

# **NAACCR Asian and Pacific Islander Identification Algorithm**

**NAACCR Annual Meeting**

**Regina, Saskatchewan**

**June 15, 2006**

## Workgroup Members

Peg Balcius

Frank Boscoe

Cheryll Cardinez

Vivien W. Chen

Catherine Grafel-Anderson

Scarlett Gomez

Michael Green

Holly Howe

Mei-Chin Hsieh

Betsy Kohler

Sandy Kwong

Andy Lake

Lihua Liu

Barry Miller

Arti Parikh

Baozhen Qiao

Maria J. Schymura

Melanie Williams

## **Purpose**

Recode cases with a race 1 code of “96” (Asian, Not Otherwise Specified) to a more specific code, making use of birth place, last name, maiden name, and first name.

## Algorithm Steps

1. Direct code single-race Asian cases  
(codes 4-32)

## Algorithm Steps

2. Identify cases with a race code of 96

- If a case has 96 AND a more specific Asian code, assign the more specific code.
- If a case has 96 AND two or more specific Asian codes, then manually review the case.
- If a case has 96 AND a non-Asian code, then manually review the case.
- Otherwise, proceed to Step 3

## Algorithm Steps

### 3. Indirect identification based on birth place

Code	Birthplace	Race
681, 682, 683, 684, 685, 686, 691	China, Taiwan, Hong Kong, <u>Macau</u> , Tibet, <u>Mongolia</u>	Chinese
693	<u>Japan</u>	Japanese
675	<u>Philippines</u>	Filipino
695	<u>Korea</u> , North Korean, South Korean	Korean
639, 641, 643, 645, 647	India, Pakistan, Nepal, Bhutan, <u>Sikkim</u> , Bangladesh, <u>Sri Lanka</u> ,	Asian Indian, Pakistani, and Other South Asian
665	<u>Vietnam</u>	Vietnamese
651	<u>Thailand</u>	Thai
661	<u>Laos</u>	Laotian
663	Cambodia, <u>Kampuchea</u>	Cambodian or Kampuchean

## Algorithm Steps

3b. Exclude Malaysian, Indonesian, or Hawaiian birth place from further consideration.

## Algorithm Steps

4. Assign more specific race based on name

MALES:

- a. Check surname against Lauderdale list
- b. Check surname against NAACCR list
- c. Check given name against Lauderdale list
- d. Check given name against NAACCR list



## Algorithm Steps

4. Assign more specific race based on name (cont.)

FEMALES without a maiden name:

- a. Check surname against Lauderdale list
- b. Check surname against NAACCR list
- c. Check given name against Lauderdale list
- d. Check given name against NAACCR list

## Algorithm Steps

4. Assign more specific race based on name (cont.)

FEMALES with a maiden name

- a. Check maiden name against Lauderdale list
- b. Check maiden name against NAACCR list
- c. Check given name against Lauderdale list
- d. Check given name against NAACCR list
- e. Check surname against Lauderdale list
- f. Check surname against NAACCR list

## Lauderdale List

Derived from Social Security records: 1.8 million persons born in Asia before 1941

Six leading countries (China, Philippines, India, Japan, Korea, Vietnam) account for over 90% of total

Names were included on the list if at least 75% of the occurrences of the name were associated with a single one of the six countries AND the name occurred at least 5 times.

## Lauderdale List

Top Ten Chinese surnames:

Wong, Chen, Chan, Wang, Chang, Lin, Wu, Liu,  
Huang, Li

Top Ten Filipino surnames:

Reyes, Santos, Garcia, Cruz, Ramos, Delacruz,  
Mendoza, Bautista, Deguzman, Fernandez



## Lauderdale List

Ha, Jung, Ko, and Lee are examples of surnames that are common but not predictive, because they are common in multiple countries.

## NAACCR List

Derived from seven NAACCR registries, cases from 1997-2001 (Hawaii, Los Angeles, Louisiana, Illinois, Nevada, New York, Texas, Los Angeles): 80 thousand Asians.

Names were included on the list if at least 75% of the occurrences of the name were associated with a single one of the six countries AND the name occurred at least 5 times.

## Comparison

Lauderdale surname list: 1600 names, 42000 cases

NAACCR surname list: 75 names, 900 cases



## Results

Over 2 million records from New York, Louisiana, Texas, Hawaii, and California were evaluated. California results were classified into Region 8 (Bay Area), Region 9 (Los Angeles), and the rest of the state.

- 7,639 records had race1=96.
- 4,759 of these were recoded to a more specific race (62%)

[Hawaii – 15%, Louisiana – 79%]



## Results

Of those cases that were recoded...

Recoded to	All Registries combined	Lowest Registry	Highest Registry
Chinese	39%	Hawaii (20%)	New York (55%)
Japanese	13%	New York (5%)	Hawaii (27%)
Filipino	19%	Louisiana (8%)	Hawaii (40%)
Korean	9%	Louisiana (5%)	Los Angeles (15%)
Asian Indian	8%	Los Angeles (3%)	New York (17%)
Vietnamese	11%	Hawaii (0%)	Louisiana (45%)

## Quality Check

- Cases with known Asian race were set to 96 and run through the algorithm. The algorithm results were then compared with the original results.
- Participating registries: New York, Louisiana, Los Angeles

# Quality Check

## RESULTS - MALES

Step	Description	N	% correct
	Birth place	10,395	98%
a	Match surname against Lauderdale list	3,788	93%
b	Match surname against NAACCR list	257	87%
c	Match first name against Lauderdale list	338	83%
d	Match first name against NAACCR list	129	81%
	<b>TOTAL</b>	<b>14,907</b>	<b>96%</b>

# Quality Check

## RESULTS – FEMALES without maiden name

Step	Description	N	% correct
	Birth place	10,081	99%
a	Match surname against Lauderdale list	3083	92%
b	Match surname against NAACCR list	221	88%
c	Match first name against Lauderdale list	414	87%
d	Match first name against NAACCR list	169	83%
	<b>TOTAL</b>	<b>13,968</b>	<b>97%</b>

# Quality Check

## RESULTS – FEMALES with maiden name

Step	Description	N	% correct
	Birth place	1,790	97%
a	Match maiden name against Lauderdale list	416	88%
b	Match maiden name against NAACCR list	40	85%
c	Match first name against Lauderdale list	54	93%
d	Match first name against NAACCR list	25	68%
e	Match surname against Lauderdale list	67	81%
f	Match surname against NAACCR list	18	83%
	<b>TOTAL</b>	<b>2,410</b>	<b>95%</b>

## Conclusions

- The NAPIIA algorithm will assign a more specific race to about 60% of cases with a code of 96.
- Of these cases, it will assign the correct race code over 90% of the time.

## Next Steps

- Additional registries to test algorithm this month
- Incorporate NAPIIA into next Call for Data
- Tackle race code 97 (Pacific Islander NOS)





## Results by Registry (not shown)

	New York	Louisiana	Texas	Hawaii
Records	<b>730,000</b>	110,000	410,000	29,000
Race1 = 96	1,812	141	1,297	100
Race2-5 = 96	0	0	27	23
Records Recoded	1,162 (64%)	<b>111 (79%)</b>	974 (75%)	15 (15%)
Recoded as Chinese	<b>642 (55%)</b>	27 (24%)	346 (36%)	3 (20%)
Japanese	53 (5%)	6 (5%)	51 (5%)	<b>4 (27%)</b>
Filipino	145 (12%)	9 (8%)	93 (10%)	<b>6 (40%)</b>
Korean	106 (9%)	5 (5%)	102 (10%)	1 (7%)
Asian Indian	<b>182 (17%)</b>	12 (11%)	115 (12%)	1 (7%)
Vietnamese	34 (3%)	<b>50 (45%)</b>	261 (27%)	0 (0%)
Other	0 (0%)	<b>2 (2%)</b>	6 (1%)	0 (0%)

## Results by Registry (not shown)

	California	Los Angeles	San Fran. Bay	TOTAL
Records	400,000	190,000	150,000	2,020,000
Race1 = 96	<b>1,996</b>	885	1,408	7,639
Race2-5 = 96	<b>55</b>	8	7	120
Records Recoded	1,300 (65%)	414 (47%)	783 (56%)	4,759 (62%)
Recoded as Chinese	413 (32%)	174 (42%)	280 (36%)	1,885 (39%)
Japanese	359 (28%)	77 (19%)	82 (10%)	632 (13%)
Filipino	239 (18%)	57 (14%)	262 (33%)	811 (19%)
Korean	106 (8%)	<b>61 (15%)</b>	44 (6%)	425 (9%)
Asian Indian	43 (3%)	13 (3%)	57 (7%)	423 (8%)
Vietnamese	122 (9%)	28 (7%)	58 (7%)	553 (11%)
Other	18 (1%)	4 (1%)	0 (0%)	30 (1%)