Original Investigation

Reliability of Adult Self-Reported Smoking History: Data from the Tobacco Use Supplement to the Current Population Survey 2002–2003 Cohort

Julia N. Soulakova, Ph.D.,¹ Anne M. Hartman, M.Sc.,² Benmei Liu, Ph.D.,³ Gordon B. Willis, Ph.D.,⁴ & Steve Augustine, M.Sc.¹

¹ Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE

² Risk Factors Monitoring and Methods Branch, Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD

³ Statistical Methodology and Applications Branch, Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD

⁴ Office of the Associate Director, Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD

Corresponding Author: Julia N. Soulakova, Ph.D., Department of Statistics, University of Nebraska-Lincoln, 340 Hardin Hall-North, Lincoln, NE 68583-0963, USA. Telephone: 402-472-7231; Fax: 402-472-5179; E-mail: jsoulakova2@unl.edu

Received July 12, 2011; accepted December 12, 2011

Abstract

Introduction: This study examined the reliability of self-reported smoking history measures. The key measures of interest were time since completely quitting smoking among former smokers; age at which fairly regular smoking was initiated among former and current smokers; the number of cigarettes smoked per day and the number of years of daily smoking among former smokers; and never smoking. Another goal was to examine sociodemographic factors and interview method as potential predictors of the odds of strict agreement in responses.

Methods: Data from the 2002–2003 Tobacco Use Supplement to the Current Population Survey were examined. Descriptive analysis was performed to detect discrepant data patterns, and intraclass and Pearson correlations and kappa coefficients were used to assess reporting consistency over the 12-month interval. Multiple logistic regression models with replicate weights were built and fitted to identify factors influencing the logit of agreement for each measure of interest.

Results: All measures revealed at least moderate levels of overall agreement. However, upon closer examination, a few measures also showed some considerable differences in absolute value. The highest percentage of these differences was observed for former smokers' reports of the number of years smoking every day.

Conclusions: Overall, the data suggest that self-reported smoking history characteristics are reliable. The logit of agreement over a 12-month period is shown to depend on a few sociodemographic characteristics as well as their interactions with each other and with interview method.

Introduction

When investigators assess current smoking prevalence and smoking history in the U.S. population, they commonly take for granted the high quality of national survey data. Indeed, data from the National Health Interview Survey (Centers for Disease Control and Prevention, 2009) and the Tobacco Use Supplement to the Current Population Survey (TUS-CPS; U.S. Department of Commerce, Census Bureau, 2007a) are used in numerous studies to determine tobacco use patterns in the U.S. adult population (Backinger et al., 2008; Osypuk & Acevedo-Garcia, 2010; Soulakova, Davis, Hartman, & Gibson, 2009; Tindle, Shiffman, Hartman, & Bost, 2009; U.S. Department of Commerce, Census Bureau, 2007b, 2007c). Nonetheless, the difficulties inherent in correctly answering some items and the likely sensitive nature of smoking-related questions could lead to data ambiguity. Therefore, inaccurate results can be observed even if the researchers use cutting-edge statistical methodology to analyze survey data.

To improve the design, administration, and data quality of national surveys, one should take into account respondents' cognitive and motivational processes when they answer smoking-related questions. Cognitive processing generally includes four stages: interpretation of the meaning of a question, memory search of all related information, integration of all related information, and report of the summary of this information (Tourangeau, Rips, & Rosinski, 2000, pp. 1–22). Difficulties can occur at any of these stages, for example, respondents may interpret a term or a question incorrectly, may have difficulty retrieving relevant information from memory, or may produce a response that fails to match the nature of the response category expected. Furthermore, motivational factors may influence

doi:10.1093/ntr/ntr313

Advance Access Published on February 7, 2012

© The Author 2012. Published by Oxford University Press on behalf of the Society for Research on Nicotine and Tobacco. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

responses. If respondents carry out these processes cautiously and comprehensively, they *optimize*; otherwise, they may make limited effort and *satisfice* (Krosnick, 1991, 1999; Krosnick et al., 2002).

Inaccurate responses also can be a direct result of a social desirability bias, in which respondents provide answers that comply with social norms. As a result, respondents might underreport undesirable behaviors, for example, smoking, drinking, and illicit drug use (Johnson & Mott, 2001; Kreuter, Presser, & Tourangeau, 2009; Sillett, Wilson, Malcolm, & Ball, 1978; Tourangeau & Yan, 2007; Velicer, Prochaska, Rossi, & Snow, 1992).

Several research papers have addressed the impact of a survey method on respondents' satisficing and social desirability bias (Kreuter et al., 2009; Holbrook, Green, & Krosnick, 2003) and overall data quality (Jäckle, Roberts, & Lynn, 2006). It was shown that a respondent is less likely to satisfice during a personal interview than during a phone interview (Holbrook et al., 2003).

Comparing self-reported current smoking habits with results of biochemical assessments have been used to assess validity of self-reports over several decades. Based on the evidence collected in two clinical trials conducted in the 1970s, it was concluded that a relatively large proportion of people who had been advised to quit smoking have provided deceptive answers regarding their quitting smoking (Sillett et al., 1978). A further study utilized the data from the Hispanic Health and Nutrition Examination Survey and investigated underreporting of cigarette consumption among Mexican-American smokers (Pérez-Stable, Marín, Marín, Brody, & Benowitz, 1990). It was found that underreporting was less common among moderate (10–19 cigarettes/day) and heavy (20 or more cigarettes/day) smokers than it was among light (less than 10 cigarettes/day) smokers. However, several other validation studies suggest that self-reports result in a valid estimate of smoking status in the population (Caraballo, Giovino, Pechacek, & Mowery, 2001; Patrick et al., 1994; Pierce, Aldrich, Hanratty, Dwyer, & Hill, 1987; Fortmann, Rogers, Vranizan, Haskell, Solomon, & Farquhar, 1984).

Test–retest reliability measures are commonly used to assess the degree of consistency or stability in response over repeated administrations (Carmines & Zeller, 1979; Groves et al., 2004). For a response to a question to be valid, it must by definition also be reliable. Hence, assessing reliability is a necessary step in measuring an item's validity. Several studies have addressed the reliability of self-reports by following up respondents to ask about current tobacco use behaviors and comparing them with their prior reports. One such study investigated the reliability of questions concerning current smoking status and smoking initiation age in young adults recruited to mandatory military service in Israel (Huerta, Chodick, Balicer, Davidovitch, & Grotto, 2005). It was concluded that females were more likely than males to provide reliable answers with respect to smoking initiation age. In addition, among people who claimed at recruitment that they currently smoked or had smoked in the past, about 8% claimed being never-smokers at the time of discharge. A study using data from the National Longitudinal Survey of Youth showed that the most consistent test–retest data were reported by adults, when compared with adolescents, and concerned questions targeting recent (past 2 years) events (Johnson &

Mott, 2001). In addition, self-reported age of initiation of tobacco use was concluded to be suitably reliable for epidemiological applications.

The aim of our research was to explore the reliability of reports concerning several specific self-reported smoking history attributes. The survey design consisted of twice administering the TUS-CPS to the same respondents. A subset of questions used the same wording, enabling us to assess test–retest reliability for those measures. In particular, we examined reliability with respect to (a) time since completely quitting smoking (as reported by former smokers), (b) the age at which fairly regular smoking was initiated (for current and former smokers), and (c) every day smoking habits (reported by former smokers). In addition, consistency of previously reported smoking habits and current smoking behaviors was assessed. In particular, we selected a subsample of respondents who reported in the second wave as being “never-smokers” and examined their smoking status as reported in the first wave.

We chose these particular measures because they fulfilled two criteria: (a) they satisfied the requirement of test–retest investigation that the true score cannot change over the period of observation, and (b) they are important measures for evaluating the success of tobacco control initiatives and/or important in epidemiologic and other studies of smoking dose and duration. To the extent possible, we also investigated the effects of sociodemographic characteristics, and data collection method, on exact agreement with respect to each measure.

Methods

Data

The TUS-CPS is a periodic National Cancer Institute supplement to the Current Population Survey (CPS). The CPS is sponsored by the U.S. Bureau of Labor Statistics and fielded by the U.S. Census Bureau. The TUS-CPS has generally been administered every 3 years since 1992–1993, to more than 240,000 persons in each cycle. TUS-CPS data are commonly used to assess national and state prevalence and patterns of use of tobacco products in the U.S. adult population and to collect information relevant to tobacco control policy (U.S. Department of Commerce, Census Bureau, 2007a, 2007b, 2007c). We analyzed a sample of TUS-CPS respondents who self-responded in February 2002 and in February 2003, with a time difference between surveys of 1 year $\pm$ 2 weeks (Davis et al., 2007).

The total sample size was 15,770. Respondents were, on average, aged 48 years ($SD = 17$), with 8.0% between 15 and 24, 36.7% between 25 and 44, 36.0% between 45 and 64, and 19.2% were 65 years or older. The sample was 58.9% female and 41.1% male: 6.6% Hispanic, 8.2% non-Hispanic Black, 81.9% non-Hispanic White, 2.2% non-Hispanic Asian and Pacific Islander, and 0.9% non-Hispanic American Indian and Alaska Native. In addition, 20.7% of the respondents resided in the Northeast, 27.2% in the Midwest, 28.8% in the South, and 23.2% in the West; 71.6% resided in a “metropolitan” area, 28.0% in a “non-metropolitan” area, and for 0.4% of respondents, metropolitan status was not identified.

TUS-CPS surveys were administered using a combination of in-person and telephone interviews: 56.0% of respondents

responded via telephone both times, 23.7% had in-person interviews both times, 5.8% had a phone interview in 2002 followed by an in-person interview in 2003, and 14.5% had an in-person interview in 2002 and a phone interview in 2003.

Smoking History Measures

The sample was analyzed with respect to several cigarette smoking history attributes, described below.

Measure 1: Time (in years) since completely quitting smoking, which was assessed for former smokers who did not relapse between the two surveys. This measure is based on the survey question "About how long has it been since you completely quit smoking cigarettes?"

Measure 2: Age at which fairly regular smoking was initiated, which was assessed for current and former smokers. The corresponding survey question is "How old were you when you first started smoking cigarettes fairly regularly?"

Measures 3.1–3.2: These measures concern attributes of everyday smoking reported by former smokers and were ascertained for respondents who indicated that they did not smoke between the surveys and who reported being former smokers at both 2002 and 2003 assessments. Measure 3.1 concerned the number of cigarettes smoked per day when last smoked every day, based on the question "When you last smoked every day, on average how many cigarettes did you smoke each day?" Measure 3.2 involved the total number of years smoked every day as determined by the item: "Altogether, about how many years did you smoke every day? Do not include any time you stayed off cigarettes for 6 months or longer."

Measure 4: Prior smoking status for "never"-smokers. This measure is based on the survey question "Have you EVER smoked at least 100 cigarettes in your entire life?" We determined the frequency of respondents who were consistent in that they in 2003 reported being never-smokers (a response of "No") and had also reported having been a never-smoker in 2002. In addition to the 100-cigarette item, the subsequent question, "Do you now smoke cigarettes every day, some days, or not at all?" determined current smoking (every day and some day choices) and former smoking (not at all choice).

Analysis

For continuous measures (Measures 1–3), we first assessed summary statistics concerning the difference between 2002 and

2003 responses. This difference was defined as the 2003 value minus the 2002 value (and corrected by 365 days for Measure 1). Due to a possible rounding error, produced when data recorded in days, months, and years were converted into years for Measure 1 (the time since completely quitting smoking), an exact match was defined as an absolute value of the difference in responses of no more than 6 months. For Measures 2 (age when first smoked fairly regularly) and 3 (the number of cigarettes smoked per day and the total number of years smoked every day), a match was assigned when the answers to 2002 and 2003 surveys were identical. For Measure 4 (prior smoking status among those reporting never smoked at a later time), a match was assigned for those reporting never smoking in both 2002 and 2003 and not a match for those reporting never smoking in 2003 who had reported any other smoking status in 2002.

The odds ratios (ORs) of *strict* agreement between 2002 and 2003 responses were examined through multiple logistic regressions, which is commonly used in addition to descriptive reliability measures (Belli, Traugott, Young, & McGonagle, 1999; Cowling, Johnson, & Holbrook, 2003; Gillum, 2005; Huerta et al., 2005; Johnson & Mott, 2001). Factors of interest included age, sex, race/ethnicity, region, and metropolitan status, recorded in 2002. In addition, 2002 and 2003 interview methods were considered.

Significance of all two-way interactions and three-way interactions of survey methods in 2002 and 2003 and other variables were explored, while controlling for all main effects mentioned above. Analysis was done using SAS and SUDAAN (Research Triangle Institute, 2008). All models incorporated adjustment for the complex sample design using replicate weights (Current Population Survey, 2006, chap. 14; Davis et al., 2007). To adjust for multiplicity, we incorporated the Bonferroni method to control the overall error rate at 5%.

Results

Exploratory Analysis

Tables 1 and 2 present the summary statistics and reliability measures for 2002 and 2003 responses and their differences, respectively, for Measures 1–3. Overall, these results suggest reasonably high reliability of Measures 1–3. As is illustrated in Supplementary Figures 1–4, the distributions of the differences for these measures appear to be symmetric about zero. In addition, McNemar's test ($p < .001$) indicates a significant asso-

Table 1. Summary Statistics for 2002 (Top Entry) and 2003 (Bottom Entry) Data for Measures 1–3

Measure	Sample size	Mean	SD	Median	Interquartile range
1: Time since completely quitting smoking (in years), unadjusted for 1-year difference	2,722	17.3 18.5	12.3 12.3	15.0 16.0	18.0 17.0
2: Age at which fairly regular smoking was initiated (in years)	5,390	17.7 17.7	4.5 4.5	17.0 17.0	4.0 4.0
3.1: The number of cigarettes smoked per day when a former smoker smoked every day	2,063	22.5 22.4	15.0 14.9	20.0 20.0	20.0 20.0
3.2: The number of years smoked every day for a former smoker	2,039	19.9 20.0	13.2 13.2	18.0 18.0	20.0 20.0

Table 2. Exploratory Analysis of Measures 1–3

Measure	Difference in responses		Median and interquartile range	Pearson correlation	Intraclass correlation and 95% confidence intervals
	Mean (<i>p</i> value)	<i>SD</i>			
1: Time since completely quit smoking (in years), adjusted for one-year difference	0.21 (.086)	6.50	0.00 2.00	0.86*	0.86 0.85, 0.87
2: Current and former smoker fairly regular smoking initiation age (in years)	0.05 (.187)	3.01	0.00 2.00	0.78*	0.78 0.77, 0.79
3.1: The number of cigarettes smoked per day when a former smoker smoked every day	0.17 (.459)	10.23	0.00 3.00	0.76*	0.77 0.75, 0.79
3.2: The number of years smoked every day for a former smoker	−0.10 (.581)	7.91	0.00 5.00	0.82*	0.82 0.81, 0.83

Note. *p* values correspond to testing a hypothesis of a zero-mean versus the two-sided alternative.

*Denotes $p < .001$, where the *p* value corresponds to testing a hypothesis of a zero-correlation versus the two-sided alternative.

ciation between Measure 4, the binary measures of reporting being a never-smoker in 2003 and 2002: Respondents who report being never-smokers in 2003 are likely to report that they never smoked in 2002. The kappa coefficient of 0.80 ($p < .001$) also suggests a relatively high level of agreement.

Despite the consistent findings that, on average, responses in 2002 and 2003 agree, we detected some differences in absolute values with respect to each measure. For time since completely quitting smoking (Measure 1), 58.9% of respondents reported times (adjusted for the 1-year difference) that are within a year, whereas 17.7% of responses were more than 5 years apart (Supplementary Figure 1). The 5th percentile of the difference distribution is −8 years and the 95th percentile is 9 years.

For age when first started smoking fairly regularly (Measure 2), we found that while 36.6% of respondents provided exactly the same age in 2002 and 2003, 64.9% of responses were no more than a year apart, and 87% were within 2 years (Supplementary Figure 2). Meanwhile only 5.1% of the responses were more than 5 years apart. The 5th percentile of the difference distribution is −4 years and the 95th percentile is 4 years.

Some inconsistencies were observed for the number of cigarettes smoked per day when a former smoker last smoked every day (Measure 3.1). Although 46.8% of respondents reported exactly the same number of cigarettes smoked per day in 2002 and 2003, only 56.0% of respondents reported the numbers with a difference of no more than four cigarettes (Supplementary Figure 3). The 5th percentile of the difference distribution is −15 cigarettes and the 95th percentile is 20 cigarettes.

More pronounced discrepancies were observed with respect to the number of years smoked every day among former smokers (Measure 3.2): only 23.3% of respondents reported exactly the same number of years smoked at both reports, 38.7% of responses were within a year apart, while 25.1% of responses were more than 5 years apart (Supplementary Figure 4). The 5th percentile of the difference distribution is −13 years and the 95th percentile is 11 years.

Finally, some discrepancies were associated with Measure 4. Interestingly, out of 9,502 subjects who reported in 2003 that

they never smoked, 1,023 (10.8%) respondents reported in 2002 being current or former smokers: 730 stated that they were former smokers, 193 identified themselves as everyday smokers, and 100 stated they were smoking some days.

Logistic Regression Analysis

All final models in Tables 3 and 4 are significant overall ($p < .001$). Table 3 presents the results for all levels of main effects not included in interactions. Table 4 presents the estimated conditional ORs, which were obtained using “effects” statements (SUDAAN; Research Triangle Institute, 2008, pp. 276–289), and corresponding 95% CIs adjusted via the Bonferroni method; only results for selected levels of interaction terms are presented and discussed below.

Measure 1: Time since completely quitting smoking. The final model included two-way interactions of age by region ($p = .004$), 2002 interview method by race/ethnicity ($p = .004$), 2002 interview method by region ($p = .006$), and 2003 interview method by region ($p < .001$). For those main effects that were not involved in any interaction terms, only sex was significant ($p = .043$): Male respondents were overall less likely to report the same time since completely quitting smoking at both reports than were female respondents.

Among the statistically significant conditional interaction effects we examined, we observed that Midwestern residents who had a phone interview in 2003 were less likely to provide consistent responses than were the ones who had an in-person interview in 2003 (adjusted $p = .048$), while Southern residents who had a phone interview in 2003 were more likely to provide consistent responses than were the ones who had an in-person interview in 2003 (adjusted $p = .043$). The interaction effects between age and region involve a large number of comparisons. Although some of the comparisons were significant, we do not report the detailed results here unless we noted a consistent pattern. We found: (a) younger age groups to be more reliable than each subsequent older age group in the South for reporting time since completely quitting smoking; (b) 15- to 24-year-olds to be more reliable than all other individual age groups in the Midwest for reporting this measure.

Measure 2: Age at which fairly regular smoking was initiated. The final model contained only the main effects, where only age

Table 3. Estimated Odds Ratios and Adjusted 95% CIs for Main Effects That Were Not Involved in Any Interaction Terms

Effect	Measure				
	1: Time since completely quit smoking	2: Current and former smoker fairly regular smoking initiation age	3.1: The number of cigarettes smoked per day when a former smoker smoked every day	3.2: The number of years smoked every day for a former smoker	4: Prior smoking status for never-smokers
Age (reference group: 65+)					
15–24		1.31	0.50		
		0.82, 2.10	0.09, 2.83		
25–44		0.84	1.29		
		0.65, 1.08	0.83, 2.00		
45–64		1.07	1.26		
		0.84, 1.37	0.90, 1.76		
Gender (reference group: female)					
Male	0.79	0.88	1.09		
	0.62, 0.99	0.77, 1.01	0.88, 1.33		
Race/origin (reference group: non-Hispanic White)					
Hispanic		0.66	0.47	0.41	
		0.40, 1.09	0.16, 1.34	0.07, 2.27	
Non-Hispanic Black		0.72	0.60	0.98	
		0.48, 1.08	0.30, 1.21	0.48, 2.01	
Non-Hispanic AIAN		0.64	0.23	0.49	
		0.21, 1.93	0.02, 2.57	0.04, 5.59	
Non-Hispanic API		0.89	1.25	0.44	
		0.38, 2.11	0.31, 5.12	0.06, 3.34	
Region (reference group: West)					
Northeast		0.95	1.06	1.00	0.81
		0.71, 1.26	0.75, 1.50	0.61, 1.63	0.56, 1.18
Midwest		0.87	1.02	1.09	0.95
		0.67, 1.13	0.68, 1.53	0.66, 1.80	0.68, 1.33
South		0.86	0.84	0.80	0.78
		0.64, 1.15	0.55, 1.27	0.48, 1.33	0.28, 1.20
Metropolitan status (reference group: metropolitan)					
Non-metropolitan	0.84	0.86	1.05	1.02	1.14
	0.62, 1.14	0.72, 1.03	0.77, 1.44	0.70, 1.48	0.90, 1.44
Not identified	1.01	0.46	3.53	7.71	15.04
	0.13, 7.65	0.14, 1.54	1.61, 7.72	2.53, 23.50	0.66, 343.81
2002 Interview method (reference group: in-person)					
Telephone		1.13	1.12	0.79	1.21
		0.95, 1.33	0.79, 1.58	0.55, 1.14	0.99, 1.47
2003 Interview method (reference group: in-person)					
Telephone		0.90	0.94	0.85	1.25
		0.76, 1.07	0.66, 1.34	0.56, 1.28	1.01, 1.53

Note. Significant at overall 5% level results (adjusted for multiplicity when needed) are in bold. API = Asian and Pacific Islander; AIAN = American Indian and Alaska native.

was significant ($p = .016$), but none of the individual comparisons revealed any significant differences.

Measure 3.1: The number of cigarettes smoked per day when former smoker smoked every day. Again, the final model contained only main effects. Only metropolitan status was significant ($p = .002$). However, the only significant individual comparison was between “not identified” metropolitan status and metropolitan status (adjusted $p = .001$).

Measure 3.2: The total number of years smoked when former smoker smoked every day. The final model included the interaction of sex by age ($p = .013$). Among the main

effects not included in interaction terms, only metropolitan status was significant ($p < .001$). Among the conditional interaction effects we examined, we observed that males aged 25–44 were less likely to report the same number of years smoked than were females aged 25–44 (adjusted $p = .020$).

Measure 4: Prior smoking status for “never”-smokers. The final model included two-way interactions of sex by age ($p = .017$) and sex by race/ethnicity ($p = .013$). Among the main effects not included in interactions, only the 2003 interview method was significant ($p = .036$). Among respondents who claimed in 2003 that they never smoked, those interviewed by phone in

Table 4. Estimated Odds Ratios and Corresponding Adjusted 95% CIs for Some Levels of Interaction Effects

Effect	Measure		
	1: Time since completely quit smoking	3.2: The number of years smoked every day for a former smoker	4: Prior smoking status for never-smokers
Sex and age			
Male vs. female: 15–24		1.79 0.09, 35.29	0.54 0.25, 1.17
Male vs. female: 25–44		0.41 0.19, 0.90	0.84 0.60, 1.17
Male vs. female: 45–64		0.78 0.52, 1.17	0.50 0.36, 0.70
Male vs. female: 65+		1.35 0.72, 2.53	0.46 0.28, 0.74
Sex and race/ethnicity			
Male vs. female: non-Hispanic White			0.46 0.28, 0.75
Male vs. female: non-Hispanic Black			0.45 0.20, 1.01
Male vs. female: Hispanic			0.25 0.11, 0.59
2002 Interview methods and race/ethnicity			
Telephone vs. in-person: non-Hispanic White	0.47 0.20, 1.09		
Telephone vs. in-person: non-Hispanic Black	2.85 0.55, 14.74		
Telephone vs. in-person: Hispanic	1.66 0.19, 14.51		
2002 Interview methods and region			
Telephone vs. in-person: Northeast	1.01 0.53, 1.92		
Telephone vs. in-person: Midwest	1.47 0.84, 2.57		
Telephone vs. in-person: South	0.51 0.23, 1.11		
Telephone vs. in-person: West	0.47 0.21, 1.06		
2003 Interview method and region			
Telephone vs. in-person: Northeast	0.53 0.24, 1.19		
Telephone vs. in-person: Midwest	0.50 0.25, 1.00		
Telephone vs. in-person: South	2.28 1.02, 5.10		
Telephone vs. in-person: West	1.88 0.82, 4.29		

Note. Significant at overall 5% level results (adjusted for multiplicity) are in bold; model corresponding to Measure 1 contains an additional interaction between age and region; see logistic regression analysis section.

2003 were more likely to report they never smoked in 2002 than those interviewed in-person in 2003.

As shown in Table 4, among the respondents who were from 45 to 64 years old or older (65+) and claimed in 2003 that they never smoked, male respondents were less likely to report this in 2002 than were females in the same age group (both adjusted $p < .001$). Similarly, White and Hispanic males who reported never smoking in 2003 were less likely to report they

never smoked 1 year prior than were females in the same race/ethnicity group (both adjusted $p < .001$).

Discussion

In this paper, we present results of analyses targeted at assessing the reliability of survey questions with respect to cigarette smoking history. Descriptive analysis generally revealed moderate to

excellent agreement for all measures of interest. Thus, we believe that TUS-CPS data can be used to obtain essential information regarding smoking history attributes for the U.S. population.

Despite the overall agreement patterns, we detected some degree of discrepant responses with respect to each measure. In particular, the highest percentage of discrepant results was observed for former smokers' reports of the number of years smoking every day. In addition, about 11% of the respondents who reported in 2002 being current or former smokers claimed in 2003 that they had never smoked.

There are several potential factors that could have contributed to the observed patterns of inconsistency. First, as is noted in the introduction, respondents could experience difficulties in processing the questions and, therefore, chose to satisfice. For example, when respondents answer a question concerning the total number of years smoked every day, they are instructed to exclude any time they stayed off cigarettes for 6 months or longer. The degree to which the respondent pays attention to this instruction, or works to exclude any periods of nonsmoking of 6 months or more, may vary between the two administrations, leading to inconsistent responses. Second, for fairly regular initiation of smoking there could be a forward telescoping bias, i.e., respondents report events closer to the time of interview than they occurred (Johnson & Schultz, 2005). Because this effect is likely to increase with time since the past event, we would expect differential error in reporting this event in 2002 relative to 2003. Third, there are several specific characteristics of the survey method or change of method (interviewer and time of day effects, competing distractions effects, etc.) which were not accounted for in our study. Finally, the obtained degree of agreement depends on the constructed response variable, for example, we can expect greater variability in Measure 3.1 responses (the number of cigarettes smoked per day) provided by former heavy smokers because 10 cigarettes of 2–3 packs a day is a smaller proportion than it is of 1 pack a day.

As is noted above, in the evaluated setting, response error could be a direct result of imprecise survey questions. Indeed, as is illustrated in Smoking History Measures section, the survey questions refer to somewhat general events. In particular, key words could be misinterpreted by respondents as signals to provide low-effort and imprecise (rather than taxing and exact) answers, such as “*about how long*,” “*about how many*,” “*fairly regularly*,” and “*on average*.” Thus, the wording of the questions themselves could be a contributor to inconsistent answers. A similar conclusion has been made in the literature with respect to questions with “do not know” as an answer category: presence of these questions can encourage respondent’s satisficing (Krosnick, 1991). However, cognitive testing of these questions suggested that these general terms be used because recall of smoking history information reflects approximations and not exact answers without accurate records. An additional goal of our study was to examine the odds of exact agreement for each measure as a function of a set of characteristics. This analysis revealed that sex, age, race/ethnicity, region, metropolitan status as well as interview administration mode may jointly influence the ORs. Among the significant individual comparisons by sex, a common result was observed for a given age group as well as race/ethnicity: males were less likely than were females to provide the consistent responses regarding the total number

of years smoked every day and reporting never smoking. Overall, interview method (telephone vs. in-person) did not produce consistent significant effects on the response, unlike the effect that has been observed with respect to reported smoking prevalence at one point in time (Soulakova et al., 2009).

Throughout the paper, we assume that a respondent’s answer may be valid only if it is also consistent across repeated survey administrations, such as here, in 2002 and 2003. In other words, reliability is necessary, although not sufficient for demonstrating validity. For example, we may find that, for a particular respondent the age of initiating fairly regular smoking reported in 2002 is exactly the same as the one reported in 2003. Such a correspondence of reports, however, does not necessarily imply that this reported age is, in fact, accurate as both reports may be incorrect.

Our results may have direct implications about choosing among alternative measures in analyses of tobacco use, for example, for estimating the total duration of smoking fairly regularly among former smokers. Given that the reliability of the age of initiation of fairly regular smoking (reported by former and current smokers), and that of the time since completely quitting smoking, were higher than the reliability for the number of years a former smoker smoked every day, our findings suggest that it may be preferable to construct the duration of total smoking by using the former two measures, rather than depending solely on former smokers’ direct reports of the number of years smoked every day—even allowing that there may be gaps of nonsmoking within this computed period for some people.

Future research on the reliability of survey reports of tobacco use can be targeted toward further examination of the odds of strict agreement, as a function of demographic or other respondent characteristics. For example, we observed that males were less likely to provide consistent answers than were females when reporting the time since completely quitting smoking. Therefore, it is essential to identify the underlying reasons for this observed gender effect, for example, it could be that males are more likely to satisfice than are females when answering smoking-related questions.

It is also important to investigate other criteria for defining agreement, such as through relaxing requirements for strict agreement (e.g., allowing a difference of up to 2 years between test and retest responses regarding the number of years smoked everyday by former smokers, when establishing matching rules). One problem with this type of relaxation, however, is that such a definition may not be reasonable across all population groups. For example, a 2-year difference in responses for older populations might not be as meaningful as it is for young adults. Thus, suitable definitions should be constructed with respect to subgroups if one wishes to examine the odds of such an agreement.

In addition, it would be of interest to investigate factors that may contribute to the odds of producing highly contradictory responses at two time points. There could be a set of sociodemographic, smoking behavioral, or survey administration characteristics associated with response discordance. For example, it might be the case that more cigarettes smoked per day, more discrepant responses about the number of cigarettes smoked per day are provided. For never smoked reporting, it could be the length of time since completely quit smoking among former smokers that effects the reliability of reporting former versus never smoking.

Supplementary Material

Supplementary Figures 1–4 can be found online at <http://www.ntr.oxfordjournals.org>

Funding

This work was supported by the NCI at the National Institutes of Health (sponsor award number HHSN261200900395P) awarded to JNS.

Declaration of Interests

None declared.

Acknowledgments

The authors would like to thank the Deputy Editor and both reviewers for their valuable feedback that helped improve the manuscript, senior program analyst James Gibson for providing the data set and science writer Anne Brown Rodgers for her editorial suggestions.

References

- Backinger, C. L., Lawrence, D., Swan, J., Winn, D. M., Breen, N., & Hartman, A., et al. (2008). Using the National Health Interview Survey to understand and address the impact of tobacco in the United States: Past perspectives and future considerations. *Epidemiologic Perspectives & Innovations*, 5, 8. doi:10.1186/1742-5573-5-8
- Belli, R. F., Traugott, M. W., Young, M., & McGonagle, K. A. (1999). Reducing vote overreporting in surveys. *Public Opinion Quarterly*, 63, 90–108. doi:10.1086/297704
- Caraballo, R. S., Giovino, G. A., Pechacek, T. F., & Mowery, P. D. (2001). Factors associated with discrepancies between self-reports on cigarette smoking and measured serum cotinine levels among persons aged 17 years and older. *American Journal of Epidemiology*, 153, 807–814. doi:10.1093/aje/153.8.807
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Sage University paper series on quantitative applications in the social sciences, 07-004. Newbury Park, CA: Sage.
- Centers for Disease Control and Prevention (2009). *About the National health interview survey*. Retrieved from http://www.cdc.gov/nchs/nhis/about_nhis.htm
- Cowling, D. W., Johnson, T. P., & Holbrook, B. C. (2003). Improving the self reporting of tobacco use: Results of a factorial experiment. *Tobacco Control*, 12, 178–183. doi:10.1136/tc.12.2.178
- Current Population Survey. (2006). *US Census Bureau. Current population survey: Design and methodology* (Technical paper 66). Retrieved from <http://www.census.gov/prod/2006pubs/tp-66.pdf>

Davis, W. W., Hartman, A. M., & Gibson, J. T. (2007). *Weighting the overlap sample obtained from two tobacco use supplements to the Current Population Survey*. Retrieved from http://riskfactor.cancer.gov/studies/tus-cps/TUS-CPS_overlap.pdf

Fortmann, S. P., Rogers, T., Vranizan, K., Haskell, W. L., Solomon, D. S., & Farquhar, J. W. (1984). Indirect measures of cigarette use: Expired-air carbon monoxide versus plasma thiocyanate. *Preventive Medicine*, 13, 127–135. doi:10.1016/0091-7435(84)90045-8

Gillum, R. F. (2005). Religiosity and the validity of self-reported smoking: The Third National Health and Nutrition Examination Survey. *Review of Religious Research*, 47, 190–196. doi:10.2307/3512050

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. New York: Wiley.

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly*, 67, 79–125.

Huerta, M., Chodick, G., Balicer, R. D., Davidovitch, N., & Grotto, I. (2005). Reliability of self-reported smoking history and age at initial tobacco use. *Preventive Medicine*, 41, 646–650. doi:10.1016/j.ypmed.2005.01.011

Jäckle, A., Roberts, C. E., & Lynn, P. (2006). *Telephone versus face-to-face interviewing: Mode effects on data quality and likely causes (report on phase II of the ESS-Gallup mixed mode methodology project)* (ISER working paper 41). Colchester, UK: University of Essex. Retrieved from <http://www.iser.essex.ac.uk/publications/working-papers/iser/2006-41.pdf>

Johnson, E. O., & Schultz, L. (2005). Forward telescoping bias in reported age of onset: An example from cigarette smoking. *International Journal of Methods in Psychiatric Research*, 14, 119–129. doi:10.1002/mpr.2

Johnson, T. P., & Mott, J. A. (2001). The reliability of self-reported age of onset of tobacco, alcohol and illicit drug use. *Addiction*, 96, 1187–1198. doi:10.1046/j.1360-0443.2001.968118711.x

Kreuter, F., Presser, S., & Tourangeau, R. (2009). Social desirability bias in CATI, IVR, and WEB surveys. The effects of mode and question sensitivity. *Public Opinion Quarterly*, 72, 1–19. doi:10.1093/poq/nfn063

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213–236. doi:10.1002/acp.2350050305

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50, 537–567. doi:10.1146/annurev.psych.50.1.537

Krosnick, J. A., Holbrook, A. L., Brent, M. K., Carson, R. T., Hanemann, W. M., Kopp, R. J., et al. (2002). The impact of “no opinion” response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly*, 66, 371–403.

- Osyupuk, T. L., & Acevedo-Garcia, D. (2010). Support for smoke-free policies: A nationwide analysis of immigrants, US-born, and other demographic groups, 1995–2002. *American Journal of Public Health, 100*, 171–181. doi:10.2105/AJPH.2009.160218
- Patrick, D. L., Cheadle, A., Thompson, D. C., Diehr, P., Koepsell, T., & Kinne, S. (1994). The validity of self-reported smoking: A review and meta-analysis. *American Journal of Public Health, 84*, 1086–1093. doi:10.2105/AJPH.84.7.1086
- Pérez-Stable, E. J., Marín, B. V., Marín, G., Brody, D. J., & Benowitz, N. L. (1990). Apparent underreporting of cigarette consumption among Mexican American smokers. *American Journal of Public Health, 80*, 1057–1061.
- Pierce, J. P., Aldrich, R. N., Hanratty, S., Dwyer, T., & Hill, D. (1987). Uptake and quitting smoking trends in Australia 1974–1984. *Preventive Medicine, 16*, 252–260. doi:10.1016/0091-7435(87)90088-0
- Research Triangle Institute. (2008). *SUDAAN language manual, release 10.0*. Research Triangle Park, NC: Author.
- Sillett, R. W., Wilson, M. B., Malcolm, R. E., & Ball, K. P. (1978). Deception among smokers. *British Medical Journal, 2*, 1185–1186. doi:10.1136/bmj.2.6146.1185
- Soulakova, J. N., Davis, W. W., Hartman, A. M., & Gibson, J. T. (2009). The impact of survey and response modes on current smoking prevalence estimates using TUS-CPS: 1992–2003. *Survey Research Methods, 3*, 123–137.
- Tindle, H. A., Shiffman, S., Hartman, A. M., & Bost, J. E. (2009). Switching to “lighter” cigarettes and quitting smoking. *Tobacco Control, 18*, 485–490. doi:10.1136/tc.2008.029314
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*, 859–883. doi:10.1037/0033-2909.133.5.859
- U.S. Department of Commerce, Census Bureau. (2007a). *National Cancer Institute and Centers for Disease Control and Prevention co-sponsored tobacco use supplement to the current population survey*. Retrieved from <http://riskfactor.cancer.gov/studies/tus-cps>
- U.S. Department of Commerce, Census Bureau. (2007b). *National Cancer Institute and Centers for Disease Control and Prevention co-sponsored tobacco use supplement to the current population survey. What are the current and past TUS survey findings?* Retrieved from <http://riskfactor.cancer.gov/studies/tus-cps/results.html>
- U.S. Department of Commerce, Census Bureau. (2007c). *National Cancer Institute and Centers for Disease Control and Prevention co-sponsored tobacco use supplement to the current population survey reports and publications using the TUS-CPS*. Retrieved from <http://riskfactor.cancer.gov/studies/tus-cps/publications.html>
- Velicer, W. F., Prochaska, J. O., Rossi, J. S., & Snow, M. D. (1992). Assessing outcome in smoking cessation studies. *Psychological Bulletin, 111*, 23–41. doi:10.1037//0033-2909.111.1.23