The Validity of Self-Reported Smoking: A Review and Meta-Analysis

ABSTRACT

Objectives. The purpose of this study was to identify circumstances in which biochemical assessments of smoking produce systematically higher or lower estimates of smoking than self-reports. A secondary aim was to evaluate different statistical approaches to analyzing variation in validity estimates.

Methods. Literature searches and personal inquiries identified 26 published reports containing 51 comparisons between self-reported behavior and biochemical measures. The sensitivity and specificity of self-reports of smoking were calculated for each study as measures of accuracy.

Results. Sensitivity ranged from 6% to 100% (mean = 87.5%), and specificity ranged from 33% to 100% (mean = 89.2%). Interviewer-administered questionnaires, observational studies, reports by adults, and biochemical validation with cotinine–plasma were associated with higher estimates of sensitivity and specificity.

Conclusions. Self-reports of smoking are accurate in most studies. To improve accuracy, biochemical assessment, preferably with cotinine–plasma, should be considered in intervention studies and student populations. (*Am J Public Health.* 1994;84:1086–1093)

Donald L. Patrick, PhD, MSPH, Allen Cheadle, PhD, Diane C. Thompson, MS, Paula Diehr, PhD, Thomas Koepsell, MD, MPH, and Susan Kinne, PhD

Introduction

Smoking continues to be the largest single preventable cause of premature mortality and morbidity in the United States, yet 29% of American adults continue to smoke.[1] Efforts to promote cessation of smoking include interventions conducted with patients in clinical practices, in group environments such as schools and work sites, and in entire communities. Self-reports of smoking behavior are often assessed to determine the efficacy of these interventions. Observational studies and epidemiological studies of risk also incorporate measures of smoking behavior. Smoking is assessed in these studies to discriminate between smokers and nonsmokers, to measure change in smoking status, or to calculate pack-years of exposure retrospectively for risk assessment.

The validity of self-reported smoking is often questioned because of the widespread belief that smokers are inclined to underestimate the amount smoked[2,3] or to deny smoking at all.[4,5] As more attention is paid to smoking in the media and in public places, work sites, and clinical practice, individuals become sensitized to socially desirable forms of behavior. Thus, smokers may be more likely to exaggerate the extent to which their behavior conforms to the perceived social norm of "not smoking." Bias may be more common wherever social desirability is greater, such as in community-based studies in which intervention programs often seek explicitly to change community norms about the social acceptability of smoking.

Biochemical assessments of smoking by-products in body substances are often made to validate self-reports of smoking. Biochemical assessments can be viewed primarily as measures of the point prevalence of current smoking.[6] Because they are believed to be more objective and less susceptible to bias, biochemical measures are most often considered the "gold standard" in validation studies (i.e., they are considered more accurate than self-reports of smoking). Cotinine (in plasma, saliva, or urine), thiocyanate (in plasma or saliva), and carbon monoxide (in expired air) are the most commonly used biochemical assessments. Participants are either told in advance that such assessments will be made or asked for informed consent and specimens "on the spot." Sometimes the bogus pipeline procedure has been used, wherein subjects are informed that their self-reports can or will be objectively verified by the researchers by means of a biochemical test. In actuality, no verification takes place, although specimens are collected and left unanalyzed.[4]

Despite their believed objectivity, biochemical measures do not provide a gold standard, nor are they perfect measures of accuracy for use in assessing criterion validity. Carbon monoxide and thiocyanate can be elevated in those who do not use tobacco, and cotinine, although a specific metabolite of nicotine, can be elevated in users of snuff and chewing tobacco. When biochemical tests are

Donald L. Patrick, Allen Cheadle, Paula Diehr, and Thomas Koepsell are with the Department of Health Services, University of Washington, Seattle. Donald L. Patrick, Thomas Koepsell, Diane C. Thompson, and Susan Kinne are with the Fred Hutchinson Cancer Center, Seattle. Paula Diehr is also with the Department of Biostatistics, and Thomas Koepsell with the Department of Epidemiology, University of Washington.

Requests for reprints should be sent to Donald L. Patrick, PhD, MSPH, Department of Health Services, SC-37, University of Washington, Seattle, WA 98195.

This paper was accepted October 27, 1993.

repeated, the results may be different even when smoking status has not changed. Biochemical measures also have practical drawbacks. Although nonreactive, these measures are obtrusive: blood, saliva, or breath samples need to be collected from the individual. Collection of samples involves more contact with respondents than usual in conducting large-scale field studies[3] and may result in increased refusals.[6] Because of the short half-life of smoking by-products in the body, biochemical assessment validates only smoking status near the time of specimen collection.[7] Costs can be considerable, ranging from less than $1 per sample for carbon monoxide to $20 per sample for cotinine analysis.[8] The cost of collecting, handling, and arranging for frozen storage of the specimens can add significantly to these estimates.

In contrast, self-reported smoking is assessed easily by using self-administered questionnaires in person or by mail or by using interviewer-administered questionnaires in person or on the phone. Questionnaires are noninvasive and inexpensive, and assurances of confidentiality of information can reduce refusals to participate. Self-reported information can be used to measure behavioral change, to calculate exposure risk, or to study pathways to smoking cessation or continuation.

The meta-analyses reported in this paper combine findings from a number of studies that validated self-reported smoking with biochemical measures, making it possible to examine the importance of different aspects of the studies, the populations, and the validation process. This paper addresses four major questions. First, what evidence exists to document the validity of self-reported measures of smoking behavior? Second, under what circumstances is it most important for investigators to consider biochemical assessment in studies of smoking behavior? Third, how do results change when using different statistical approaches for analyzing variation in the measures of accuracy and for pooling information across studies? Finally, how does this literature review inform the conduct of future validation studies, reports of smoking behavior, and the publication of results?

Methods

Description of Meta-Analysis Procedures

Meta-analyses are becoming common in both clinical and social science research.[9–11] Meta-analytic techniques were applied in this study to observations of the association between biochemical measures and self-reported smoking, similar to meta-analyses of diagnostic tests. Similar applications, such as the accuracy of the exercise electrocardiogram[12] and the human immunodeficiency virus antibody test,[13] have appeared in the literature. This application, like meta-analyses of randomized clinical trials, is an observational study of previously published studies.

Standard procedures were followed in accumulating and evaluating research studies for the meta-analysis.[9–11] We defined the problem as accuracy of self-reported smoking, with biochemical assessment as the criterion or concordance measure for evaluating validity. Bibliographic searches were conducted on all articles published between 1982 and 1991. Initially the MEDLINE database was used, with "smoking" as the subject head and the keywords "intervention studies," "evaluation," "community-based programs," and "education" as subheads. The Science Citation Index was used to trace articles referenced in studies previously identified in the bibliographic search. The Current Contents database was scanned for more recent articles through mid-1991, and references in these articles were also evaluated. Investigators familiar with smoking research at the Fred Hutchinson Cancer Research Center were also asked to identify appropriate studies from their literature files.

Thirty studies containing comparisons between self-reported smoking and biochemical assessments were identified. Studies confined to pregnant women were excluded. All studies were reviewed for information on the following characteristics: method of administration (self-administered vs interviewer administered), biochemical measures (cotinine, thiocyanate, or carbon monoxide), type of sample (air, blood, saliva), cutoff value used for biochemical definition of smoking, population (student vs general population), study design (intervention vs observational), sample size, ability to classify participants according to a 2 × 2 table based on self-report of smoking (yes/no) and the gold-standard definition of exceeding the defined cutoff level on the biochemical measure, and the smoking rate (i.e., prevalence of smoking), defined by the gold standard and self-report measures. The 2 × 2 table for calculating the accuracy of reports is shown in Figure 1.

		Respondent Exceeds Cut-Off on Biochemical Measure?	
		Yes	No
Self-Reported Smoking?	Yes	a	b
	No	c	d

FIGURE 1—The 2 × 2 table.

All the required data were available in 26 of the 30 articles identified as validity studies of self-reported smoking.[4,5,14–37] Four studies did not contain sufficient information to calculate accuracy measures, and these studies were eliminated from further analysis. Three members of the study team (Donald L. Patrick, Diane C. Thompson, and Susan Kinne) abstracted data independently to ensure quality control of the data used in the analyses. Discrepancies among the three abstractors were investigated and resolved after discussion. The 26 studies contained 32 comparisons based on independent samples and 51 comparisons wherein 2 or more comparisons were made on partial or total analyses of the same individuals. Table 1 contains the essential data abstracted from the studies included in the meta-analysis.

Measures of Accuracy

Data abstracted from the 26 studies were used to calculate two measures of self-report accuracy. For the purposes of this study, in which biochemical measures were considered the criterion measure, *sensitivity* was defined as $a/(a + c)$ in Figure 1, or the proportion of respondents with a positive level on the biochemical measure that reported smoking. *Specificity* was defined as $d/(b + d)$, or the proportion of respondents with a negative level on the biochemical measure that reported absence of smoking.

Analytic Models and Procedures

Cutoff levels were standardized into comparable units of measurement across studies for each type of biochemical measure and biological specimen. Studies using thiocyanate–saliva samples were eliminated from our analyses because of their outlying values for sensitivity and specificity after such standardization. We thus analyzed 47 of the 51 comparisons (allowing more than 1 comparison per

TABLE 1—Data Used in Meta-Analysis of 26 Published Studies on Biochemical Validation of Self-Reported Smoking

Author(s) (Publication Year)	Study No.	Comparison No.	Method of Administration	Biochemical Measure	Sample	Cutoff Value	Population	Study Design	Sample Size	a + c[a]	Sensitivity	b + d[b]	Specificity	Gold Standard Smoking Rate	Self-Report Smoking Rate
Bauman & Dent (1982)[14]	1	1	SAQ	CO	Air	9	S	O	389	36	58	352	92	9.3	12.6
Bauman & Dent (1982)	1	2	SAQ	CO	Air	9	S	O	1189	100	90	1089	89	8.4	17.7
Bauman & Dent (1982)	1	3	SAQ	CO	Air	9	G	O	370	112	93	258	90	30.3	35.1
Bauman & Dent (1982)	1	4	SAQ	CO	Air	9	G	O	1115	350	95	765	88	31.4	38.1
Bauman et al. (1982)[15]	2	1	SAQ	CO	Air	9	S	O	82	3	100	79	97	3.7	6.1
Bauman & Koch (1983)[16]	3	1	SAQ	CO	Air	9	G	O	1439	460	95	979	92	32.0	36.3
Bauman & Koch (1983)	3	2	SAQ	CO	Air	9	G	O	386	36	64	350	95	9.3	10.9
Cohen & Bartsch (1980)[17]	4	1	SAQ	COHb	Blood	2	G	O	617	385	91	232	67	62.4	69.0
Cohen & Bartsch (1980)	4	2	SAQ	SCN	Blood	100	G	O	617	431	92	186	83	69.9	69.0
Fortman et al. (1984)[18]	5	1	IAQ	CO	Air	9	G	O	1279	416	91	863	99	32.5	30.4
Fortman et al. (1984)	5	2	IAQ	SCN	Blood	100	G	O	1673	550	87	1123	96	32.9	31.4
Gillies et al. (1982)[19]	6	1	SAQ	SCN	Saliva	100	S	O	252	48	6	204	97	18.7	3.6
Haddow et al. (1986)[20]	7	1	SAQ	COT	Blood	10	G	O	296	64	92	232	98	21.3	21.3
Luepker et al. (1981)[21]	8	1	SAQ	SCN	Saliva	100	S	O	1205	74	37	1258	98	5.0	3.5
Luepker et al. (1989)[5]	9	1	SAQ	COT	Saliva	20	S	O	263	88	92	175	97	33.5	33.1
Luepker et al. (1989)	9	2	SAQ	COT	Saliva	20	S	O	290	108	95	182	93	37.2	39.7
Luepker et al. (1989)	9	3	SAQ	COT	Saliva	20	S	O	325	105	77	220	97	32.3	26.8
Luepker et al. (1989)	9	4	SAQ	COT	Saliva	20	S	O	331	111	93	220	95	33.5	34.7
McNeill et al. (1987)[22]	10	1	SAQ	COT	Saliva	15	S	O	508	125	96	383	86	24.6	34.1
Murray et al. (1987)[24]	11	1	SAQ	CO	Air	9	S	O	245	46	46	199	89	18.8	17.6
Murray et al. (1987)	11	2	SAQ	CO	Air	9	S	O	248	15	100	233	85	6.0	20.6
Murray et al. (1987)	11	3	SAQ	CO	Air	9	S	O	246	19	89	227	88	7.7	18.3
Noland et al. (1988)[23]	12	1	SAQ	COT	Saliva	25	S	O	308	121	99	187	79	39.0	51.6
Pechacek et al. (1984)[24]	13	1	SAQ	SCN	Saliva	100	S	O	1183	184	72	999	89	15.6	20.8
Pechacek et al. (1984)	13	2	SAQ	CO	Air	9	S	O	1221	166	95	1055	91	13.6	20.7
Petitti et al. (1981)[25]	14	1	SAQ	SCN	Blood	100	G	O	267	55	100	212	85	20.6	32.2
Petitti et al. (1981)	14	2	SAQ	CO	Air	9	G	O	267	58	98	209	86	21.7	32.2
Pierce et al. (1987)[26]	15	1	IAQ	COT	Saliva	40.5	G	O	975	336	93	639	93	34.5	36.2
Pojer et al. (1984)[27]	16	1	SAQ	COHb	Blood	2	G	O	368	166	98	202	88	45.1	50.8
Pojer et al. (1984)	16	2	SAQ	COT	Blood	40.5	G	O	368	181	98	187	95	49.2	50.8
Pojer et al. (1984)	16	3	SAQ	SCN	Blood	70	G	O	369	144	98	225	80	39.0	50.7
Prignot (1987)[28]	17	1	IAQ	SCN	Blood	80	G	O	197	31	90	166	99	15.7	15.2
Saloojee et al. (1982)[29]	18	1	IAQ	COHb	Blood	1.6	G	O	439	349	99	90	86	79.5	82.0
Saloojee et al. (1982)	18	2	IAQ	SCN	Blood	73	G	O	439	339	99	100	75	77.2	82.0
Sillett et al. (1978)[30]	19	1	SAQ	COHb	Blood	1.7	G	I	79	36	69	43	93	45.6	35.4
Sillett et al. (1978)	19	2	SAQ	COHb	Blood	1.7	G	I	140	89	63	51	96	63.6	41.4
Slattery et al. (1989)[31]	20	1	IAQ	COT	Blood	15	G	O	542	157	95	385	96	29.0	30.1
Stookey et al. (1987)[32]	21	1	IAQ	COT	Saliva	10	G	I	338	260	82	78	97	76.9	63.9
Stookey et al. (1987)	21	2	IAQ	CO	Air	9	G	I	338	234	88	104	91	69.2	63.9
Stookey et al. (1987)	21	3	IAQ	COT	Saliva	10	G	I	236	214	100	22	91	100.0	99.0
Stookey et al. (1987)	21	4	IAQ	SCN	Saliva	80	G	I	236	193	97	43	33	96.8	86.5
Stookey et al. (1987)	21	5	IAQ	CO	Air	9	G	I	236	207	100	29	69	100.0	95.8
Van Vunakis et al. (1989)[33]	22	1	IAQ	COT	Blood	8	G	O	318	137	92	181	100	43.1	39.6
Van Vunakis et al. (1989)	22	2	IAQ	COT	Saliva	25	G	O	303	122	94	181	99	40.3	38.6
Vogt et al. (1977)[34]	23	1	IAQ	CO	Air	9	G	I	123	79	97	44	98	64.2	63.4
Vogt et al. (1977)	23	2	IAQ	CO	Air	9	G	I	139	85	98	54	80	61.2	67.6
Vogt et al. (1977)	23	3	IAQ	SCN	Blood	100	G	I	123	74	96	49	86	60.2	63.4
Vogt et al. (1977)	23	4	IAQ	SCN	Blood	100	G	I	139	79	96	60	70	56.8	67.6
Wagenknecht et al. (1990)[35]	24	1	IAQ	COT	Blood	13	G	O	4932	1542	88	3390	98	31.3	28.9
Wald et al. (1981)[36]	25	1	IAQ	COHb	Blood	2	G	O	8724	1666	99	7055	94	19.1	23.9
Williams et al. (1979)[37]	26	1	SAQ	COT	Blood	3	S	O	118	21	90	97	99	17.8	16.9

Note. SAQ = self-administered questionnaire; IAQ = interviewer-administered questionnaire. CO = carbon monoxide measured in parts per million; COHb = carboxyhemoglobin, measured in percent; SCN = thiocyanate measured in micromoles per liter; COT = cotinine measured in nanograms per milliliter. S = student; G = general population; O = observational; I = intervention.

[a]Respondent exceeded cutoff on biochemical measure (see Figure 1).

[b]Respondent did not exceed cutoff on biochemical measure (see Figure 1).

study) and 30 of the 32 strictly independent comparisons (allowing only 1 comparison per study, chosen at random).

Since there is uncertainty about how best to combine proportions across studies,[40] several analytic methods were used to obtain point estimates and confidence intervals of average sensitivity and specificity. The fixed-effects approach assumes that all studies are homogeneous and that the "true" sensitivity and specificity are the same for each study. Under this model, estimates from each study differ from this assumed true value only because of sampling variation, and the estimates from small studies tend to be worse than those from larger studies. In the fixed-effects model, the most efficient point estimate for sensitivity, for example, is the mean of the sensitivities weighted by sample size. This is equivalent to putting all of the people from all studies into one 2×2 table and estimating sensitivity and specificity from it. Confidence intervals are calculated from the weighted standard error. Regression analyses to determine correlates of accuracy can be achieved by using weighted ordinary least squares, with the study as the unit of analysis, or by using logistic regression. We used logistic regression to conduct fixed-effects analyses of the correlates of accuracy.

In actuality, however, studies are not homogeneous: they involve different study populations and different data collection methods. The "true" sensitivity and specificity values being estimated probably do vary among studies because of this heterogeneity. Under a random-effects model, variation among the studies is assumed to be due, in part, to the variation among the "true" values, as well as to sampling variation within each study. The correct weight in this case depends on both sources of variation. If the between-studies heterogeneity is large, this results in approximately equal weights for each study; if it is small, then the weights are approximately proportional to the sample size, as in the fixed-effects model. Laird and Mosteller[38] suggested a moment-based estimate for the mean, with its associated variance, that can be used to obtain confidence intervals for the random-effects method. This method computes appropriate weights for each study; these weights are a combination of the variance within studies and the variances among studies. If there is almost no variation among studies, then each study is weighted by its sample size. If there is great variance among studies, then each study will receive approximately equal weight. We computed the weights and found that the latter case was true for our data. The sensitivity and specificity estimates for our studies were so variable that the random-effects analysis essentially gave all studies equal weight. For this reason, to examine the effects of covariates under a random-effects assumption, we used simple ordinary least squares regression with the study as the (unweighted) unit of analysis and sensitivity and specificity as the dependent variables.

The sensitivity and specificity measures were negatively skewed. To test the sensitivity of the findings to this deviation from normality, we analyzed the logarithm of 100 minus sensitivity plus 1 and 100 minus specificity plus 1, which did have a reasonably normal distribution. The results of analyses using the logarithmically transformed measures were substantially the same as analyses with untransformed data; hence, for ease of interpretation, we report the original analyses.

We compared fixed-effects and random-effects models. For the purposes of this paper, we present unadjusted, bivariate results and multivariate analyses using the ordinary least squares random-effects regression (most conservative) and the fixed-effects logistic regression (least conservative). Independent variables in the regression analyses were method of administration (self-administered vs interviewer administered), study design (observational vs intervention), population type (student vs general population), and type of biochemical measure and specimen (cotinine–plasma as the reference vs cotinine–saliva, thiocyanate–plasma, or carbon monoxide [air and blood]). Two-way interactions among the study characteristics were examined in the random-effects ordinary least squares analyses of 47 comparisons; one interaction was entered at a time.

Results

A total of 36 830 respondents were included in the 26 studies and 51 comparisons. Of the comparisons reported in these studies, 37.7% were obtained from self-administered questionnaires and 62.3% were obtained by interviewers. Students represented 22.8% of respondents for the comparisons; 77.2% of respondents were drawn from the general population. Only 5.8% were respondents in intervention studies in which biochemical assessments were used. Carbon monoxide, thiocyanate, and cotinine were used as the biochemical measure in 54.8%, 18.2%, and 27% of the comparisons, respectively.

As shown in Table 1, sensitivity values ranged from 6 to 100 in the 51 comparisons; specificities ranged from 33 to 100. The average sensitivity for these studies, unweighted by sample size, was 87.5; the average specificity was 89.2. Tables 2 and 3 show results for sensitivity and specificity for all 47 comparisons and for the 30 strictly independent comparisons, respectively. For each sample, results are shown for the ordinary least squares random-effects and logistic fixed-effects models. The bivariate analyses show the effect on sensitivity or specificity of each study characteristic by itself. The multivariate analysis results come from regression models that included as predictors all of the study characteristics listed. Thus, they show the effect of each predictor on sensitivity or specificity after adjusting statistically for all of the other predictors.

For the fixed-effects analysis, the logistic regression coefficients were used to calculate the effect, in percentage points, of the study characteristic of interest when other characteristics (if any) in the model were held constant at their mean values. Effects for the ordinary least squares random-effects regression were the coefficients themselves. For example, per the random-effects ordinary least squares regression treating comparisons as independent (Table 2), interviewer-administered survey self-reports had a sensitivity that was 5.2 percentage points higher than that of self-administered survey reports, but this difference was not significantly different from zero ($P = .135$).

For sensitivity, the fixed-effects result for interviewer-administered studies was a 1.1 percentage point increase, which was significant at the .05 level. After controlling for all other study characteristics, the effect was a 4.0 percentage point increase in self-report sensitivity for the random-effects model (not significant) and a 1.2 percentage point increase in the fixed-effects analysis (marginally significant at $P = .066$). None of the study characteristics showed significantly different estimates for sensitivity in either the bivariate or multivariate random-effects analyses. For the fixed-effects approach, student populations and the type of sample were significantly related to sensitivity. Student populations yielded significantly lower sensitivity than general population studies. All biochemical samples yielded higher

TABLE 2—Estimated Effects of Study Characteristics on Sensitivity and Specificity of Self-Reported Smoking: All Comparisons (n = 47)

	Sensitivity				Specificity			
	Ordinary Least Squares Random Effects		Logistic: Fixed Effects		Ordinary Least Squares Random Effects		Logistic: Fixed Effects	
Study Characteristic	Effect[a]	*P*	Effect[a]	*P*	Effect[a]	*P*	Effect[a]	*P*
Bivariate analyses								
Interviewer administered	5.2	.135	1.1	.047	2.5	.313	6.3	<.001
Student population	−4.6	.227	−3.7	<.001	1.8	.510	−3.4	<.001
Observational study	−0.3	.937	−1.0	.135	4.2	.124	4.8	<.001
Cotinine–saliva[b]	−0.4	.944	2.3	.017	−4.9	.212	−5.5	<.001
Thiocyanate–plasma[b]	2.1	.750	3.0	.001	−13.5	.001	−7.6	<.001
CO[b]	−4.5	.441	2.6	.003	−8.5	.023	−6.8	<.001
COHb[b]	−6.2	.382	6.3	<.001	−10.6	.019	−5.0	<.001
Multivariate analyses								
Interviewer administered	4.0	.402	1.2	.066	3.8	.204	8.1	<.001
Student population	−5.2	.301	−3.3	.001	0.1	.983	−0.3	.551
Observational study	3.7	.480	0.8	.310	3.5	.286	5.2	<.001
Cotinine–saliva[b]	1.4	.845	2.9	<.001	−5.4	.222	−8.6	<.001
Thiocyanate–plasma[b]	2.5	.716	2.6	<.001	−12.3	.007	−10.7	<.001
CO[b]	−1.5	.809	2.9	<.001	−6.9	.085	−3.8	<.001
COHb[b]	−4.5	.556	5.6	<.001	−8.2	.097	−9.2	<.001

Note. Four studies using thiocyanate–saliva samples were omitted from analysis. CO = carbon monoxide; COHb = carboxyhemoglobin, measured in percent.
[a]Estimated percentage point difference in sensitivity or specificity due to the study characteristic.
[b]Cotinine–plasma is the reference.

TABLE 3—Estimated Effects of Study Characteristics on Sensitivity and Specificity of Self-Reported Smoking: Independent Comparisons Only (n = 30)

	Sensitivity				Specificity			
	Ordinary Least Squares Random Effects		Logistic: Fixed Effects		Ordinary Least Squares Random Effects		Logistic: Fixed Effects	
Study Characteristic	Effect[a]	*P*	Effect[a]	*P*	Effect[a]	*P*	Effect[a]	*P*
Bivariate analyses								
Interviewer administered	4.9	.326	2.4	.001	4.2	.056	6.2	<.001
Student population	−4.7	.351	−4.6	<.001	−1.5	.508	−4.7	<.001
Observational study	5.3	.389	2.1	.033	0.9	.748	1.3	.239
Cotinine–saliva[b]	0.1	.993	3.3	.010	−8.1	.026	−7.1	<.001
Thiocyanate–plasma[b]	2.5	.804	3.6	.017	−9.0	.031	−9.2	<.001
CO[b]	−4.5	.546	1.6	.128	−8.4	.007	−7.0	<.001
COHb[b]	−5.9	.511	8.0	<.001	−7.0	.051	−4.3	<.001
Multivariate analyses								
Interviewer administered	4.0	.554	2.9	.002	4.6	.088	6.7	<.001
Student population	−7.1	.364	−3.7	.003	2.3	.454	0.4	.469
Observational study	9.3	.272	8.7	<.001	1.3	.688	0.8	.406
Cotinine–saliva[b]	4.4	.667	3.9	<.001	−9.5	.022	−10.7	<.001
Thiocyanate–plasma[b]	5.3	.639	3.9	<.001	−6.9	.117	−8.9	<.001
CO[b]	1.2	.888	3.9	<.001	−7.4	.030	−5.1	<.001
COHb[b]	−0.9	.931	−8.5	<.001	−4.9	.245	−7.5	<.001

Note. Two studies using thiocyanate–saliva samples were omitted from analysis. CO = carbon monoxide; COHb = carboxyhemoglobin, measured in percent.
[a]Estimated percentage point difference in sensitivity or specificity due to the study characteristic.
[b]Cotinine–plasma is the reference.

sensitivity than cotinine–plasma, and interviewer-administered studies provided marginally greater sensitivity ($P = .047$).

For specificity, the random-effects model showed statistically significant lower estimates for thiocyanate–plasma, saliva cotinine, and carbon monoxide (both air and plasma) in comparison with plasma cotinine. When all study characteristics

were entered into the regression equation, significant beneficial effects for specificity remained only for cotinine–plasma in comparison with plasma thiocyanate samples.

Estimates of self-report accuracy using the fixed-effects analyses were much more likely to show statistically significant differences due to study and population characteristics than were the results of random-effects analyses. The ordinary least squares random-effects results in Table 2 (columns 2 and 4) illustrate the most conservative approach; the logistic fixed-effects results (columns 3 and 5) are less conservative under the assumption that studies are homogeneous. The ordinary least squares random-effects analyses, both bivariate and multivariate, showed no significant effects for sensitivity and significantly lower estimates of specificity for plasma thiocyanate samples using statistical significance as the criterion.

Table 3 shows the results of even more conservative analyses using only 30 independent comparisons, with 1 comparison chosen at random from each study eligible for analyses. Results were similar for sensitivity using the random-effects or the fixed-effects approach in both bivariate and multivariate analyses. No study characteristics produced significantly higher or lower estimates of self-report accuracy. Fixed-effects analyses of sensitivity were also similar for bivariate and multivariate analyses, indicating that all study characteristics produced the same higher or lower estimates. Results of studies with student populations yielded lower estimates of sensitivity. For specificity and the random effects analyses, significantly lower estimates were obtained for the different samples in comparison with plasma cotinine. All study characteristics were significantly different in the bivariate, fixed-effects analysis of specificity, except for observational studies (in comparison with intervention studies). This same pattern emerged with multivariate analyses of specificity using the fixed-effects approach, although the differences obtained from student populations no longer were evident.

An analysis of two-way interaction effects using the random-effects model and 47 comparisons yielded only one significant interaction each for sensitivity and specificity (interviewer-administered questionnaires in observational studies). Our power to detect interaction effects, however, was small given the small number of studies included in the analysis.

Discussion

This meta-analysis of published studies comparing self-reported smoking status with results of biochemical validation suggests generally high levels of sensitivity and specificity for self-report. Across all studies, the sensitivity of self-report was 87%, and the specificity was 89%. Nonetheless, both measures of accuracy proved quite variable among studies, as shown in Table 1, suggesting that specific aspects of the setting, study population, measurement methods, and study purpose are important to the accuracy of smoking self-reports.

Our search for systematic patterns of variation in sensitivity and specificity across studies was only partially successful. Two different methods of meta-analysis yielded generally similar results on the sign and magnitude of the effect of each study characteristic on sensitivity or specificity, but they often produced widely divergent verdicts on the statistical significance of those effects. The fact that the random-effects model produced *P* values that were usually much higher than those from the fixed-effects model indicates a large amount of between-study variability that could not be accounted for by the study characteristics measured. This conclusion, in turn, suggests that other unmeasured study characteristics may confound our results about the effects of method of administration, study population, study type, and biochemical test and specimen.

One of the most important unmeasured study characteristics is the specific wording of questions on smoking status. Very few studies reported this critical information, despite considerable evidence from survey research that responses are heavily influenced by how a question is phrased and the order in which questions are asked.

Because of these limitations, the observed patterns of association between accuracy of self-report and study characteristics must be interpreted with caution. The results suggest that interviewer-administered questionnaires yielded higher estimates of sensitivity and specificity than did self-administered questionnaires. Interviews identified more of the smokers correctly and classified nonsmokers more accurately. This may reflect smokers' awareness of sensory cues about their smoking (visible cigarettes, nicotine stains on teeth or hands, smoke odor) that would be obvious to an interviewer. More respondents may attempt to hide smoking behavior in self-administered questionnaires, even when biochemical validation is known.

Even in the most conservative analyses, student self-reports had lower sensitivity than studies using reports from subjects in the general population; however, the results were not always statistically significant. That is, students appear more likely to deny smoking, even when biochemical measures classify them as smokers. This is not surprising, since smoking by minors is illegal in most states and many young tobacco users have not yet defined themselves as smokers. Both of these conditions would contribute to a tendency, whether conscious (fear of being found out) or unconscious (self-definition inconsistent with behavior), to underreport. The different analyses do not suggest, however, that student self-reports have higher specificity than do those of subjects from the general population. Unfortunately, all studies of student populations reported here were observational in nature; no intervention studies with students that reported biochemical results were found among published studies, although the bogus pipeline procedures have been used with students. Reports of accuracy from intervention studies with student populations might have yielded lower estimates of sensitivity given the results from observational studies.

In the most conservative analyses, using only one comparison from each study and the random-effects approach, observational studies had higher levels of sensitivity than intervention studies, a conclusion supported by a qualitative review of studies in the 1990 surgeon general's report.[3] Self-reports from subjects in intervention studies, in which there is an expectation of cessation of smoking, are more likely to involve underreporting of actual smoking.

Self-reports of participants in intervention programs also have lower specificity, meaning that more biochemically validated nonsmokers reported smoking. Biochemical tests have limited ability to detect the very low levels of smoking that would be expected from recent quitters who "slip" and smoke an occasional cigarette.[34] Discussions of quitters' reactions to such slips (the abstinence violation effect) suggest that the would-be quitter is likely to exaggerate the importance of a few cigarettes under these conditions.[39] This can produce a report of "smoking," although its magnitude is smaller than the test can detect. In observational studies, with little focus on

cessation, this reaction may not be triggered.

Consistent with previous reports,[40] self-reports of smoking validated by means of cotinine–plasma biochemical measures appear to have higher specificity than those reports validated by other biochemical tests and specimens. The *P* values for tests comparing cotinine–plasma and other biochemical measures were not always significant in the more conservative random effects model, but the signs and magnitudes of the coefficients were mostly consistent across analytic models and between Tables 2 and 3. The observed pattern shows that more false-positive self-reports (i.e., respondents report that they are smokers when the biochemical test does not confirm such a report) tend to be observed in studies that use methods other than cotinine–plasma. This finding may reflect variation in the accuracy of the other biochemical tests in relation to self-reported smoking. These so-called false-positives may actually be smokers, but the poorer biochemical tests were too insensitive to detect relatively low levels of biochemical abnormality. There is no substantive reason to expect differences in self-report behavior due to the form of biochemical validation.

The fixed-effects and random-effects analyses yielded both different point estimates of the effects and different significance levels. Between-study variation was highly significant in all analyses, calling into question a key premise underlying the fixed-effects model. The results obtained from using multiple nonindependent comparisons in the different studies (Table 2) and from using only independent measures (Table 3) were substantially similar. Thus, the lack of independence in samples did not prove to be a serious problem for this meta-analysis.

In summary, our results suggest that biochemical validation may be more important in intervention studies, in studies with student populations, and in studies using self-administered rather than interviewer-administered questionnaires. For greatest accuracy, self-administered questionnaires given to students might benefit from biochemical validation, given the lower estimates obtained from these groups. Cotinine–plasma may be the biochemical test of choice if adequate resources are available for collection and analysis.

The decision to use biochemical validation is not as straightforward as it might appear. Biochemical validation is costly and sometimes difficult to obtain for all participants.[6] The bogus pipeline procedure and use of biochemical assessments with random subsamples of the target population are alternative strategies.[4]

The conclusions from this meta-analysis are subject to three major cautionary notes that, in turn, indicate needed improvements in the conduct and reporting of future studies of smoking behavior. First, it is known that the form and content of self-report questions about smoking influence the responses given and, hence, the categorization of respondents as smokers.[25] This observation argues that studies asking about smoking should report or reference the exact questions used so that this source of misclassification can be controlled. Second, authors of published studies need to make clear how biochemical validation was presented to study participants. In studies in which participants know that biochemical assessment will occurr, such as in the bogus pipeline procedure, self-reported smoking rates may be different from those in studies in which biochemical assessment is presented only at the point of collection. The exact procedures used in studies should be identified in the methods sections of articles to permit evaluation of the results according to the potential bias introduced by precollection announcement of biochemical validation.

Finally, in any meta-analysis, publication bias and the "file drawer phenomenon," the failure to submit for publication studies that do not produce effects, have an impact on the data available for analysis and can bias the outcome.[41,42] Although this bias is likely to be reduced in smoking studies that do not seek to test a specific hypothesis about validity, the universe of potential data sets on the validity of self-reports is still influenced by investigator analysis and submission of data for publication. This meta-analysis also excluded studies of pregnant women, and, thus, the generalizability of the results must be considered in comparison with the population under investigation.

The failure to include sufficient data on biochemical assessment resulted in the exclusion of several studies that reported validations of self-report. This suggests that in future studies with smoking validation, sufficient data should be published for investigators to confirm and evaluate the accuracy of self-reports. □

Acknowledgments

This work was supported by grant CA 34847 from the National Cancer Institute awarded to the Fred Hutchinson Cancer Center Public Health Sciences Division, Cancer Prevention Research Unit, for work on the methodology of community-based studies.

An earlier version of this paper was presented at the International Conference on Measurement Errors in Survey, Tucson, Ariz, November 1990.

References

1. *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.
2. Haley NJ, Hoffmann D. Analysis for nicotine and cotinine in hair to determine cigarette smoker status. *Clin Chem.* 1985;31: 1598–1600.
3. *The Health Benefits of Smoking Cessation.* Washington, DC: US Dept of Health and Human Services; 1990. DHHS publication CDC 90-8416.
4. Murray DM, O'Connell CM, Schmidt LA, Perry CL. The validity of smoking self-reports by adolescents: a reexamination of the bogus pipeline procedure. *Addict Behav.* 1987;12:7–15.
5. Luepker RV, Pallonen UE, Murray DM, Pirie PL. Validity of telephone surveys in assessing cigarette smoking in young adults. *Am J Public Health.* 1989;79:202–204.
6. Velicer WF, Prochaska JO, Rossi JS, et al. Assessing outcome in smoking cessation studies. *Psychol Bull.* 1992;111:23–47.
7. Benowitz NL. The use of biologic fluid samples in assessing tobacco smoke consumption. In: Grubowski J, Bell CS, eds. *Measurement in the Analysis and Treatment of Smoking Behavior.* Washington, DC: US Dept of Health and Human Services; 1983;6–26. NIDA research monograph 48.
8. Windsor T, Morris J, Cutter G, et al. Sensitivity, specificity and predictive value of saliva thiocyanate among pregnant women. *Addict Behav.* 1989;14:447–452.
9. Light RJ, Pillemer DB. *Summing Up: The Science of Reviewing Research.* Cambridge, Mass: Harvard University Press; 1984.
10. Rosenthal R. *Meta-Analytic Procedures for Social Research.* Newbury Park, Calif: Sage; 1984. Applied Social Research Methods Series Volume 6.
11. Hunter JE, Schmidt FL. *Methods of Meta-Analysis.* Newbury Park, Calif: Sage; 1990.
12. Gianrossi R, Detrano R, Mulvihill D, et al. Exercise-induced ST depression in the diagnosis of coronary artery disease. A meta-analysis. *Circulation.* 1989;80:87–98.
13. Phillips KA. The use of meta-analysis in technology assessment: a meta-analysis of the enzyme immunosorbent assay human immunodeficiency virus antibody test. *J Clin Epidemiol.* 1991;44:925–931.
14. Bauman KE, Dent CW. Influence of an objective measure on self-reports of behavior. *J Appl Psychol.* 1982;67:623–628.
15. Bauman KE, Koch GG, Bryan E. Validity of self-reports of adolescent cigarette smoking. *Int J Addict.* 1982;17:1131–1136.
16. Bauman KE, Koch GG. Validity of self-

reports and descriptive and analytical conclusions: the case of cigarette smoking by adolescents and their mothers. *Am J Epidemiol.* 1983;118:90–98.
17. Cohen JD, Bartsch GE. A comparison between carboxyhemoglobin and serum thiocyanate determinations as indicators of cigarette smoking. *Am J Public Health.* 1980;70:284–286.
18. Fortmann SP, Rogers T, Vranizan K, et al. Indirect measures of cigarette use: expired-air carbon monoxide versus plasma thiocyanate. *Prev Med.* 1984;13:127–135.
19. Gillies PA, Wilcox B, Coates C, et al. Use of objective measurement in the validation of self-reported smoking in children aged 10 and 11 years: saliva thiocyanate. *J Epidemiol Community Health.* 1982;36:205–208.
20. Haddow JE, Palomakii GE, Knight GJ. Use of serum cotinine to assess the accuracy of self-reported non-smoking. *BMJ.* 1986;293:1306. Letter.
21. Luepker RV, Pechacek TF, Murray DM, et al. Saliva thiocyanate: a chemical indicator of cigarette smoking in adolescents. *Am J Public Health.* 1981;71:1320–1324.
22. McNeill AD, Jarvis MJ, West R, et al. Saliva cotinine as an indicator of cigarette smoking in adolescents. *Br J Addict.* 1987;82:1355–1360.
23. Noland MP, Kryscio RJ, Riggs RS, et al. Saliva cotinine and thiocyanate: chemical indicators of smokeless tobacco and cigarette use in adolescents. *J Behav Med.* 1988;11:423–433.
24. Pechacek TF, Murray DM, Luepker RV, et al. Measurement of adolescent smoking behavior: rationale and methods. *J Behav Med.* 1984;7:123–130.
25. Petitti DB, Friedman GD, Kahn W. Accuracy of information on smoking habits provided on self-administered research questionnaires. *Am J Public Health.* 1981;71:308–311.
26. 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.
27. Pojer R, Whitefield JB, Poulos V, et al. Carboxyhemoglobin, cotinine, and thiocyanate assay compared for distinguishing smokers from non-smokers. *Clin Chem.* 1984;30:1377–1380.
28. Prignot J. Quantification and chemical markers of tobacco-exposure. *Eur J Respir Dis.* 1987;70:1–7.
29. Saloojee Y, Vesey CJ, Cole PV. Carboxyhemoglobin and plasma thiocyanate: complementary indicators of smoking behavior? *Thorax.* 1982;37:521–525.
30. Sillett RW, Wilson MB, Malcolm RE, et al. Deception among smokers. *BMJ.* 1978;2:1185–1186.
31. Slattery ML, Hunt SC, French TK, et al. Validity of cigarette smoking habits in three epidemiologic studies in Utah. *Prev Med.* 1989;18:11–19.
32. Stookey GK, Katz BP, Olson BI, et al. Evaluation of biochemical validation measures in determination of smoking status. *J Dent Res.* 1987;66:1597–1601.
33. Van Vunakis H, Tashkin DP, Rigas B, et al. Relative sensitivity and specificity of salivary and serum cotinine in identifying tobacco-smoking status of self-reported nonsmokers and smokers of tobacco and/or marijuana. *Arch Environ Health.* 1989;44:53–58.
34. Vogt TM, Selvin S, Widdowson G, et al. Expired air carbon monoxide and serum thiocyanate as objective measures of cigarette exposure. *Am J Public Health.* 1977;67:545–549.
35. Wagenknecht LE, Cutter GR, Haley NJ, et al. Racial differences in serum cotinine levels among smokers in the coronary artery risk development in (young) adults study. *Am J Public Health.* 1990;80:1053–1056.
36. Wald NJ, Idle M, Boreham J, et al. Carbon monoxide in breath in relation to smoking and carboxyhemoglobin levels. *Thorax.* 1981;36:366–369.
37. Williams CL, Eng A, Botwin GJ, et al. Validation of students' self-reported cigarette smoking levels with plasma cotinine levels. *Am J Public Health.* 1979;69:1272–1274.
38. Laird NM, Mosteller F. Some statistical methods for combining experimental results. *Int J Technol Assess Health Care.* 1990;6:5–30.
39. Curry SG, Marlatt GA, Gordon JR. Abstinence violation effect: validation of an attributional construct with smoking cessation. *J Consult Clin Psychol.* 1987;55:145–149.
40. Murray RP, Connett JE, Lauger GG, Voelker HT. Error in smoking measures: effects of intervention on relations of cotinine and CO to self-reported smoking. *Am J Public Health.* 1993;83:1251–1257.
41. Jenicek M. Meta-analysis in medicine: where we are and where we want to go. *J Clin Epidemiol.* 1989;42:35–44.
42. Fleiss JL, Gross AJ. Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental tobacco smoke and lung cancer: a critique. *J Clin Epidemiol.* 1991;44:127–139.