

Smoking status by proxy and self report: rate of agreement in different ethnic groups

Ana M Navarro

Abstract

Objective—To examine the rate of agreement between proxy and self report of smoking status in Hispanics compared with other ethnic groups.

Design—Data source is the 1990 California Tobacco Survey (CTS) which includes proxy and self reported smoking status. The CTS is a random digit dialled survey conducted in 57 244 households. A sample of 10 011 adults was included in the analyses.

Main outcome measure—Percentages of agreement and disagreement between self report and report by other member of the household on smoking status.

Results—Cohen's κ coefficients of agreement on smoking status between self report and proxy report was highest in non-Hispanic whites and African Americans ($\kappa = 0.91$), followed by Asian Americans ($\kappa = 0.82$) and Hispanics ($\kappa = 0.76$). Among adults identified as current smokers by proxy, a lower percentage of Hispanics compared with non-Hispanics indicated that they were current smokers (odds ratio (OR) = 3.74, 95% confidence intervals (CI) = 3.28 to 4.20). Furthermore, agreement between proxy and self report was also lower in Hispanics of low acculturation compared with Hispanics with a high level of acculturation (OR = 0.40, 95% CI = 0 to 0.94).

Conclusions—The agreement between self reported and proxy reported smoking status is higher among non-Hispanics compared with Hispanics. Smoking rates in different ethnic groups that are estimated by telephone surveys including proxy and self report might not be comparable.

(*Tobacco Control* 1999;8:182-185)

Keywords: smoking status; Hispanics; self reporting

Introduction

Much of our knowledge about smoking behaviour relies on self report. Estimates of the prevalence of cigarette smoking as well as the consequences of smoking on public health are derived from self reported smoking behaviour. Further, the planning of activities that aim to reduce the effects of tobacco use is also primarily based on self report data. Previous investigations have examined different aspects of the validity of self report¹ as well as methods to enhance the validity of self report.² In addition, previous studies with Hispanics comparing self report with biological markers of cigarette

smoking suggest that underreporting may contribute to lower estimates of cigarette smoking prevalence³ and tobacco dependence⁴ among Hispanics compared with other ethnic groups.

Some epidemiological surveys also rely on the reports from surrogate respondents, typically members of the household or the family. Previous studies have shown a satisfactory percentage of agreement between proxy and self report of health relevant personal behaviours such as smoking.^{5,6} Indeed, the use of proxy reports seems most appropriate when their inclusion can improve the size and the representativeness of the sample.^{5,7}

In this paper, we studied the extent to which self report and report of smoking status by other members of the household are consistent across different ethnic groups. We used the 1990 California tobacco survey (CTS) to examine the rate of agreement between proxy and self report of smoking status in Hispanics compared with non-Hispanics. The CTS constitutes a unique dataset because of the large number of Hispanics in the sample. The sample size allows for comparisons between subgroups of demographic characteristics (gender and acculturation) that are particularly relevant to smoking behaviour in Hispanics. Further, the CTS includes data that allow comparisons in smoking status based on proxy and self report. In this study, we examine the extent to which the agreement between proxy and self report of smoking status is similar in Hispanics compared with non-Hispanics and in Hispanics of low acculturation compared with Hispanics with a high level of acculturation.

Methods

SURVEY DESIGN

This study used the data from the California tobacco survey that was conducted between June 1990 and February 1991.⁸ Separate samples were drawn to be representative of each of the 10 largest counties in California by population, and the remaining counties were grouped into eight regions with separate samples being drawn for each region. Statewide prevalence estimates can be obtained by combining the weighted data for all 18 regions.

The survey was conducted by telephone in either English or Spanish. Data on smoking prevalence were collected in 57 244 households that enumerated a total of 118 448 adults in the screener survey. Adults were stratified by smoking status and then randomly selected within stratum for an in-depth interview, with individuals who had smoked in the past five years having 4.3 higher probability

Department of Family and Preventive Medicine, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0622, USA; anavarro@ucsd.edu

Received 11 March 1998 and in revised form 23 December 1998
Accepted 24 January 1999.

of being selected. Detailed telephone interviews on smoking behaviour and attitudes were completed by a total of 26 815 adults over 18 years of age. The present study used a sample of 5346 non-Hispanic whites, 651 African Americans, 1918 Asian Americans, and 2096 Hispanics who met the selection criteria described in the statistical analysis section below. In addition to the survey design and to the selection criteria, the difference in sample size reflects the ethnic distribution of the population in California.⁹

SURVEY INSTRUMENTS AND OPERATIONALISATION OF MEASURES

The CTS included two survey instruments that are relevant to this study: the screener survey and the adult attitudes and practices survey. The screener survey (SCS) included questions concerning demographic information and the current smoking status of every member of a selected household. The survey was answered by one of the members of the household who was at least 18 years old. This survey constituted the source of proxy data on smoking status for this study. The adult attitudes and practices survey (AAPS) collected detailed self report data for adults 18 years and older on tobacco use and related topics as well as sociodemographic information.

SOCIODEMOGRAPHIC VARIABLES

Sociodemographic variables collected in the CTS that are relevant to this paper include gender, age, educational attainment, ethnicity, and acculturation. Self report data from the AAPS were used to operationalise these variables.

Ethnicity was defined using the following two AAPS questions: "Are you Hispanic?" and "Which of the following categories best describes your racial background? White, Black, Japanese, Chinese, Filipino, Korean, Other Asian or Pacific Islander, or American Indian or Alaskan Native." The category "Other Asian or Pacific Islander" included individuals who responded "Asian", "Mixed Asian and White", "Eurasian", "Asian Hispanic", "Samoan", "Vietnamese", or "Indonesian". For the purposes of this study, respondents were considered Hispanic if they identified themselves as such, independent of their racial background. Moreover, the ethnic background of respondents who did not identify themselves as Hispanics was classified as non-Hispanic white or African American if the respondent described her or his racial background as white or black, respectively. Individuals who described their racial background as Japanese, Chinese, Filipino, Korean, other Asian, or Pacific Islander were classified as Asian Americans. Among Hispanics, an additional question explored whether the respondents identified themselves as of Mexican origin or not.

AAPS questions allowed the operationalisation of acculturation based on the use of a native language other than English. An acculturation index on a scale from 1 to 5 was

created, averaging the answers to five questions on the use of native language among Hispanics and Asian Americans. These questions were adopted from the short acculturation scale for Hispanics.¹⁰ A value of 1 indicates exclusive use of a native language other than English and a value of 5 shows exclusive use of the English language. The acculturation index was used to differentiate individuals with low and high levels of acculturation. Hispanics were considered as having a low level of acculturation if the acculturation index was less than 3 and high acculturation if the index was 3 or more. The acculturation scale was based exclusively on language use because the only CTS questions related to acculturation were those that explored the use of native language other than English. However, as indicated in a previous study,¹¹ the acculturation index based on language use was found to be highly correlated with an alternative index including other variables relevant to acculturation (country of birth, native country of mother's respondent, and age at which individuals born abroad came to the United States.)

TOBACCO USE

The CTS included questions that allowed the operationalisation of smoking status in the SCS and the AAPS. Relevant questions in the SCS were "As far as you know, has [person] smoked at least 100 cigarettes during his/her lifetime?" and "Does [person] smoke cigarettes now?" Corresponding questions in the AAPS to establish self reported smoking status were: "Have you smoked at least 100 cigarettes in your entire life?" and "Do you smoke cigarettes now?" Individuals were considered current smokers if they had smoked at least 100 cigarettes in their lives and the answer to the CTS question about smoking now was affirmative.

STATISTICAL ANALYSES

For this study, a sample of CTS survey respondents was selected from the respondents that met the following criteria: (a) 18 years old or older; (b) self reported smoking status could be ascertained based on the AAPS; and (c) smoking status from the SCS was determined based on proxy report by another member of the household. That is, in this study we did not include individuals who had responded to the SCS themselves as only self report and not proxy information about smoking status was available for them. Consequently, the sample in this study necessarily excluded the 13.6% of respondents who were the only adult living in their households.

Statistical analyses included crosstabulations, Cohen's κ coefficients, and multiple logistic regression. Pertinent sampling weights were used in all statistical analyses. Furthermore, the 95% confidence intervals around pertinent statistics were determined using a system of replicate weights.⁸ Crosstabulations of smoking status as determined by proxy and by self report were performed separately for non-Hispanic whites, African Americans, Asian Americans, and Hispanics. A

Cohen's κ coefficient was calculated in each ethnic group to estimate inter-rater (proxy *v* self) agreement on smoking status.

Multiple logistic regression analyses were performed to compare agreement on smoking status as determined by self report and by proxy in different ethnic groups (Hispanics *v* non-Hispanics) and in Hispanics with different levels of acculturation (high *v* low). Gender was included in all analyses as an independent variable where a value of 1 was assigned to men and a value of 0 to women. In addition, when comparing Hispanics with non-Hispanics, a dichotomous variable Hispanic (value of 1) *v* non-Hispanic (value of 0) was included as an independent variable. To establish comparisons among Hispanics, two additional independent variables were included in the analyses: (a) a dichotomous variable indicating whether the respondent was of Mexican origin (value of 1) or not (value of 0), and (b) acculturation coded as high (value of 1) or low (value of 0). The dependent variable in the logistic regression analyses was smoking status by self report, that is, whether AAPS data indicated that the respondent was a current smoker (value of 1) or not (value of 0). The logistic regression modelled the probability of being a non-smoker by self report. All logistic regression analyses were adjusted for age and years of formal education and included exclusively individuals who had been identified as current smokers by proxy (SCS).

Results

Table 1 shows the interrater Cohen's κ coefficients and the percentages of agreement and disagreement between self report and proxy report in different ethnic groups. The results indicate that overall, consistency between self report and proxy report of smoking status is highest in non-Hispanic whites and African Americans ($\kappa = 0.91$) followed by Asian Americans ($\kappa = 0.82$), and Hispanics ($\kappa = 0.76$). Differences between the various ethnic groups in the percentages of agreement reflect the existing differences between ethnic groups in smoking rates. The results of the crosstabulations presented in table 1 indicate the nature of the disagreements. Across ethnic groups, the percentage of smokers by proxy and non-smokers by self report is higher than the percentage of non-smoker by proxy and smoker by self report. There are statistically significant differences between Hispanics (4.82%, 95% confidence intervals (CI) = 3.53 to 6.11) and non-Hispanic whites (1.67%, 95% CI = 1.36 to 1.98) in the percentages of individuals who were identified as smokers by

proxy and as non-smokers by self report. Specifically, among smokers by proxy, the percentage of smokers by self report is highest among African Americans (92.9%, 95% CI = 88.2 to 97.6) and non-Hispanic whites (92.9%, 95% CI = 91.7 to 94.1), followed by Asian Americans (82.8%, 95% CI = 75.6 to 90.0) and Hispanics (76.7%, 95% CI = 71.4 to 82.0).

Multivariate analyses modelled the probability of being a non-smoker by self report in individuals who had been identified as smokers by proxy. The results of the logistic regression analysis with Hispanics and non-Hispanics ($n = 5222$) indicated that disagreement between proxy and self report was statistically significantly higher among Hispanics when compared with non-Hispanics (odds ratio (OR) = 3.74, 95% CI = 3.28 to 4.20, $p < 0.001$). Differences between men and women were not statistically significant (OR = 1.02, 95% CI = 0.72 to 1.32).

The results of the logistic regression analysis in Hispanics only ($n = 770$) showed that disagreement between self and proxy report on current smoking status was statistically significantly lower in Hispanics with a high level of acculturation when compared with Hispanics with a low level of acculturation (OR = 0.40, 95% CI = 0.00 to 0.94, $p < 0.002$). The results did not indicate statistically significant differences in any of the other variables included in the model (gender and country of origin.)

Discussion

Estimates of cigarette smoking prevalence are based on surveys that largely rely on self report data. Epidemiological surveys also rely on the reports from proxy respondents, typically members of the household or the family. The results of this paper indicate that the consistency between self report and proxy report of smoking status is lower among Hispanics compared with other ethnic groups. It is unlikely that the differences in agreement between proxy and self report on smoking status can be explained by the fact that respondents had quit in the period of time that lapsed between completion of the SCS and AAPS surveys. In more than 50% of the cases the SCS (proxy) was conducted the same day as the AAPS (self report), and 90% of the AAPS questionnaires were completed no later than 15 days after the SCS.

Our results indicate that the percentage of individuals identified as current smokers by proxy and as non-smokers by self report is higher among Hispanics compared with other

Table 1 Interrater Cohen's κ coefficient and percentage of agreement and disagreement on smoking status by proxy (SCS) and self report (AAPS) in different ethnic groups

Ethnic group	Smoker by proxy/Smoker by self report		Smoker by proxy/Non-smoker by self report		Non-smoker by proxy/Smoker by self report		Non-smoker by proxy/Non-smoker by self report		κ
	%	95% CI	%	95% CI	%	95% CI	%	95% CI	
Hispanics (n = 2096)	15.87	13.79–17.94	4.82	3.53–6.11	2.74	1.06–4.42	76.58	73.51–79.64	0.761
Non-Hispanic whites (n = 5346)	21.79	20.54–23.05	1.67	1.36–1.98	1.29	1.01–1.58	75.25	73.90–76.60	0.912
African Americans (n = 651)	28.36	23.20–33.52	2.17	0.73–3.61	1.45	0.09–2.82	68.02	62.46–73.58	0.914
Asian Americans (n = 1918)	11.83	10.21–13.45	2.46	1.27–3.65	1.86	1.04–2.67	83.86	81.69–86.04	0.821

CI = confidence intervals.

ethnic groups. It is possible that the knowledge of current smoking status of the household members is less accurate among proxy in Hispanics compared with non-Hispanic households. The average number of individuals in Hispanic and Asian American households tends to be greater than in non-Hispanic white and African American households. It is possible that the knowledge of proxy informants about smoking status of other members of the household is less accurate in larger households compared with smaller households. If so, our results could be confounding ethnic background with size of the household. To explore this possibility, three subsamples were identified based on the number of adults enumerated in the household (two adults, three or four adults, and five or more adults). Crosstabulations analyses were conducted in each subsample for each of the ethnic and gender groups. Not surprisingly, the results of crosstabulations in the identified subsamples indicated that the agreement between proxy and self report tends to be greater in smaller size households. However, the robustness of the main finding of this study was confirmed because greater disagreement in Hispanics compared with non-Hispanic whites was found in all the subsamples of households of different sizes.

The discrepancy between proxy and self report is most intriguing in the cases in which the proxy describes an individual as a current smoker whereas the individual answers the survey questions as a person who has never been a smoker. Quantitative and qualitative reviews^{1 2} of studies containing comparisons between self report and biochemical measures have shown that the strongest evidence of underreporting of smoking comes from studies on reports of smoking status among adolescents or individuals in situations where there is strong pressure to stop smoking. Means and colleagues² also pointed out in their review that it would be expected that underreporting of smoking would become increasingly frequent as smoking becomes less acceptable socially. Further, there is some evidence that social desirability is a particularly important factor operating in the way Hispanics answer behavioural surveys.^{12 13} Presenting oneself in a good light by claiming socially desirable traits is an image management technique that is likely to be used by people in relatively powerless positions, or in groups that stress the presentation of a good face to outsiders,¹³ as it is the case in Hispanics.¹⁴ Therefore, it is plausible that Hispanics in general and Hispanic occasional smokers in particular are more likely than non-Hispanic whites to answer questions about smoking in a way that responds to the demand characteristics of the situation.

In summary, there is a significantly higher discrepancy in smoking status as ascertained

by self report and proxy in Hispanics compared with non-Hispanic whites. Telephone surveys including self report and surrogate respondents constitute a common method to collect information about the use of tobacco and to inform policymakers about the need to implement public health programmes in particular populations. This underlines the importance of studying whether data collected using the same procedures have a similar degree of validity and reliability in different ethnic populations. Our results indicate that there are indeed differences. Further studies including biological markers of smoking are needed to determine whether or not proxy reports of current smoking in Hispanics provide more accurate information on tobacco use prevalence in this population than self report. Public health professionals need to be aware of these issues when designing and interpreting the results of surveys. Formulating survey questions that clarify the meaning of smoking and in particular occasional smoking by proxy and self report is of particular relevance to understanding the impact of smoking on Hispanics.

This work was funded by the Tobacco Tax Health Protection Act of 1988, Proposition 99, Tobacco-Related Disease Research Program, grant number 3KT-0031. The author wishes to thank Dr Charles C Berry, Ms Elizabeth A Gilpin, Dr Robert M Kaplan, Dr Richard G Kronick, and Dr John P Pierce for their thoughtful comments on an earlier version of this paper.

- Patrick DL, Cheadle A, Thompson DC, *et al.* The validity of self-reported smoking: a review and meta-analysis. *Am J Public Health* 1994;**84**:1086-93.
- Means B, Habina K, Swan GE, *et al.* Cognitive research on response error in survey questions on smoking. *National Center for Health Statistics, Vital and Health Statistics* 1992; **6**(5).
- Coultas DB, *et al.* Discrepancies between self-reported and validated cigarette smoking in a community survey of New Mexico Hispanics. *Am Rev Respir Dis* 1988;**137**:810-14. Ref 3: list 2nd and 3rd author please
- Pérez-Stable EJ, Marin BV, Marin G, *et al.* Apparent underreporting of cigarette consumption among Mexican American smokers. *Am J Public Health* 1990;**80**:1057-61.
- Kolonel LN, Hirohata T, Nomura AM. Adequacy of survey data collected from substitute respondents. *Am J Epidemiol* 1977;**106**:476-84.
- Herrmann N. Retrospective Information from Questionnaires. I. Comparability of primary respondents and their next of kin. *Am J Epidemiol* 1985;**121**:937-47.
- Gilpin EA, Pierce JP, Cavin SW, *et al.* Estimates of population smoking prevalence: self- vs. proxy reports of smoking. *Am J Public Health* 1994;**94**:1576-9.
- Pierce J, Goodman J, Gilpin E, *et al.* *Technical report on analytic methods and approaches used in the Tobacco Use in California, 1990-1991 Report. Book 1.* San Diego, California: University of California, 1992.
- US Bureau of the Census. *Statistical abstracts of the United States.* Washington, DC: US Department of Commerce, 1993.
- Marin G, Sabogal F, Marin BV, *et al.* Development of a short acculturation scale for Hispanics. *Hisp J Behav Sci* 1987;**9**:183-205.
- Navarro AM. Cigarette smoking among Latinos in California: the California Tobacco Baseline Survey. *Ann Behav Med* 1996;**18**:238-45.
- Marin G, Triandis HC, Kashima Y, *et al.* Ethnic affirmation versus social desirability: explaining discrepancies in bilinguals' responses to a questionnaire. *J Cross-Cult Psychol* 1983;**14**:173-86.
- Ross CE, Mirowsky J. Socially-desirable response and acquiescence in a cross-cultural survey of mental health. *J Health Soc Behav* 1984;**25**:189-97.
- Marin G, Marin BV. *Research with Hispanic populations.* Newbury Park, California: Sage, 1991.