

## Positive Predictive Values of Name Lists to Identify Persons of Asian Descent in a Clinical Population in Northern California

Palaniappan LP<sup>1</sup>, Wong E<sup>2</sup>, Fortmann SP<sup>2</sup>, Lauderdale DS<sup>3</sup>

<sup>1</sup>Palo Alto Medical Foundation Research Institute, Palo Alto, CA. <sup>2</sup>University of Chicago, Chicago, IL. <sup>3</sup>Stanford Medical School, Stanford, CA.

### BACKGROUND

- Race/Ethnicity is useful for the study of health disparities, however, race/ethnicity information is often not available in clinical and administrative data.
- Name lists can be used to identify persons with high likelihood of belonging to a certain race/ethnic group based on their having a name that is essentially exclusive to that race/ethnic group. For instance, "Kim" is a distinctive Korean surname, while "Lee" is not.
- The aim of name list identification is to maximize positive predictive value, with the assumption that those that are identified ("Kim") do not differ systematically from those who are not identified("Lee").
- Lauderdale and Kestenbaum developed surname lists [1] for the six largest Asian origin subgroups (Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese), based on all social security number applications (which include place of birth).
- These surname lists were validated using Census records [1], with positive predictive values ranging from 0.82 (Asian Indian) to 0.93 (Japanese).
- Lauderdale and Kestenbaum have also developed given name lists for the six largest Asian origin subgroups (Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese), which have not previously been validated
- It is unknown how well these combined given and surname lists would perform in a younger clinical population.

### OBJECTIVE

- To determine if the combined given and surname lists can correctly identify persons of specific Asian descent (Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese) in a clinical population in Northern California.

### METHODS

The Palo Alto Medical Foundation is an outpatient, multi-specialty group of clinics in the San Francisco Bay Area, CA with over 300 physicians, serving more than 200,000 active patients.

- There were 137,794 adults 35 and older who were clinic-enrolled from April 1, 2005 – March 31, 2007.
- 19,967 Persons of Asian descent were identified using given and surname lists - Asian Indian (4618), Chinese (10,236), Filipino (956), Japanese (2158), Korean (806) and Vietnamese (1193).
- 8152 people were identified using a Spanish surname list developed by the U.S. Census Bureau.[2]
- The remaining 109,675 patients, whose names were non-Hispanic and non-Asian, were inferred as Non-Hispanic White (NHW), as the clinic and surrounding community includes relatively few African Americans. The positive predictive values for the Spanish surname list and the NHW list served as a comparison group for this study.
- Patients were assigned an "Inferred Race/Ethnicity" based on the classification above. Those with discordant first and last name classifications were excluded (1.5%).
- A random sample of 400 per race/ethnic group (3200 total) were selected for a mailed race/ethnicity survey (hardcopy available below), and a separate random sample of 400 per race/ethnic group (3200 total) were selected for a telephone survey.



- Positive predictive value = true positives (self identified race/ethnicity)/ all positives (inferred race/ethnicity).
- Cost per response was calculated by dividing the total cost (postage, labor) to administer each survey (telephone vs. mail) by the response rate.

### RESULTS

- The response rates and positive predictive values for each race and ethnic group are shown below (Table)
- In the telephone survey, 2,671 individuals were contacted, and the overall response rate was 31.4%. The cost of the telephone survey was \$4,579.00, and cost per response was \$5.46.
- In the mail survey, 3,200 letters were mailed, and the overall response rate was 5.8%. The cost of the mail survey was \$1767.80, and cost per response was \$9.50.

Table: Positive Predictive Values for Telephone and Mail Surveys by Race/Ethnic Group

	Telephone Survey		Mail Survey	
	Response Rate	PPV	Response Rate	PPV
Asian Indian	32.7%	0.77	3.3%	0.85
Chinese	29.7%	0.92	4.3%	1.00
Filipino	32.8%	0.81	8.8%	0.94
Japanese	30.6%	0.80	9.8%	0.85
Korean	32.7%	0.78	5.5%	0.86
Vietnamese	34.1%	0.74	6.5%	0.88
Non-Hispanic White	28.1%	0.83	8.0%	0.88
Spanish (Hispanic)	31.0%	0.67	3.8%	0.53

### CONCLUSIONS

- Despite the low response rates in the validation study, the given and surname lists have been successfully validated, with high positive predictive value, in this clinical population.
- Positive predictive values for the Asian surname lists were higher than the widely used Spanish surname lists developed by the U.S. Census Bureau in this population.
- Telephone surveys for race/ethnicity have 5 times higher response rates than mailed surveys, and half the cost per response.

### LIMITATIONS

- The sensitivity of these lists are unknown. Low sensitivity (missing members of a certain race/ethnic group) would underestimate race/ethnic group population size.

### IMPLICATIONS

- Asian given and surname lists can be used with confidence to infer race/ethnicity information for Asian subgroups (Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese), when race/ethnicity information is not available.

### REFERENCES

- Lauderdale DS, Kestenbaum B. Asian American ethnic identification by surname. *Population Research and Policy Review*. June 2000;19(3):283-300.
- Word D, Perkins C. *Building a Spanish Surname List for the 1990's - A New Approach to an Old Problem*. Washington DC: U.S. Census Bureau; March 1996.