



## Practice of Epidemiology

### Validity of Recall of Tobacco Use in Two Prospective Cohorts

**Janet Brigham\***, Christina N. Lessov-Schlaggar, Harold S. Javitz, Ruth E. Krasnow, Elizabeth Tildesley, Judy Andrews, Hyman Hops, Marie D. Cornelius, Nancy L. Day, Mary McElroy, and Gary E. Swan

\* Correspondence to Dr. Janet Brigham, Center for Health Sciences, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025 (e-mail: [janet.brigham@sri.com](mailto:janet.brigham@sri.com)).

*Initially submitted January 28, 2010; accepted for publication May 17, 2010.*

This project studied the convergent validity of current recall of tobacco-related health behaviors, compared with prospective self-report collected earlier at two sites. Cohorts were from the Oregon Research Institute at Eugene ( $N = 346$ , collected 19.5 years earlier) and the University of Pittsburgh, Pennsylvania ( $N = 294$ , collected 3.9 years earlier). Current recall was examined through computer-assisted interviews with the Lifetime Tobacco Use Questionnaire from 2005 through 2008. Convergent validity estimates demonstrated variability. Validity estimates of some tobacco use measures were significant for Oregon subjects (age at first cigarette, number of cigarettes/day, quit attempts yes/no and number of attempts, and abstinence symptoms at quitting; all  $P < 0.03$ ). Validity estimates of Pittsburgh subjects' self-reports of tobacco use and abstinence symptoms were significant ( $P < 0.001$ ) for all tobacco use and abstinence symptoms and for responses to initial use of tobacco. These findings support the utility of collecting recalled self-report information for reconstructing salient lifetime health behaviors and underscore the need for careful interpretation.

data collection; mental recall; prospective studies; reproducibility of results; retrospective studies; tobacco use disorder

Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient; LTUQ, Lifetime Tobacco Use Questionnaire.

Public health strategies to estimate tobacco use, its precursors, its correlates, and its sequelae are of considerable concern not only because of adverse effects of tobacco use, but also because tobacco use is intertwined with other health-related behaviors (1). Accurate tobacco use history is necessary for developing biomarkers of harm and exposure critical in clinical and regulatory research. Tobacco research often uses systematic measures (2) to facilitate understanding of the natural history of usage trends, to identify research and treatment needs, and to evaluate prevention and control programs.

Another example of heavy reliance on recall of smoking lies in recent genome-wide association studies of measures of nicotine dependence (3–6). While engaging in high-level quality control of genomic data, all of the studies relied on retrospective self-report of nicotine dependence or smoking quantity without reference to the reliability, validity, or re-

call interval of these phenotypes. Questions remain regarding whether the reported association between variation in the nicotinic acetylcholine receptor gene cluster on chromosome 15q25.1 and measures of nicotine dependence are independent of the locus' association with diseases such as lung cancer. Variation in unreported measurement properties of methods of assessing these phenotypes may partly explain the difficulty in teasing apart the underlying causal model.

Tobacco use is one of many health-related behaviors for which epidemiologic research often relies on self-report and recall, with or without biologic verification. Surveillance surveys often include questions about ever use of common forms of tobacco (cigarettes, smokeless, cigars, pipe), age at first tobacco use or first cigarette (initiation), extent of current use, number of quit attempts, and related health behaviors (7).

Most research in tobacco use history relies on recall of experiences. Although the quality of self-reports may depend on strategies to reduce reporting error and optimize recall, retrospective research can overcome some of the limitations of prospective research, in part because perceptions of privacy and confidentiality can affect prospective reporting (8). Underreporting of tobacco use can be related to interaction with an interviewer (e.g., social desirability bias). Even the standard of biochemical testing, which remains the most common method of validating tobacco self-report, is useful only for validating or quantifying exposure for a specific duration. Thus, a repeated-measures prospective study with biologic testing might not capture episodic use of tobacco (8).

The purpose of this project was to examine the validity of questions in the authors' Lifetime Tobacco Use Questionnaire (LTUQ). To accomplish this, the investigators examined convergent validity of separate measures of the same events, studied first in adolescence and later in young adulthood or adulthood. Specifically, the present analyses examined the convergent validity of a) tobacco use measures administered initially in two separate prospective, repeated-measures cohort studies, compared with b) similar measures administered to the same cohorts with the LTUQ between 2005 and 2008. Because of limitations inherent in self-report, neither the prospective data nor the recent interview data were considered a "gold standard." Only a small portion of the prospective data reported real-time measures and in most cases involved near-term retrospective recall of events occurring within a period ranging from a week to several years.

This study builds on earlier psychometric work (9, 10) that identified moderate to high test-retest reliability of self-administration of earlier and later versions of the study questionnaire in separate Web samples. The present analyses addressed the following research questions: a) What is the convergent validity of tobacco use responses when original prospective responses are compared with later LTUQ responses? and b) What question-related factors appear to moderate the validity of recall? A related goal was to explore characteristics of the instrument.

## MATERIALS AND METHODS

### Recruitment

Subjects were recruited from cohorts previously studied in two prospective projects: 1) the Smoking in Families Study (346 of 483 original subjects, 71.6%) at the Oregon Research Institute in Eugene; and 2) a prenatal tobacco exposure group in Pittsburgh, Pennsylvania (301 of 426 original subjects, 70.7%). The use of these two cohorts provided long-term (Oregon) and near-term (Pittsburgh) estimates of convergent validity. Started in 1981, the Oregon study was a repeated-measures (10 time points at approximately one-year intervals) cohort study of substance use risk factors. The Pittsburgh study was part of the Maternal Health Practices and Child Development Study. The cohort was selected at the fourth month of gestation and was followed into young adulthood (11 time points) (11–13).

The institutional review boards of SRI International, Oregon Research Institute, and the University of Pittsburgh approved the study. Written informed consent was obtained from all participants, in addition to the informed consent obtained for their original participation in the Oregon and Pittsburgh research studies. Previous cohort members were contacted by written invitations. Oregon participants were contacted with follow-up telephone contact as needed.

Because the two cohorts differed substantially in demographics, as indicated in Table 1, the data sets were analyzed separately.

### Administration

*Original data collection.* The original Oregon data were collected through in-person interviews, with questionnaires completed by child/adolescent participants and their families. The average interval between the first Oregon testing and the LTUQ was 19.5 years (standard deviation, 0.6 years). The original Pittsburgh data were collected through in-person interviews. Average time from original Pittsburgh adolescent testing to LTUQ administration was 3.9 years (standard deviation, 0.8).

*Data collection using the LTUQ.* Programming of the Web-based LTUQ was started in 2004, with major programming revisions and minor text revisions following alpha and beta testing. Programming and hosting were completed by research software company CfMC, San Francisco, California.

Interviewers were trained during on-site visits by the investigators in September and October 2004. One Oregon interviewer was trained at SRI. Data collection began in 2005 and continued through August 2008. Interviewers were monitored, interim data were analyzed, and results were discussed at annual on-site visits. Minor interviewing modifications were made at the 2006 site visits to increase equivalence of participant prompting. Although data were not analyzed across both studies, the investigators and staff endeavored to achieve equivalent interview practices.

The LTUQ was administered primarily by telephone by trained interviewers. Forty-nine Pittsburgh subjects were interviewed in person because of inadequate telephone access or privacy concerns. Seven Oregon subjects self-administered the LTUQ online because of miscommunication. Since the LTUQ was written for both interviewer and self-administration, and since the data from those 7 were not outliers, the data were retained.

Interviewers entered a unique pass code for each subject at the outset of administering the questionnaire, which was hosted on a secure Web site. Researchers received all data without personal identifiers. Responses were encoded and collected on secure central servers and decoded offline before the data were provided to the investigators.

### Measures

*Prospective.* Data used in the analysis of the Oregon prospective data were a subset of the 10 Oregon time points. Prospective Pittsburgh data involved the most participating subjects at about age 16 years, at time point 10. The time

**Table 1.** Subject Characteristics and Ever Use of Tobacco, Study 1 and Study 2, Oregon Research Institute, Eugene and University of Pittsburgh, Pennsylvania, 2005–2008

	Study 1: Oregon					Study 2: Pittsburgh					
	Initial Testing <sup>a</sup>		Subsequent Testing <sup>b</sup>			Initial Testing <sup>a</sup>			Subsequent Testing <sup>b</sup>		
	No.	%	No.	%	Mean (SD)	No.	%	Mean (SD)	No.	%	Mean (SD)
Sex: male	346	41.3	346	41.2		294	58.8		294	59.2	
Age, years <sup>c</sup>			346		32.7 (1.5)	283		16.8 (0.6)	296		20.8 (0.8)
Race/ethnicity: white, non-Hispanic	346	87.9	297	85.6		294	48.0		294	53.4	
Education: post-high school degree <sup>d</sup>			346	41.9					296	0.7	
Tobacco ever use											
Cigarettes	346	59.5	346	77.5		283	91.9 <sup>e</sup>		296	88.5 <sup>e</sup>	
Smokeless			346	33.8					296	18.6	
Cigars			346	37.0					296	25.0	
Pipe			346	13.9					296	3.0	

Abbreviation: SD, standard deviation.

<sup>a</sup> Initial testing was with the Smoking in Families Study (Oregon) and the Maternal Health Practices and Child Development Study (Pittsburgh).

<sup>b</sup> Subsequent testing was with the Lifetime Tobacco Use Questionnaire.

<sup>c</sup> Since the prospective data collection period for study 1 spanned ages 10 to 27 years, initial self-report age was not calculated for this table.

<sup>d</sup> At initial testing, the low mean age of Pittsburgh participants precluded post-high school education. An analysis of educational-level effects (Web Table 4) uses Pittsburgh ages at subsequent, rather than initial, testing.

<sup>e</sup> At initial testing, 260 of 283 endorsed ever use of cigarettes; at subsequent testing, 262 of 296 endorsed ever use.

points and ages used in the present analyses are noted in Tables 1 through 4.

All prospective data were collected by in-person interviews. Prospective questionnaires included items listed in Tables 2 and 3. Specific wording for prospective and LTUQ questions is available online in Web Table 1, posted on the *Journal's* Web site, <http://aje.oupjournals.org>.

**LTUQ.** The LTUQ retrospectively assessed the use of tobacco or nicotine across the lifespan. Developed initially in 1997–1998, the LTUQ was tested in 3 earlier versions on approximately 1,700 respondents through computer-assisted self-interviewing, computer-assisted telephone interviewing, computer-assisted personal interviewing, and usability testing. Previously published reports on two-year test-retest reliability (9) and two-month test-retest reliability (10) indicated that the LTUQ had high reliability for salient tobacco-related questions.

LTUQ programming utilized computerized features including skip logic, branching, and loops to shorten testing time and minimize attrition. Response options were randomized and rotated to reduce sequence effects and carry-over/practice effects, with some response options anchored for consistency. The questionnaire included internal validity checks, accuracy checks, and response limitations that either prevented respondents from entering certain types of inaccurate data or flagged those responses for later examination. The LTUQ was available only in computerized administration mode.

The LTUQ was structured around a *core questionnaire* that assessed the extent and nature of tobacco use from earliest exposure to the point of testing. Questions covered 4 major types of tobacco—cigarettes, cigars, smokeless, and

pipe—and included an open-ended response option for other tobacco delivery methods. In addition to the core questions, *module* questions examined risk and protective factors related to tobacco use. The core tobacco-use questions assessed initial use, transition to weekly and daily use, quit attempts, and abstinence. Modules of additional questions addressed subjective reactions to initial use (9, 10).

### Statistical analyses

Because of heterogeneity between the two prospective cohorts, data were not analyzed across studies. Within each study, the original data (from age ranges specified in Tables 2 through 4) were compared with LTUQ responses on similar questions with Spearman's rank correlation coefficient ( $r_s$ ), a nonparametric alternative to the Pearson's  $r$  correlation; with intraclass correlation coefficient (ICC), used for group comparisons of consistency; or with polychoric correlation, used in ordinal scales and in self-report scales with limited option ranges. Confidence intervals and probability values also were calculated.

Equality of the Spearman's correlation was tested by applying a variance-stabilizing arctangent transformation and calculating a two-sample  $t$  statistic for the equality of means of the transformed variables. The equality of ICCs was tested by calculating a two-sample  $t$  statistic using the estimated ICCs and their estimated standard error.

The two prospective data sets differed in question content, although both overlapped with the LTUQ. Pittsburgh data included some question sets that were not in the Oregon data, such as the response to initial use of tobacco. Similarly, specific analyses were conducted on the Oregon data, but

**Table 2.** Validity of Tobacco Use History Measures, Study 1: Oregon Research Institute, Eugene, Oregon, 2005–2008

	No.	Initial Testing <sup>a</sup>		Subsequent Testing <sup>b</sup>		Spearman's $r_s$ or ICC	95% Confidence Interval
		Mean (SD)	%	Mean (SD)	%		
Age at first cigarette	163	10.3 (2.5)		12.6 (3.4)		ICC = 0.33*	0.19, 0.47
Age at first smokeless	14	10.4 (2.87)		12.1 (5.6)		ICC = 0.70*	0.50, 0.90
Age at first weekly smoking	59	14.9 (1.3)		15.0 (3.7)		$r_s = 0.07$	-0.19, 0.32
Number of cigarettes/week, weekly smoking	164	18.3 (43.0)		38.1 (38.0)		$r_s = 0.14$	-0.02, 0.28
Age at first daily smoking	49	15.1 (1.2)		14.4 (2.4)		$r_s = 0.01$	-0.28, 0.28
Number of cigarettes/day, daily smoking	89	10.4 (9.9)		7.4 (5.7)		$r_s = 0.31^{**}$	0.11, 0.49
Ever tried to quit	157		73.9		65.0	$r_s = 0.38^{**}$	0.14, 0.62
Number of times tried to quit	176	0.3 (0.5)		2.5 (2.8)		$r = 0.33^*$	0.19, 0.46
Sought help in quitting, compared with first LTUQ quit attempt	142		14.1		12.0	$r_s = 0.06$	-0.49, 0.36
Sought help in quitting, compared with other LTUQ quit attempt	142		14.1		21.1	$r_s = 0.27$	-0.06, 0.60
Ever experienced abstinence symptoms	147		38.1		88.4	$r_s = 0.46^{**}$	0.15, 0.77

Abbreviations: ICC, intraclass correlation coefficient; LTUQ, Lifetime Tobacco Use Questionnaire; SD, standard deviation.

\*  $P = 0.00$ ; \*\* $P < 0.01$ , comparing initial testing with subsequent testing.

<sup>a</sup> Initial testing was with the Smoking in Families Study.

<sup>b</sup> Subsequent testing was with the Lifetime Tobacco Use Questionnaire.

not the Pittsburgh data, to study Oregon questions that did not have direct corollaries between the prospective testing and the LTUQ administration. Some measures from initial

testing were imputed from repeated questions that indicated the year that specific tobacco events occurred (e.g., initiation of daily smoking). Some Oregon questions were not closely

**Table 3.** Validity of Tobacco Use History Measures, Study 2: University of Pittsburgh, Pittsburgh, Pennsylvania, 2005–2008

	No.	Initial Testing <sup>a</sup> , Mean (SD)	Subsequent Testing <sup>b</sup> , Mean (SD)	Spearman's $r$ or ICC	95% Confidence Interval
Age at first cigarette	149	13.1 (2.3)	13.4 (2.6)	ICC = 0.58*	0.48, 0.68
Age at first smokeless <sup>c</sup>	7				
Age at first weekly smoking	106	14.6 (1.4)	15.0 (2.0)	ICC = 0.49*	0.35, 0.63
Number of cigarettes/week at first weekly smoking	127	8.2 (7.2)	5.7 (5.3)	$r_s = 0.40^*$	0.25, 0.54
Age at first daily smoking	95	14.6 (1.5)	15.0 (2.0)	ICC = 0.52*	0.38, 0.66
Number of cigarettes/day at first daily smoking	111	8.8 (7.4)	7.1 (5.4)	$r_s = 0.32^*$	0.14, 0.48
Minutes to first cigarette of day in daily smoking	120	125.7 (177)	109.3 (113)	$r_s = 0.37^*$	0.20, 0.50
Response to first tobacco <sup>d</sup>					
Dizzy	152	1.44 (1.2)	3.26 (1.5)	$r_s = 0.49^*$	0.35, 0.64
Nausea	148	0.72 (1.1)	2.49 (1.6)	$r_s = 0.51^*$	0.34, 0.67
Coughing	149	1.24 (1.2)	3.09 (1.5)	$r_s = 0.42^*$	0.26, 0.58
Relaxed	144	0.86 (1.0)	2.60 (1.4)	$r_s = 0.16^*$	-0.03, 0.35
Rush/buzz	148	0.91 (1.1)	3.07 (1.5)	$r_s = 0.31^*$	0.13, 0.49
Difficulty inhaling	146	0.83 (1.1)	2.62 (1.6)	$r_s = 0.32^*$	0.14, 0.51

Abbreviations: ICC, intraclass correlation coefficient; SD, standard deviation.

\*  $P = 0.00$ , comparing initial testing with subsequent testing.

<sup>a</sup> Initial testing was with the Maternal Health Practices and Child Development Study.

<sup>b</sup> Subsequent testing was with the Lifetime Tobacco Use Questionnaire.

<sup>c</sup> No calculations reported because of small cell size.

<sup>d</sup> Initial measures were rated on a 0–3 scale; LTUQ measures were rated on a 1–5 scale.

**Table 4.** Consistency of Tobacco Use History Responses, Study 1: Oregon Research Institute, Eugene, Oregon, 2005–2008<sup>a</sup>

Age at Initial Self-report <sup>b</sup> , Compared with Subsequent Report <sup>c</sup>	No.	Actual % Consistent Responses	Random % Consistent
Ever used cigarettes			
Age 10–16 years	339	94.1*	54.3
Age 11–17 years	335	92.2*	46.6
Age 12–18 years	329	92.2*	43.5
Amount smoked			
Age 10–16 years	339	94.1*	59.6
Age 11–17 years	335	96.1*	49.7
Age 12–18 years	329	94.1*	48.3
Ever used smokeless			
Age 10–16 years	340	100.0*	77.9
Age 11–17 years	335	90.2*	70.2
Age 12–18 years	328	96.1*	69.2
Amount of smokeless tobacco used			
Age 10–16 years	186	94.1*	68.3
Age 11–17 years	182	96.1*	60.4
Age 12–18 years	175	94.1*	57.1

\*  $P = 0.00$ , comparing actual percentage consistent with random percentage consistent.

<sup>a</sup> No parallel analyses were conducted on the Pittsburgh, Pennsylvania, data.

<sup>b</sup> Initial testing was with the Smoking in Families Study.

<sup>c</sup> Subsequent testing was with the Lifetime Tobacco Use Questionnaire.

related enough to the LTUQ for validity analysis per se. Consequently, for some questions (Table 4), an LTUQ question was compared with available responses from the Oregon data. For example, the LTUQ question *Have you ever used cigarettes?* had no direct corollary in the sequence of Oregon questionnaires, but a positive or negative response could be implied from other Oregon responses indicating cigarette use. (See Web Table 1.) The consistency of responses was evaluated (Table 4) to calculate the percentage of responses consistent with LTUQ responses and to compare that percentage with the expected percentage that would be consistent if responding were random.

## RESULTS

The results supported the variability of validity estimates of retrospective associations and considered the utility of using retrospective assessment to study specific earlier events and behaviors. Results were interpreted in light of general guidelines (14, p. 133) establishing kappa and correlation values as follows: <0.40, poor; 0.40 to 0.59, fair; 0.60 to 0.74, good;  $\geq 0.75$ , excellent. (All  $P$  values are 2-sided.)

No systematic differences related to education, race, and sex were found; the only comparison differing at the  $P < 0.01$  level was a race difference in difficulty inhaling at first use of tobacco, with nonwhites showing less consistent responding across testing times. These results are in-

cluded as online-only Web Tables 2–4. In the Pittsburgh education comparison, age reporting was from the LTUQ, since the young adults were pre–high school graduation age at their original assessment.

## Demographic comparisons

The convergent validity of initial and LTUQ measures of milestones in tobacco use history differed in minor ways by sex (Web Table 2), race (Web Table 3), or educational level (Web Table 4). Male participants in the Pittsburgh study ( $t = 2.11$ ,  $P = 0.04$ ) and white participants in both Pittsburgh and Oregon studies ( $t = 2.17$ ,  $P = 0.03$ ) demonstrated higher validity reporting the age at first weekly smoking. White participants ( $t = 2.30$ ,  $P = 0.02$ ) and those with post–high school education ( $t = 2.00$ ,  $P = 0.05$ ), all from Pittsburgh, reported the “dizzy” response with higher validity. White participants from Pittsburgh also reported “difficulty inhaling” with higher validity ( $t = 2.60$ ,  $P = 0.01$ ). Pittsburgh participants with a high school education or less demonstrated higher validity reporting age at first cigarette use ( $t = 2.07$ ,  $P = 0.04$ ). No comparisons by sex, race, or educational level indicated a  $P$  value less than 0.01.

## Oregon ( $N = 346$ )

Tobacco use showing consistency between the original data collection and the LTUQ administration included age at first cigarette (ICC = 0.33; 95% confidence interval (CI): 0.19, 0.47) and number of cigarettes/day during daily smoking (Spearman’s  $r_s = 0.31$ ; 95% CI: 0.11, 0.49); history of quit attempts ( $r_s = 0.38$ ; 95% CI: 0.14, 0.62); and abstinence symptoms ( $r_s = 0.46$ , 95% CI: 0.15, 0.77; Table 2).

Responses for less directly related measures were consistent regarding ever use of cigarettes and amount smoked (both, 94.1% consistency,  $P < 0.01$ ). Responses to basic tobacco use questions, including ever use of cigarettes and smokeless tobacco, indicated that the percentage of consistent responses differed significantly from the percentage that would be consistent if responding were random. (See Table 4.)

## Pittsburgh ( $N = 296$ )

Tobacco-use variables measured prospectively and with the LTUQ were age at first cigarette (ICC = 0.58; 95% CI: 0.48, 0.68), age at first weekly smoking (ICC = 0.49; 95% CI: 0.35, 0.63), cigarettes per week during weekly smoking ( $r_s = 0.40$ ; 95% CI: 0.25, 0.54), age at first daily smoking (ICC = 0.52; 95% CI: 0.38, 0.66), cigarettes per day ( $r_s = 0.32$ ; 95% CI: 0.14, 0.48), and time to first cigarette of the day during daily smoking ( $r_s = 0.37$ ; 95% CI: 0.20, 0.50). Similarly, convergent validity for most subjective responses to first use of tobacco was significantly associated, ranging from nausea ( $r_s = 0.51$ ; 95% CI: 0.34, 0.67) to rush/buzz ( $r_s = 0.31$ ; 95% CI: 0.13, 0.49). (See Table 3.)

## DISCUSSION

### Research goals

This project examined the convergent validity of tobacco use measures administered initially in two separate

repeated-measures cohort studies, then compared with similar measures administered to the same cohorts years later. Although the need for establishing the validity of basic health-related measures is obvious, the present analyses are among the few tobacco-related validity studies. The present studies examined convergent validity of prospective and retrospective self-report interview responses, indicating that recalled responses demonstrated reasonable convergent validity with prospective responses, particularly in the near-term estimates of convergent validity.

### Strengths and limitations

Strengths of the research included the importance of providing guidelines for interpreting research using these commonly used measures; the use of a prenatal tobacco exposure at Pittsburgh and an adolescent-emerging adult cohort at Oregon; and the lengthy time intervals between the initial research and the administration of the LTUQ. Limitations, which may have influenced the strength of the associations, related to differences in question wording from the initial testing to the subsequent LTUQ testing (Web Table 1; <http://aje.oupjournals.org>); differences in the mode of completion (initial in-person interview and telephone interview); lack of a definitive gold standard for establishing validity of the LTUQ; and possible bias related to nonresponse of initial participants, which could result in an overestimate of the level of agreement.

### Nature of self-report

On some key elements of tobacco use, responses from participants in both the Oregon and Pittsburgh studies demonstrated adequate convergent validity across time and across instruments. Tobacco use responses from the Pittsburgh study, with a considerably shorter duration between original testing and LTUQ administration, all had better validity estimates. Notably, the Oregon participants' prospective and retrospective reports of quit attempts were consistent, despite the metric defining quitting differing between the original data and the LTUQ.

It would be inappropriate to assume that proximity in time to an event invariably increases the accuracy of measurement. As one review (8) concluded, numerous factors can affect the validity of adolescents' self-reports of health-risk behaviors. This may be because smoking "tends to be habitual, repetitious, and almost unconscious" (15 p. 8) and because the episodic nature of adolescent smoking defies description of smoking patterns (16). Also, direct interaction with an interviewer may lead to underreporting, even under optimal reporting and interview conditions (8). Additionally, studies using measures collected in real time, also referred to as ecologic momentary assessment techniques, may be limited in sample size and duration because of event-based and time-based designs (17).

It is difficult to estimate whether prospective data were affected by changes in levels of social and legal stigma across time. A possible underreporting of tobacco use in childhood and adolescence could have contributed to bias and error. Additionally, reports of present use of tobacco in

the later testing could be affected by social desirability because the LTUQ was administered by interviewers (18).

Prior studies (9, 10) indicated that salience of questioned events appeared to affect reliability of test-retest recall. Although respondents in those two studies of separate Web-based self-administration cohorts were able to answer many questions consistently, events that were less well defined or salient resulted in lower reliability. For example, respondents reliably remembered details about their first use of tobacco but could not calculate reliably how many cigarettes they smoked between experimentation and monthly use.

### Quality of recall

Previous work has shown valid reporting of tobacco dependence across a period of 5 to 12 years (19). Self-report of tobacco use has shown consistency in review and meta-analysis (20), in smoking and smokeless tobacco use (21, 22), in lung cancer screening with self-report and urinary cotinine (23), and in saliva cotinine verification of audio computer-assisted self-report (24). Despite these positive indications, the limitations of retrospective collection of information are well known, and retrospective techniques often are affected by overall informant inaccuracy. In a frequently cited 1984 review, Bernard et al. summarized: "Informants are inaccurate; memory does decay exponentially with time. . . . And on top of this, there appears to be systematic distortion in how informants recall just about everything" (25, p. 509).

Scott and Alwin (26) approached the limitations of informant accuracy by examining types of retrospective information: a) recollections of past experiences that involve reconsidering the past and reporting present reactions, and b) reviewing or contemplating the past rather than simply recalling events. "In this sense," Scott and Alwin noted, "retrospections are not longitudinal at all; they are 'current' or rooted in the present" (26, p. 104). Two limitations of retrospective data are 1) measures of past experiences may be unavailable because of lapse in memory, or because the information can no longer be retrieved or accessed; and 2) recollections may be biased by more current experiences and events. Previous analyses of LTUQ retrospective data (9, 10, 27) indicated that while reliability was high for many questions about tobacco use, the overall salience of the recalled events was critical. This concurred with the summation of Scott and Alwin that "even retrospective attitudinal data can be quite reliable if the attitudes concerned are highly salient" (26, p. 121). Therefore the success of studying tobacco-related life-history events appears to depend on whether events are salient when they occur. Additionally, the salience of historical personal events can override current interpretation bias.

Implications for the study of other health-related behaviors are evident because behaviors and conditions that influence health are related. An example is the intertwining of tobacco use with other conditions in pregnancy, childbirth, and early childhood. Shenassa et al. (28) examined the validity of adult 40-year recall of maternal smoking during pregnancy. They reported that higher socioeconomic status and recall of specific situations resulted in more accurate

recall. A review of more than 60 studies of birth certificate information (29) found the certificate information to be neither reliable nor valid regarding tobacco and alcohol use, prenatal care, maternal risk, delivery complications, labor, and delivery. A review (30) of articles on maternal recall of breastfeeding practices showed validity for recall of initiation and duration of breastfeeding, but less satisfactory recall of practices with less distinct boundaries or impact. Similarly, a review (31) of adult retrospective reports of adverse childhood experiences indicated that reporting bias does not invalidate case-control retrospective studies of readily defined major adversity. Even so, details might not be reported accurately, particularly if judgment and interpretation are required. Stanton et al. (32), studying recanting of earlier reports of smoking status, found that minority status and reports of earlier smoking frequency could result in misclassification. These mixed findings underscore the need for psychometrically valid measurement of recall of health behaviors in these critical periods.

### Conclusion and recommendations

Responses are rarely better than the questions asked. As detailed in two published LTUQ reliability studies and the present analyses, the accuracy and psychometric quality of the responses depended on the clarity of the questions and salience of the information. At issue is whether an event was sufficiently notable that it could be remembered or calculated decades later.

Conversely, the validity (and reliability, in prior studies) of responses about the age at first cigarette use likely reflected the milestone nature of the event. Pittsburgh subjects' consistent recall years later of their subjective response to tobacco initiation reflected the salience of an event that occurred several years earlier.

The emergence of withdrawal symptoms in abstinence appeared to be sufficiently salient for high validity in the Oregon study. It may be reasonable to expect accuracy when dependent users recall the experience of trying to quit. It may be less reasonable to expect that smokers will be able to recall accurately how many cigarettes they used per week at various time points across their lives. Milestones that are important to researchers are not necessarily notable to tobacco users at the time the events occur.

An important future direction for this line of health-related psychometric work would be the systematic study of the effect of duration between an original event and its later recall, to elucidate question features that can be anticipated to yield valid, reliable responses across the lifespan.

### ACKNOWLEDGMENTS

Author affiliations: Center for Health Sciences, SRI International, Menlo Park, California (Janet Brigham, Harold S. Javitz, Ruth E. Krasnow, Mary McElroy, Gary E. Swan); Washington University School of Medicine, St. Louis, Missouri (Christina N. Lessov-Schlaggar); Oregon Research Institute, Eugene, Oregon (Elizabeth Tildesley, Judy

Andrews, Hyman Hops); and University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania (Marie D. Cornelius, Nancy L. Day).

This work was supported by the National Institutes of Health (NIH) (DA018019 to G. E. S., CA75581 (subcontract to G. E. S.), DA11795 to J. B., DA003706 to H. H., DA009275 to M. D. C., HD036890 to N. L. D).

Psychometric work on the LTUQ was funded by NIH grant DA018019 to G. E. S. Initial development work on the LTUQ was under subcontract to University of Michigan, NIH grant CA75581 to Ovide F. Pomerleau. Development of the computerized precursor of the LTUQ was funded by NIH grant DA11795 to J. B. Prospective SMOFAM data were collected under NIH grant DA003706 to H. H. Prospective Pittsburgh data were collected under NIH grant DA009275 to M. D. C. and HD036890 to N. L. D.

The authors express thanks to programmer Nancy Chong; Dale McBride, Sandra Sterry, Oregon Research Institute; and Young Jhon and Margaret Watson, University of Pittsburgh.

The findings were presented in part at the 15th annual meeting of the Society for Research on Nicotine and Tobacco, Baltimore, Maryland, February 28, 2010.

Conflict of interest: none declared.

### REFERENCES

1. Danaei G, Vander Hoorn S, Lopez AD, et al. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *Lancet*. 2005;366(9499):1784–1793.
2. Giovino GA, Biener L, Hartman AM, et al. Monitoring the tobacco use epidemic I. Overview: optimizing measurement to facilitate change. *Prev Med*. 2009;48(1 suppl):S4–S10.
3. Saccone SF, Hinrichs AL, Saccone NL, et al. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet*. 2007;16(1):36–49.
4. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*. 2008;452(7187):638–642.
5. Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet*. 2008;40(5):616–622.
6. Hung RJ, McKay JD, Gaborieau V, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature*. 2008;452(7187):633–637.
7. Delnevo CD, Bauer UE. Monitoring the tobacco use epidemic III: the host: data sources and methodological challenges. *Prev Med*. 2009;48(1 suppl):S16–S23.
8. Brener ND, Billy JO, Grady WR. Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: evidence from the scientific literature. *J Adolesc Health*. 2003;33(6):436–457.
9. Brigham J, Lessov-Schlaggar CN, Javitz HS, et al. Reliability of adult retrospective recall of lifetime tobacco use. *Nicotine Tob Res*. 2008;10(2):287–299.
10. Brigham J, Lessov-Schlaggar CN, Javitz HS, et al. Test-retest reliability of web-based retrospective self-report of tobacco exposure and risk. *J Med Internet Res*. 2009;11(3):e35. (doi:10.2196/jmir.1248).

11. Day N, Cornelius M, Goldschmidt L, et al. The effects of prenatal tobacco and marijuana use on offspring growth from birth through 3 years of age. *Neurotoxicol Teratol.* 1992;14(6):407–414.
12. Cornelius MD, Leech SL, Goldschmidt L, et al. Prenatal tobacco exposure: is it a risk factor for early tobacco experimentation? *Nicotine Tob Res.* 2000;10(1):45–52.
13. Cornelius MD, Leech SL, Goldschmidt L, et al. Is prenatal tobacco exposure a risk factor for early adolescent smoking? A follow-up study. *Neurotoxicol Teratol.* 2005;27(4):667–676.
14. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic.* 1981;86(2):127–137.
15. Means B, Habina K, Swan GE, et al. Cognitive research on response error in survey questions on smoking. National Center for Health Statistics. *Vital Health Stat 6.* 1992;6(5). ([http://www.cdc.gov/nchs/data/series/sr\\_06/sr06\\_005.pdf](http://www.cdc.gov/nchs/data/series/sr_06/sr06_005.pdf)). (Accessed July 15, 2010).
16. Pokorski TL, Chen WW, Bertholf RL. Use of urine cotinine to validate smoking self-reports in U.S. Navy recruits. *Addict Behav.* 1994;19(4):451–454.
17. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol.* 2008;4:1–32.
18. Kreuter F, Presser S, Tourangeau R. Social desirability bias in CATI, IVR, and web surveys. *Public Opin Q.* 2008;72(5):847–865.
19. Hudmon KS, Pomerleau CS, Brigham J, et al. Validity of retrospective assessments of nicotine dependence: a preliminary report. *Addict Behav.* 2005;30(3):613–617.
20. 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(7):1086–1093.
21. Post A, Gilljam H, Rosendahl I, et al. Validity of self reports in a cohort of Swedish adolescent smokers and smokeless tobacco (snus) users. *Tob Control.* 2005;14(2):114–117.
22. Post A, Gilljam H, Bremberg S, et al. Maternal smoking during pregnancy: a comparison between concurrent and retrospective self-reports. *Paediatr Perinat Epidemiol.* 2008;22(2):155–161.
23. Studts JL, Ghate SR, Gill JL, et al. Validity of self-reported smoking status among participants in a lung cancer screening trial. *Cancer Epidemiol Biomarkers Prev.* 2006;15(10):1825–1828.
24. Fendrich M, Mackesy-Amiti ME, Johnson TP, et al. Tobacco-reporting validity in an epidemiological drug-use survey. *Addict Behav.* 2005;30(1):175–181.
25. Bernard HR, Killworth P, Kronenfeld D, et al. The problem of informant accuracy: the validity of retrospective data. *Annu Rev Anthropol.* 1984;13:495–517.
26. Scott J, Alwin D. Retrospective versus prospective measurement of life histories in longitudinal research. In: Giele JZ, Elder GH, eds. *Methods in Life Course Research: Qualitative and Quantitative Approaches.* Thousand Oaks, CA: Sage; 1998:98–127.
27. Javitz HS, Brigham J, Lessov-Schlaggar CN, et al. Association of tobacco dependence and quit attempt duration with Rasch-modeled withdrawal sensitivity using retrospective measures. *Addiction.* 2009;104(6):1027–1035.
28. Shenassa E, Wen X, Paradis A, et al. Validity of adults' 40-year recall of exposure to maternal smoking during early childhood. Symposium paper accepted for presentation at the 2010 Annual Meeting of the Society for Research on Nicotine and Tobacco, Baltimore, MD, February 24–27, 2010.
29. Northam S, Knapp TR. The reliability and validity of birth certificates. *J Obstet Gynecol Neonatal Nurs.* 2006;35(1):3–12.
30. Li R, Scanlon KS, Serdula MK. The validity and reliability of maternal recall of breastfeeding practice. *Nutr Rev.* 2005;63(4):103–110.
31. Hardt J, Rutter M. Validity of adult retrospective reports of adverse childhood experiences: review of the evidence. *J Child Psychol Psychiatry.* 2004;45(2):260–273.
32. Stanton CA, Papandonatos G, Lloyd-Richardson EE, et al. Consistency of self-reported smoking over a 6-year interval from adolescence to young adulthood. *Addiction.* 2007;102(11):1831–1839.