

Estimates of Population Smoking Prevalence: Self- vs Proxy Reports of Smoking Status

ABSTRACT

Objectives. In the face of rising costs of surveillance systems, it is time to reexamine the feasibility of including proxy respondents in surveys designed to provide population estimates of smoking prevalence.

Methods. Data are from the California Tobacco Surveys, which are random-digit dialed telephone surveys. One adult provided demographic information and smoking status for all household residents. Additionally, some adults were selected for in-depth interviews that also included smoking status questions. We matched information from proxy respondents and self-respondents and evaluated smoking status discrepancies between them relative to demographic and other factors ($n = 2930$ matched pairs) in 1992. We address the potential bias these discrepancies might introduce into the population estimate of smoking prevalence.

Results. Overall, the discrepancy between proxy report and self-report was 4.3%, and it increased particularly when the self-respondent reported nondaily smoking or recent quitting. Discrepancies acted in both directions, and the net bias was that the screener survey overestimated smoking prevalence by 0.1% in 1992 (0.3% in 1990).

Conclusions. Smoking status questions can be added to ongoing surveys such as the census or labor force surveys; one adult could provide smoking status for all household members. (*Am J Public Health.* 1994; 84:1576-1579)

Elizabeth A. Gilpin, MS, John P. Pierce, PhD, Shirley W. Cavin, MS, Charles C. Berry, PhD, Nicola J. Evans, PhD, Michael Johnson, PhD, and Dileep G. Bal, MD

Introduction

Cigarette smoking has been identified as one of the most preventable causes of death and disability.^{1,2} Accordingly, a surveillance system for smoking prevalence is important for many public health authorities at the state level. Establishing such a surveillance system requires addressing the issues of validity and cost.

The primary survey used to track smoking prevalence in the United States has been the National Health Interview Survey (NHIS).²⁻⁴ Because the NHIS design provides population estimates representative of four regions of the continental United States, it cannot provide reliable state-specific estimates of smoking prevalence. In 1974 the NHIS decided to collect data on smoking status only from self-respondents.⁵ Given that other types of surveys use proxy respondents, it is timely to consider the impact of including proxy respondents in the overall estimate of smoking prevalence.

Cost can be minimized if the proposed surveillance system can be incorporated into ongoing surveys, such as the population census or labor force participation surveys. In the United States, the Bureau of the Census conducts the Current Population Survey, which serves as the vehicle for evaluating state-level interventions in the large-scale American Stop Smoking Intervention Study project (ASSIST), which is a joint effort of the National Cancer Institute and the American Cancer Society.⁶ These surveys typically seek information on all household members from one respondent; that is, they include information from a proxy respondent. Most households in the United States include two or more adults, and the second adult usually represents either a different sex or a different age

group. In surveys that have sought to interview these other adults, the drop in response rate was considerable.⁷ Thus, surveys that include information from proxy respondents may lead to more representative population estimates achieved with minimal statistical adjustment. Also, the sample size is increased considerably for only a marginal increase in the cost of a single interview. This advantage becomes particularly important when the survey is to be used to estimate smoking prevalence in demographic subgroups.

In the California Tobacco Surveys, one adult reported on his or her own smoking status and acted as a proxy respondent for other adults in the household. In addition, a subsequent in-depth survey of tobacco use was undertaken in some adults who had their smoking status previously reported by proxy on the screener survey.

This study examines the extent of discrepancies in current smoking status between the proxy-self pairs and explores demographic and other factors that might be associated with higher discrepancy levels. In addition, we computed adjusted estimates of smoking prevalence based on the observed discrepancy levels so that the

Elizabeth A. Gilpin, John P. Pierce, Shirley W. Cavin, and Nicola J. Evans are with the Cancer Prevention and Control Program, Cancer Center, and Charles C. Berry is with the Department of Family and Preventive Medicine, University of California-San Diego, La Jolla, Calif. Michael Johnson and Dileep G. Bal are with the Tobacco Control Section, California Department of Health Services, Sacramento, Calif.

Requests for reprints should be sent to John P. Pierce, PhD, Cancer Prevention and Control Program, Cancer Center, University of California-San Diego, La Jolla, CA 92093-0901.

This paper was accepted February 23, 1994.

degree of bias attributable to the inclusion of proxy responses could be determined.

Methods

Surveys

The California Tobacco Surveys are population-based, random-digit dialed telephone surveys.^{8,9} In each household (n = 10 774 in 1992), an adult was administered a screener questionnaire that collected basic demographic information, including smoking status, for each household member. Each person enumerated (n = 21 870) was given a weight so that population-based estimates could be computed.⁹ Briefly, the base weight reflects the probability that the household was selected, and it is adjusted to ensure that the sample is representative of the California population with respect to age, sex, county/region, education, and race/ethnicity.

From the information enumerated on the screener interview, some adults (≥ 18 years) were selected for in-depth interviews. All persons who had smoked in the last 5 years were selected, as were a random sample of 28% of those who had not (total n = 7263). A weighting procedure similar to that used for the screener respondents was also applied to the respondents who had the in-depth interview; it further adjusted for the fact that proportionately more smokers or persons who had quit smoking in the last 5 years than nonsmokers were interviewed in depth.⁹ With this survey design, some individuals had smoking status reported both by proxy from the screener respondent and by self from the in-depth interview (n = 2930). It should be noted that 51.3% of the persons enumerated on the screener interview provided their own smoking status. In the analysis that examines the factors related to discrepancies (described below), the respondent weights from the in-depth interview are used.

Smoking Status

In the screener interview, the respondent was asked, "[As far as you know], have you [has person] smoked at least 100 cigarettes during your [his/her] lifetime?" If the answer to this question was "yes," the respondent was asked, "Do you [does person] smoke cigarettes now?" If the answer to this question was "no," the respondent was asked, "Have you [has person] quit smoking during the past 5 years, [as far as you know]?"

In the in-depth interview, each respondent was asked the same set of

TABLE 1—Proxy-Reported Smoking Status for Each Group of Self-Reported Smoking Status

Smoking Status: Proxy Report	Smoking Status: Self-Report, No. (%)				Total
	Current Daily	Current Occasional	Quit Less than 1 Year Ago	Other ^a	
Current	911 (96.3)	154 (65.3)	46 (26.4)	59 (1.6)	1170
Quit less than 5 years ago	13 (1.2)	29 (10.5)	100 (49.6)	388 (19.2)	530
Other ^a	8 (2.5)	20 (24.2)	14 (24.0)	1188 (79.2)	1230
Total	932	203	160	1635	2930

Note. Percentages are weighted with in-depth interview weights (see Methods section).
^aOther includes never smokers and longer-term former smokers.

questions. Those who had smoked 100 cigarettes and who were not smoking when surveyed were asked the date when they last smoked. In addition, current smokers were asked, "On how many days in the last month did you smoke?" Those smoking 25 or fewer days in the last month were considered occasional rather than daily smokers.¹⁰ All persons who had smoked in the last year were asked the date and length of any quit attempts during the last year.

Analytical Methods

Proxy-self pair discrepancies. For each category of self-reported smoking status (current daily, current occasional, quit in last year, nonsmoker for at least 1 year), we computed the weighted percentages of proxies reporting current smoking, quit in last 5 years, and nonsmoker for at least 5 years. A logistic regression of the likelihood of discrepancy was conducted to determine whether factors that appeared related to an increased discrepancy rate were independently related after controlling for all the other factors.¹¹ Because of the complex survey design, a jackknife resampling procedure was used to obtain variance estimates for the regression coefficients so that the independent significance of each variable could be assessed.⁹

Effect of proxy errors on estimate of smoking prevalence. To calculate the impact of discrepancies on the overall estimate of smoking prevalence, we estimated discrepancy percentages for each smoking status group that could be determined from the screener data: proxy-reported current smokers who self-reported nonsmoking (D1), proxy-reported quitters in the last 5 years who self-reported current smoking (D2), proxy-reported former smokers for at least 5 years who self-reported current smoking (D3), and proxy-reported never smokers who self-reported current smoking

(D4). In order to obtain estimates that are as accurate as possible, data from the 1990 California Tobacco Surveys were pooled with the 1992 data to estimate these discrepancy rates. (The 1990 California Tobacco Surveys surveyed 32 125 households with 65 139 adults enumerated. A total of 24 296 adults received the extended interview, yielding 9384 self-proxy pairs. The 1990 survey did not ascertain the relationship of the self- and proxy respondents, so this paper reports mainly the 1992 results.) Screener weights for the discrepant pairs were summed across surveys and divided by the sum of the sums of weights for the two surveys.

The estimate of smoking prevalence from the screener survey was computed as follows:

$$p = \frac{\sum_{i=1}^K w_i \cdot s_i}{\sum_{i=1}^K w_i}$$

where $s_i = 1$ if the respondent is a current smoker and 0 otherwise, and w_i is the weight for a person enumerated on the screener. K is the total number of adults from the screener interview. The four discrepancy rates estimated from the proxy-self pairs were used to correct the prevalence estimate from the screener survey as follows:

$$p = \left[\sum_{i=1}^{n0} w_i \cdot s_i + (1 - D1) \cdot \sum_{i=1}^{n1} w_i + D2 \cdot \sum_{i=1}^{n2} w_i + D3 \cdot \sum_{i=1}^{n3} w_i + D4 \cdot \sum_{i=1}^{n4} w_i \right] / \sum_{i=1}^K w_i$$

where $n0$ = number of self-respondents on screener (reported own smoking status and served as proxy respondent for other adult household members); $n1$ = number

TABLE 2—Proxy–Self Discrepancies in Subgroups Defined by Various Factors Related to the Household and Proxy and Self-Respondents

	No. Pairs	Total Percentage (n = 175) Discrepant ^a	<i>P</i> ^b
Overall	2930	4.3	
Age of self-respondent, y			
18–24	503	8.1	Ref
25–44	1411	3.8	.348
45+	1016	3.5	.245
Sex of self-respondent			
Male	1738	5.2	Ref
Female	1192	3.2	.451
Education of self-respondent, y			
< 12	441	8.5	Ref
12+	2489	3.9	.327
Race/ethnicity of self-respondent			
White	2003	3.3	Ref
African American	138	2.7	.320
Hispanic	596	7.3	.562
Asian/other	193	2.3	.662
Relationship of self-respondent to proxy respondent			
Child	357	6.3	Ref
Parent/guardian	303	2.7	.671
Spouse/partner	1678	2.9	.170
Sibling	147	3.7	.443
Other relative	112	6.3	.295
Unrelated	322	8.5	.439
No. adults in household			
2	1711	2.4	Ref
3–4	1052	6.3	.547
5+	165	7.5	.303
Elapsed time between interviews			
≤ 1 wk	1626	4.8	Ref
> 1 wk	1304	3.8	.263
Smoking status of proxy			
Nonsmoker	2213	4.0	Ref
Smoker	717	5.8	.429
Smoking status of self-respondent			
Never, quit for > 12 mo	1645	1.4	Ref
Quit for 1–12 mo	138	22.7	.006
Quit in last month	22	45.8	.060
Occasional	203	35.0	.001
Daily	922	4.3	.267

^aPercentages are weighted with in-depth interview weights (see Methods section).

^b*P* value gives the independent significance of the variable, adjusting for all others. Variables were coded so that variable categories can be compared to a reference ("Ref").

Factors Related to Proxy–Self Discrepancy

Overall, among the 2930 proxy–self pairs for 1992, there were 175 (4.3%, weighted percentage) discrepancies concerning current smoker or nonsmoker status (Table 2). In 105 of these cases the proxy called a self-reported nonsmoker a current smoker, and in 70 cases the proxy called a self-reported smoker a nonsmoker. When weighted, these two discrepancies represent nearly equal percentages of the proxy–self pairs, 2.1% and 2.2%, respectively. In addition to the overall discrepancy rate for each demographic category, Table 2 shows a *P* value that was obtained by using specially coded variables contrasting each category to the indicated reference in the logistic regression analysis. This *P* value indicates the independent significance of a given category after adjusting for all the other variables.

Although there were trends for more discrepancies for younger, male, Hispanic, and less-educated self-respondents, none of these factors was significantly related to the likelihood of discrepancy. Other nonsignificant trends included higher discrepancy rates when the self-respondent and proxy-respondent were unrelated or when the relationship was other than spouse/partner, parent/guardian, or sibling. Also, discrepancy rates increased with larger household size.

The discrepancy rate was not related to the smoking status of the proxy nor to the time elapsed between the screener and in-depth interviews. Some persons selected for an in-depth interview were not available when the screener interview was administered and were contacted later. If over a week elapsed between interviews, the self-respondent's smoking status might change.

When the self-respondent reported quitting in the last year, the discrepancy rate was significantly greater than that in the reference group of those who had not smoked in the last year. Finally, self-respondents who were occasional smokers showed a significantly higher rate of proxy discrepancy.

Effect of Potential Discrepancy on Estimate of California Smoking Prevalence

Smoking prevalence for California estimated from the screener survey was 20.0 ± 0.7% in 1992. The error rates D1, D2, D3, and D4 were estimated to be 12.0%, 7.8%, 1.9%, and 2.4%, respec-

of proxy-designated current smokers; *n*2 = number of proxy-designated former smokers within the last 5 years; *n*3 = number of proxy-designated former smokers who quit more than 5 years ago; and *n*4 = number of proxy-designated never smokers. Note that *K* = *n*0 + *n*1 + *n*2 + *n*3 + *n*4.

Results

Smoking Status Agreement Among Proxy–Self Pairs

Table 1 shows the number of proxy–self pairs and how they are distributed

with respect to agreement and disagreement. The percentages shown are weighted column percentages of the distribution of proxy classification for each category of self-classification. Self-respondent and proxy-respondent reports of current daily smoking agreed in 96.3% of the cases. Percentages for the "Other" possible smoking status classification are also shown. Of particular interest is that proxy respondents classified as current smokers 26.4% of self-reported quitters in the last year.

tively. The corrected smoking prevalence estimate with these percentages was 19.9%. Thus, the screener tended to overestimate smoking prevalence by 0.1% in 1992 (0.3% in 1990).

Discussion

The use of proxy respondents in population surveys of smoking status can substantially decrease the cost of obtaining a smoking prevalence estimate with minimal bias caused by potential error on the part of the proxy respondent. Inclusion of proxy respondents led to a more conservative estimate of smoking prevalence compared with the corrected estimates. The survey results are likely to be more representative because the problem of nonresponse of individuals selected for subsequent self-report does not occur. Furthermore, such a cluster sampling approach has been demonstrated to produce unbiased population estimates.¹²

The two types of discrepancies with respect to current smoking status tended to cancel one another; however, the net effect was a slight increase in the estimate of smoking prevalence when proxy respondents were included. As in a previous study,¹³ the degree of discrepancy was not great, and it should be remembered that about half of the screener adults were self-respondents. The overall nature of the discrepancies was very similar for both the 1990 and 1992 surveys (Table 1), so if the same type of screener survey instrument is used for surveillance of smoking prevalence, any bias that is introduced should be present to the same degree. Thus, any change in smoking prevalence detected with such an instrument should be unaffected by the bias.

Although the proxy-self discrepancy rate was particularly high in some cases (nondaily smokers and recent quitters), these groups comprise a relatively small fraction of the total adult population. In agreement with a previous study,¹³ we found that spouses and children tended to provide more accurate information than other relatives or unrelated individuals. This was expected as the immediate family would typically have a closer relationship with the individual and hence have more complete knowledge of his or her behavior. Hispanics appear to have a higher discrepancy rate than other racial/ethnic groups. Hispanic households are typically large, so household size may have partly accounted for this result. The more likely factor is occasional smoking, which we have previously shown to be particularly high among Hispanic smokers.^{7,10}

We assumed that the self-report of smoking status was the correct one; however, it has been demonstrated that around 30% of self-reported quitters who participated in cessation programs have positive biochemical tests indicating that they have recently smoked.¹⁴⁻¹⁶ Some programs validate self-report of recent quitting with the report of a significant other.¹⁷ In our study, 26.4% of self-reported recent quitters were reported by the proxy to be smoking (Table 1). These reporting discrepancies are similar to those identified in clinic programs. Smokers are faced with considerable social pressure to quit, which may lead to a "wish bias" in self-reporting or desire to please the interviewer. Finally, those who are trying to quit or who do not smoke daily simply might not consider themselves smokers although the proxy does.

Although smoking prevalence is usually the most important issue that needs to be addressed in assessing a population's smoking problem, information on initiation, quitting, and relapse patterns is also of interest. Others have shown that proxy respondents can be reliable for current status of health issues,¹⁸ and the screener survey instrument we described appears reliable for estimating smoking prevalence. However, because of concerns about accuracy,¹⁹⁻²³ we did not obtain detailed smoking histories and consumption levels from proxy respondents.

Our results from two large population surveys indicate that the use of proxy reporting of smoking status is cost-effective and minimally biased. Thus, it is recommended that surveillance systems for smoking prevalence use the approach of having one adult report smoking status for all adults in the household. □

Acknowledgment

This study was supported by contract 89-97872 from the California Department of Health Services, Tobacco Control Section, Sacramento, Calif.

References

1. *Smoking and Health. A Report of the Surgeon General.* Washington, DC: US Dept of Health, Education and Welfare; 1979. DHEW publication PHS 79-50066.
2. *Reducing the Health Consequences of Smoking: 25 Years of Progress. A Report of the Surgeon General.* Washington, DC: US Dept of Health and Human Services; 1989. DHHS publication CDC 89-8411.
3. *The Health Benefits of Smoking Cessation: A Report of the Surgeon General.* Washington, DC: US Dept of Health and Human Services; 1990:25-59. DHHS publication CDC 90-8416.
4. Pierce JP, Fiore MC, Novotny TE, Hatzian-

- dreu EJ, Davis RM. Trends in cigarette smoking in the United States: projections to the year 2000. *JAMA.* 1989;261:49-55.
5. National Center for Health Statistics. The National Health Interview Survey design, 1973-84, and procedures, 1975-83. *Vital Health Stat [1].* 1985;18. DHHS publication PHS 85-1320.
6. American Stop Smoking Intervention Study (ASSIST) for Cancer Prevention. December 1989. Proposal in response to RFP NCI-CN-95165-38.
7. Pierce JP, Farkas A, Evans N, et al. *Tobacco Use in California 1992. A Focus on Preventing Uptake in Adolescents.* Sacramento, Calif: California Department of Health Services; 1993.
8. Waksberg J. Sampling methods for random-digit dialing. *J Am Stat Assoc.* 1978;73:40-46.
9. Pierce J, Goodman J, Gilpin E, Berry C. *Technical Report on Analytic Methods and Approaches Used in the Tobacco Use in California, 1990-1991 Report.* Sacramento, Calif: California Department of Health Services; 1992.
10. Evans N, Gilpin E, Pierce JP, et al. Occasional smoking among adults: evidence from the California Tobacco Survey. *Tobacco Control.* 1992;1:169-175.
11. SAS Institute Inc. *SAS/STAT User's Guide,* Release 6.03 Edition. Cary, NC: SAS Institute Inc; 1988.
12. Cochran WG. *Sampling Techniques.* 2nd ed. New York, NY: John Wiley & Sons; 1963.
13. McLaughlin JK, Diety MS, Mehl EX, Blot WJ. Reliability of surrogate information on cigarette smoking by type of informant. *Am J Epidemiol.* 1987;126:144-146.
14. Ohlin P, Lundh B, Westling H. Carbon monoxide blood levels and reported cessation of smoking. *Psychopharmacology.* 1976; 49:263-265.
15. Brockway BS. Chemical validation of self-reported smoking rates. *Behav Ther.* 1978;9: 685-686.
16. Pierce JP, Dwyer T, DiGiusto E, et al. Cotinine validation of self-reported smoking in commercially run community surveys. *J Chronic Dis.* 1987;40:689-695.
17. Swartz JL. *Review and Evaluation of Smoking Cessation Methods: The United States and Canada, 1978-1985.* Washington, DC: Division of Cancer Prevention and Control, National Cancer Institute; 1987. NIH publication 87-2940.
18. Moore JC. Self/proxy response status and survey response quality: a review of the literature. *J Off Stat.* 1988;4:155-172.
19. Kolonel LN, Hirohata T, Nomura AMY. Adequacy of survey data collected from substitute respondents. *Am J Epidemiol.* 1977;106:476-489.
20. Rogot E, Reid D. The validity of data from next-of-kin in studies of mortality among immigrants. *Int J Epidemiol.* 1975;4:51-54.
21. Pickle LW, Morris Brown L, Blot WJ. Information available from surrogate respondents in case-control interview studies. *Am J Epidemiol.* 1983;118:99-108.
22. Lerchen ML, Samet JM. An assessment of the validity of questionnaire responses provided by a surviving spouse. *Am J Epidemiol.* 1986;123:481-489.
23. Machlin SR, Kleinman JC, Madans JH. Validity of mortality analysis based on retrospective smoking information. *Stat Med.* 1989;8:997-1009.