Published in final edited form as:
J Immigr Minor Health. 2011 April ; 13(2): 345–351. doi:10.1007/s10903-009-9296-x.

Lessons Learned from the Application of a Vietnamese Surname List for Survey Research

Victoria M. Taylor[1], **Tung T. Nguyen**[2], **H. Hoai Do**[1], **Lin Li**[1], and **Yutaka Yasui**[3]

Victoria M. Taylor: vtaylor@fhcrc.org

[1] Division of Public Health Sciences, Fred Hutchinson Cancer Research Center (M3-B232), 1100 Fairview Avenue North, Seattle, WA 98109, USA

[2] Division of General Internal Medicine, University of California at San Francisco, San Francisco, CA, USA

[3] Department of Public Health Sciences, University of Alberta, Edmonton, AB, Canada

Abstract

Surname lists are increasingly being used to identify Asian study participants. Two Vietnamese surname lists have previously been published: the Vietnamese Community Health Promotion Program (VCHPP) list and the Lauderdale list. This report provides findings from a descriptive analysis of the performance of these lists in identifying Vietnamese. To identify participants for a survey of Vietnamese women, a surname list (that included names that appear on the VCHPP list and/or Lauderdale list) was applied to the Seattle telephone book. We analyzed surname data for all addresses in the survey sample, as well as survey respondents. The VCHPP list identified 4,283 potentially Vietnamese households, and 79% of the households with established ethnicity were Vietnamese; and the Lauderdale list identified 4,068 potentially Viet-namese households, and 80% of the households with established ethnicity were Vietnamese. However, the proportions of contacted households that were Vietnamese varied significantly among commonly occurring surnames. The characteristics of women with surnames on the VCHPP and Lauderdale lists were equivalent. The two lists performed equally well in identifying Vietnamese households. Researchers might consider using different combinations of Vietnamese surnames, depending on whether accuracy or high population coverage is the more important consideration.

Keywords

Vietnamese; Surname lists

Introduction

The United States (US) has recently experienced one of the largest immigration waves in history. Consequently, researchers are increasingly focusing on health disparities experienced by immigrant populations [1]. According to Census data, 11% of Asian Americans are of Vietnamese descent and the Vietnamese population now exceeds 1,400,000 [2,3]. A majority of Vietnamese Americans came to the US as refugees or immigrants over the last three decades [4,5]. Vietnamese are the second fastest growing Asian American group, after South Asian Indians [4]. California, Texas, and Washington State have the largest Vietnamese communities [2].

Correspondence to: Victoria M. Taylor, vtaylor@fhcrc.org.

Commonly used methods for obtaining a representative sample of the general population (e.g., random digit dialing) are usually cost-prohibitive when researchers are attempting to identify members of smaller ethnic groups. Further, use of non-representative sampling methods (e.g., client lists provided by organizations serving one ethnic group) affects the extent to which study findings can be generalized to an ethnic group as a whole [6]. Therefore, surname lists are increasingly being used to identify people of a particular ethnicity for research with Asian immigrant populations [7].

Previous studies have developed and evaluated surname lists for identifying Chinese in Australia and Canada; South Asians in Canada, England, and Scotland; and Turks in Germany and The Netherlands [8–15]. Surname lists have been used in both epidemiologic and health services research [7,9]. As examples, surname lists have recently been applied to the California cancer registry to examine cancer incidence and stage at diagnosis patterns among Hmong immigrants to California, and to the Vancouver telephone book to identify participants for a community-based survey addressing hepatitis B knowledge and practices among Chinese Canadians [16,17].

Vietnamese surnames are not related to geographic region of origin in Vietnam or membership of a clan network, and are maintained over generations. In Vietnam, names generally consist of three parts: a family name, a middle name, and a given name, used in this order. However, Vietnamese Americans reverse this name order to be consistent with the US naming system. Vietnamese children take their father's last name and Vietnamese women keep their own family name upon marriage. Parents avoid giving Vietnamese children first names that belonged to ancestors because this is considered disrespectful of elders. Certain first names can be used for both genders but middle names are uniquely male or female. Vietnamese Americans have begun to give their children American first names but have not changed their last names from Vietnamese to American names.

Two lists of surnames that can be used to identify Vietnamese Americans have been published [18,19]. First, the Vietnamese Community Health Promotion Project (VCHPP) at the University of California, San Francisco uses a Vietnamese surname list to identify Vietnamese Americans for its research projects [19]. The VCHPP name list was initially developed with input from members of the Vietnamese community in California, and has been expanded upon and refined over time [19,20]. Second, Lauderdale and her colleague used US Social Security Administration files and Census records to develop a list of Vietnamese surnames and rank them by frequency [18]. We recently used a list of Vietnamese surnames (that includes names from the VCHPP and Lauderdale lists) to identify participants for a women's health study in Seattle, Washington. This report provides findings from a descriptive analysis of the performance of this Vietnamese surname list in identifying Vietnamese Americans.

Methods

Vietnamese Name Lists

Our Vietnamese surname list included 55 names (all 37 names that are on the VCHPP list, as well as 49 of 50 names that are on the Lauderdale list). The Lauderdale list and its frequency rankings are shown in Table 1, and the VCHPP list is shown in Table 2 [18,19]. Nineteen names—Banh, Buu, Dam, Giang, Hoa, La, Luc, Nghiem, Phu, Quang, Thach, Thai, Thanh, Thi, Tieu, Trang, Van, Vinh, and Vong—are on the Lauderdale list but not the VCHPP list; and six names—Ha, Ho, Khuu, Lam, Quan, To—are on the VCHPP list but not the Lauderdale list. The top 22 ranked Lauderdale names are all included on the VCHPP list.

Survey Methods

Our survey methods have been described in detail elsewhere [21]. The study surname list was applied to an electronic version of the 2005 telephone book for metropolitan Seattle. We identified all addresses (in 33 zip codes) with one of the Vietnamese surnames. Each address received an introductory mailing about the project that specifically asked the household to call the project if it did not consider itself to be Vietnamese. Bilingual (Vietnamese and English) survey workers subsequently visited each address and completed an in-person survey with a female household member (if the household verified it was Vietnamese and included a Vietnamese woman who was in the 20–79 age group). Because Vietnamese women generally keep their own surname upon marriage, each survey respondent was asked to specify her own surname (regardless of whether it was the same as the surname used to identify the household as potentially Vietnamese or not). Survey respondents were also asked to provide demographic information about themselves.

Surname Groupings

In this study, we considered four different surname combinations: all names from the combined VCHPP and Lauderdale study list, all names from the VCHPP list, all names from the Lauderdale list, and the top 15 ranked Lauderdale names. We also considered the more commonly occurring Vietnamese surnames individually.

Household Sample Analyses

We conducted an analysis of all addresses that were in the original sample. Specifically, we examined the numbers and proportions of potentially Vietnamese households that were identified by each of the surname combinations and commonly occurring individual surnames. Additionally, we conducted an analysis of households where ethnicity was established. This analysis examined the proportions of households that were actually Vietnamese by name combinations and individual names. We used chi square tests to compare the proportion of households that were identified using the very common surname of Nguyen with the proportions of households that were identified using each of the other most commonly occurring surnames that were actually Vietnamese.

Survey Respondent Analyses

Another analysis focused on women who responded to the survey. This analysis used each woman's surname (provided at the time of her survey) rather than the surname that was originally used to identify her household as potentially Vietnamese. We examined selected demographic characteristics by each of the four surname combinations. Specifically, we examined age (<50 vs. ≥50 years), educational level (<12 vs. ≥12 years), duration of time since immigration (<15 vs. ≥15 years), and English proficiency (did not speak well or at all vs. spoke well or fluently). To provide an assessment of Vietnamese surnames that are not included on the VCHPP list or Lauderdale list, we described the surnames of Vietnamese survey respondents who had surnames that are not on either of the lists. (This assessment was possible because married Vietnamese women usually keep their own surname.)

Ethics Committee Review

The Fred Hutchinson Cancer Research Center Institutional Review Board approved all the study procedures.

Results

Study Samples

As shown in Fig. 1, 4,437 potentially Vietnamese household addresses were identified in the study zip codes (using the study surname list), and we were able to establish ethnicity for 3,017 (68%) of these addresses. Of the 3,017 households where ethnicity was established, 2,353 (78%) were Vietnamese. Finally, 1,532 Vietnamese women completed surveys. Among the 1,532 households that participated in the survey, the surname used to identify the household as potentially Vietnamese and the responding woman's surname were discordant (e.g., Nguyen and Pham) in 912 (60%) of cases and concordant (e.g., Nguyen and Nguyen) in 620 (40%) of cases.

Household Sample Results

Results from our analysis of all addresses in the original sample for the four name combinations are given in Table 3. The numbers of potentially Vietnamese households that were identified by the VCHPP list (4,283) and Lauderdale list (4,068) were similar. Additionally, the proportions of contacted households that were verified to be Vietnamese were almost identical for names on the VCHPP list (79%) and names on the Lauderdale list (80%). The top 15 Lauderdale names identified 3,388 potentially Vietnamese households, and 82% of contacted households with one of these 15 names were verified to be Vietnamese.

Table 4 gives results for the most commonly occurring names in the original sample. One name (Nguyen) identified 28% of all the addresses. The top five Lauderdale names in the Seattle sample were identical to the top five names in Lauderdale's national study (and had the same frequency rankings). Similarly, the top 15 Lauderdale names in the Seattle sample were identical to the top 15 names in Lauderdale's national study (but had different frequency rankings) [18]. Two names that are on the VCHPP list but not the Lauderdale list —Ho and Lam—were the sixth and seventh most commonly occurring names in the Seattle sample.

As previously stated, we compared the proportion of households that were identified using the very common surname of Nguyen with the proportions of households that were identified using each of the other most commonly occurring surnames that were actually Vietnamese. Households that were identified using five names—Lam, Ho, Dang, Ly, and Ngo—were significantly less likely ($p < 0.05$) to be Vietnamese than households that were identified using the name Nguyen.

Survey Sample Results

Over 98% of the women who completed a survey indicated they were born in Vietnam. The remaining women were born in the US or another Asian country (e.g., the Philippines). Additionally, over 98% of the survey respondents completed their survey in Vietnamese (rather than English). One-half (50%) of the responding women were less than 50 years old, 50% had less than 12 years education, 63% had been in the US for less than 15 years, and 41% had limited English proficiency (Table 5).

The four surname combinations considered in this study generated sub-groups of women who were very similar with respect to their demographic characteristics (age, educational level, years since immigration, and English proficiency). Only 81 of the Vietnamese women who responded to our survey had surnames that are not on the VCHPP list or Lauderdale list. Further, no individual surnames that are not on the VCHPP list or Lauderdale list occurred frequently (more than six times) among survey participants.

Discussion

We descriptively evaluated two published Vietnamese surname lists that were developed using different approaches. One of these lists (Lauderdale) was developed systematically using automated databases, while the other list (VCHPP) was developed using a community-based participatory approach [18–20]. Interestingly, these two lists performed very similarly when applied in metropolitan Seattle. Specifically, the VCHPP list identified 4,283 potentially Vietnamese households, and 79% of the households with established ethnicity were Vietnamese; and the Lauderdale list identified 4,068 potentially Vietnamese households, and 80% of the households with established ethnicity were Vietnamese.

Mateos has recently summarized the potential limitations of using surname lists to identify ethnic population subgroups. Temporal differences in name distributions between a reference population (used to develop a list) and a target population can occur because of different migration waves and variations in geographic distribution patterns over time. Additionally, regional differences in the frequency distribution of names can occur between and within countries because of differential historic processes and migration flows [7]. For example, the frequencies of Pakistani names occurring in the north and south east of England differ [10]. However, the distribution of Vietnamese surnames in our recent Seattle study did not differ meaningfully from the distribution of Vietnamese surnames in the Lauderdale study that used a list of Vietnamese Americans who were born in North or South Vietnam before 1941 [18].

The positive predictive value of any individual Vietnamese surname (the proportion of people with the name who are actually Vietnamese) depends on the proportion of Vietnamese in a particular geographic area [18]. Our results indicate that some of the names on the VCHPP and/or Lauderdale lists also occur relatively frequently in other Asian ethnic groups. For example, less than one-half (45%) of the households in the original sample with the name Lam (which is also a relatively common Chinese name) were actually Vietnamese [9,18]. If a study is being conducted in an area where the Asian population is predominantly Vietnamese, the positive predictive value of names such as Lam will be higher than if a study is being conducted in a geographic area where multiple Asian groups are represented.

In our study, the household surname and survey respondent surname could have been concordant because a husband and wife had the same commonly occurring name (e.g., Nguyen) prior to marriage, a woman had adopted the western custom of changing surname upon marriage, an adult daughter (with the same name as her father) completed the survey, or the household was headed by a woman. However, in 912 cases the household surname and respondent surname were discordant. Of the 912 respondents with a name that differed from the corresponding household name, only 81 (9%) had a name that is not on the VCHPP list or Lauderdale list. Further, none of these names occurred more than six times. This suggests that our combined VCHPP and Lauderdale list identifies a high proportion of Vietnamese Americans and includes all commonly occurring Vietnamese surnames.

Researchers rarely know whether persons with names that are on an ethnic surname list differ from persons of the same ethnicity with names that are not on the ethnic surname list, and few studies have addressed this issue. However, one study found that Korean Americans with the very common Korean surname of Kim did not differ demographically from all Korean Americans [22]. Similarly, we found that the demographic characteristics of Vietnamese Americans with the very common Vietnamese surname of Nguyen, one of the top 15 Lauderdale names, one of the Lauderdale names, and one of the VCHPP names were very similar. This finding suggests that the demographic characteristics of Vietnamese

Americans do not differ meaningfully by surname, and researchers could use a subset of commonly occurring Vietnamese surnames without introducing bias.

A recent analysis of 2000 Census data found that 94% of Americans with the surname of Nguyen are Vietnamese [23]. In our study, only 84% of households that were identified using the surname of Nguyen were Vietnamese American. One limitation of our study is that some households probably moved between the time we conducted our sampling procedures and attempted to contact study households (the time lag ranged from one to 6 months). For example, a non-Vietnamese family may have replaced a Vietnamese family at a particular address. Therefore, it is likely that our study under-estimates the performance of Vietnamese names in identifying individuals of Vietnamese descent. Further, the interviewers were unable to establish the household ethnicity for 1,177 of the residential addresses in our original sample (because they were unable to access a secure complex or unable to contact a household after five attempts). However, there is no reason to believe that mobility (the frequency of address changes), residence in a secure apartment or condominium complex, and the likelihood of not being at home is related to Vietnamese surname. Finally, our study did not include households that have unlisted land phones or exclusively use cell phones. Younger people (<35 years of age) are more likely to use a cell phone as their primary telephone than older people (35 years of age or older) [24].

New Contribution to the Literature

If accuracy (being sure of a person's Vietnamese ethnicity) is the most important consideration in a study, researchers might want to consider using a subset of surnames from the VCHPP and Lauderdale lists. For example, they could use the following 12 commonly occurring Vietnamese names that are all associated with a high positive predictive value: Nguyen, Tran, Le, Pham, Huynh, Do, Bui, Truong, Hoang, Phan, Vo, and Vu. In contrast, if detecting the highest possible number and/or proportion of Vietnamese in a particular geographic area is the most important consideration in a study, researchers might want to use all the surnames that are included on the VCHPP and Lauderdale lists.

Acknowledgments

This publication was supported, in part, by grant R01-CA-115564 from the National Cancer Institute, cooperative agreement U01-CA-114640 from the National Cancer Institute, and cooperative agreement U48-DP-000050 from the Centers for Disease Control and Prevention. The contents of the article are solely the responsibility of the authors and do not necessarily represent the views of the National Cancer Institute or the Centers for Disease Control and Prevention.

References

1. Kandula NR, Kersey M, Lurie N. Assuring the health of immigrants: what the leading health indicators tell us. Annu Rev Public Health 2004;25:357–76. [PubMed: 15015925]
2. Pfeiffer, ME. US Census releases 2005 American Community Survey data for Southeast Asian Americans. [Accessed 11 Oct 2008]. www.hmongstudies.org
3. Census Bureau US. The American community-Asians, 2004. Washington: US Department of Commerce; 2007.
4. Chan, S.; Lee, E. Families with Asian roots. In: Lynch, EW.; Hanson, MJ., editors. Developing cross-cultural competency. Baltimore: Paul Brookes Publishing; 2004.
5. Pham, L. The Vietnamese community in the United States. Seattle: The Cross Cultural Health Care Program; 1999.
6. Tjam EY. How to find Chinese research participants: use of a phonologically based surname search method. Can J Public Health 2001;92:138–42. [PubMed: 11338153]
7. Mateos P. A review of name-based ethnicity classification methods and their potential in population studies. Popul Space Place 2007;13:243–63.

8. Browhuis CB, Moll HA. Determination of ethnicity of children in The Netherlands: two methods compared. European J Epidemiol 2003;13:243–63.

9. Choi BC, Hanley AJ, Holowaby EJ, Dale D. Use of surnames to identify individuals of Chinese ancestry. Am J Epidemiol 1993;138:723–34. [PubMed: 8237987]

10. Cummins C, Winter H, Cheng KK, Maric R, Silcocks P, Varghese C. An assessment of the Nam Pehchan computer program for the identification of names of south Asian ethnic origin. J Public Health Med 1999;21:401–6. [PubMed: 11469361]

11. Hage BK, Oliver RJ, Powles JW. Telephone directory listings of presumptive Chinese surnames: an appropriate sampling frame for a dispersed population with characteristic surnames. Epidemiol 1990;1:405–8.

12. Nanchahal K, Mangtani P, Alston M, dos Santos Silva I. Development and validation of a computerized South Asian names and group recognition algorithm (SANGRA) for use in British health-related studies. J Public Health Med 2001;23:278–85. [PubMed: 11873889]

13. Quan H, Wang F, Schopflocher D, Norris C, Galbraith PD, Faris P, et al. Development and validation of a surname list to define Chinese ethnicity. Med Care 2006;44:328–33. [PubMed: 16565633]

14. Razum O, Zeeb H, Akgun S. How useful is a name-based algorithm in health research among Turkish migrants in Germany? Trop Med Int Health 2001;6:654–61. [PubMed: 11555431]

15. Sheth T, Nair C, Nargundkar M, Anand S, Yusuf S. Cardiovascular and cancer mortality among Canadians of European, south Asian and Chinese origin from 1979 to 1993: an analysis of 1.2 million deaths. Can Med Assoc J 1999;161:132–8. [PubMed: 10439820]

16. Hislop TG, Teh C, Low A, Li L, Tu SP, Yasui Y, et al. Hepatitis B knowledge, testing and vaccination levels in Chinese immigrants to British Columbia, Canada. Can J Public Health 2007;98:125–9. [PubMed: 17441536]

17. Mills PK, Yang RC, Riordan D. Cancer incidence in the Hmong in California, 1988–2000. Cancer 2005;104:2969–74. [PubMed: 16247793]

18. Lauderdale DS, Kestenbaum B. Asian American ethnic identification by surname. Popul Res Policy Rev 2000;19:283–300.

19. Nguyen T, McPhee S, Lam T, Mock J. Predictors of cervical Pap smear screening awareness, intention, and receipt among Vietnamese-American women. Am J Prev Med 2002;23:207–14. [PubMed: 12350454]

20. McPhee SJ. Promoting breast and cervical cancer screening among Vietnamese American women: two interventions. Asian Am Pac Isl J Health 1998;6:344–50. [PubMed: 11567460]

21. Taylor VM, Yasui Y, Nguyen T, Woodall E, Do H, Acorda E, Li L, Choe J, Jackson JC. Pap smear receipt among Vietnamese immigrants: the importance of health care factors. Ethn Health. (in press).

22. Shin SH, Yu EY. Use of surnames in ethnic research: the case of Kims in the Korean-American population. Demography 1984;21:347–60. [PubMed: 6479394]

23. Falkenstein, MR.; Word, DL. The Asian and Pacific Islander surname list: as developed from Census 2000. Washington: US Census Bureau; 2002.

24. Making a connection—survey results concerning mobile phone usage. American Demographics. 2001 December 1;

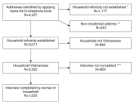

Fig. 1.

overview of survey recruitment. * The interviewers were unable to access a secure apartment or condominium complex (*N* = 202) or unable to contact the household after five attempts (*N* = 975), ** The address was a vacant dwelling (*N* = 190) or business (*N* = 53), *** The household was Vietnamese but refused to participate (*N* = 596) or did not include a woman who was age-eligible for survey participation (*N* = 224)

Table 1

Lauderdale name list with frequency rankings [18]

1. Nguyen	11. Dang	21. Doan	31. Diep	41. Phu
2. Tran	12. Do	22. Dao	32. Ton	42. Vinh
3. Le	13. Bui	23. Thai	33. La	43. Quang
4. Pham	14. Vo	24. Mai	34. Thach	44. Tieu
5. Huynh	15. Ly	25. Van*	35. Thi	45. Hoa
6. Vu	16. Duong	26. Cao	36. Thanh	46. Trang
7. Phan	17. Luong	27. Vuong	37. Dam	47. Giang
8. Truong	18. Dinh	28. Phung	38. Vong	48. Luc
9. Hoang	19. Trinh	29. Quach	39. Trieu	49. Banh
10. Ngo	20. Luu	30. Ta	40. Buu	50. Nghiem

*Van was not included in the study sampling because Dutch-origin names (e.g., Van Buren) were inadvertently selected during electronic sampling procedures

Table 2

VCHPP name list [19]

Bui	Duong	Luong	Phung	Trinh
Cao	Ha	Luu	Quach	Truong
Dang	Ho	Ly	Quan	Vo
Dao	Hoang	Mai	Ta	Vu
Diep	Huynh	Ngo	To	Vuong
Dinh	Khuu	Nguyen	Ton	
Do	Lam	Pham	Tran	
Doan	Le	Phan	Trieu	

Table 3

Household sample results—name combinations

Name combination	All addresses	Households with established ethnicity	
	N	*N*	% Vietnamese
All VCHPP and Lauderdale names	4,437	3,017	78
All VCHPP names	4,283	2,900	79
All Lauderdale names	4,068	2,767	80
Top 15 Lauderdale names	3,388	2,289	82

Table 4

Household sample results—individual names

Name	Seattle rank	Lauderdale rank	All addresses			Households with established ethnicity	
			N	%	Cumulative %	*N*	% Vietnamese
Nguyen	1	1	1,228	27.7	28	822	84
Tran	2	2	496	11.2	39	329	83
Le	3	3	319	7.2	46	215	83
Pham	4	4	228	5.1	51	159	86
Huynh	5	5	157	3.5	55	104	79
Lam	6	–	133	3.0	58	96	45*
Ho	7	–	130	2.9	61	89	55*
Do	8	12	105	2.4	63	72	82
Bui	9	13	103	2.3	65	71	83
Truong	9	8	103	2.3	68	65	75
Hoang	11	9	101	2.3	70	74	82
Phan	11	7	101	2.3	72	72	76
Vo	13	14	100	2.3	75	73	92
Dang	14	11	93	2.1	77	61	62*
Ly	15	15	89	2.0	79	66	56*
Vu	16	6	89	2.0	81	57	89
Ngo	17	10	76	1.7	82	49	65*

*Significantly lower proportion than Nguyen ($P < 0.05$)

Table 5

Survey sample results—characteristics of respondents

Name combination	*N*	Characteristic			
		< 50 Years of age (%)	< 12 Years education (%)	< 15 Years since immigration (%)	Limited English proficiency (%)
All survey respondents	1,532	50	50	63	41
All VCHPP and Lauderdale names	1,451	51	50	64	41
All VCHPP names	1,413	50	50	64	41
All Lauderdale names	1,395	51	50	64	41
Top 15 Lauderdale names	1,179	50	50	65	41
Nguyen	459	52	49	66	40