Conditions Under Which a Bogus Pipeline Procedure Enhances the Validity of Self-Reported Cigarette Smoking: A Meta-Analytic Review[1]

HERMAN AGUINIS,[2] CHARLES A. PIERCE, AND BRIAN M. QUIGLEY
University at Albany, State University of New York

A meta-analysis was conducted to test whether the use of self-report measures within the bogus pipeline (BPL) paradigm yields more valid responses than the use of self-report measures alone for assessing cigarette smoking behavior. The meta-analytic results indicate that, overall, a BPL condition resulted in a larger proportion of subjects reporting that they are frequent smokers, as compared to a self-report measure only (no pipeline) condition. Tests of categorical models indicate that the enhanced validity of self-reports within the BPL paradigm is moderated by the following variables: (a) type of BPL presentation employed, (b) type of self-report measure to which the BPL technique is compared, and (c) whether most participants are smokers (as indicated by a biochemical marker).

The consumption of tobacco continues to be a serious public health problem. Smoking is strongly associated with emphysema, heart disease, and a number of cancers (Fielding, 1985a; U.S. Department of Health and Human Services, 1989). It has been reported that the consumption of tobacco causes approximately half a million deaths annually in the United States (U.S. Department of Health and Human Services, 1989), and the total cost in health care expenses and loss of productivity resulting from smoking is about $60 billion each year (Manley, Epps, Husten, Glynn, & Shopland, 1991; Silvis & Perry, 1987). From the perspective of organizations, an employee who smokes costs an employer between $274 and $287 in excess insurance costs (Kristein, 1982). Also, smokers are absent from work 33% to 45% more than nonsmokers, which amounts to approximately two to three days' excess absenteeism per year (Kristein, 1982). More generally, Mahler (1990) estimates that the loss of productivity due to smoking and its effects amount to be between 30 and 43 billion dollars per year.

Because of the importance of the issue, a number of long-term research

[1]We thank George M. Alliger, Donn Byrne, Eugene F. Stone, and James T. Tedeschi for their insightful comments on a previous draft of this article. We acknowledge James T. Tedeschi for his help in locating articles, and are also grateful to two anonymous reviewers for their helpful suggestions on a previous version.

[2]The order of authorship was determined alphabetically. Correspondence concerning this article should be addressed to Herman Aguinis, Charles A. Pierce, or Brian M. Quigley, Department of Psychology, University at Albany, State University of New York, Albany, NY 12222.

Journal of Applied Social Psychology, 1993, **23**, 5, pp. 352–373.

programs have shown concern with the factors affecting the acquisition, cessation, and prevention of smoking behavior (DiClemente & Prochaska, 1982, 1985; Elder et al., 1990; Evans et al., 1981; Fielding, 1985b; Prochaska & DiClemente, 1983; Prochaska, DiClemente, Velicer, Ginpil, & Norcross, 1985; Stern, Prochaska, Velicer, & Elder, 1987; Taylor, 1986).

Recent reviews on the measures utilized in smoking cessation studies (Velicer, Prochaska, Rossi, & Snow, 1992), and on the effectiveness of such interventions (Viswesvaran & Schmidt, 1992), indicate that the relative validities of the measures that assess cigarette smoking behavior is a questioned and unresolved issue.

One of the dilemmas faced by investigators is whether to utilize self-report measures of smoking behavior. Self-report measures are frequently criticized because they are subject to socially desirable responses. Results from extensive research programs on the social approval motive (Crowne & Marlowe, 1964; Edwards, 1957; Kenny, 1956; Rosen, 1956), demand characteristics (Orne, 1962, 1969), evaluation apprehension (Rosenberg, 1965, 1969; Sigall, Aronson, & Van Hoose, 1970), and impression management (Leary & Kowalski, 1990; Tedeschi, 1981) indicate that subjects will tend to underreport behaviors that are socially defined as undesirable, or that might make them "look bad" in the eyes of the experimenter. This underreporting bias can occur when subjects are asked to respond about sensitive topics such as cigarette smoking on self-report questionnaires.

Even though false physiological feedback had been previously used as an independent experimental variable (Behar, 1967; Bramel, 1962; Gerard & Rabbie, 1961; Valins, 1966), Jones and Sigall (1971) first suggested the bogus pipeline (BPL) paradigm to minimize socially desirable responses in social psychological research. The paradigm is based on the premise that subjects do not want their questionnaire responses to be second-guessed by a machine. The BPL paradigm is not a single method, but rather a set of procedures that lead subjects to believe that their responses to a paper and pencil questionnaire can and will be independently verified by a biochemical or physiologically based measurement instrument, thus motivating subjects to provide more valid responses on the paper and pencil measure. The physiologically based measure is portrayed as a powerful, sophisticated, and practically infallible lie detector. If respondents are convinced that the BPL is effective, they will be faced with a dilemma when they have to respond in a way that might be considered socially undesirable. "If they tell the truth about the way they feel, others may be offended and evaluate them negatively. If they lie by giving a socially desired response, the subjects presumably believe that the BPL will detect the lie; therefore, they will be revealed as having a socially undesirable attitude, *and* as being liars" (Tedeschi, Lindskold, & Rosenfeld, 1985, p. 164). In actuality, however, some researchers evaluate the physiological measure, and others do not. The assumption underlying the method is that, given that the respondents are led to believe that the physiologically based measure can assess their true behaviors and attitudes, they will respond more honestly on the paper and pencil scales.

In a seminal study published in 1977, Evans, Hansen, and Mittlemark utilized the BPL paradigm to measure smoking behavior. They found that adolescents reported higher frequency of smoking in a BPL than in the control (paper and pencil only) condition. Evans et al. (1977) interpreted this result as an indication that the BPL methodology yields more accurate and valid self-reports of smoking behavior than a self-report measure alone.

Evans et al.'s and others' (e.g., Bauman & Dent, 1982; Berman, McCombs, & Boruch, 1977; Cherry, Byrne, & Mitchell, 1976; Gaes, Quigley-Fernandez, & Tedeschi, 1978; Hill, Dill, & Davenport, 1988; Jones & Sigall, 1971) inference concerning the greater validity of self-report measures as part of the BPL paradigm is congruent with Cook and Campbell's (1979) definition of construct validity. Construct validity "involves the fit between operations and referent constructs" (Cook & Campbell, 1979, p. 63). Construct validity will be threatened to the extent that the operationalizations (i.e., measurement procedures) are influenced by confounding factors (e.g., social desirability). Social desirability would result in underreporting an undesirable behavior such as cigarette smoking. If the BPL procedure indeed enhances the validity of self-reported smoking behavior, then the number of individuals reporting that they smoke will be greater in a BPL than a no-BPL condition. In terms of proportions, the number of self-reported smokers in the BPL group relative to the total number of subjects in this condition will be larger than the number of self-reported smokers in the no-BPL group relative to the total number of subjects in the no-BPL condition.

The results reported by Evans et al. (1977) were received as a major methodological advance, and a number of researchers conducted replications in which variations of the original BPL procedure were implemented. However, even though the BPL seemed promising, the comparison of self-reported smoking behavior as measured with the BPL versus a self-report measure alone yielded mixed results. Some researchers reported that utilizing the BPL paradigm increases the number of subjects who report that they smoke (e.g., Bauman & Dent, 1982; Evans et al., 1977; Luepker et al., 1981). On the other hand, several studies (e.g., Akers, Massey, Clarke, & Lauer, 1983; Bauman, Koch, & Bryan, 1982; Williams & Gillies, 1984) yielded a similar proportion of subjects who reported being smokers as measured with a BPL procedure in comparison to the proportion of subjects reporting that they smoke on self-report measures alone.

Given the need to account for the inconsistency in these results, Murray, O'Connell, Schmid, and Perry (1987) conducted a narrative review of the literature. They suggested that possible factors influencing the evaluation of the effectiveness of the BPL paradigm were whether there was an equal number of smokers in the groups, the statistical power of the analyses performed, the credibility of the pipeline message, how the samples were selected, and the level of social pressure to underreport smoking behavior. Even though this qualitative review was conducted, a quantitative review is needed in order to assess the merit of the BPL procedure in measuring tobacco consumption as compared to the use of self-report measures alone.

Relatively novel statistical techniques permit the utilization of meta-analytic reviews, which integrate and summarize existing evidence in a systematic and quantitative fashion (Hedges & Olkin, 1985; Rosenthal, 1984). A quantitative review differs from a qualitative review in that the latter might be more vulnerable to error because of its relative informality, and its incomplete and unsystematic sampling of the available studies (Eagly, Makhijani, & Klonsky, 1992). Because a quantitative approach permits the computation of effect size estimates across a large number of studies, the statistical and inferential power are increased (Johnson, 1989). Also, meta-analysis permits formal testing of hypothesized moderator variables. Narrative reviews have been criticized because they rely too heavily on the findings of individual studies and do not consider the magnitude of the relationship between variables (Glass, McGaw & Smith, 1981; Hedges, 1986; Mullen, 1989). Wolf (1986) provides a summary of the major drawbacks of traditional narrative reviews that can be surmounted by meta-analysis; these limitations include selective inclusion of studies, differential subjective weighting of studies while interpreting a set of findings, misleading interpretations, and the failure to consider study characteristics as potential moderating variables in the relationship examined.

Overall Prediction for the Evaluation of the BPL Paradigm

A seemingly clear prediction for the meta-analytic review is that the use of a BPL procedure should yield a larger proportion of subjects reporting that they smoke, as compared to a no-BPL (self-report measure only) condition. As noted above, however, results might be inconsistent because there are a number of factors influencing the effectiveness of the BPL procedure (e.g., the operationalization of the BPL paradigm).

Predictions About Moderating Variables

The inconsistencies in the results reported in previous studies and the review by Murray et al. (1987) suggest that a number of variables might affect the relative validity of self-reports in a BPL condition as compared to using self-report measures alone. In this section, potential moderating variables are examined, and predictions are made regarding their impact.

Credibility of the BPL procedure. The degree to which subjects believe that the BPL is an infallible method for detecting smoking behavior will increase the likelihood that they will provide more honest answers. The credibility of the pipeline message may vary across studies because (a) different types of BPL presentations have been employed, (b) different biochemical measures have been used as part of the BPL paradigm, and (c) the role identity of the BPL administrator has differed across studies.

The manner in which the BPL is presented (e.g., live, verbal, video) to the subjects will have an impact on the credibility of the procedures. For example, a live or video demonstration of how the pipeline can accurately detect smoking behavior would be more convincing than merely verbalizing a passage of how the method operates (Murray et al., 1987). Thus, it is predicted that the more vivid and detailed the description and explanation about the accuracy of the pipeline, the more valid the self-reports will be in a BPL condition as compared to a no-BPL condition.

The type of biochemical measure utilized as part of the BPL methodology might have an impact on perceptions about the method's accuracy in detecting true frequency of smoking. Typically, researchers either collect saliva specimens or air samples from subjects.[3] Subjects are told that the former is utilized to measure the level of saliva thiocyanate (SCN) and the latter is used to measure the level of expired air carbon monoxide (CO). The collection of one of these samples may be seen as more "scientific" than the collection of the other, thereby increasing the credibility of the BPL, and the disclosure of smoking behavior.

The BPL procedure may be more credible when described by a senior researcher rather than a research assistant. There is evidence that suggests a positive relationship between the credibility of a message and source characteristics such as age (Goodman, Golding, & Haith, 1984), expertise (Olson & Cal, 1984), and experience (Petty & Cacioppo, 1981; Wu & Shaffer, 1987). Thus, it is hypothesized that the use of the BPL will enhance the validity of self-reported smoking when the principal investigators, as opposed to assistants with lower status, are responsible for administering the BPL manipulation and collecting the data.

Pressure to underreport smoking behavior. The BPL procedure will be most useful when smoking is perceived negatively, and thus the pressure to

[3]Plasma cotinine (the major metabolite of nicotine) is also a biochemical measure utilized by some researchers in the field. However, the use of this biochemical measure will not be considered because no studies were located that compared a BPL versus a no-BPL condition using cotinine as part of the procedures.

underreport it will be greater. The pressure to underreport may be affected by (a) the type of self-report measure used, (b) the degree to which subjects believe that their responses will remain confidential, and (c) the age of the respondents.

Comparing the BPL procedure to different types of self-report measures may yield different results. Some self-report measures (e.g., randomized response technique, Himmelfarb & Lickteig, 1982; Warner, 1965) assure more confidentiality than others (e.g., traditional self-report questionnaire). When the randomized response technique is utilized in a no-BPL condition, subjects randomly choose to respond to either a sensitive (e.g., frequency of smoking) or a nonsensitive (e.g., month of birth) item. Based on a random probability distribution, and unbeknownst to the respondents, the researcher can compute the proportion of *yes* responses to the sensitive item (Martin & Newman, 1988). Because assured confidentiality might have an impact on the effectiveness of the BPL paradigm, it is hypothesized that the use of a BPL will enhance the validity of self-reported smoking when compared to *regular* self-reports alone, but not when compared to a randomized response technique.

If the subjects' responses are made available to individuals for evaluation (e.g., parents, teachers, supervisors), then the pressure to underreport will be heightened. Alternatively, if subjects are convinced that their responses to the self-report questionnaire are confidential, they might not feel as pressured to underreport smoking behavior. Thus, it is predicted that when subjects are not assured that their responses will be kept confidential, a BPL condition will result in a larger proportion of subjects reporting their smoking behavior as compared to self-report measures alone.

The age range of subjects might have an impact on the pressure to underreport smoking behavior. More specifically, it is hypothesized that younger respondents (i.e., junior high school students) will feel more pressure to underreport smoking behavior. Adults, on the other hand, might not feel such a pressure because they would not expect negative consequences for reporting their smoking behavior. Thus, it is predicted that for younger populations, the pressure to underreport smoking behavior will be heightened, and thus the validity of self-reports in a BPL as compared to a no-BPL condition will be greater.

Use of a biochemical marker to detect smokers. The statistical power of the tests performed to detect a possible difference between the proportion of subjects reporting that they smoke in the BPL versus the no-BPL condition will be increased when only smokers are considered. Conversely, if the groups include both smokers and nonsmokers and the number of smokers is only a very small proportion of the total sample size, the likelihood of committing a Type II error will be increased (i.e., we might falsely conclude that the BPL technique does not enhance the validity of self-reports, when in actuality it does). Because the power of the statistical test will be increased when only smokers are examined, it is hypothesized that the reports from a BPL condition will be more valid than those from a no-BPL condition when the biochemical measure is not only used as part of the BPL paradigm, but also used to exclude from the analysis those individuals who are not smokers.[4]

Method

Sample of Studies

Initial computer-based searches were conducted using the keywords *bogus pipeline* paired together. These keywords were searched on the following computer databases: *Psychological Abstracts* (PsycLIT, 1974 to 1991), *Educational Resources Information Center* (ERIC, 1966 to 1991), *Sociological Abstracts* (Sociofile, 1974 to 1991), *Dissertation Abstracts International* (DISS, 1985 to 1991), and MEDLINE (1985 to 1991). These searches included information available through September 1991. Subsequent to the initial computer-based searches, reference lists of the relevant located articles were also examined for potential study candidates.

The criteria used for including studies in the final sample were that (a) a bogus pipeline technique was used within the realm of cigarette smoking behavior, (b) a measure of smoking behavior was collected (i.e., subjects responded whether they smoked daily or weekly), (c) at least two conditions were used, one using a bogus pipeline procedure and the other a control condition (i.e., not using any form of a bogus pipeline), and (d) primary statistics were reported. It should be noted that the dependent measure of interest is the proportion of subjects who report that they consistently smoke at least once a week. Using the aforementioned selection criteria, the final sample consisted of 15 documents. Most of the studies allowed for the computation of multiple effect size estimates and, therefore, 30 effect sizes were calculated.

One of the assumptions of meta-analysis is independence of effect sizes. In the current meta-analysis, it was not uncommon to compute more than one effect size from a single study. Multiple effect sizes were calculated when two

[4]The validity of biochemical measures is also a controversial issue (Abrams, Emmons, Niaura, Goldstein, & Sherman, in press; Velicer et al., 1992). However, even though there might be a number of false positives (nonsmokers classified as smokers) when the biochemical measure is used to classify subjects, the proportion of true smokers in the BPL and no-BPL conditions will be larger than in the original samples, and the power of the statistical test will be increased.

or more experimental groups (i.e., use of a BPL procedure) were compared to a no-BPL control group. For instance, some of the studies manipulated more than one kind of BPL procedure using two independent experimental conditions. Comparisons were then made to a no-BPL control group. Assuming each of the subjects' responses did not influence any of the other subjects' responses, the observations made in each of the conditions (three in this case, two experimental and one control) should be independent of one another. Therefore, two independent effect sizes could be computed here, one for each of the experimental groups being compared to the same control group.

Variables Coded From Each Study

The following information was coded from each of the studies in the final sample: (a) cartoon, video, verbal, live, or live and video presentation of the BPL; (b) type of biochemical measure used, CO or SCN; (c) who actually manipulated the BPL condition, either the researcher, an assistant, a student, or nonspecified; (d) whether the no-BPL condition consisted of a randomized response technique or a regular self-report questionnaire; (e) assured confidentiality of responses, either yes, or nonspecified; (f) age range of subjects used, either junior high, high school, or adults; and (g) whether the biochemical measure (e.g., level of CO in the air bag) was actually used to exclude nonsmokers from the analyses. These variables were independently coded by the three authors. Overall agreement among the raters exceeded 95%. The few discrepancies were resolved by consensus.

Computation and Analysis of Effect Sizes

All of the meta-analytic statistics were computed following the Hedges and Olkin (1985) framework to meta-analysis. The effect size calculated is *g*, which is the standardized difference between the mean proportion of self-reported frequent smokers in the BPL and no-BPL groups, divided by the pooled standard deviation. The sign of the difference was positive when the percent of individuals in the BPL condition reporting frequent cigarette smoking was greater than the percent of individuals in the no-BPL condition reporting frequent smoking. Conversely, if the percent of individuals in the no-BPL condition was greater, then the sign of *g* was negative. The computations of all 30 effect sizes were based on either frequencies or proportions.

The *g*s were converted to *d*s since *g*s tend to overestimate the population effect size when small sample sizes are used (Hedges & Olkin, 1985). All of the study-level *d*s were then combined by calculating both an unweighted and weighted (by the reciprocal of the effect size variance) mean effect size.

The weighted mean effect size places more emphasis on effect sizes estimated from larger sample sizes. A homogeneity statistic Q, which has an approximate chi-square distribution with $k - 1$ degrees of freedom (k is the number of effect sizes), was calculated in order to assess whether the study-level ds share a common population effect size (i.e., were consistent across the studies).

Characteristics from each of the studies were coded and categorically tested. The categorical tests allow for the examination of the relation between study characteristics and the magnitude of the effect sizes. According to Hedges and Olkin (1985), the testing of categorical models is similar to conducting analyses of variance (ANOVAs). Categorical model test procedures include a homogeneity test for the between-class effect (in ANOVA, this is analogous to a main effect) and a homogeneity test for the effect sizes within a given subgroup or class. The homogeneity statistic for the between-class effect is Q_B, which is approximately distributed as a chi-square with $p - 1$ degrees of freedom (p is the number of classes). The within-class homogeneity test statistic used is Q_{Wi}, which is approximately distributed as a chi-square with $m - 1$ degrees of freedom (m is the number of effect sizes in the class).

The purpose of testing categorical models is to detect the presence of potential moderator variables. The categorical variables that were coded are more easily interpreted as moderators in the following situation: (a) if the effect sizes *within* a particular class of a variable result in a nonsignificant Q_{Wi} homogeneity statistic, suggesting within-class homogeneity; and (b) if the effect sizes *between* classes of a given variable result in a significant Q_B homogeneity statistic, indicating heterogeneity across the levels of a subgrouping variable.

All of the effect size calculations and categorical model testing were performed utilizing Johnson's (1989) DSTAT meta-analytic software.

Results

Overall Evaluation of the BPL Paradigm

The overall prediction that a larger proportion of subjects would accurately report that they smoke if a BPL procedure is used as compared to a self-report measure only was supported. The overall unweighted mean effect size was 0.13. The weighted mean effect size was 0.05, and its 95% confidence interval did not include zero (95% CI = 0.01–0.09). However, both the unweighted and weighted mean effect sizes are small as defined by Cohen (1988). Homogeneity of ds was rejected, $\chi^2(29, N = 6{,}437) = 79.41, p < .001$, suggesting that the effect of the BPL was not uniform across studies. Table 1 presents the study name, year of publication, sample sizes, and effect size estimate for each of the studies used in the meta-analysis.

Moderating Variables

Table 2 displays the results of the categorical models tested, including all of the potential moderator variables examined, along with their classes, between-class homogeneity statistic, within-class weighted mean effect sizes (along with 95% confidence intervals), and within-class homogeneity statistics.

Credibility of the BPL procedure. As predicted, the between-class homogeneity statistic for type of BPL presentation used was significant, $\chi^2(4) = 39.87, p < .001$, suggesting that the type of BPL presentation has an impact on the disclosure of smoking behavior. Follow-up pairwise contrasts between all of the within-class weighted mean *d*s for type of BPL presentation were conducted. Significant post hoc comparisons (based on z^2, which is distributed as chi-square with 1 degree of freedom, Johnson, 1989) were found for video versus verbal, video versus live, cartoon versus live, cartoon versus live and video, verbal versus live, verbal versus live and video, and live versus live and video. All of these comparisons were significant at the .05 level, and the pattern of *d*s suggests that a more vivid presentation enhances the disclosure of cigarette smoking.

The between-class effect of type of biochemical measure used as part of the BPL technique was nonsignificant, $\chi^2(2) = 0.93, p = .63$, suggesting that the two BPL operationalizations (SCN and CO) elicit similar responses on self-reports of smoking behavior.

A nonsignificant between-class effect was observed for role identity of the BPL administrator ($\chi^2(1) = 2.14, p = .14$). However, because there was only one effect size in the "assistant" category (and 29 in the "nonspecified" category), no conclusions should be drawn.

Pressure to underreport smoking behavior. As hypothesized, a significant between-class homogeneity statistic was found for type of self-report measure used in the no-BPL condition ($\chi^2(1) = 6.64, p < .01$). In the only study that compared the BPL to a randomized response technique, no differences were found in the proportions of self-reported smokers. Conversely, the 95% confidence interval for the mean effect size ($d = 0.07$) for those studies that used a regular self-report measure in the no-BPL condition did not include zero (95% CI = 0.03–0.11). This suggests that the BPL technique enhances the validity of self-reported cigarette smoking when compared to a traditional self-report measure, rather than a randomized response questionnaire.

There was not a significant between-class effect for confidentiality ($\chi^2(1) = 0.35, p = .55$). However, none of the studies specifically reported that they

Table 1

Study Name, Sample Size, and Effect Size Estimates for Studies Used in the Meta-Analysis

Study	*N* BPL	*N* No-BPL	*d*
Lauer, Akers, Massey, & Clarke (1982)	80	76	−0.20
Werch, Gorman, Marty, Forbess, & Brown (1987)	62	63	−0.15
Lauer, Akers, Massey, & Clarke (1982)	166	178	−0.14
Werch, Gorman, Marty, Forbess, & Brown (1987)	60	63	−0.14
Werch, Lundstrum, & Moore (1989)	50	52	−0.13
Bauman, Koch, & Bryan (1982)	43	39	−0.13
Botvin, Botvin, Renick, Filazzola, & Allegrante (1984)	153	159	−0.10
Martin & Newman (1988)	515	630	−0.10
Williams & Gillies (1984)	148	147	−0.09
Akers, Massey, Clarke, & Lauer (1983)	74	76	−0.04
Lauer, Akers, Massey, & Clarke (1982)	75	76	−0.04
Akers, Massey, Clarke, & Lauer (1983)	221	178	−0.02
Lauer, Akers, Massey, & Clarke (1982)	289	178	0.01
Martin & Newman (1988)	603	552	0.02
Botvin, Botvin, Renick, Filazzola, & Allegrante (1984)	152	159	0.06
Lauer, Akers, Massey, & Clarke (1982)	207	178	0.06
Hansen, Malotte, & Fielding (1985)	211	299	0.08
Botvin, Botvin, Renick, Filazzola, & Allegrante (1984)	174	159	0.08
Lauer, Akers, Massey, & Clarke (1982)	68	76	0.09
Bauman & Dent (1982)	350	112	0.09
Hansen, Malotte, & Fielding (1985)	440	585	0.11
Gillies, Wilcox, Coates, Kristmundsdottir, & Reid (1982)	141	138	0.11
Werch, Lundstrum, & Moore (1989)	50	52	0.14
Hill, Henderson, Bray, & Evans (1981)	476	213	0.23
Luepker, Pechacek, Murray, Johnson, Hund, & Jacobs (1981)	187	47	0.29
Evans, Hansen, & Mittlemark (1977)	54	55	0.38

Table 1. *Continued*

Study	N BPL	N No-BPL	d
Evans, Hansen, & Mittlemark (1977)	32	29	0.57
Bauman & Dent (1982)	100	36	0.89
Murray, O'Connell, Schmid, & Perry (1987)	21	101	0.92
Murray, O'Connell, Schmid, & Perry (1987)	15	101	1.16
Unweighted mean effect size			0.13
Weighted mean effect size			0.05

Note. The effect size (*d*) is the standardized difference between the mean proportion of self-reported smokers in the BPL group and the no-BPL group, divided by the pooled standard deviation.

did not assure confidentiality. Thus, the categories included in the analysis were "confidentiality assured" and "not specified." Consequentially, it cannot be assumed that the studies which were nonspecified with respect to assuring confidential responses actually did *not* assure subjects that their reports would remain confidential.

A nonsignificant between-class effect was observed for the age range of subjects ($\chi^2(2) = 0.76, p = .68$). As expected, however, for the class of studies employing junior high school students, there was a significant effect of using a BPL procedure ($d = 0.07$, 95% CI = 0.01–0.13). The confidence interval for studies that used adults included zero (95% CI = –0.11–0.22). The between-class comparison, although statistically nonsignificant, is in the predicted direction and therefore suggests that the BPL may be most effective when used with younger adolescents (i.e., junior high school age) for the purpose of detecting cigarette consumption.

Use of a biochemical marker to detect smokers. The between-class homogeneity statistic was significant, $\chi^2(1) = 19.03, p < .001$, suggesting that the use of a biochemical index to detect smokers prior to group assignment (i.e., BPL, no-BPL) significantly increases the validity of the BPL paradigm.

Discussion

Research continues to demonstrate the health risks associated with smoking (e.g., U.S. Department of Health and Human Services, 1989). The increased risks of heart disease and cancer that accompany the use of tobacco make smoking an important issue to individuals and organizations.

Table 2

Categorical Models on Study-Level Effect Sizes

Variable and class	Q_B[a]	n[b]	d_{i+}[c]	95% CI for d_{i+} Lower	Upper	Q_{Wi}[d]
Type of BPL used	39.87***					
Live		1	1.16	0.59	1.72	0.00
Live and video		1	0.92	0.44	1.41	0.00
Video		5	0.22	0.11	0.33	4.66
Verbal		19	0.01	−0.03	0.06	32.74*
Cartoon		4	−0.03	−0.18	0.12	2.14
Type of biochemical measure used	0.93					
SCN		19	0.06	0.01	0.11	24.65
CO		7	0.05	−0.03	0.12	51.50***
None		4	−0.01	−0.12	0.12	2.32
Experimenter(s)	2.14					
Assistant		1	0.29	−0.04	0.61	0.00
Not specified		29	0.05	0.00	0.09	77.26***
Self-report	6.64**					
Regular		29	0.07	0.03	0.11	72.76***
Random response		1	−0.10	−0.21	0.02	0.00
Confidentiality assured	0.35					
Yes		19	0.06	0.01	0.10	65.73***
Not specified		11	0.03	−0.06	0.11	13.32
Age range of subjects	0.76					
Junior high		18	0.07	0.01	0.13	35.92**
Adults		3	0.06	−0.11	0.22	1.16
Senior high		9	0.03	−0.03	0.09	41.56***
Biochemical marker used to detect smokers	19.03***					
Yes		6	0.31	0.19	0.44	30.07***
No		24	0.02	−0.03	0.06	30.30

Note. [a]Between-class homogeneity statistic. [b]Number of effect sizes in each class. [c]Weighted mean effect size within a class. [d]Within-class homogeneity statistic. CI represents a confidence interval for d_{i+}. Asterisks indicate rejection of the hypothesis of homogeneity (between-class for Q_B and within-class for Q_{Wi}), where $^*p < .05$, $^{**}p < .01$, and $^{***}p < .001$.

Identifying smokers allows for timely and more effective implementation of intervention programs. Because investigators interested in evaluating the effectiveness of intervention programs need valid measures of smoking behavior (Velicer et al., 1992), the present meta-analysis was conducted to assess whether the bogus pipeline is a paradigm that allows for the identification of smokers who may not be detected with a self-report questionnaire alone.

Self-report measures administered as part of a BPL procedure produced more valid disclosures about smoking behavior than self-report measures used alone. More interestingly, the present meta-analytic findings help to explain some of the inconsistencies in the literature concerning the use of the BPL to produce more valid responses of smoking behavior. The identified moderating variables suggest the conditions under which a BPL manipulation is most effective.

The perceived credibility of the BPL has an impact on the effectiveness of the procedure. The manipulation of the bogus pipeline scenario must be believable to the subjects in order for the procedure to be useful (Jones & Sigall, 1971). In laboratory studies, detailed procedures are used to convince the subject that the apparatus has the ability to detect a lie (e.g., Quigley-Fernandez & Tedeschi, 1978; Sigall & Page, 1971). These elaborate presentations were not employed in many of the studies examined by the current quantitative review.

The BPL was presented to the subjects in several different modes, including verbal, video, cartoon, and live demonstrations. The results indicate that a more vivid presentation of the BPL enhances the validity of self-reported smoking. Live demonstrations and video presentations were the most effective. Even though only one study used a live presentation, this method (or in conjunction with a videotape) is the most effective way to maximize the validity of self-reports. Making use of a videotape alone to show the accuracy of the BPL also appears to be effective.

The pressure to underreport socially undesirable behaviors has an impact on the effectiveness of the BPL paradigm (Jones & Sigall, 1971). The social undesirability of smoking may not have been a concern to the respondents in all of the BPL evaluations; thus the use of the BPL would not yield an increase in the proportion of subjects admitting that they are smokers as compared to a self-report measure alone.

The pressure to underreport socially undesirable behaviors such as cigarette smoking may be diminished when subjects believe that their responses cannot be attributed individually. By using a randomized response procedure in the no-BPL condition, confidentiality of the subjects' responses is assured, and thus socially desirable responses are diminished. The BPL paradigm is more effective when compared to a regular self-report alone than when compared to a randomized response questionnaire.

Adults may not perceive their smoking as socially undesirable and may not attempt to conceal their behavior, whereas adolescents may fear the repercussions that may occur if they admit to smoking. It is possible that smoking may only have been a socially undesirable response for younger populations. Even though the between-group homogeneity statistic for age range was nonsignificant, the results were in the hypothesized direction. The junior high school students' reports appear to be more affected by the pipeline than the reports solicited from high school students and adults.

A methodological factor was found to moderate the effectiveness of the BPL. The statistical power of the test for the difference between the proportion of individuals reporting frequent smoking in the two conditions will be increased when a biochemical measure is utilized to detect smokers (see Footnote 4). When only subjects who presented high saliva thiocyanate or carbon monoxide levels were employed, a larger proportion of smokers was detected by self-report measures within the BPL paradigm than by using self-reports alone. In studies that did not use biochemical measures to eliminate nonsmokers from the sample, the proportion of reported frequent smokers did not differ between conditions. The potential effectiveness of the BPL is diminished because of low power to detect the few smokers in each condition.

Prescriptions for Use of the Bogus Pipeline

In contrast to a previous narrative review (Murray et al., 1987), the current meta-analysis *quantitatively* examined the conditions under which a BPL procedure enhances the validity of self-reported smoking. The results are a unique contribution to BPL research because (a) an overall effect size estimate was computed, and (b) categorical tests of presupposed moderating variables were conducted. The overall d for the effectiveness of the BPL was based on a sample size of nearly 6,500 participants. The results of the categorical models tested allow for the following prescriptions to be made regarding the use of the bogus pipeline in self-reported cigarette smoking assessments:

(1) When using the BPL, it is imperative that subjects believe the procedure can accurately detect smoking behavior. If the procedure is not credible, subjects will not be compelled to admit socially undesirable behaviors. A live demonstration of the technique is the most effective. When a live demonstration is not feasible, a video presentation can also be used effectively.

(2) Unless respondents experience pressure to underreport their smoking behavior, the BPL will not enhance the validity of self-reports. The effectiveness of the BPL is maximized when compared to a traditional self-report

only condition. Conversely, the BPL will be least effective when participants believe their self-reports will not be individually attributed (e.g., using a randomized response technique).

(3) Unless there is a large proportion of smokers in each group, the statistical power to detect differences between a BPL and no-BPL condition will not be of sufficient magnitude. Therefore, the BPL should only be utilized in populations suspected to include a substantial proportion of frequent smokers.

Finally, future research on the BPL paradigm and its enhancement of the validity of self-report measures should include both primary and meta-analytic investigations. First, primary research could be conducted to examine potential moderating variables that could not be tested in the present meta-analysis due to a lack of accumulated evidence. For example, presenting the BPL procedure to subjects individually (as opposed to collectively) could increase the credibility and, hence, the effectiveness of the BPL procedure. Second, future meta-analytic reviews need to assess the effectiveness of the BPL paradigm in enhancing the validity of self-report measures focusing on other plausible socially undesirable behaviors and attitudes. For example, alcohol consumption, drug use, and prejudices related to gender and race are important research domains that need to be quantitatively reviewed.

References

Abrams, D. B., Emmons, K. M., Niaura, R. S., Goldstein, M. G., & Sherman, C. B. (in press). Tobacco dependence: An integrative individual and public health perspective. In P. Nathan (Ed.), *Annual review of addictions: Research and treatment*. New York: Pergamon Press.

Akers, R. L., Massey, J., Clarke, W., & Lauer, R. M. (1983). Are self-reports of adolescent deviance valid? Biochemical measures, randomized response, and the bogus pipeline in smoking behavior. *Social Forces,* **62**, 234–251.

Bauman, K. E., & Dent, C. W. (1982). Influence of an objective measure on self-reports of behavior. *Journal of Applied Psychology,* **67**, 623–628.

Bauman, K. E., Koch, G. G., & Bryan, E. S. (1982). Validity of self-reports of adolescent cigarette smoking. *International Journal of the Addictions,* **17**, 1131–1136.

Behar, L. B. (1967). The effects of cognitive dissonance on inappropriate emotional reactions. *Journal of Personality,* **35**, 505–519.

Berman, J., McCombs, H., & Boruch, R. (1977). Notes on the contamination method: Two small experiments in assuring confidentiality of responses. *Sociological Methods and Research,* **6**, 45–62.

Bramel, D. A. (1962). A dissonance theory approach to defensive projection. *Journal of Abnormal and Social Psychology,* **64**, 121–129.

Cherry, F., Byrne, D., & Mitchell, H. E., (1976). Clogs in the bogus pipeline: Demand characteristics and social desirability. *Journal of Research in Personality,* **10**, 69–75.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston, MA: Houghton Mifflin.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive.* New York: Wiley.

DiClemente, C. C., & Prochaska, J. O. (1982). Self-change and therapy change of smoking behavior: A comparison of processes of change in cessation and maintenance. *Addictive Behaviors,* **7**, 133–142.

DiClemente, C. C., & Prochaska, J. O. (1985). Processes and stages of change. Coping and competence in smoking behavior change. In S. Shiffman & T. A. Wills (Eds.), *Coping and substance abuse* (pp. 319–343). San Diego, CA: Academic Press.

Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin,* **111**, 3–22.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden.

Elder, J. P, de Moor, C., Young, R. L., Wildey, M. B., Molgaard, C. A., Golbeck, A. L., Sallis, J. F., & Stern, R. A. (1990). Stages of adolescent tobacco-use acquisition. *Addictive Behaviors,* **15**, 449–454.

Evans, R. I., Hansen, W. B., & Mittlemark, M. B. (1977). Increasing the validity of self-reports of smoking behavior in children. *Journal of Applied Psychology,* **62**, 521–523.

Evans, R. I., Rozelle, R. M., Maxwell, S. E., Raines, B. E., Dill, C. A., Guthrie, T. J., Henderson, A. H., & Hill, P. C. (1981). Social modeling films to deter smoking in adolescents: Results of a three-year field investigation. *Journal of Applied Psychology,* **66**, 399–414.

Fielding, J. (1985a). Smoking: Health effects and control, part I. *New England Journal of Medicine,* **313**, 491–498.

Fielding, J. (1985b). Smoking: Health effects and control, part II. *New England Journal of Medicine,* **313**, 555–561.

Gaes, G. G., Quigley-Fernandez, B., & Tedeschi, J. T. (1978). Unclogging the bogus pipeline: A critical reanalysis of the Cherry, Byrne, and Mitchell study. *Journal of Research in Personality,* **12**, 189–192.

Gerard, H. B., & Rabbie, J. M. (1961). Fear and social comparison. *Journal of Abnormal and Social Psychology,* **62**, 586–592.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Goodman, G. S., Golding, J. M., & Haith, M. M. (1984). Jurors' reactions to child witnesses. *Journal of Social Issues,* **40**, 139–156.

Hedges, L. V. (1986). Issues in meta-analysis. In E. Z. Rothkopf (Ed.), *Review of Research in Education* (Vol. 13, pp. 353–398). Washington, DC: American Educational Research Association.

Hedges, L. V. (1986). Issues in meta-analysis. In E. Z. Rothkopf (Ed.), *Review of research in education* (Vol. 13, pp. 353–398). Washington, DC: American Educational Research Association.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Himmelfarb, S., & Lickteig, C. (1982). Social desirability and the randomized response technique. *Journal of Personality and Social Psychology,* **43**, 710–717.

Johnson, B. T. (1989). *Software for the meta-analytic review of research literatures.* Hillsdale, NJ: Erlbaum.

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin,* **76**, 349–364.

Kenny, D. T. (1956). The influence of social desirability on discrepancy measures between real and ideal self. *Journal of Counseling Psychology,* **20**, 315–318.

Kristein, M. M. (1982). The economics of health promotion at the worksite. *Health Education Quarterly,* **9**, 27–36.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin,* **107**, 34–47.

Luepker, R. V., Pechacek, T. F., Murray, D. M., Johnson, C. A., Hund, F., & Jacobs, D. R. (1981). Saliva thiocyanate: A chemical indicator of cigarette smoking in adolescents. *American Journal of Public Health,* **12**, 1320–1324.

Mahler, D. M. (1990). Health resource allocation in the 1990's. *Business Horizons,* **33**, 118–137.

Manley, M. , Epps, R. P., Husten, C., Glynn, T., & Shopland, D. (1991). Clinical interventions in tobacco control: A national cancer institute training program for physicians. *Journal of the American Medical Association,* **266**, 3172–3173.

Martin, G. L., & Newman, I. M. (1988). Assessing the validity of self-reported adolescent cigarette smoking. *Journal of Drug Education,* **18**, 275–284.

Mullen, B. (1989). *Advanced BASIC meta-analysis: Procedures and programs.* Hillsdale, NJ: Erlbaum.

Murray, D. M., O'Connell, C. M., Schmid, L. A., & Perry, C. L. (1987). The validity of smoking self-reports by adolescents: A reexamination of the bogus pipeline procedure. *Addictive Behaviors,* **12**, 7–15.

Olson, J. M., & Cal, A. V. (1984). Source credibility, attitudes, and the recall of past behaviours. *European Journal of Social Psychology,* **14**, 203–210.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist,* **17**, 776–783.

Orne, M. T. (1969). Demand characteristics and the concept of quasi-control. In R. Rosenthal & R.L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 143–179). San Diego, CA: Academic Press.

Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches.* Dubuque, IA: Wm. C. Brown.

Prochaska, J. O., & DiClemente, C. C. (1983). Stages and processes of self-change of smoking: Toward an integrative model of change. *Journal of Consulting and Clinical Psychology,* **51**, 390–395.

Prochaska, J. O., DiClemente, C. C., Velicer, W. F., Ginpil, S., & Norcross, J. C. (1985). Predicting change in smoking status for self-changers. *Addictive Behaviors,* **10**, 395–406.

Quigley-Fernandez, B. M., & Tedeschi, J. T. (1978). The bogus pipeline as a lie detector: Two validity studies. *Journal of Personality and Social Psychology,* **36**, 247–256.

Rosen, E. (1956). Self-appraisal, personality desirability and perceived desirability of personality traits. *Journal of Abnormal and Social Psychology,* **52**, 151–158.

Rosenberg, M. J. (1965). When dissonance fails: Eliminating evaluation apprehension from attitude measurement. *Journal of Abnormal and Social Psychology,* **56**, 28–43.

Rosenberg, M. J. (1969). The conditions and consequences of evaluation apprehension. In R. Rosenthal & R.L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 279–349). New York: Academic Press.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.

Sigall, H., Aronson, E., & Van Hoose, T. (1970). The cooperative subject: Myth or reality? *Journal of Experimental Social Psychology,* **6**, 1–10.

Sigall, H., & Page, R. (1971). Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology,* **18**, 247–255.

Silvis, G. L., & Perry, C. L. (1987). Understanding and deterring tobacco use among adolescents. *Chemical Dependency,* **34**, 363–379.

Stern, R. A., Prochaska, J. O., Velicer, W. F., & Elder, J. P. (1987). Stages of adolescent cigarettes smoking acquisition: Measurement and sample profiles. *Addictive Behaviors,* **12**, 319–329.

Taylor, S. E. (1986). *Health psychology.* New York: Random House.

Tedeschi, J. T. (Ed.) (1981). *Impression management theory and social psychological research.* New York: Academic Press.

Tedeschi, J. T., Lindskold, S., & Rosenfeld, P. (1985). *Introduction to social psychology*. St. Paul, MN: West Publishing Co.

U.S. Department of Health and Human Services. (1989). *Reducing the health consequences of smoking: 25 years of progress: A report of the surgeon general* (DHHS Publication No. (PHS) 89-8411). Washington, DC: U.S. Government Printing Office.

Valins, S. (1966). Cognitive effects of false heart-rate feedback. *Journal of Personality and Social Psychology,* **4**, 400–408.

Velicer, W. F., Prochaska, J. O., Rossi, J. S., & Snow, M. G. (1992). Assessing outcome in smoking cessation studies. *Psychological Bulletin,* **111**, 23–41.

Viswesvaran, C., & Schmidt, F. L. (1992). A meta-analytic comparison of the effectiveness of smoking cessation methods. *Journal of Applied Psychology,* **77**, 554–561.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association,* **73**, 35–39.

Williams, R., & Gillies, P. (1984). Do we need objective measures to validate self-reported smoking? *Public Health London,* **98**, 294–298.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Beverly Hills, CA: Sage Publications.

Wu, C., & Shaffer, D. R. (1987). Susceptibility to persuasive appeals as a function of source credibility and prior experience with the attitude object. *Journal of Personality and Social Psychology,* **52**, 677–688.

Appendix

Studies Included in the Meta-Analysis

Akers, R. L., Massey, J., Clarke, W., & Lauer, R. M. (1983). Are self-reports of adolescent deviance valid? Biochemical measures, randomized response, and the bogus pipeline in smoking behavior. *Social Forces,* **62**, 234–251.

Bauman, K. E., & Dent, C. W. (1982). Influence of an objective measure on self-reports of behavior. *Journal of Applied Psychology,* **67**, 623–628.

Bauman, K. E., Koch, G. G., & Bryan, E. S. (1982). Validity of self-reports of adolescent cigarette smoking. *International Journal of the Addictions,* **17**, 1131–1136.

Botvin, E. M., Botvin, G. J., Renick, N. L., Filazzola, A. D., & Allegrante, J. P. (1984). Adolescents' self-reports of tobacco, alcohol, and marijuana use: Examining the comparability of video tape, cartoon and verbal bogus-pipeline procedures. *Psychological Reports,* **55**, 379–386.

Evans, R. I., Hansen, W. B., & Mittlemark, M. B. (1977). Increasing the validity of self-reports of smoking behavior in children. *Journal of Applied Psychology,* **62**, 521–523.

Gillies, P. A., Wilcox, B., Coates, C., Kristmundsdottir, F., & Reid, D. J. (1982). Use of objective measurement in the validation of self-reported smoking in children aged 10 and 11 years: Saliva thiocyanate. *Journal of Epidemiology and Community Health,* **36**, 205–208.

Hansen, W. B., Malotte, C. K., & Fielding, J. E. (1985). The bogus pipeline revisited: The use of the threat of detection as a means of increasing self-reports of tobacco use. *Journal of Applied Psychology,* **70**, 789–792.

Hill, P. C., Henderson, A. H., Bray, J. H., & Evans, R. I. (1981). Generalizing a self-report validator of cigarette smoking to older adolescents. *Replications of Social Psychology,* **1**, 38–40.

Lauer, R. M., Akers, R. L., Massey, J., & Clarke, W. R. (1982). Evaluation of cigarette smoking amongst adolescents: The Muscatine study. *Preventive Medicine,* **11**, 417–428.

Luepker, R. V., Pechacek, T. F., Murray, D. M., Johnson, C. A., Hund, F., & Jacobs, D. R. (1981). Saliva thiocyanate: A chemical indicator of cigarette smoking in adolescents. *American Journal of Public Health,* **12**, 1320–1324.

Martin, G. L., & Newman, I. M. (1988). Assessing the validity of self-reported adolescent cigarette smoking. *Journal of Drug Education,* **18**, 275–284.

Murray, D. M., O'Connell, C. M., Schmid, L. A., & Perry, C. L. (1987).

Appendix *Continued*

The validity of smoking self-reports by adolescents: A reexamination of the bogus pipeline procedure. *Addictive Behaviors,* **12**, 7–15.

Werch, C. E., Gorman, D. R., Marty, P. J., Forbess, J., & Brown, B. (1987). Effects of the bogus-pipeline on enhancing validity of self-reported adolescent drug use measures. *Journal of School Health,* **57**, 232–236.

Werch, C. E., Lundstrum, R. H., & Moore, A. (1989). Bogus-pipeline effects on self-reported college student drug use, problems, and attitudes. *International Journal of the Addictions,* **24**, 1003–1010.

Williams, R., & Gillies, P. (1984). Do we need objective measures to validate self-reported smoking? *Public Health London,* **98**, 294–298.