NAACCR Hispanic Identification Algorithm

Holly L. Howe, PhD

Chair *of the* Expert Panel on

Hispanic Identification

NAACCR Annual Meeting
June 12, 2003

North American Association of Central Cancer Registries

Census Facts

- In the 2000 United States Census, 13 percent of the population stated they were of Hispanic Origin.

- The Hispanic population has grown by 58% since 1990, while the total U.S. population has grown by only 13%.

More Census Facts

- One-half of all U.S. Hispanics live in Texas and California.
- All states are experiencing a growth in Hispanics in their population.
- Texas, Florida, and California account for about half the total growth.
- Largest percentage growth occurred in: NC, AR, GA, TN, NV, SC, AL, KY, MN, and NE.
- Mexican Hispanics comprise 59% of the U.S. Hispanics.
- Median age of Hispanics was 25.9 years *c.f.* 35.3 years in general pop.

North American Association of Central Cancer Registries

Immediate need …

- **NAACCR has had a longer history of addressing need;**
- **Method now developed so that we can use current 2000 population counts and produce cancer incidence rates for the Hispanic population in the United States**

History of NAACCR Addressing Hispanic Identification

- *1995-1997*. A UDS sub-committee was formed to consider issues surrounding Hispanic/Latino ethnicity coding in central cancer registries.

History of NAACCR Addressing Hispanic Identification

- *1998-2000.* The Collaborative Research Working Group formed a subcommittee to identify issues surrounding publication of cancer rates by racial and ethnic groups.
- They initiated an assessment project.

History: 1998-2000 Assessment Project

- **Focused on classification issues at the central registry**

- **Repeated the 1995 NAACCR survey regarding ethnicity classification procedures at the central registry**

History: 1998 Survey Results

- **Registries used different methods to determine Hispanic ethnicity; some did not follow NAACCR guidelines.**

- **Use of death certificates varied: some used the DC as a gold standard & others used it only to supplement missing data.**

- **Race, birthplace, & maiden name were used differently.**

- **Registries with a protocol were confident in their data.**

- **Use of surname lists and algorithms varied.**

2001-2003. An Expert Panel for Hispanic identification convened.

- **Represented by states with ~57% U.S. Hispanics:**

Los Angeles and the Greater San Francisco Bay Area (GBA) of California, Colorado, Texas, Illinois, New York, and New Jersey.

North American Association of Central Cancer Registries

Expert Panel

- **Assess the reliability of a method to enhance Hispanic identification; and**
- **Determine if a recommendation for a uniform method was feasible among all central cancer registries**

Method

- **The Illinois State Cancer Registry Hispanic Algorithm procedure was selected as the standard approach**
  - theoretically robust
  - published in scientific in sufficient detail
- **Comparability/ reliability of the outcomes produced by the various registries could be assessed in conjunction with the registry's method for enhancing identification of Hispanic persons**

North American Association of Central Cancer Registries

Method

- Total counts for all 1998-1999 cancers and 10 selected cancer sites by gender were generated for Hispanics.

- First, using information on original reports;

- Second, using current registry protocol to enhance Hispanic identification of cases.

- Third, by applying the Illinois State Cancer Registry Hispanic Algorithm.

- **2000 US standard population was used in the calculation of age-adjusted rates.**

CANCER SITES

- All Sites
- Lung
- Cervix
- Colo-rectal
- Prostate
- Stomach

- Breast (female only)
- Bladder
- Non-Hodgkin lymphoma
- Gall bladder
- Liver

Age Distribution of Estimated Hispanic (all races) Population, 1998-1999, Males

Legend:
- Colorado
- Greater Bay Area
- Illinois
- Los Angeles
- New Jersey
- New York
- Texas

Y-axis: %

X-axis: Age Groups (0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
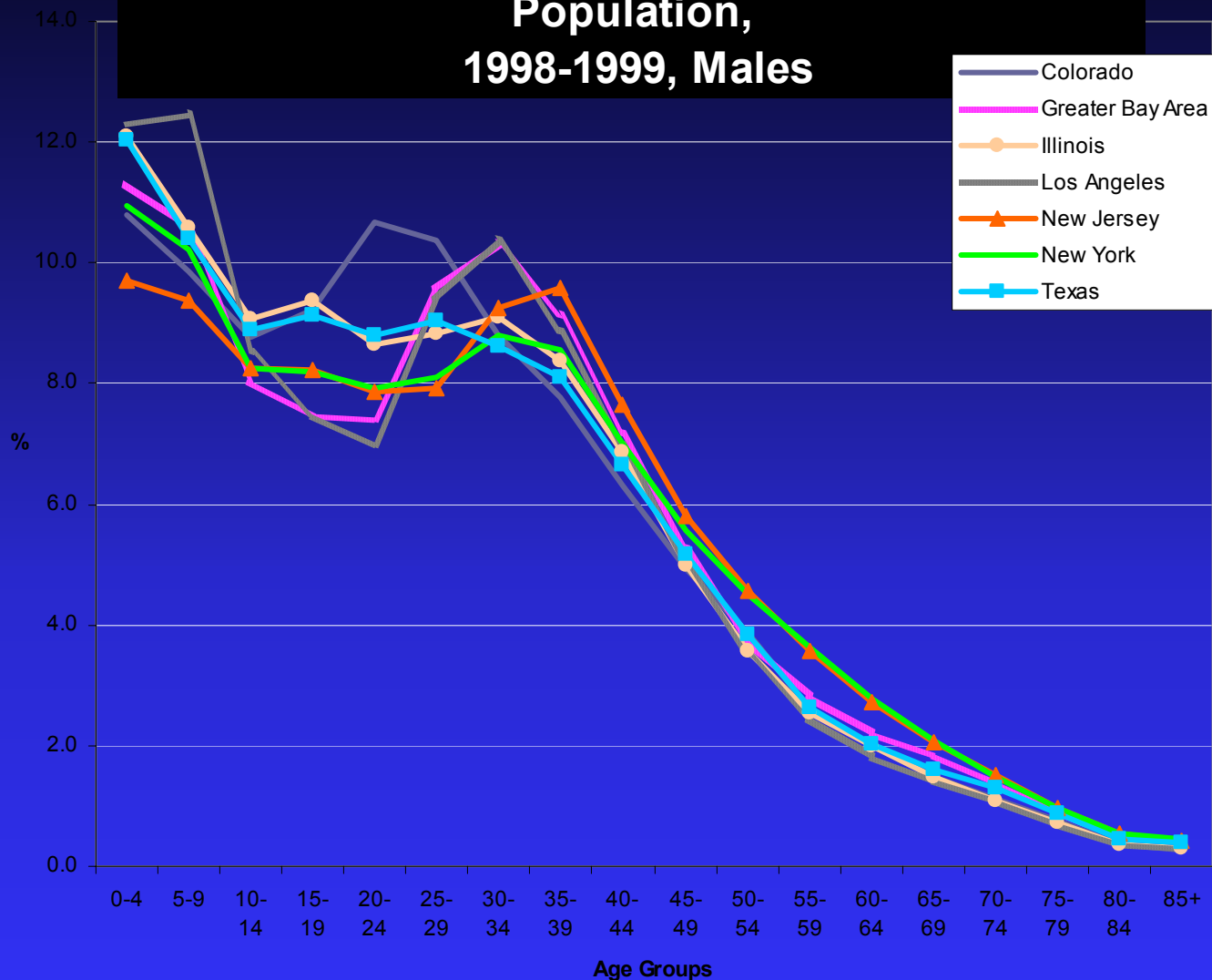
North American Association of Central Cancer Registries

Age Distribution of Estimated Hispanic (all races) Population, 1998-1999, Females
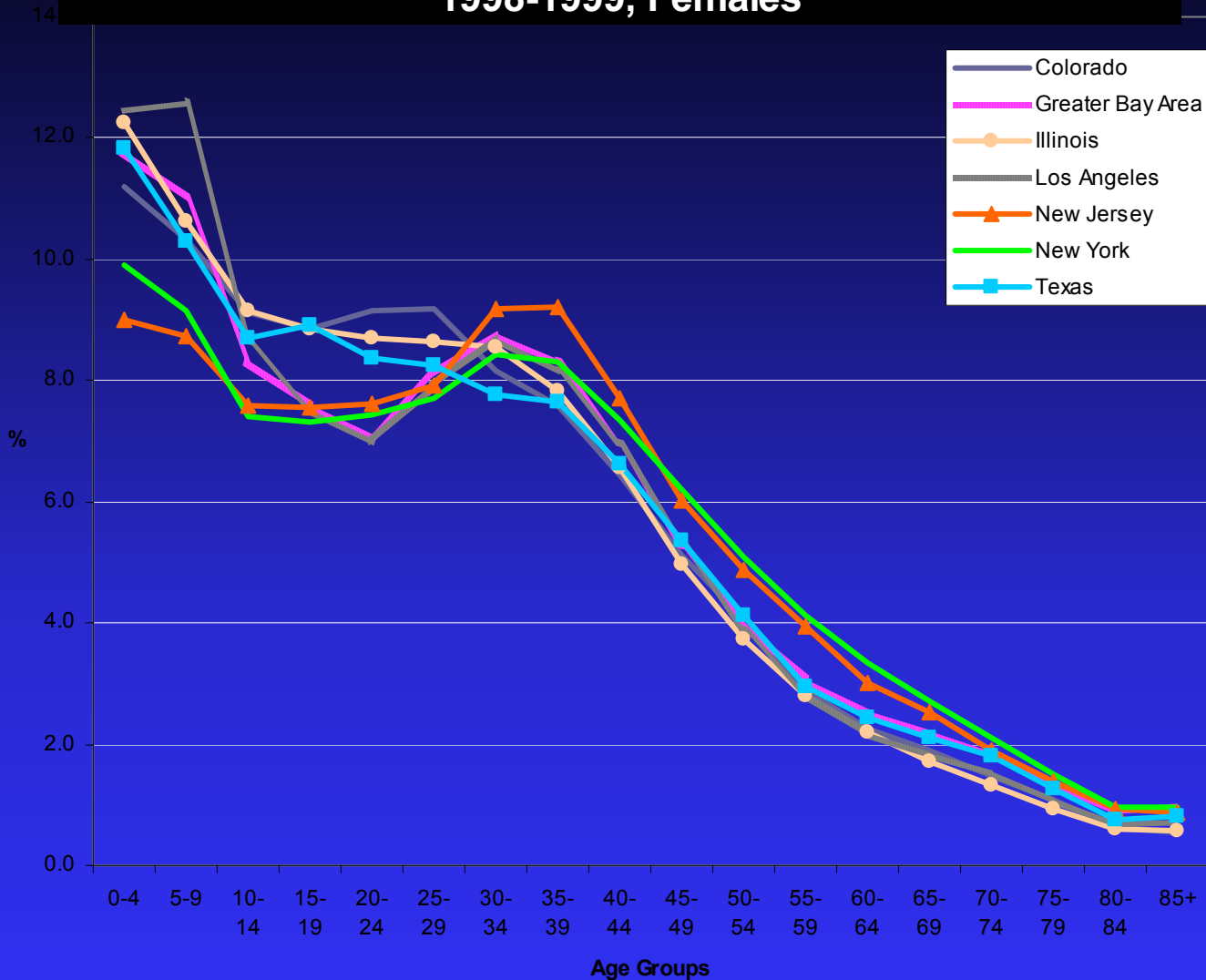
North American Association of Central Cancer Registries

General Findings

- **Method fairly reliable with historical methods; a few concerns.**

- **The age-adjusted rates varied more by registry than we expected.**

- **Cancer incidence rates were highest in Colorado and the rates in New York and New Jersey were more similar to each other than rates from either the West Coast or the Central United States.**

Other Explanations

- **Inconsistent results when algorithm applied to different populations.**

However,

- **The age-specific patterns were similar, with the exception of Colorado where the rates begin to diverge for men and women by the late 50s and continue to widen throughout the remaining life span.**

All sites age-specific incidence rates of Hispanics defined by the Illinois algorithm, 1998-1999, Males

Legend:
- Colorado
- Greater Bay Area
- Illinois
- Los Angeles
- New Jersey
- New York
- Texas

Y-axis: Rate per 100,000
X-axis: Age Groups (0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)
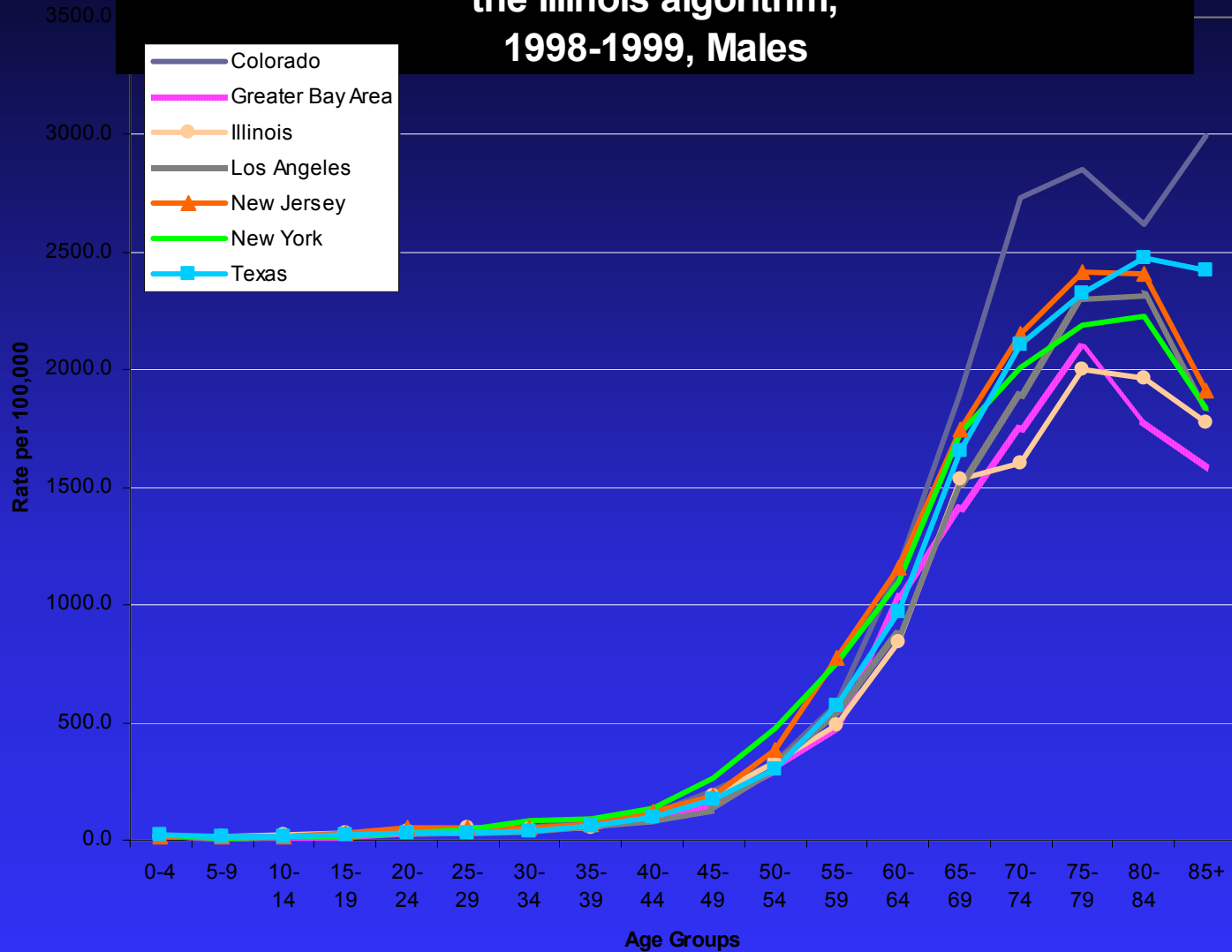
North American Association of Central Cancer Registries

All sites age-specific incidence rates of Hispanic defined by the Illinois algorithm, 1998-1999, Females

North American Association of Central Cancer Registries

ELEMENTS OF NAACCR HISPANIC IDENTIFICATION PROCESS

NAACCR Case Record
- Version = 10
- Date = January, 2003

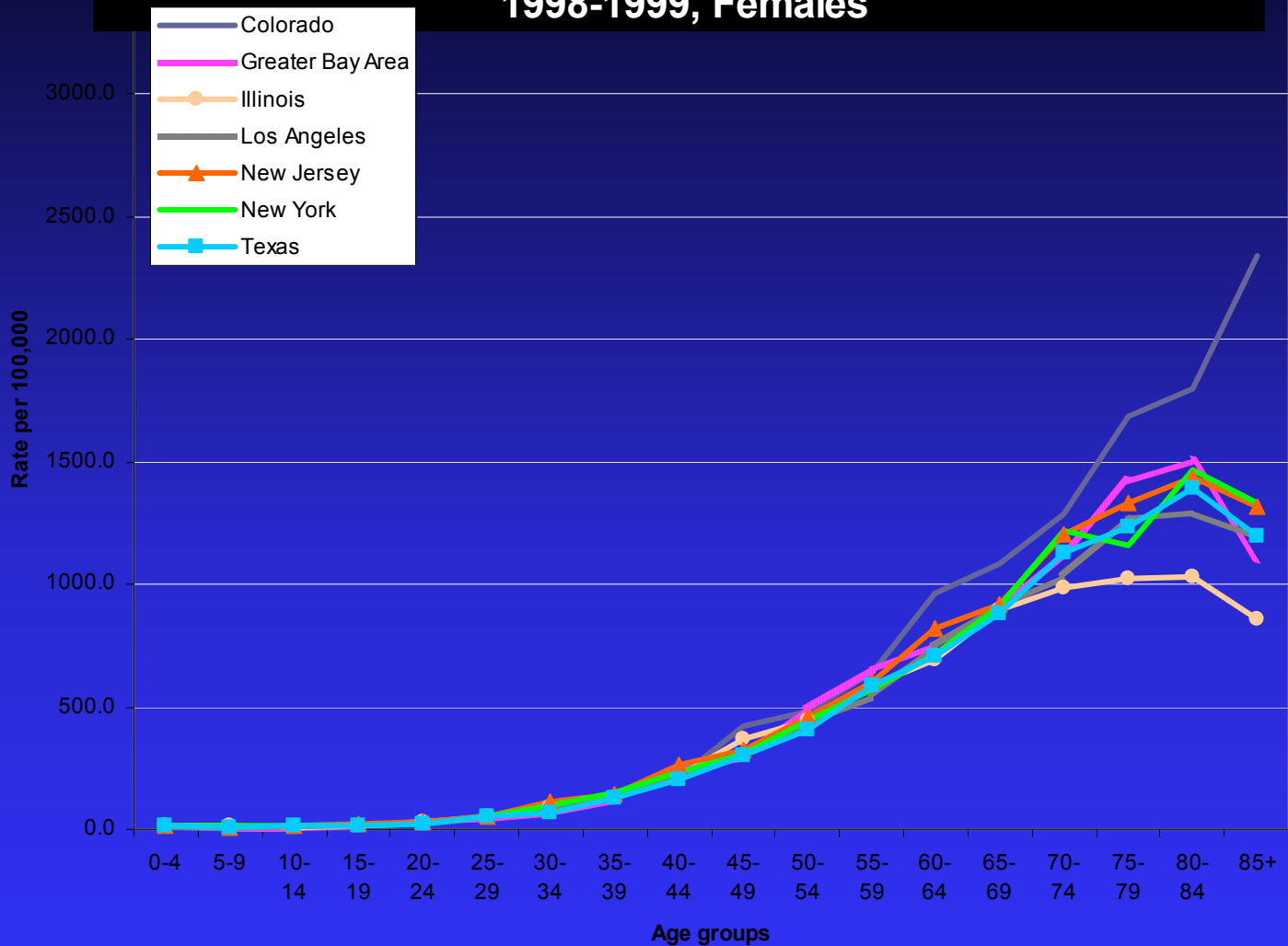**Data Item 190 Codes**
- 0 = non-Spanish; non-Hispanic
- 1 = Mexican (includes Chicano)
- 2 = Puerto Rican
- 3 = Cuban
- 4 = S. or C. American (not Brazil)
- 5 = Other specified Sp/Hisp origin (incl Europe)
- 6 = Sp/Hisp/Lat NOS
- 7 = Surname only
- 9 = Unknown whether Hispanic or not

Data Item 160 Codes
- 01 = White
- 02 = Black
- ...
- 06 = Filipino
- ....
- 88 = No further race documented
- .....
- 98 = Other
- 99 = Unknown

New Data Item (190-like) Codes
- 0 = non-Spanish; non-Hispanic
- 1 = Mexican (includes Chicano)
- 2 = Puerto Rican
- 3 = Cuban
- 4 = S. or C. American (not Brazil)
- 5 = Other specified Sp/Hisp origin (incl Europe)
- 7 = Surname - NAACCR algorithm

Data Item 190
- name = Spanish/Hispanic Origin
- code = <number>

Data Item 160
- name = Race 1
- code = <number>

<<external document / file>>
Spanish Surname List

Source = 1990 U.S. Bureau of the Census
1 Heavily Hispanic = 101, 102, 105, 110, 115, 125
2 Generally Hispanic = 201, 202, 205, 210, 215, 225
3 Moderately Hispanic = 301, 302, 305, 310, 315, 325
4 Occasionally Hispanic = ---, ---, 405, 410, 415, 425
5 Rarely Hispanic = 5001, 5005, 5010, 5025, 5100, 5500
6 Indeterminate = ---, ---, ---, ---, ---, 6000
9 Missing = ---, ---, ---, ---, ---, 9000

New Data Item (190-like)
- name = <text>
- code = <number>

Data Item 220
- name = Sex
- code = <number>

HISPANIC IDENTIFICATION PROCESS

Data Item 250
- name = Birthplace
- code = <number>

Data Item 250 Codes
- 000 = United States
- 001 = New England and NJ
- ...
- 230 = Mexico
- ....
- 443 = Spain
- .....

Data Item 220 Codes
- 1 = Male
- 2 = Female
- 3 = Other (hermaphrodite)
- 4 = Transsexual
- 9 = Not stated/Unknown

Data Item 2230
- name = Name--Last

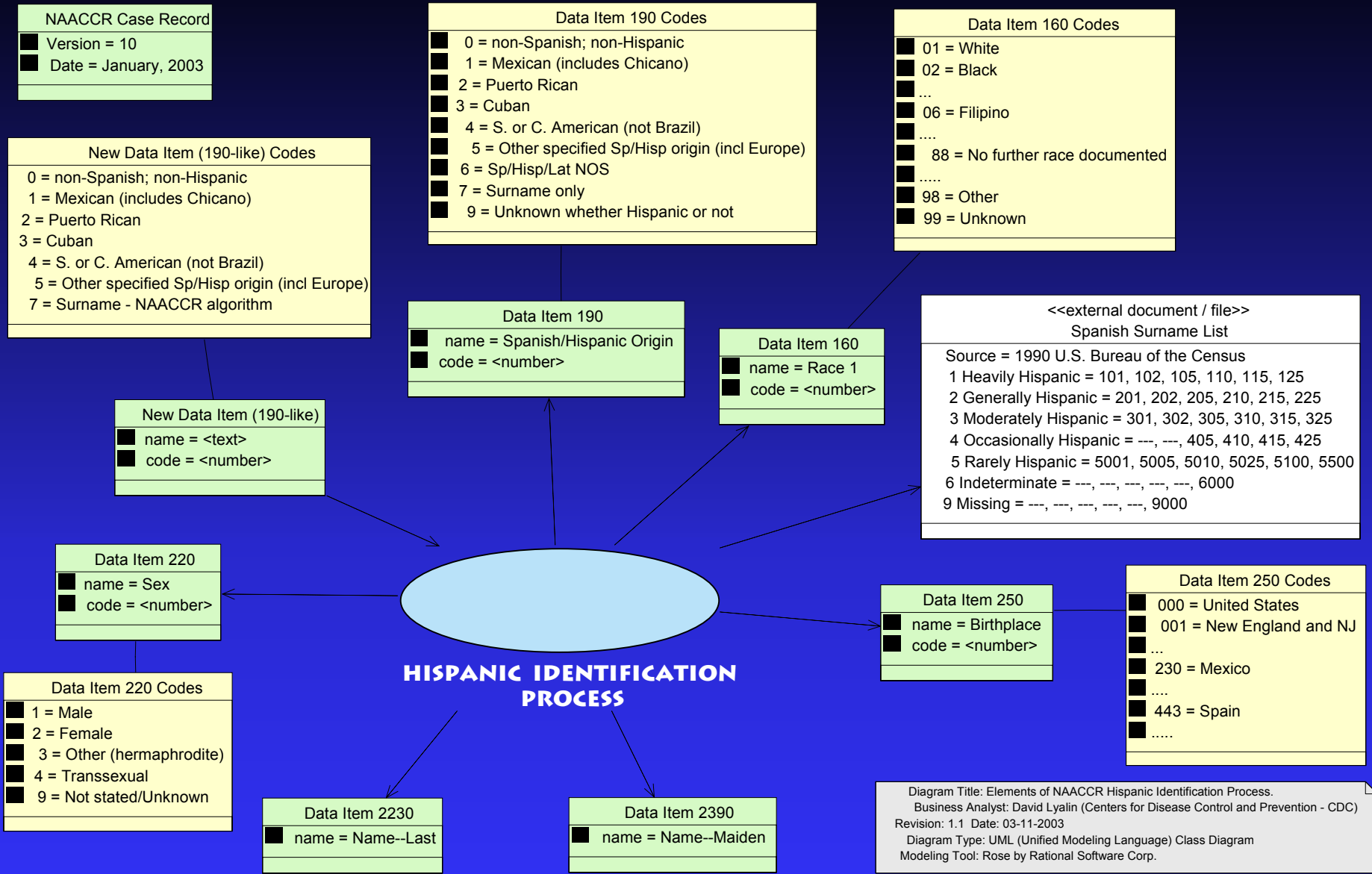**Data Item 2390**
- name = Name--Maiden

Diagram Title: Elements of NAACCR Hispanic Identification Process.
Business Analyst: David Lyalin (Centers for Disease Control and Prevention - CDC)
Revision: 1.1 Date: 03-11-2003
Diagram Type: UML (Unified Modeling Language) Class Diagram
Modeling Tool: Rose by Rational Software Corp.

Expert Panel Activities

- **April 2003. A best practices guideline released.**

- **Minor modifications were made to the Illinois State Cancer Registry Hispanic Identification Algorithm.**

- **The modified approach will be referred to as the NAACCR Hispanic Identification Algorithm (NHIA).**

North American Association of Central Cancer Registries

Tools for NHIA

- **The narrative description and visual diagrams provided in the NHIA report;**
- **The computerized program (electronic version) of the Hispanic algorithm; and**
- **The list of 1990 Spanish Surnames from the US Census.**

All are available for download: http://www.naaccr.org/news/index.html

Next Step

- **Call for Hispanic Data issued AND the deadline is July 1, 2003.**
- **All registries are invited to participate, however, only those registries who apply the NAACCR Hispanic Identification Algorithm will be eligible for inclusion.**
- **Aim is for at least 85% U.S. Hispanic population coverage.**
- **The expected date of release for the report will be December 2003.**

Hispanic Data

- **Use new population estimates**
- **Continue to evaluate data submitted through special assessments.**
- **Expert Panel will conduct several special studies:**
  - **Childhood**
  - **Urban-rural variations**
  - **Geographic variations/ Specific Hispanic populations**

North American Association of Central Cancer Registries

Expert Panel Members

- Dee West
- Cynthia O'Malley
- Lihua Liu
- Jack Finch
- Cheryl Bowcock
- Betsy Kohler
- Toshi Abe

- Pam Agovino
- Maria Schymura
- Terry Dolecek
- Vivien Chen
- Holly Howe

Diagrams:

David Lyalin