

Letter to the Editor

Validity of using surname to define Chinese ethnicity

To the Editor:

Many secondary databases cannot readily be used to study ethnic/racial variation in population health because they lack information on the ethnicity of individual patients. To overcome this limitation, some have proposed using surnames as an alternative source of information for defining ethnicity since these are usually passed on from generation to generation. Three Chinese surname lists have been independently developed as indicators of Chinese ethnicity by Coldman et al.¹ in 1988, Choi et al.² in 1993 and Tjam³ in 2001. While a growing number of studies have employed these methodologies,⁴⁻⁸ their validity has not been fully characterized.

We have access to a unique clinical database of all cardiac catheterization procedures performed in the province of Alberta, Canada [the Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease (APPROACH) registry].⁹ This database contains both patient surnames and self-reported ethnicity that was collected through a mailing survey. In the survey, patients were asked, "Please mark one or more, You are?" Following the question, a number of ethnicities were listed, among which "Chinese" was listed as an option. Using this data resource, we assessed the agreement between self-perceived Chinese ethnicity and the ethnicity assigned using a Chinese surname list developed by Tjam.³ We chose Tjam's list because it is the latest and most comprehensive algorithm. We also assessed the effect of slight modifications to the surname list to derive a

more sensitive surname method for defining Chinese ethnicity.

Of the 3,059 patients reporting information on ethnicity, 0.6% reported themselves to be Chinese. That proportion was slightly lower than the 1.1% of patients defined to be Chinese in our database according to Tjam's surname list. When compared with self-reported Chinese ethnicity from the follow-up survey, the surname list had a sensitivity of 66.7% [95% confidence interval (95% CI): 46.5-80.3], positive predictive value of 35.3% (95% CI: 19.7-53.5), specificity of 99.3% and negative predictive value of 99.8%.

We revised the surname list by removing surnames from Tjam's list that are common in both Chinese and other ethnic groups and by adding other Chinese surnames to Tjam's list. The revised surname list (See Table I) had 83.3% (95% CI: 58.6-96.4) sensitivity, 83.3% (95% CI: 58.6-96.4) positive predictive value, 99.9% specificity and 99.9% negative predictive value.

Comments

Using surnames to define ethnicity is a promising approach to using existing health databases to define individual patients' ethnicities. Our study findings support that the Chinese surname list developed by Tjam³ is reasonably sensitive for identifying individuals of Chinese ethnicity. The revised surname list is more valid than Tjam's list and could accurately identify over 80% of Chinese individuals in databases like ours. Of course, any approach that relies on surnames to classify ethnicity will occasionally misclassify individuals due to the various name translations that arise

when converting a Chinese name into English, the common spelling with names from other ethnicities (e.g., Koreans), name changes associated with inter-ethnic marriages, and multiple-cultural identities.

Our study is limited by a relatively small sample size, and by the fact that we have not validated our slight revision of Tjam's list in a large general population. Nevertheless, this methodological study assessing the validity of an existing surname algorithm for defining Chinese ethnicity builds on a growing methodological foundation for future research on ethnicity and health.

Hude Quan, MD, PhD,^{1,2} William A. Ghali, MD, MPH,^{1,3} Stafford Dean, MA,⁴ Colleen Norris, PhD,⁵ P. Diane Galbraith, BN,^{1,2} Peter Faris, PhD,⁴ Michelle M. Graham, MD,⁵ Merrill L Knudtson, MD³

1. Department of Community Health Sciences, University of Calgary, Calgary, AB
2. Centre for Health and Policy Studies, University of Calgary
3. Department of Medicine, University of Calgary
4. Calgary Health Region
5. Department of Medicine, University of Alberta, Edmonton, AB

Correspondence: Dr. Hude Quan, Department of Community Health Sciences, University of Calgary, 3330 Hospital Dr. NW, Calgary, AB T2N 4N1, Tel: (403) 944-8912, Fax: (403) 944-8950, E-mail: hquan@ucalgary.ca

Acknowledgement: The conduct of this study was supported by operating grants from the Calgary Health Region, Calgary, and the Institute of Health Economics, Alberta, Canada.

References

1. Coldman AJ, Braun T, Gallagher RP. The classification of ethnic status using name information. *J Epidemiol Community Health* 1988;42:390-95.
2. Choi BC, Hanley AJ, Holowaty EJ, Daley D. Use of surname to identify individuals of Chinese ancestry. *Am J Epidemiol* 1993;138:723-34.
3. Tjam EY. How to find Chinese research participants: Use of a phonologically based surname search method. *Can J Public Health* 2001;92(2):138-42.
4. Ralston JD, Taylor VM, Yasui Y, Kuniyuki A, Jackson JC, Tu SP. Knowledge of cervical cancer risk factors among Chinese immigrants in Seattle. *J Community Health* 2003;28:41-57.
5. Taylor VM, Jackson JC, Tu SP, Yasui Y, Schwartz SM, Kuniyuki A, et al. Cervical cancer screening among Chinese Americans. *Cancer Detection Prevention* 2002;26:139-45.
6. Satia-Abouta J, Patterson RE, Kristal AR, Teh C, Tu SP. Psychosocial predictors of diet and acculturation in Chinese American and Chinese Canadian women. *Ethnicity Health* 2002;7:21-39.
7. Ounpuu S, Negassa A, Yusuf S. Inter-Heart: A global study of risk factors for acute myocardial infarction. *Am Heart J* 2001;141:711-21.
8. Lauderdale DS, Jacobsen SJ, Fumer SE, Levy PS, Brody JA, Goldberg J. Hip fracture incidence among elderly Asian-American populations. *Am J Epidemiol* 1997;146:502-9.
9. Ghali WA, Knudtson ML. Overview of the Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease. On behalf of the APPROACH investigators. *Can J Cardiol* 2000;16:1225-30.

TABLE I

Revised Chinese Surname List

Au, Bean, Bui, Cai, Cao, Cha, Chai, Chak, Chan, Chang, Chao, Chau, Che, Chen, Cheng, Chiang, Chieh, Chien, Chim, Chin, Ching, Chiu, Cho, Choi, Chong, Choo, Choong, Chow, Chown, Choy, Chu, Chuang, Chui, Chun, Chung, Chui, Cui, Dai, Dong, Du, Duong, Egan, Ek, Eng, Fan, Fay, Feng, Fok, Fong, Fu, Fung, Funk, Gao, Ge, Geng, Gong, Gow, Gu, Guan, Guo, Guy, Ha, Hak, Han, Hang, Hau, He, Heng, Heu, Heung, Ho, Hon, Hong, Hou, Hsieh, Hsiung, Hsu, Hsueh, Hu, Hua, Huan, Huang, Hui, Hung, Hwang, Ing, Ip, Jan, Jain, Jiang, Jin, Joie, Joo, Jung, Kam Kan, Kang, Kao, Kat, Kau, Kew, Kim, Kip, Kit, Ko, Kok, Kong, Koo, Koy, Kuan, Kun, Kwan, Kwok, Kwong, Ky, Lai, Laing, Lam, Lan, Lau, Lem, Leu, Leung, Lew, Li, Lian, Liang, Liao, Lieu, Lim, Lin, Ling, Liou, Liu, Lo, Loi, Loo, Lou, Low, Loy, Lu, Lui, Luk, Lum, Lung, Luo, Luong, Ly, Ma, Mah, Mai, Mak, Man, Mar, Mau, Maw, Mei, Mo, Moh, Moi, Mok, Monk, Moon, Mua, Mui, Nam, Neu, Ning, Noi, Pak, Pan, Pang, Pau, Peng, Ping, Pong, Qu, Sa, Seng, Shan, Shen, Shi, Shue, Sieu, Sim, Sin, Siu, So, Sok, Song, Soo, Su, Sun, Sze, Ta, Tai, Tam, Tan, Tang, Tao, Teng, Teoh, Tiew, Ting, Tiong, To, Tom, Tsang, Tse, Tseng, Tso, Tsoi, Tsui, Tu, Tung, Wan, Wang, Way, Wei, Wey, Wing, Won, Wong, Woo, Wu, Xia, Xiao, Xie, Xing, Xiong, Xu, Yam, Yan, Yang, Yap, Yau, Ye, Yee, Yeo, Yeung, Yim, Yin, Yip, Yiu, Yo, Yong, Yoon, Yu, Yuan, Yue, Yuen, Yung, Zai, Zee, Zhai, Zhan, Zhang, Zhao, Zheng, Zhong, Zhou, Zhu, Zuk