asks two part race question, the first part of which allows multiple races to reported, and the second of which asks multiple race respondents to select the single group they think best describes them - **built-in method for bridging back to single race group**.
- **Linked NHIS data to 1990 public-use STF3A files that contain county-level population characteristics (linkage done by state and count FIPS codes)**

Methodology

- **We estimate multinomial logistic regression models to develop predicted probabilities of selecting a primary race as a function of selected individual and contextual-level characteristics (using STATA software)**

- **The statistical models take the following form:**

-

Methodology, Cont'd

- $\ln[P(Y_i=r_k|X_i)/P(Y_i=r_m|X_i)]$ ➔
- the log odds of choosing one 'primary' race vs. another, where:
  - r_k = primary race category of interest;
  - r_m=comparative primary race category.

Methodology, Cont'd

- Probability Profiles:
  - We use results of estimated multinomial logistic regressions to produce predicted probability profiles (using Lotus software)
  - We estimate probability profiles in the following manner:
  - $P(Y_i=r|\text{☞}X_i) \Rightarrow$
  - $\exp(\text{✂} + \text{☞}_j \text{🖇}_{jr} X_{ji}) / \text{☞}\exp(\text{✂} + \text{☞}_j \text{🖇}_{jm} X_{ji})$

Variables

- **Dependent Variable (Y) => 'Primary Race'** :
- White-AIAN: White, AIAN, Multiple
- White-Black: White, Black, Multiple
- White-API: White, API, Multiple
- AIAN-Black: Black, AIAN, Multiple

- **Independent Variables(X)**:
- Hispanic Origin:
  - Hispanic
  - Non-Hispanic
- Age: <18, 18-44, 45-+
- Sex: Male, Female
- Education (of household adult):
  - <High School
  - High School Graduate
  - > High School

Variables, Cont'd

- Family Income: four category variable reflecting the four quartiles of the distribution of total family income.
- Geographic Region of Residence:
  - West
  - Northeast
  - South
  - Midwest
- County Racial Composition: A four category variable reflecting the four quartiles of the distribution of the ratio of one single race to another in country of residence.

Variables, Cont'd

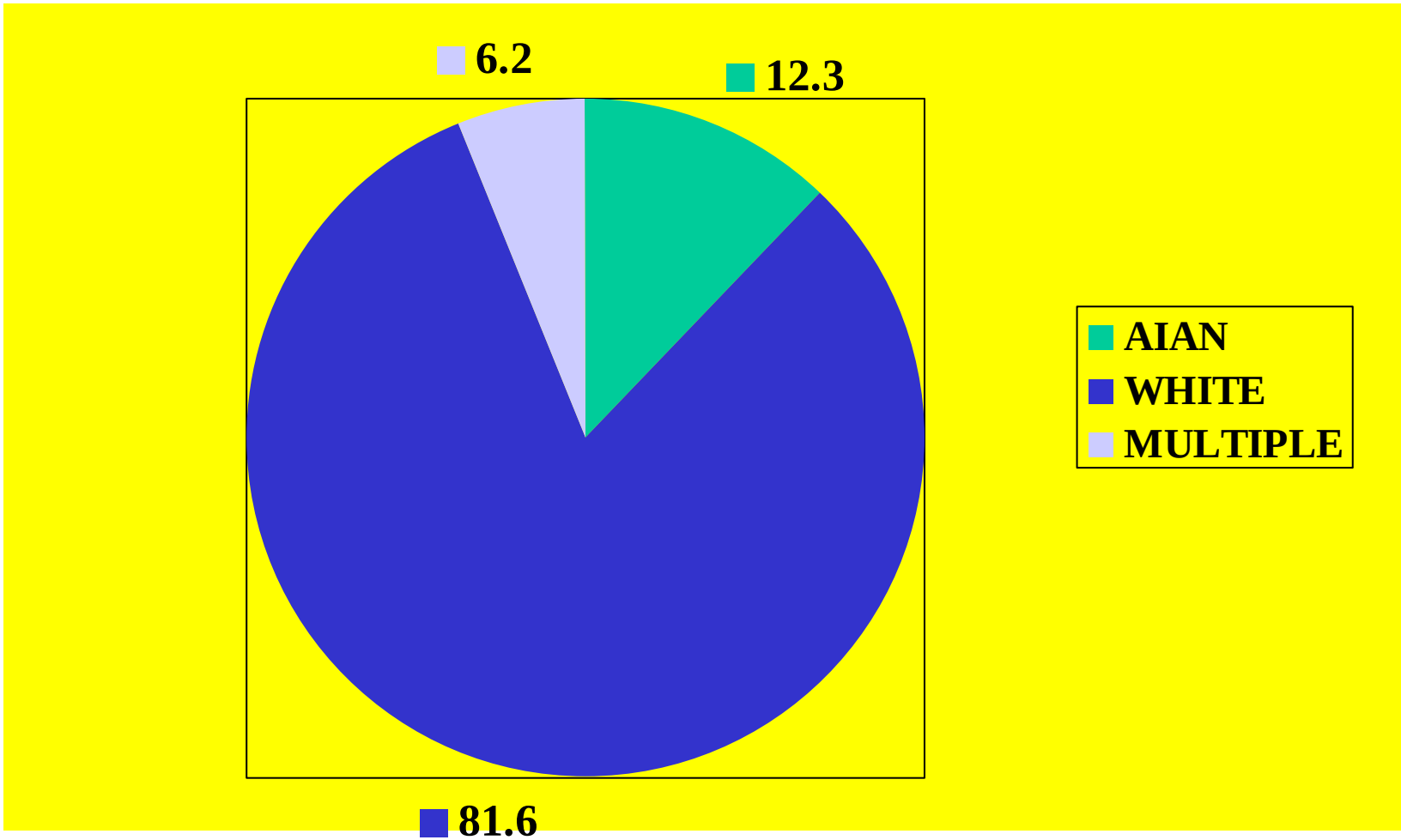
. *Respondent status*: self or proxy; included as a control because the NHIS responses to survey questions are given by an adult in the household for most individuals under age 18.

- *Survey Year:* a continuous variable included to control for period effects of race perceptions as well as survey questionnaire effects.

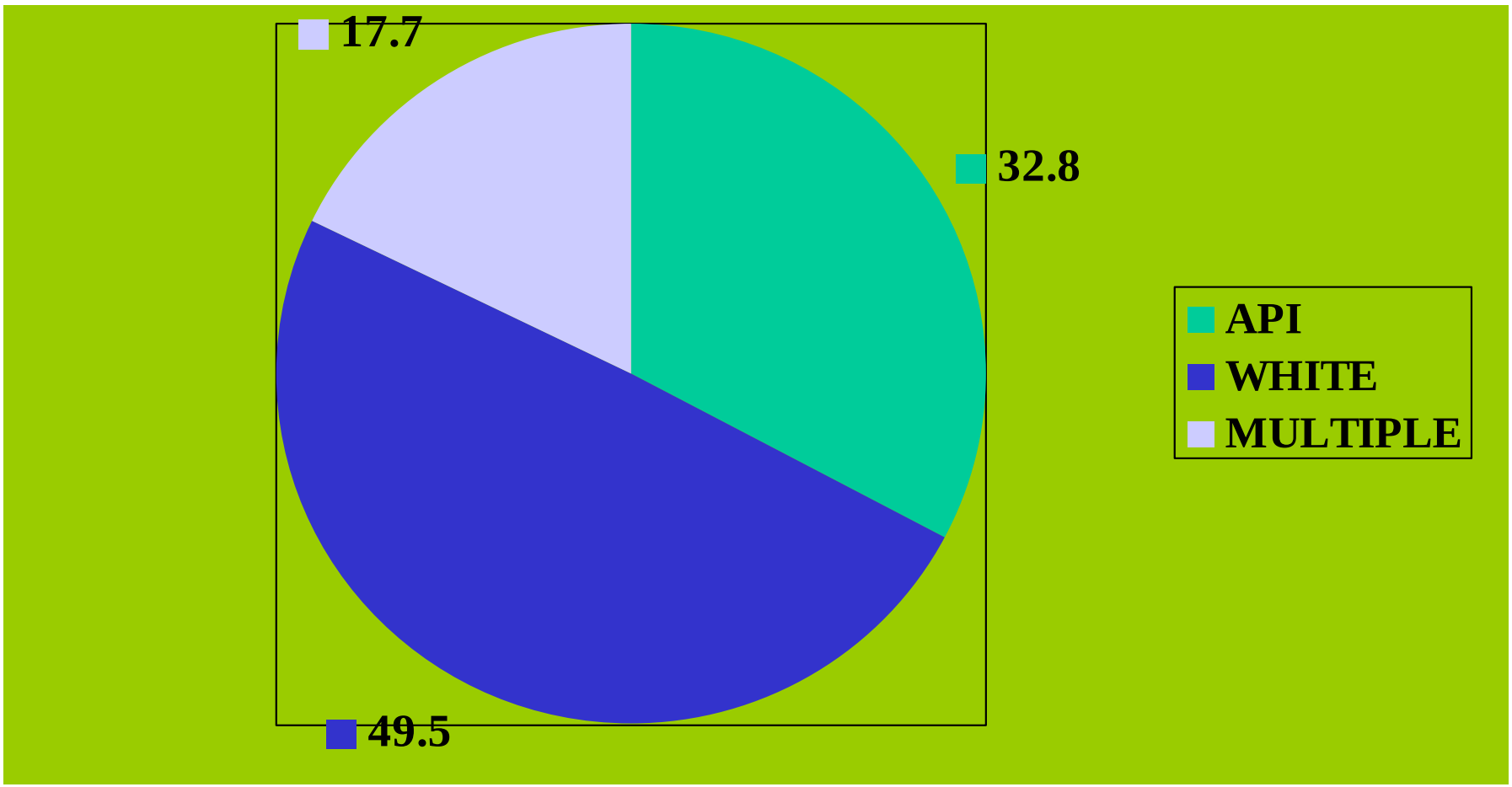
Results

- **Total sample size is approximately 14,000 bi-racial individuals**
- **The primary bi-racial groups examined are AIAN-White; API-White; AIAN-Black; and Black-White (largest multiple race groups)**
- **Descriptive results are shown first, followed by results of multinomial logistic regressions and estimated predicted probabilities**

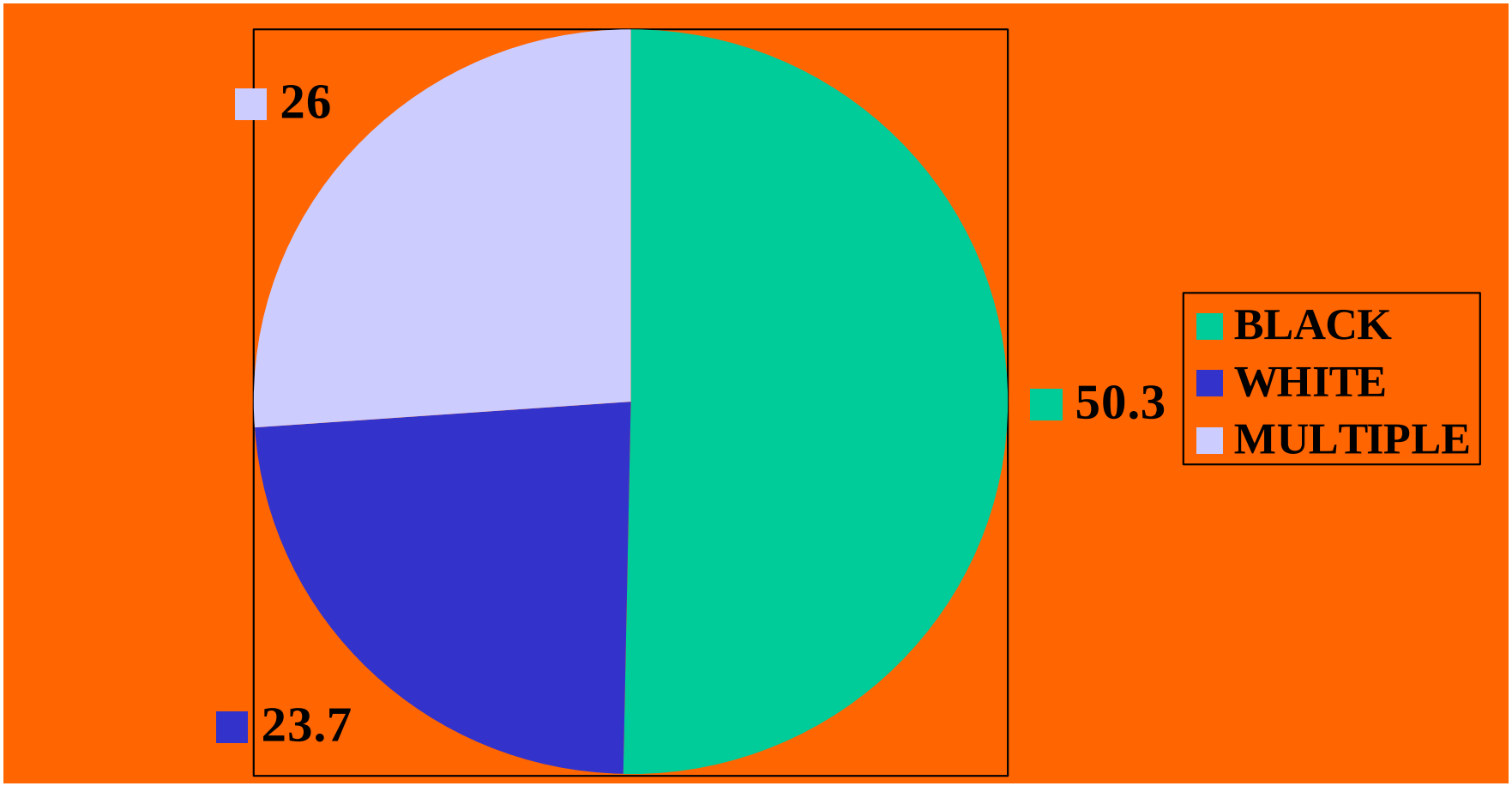
AIAN/White:
N = 7,646



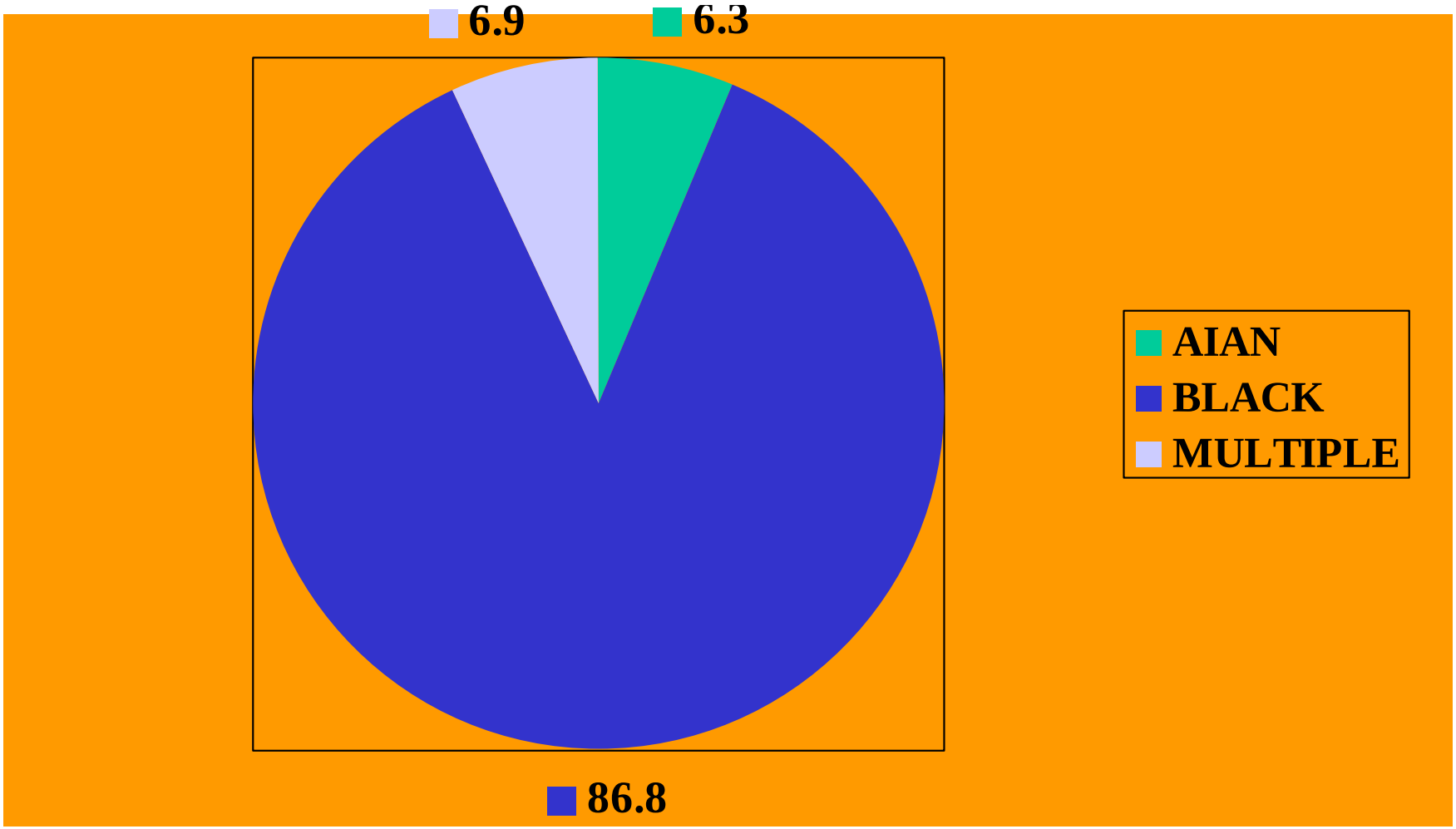
API/White:
N = 1,797



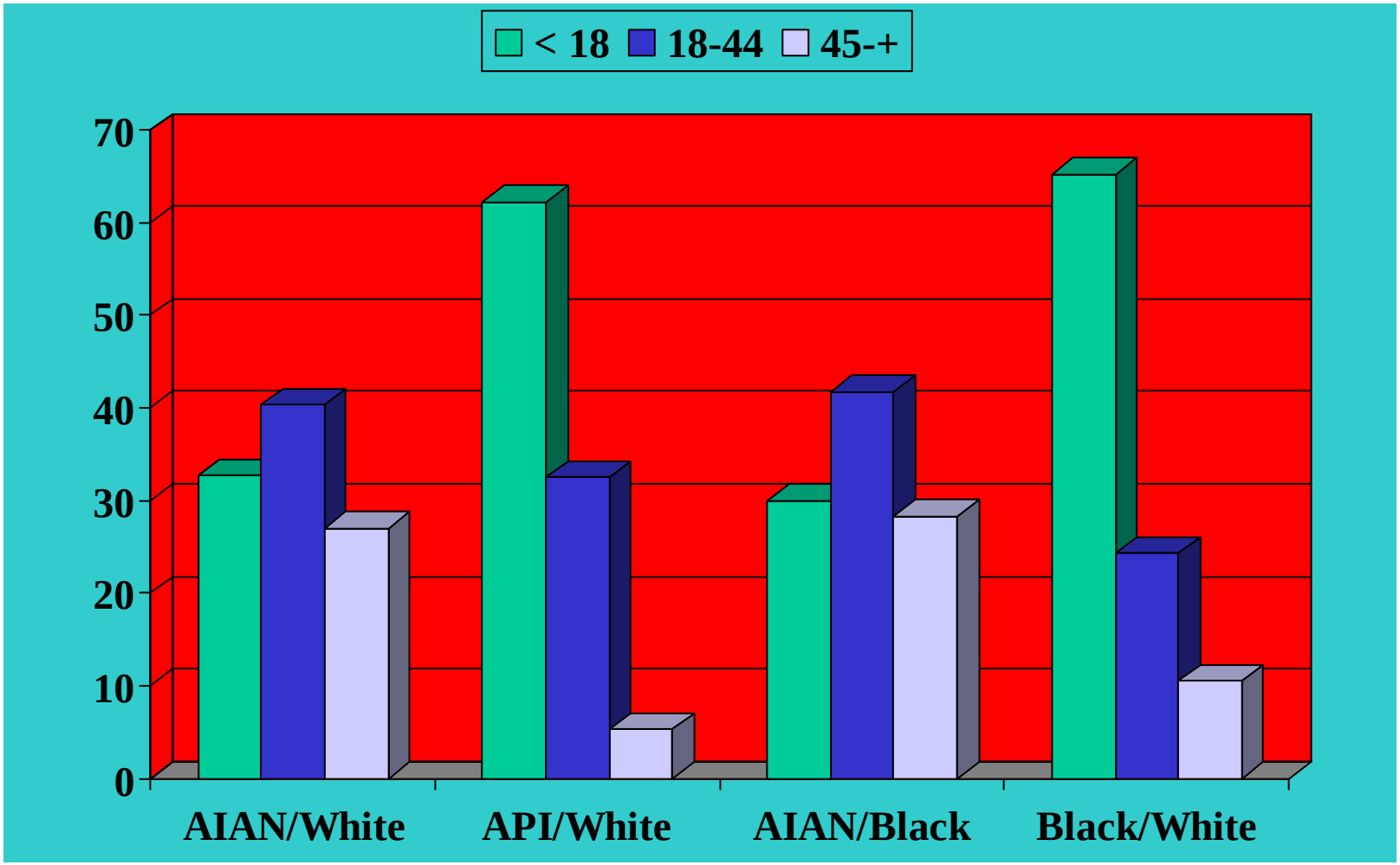
Black/White:
N = 1,555



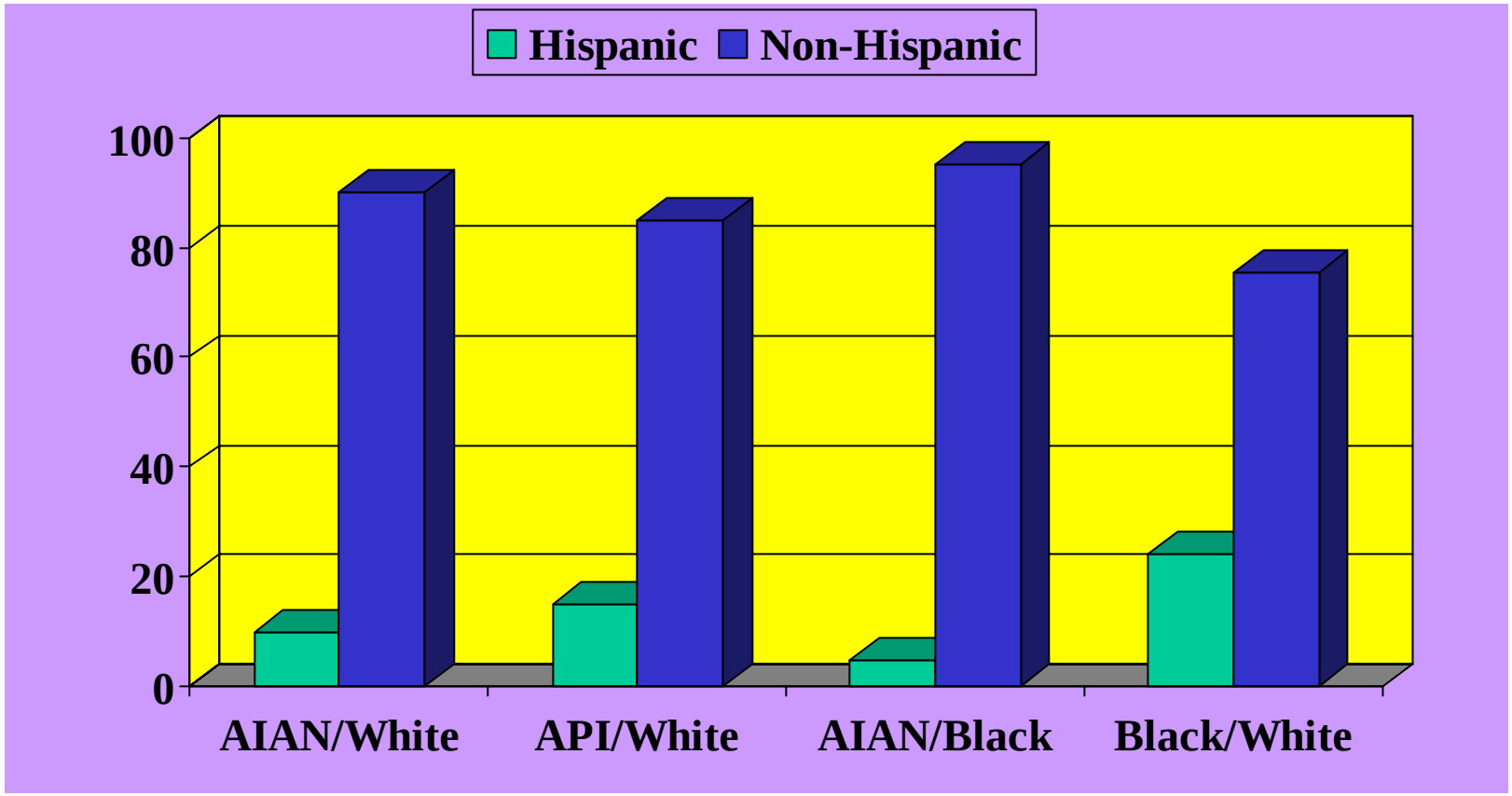
Black/AIAN
N = 1,525



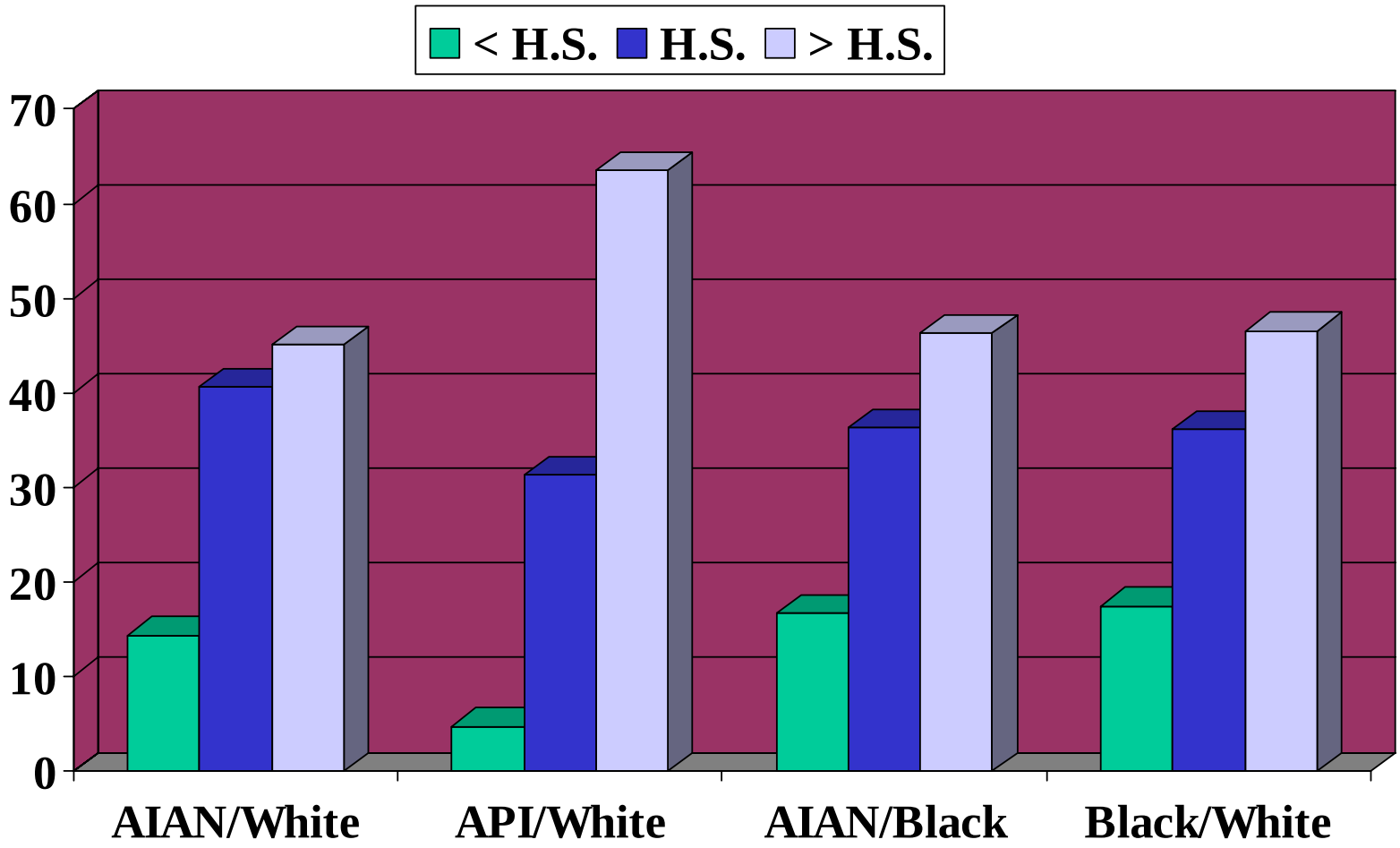
Age Composition



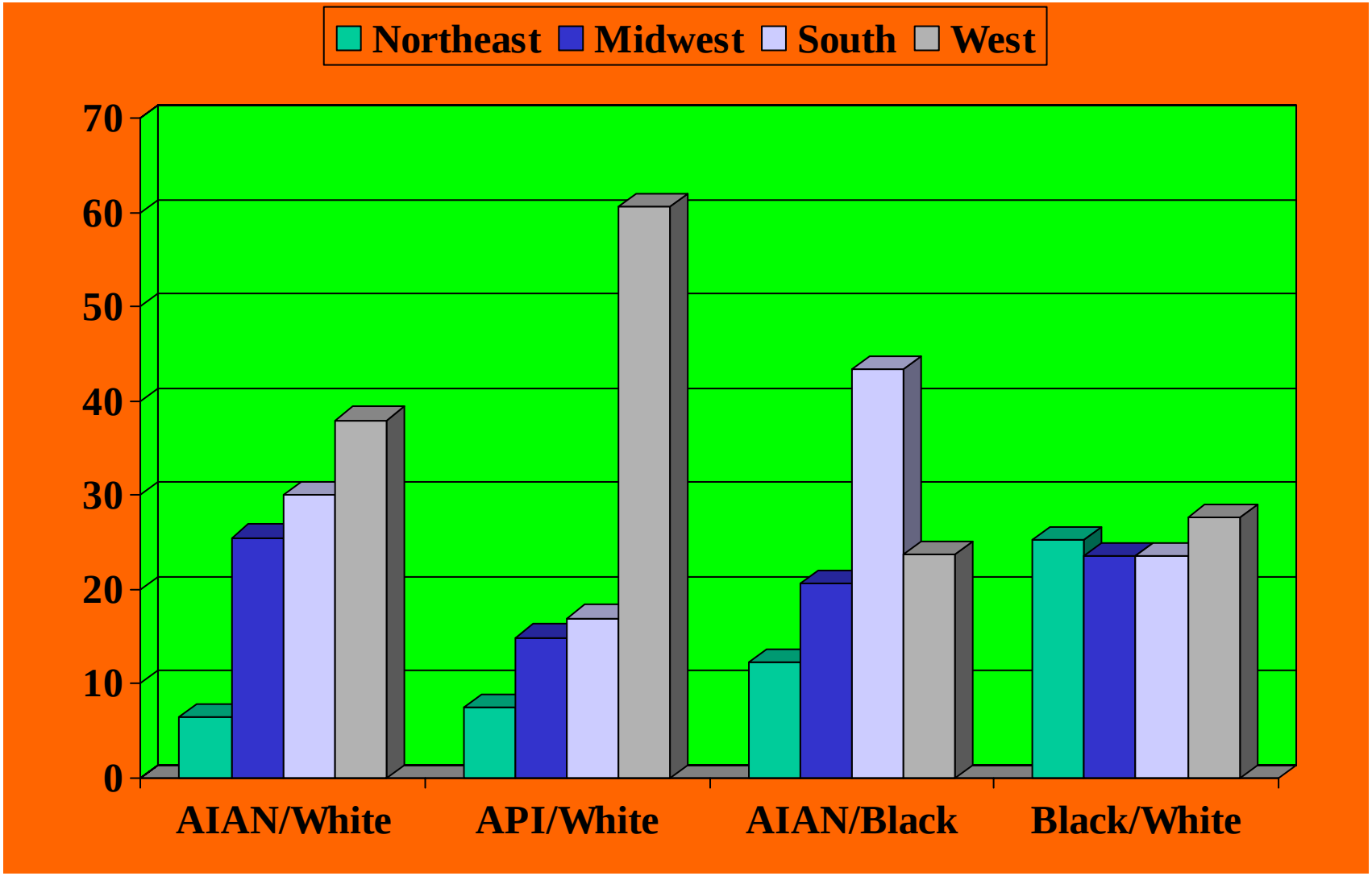
Hispanic Origin



Education Of Household Adult



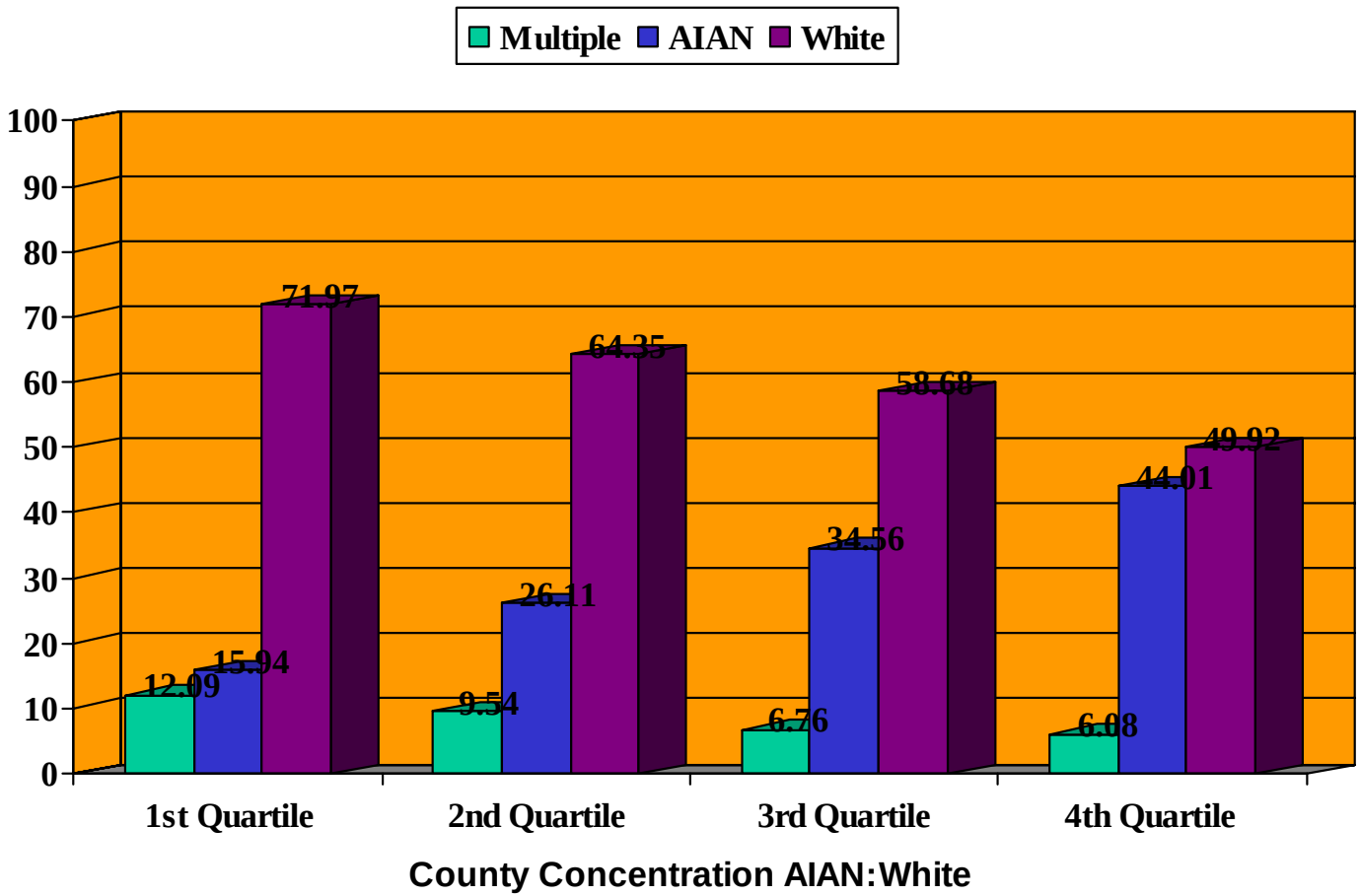
Region Of Residence



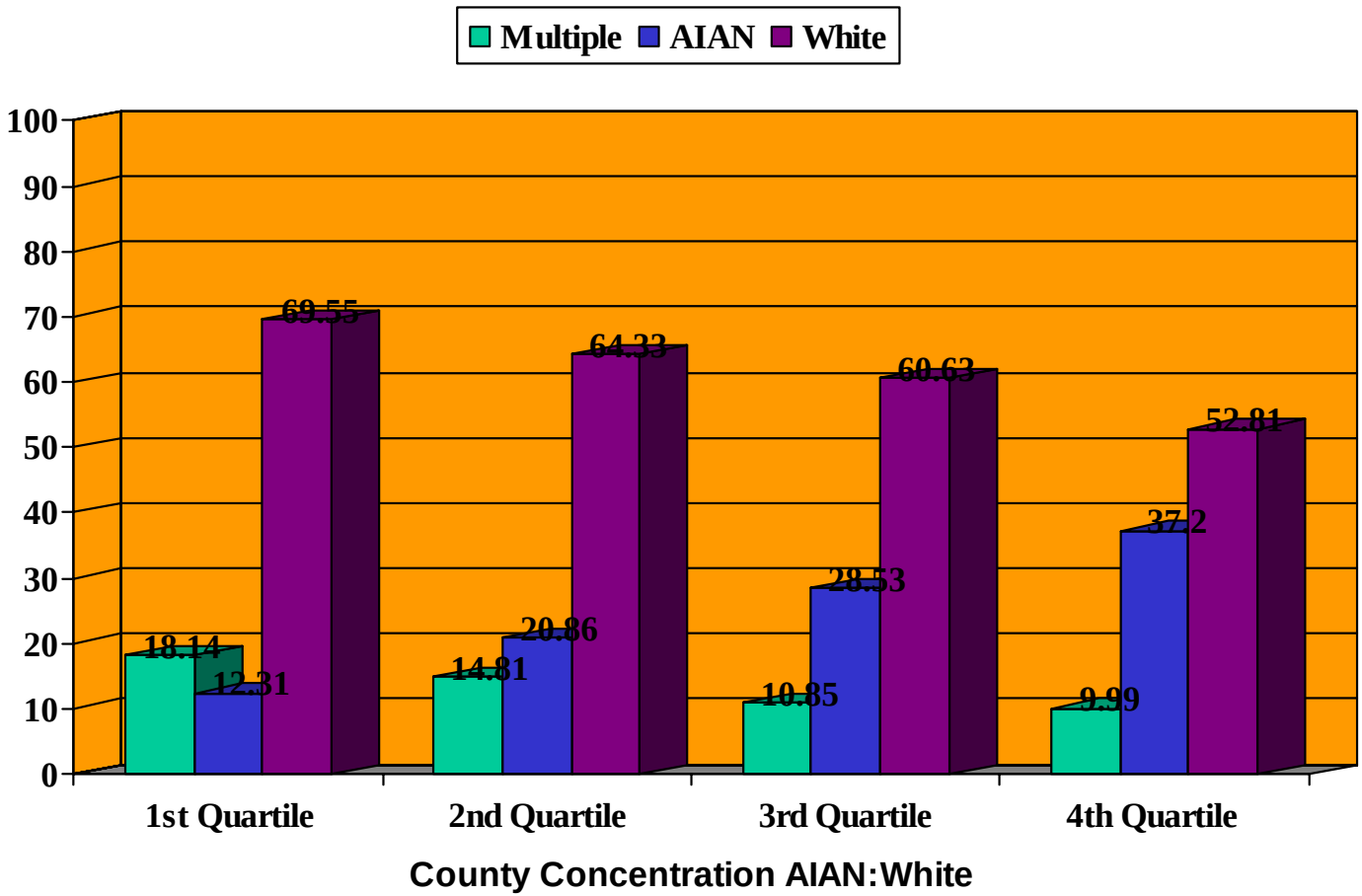
Multinomial Logistic Regression Results

	AIAN-White	Multiple-White	API-White	Multiple-White	AIAN-Black	Multiple-Black	White-Black	Multiple-Black
Hispanic	H	H	H		H	H		
Age								
Sex		F	M					M
Education							H	
Income	L						H	H
Region	M			NE				NE
Minority| Majority	HC		HC				LC	
Survey Year						+YR		
Resp		S	S					

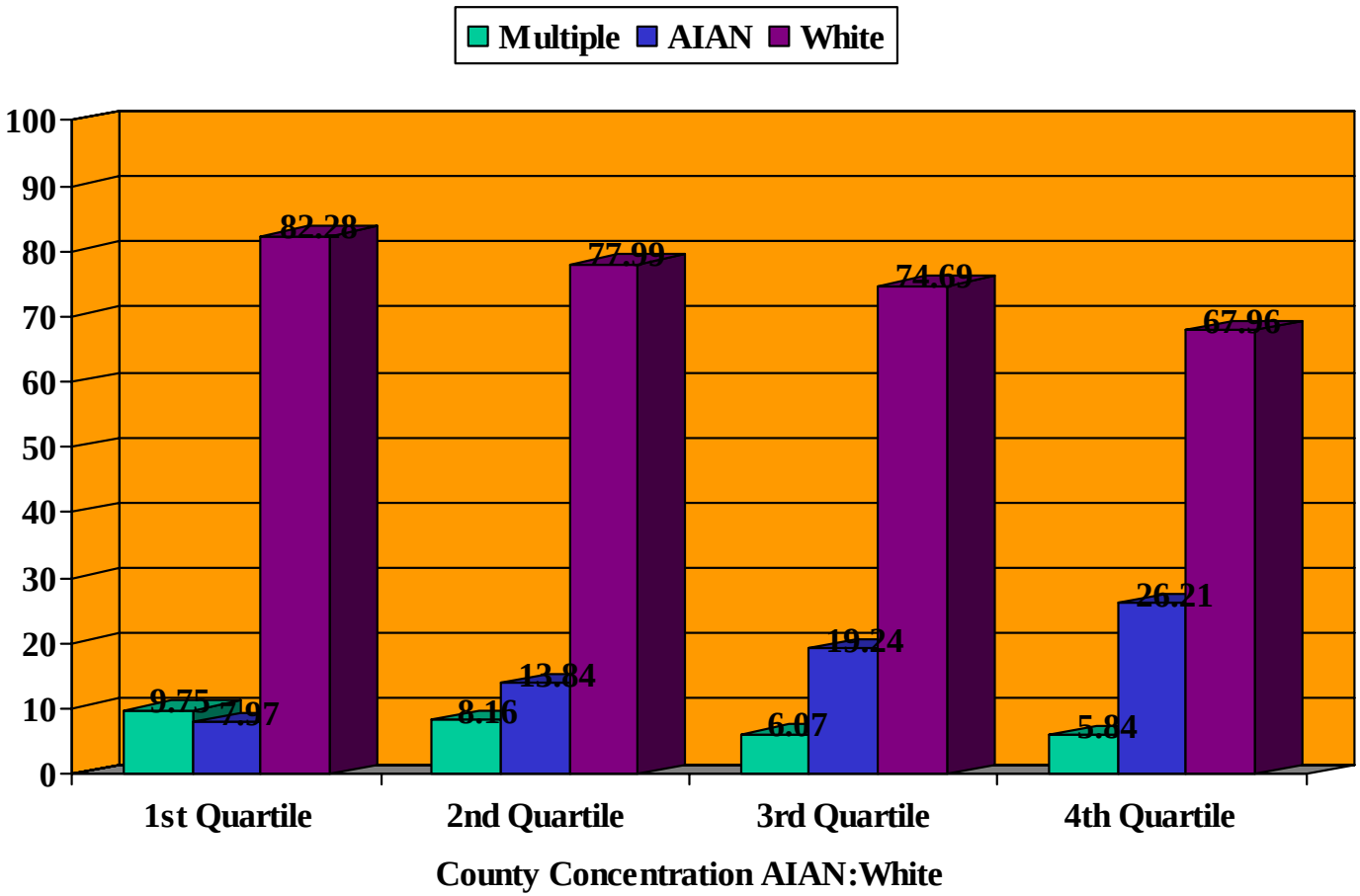
AIAN/White Hispanic Males, Age < 18, Low SES, Northeast



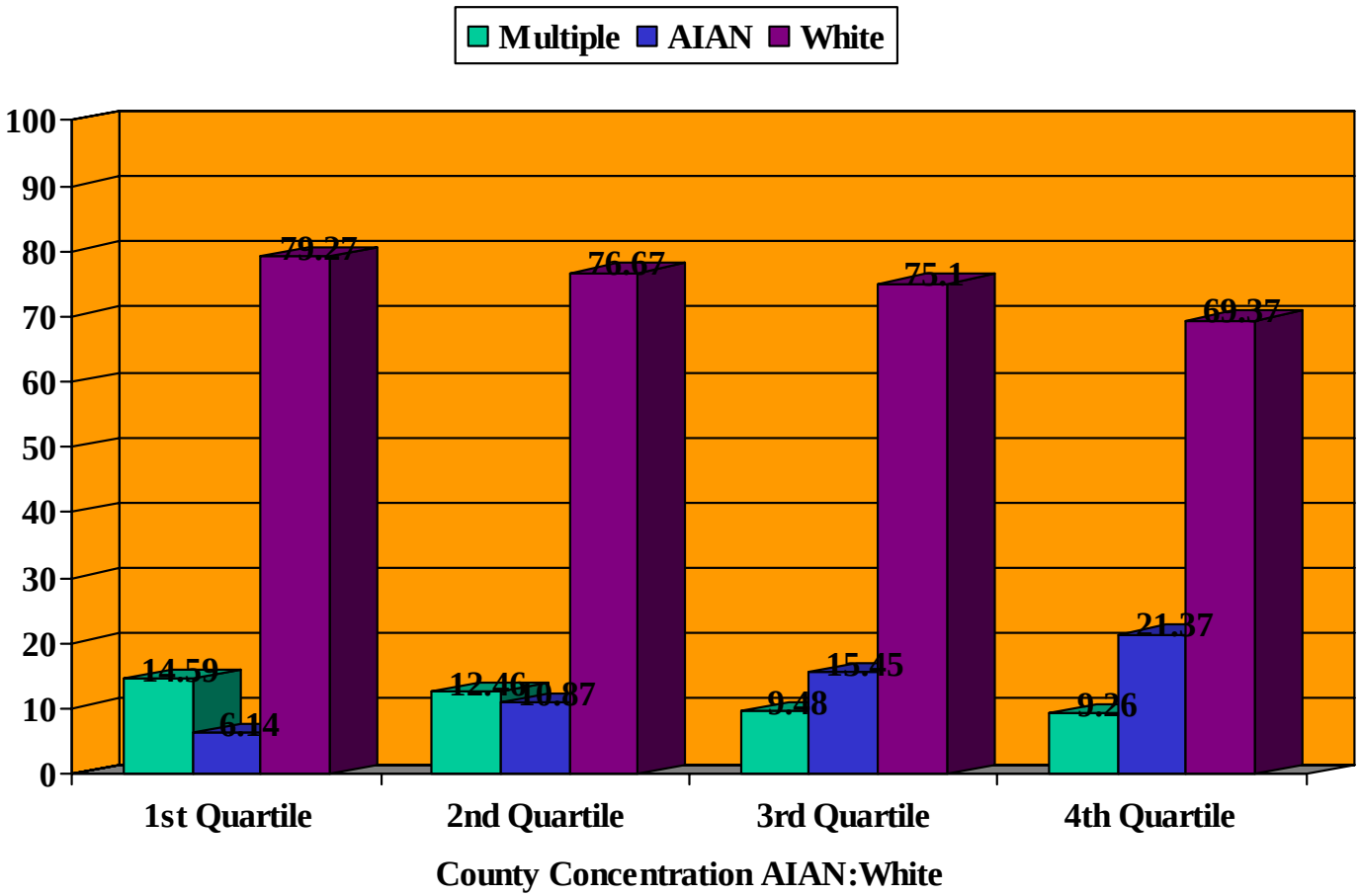
AIAN/WHITE Hispanic Males, Age < 18, Low SES, West



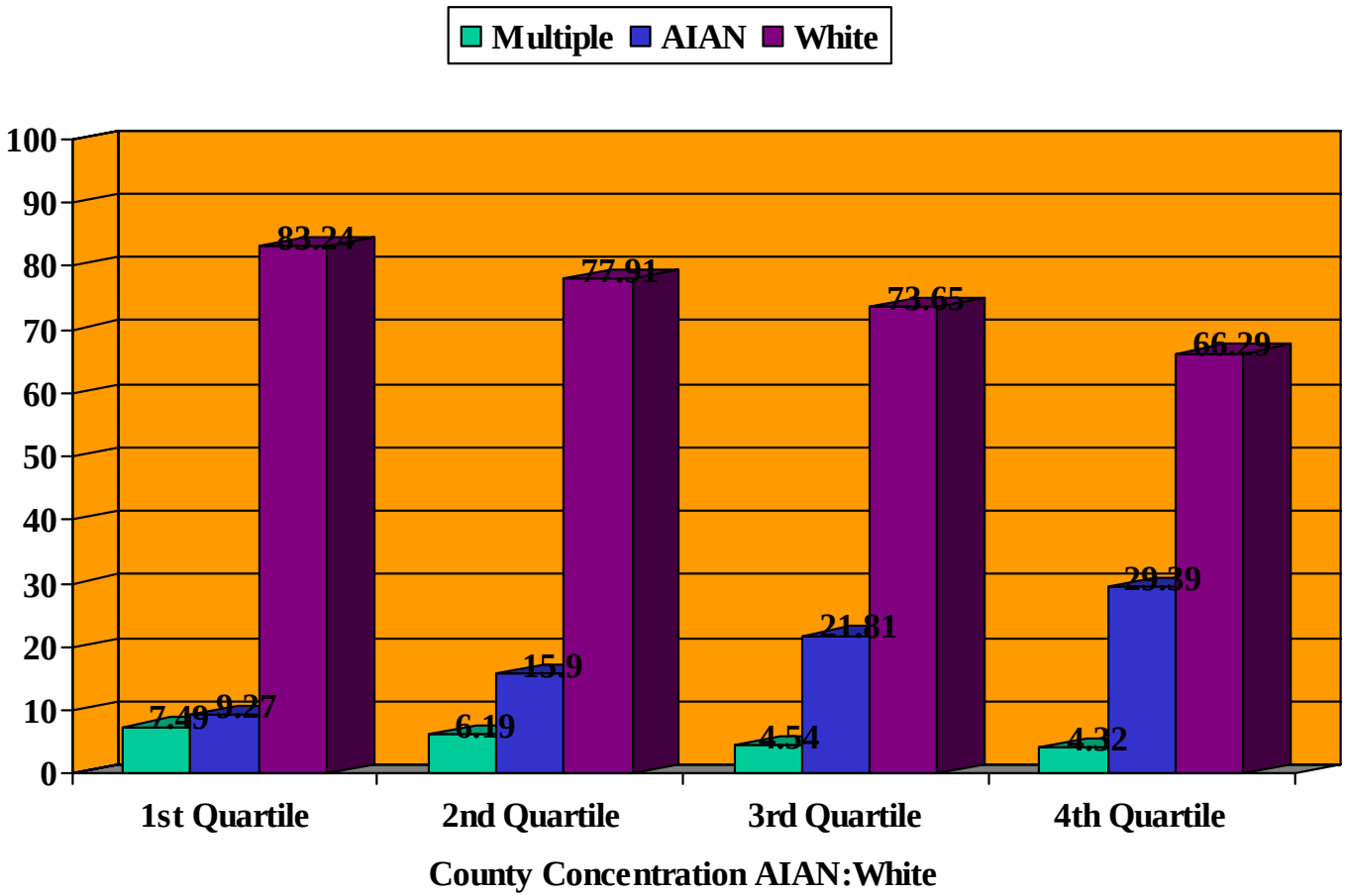
AIAN/White Hispanic Males, AGE < 18, High SES, Northeast



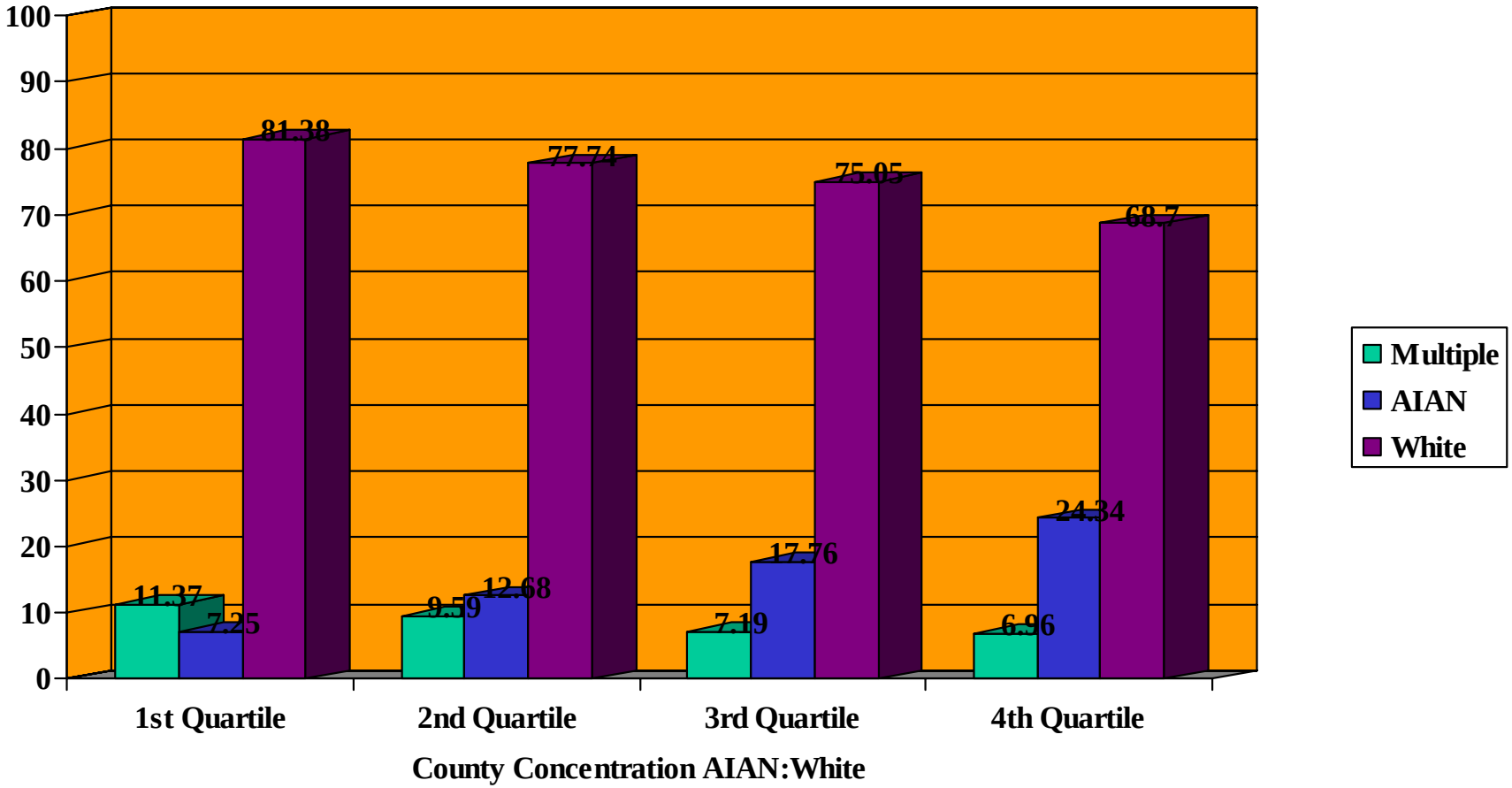
AIAN/White Hispanic Males, Age < 18, High SES, West



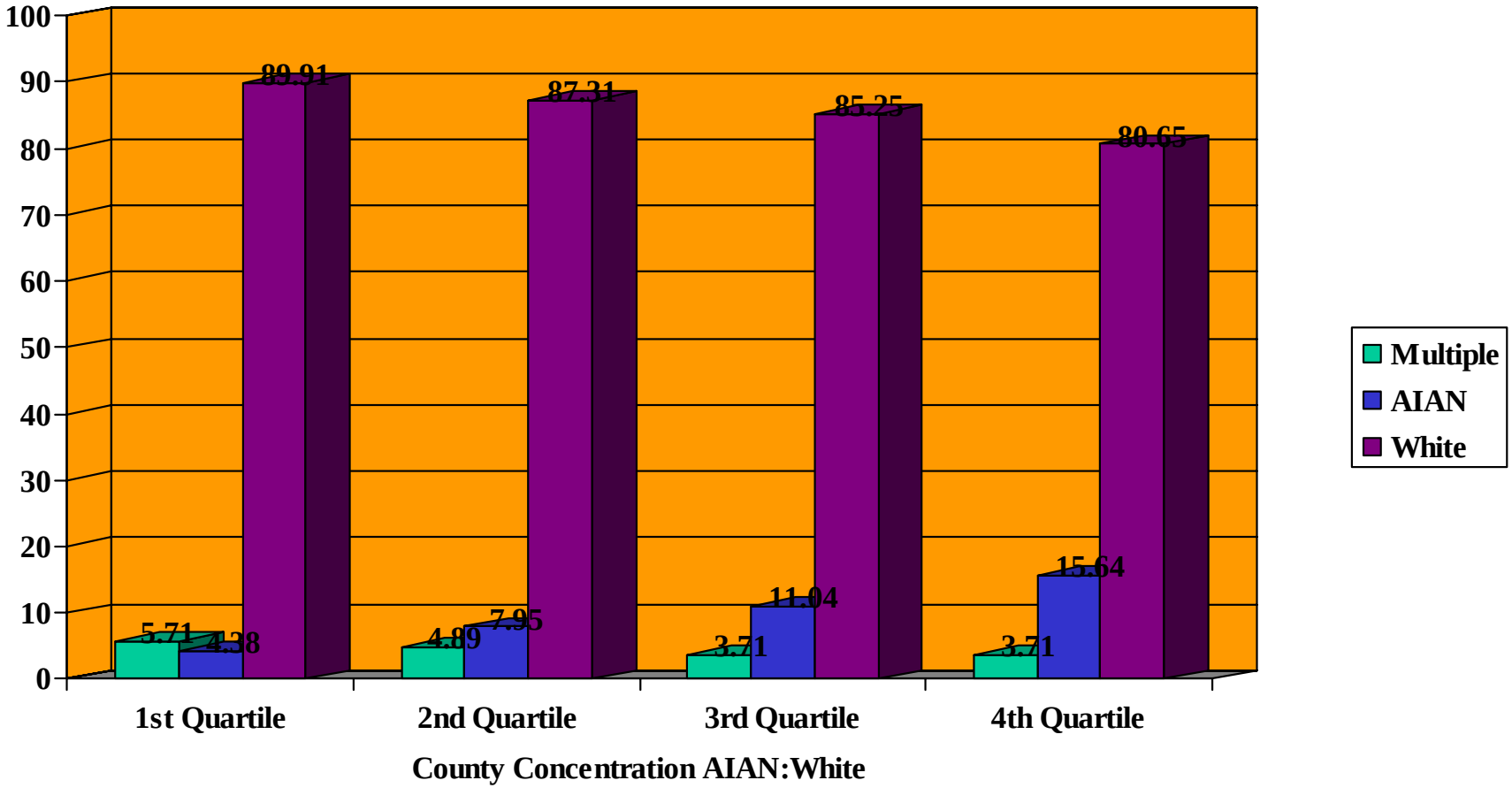
AIAN/White Non-Hispanic Males, Age < 18, Low SES, Northeast



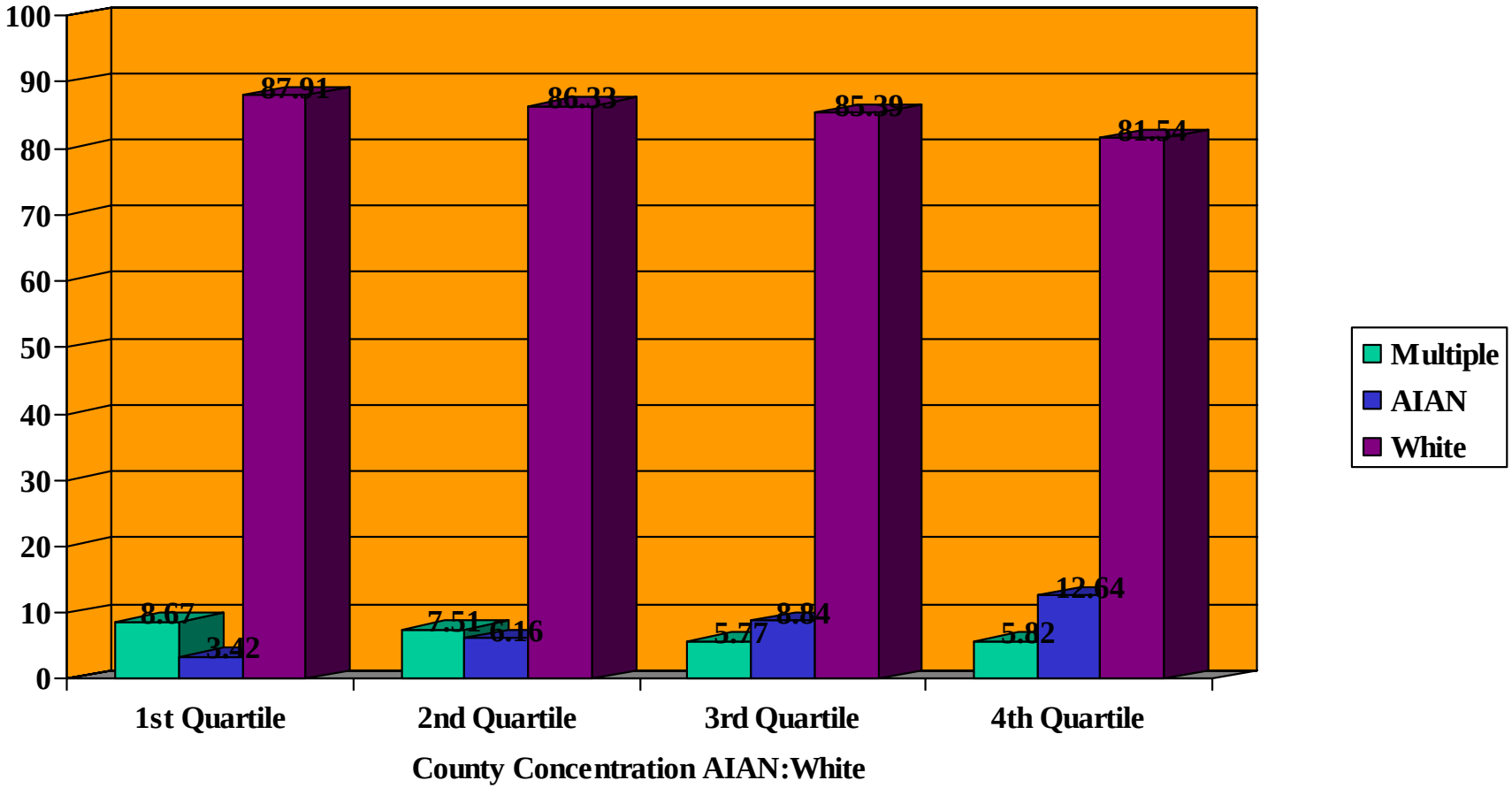
AIAN-WHITE NON-HISPANIC MALES, AGE < 18, LOW SES, WEST



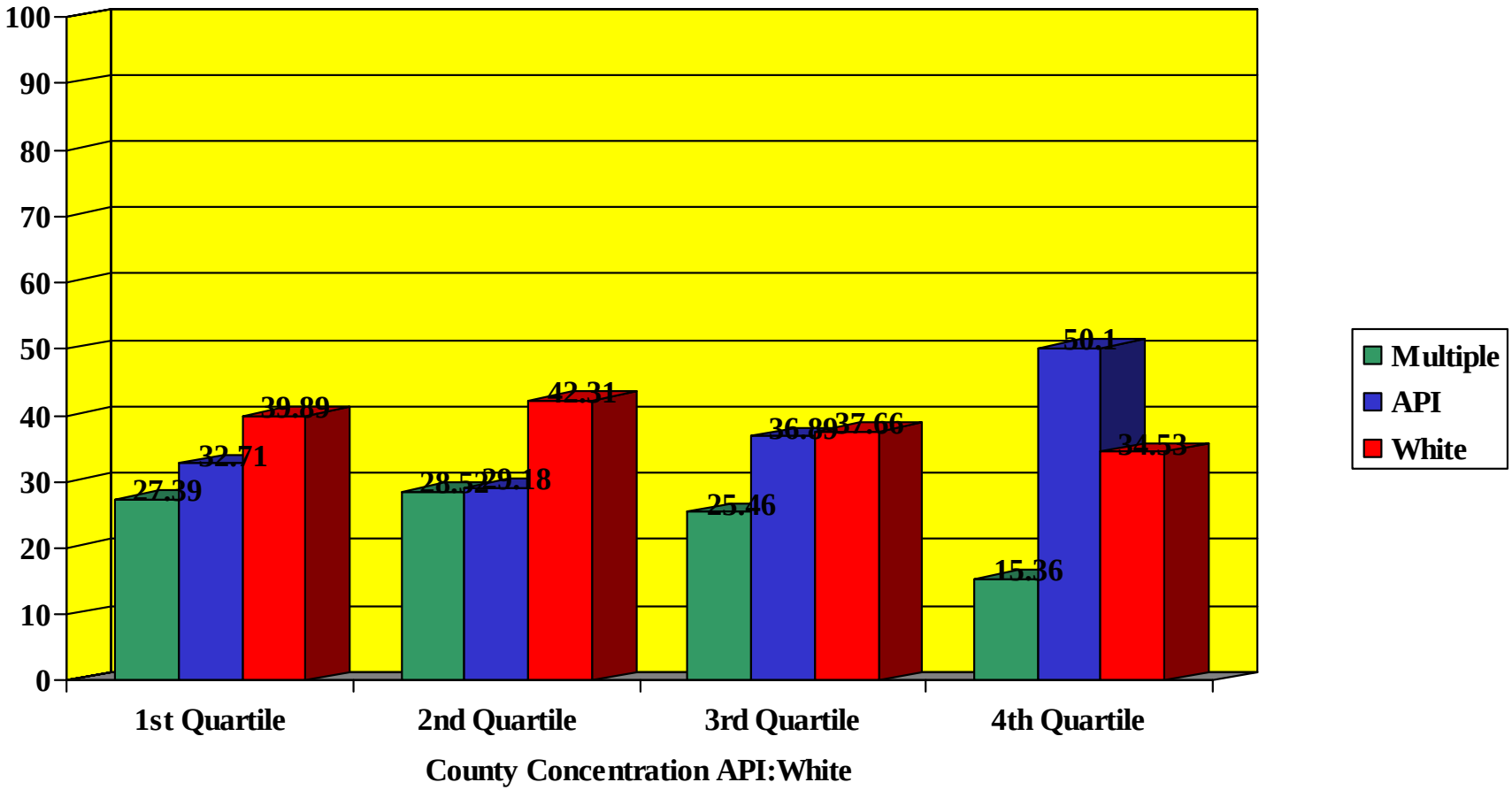
AIAN-WHITE NON-HISPANIC MALES, AGE < 18, HIGH SES, NORTHEAST



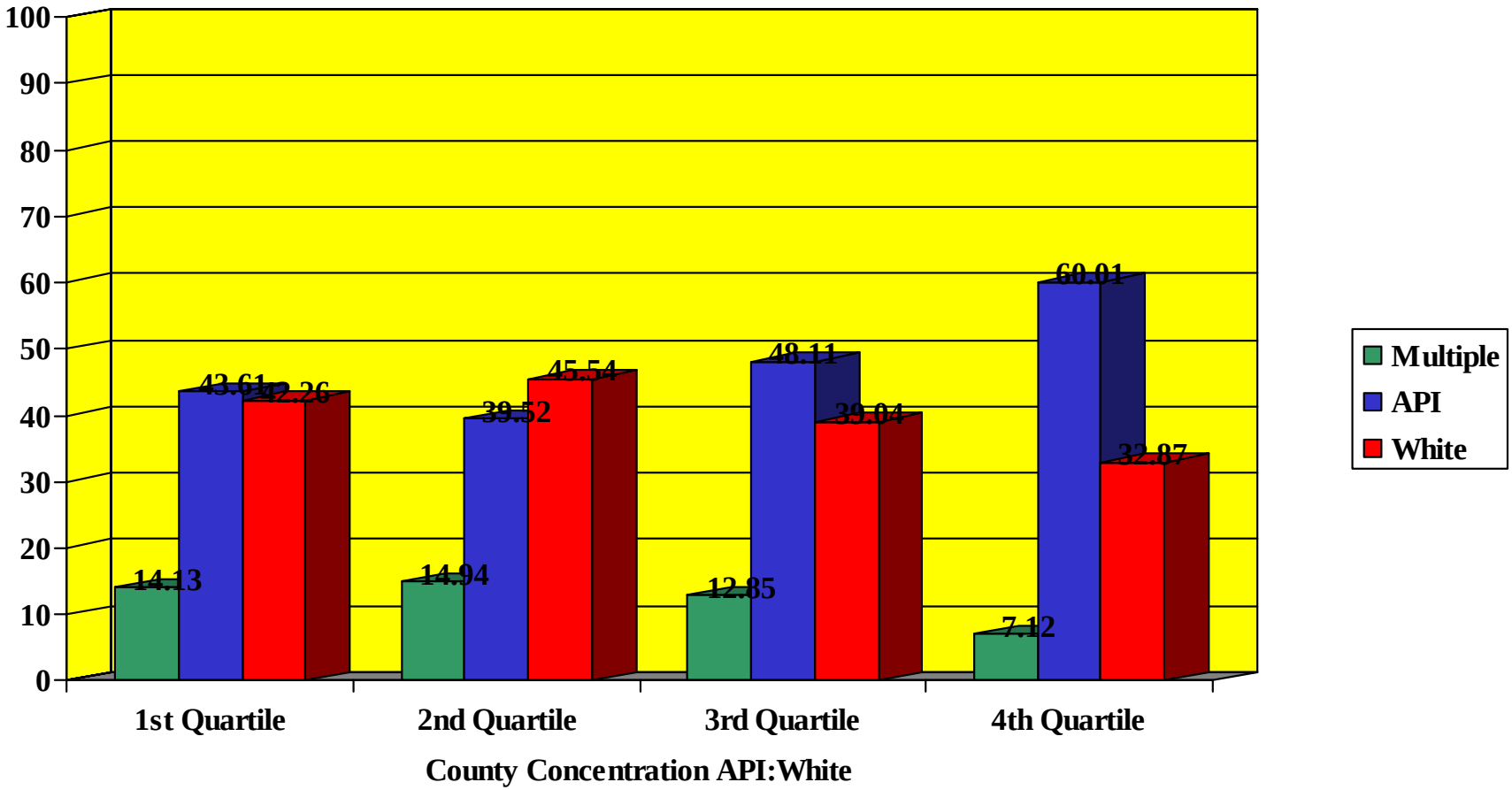
AIAN-WHITE NON-HISPANIC MALES, AGE < 18, HIGH SES, WEST



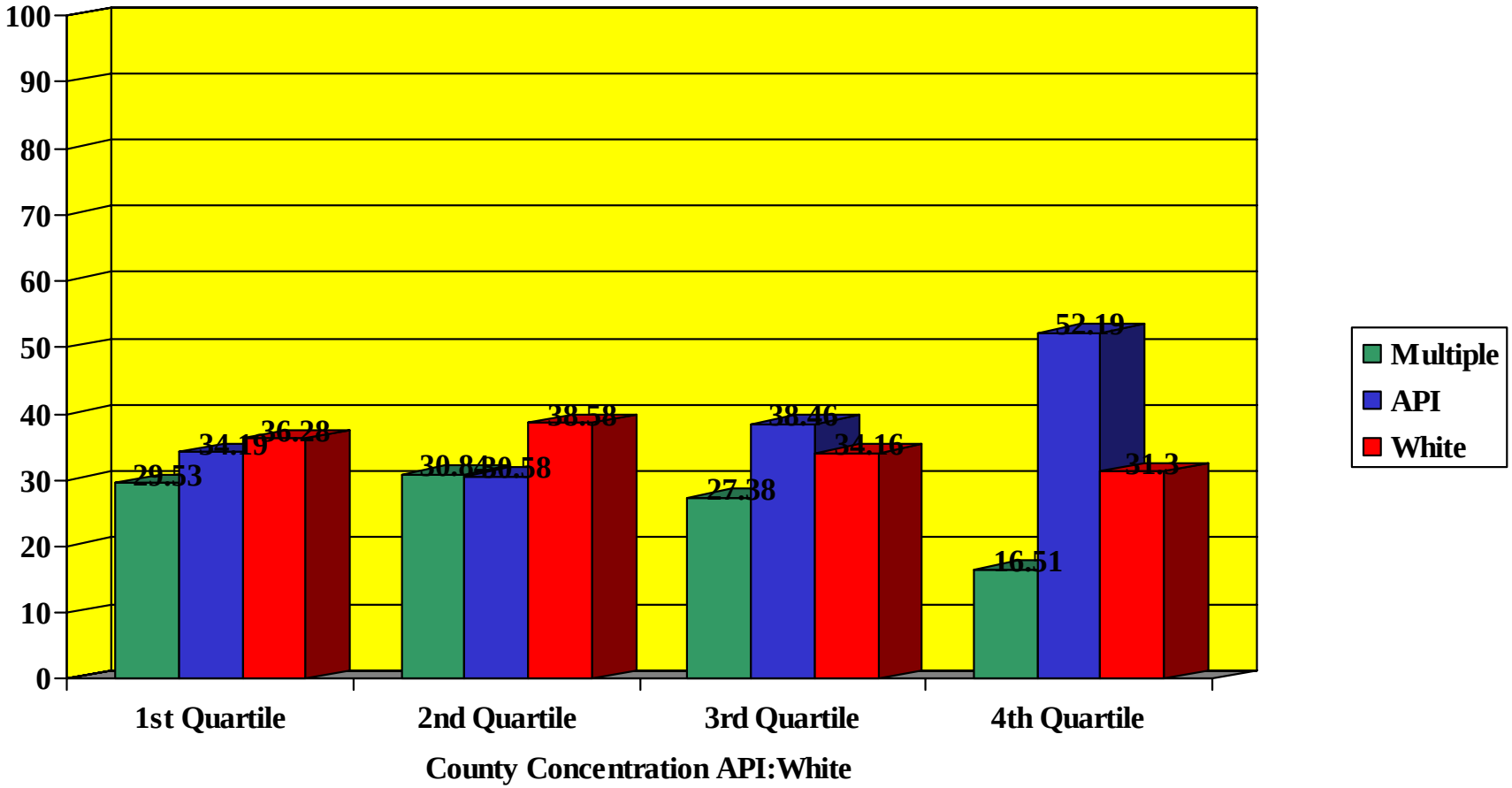
API-WHITE HISPANIC MALES, AGE < 18, LOW SES, NORTHEAST



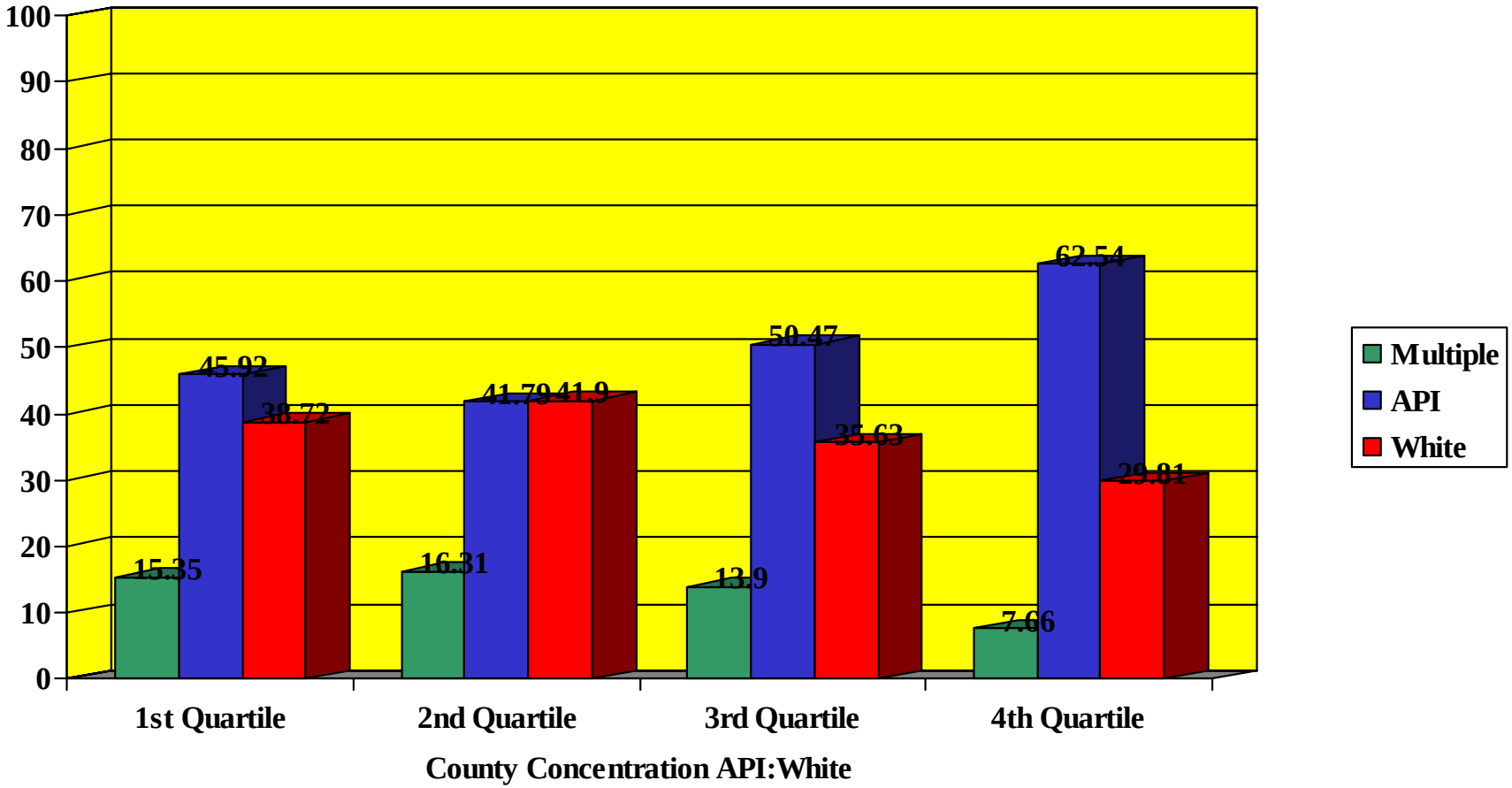
API-WHITE HISPANIC MALES, AGE < 18, LOW SES, WEST



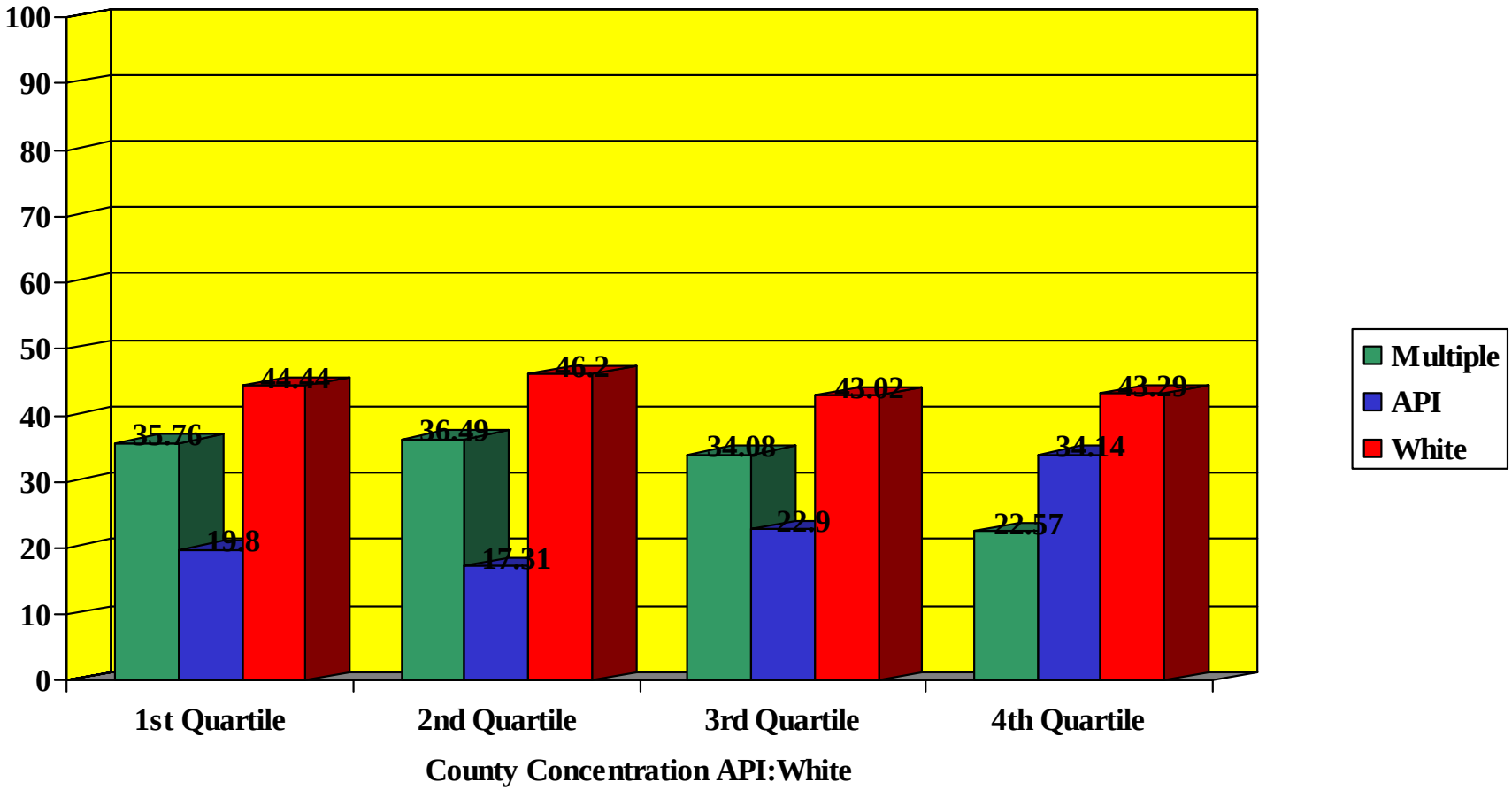
API-WHITE HISPANIC MALES, AGE < 18, HIGH SES, NORTHEAST



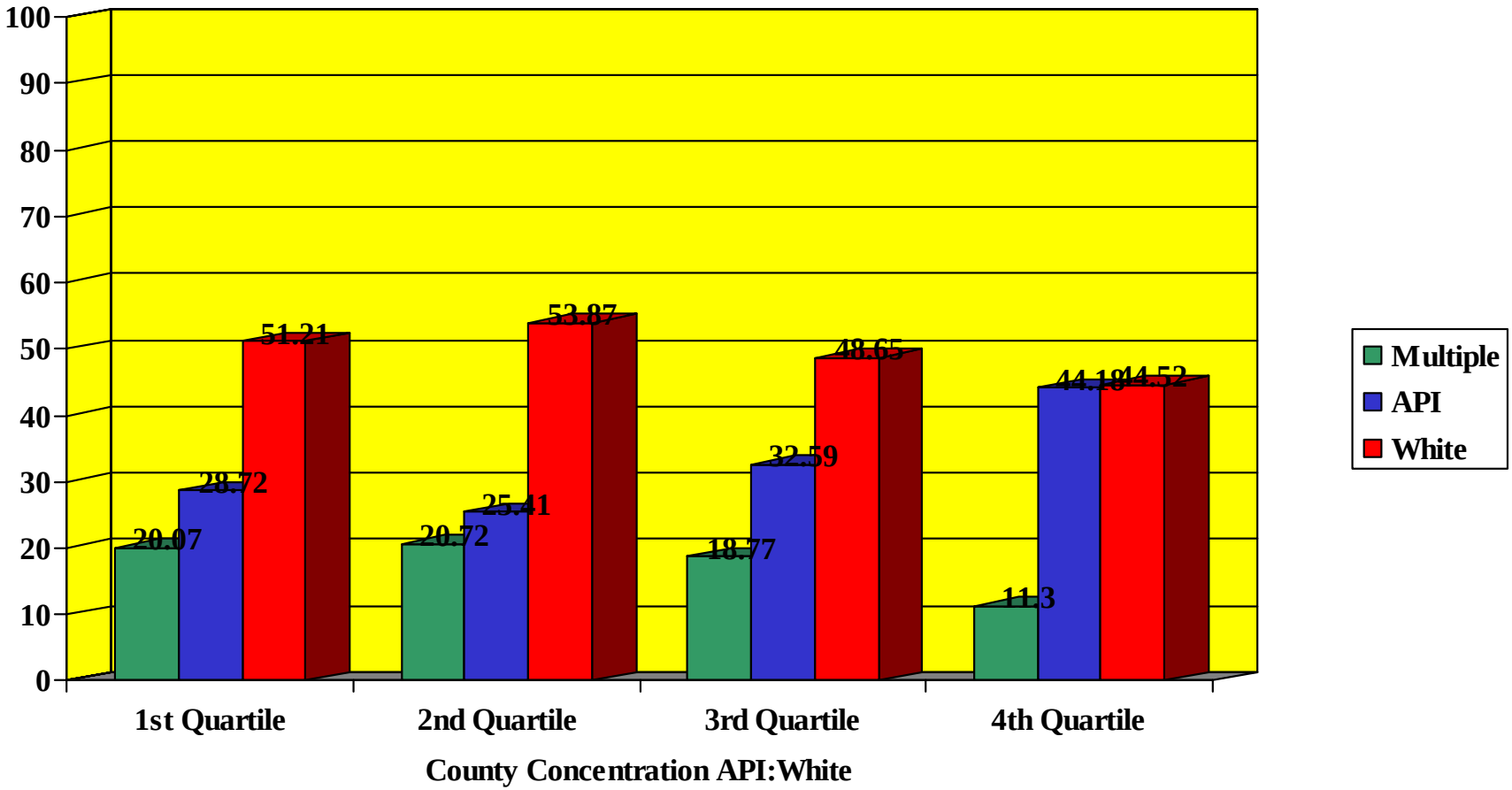
API-WHITE HISPANIC MALES, AGE < 18, HIGH SES, WEST



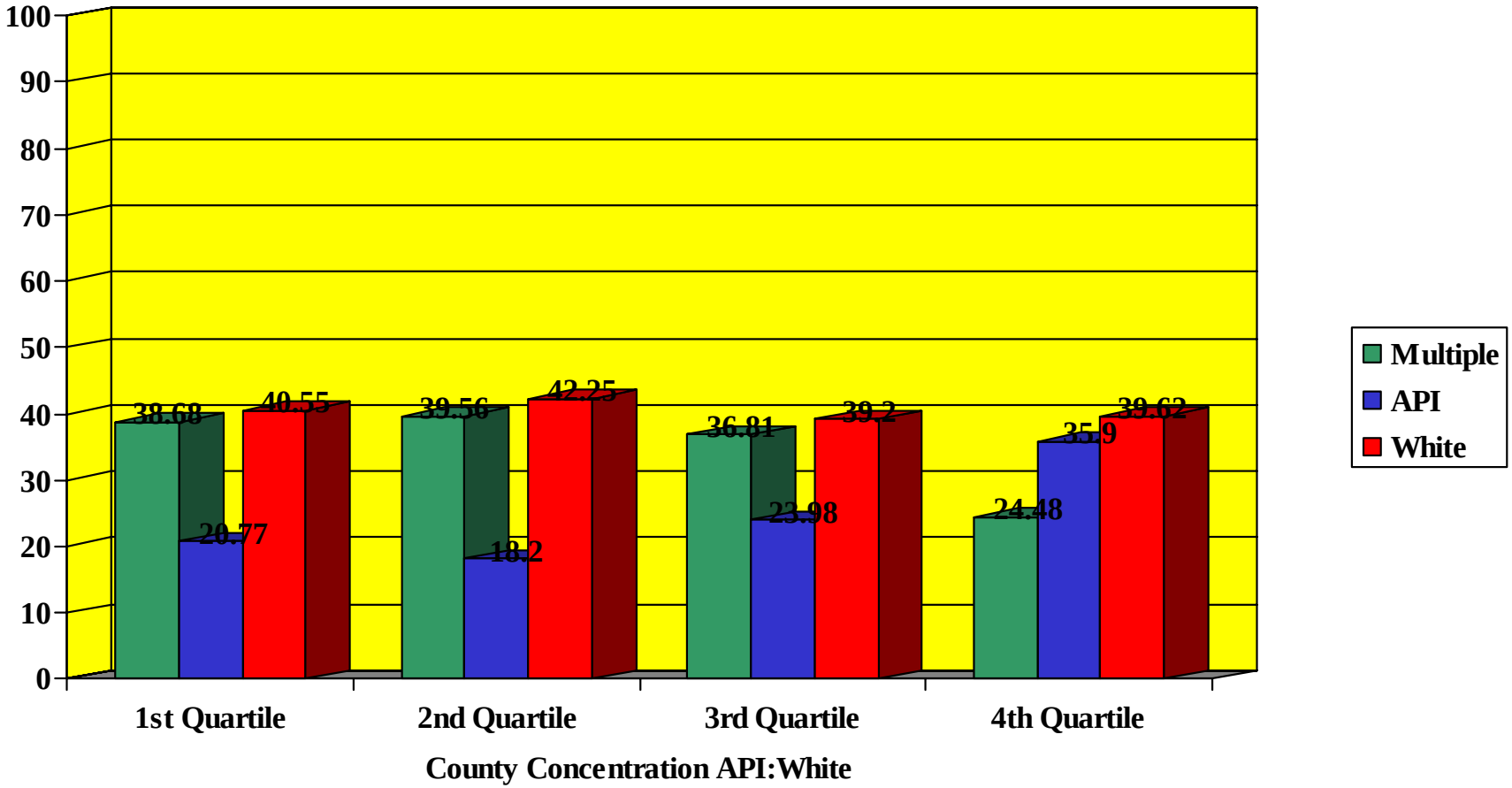
API-WHITE NON-HISPANIC MALES, AGE < 18, LOW SES, NORTHEAST



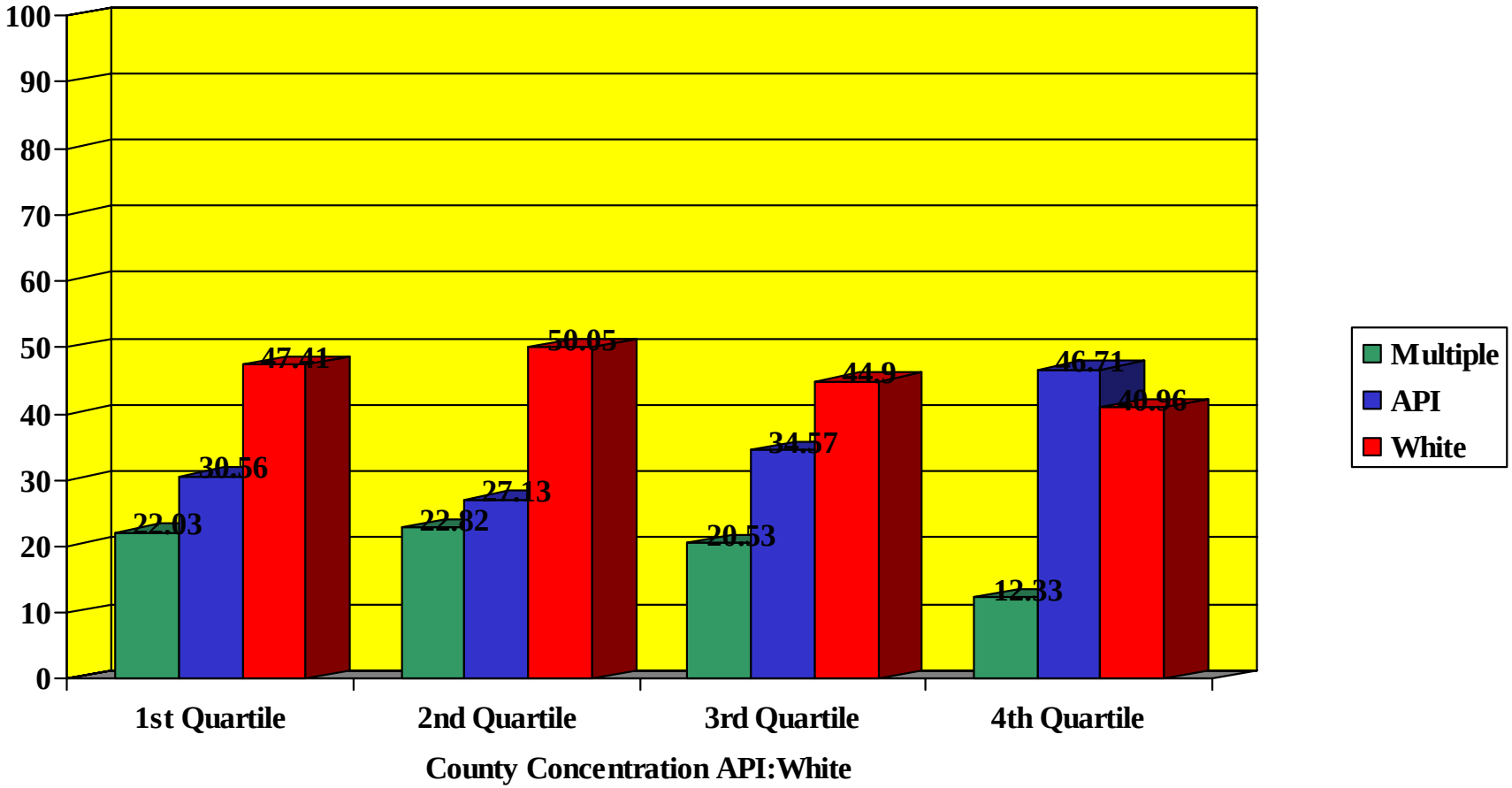
API-WHITE NON-HISPANIC MALES, AGE < 18, LOW SES, WEST



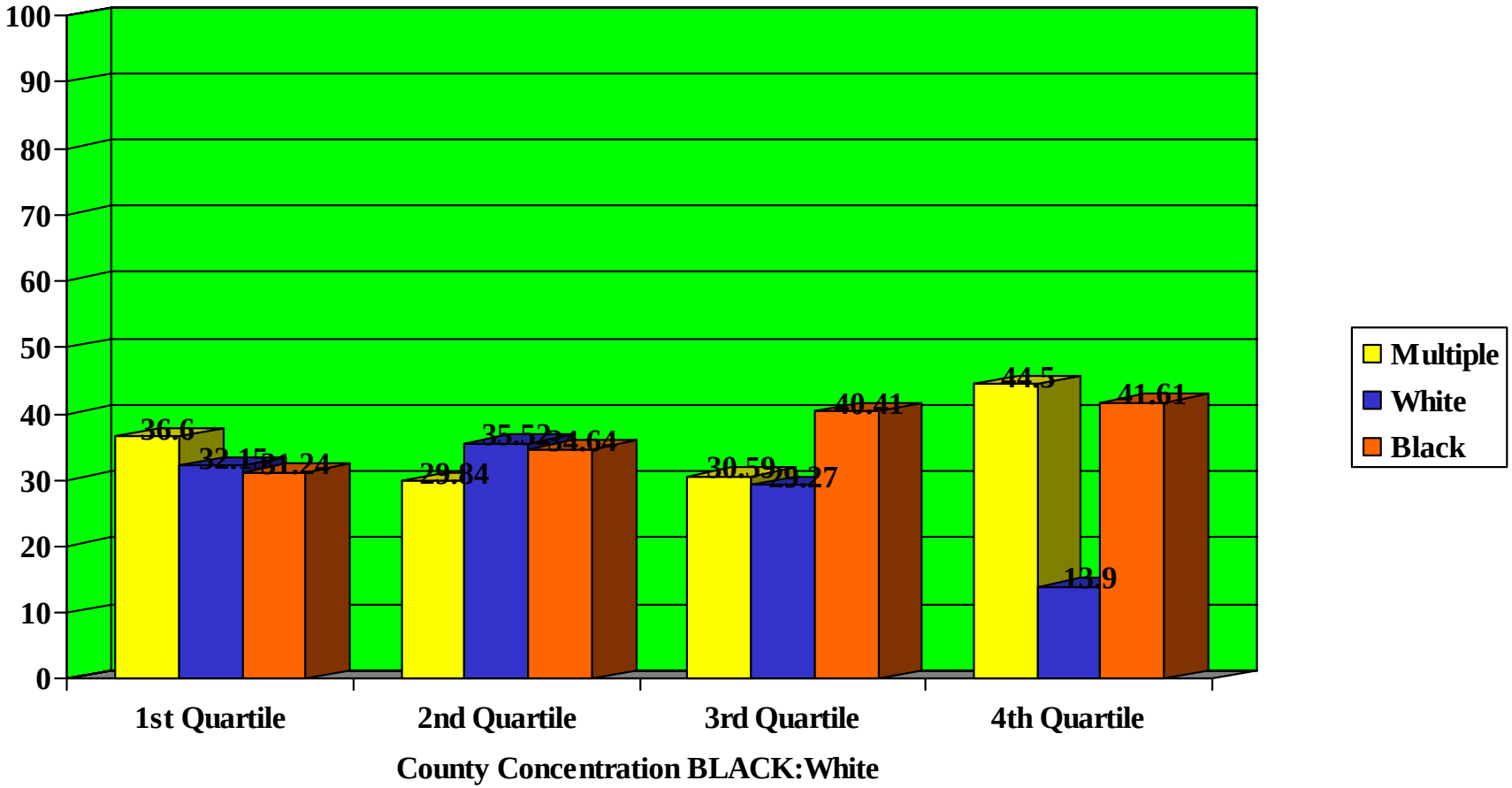
API-WHITE NON-HISPANIC MALES, AGE < 18, HIGH SES, NORTHEAST



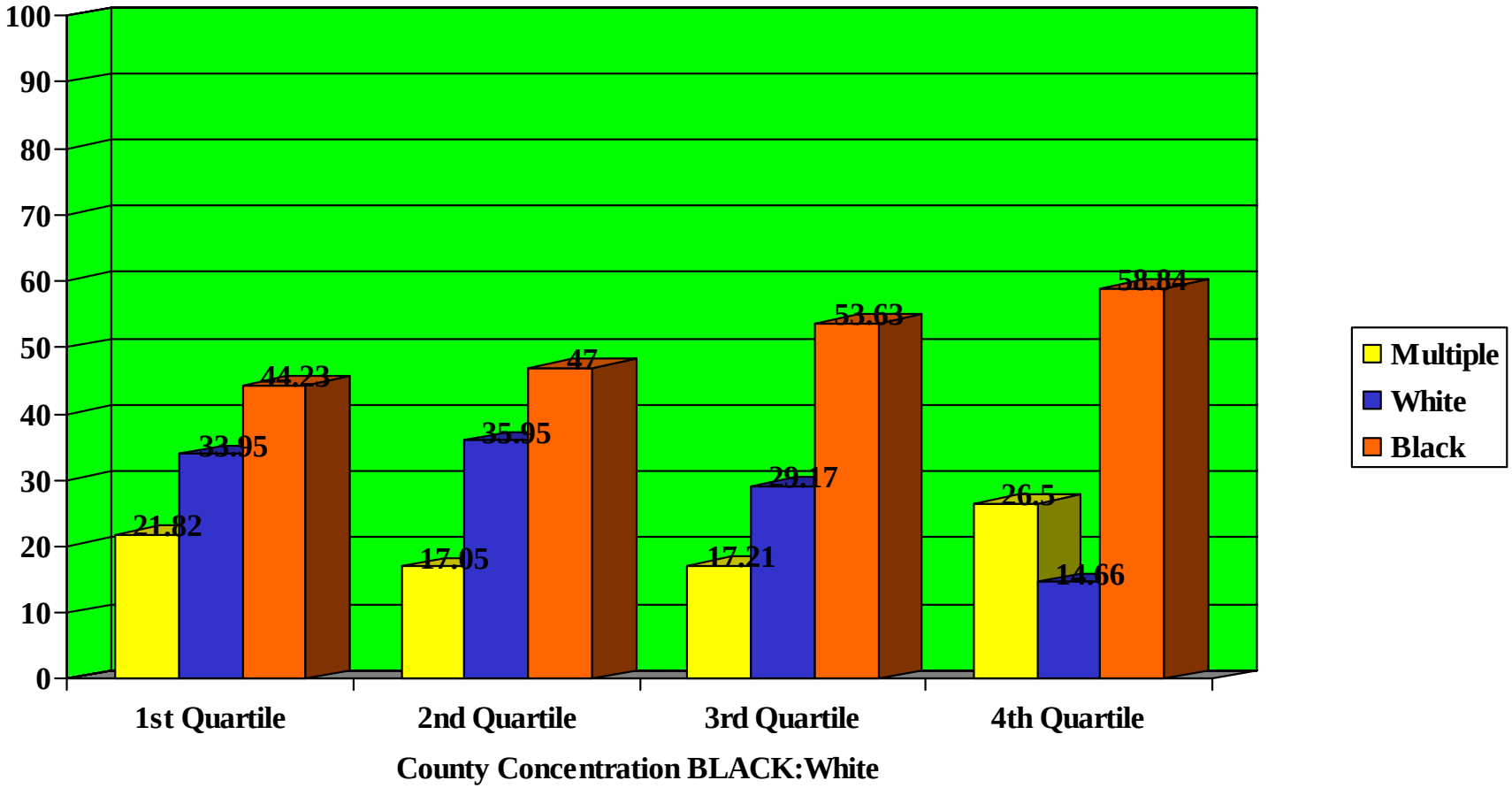
API-WHITE NON-HISPANIC MALES, AGE < 18, HIGH SES, WEST



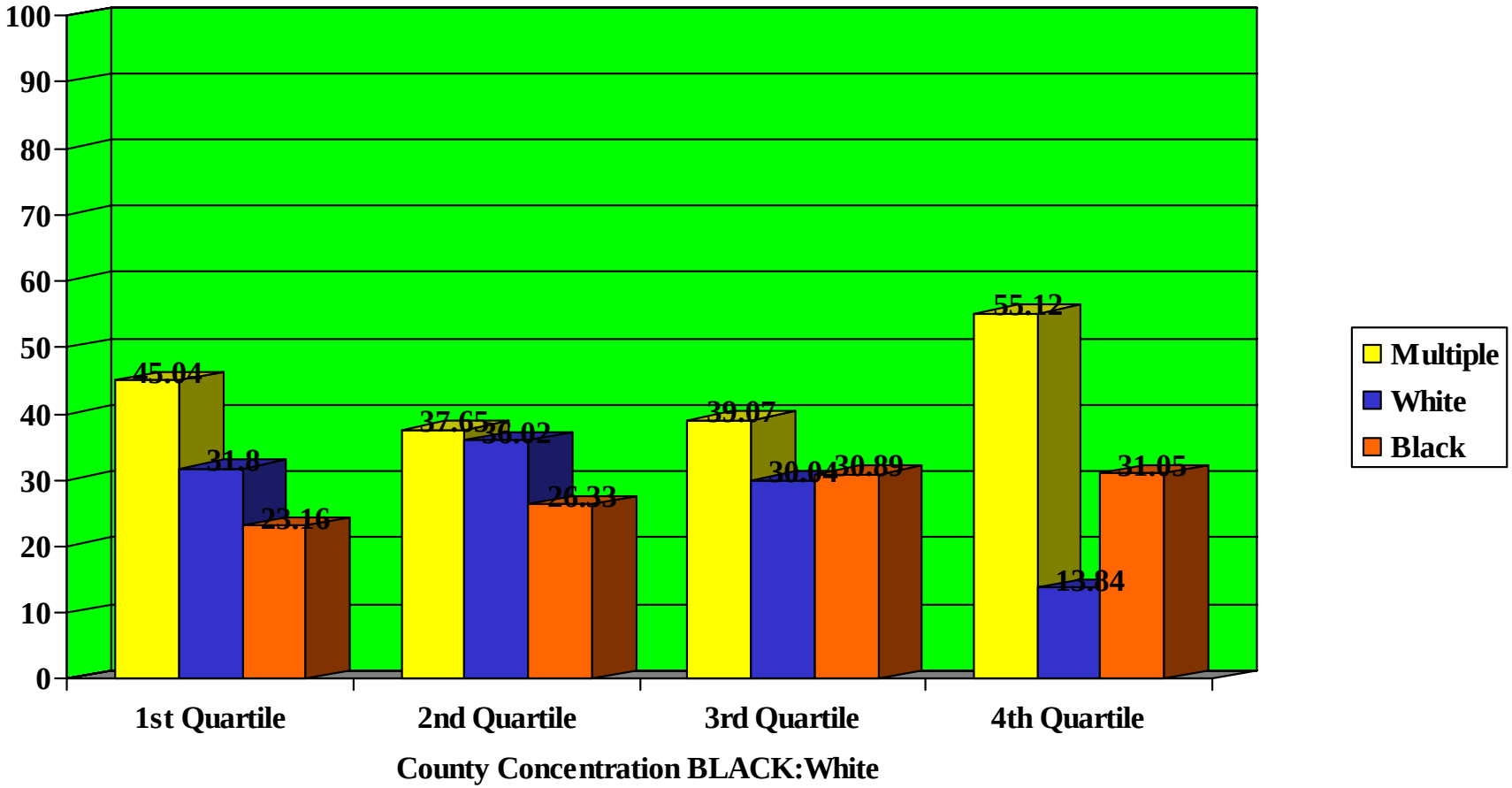
BLACK-WHITE HISPANIC MALES, AGE < 18, LOW SES, NORTHEAST



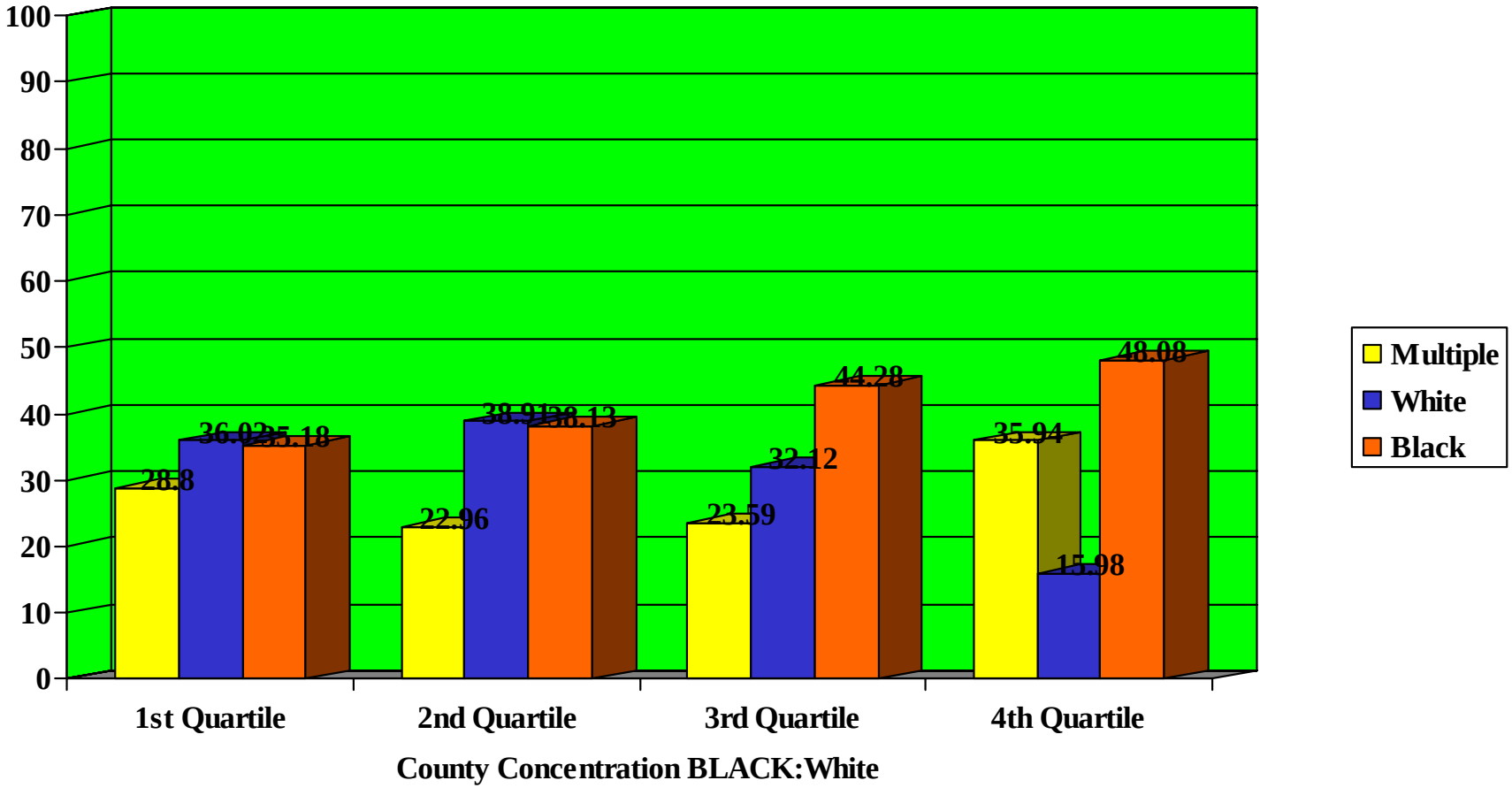
BLACK-WHITE HISPANIC MALES, AGE < 18, LOW SES, WEST



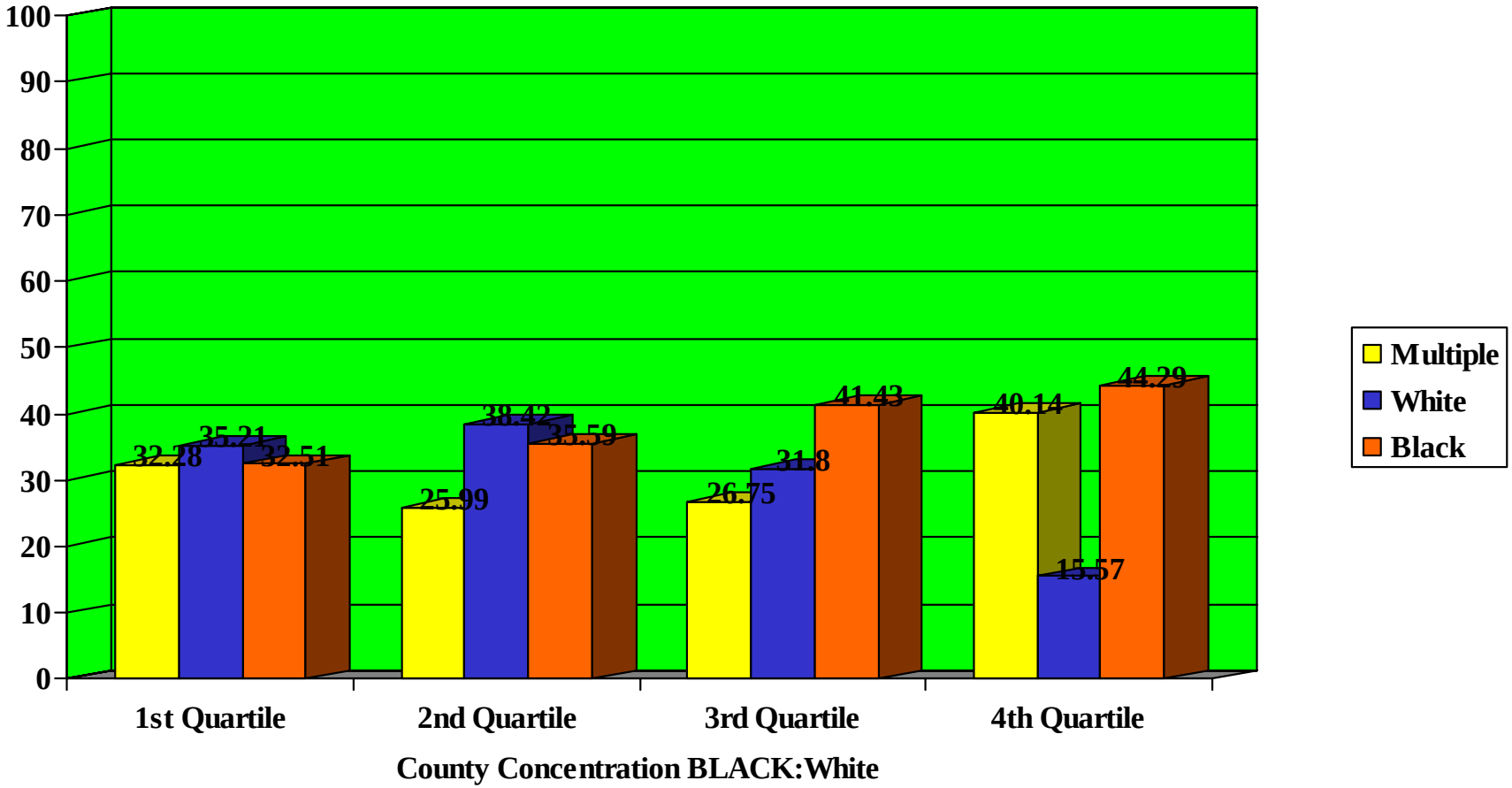
BLACK-WHITE HISPANIC MALES, AGE < 18, HIGH SES, NORTHEAST



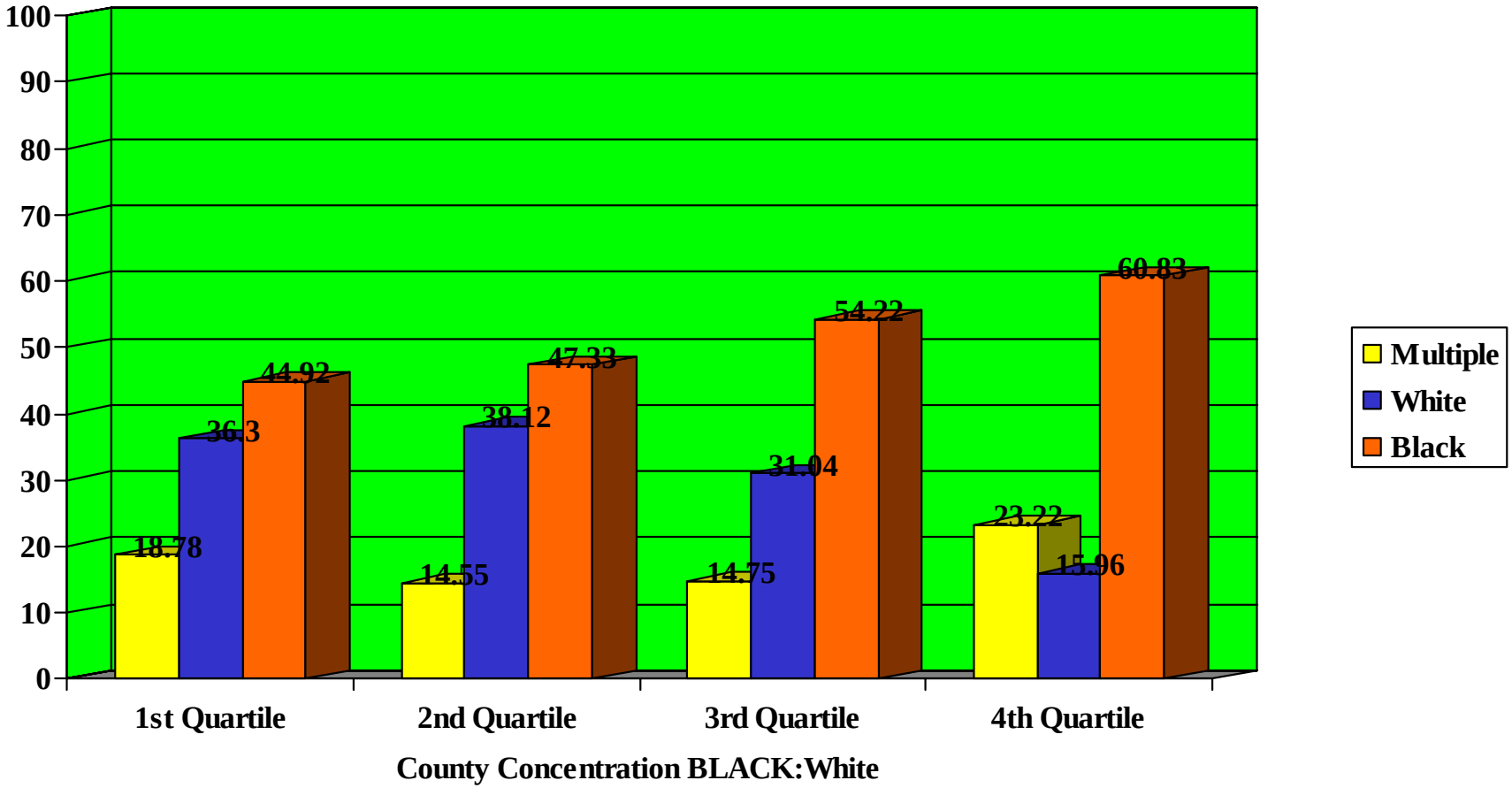
BLACK-WHITE HISPANIC MALES, AGE < 18, HIGH SES, WEST



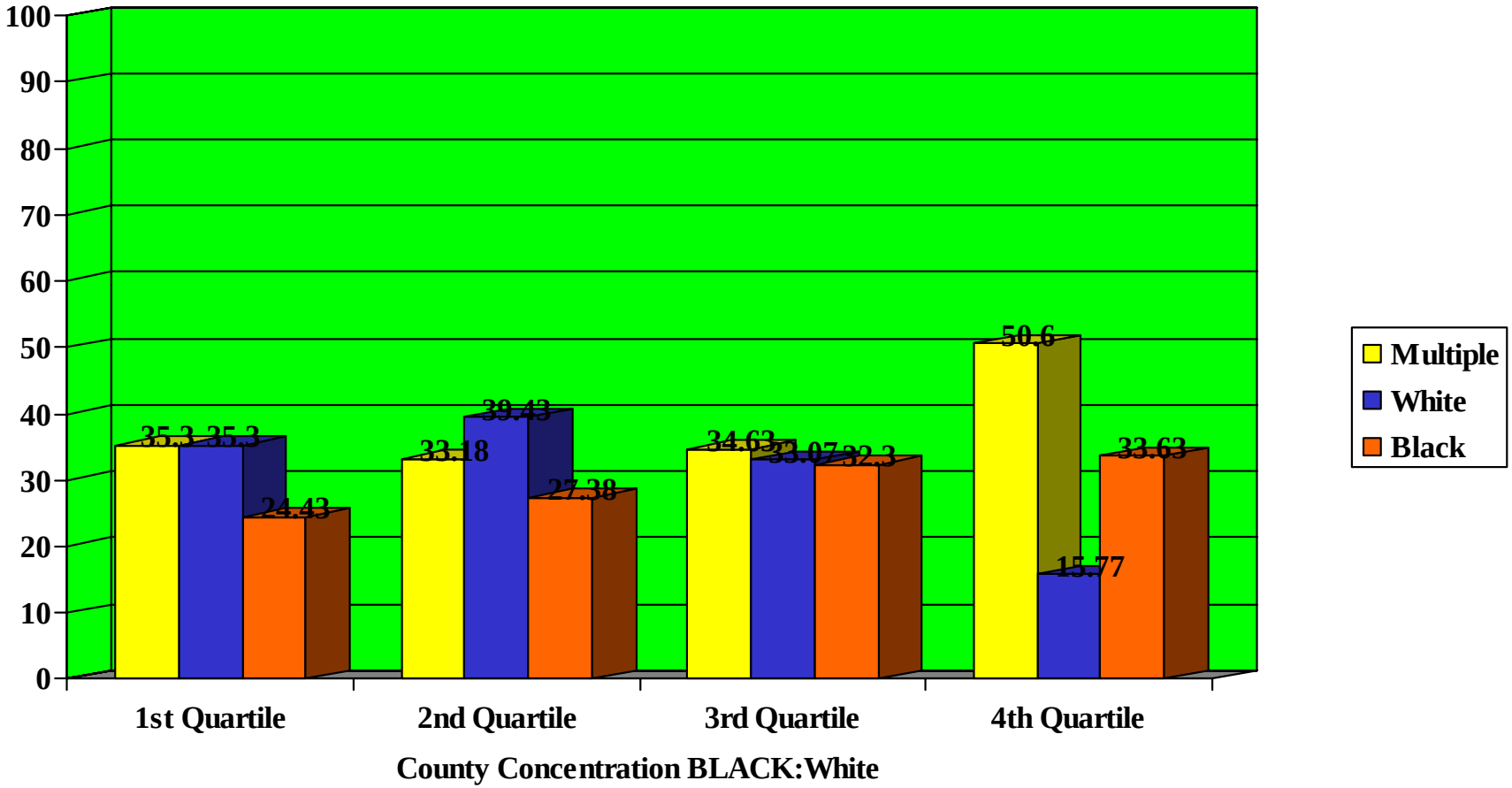
BLACK-WHITE NON-HISPANIC MALES, AGE < 18, LOW SES, NORTHEAST



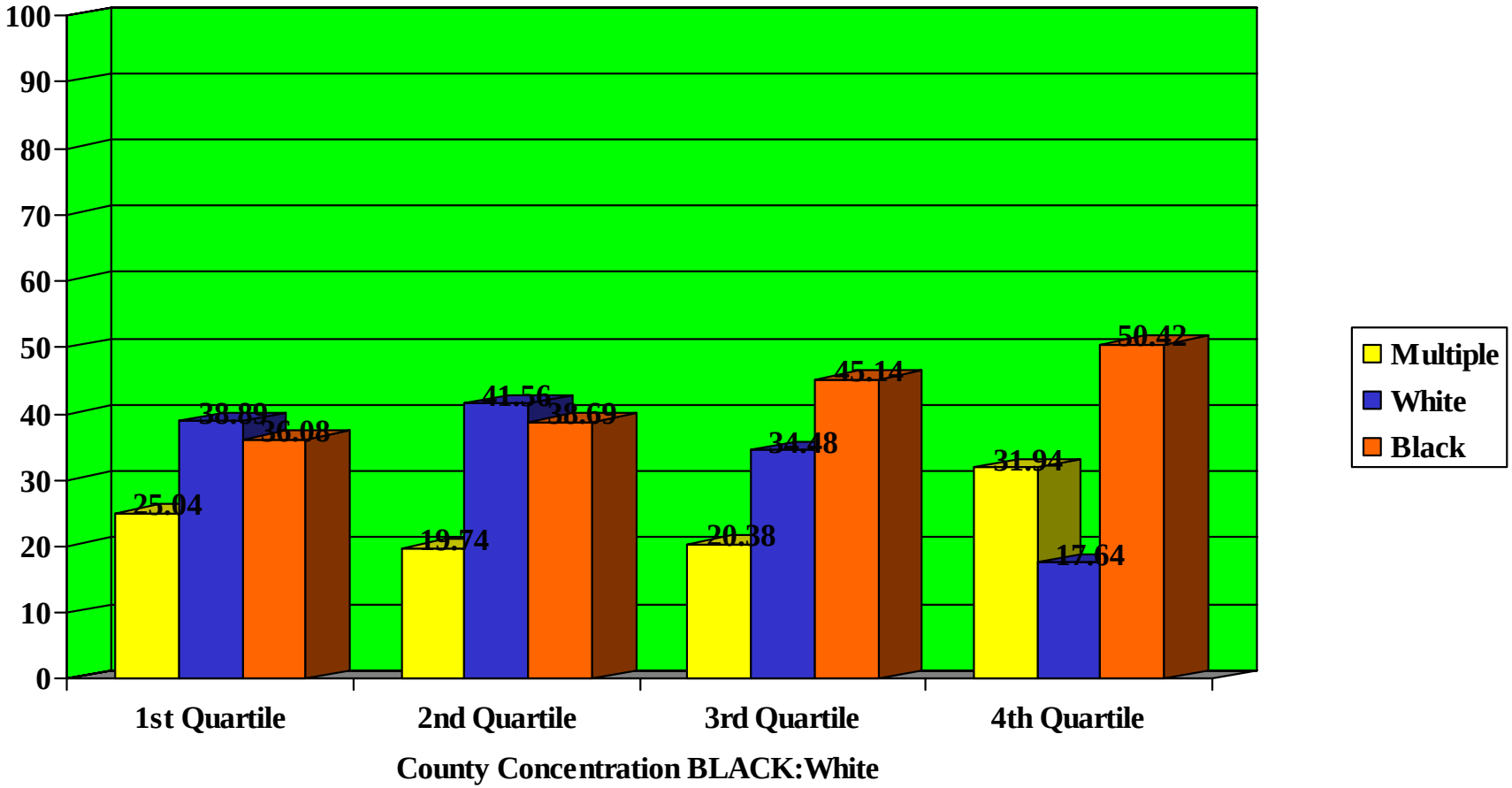
BLACK-WHITE NON-HISPANIC MALES, AGE < 18, LOW SES, WEST



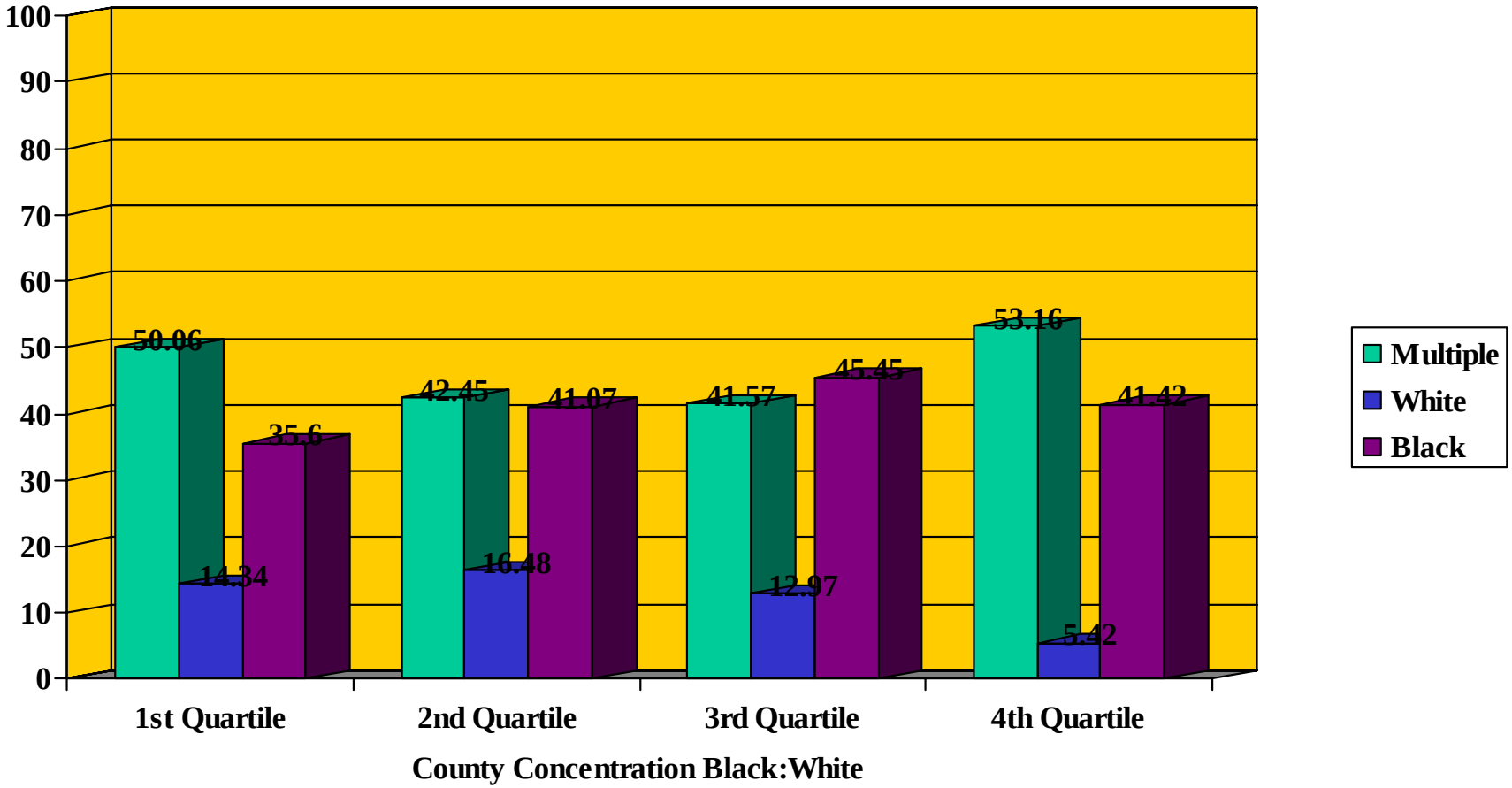
BLACK-WHITE NON-HISPANIC MALES, AGE < 18, HIGH SES, NORTHEAST



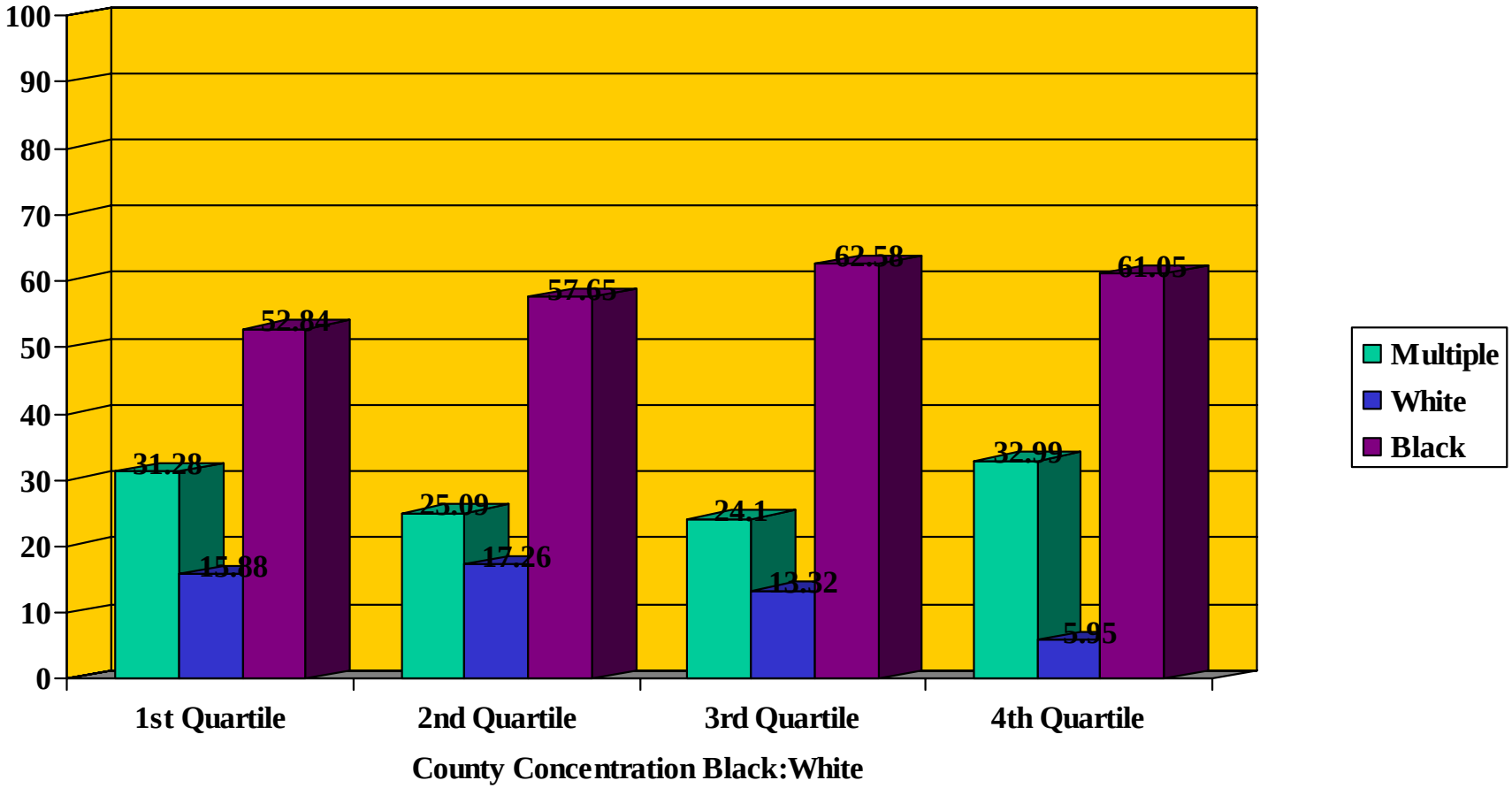
BLACK-WHITE NON-HISPANIC MALES, AGE < 18, HIGH SES, WEST



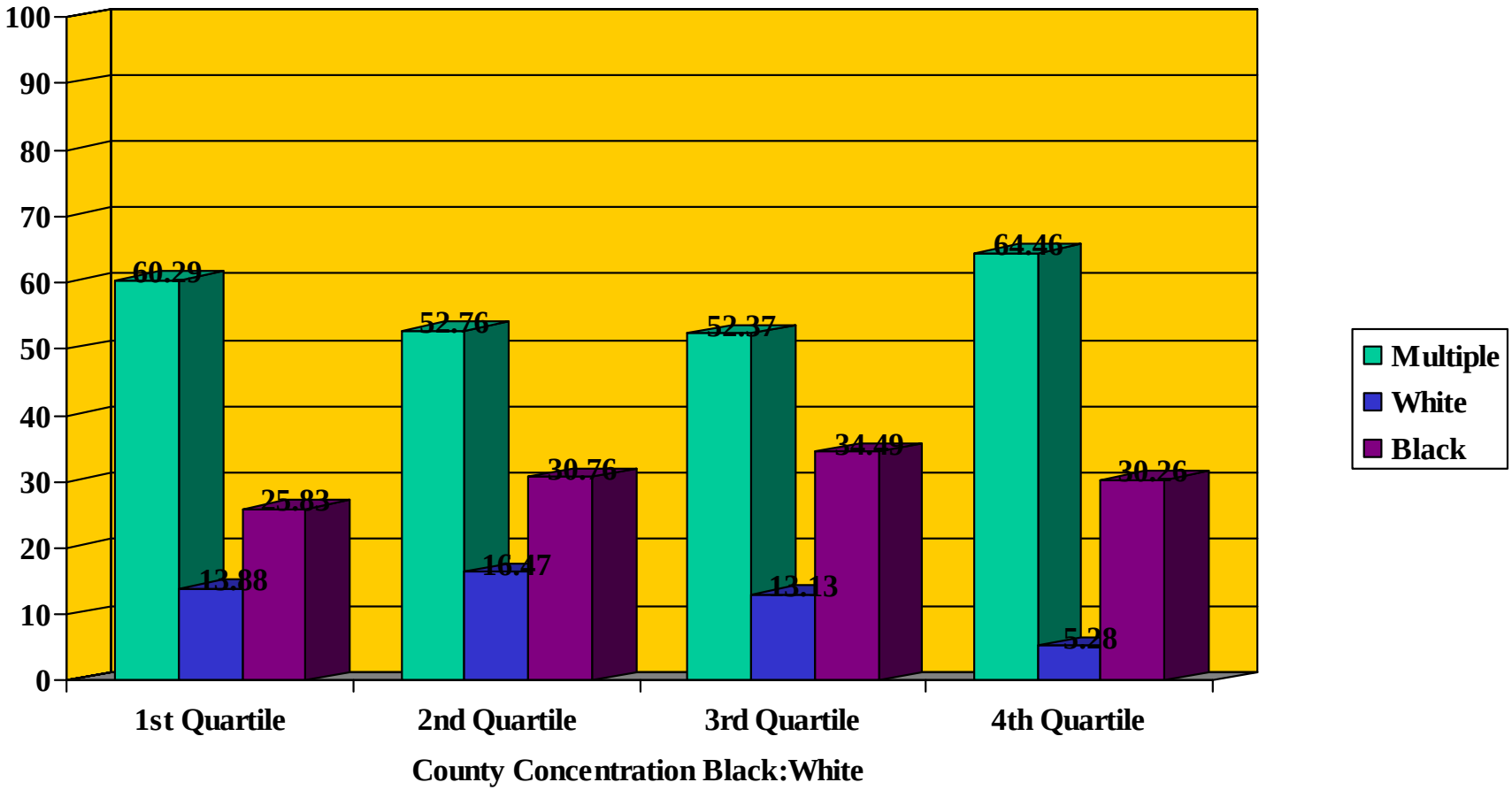
BLACK-WHITE HISPANIC MALES, AGE 45+, LOW SES, NORTHEAST



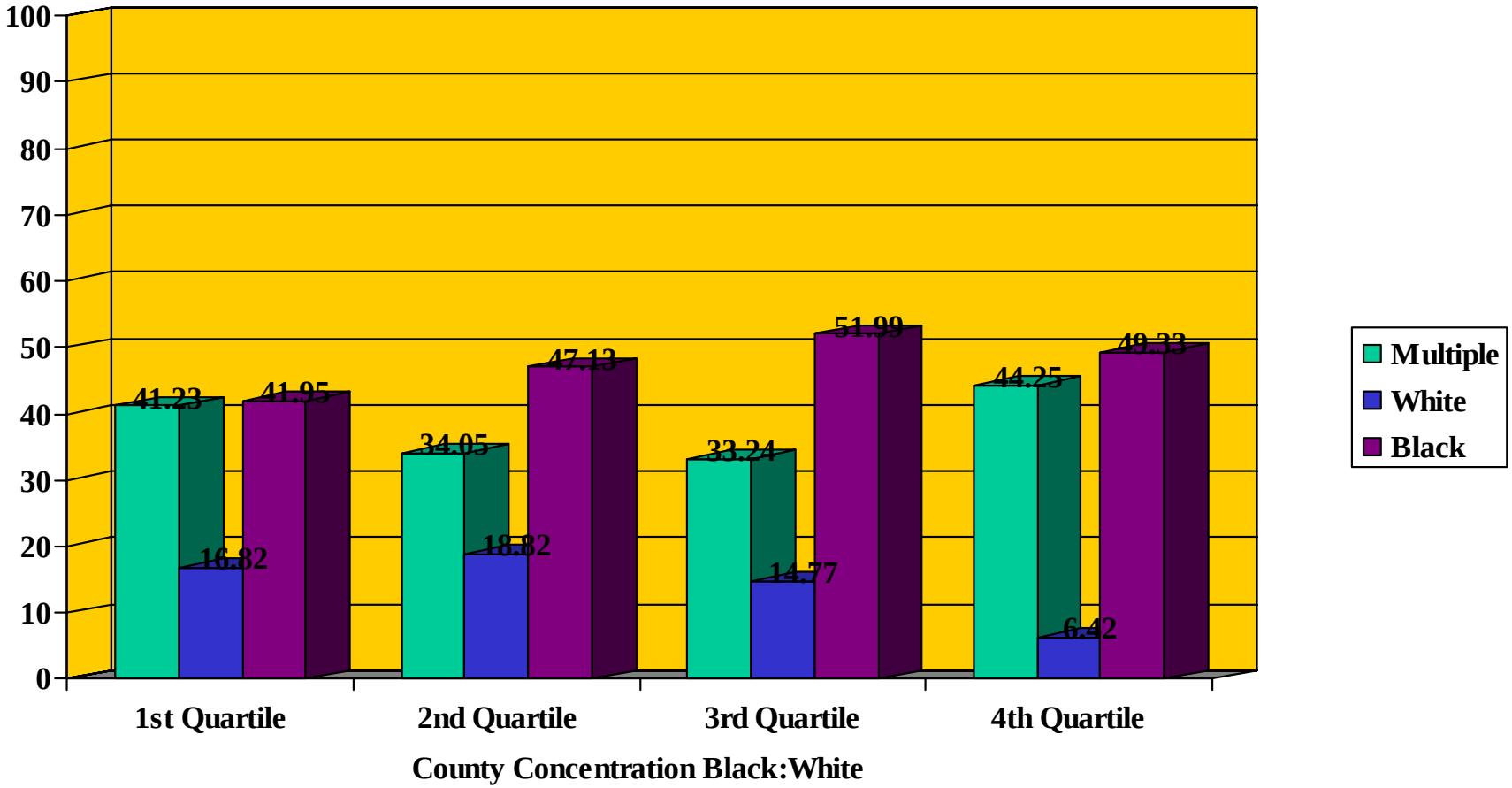
BLACK-WHITE HISPANIC MALES, AGE 45+, LOW SES, WEST



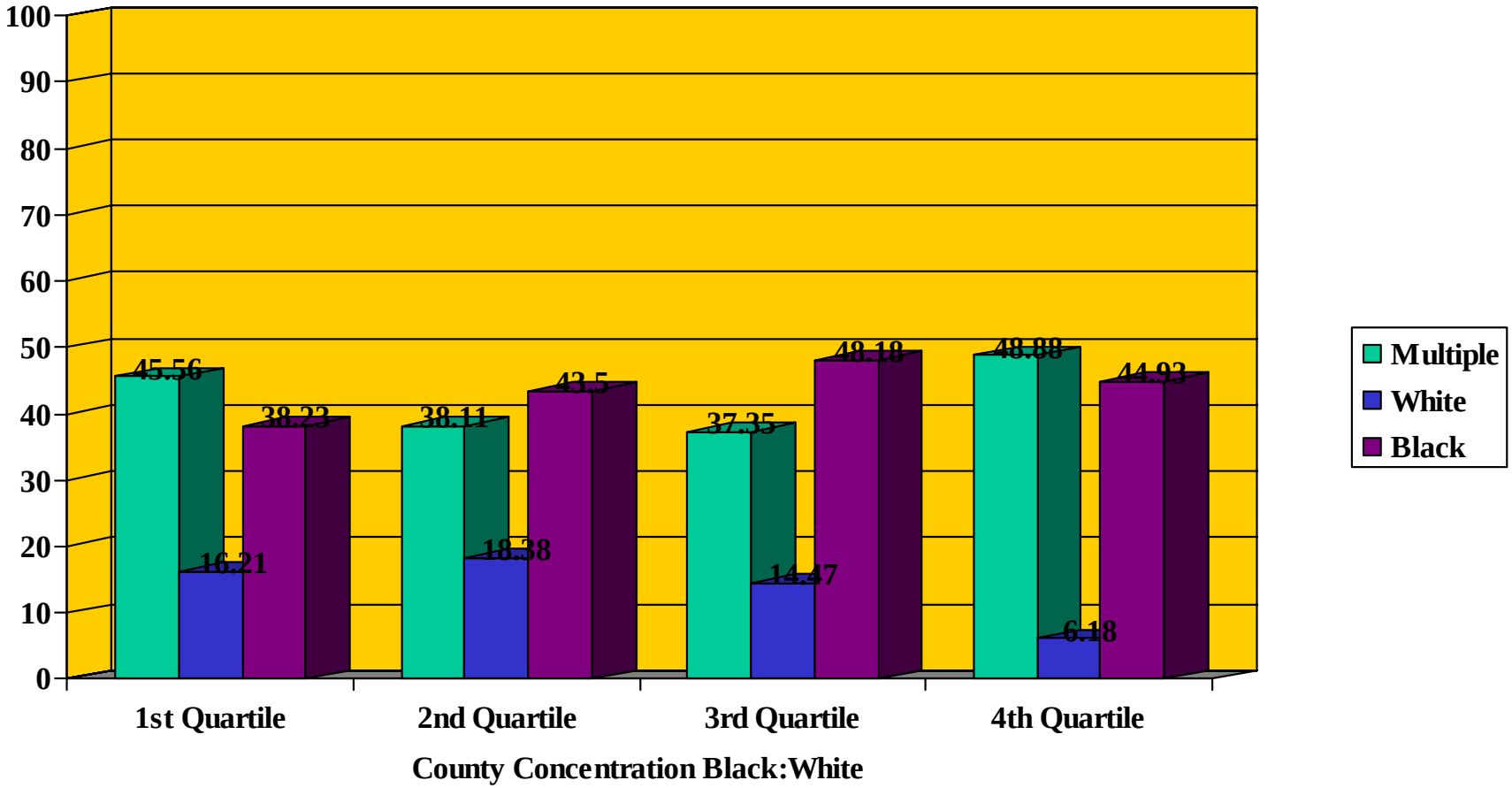
BLACK-WHITE HISPANIC MALES, AGE 45+, HIGH SES, NORTHEAST



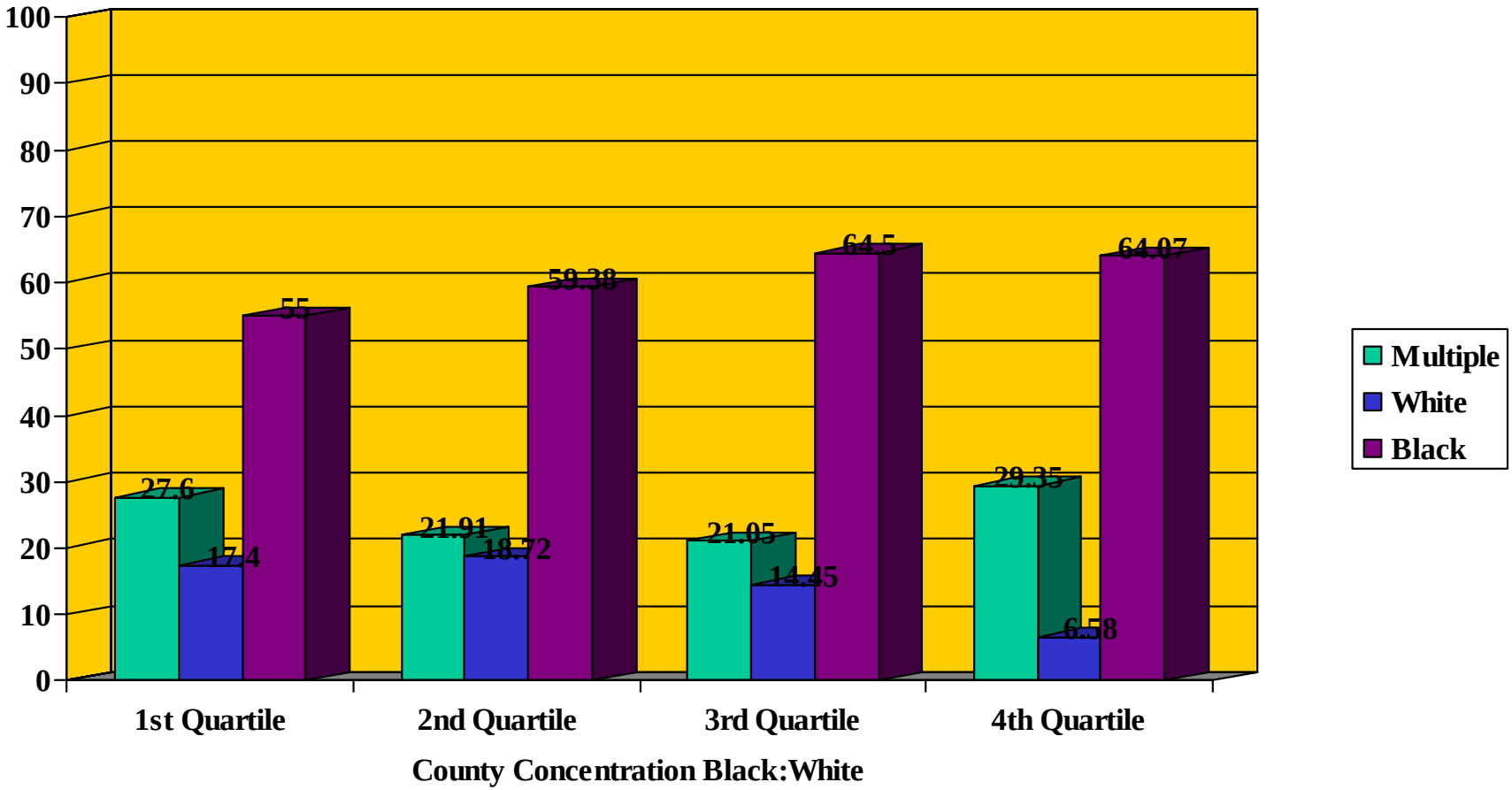
BLACK-WHITE HISPANIC MALES, AGE 45+, HIGH SES, WEST



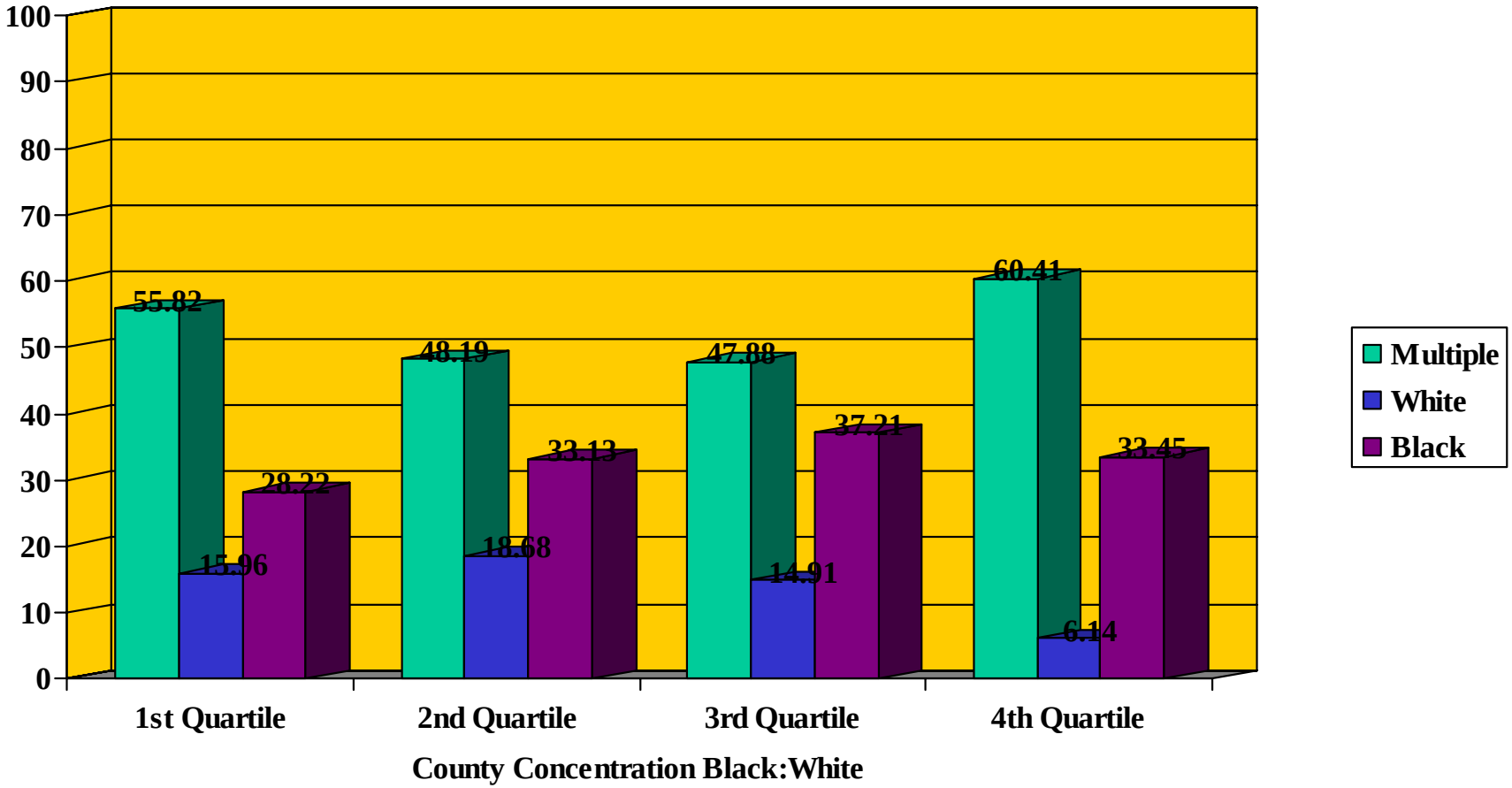
BLACK-WHITE NON-HISPANIC MALES, AGE 45+, LOW SES, NORTHEAST



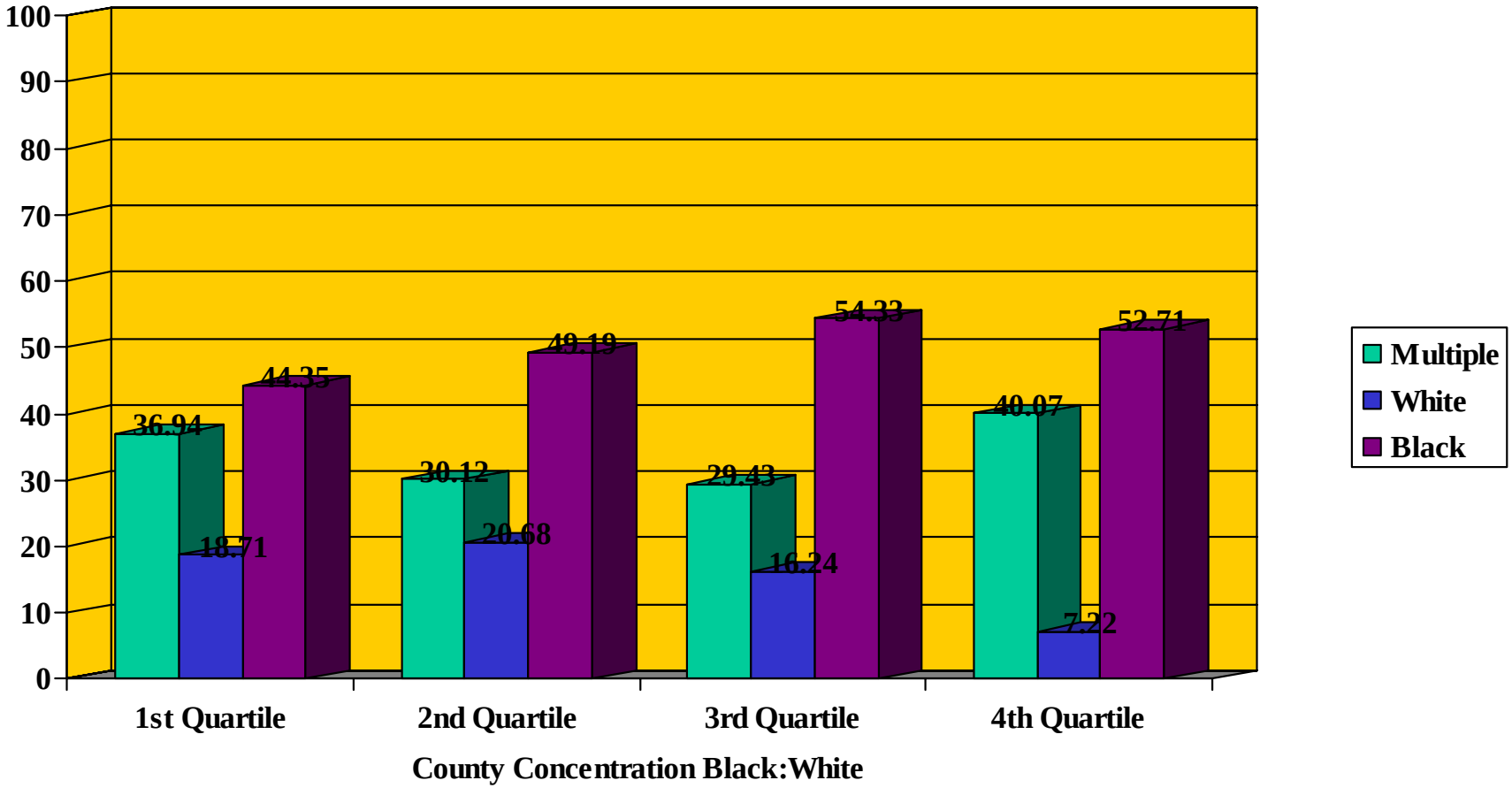
BLACK-WHITE NON-HISPANIC MALES, AGE 45+, LOW SES, WEST



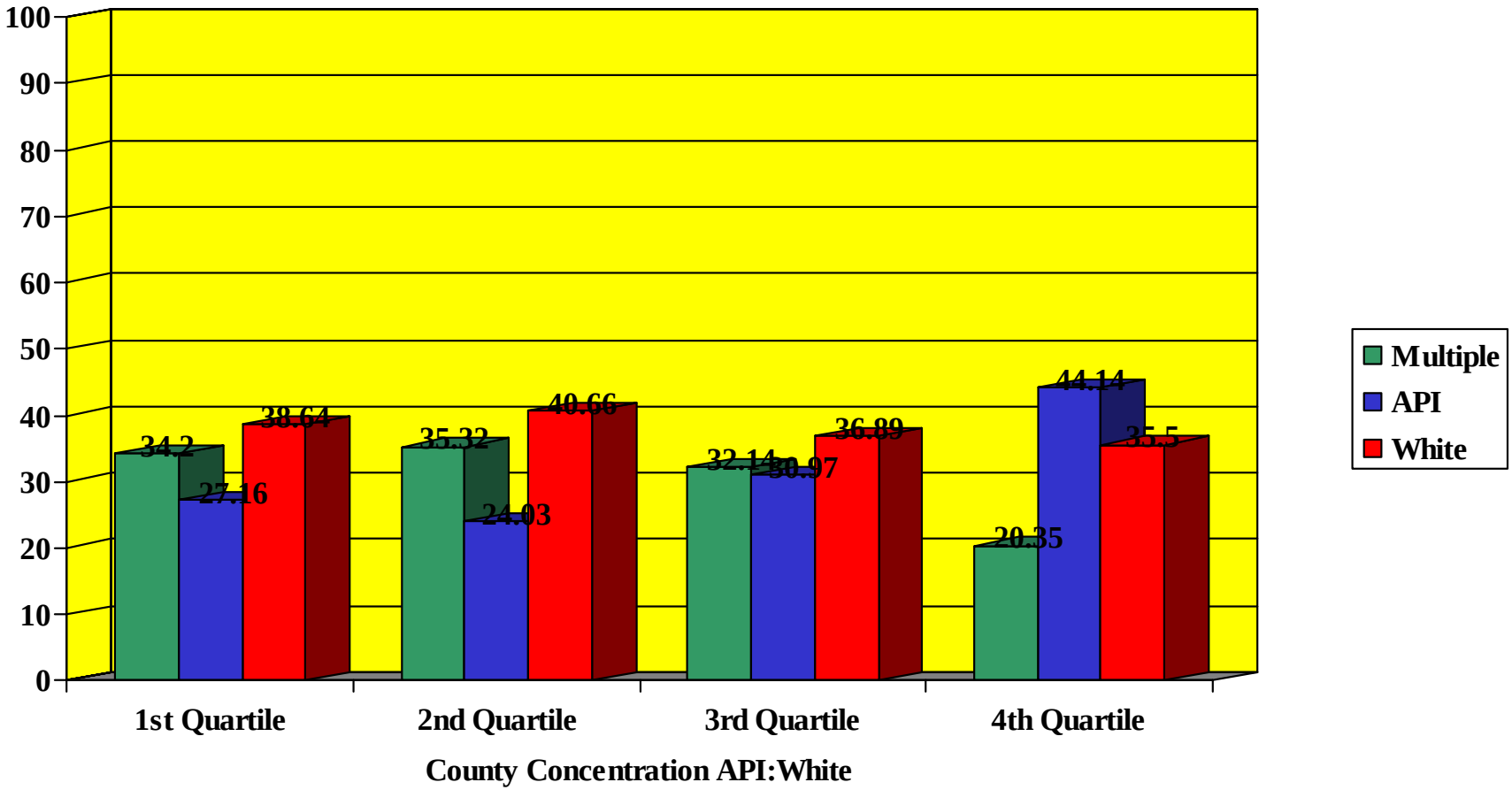
BLACK-WHITE NON-HISPANIC MALES, AGE 45+, HIGH SES, NORTHEAST



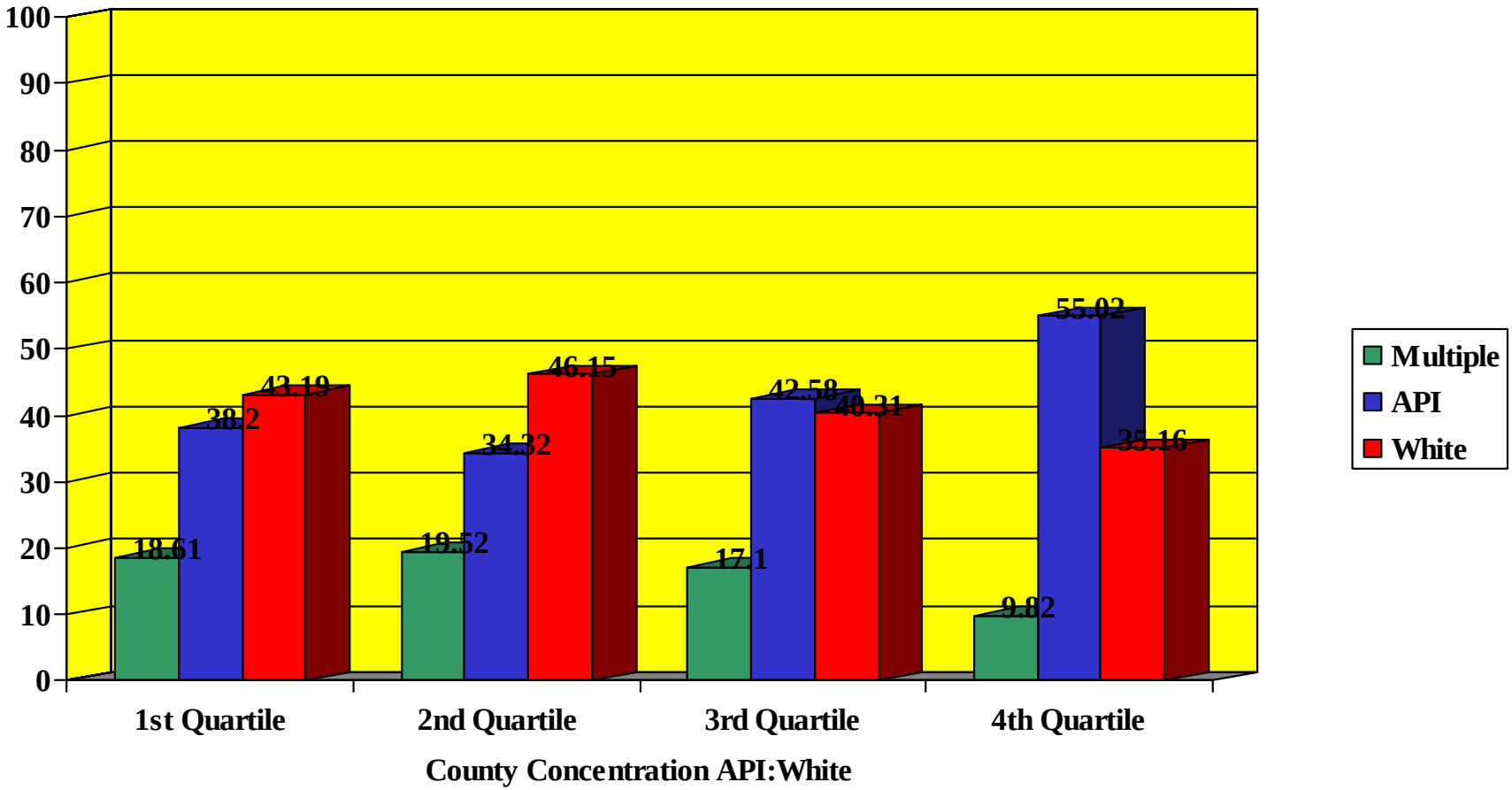
BLACK-WHITE NON-HISPANIC MALES, AGE 45+, HIGH SES, WEST



API-WHITE HISPANIC MALES, AGE 45+, LOW SES, NORTHEAST



API-WHITE HISPANIC MALES, AGE 45+, LOW SES, WEST



Summary

- The combination of factors associated with selecting a primary race group differs greatly for the various multiple race groups.
- There are also significant and important differences within each bi-racial group in the selection of ‘primary’ vs. ‘multiple’ or refusal to select a single race. For example:
- **AIAN/White**: Hispanic Origin, Family Income, Region of Residence and Concentration of Minority to Majority population at the County level all have significant net effects on the choice of AIAN vs. White as a primary race.
- Hispanic Origin, Sex and Respondent all have significant net effects on the choice of Multiple vs. White

Summary, Continued

- **API/White**: Hispanic Origin, Sex, Concentration of Minority to Majority population at the County level all have significant net effects on the choice of API vs. White.
- Region of Residence has a significant net effect on the choice of Multiple vs. White

Summary, Continued

- **Black/White**: Education, Family Income and Concentration of Minority to Majority population at the County level all have significant net effects on the choice of White vs. Black
- Sex, Family Income, and Region of Residence have significant net effects on the choice of Multiple vs. Black
- **AIAN/Black**: Only Hispanic Origin has a significant net effect on the choice of AIAN vs. Black
- Sex Family Income and Region of residence have a significant net effect on the choice of Multiple vs. Black

Summary, Continued

-The estimated probability profiles make even more clear the differences between the groups. Each group exhibits distinct profiles by Hispanic Origin, Age, SES and Minority concentration at the county-level.

-However, there are also important similarities among the groups.

-With the exception of the AIAN/Black group, Minority concentration at the county level is a consistently significant factor.

-Similarly, with the exception of the Black/White group, Hispanic Origin is a consistently significant factor.

Implications of findings

- Probability profiles suggest that developing algorithms for assigning multiple race persons to a single race group may not be straightforward:
  - Group-specific algorithms should be developed that take into account the relevant factors for each group
- The results of these analyses should be compared to those obtained from other analyses (e.g., Schenker, Parker, Ingram)
- Results should be tested on various data systems – vital statistics and survey data

Limitations

- Older NHIS data used – more recent data might produce different results (survey year controlled for in analysis)

Future analyses

- Creation of additional probability profiles for other combinations of age, sex, SES and Hispanic origin groups
- Determine what to do with individuals who refuse to select a ‘primary’ race
- Assess whether results obtained from 1986-1994 NHIS data are replicable with 1997-2000 NHIS data (Schenker, Parker and Ingram).