Sankhyā : The Indian Journal of Statistics
Special Issue on Sample Surveys
1999, Volume 61, Series B, Pt. 1, pp. 91-105

AN ANALYSIS OF SURVEY DATA ON SMOKING USING PROPENSITY SCORES

By MICHAEL D. LARSEN
Harvard University, Cambridge

SUMMARY. Responses to questions on perceived health risks from smoking, amount of exercise, overall health, and desire to quit smoking vary across smoking groups. The smoking groups also have very different distributions of covariates. Categorizing respondents based on estimated propensities to smoke, as in blocking, creates groups with less measurable imbalance in distributions of covariates. Adjustment of results using propensity score classes lessens some effect estimates while increasing others and changes results of some significance tests. However, even after adjusting for covariates through the use of propensity scores, smokers report a higher risk of disease than do nonsmokers.

1. **Introduction**

The Gallup Organization conducted a survey of 3019 adults in May, 1994, in which 3017 respondents answered questions regarding personal smoking behavior. Simple comparisons among smoking groups reveal big, significant differences in health and attitude measures. However, smokers and nonsmokers have very different distributions of important demographic characteristics (e.g., sex, age, education). A further analysis, intended to control simultaneously for several continuous and discrete covariates, estimates the probability of membership in smoking groups and compares smoking groups within propensity score classes (Rosenbaum and Rubin 1983a, 1984).

Section 2 describes the survey instrument and results of some simple comparisons. Section 3 briefly reviews propensity scores. Section 4 discusses logistic regression models used to estimate propensity scores and the effect of subclassification on covariate balance between pairs of groups. Section 5 presents outcomes adjusted for propensity score subclassifications.

AMS (1991) *subject classification.* 62D05, 62P99.
Key words and phrases. Logistic regression, subclassification, blocking, observational study.

2. Survey Questions and Initial Results

Telephone numbers were selected according to Random Digit Dialing (RDD) procedures, which enable contacting listed and unlisted household telephone numbers. Survey weights were calculated as inverse probabilities of selection and to adjust the sample to be representative of 1990 census demographics (Chattopadhyay *et al.* 1995).

2.1 *Definition of smoking status.* Respondents who smoked during the week prior to interview are considered to be Current Smokers ($n = 729$, unweighted, abbreviated by S in tables). Current Smokers who said they had quit in the past are called Recidivists ($n = 573$, R), while subjects who had never tried to quit are Never Quitters ($n = 156$, NQ). Respondents who did not smoke during the week prior to interview are Nonsmokers ($n = 2288$, NonS). If a Nonsmoker smoked regularly in the past, the respondent is a Former Smoker ($n = 774$, F). Nonsmokers who never smoked regularly in the past are Never Smokers ($n = 1514$, NS). Ever Smokers (ES) are individuals who are now or have been smokers (all but the Never Smokers).

Other definitions of smoking status exist. Gallup used these definitions in this survey, because they wanted little chance of recall bias. However, under this definition, some Current Smokers smoke relatively few cigarettes each day and smoking regularly is a subjective assessment. Estimates by the Gallup Organization of smoking rates for men and women have been reported since the 1940s (Gallup Poll Monthly 1991 and 1988, Gallup Report 1989) and results are similar to those obtained by the National Health Survey (1989) and Met Life (1992).

2.2 *Questions.* Respondents were asked about their own likelihood of developing heart disease, lung disease such as cancer or emphysema, and mouth or throat cancer from cigarette smoking. They also were asked to evaluate their present overall health and how much they exercise. Ever Smokers were asked whether they think cigarette smoking has already affected their health and would they start smoking again if they had it to do over. Current Smokers were asked whether they would like to quit and consider themselves to be addicted. Weighted means and proportions are presented in tables 1 and 2.

Responses to the questions on risk of specific diseases were very likely (VL), somewhat likely (SL), not very likely (NVL), and not at all likely (NAL). The question should have been clear to smokers, but nonsmokers might have interpreted this question to be concerned with their own risk if they were cigarette smokers. Since it is unlikely that many Nonsmokers believe smoking prevents the diseases, the reported likelihood of disease might be inflated for Nonsmokers because of a misinterpretation of the question wording. Thus, the observed large difference in attitudes between smokers and nonsmokers can be treated as possibly an underestimate of a true, large difference.

Table 1. Outcomes for four smoking groups. perceived risk of diseases, rating of overall health, and amount exercised per week. percentages and means are calculated using survey weights. NAL=not at all likely, NVL=not very likely, SL=somewhat likely, VL=very likely, DK=do not know, AH=already have disease, RF=refuse to answer.

		Smoking Groups			
Outcome	Response	Never Quitters	Recidivists	Former Smokers	Never Smokers
Heart Disease	NAL	0.17	0.12	0.24	0.28
	NVL	0.19	0.16	0.25	0.25
	SL	0.28	0.40	0.20	0.20
	VL	0.29	0.25	0.23	0.24
	DK	0.05	0.05	0.03	0.01
	AH	0.01	0.01	0.04	0.01
Lung Disease	NAL	0.14	0.11	0.24	0.27
	NVL	0.10	0.11	0.26	0.26
	SL	0.33	0.40	0.20	0.16
	VL	0.37	0.34	0.27	0.29
	DK	0.06	0.03	0.02	0.01
	AH	0.00	0.00	0.01	0.00
Mouth or Throat Cancer	NAL	0.21	0.16	0.30	0.35
	NVL	0.21	0.20	0.28	0.25
	SL	0.24	0.40	0.18	0.17
	VL	0.25	0.20	0.20	0.22
	DK	0.08	0.04	0.03	0.02
	RF	0.01	0.00	0.01	0.00
Overall Health	Excellent	0.27	0.24	0.31	0.36
	Good	0.55	0.51	0.51	0.52
	Fair	0.13	0.22	0.16	0.11
	Poor	0.05	0.03	0.02	0.01
Amount Exercised	Mean Days	3.1	2.5	3.1	2.9
	Males	3.9	2.7	3.4	3.3
	Females	2.7	2.3	2.6	2.6

The remaining questions should not be confusing to any respondents. Overall health was rated excellent, good, fair, or poor. The amount of exercise was reported as days per week. A respondent's willingness to "do it again" was recorded as very, somewhat, not very, or not at all likely. The other questions reported in table 2 elicited yes or no answers.

Few respondents said do not know (DK) or refused to answer (RF). Some respondents already have (AH) various diseases. These responses are relatively infrequent and are highest in smoking groups. It is probable that DK, RF, and AH are related to smoking status and are nonignorably missing in the sense of Rubin (1976). Future analyses should study the relationship between these responses and other variables.

2.3 *Description of outcomes.* Never Quitters and Recidivists are more likely than are Former and Never Smokers to say they are very or somewhat likely to get the three diseases. Former Smokers are most likely to already have heart disease. Current Smokers are most likely to respond "do not know" or to refuse

to answer. The groups in order of decreasing overall health are Never Smokers, Former Smokers, Never Quitters, and Recidivists. Recidivists exercise the least. Males report exercising more than females, but other comparisons between males and females are not as striking or consistent across smoking groups.

Table 2. Outcomes for current and former smokers. percentages are calculated using survey weights. questions are: has smoking already affected your health? would you start smoking again if you could do it again? would you like to quit? do you consider yourself to be addicted? *NA*=not applicable.

		Smoking Groups		
Outcome	Response	Never Quitters	Recidivists	Former Smokers
Affect Health	Yes	0.38	0.58	0.32
	No	0.57	0.39	0.62
	DK	0.04	0.04	0.06
	RF	0.02	0.00	0.00
Do It Again	NAL	0.44	0.54	0.81
	NVL	0.10	0.19	0.09
	SL	0.17	0.12	0.04
	VL	0.26	0.12	0.05
	DK	0.02	0.02	0.01
	RF	0.01	0.00	0.00
Like to Quit	Yes	0.54	0.74	*NA*
	No	0.45	0.23	*NA*
	DK	0.01	0.03	*NA*
Addicted	Yes	0.69	0.63	*NA*
	No	0.28	0.35	*NA*
	DK	0.03	0.02	*NA*

A Recidivist is more likely than a Never Quitter or a Former Smoker to say smoking already has had a health impact. Former Smokers are less likely than Recidivists who are less likely than Never Quitters to say they would start smoking again if they could choose. Three-quarters of Recidivists and one-half of Never Quitters say they desire to quit smoking, while two-thirds of these two groups say they are addicted.

Tables 3 and 4 present odds ratios and mean differences. Odds ratios are ratios of odds for saying very or somewhat likely (VL/SL) versus not very or not at all likely (NVL/NAL) on the heart, lung, mouth/throat, and "do it again" questions, for saying excellent versus good, fair, or poor regarding overall health, and for saying yes versus no on the other questions. The divisions of responses for calculating odds were chosen after looking at tables 1 and 2. Mean differences are reported for exercise. All estimates use weights.

2.4 *Assessment of significance.* Overall results are very significant, so results are presented for several comparisons in tables 3 and 4. No adjustment has been made for multiple comparisons. This is not a formal epidemiologic analysis. Ninety-percent confidence intervals for the odds ratio are based on percentiles of

a nonparametric bootstrap procedure (Efron and Tibshirani 1993) in which one-thousand bootstrap samples were generated by sampling rows with replacement from the complete data set. Estimates for all comparisons were calculated, as they were for the real data, for each replicate data set.

Table 3. INITIAL COMPARISON OF SMOKING-STATUS GROUPS. ODDS RATIOS ARE RATIOS OF ODDS WHERE ODDS ARE FOR VL/SL VERSUS NVL/NAL FOR HEART, LUNG, AND MOUTH/THROAT DISEASE AND EXCELLENT VERSUS GOOD/FAIR/POOR FOR OVERALL HEALTH. A RATIO GREATER THAN ONE MEANS THE FIRST GROUP LISTED HAS A HIGHER PROPORTION OF RESPONSES IN THE FIRST CATEGORY THAN DOES THE SECOND GROUP. MEAN DIFFERENCE IS REPORTED FOR EXERCISE. THE GROUPS ARE NEVER QUITTERS (NQ), RECIDIVISTS (R), FORMER SMOKERS (F), NEVER SMOKERS (NS), CURRENT SMOKERS (S), NON SMOKERS (NONS), AND EVER SMOKERS (ES).

	Comparisons of Smoking-Status Groups							
Outcome	NQ/R	NQ/F	NQ/NS	R/F	R/NS	F/NS	S/NonS	ES/NS
Heart Disease								
Odds Ratio	0.69	1.79	1.94	2.60	2.82	1.09	2.52	1.66
Interval	(0.48, 1.02)	(1.27, 2.69)	(1.40, 2.81)	(2.07, 3.26)	(2.33, 3.45)	(0.91, 1.30)	(2.12, 3.05)	(1.43, 1.92)
P-value, χ^2	0.04	0.01	0.00	0.00	0.00	0.48	0.00	0.00
P-value, boot	0.11	0.05	0.00	0.00	0.00	0.59	0.00	0.00
Lung Disease								
Odds Ratio	0.85	3.01	3.38	3.56	3.99	1.12	3.70	1.97
Interval	(0.57, 1.29)	(2.11, 4.51)	(2.41, 4.95)	(2.80, 4.66)	(3.20, 5.02)	(0.94, 1.33)	(3.09, 4.53)	(1.73, 2.26)
P-value, χ^2	0.39	0.00	0.00	0.00	0.00	0.04	0.00	0.00
P-value, boot	0.55	0.00	0.00	0.00	0.00	0.10	0.00	0.00
Mouth/Throat								
Odds Ratio	0.71	1.81	1.83	2.56	2.59	1.01	2.39	1.55
Interval	(0.49,0.99)	(1.27, 2.56)	(1.30, 2.57)	(2.07, 3.22)	(2.13, 3.16)	(0.86, 1.20)	(2.02, 2.86)	(1.36, 1.79)
P-value, χ^2	0.00	0.01	0.00	0.00	0.00	0.11	0.00	0.00
P-value, boot	0.02	0.04	0.01	0.00	0.00	0.22	0.00	0.00
Overall Health								
Odds Ratio	1.21	0.84	0.66	0.69	0.54	0.79	0.61	0.67
Interval	(0.79, 1.81)	(0.57, 1.21)	(0.44, 0.92)	(0.54, 0.89)	(0.43, 0.67)	(0.65, 0.94)	(0.51, 0.73)	(0.57, 0.77)
P-value, χ^2	0.03	0.20	0.00	0.00	0.00	0.00	0.00	0.00
P-value, boot	0.09	0.34	0.03	0.02	0.00	0.01	0.00	0.00
Exercise								
Difference	0.55	0.03	0.19	-0.52	-0.37	0.15	-0.30	-0.05
Interval	(0.06, 1.05)	(-0.43, 0.46)	(-0.25, 0.62)	(-0.78, -0.25)	(-0.61, -0.13)	(-0.06, 0.35)	(-0.51, -0.09)	(-0.23, 0.12)
P-value, *t*	0.02	0.88	0.41	0.00	0.00	0.17	0.01	0.59
P-value, boot	0.05	0.90	0.47	0.00	0.01	0.22	0.01	0.63

Tests of hypotheses were conducted by two methods, which produce quantitatively similar results. First, parametric approximations were used. *P*-values for the categorical variables are computed based on comparing a likelihood ratio statistic for homogeneity in a 2×4 or 2×2 table to a chi-square distribution with three degrees or one degree of freedom, respectively. *P*-values for the exercise question are based on a two-sample *t*-test. The post-stratification weighting reflects the fact that the observations do not have equal probabilities of selection. Familiar statistics do not have usual asymptotic distributions in non i.i.d. sampling (see, e.g., Rao and Scott 1981).

Second, for comparison, one-thousand bootstrap replicates were generated based on the hypothesis that group membership and outcomes are independent. That is, for each replicate data set, outcome vectors were sampled with replacement from the observed data and paired with vectors of randomly sampled indicators of smoking status. *P*-values computed from the bootstrap reference distribution are similar to, but larger than those based on the parametric approximation.

Another alternative for performing significance tests is the method of Rao and Thomas (1989). The bootstrap method used here provides interval estimates. The bootstrap also is used in section 5 after propensity-score adjustment.

Table 4. Initial comparison of outcomes for smokers. individuals who have tried to quit are recidivists or former smokers. odds ratios are ratios of odds where odds are for VL/SL versus NVL/NAL on "do it again" question and yes versus no for other questions.

	Smoking Group Comparisons				
Outcome	NQ/R	NQ/F	R/F	S/F	NQ/Try to quit
Affect Health					
Odds Ratio	0.44	1.28	2.88	2.42	0.79
Interval	(0.31, 0.65)	(0.91, 1.86)	(2.30, 3.61)	(1.98, 2.99)	(0.57, 1.12)
P-value, χ^2	0.00	0.18	0.00	0.00	0.18
P-value, boot	0.00	0.24	0.00	0.00	0.24
Do It Again					
Odds Ratio	2.36	8.04	3.41	4.19	4.16
Interval	(1.62, 3.47)	(5.49, 11.71)	(2.54, 4.51)	(3.22, 5.48)	(2.88, 6.02)
P-value, χ^2	0.00	0.00	0.00	0.00	0.00
P-value, boot	0.00	0.00	0.00	0.00	0.00
Like to Quit					
Odds Ratio	0.38	*NA*	*NA*	*NA*	*NA*
Interval	(0.25, 0.54)	*NA*	*NA*	*NA*	*NA*
P-value, χ^2	0.00	*NA*	*NA*	*NA*	*NA*
P-value, boot	0.00	*NA*	*NA*	*NA*	*NA*
Addicted					
Odds Ratio	1.39	*NA*	*NA*	*NA*	*NA*
Interval	(0.94, 2.05)	*NA*	*NA*	*NA*	*NA*
P-value, χ^2	0.09	*NA*	*NA*	*NA*	*NA*
P-value, boot	0.15	*NA*	*NA*	*NA*	*NA*

2.4.1 *Comparisons of smokers and nonsmokers.* In general, when any smoking group is compared to any nonsmoking group, results on health questions are significant. Smokers tend to be more than twice as likely as Nonsmokers to say they are at risk from diseases and to be in worse overall health. When Never Quitters are compared to Recidivists (both Current Smokers) and when Former Smokers are compared to Never Smokers (both Nonsmokers) results are not significant and odds ratios are closer to one. The division between Smokers and Nonsmokers produces larger differences than between Ever and Never Smokers, because Former Smokers tend to be more similar to Nonsmokers than Current Smokers.

The Recidivists exercised the least amount per week and in terms of exercise are significantly different from the Former and Never Smokers and at the borderline of significance versus the Never Quitters. Smokers exercised significantly less than nonsmokers.

2.4.2 *Comparisons of current and former smokers.* Results on questions of Current and Former Smokers are presented in table 4. Recidivists are significantly more likely than are Never Quitters and Former Smokers to say smoking already has affected their health. All comparisons of the groups on whether respondents would start smoking again if they had another chance are significant, with Never Quitters being the most and Former Smokers being the least likely to start again. The odds a Never Quitter wants to quit are about one-third the odds for a Recidivist, however the level of addiction is not quite significantly different.

One strategy for continuing the analysis is to extensively poststratify the sample on many qualitative and categorized variables and analyze outcomes within subgroups of respondents. However, sample sizes quickly become prohibitively small and simple divisions simultaneously adjust for very few covariates. The alternative pursued here is to estimate propensity scores conditional on several covariates and poststratify according to estimated scores.

3. **Propensity Scores**

A propensity score is the probability of membership in a subpopulation given values for covariates (Rosenbaum and Rubin 1983a). In practice, the probabilities can be estimated via a multiple logistic regression that uses as a response variable a dichotomous indicator of membership in a group versus an alternative group. Separate sets of propensity scores are estimated for every pair of groups being compared, since covariate differences vary across pairings of groups (Rubin 1997).

Theory supports the use of propensity scores and their estimates in subclassifying subjects in observational studies (Rosenbaum and Rubin 1983a, 1984). The propensity score is a scalar summary of several covariates. In large samples, dividing subjects into five or six equal-sized groups according to propensity score values should reduce the imbalance between groups in the distributions of covariates used to estimate the propensity scores by ninety percent (Rosenbaum and Rubin 1984). The idea of propensity score subclassification is analogous to matching or blocking to deal with bias due to confounding variables in observational studies, which have been discussed in sources such as Cochran (1953), Greenberg (1953), Cochran and Rubin (1973), and Rubin (1979). While actual bias cannot be measured, observed balance on covariates can be measured by comparing distributions in pairs of subgroups.

The choice of covariates to use in estimating propensity scores is important (see, e.g., Rosenbaum and Rubin 1983b and 1985 and Rosenbaum 1984). For

example, amount smoked and years smoked have prima facie relevance in estimating the probability of being a Former Smoker versus a Recidivist. However, it is difficult to quantify a person's smoking history by responses to a few survey questions. Also, variables better viewed as outcomes rather than as covariates should not be used to estimate propensity scores.

4. **Propensity Score Estimation**

Different logistic regressions were fit for each comparison of two smoking status subpopulations. Simple rules were followed for building the regression models. A set of main effects including indicators of missing values were defined and selected using an automated step-wise procedure. The percent of "do not know" and refuse responses on almost any particular variable is generally less than two percent and does not vary much by smoking group. Income is not reported by almost ten percent of the sample. The stepwise selection procedure was begun with all main effects in the model.

Terms were added to decrease imbalance in resulting propensity subclasses. Variables considered for addition to the model included omitted main effects and interactions between and quadratic terms for the variables in the model. Additionally, the survey weights, which themselves summarize information such as number of adults in the household, were tried as an additional covariate. Section 4.1 discusses model building in estimating propensity scores for distinguishing Recidivists versus Former Smokers. Section 4.2 presents measurements of covariate imbalance before and after subclassification on the estimated propensity scores for all comparisons.

No claim is being made that the regressions identify factors that cause respondents to smoke or not to smoke. The logistic regression models are simply being used to describe categories of respondents.

4.1 *Model building example.* In the comparison of Recidivists versus Former Smokers, the main effects model includes 12 out of 22 possible predictors. Other things being equal, the younger the respondents are, the more likely they are to be Recidivists rather than Former Smokers. Also predictive of being a Recidivist rather than a Former Smokers is having less education, not being Hispanic, being white, and earning less than $15,000. The older respondents were when they started smoking, the more likely they are to be Recidivists. If respondents did not answer the phone question or have more than one phone line in the household, they are more likely to be Former Smokers.

When interactions and other terms are allowed, the model expands to include twenty terms and fits significantly better than the main effects model with twelve predictors. Models were developed in the same manner for the other pairs of comparison groups. Nonsmokers were not asked as many questions as were smokers. Never smokers in particular did not have any smoking history to report.

Table 5. Initial balance on 22 variables: recidivists versus former smokers. differences of weighted means are reported for quantitative variables. all categorical variables are dichotomous.

	Measurements and Significance of Imbalance			
Quantitative Variables	Difference		*t*-statistic	*P*-value
Age	-12.19		-14.67	0.00
Age Start Smoking	-0.17		-0.58	0.56
Income	-6.78		-4.02	0.00
Categorical Variables	Difference	Odds Ratio	LRT statistic	*P*-value
Respond to Age Start Smoking	-0.01	0.55	2.56	0.11
Respond to How Start Smoking	0.01	2.34	1.58	0.21
How Start Smoking	0.06	1.30	4.85	0.03
Respond to Education	0.00	0.71	0.33	0.57
Less than High School Education	0.06	1.68	1.43	0.00
College Education	-0.12	0.51	26.73	0.00
Graduate School	-0.06	0.37	19.99	0.00
Respond to Hispanic Question	0.01	1.62	5.03	0.02
Hispanic Ethnicity	0.00	1.00	0.00	1.00
Respond to Race Question	0.01	2.81	3.87	0.05
White Race	-0.01	0.85	0.83	0.36
Black Race	0.02	1.39	2.28	0.13
Respond to Income Question	0.04	1.68	6.95	0.01
Less than $15,000 Income	0.06	1.64	9.55	0.00
More than $100,000 Income	-0.02	0.70	2.33	0.13
Respond to Number of Phones	0.01	11.77	5.78	0.02
One Phone Line in Household	0.01	1.13	0.44	0.51
Two Phone Lines in Household	-0.01	0.89	0.39	0.53
Sex	-0.09	0.70	1.24	0.00

4.2 *Initial and resulting balance.* The balance on covariates was assessed by comparing distributions in pairs of smoking groups. Results for the comparison of Recidivists and Former Smokers are in table 5. Weighted means and standard deviations were calculated and used in two-sample *t*-tests for variables age, age started smoking, and amount smoked, which is relevant when comparing groups of current smokers. For categorical variables, likelihood ratio test statistics for homogeneity of proportions using weighted cell counts were computed and compared to chi-square distributions. The largest single difference between Recidivists and Former Smokers occurs for the variable Age. Several variables have large, significant imbalance.

Table 6 summarizes the balance on variables across all comparison pairs initially and after the use of propensity scores. The numbers of variables that are different at four levels of significance between groups are reported. For example, for the comparison of Recidivists versus Former Smokers, eight variables have *P*-values for comparisons that are 0.01 or less, four are above 0.01 but 0.05 or less, none are above 0.05 but 0.10 or less, and ten are above 0.10. All comparison groups have some covariate distributions initially that are significantly out of balance.

Table 6. Summary of balance on covariates before and after subclassification on propensity scores. The number of variables that are significantly out of balance when comparing pairs of groups. The significance levels are (a) .01 or less, (b) above .01 but .05 or less, (c) above .05 but .10 or less, and (d) above .10.

		Balance Initially and Under Two Models: Number of Variables with More to Less Imbalance											
	Total Variables	Initial Model				Main Effects Model				Expanded Model			
		a	b	c	d	a	b	c	d	a	b	c	d
NQ/R	24	2	0	1	21	0	1	3	20	0	0	2	22
NQ/F	22	6	3	0	13	0	1	3	18	0	0	1	21
NQ/NS	18	4	2	2	10	0	0	1	17	0	0	1	17
R/F	22	8	4	0	10	0	3	3	16	0	0	3	19
R/NS	18	7	4	1	6	4	0	2	12	0	1	3	14
F/NS	18	4	1	0	13	2	2	1	13	1	0	2	15
S/NonS	18	7	3	1	7	1	4	1	12	0	0	1	17
ES/NS	18	8	1	0	9	2	2	3	11	0	3	1	14
S/F	22	9	4	0	9	2	2	2	16	0	0	0	22
NQ/Try to Quit	22	2	4	1	15	1	2	0	19	0	0	2	20

Respondents were subclassified into six groups according to sextiles of estimated propensity scores. Balance after subclassification is measured by looking at the association between smoking group and question response conditional on propensity score class. *P*-values in table 6 for the two models are assessments of whether or not the interaction between smoking group and question response is significantly different from zero in a log linear model for the categorical responses and in an analysis of variance model for exercise. Substantial imbalance was eliminated under both the main effects and expanded models in all comparisons. The variables that are not balanced by subclassification using the expanded model were severely out of balance initially. For example, the variable with significant differences after subclassification on propensity scores estimated using the expanded model between the Former Smokers, who were on average eleven years older than Never Smokers, and Never Smokers is age. Perhaps if the variable age were transformed or used differently in the model, balance would have been achieved.

5. **Results**

The outcome measures are reanalyzed here using the subclassifications on the propensity scores estimated from the expanded models. Adjustment based on propensity-score classes can change results if covariates such as age, sex, education, and income that are used as predictors are related to smoking status and outcomes. No adjustment for making multiple comparisons has been made. Different logistic regression models are used to estimate propensity scores are for the various comparison groups. Results are presented in tables 7 and 8.

Table 7. COMPARISON OF OUTCOMES FOR SMOKING GROUPS AFTER SUBCLASSIFICATION ON PROPENSITY SCORES FROM THE EXPANDED MODEL.

	Comparisons of Smoking-Status Groups							
Outcome	NQ/R	NQ/F	NQ/NS	R/F	R/NS	F/NS	S/NonS	ES/NS
Heart Disease								
Odds Ratio	0.51	1.30	1.92	2.09	2.56	1.02	2.35	1.59
Interval	(0.34, 0.80)	(0.84, 2.44)	(1.30, 2.71)	(1.55, 2.63)	(2.07, 3.16)	(0.83, 1.27)	(1.96, 2.83)	(1.39, 1.85)
P-value, χ^2	0.79	0.45	0.18	0.00	0.00	0.97	0.00	0.00
Lung Disease								
Odds Ratio	0.54	2.25	3.38	3.22	3.50	1.04	3.38	1.90
Interval	(0.34, 0.82)	(1.31, 3.81)	(2.21, 4.76)	(2.38, 4.10)	(2.79, 4.41)	(0.86, 1.28)	(2.74, 4.11)	(1.65, 2.20)
P-value, χ^2	0.78	0.02	0.00	0.00	0.00	0.09	0.00	0.00
Mouth/ Throat								
Odds Ratio	0.53	1.18	2.19	2.27	2.31	0.89	2.20	1.47
Interval	(0.36, 0.77)	(0.77, 1.66)	(1.51, 2.93)	(1.71, 2.88)	(1.85, 2.80)	(0.73, 1.10)	(1.81, 2.64)	(1.26, 1.70)
P-value, χ^2	0.47	0.80	0.08	0.00	0.00	0.20	0.00	0.00
Overall Health								
Odds Ratio	1.47	0.81	0.69	0.69	0.59	0.75	0.66	0.71
Interval	(0.91, 2.16)	(0.49, 1.57)	(0.45, 1.03)	(0.54, 0.90)	(0.47, 0.74)	(0.62, 0.92)	(0.54, 0.78)	(0.60, 0.83)
P-value, χ^2	0.75	0.51	0.15	0.23	0.00	0.37	0.01	0.00
Exercise								
Difference	0.75	0.35	0.23	-0.59	-0.34	-0.08	-0.35	-0.17
Interval	(0.23, 1.24)	(-0.68, 0.76)	(-0.37, 0.69)	(-0.88, -0.29)	(-0.60, -0.08)	(-0.33, 0.18)	(-0.55, -0.11)	(-0.35, 0.01)
P-value, *t*	0.01	0.53	0.39	0.00	0.01	0.55	0.29	0.53

The overall adjusted estimate for a smoking group is the direct average of its propensity-score subclass means or percentages, since subclasses are chosen to be the same size. An alternative method of using the propensity scores would be as the basis for new weights (Czajka *et al.* 1992). The odds ratios that are reported are computed from adjusted proportions. Tests for health questions in table 7 and questions asked of smokers in table 8 are for independence between response and smoking status conditional on propensity score class. Tests for exercise are based on contrasts among means in subgroups defined by smoking status and propensity class. The variance for the contrast is estimated as a equally-weighted sum of sample variances, one for each combination of propensity-score class and smoking group. The sample variances are computed using the sample weights and treat the propensity score classes as fixed.

A nonparametric bootstrap procedure was used to assess significance as it was in tables 3 and 4. Five-hundred replicate data sets were generated. Variables were not re-selected and propensity scores were not re-estimated for each bootstrap sample and pair of comparison groups. Conditional on the values of the propensity scores, confidence intervals were calculated and are reported in tables 7 and 8. Interval widths are similar to those in the previous tables.

Table 8. Comparison of outcomes for smokers after subclassification on propensity scores from the expanded model.

	Smoking Group Comparisons				
Outcome	NQ/R	NQ/F	R/F	S/F	NQ/Try to quit
Affect Health					
Odds Ratio	0.35	1.22	2.98	2.45	0.72
Interval	(0.29, 0.51)	(0.70, 2.70)	(2.33, 3.88)	(1.98 , 3.07)	(0.44, 1.20)
P-value, χ^2	0.00	0.49	0.00	0.00	0.76
Do It Again					
Odds Ratio	3.05	8.26	2.84	3.52	4.82
Interval	(2.01, 4.37)	(4.43, 19.13)	(2.10, 3.86)	(2.59, 4.80)	(3.01, 7.39)
P-value, χ^2	0.07	0.00	0.00	0.00	0.00
Like to Quit					
Odds Ratio	0.28	*NA*	*NA*	*NA*	*NA*
Interval	(0.19, 0.41)				
P-value, χ^2	0.00	*NA*	*NA*	*NA*	*NA*
Addicted					
Odds Ratio	0.98	*NA*	*NA*	*NA*	*NA*
Interval	(0.64, 1.39)				
P-value, χ^2	0.64	*NA*	*NA*	*NA*	*NA*

5.1 *Health and exercise questions.* After adjustment for covariates through propensity score subclassification, odds ratios for Never Quitters versus Recidivist on disease questions decease away from one, but less frequently are significant. For example, the odds ratio comparing Never Quitters to Recidivists for being very or somewhat likely to have heart disease from smoking decreases from 0.69 to 0.51. However, the P-value based on the adjusted estimate clearly is not significant. A result with more extreme odds ratio, but decreasing significance is possible because the P-value is based on a test that does not use the odds ratio as a statistic and also because it is a significance level for a test of conditional independence rather than of independence as for the unadjusted data.

The odds ratios for risk of disease comparing Never Quitters to Former Smokers decrease towards one and are less significant. Odds ratios do not change in a consistent direction for Never Quitters versus Former Smokers. The results for Recidivists versus Former or Never Smokers decrease in magnitude, but remain significant. When Smokers and Nonsmokers or Ever and Never Smokers are compared, results again are decreased but still significant. There is little change in the Former versus Never Smokers, the comparison for which balance on age was not achieved through subclassification.

On the question of overall health, results for the pair-wise comparisons of smoking groups generally are less significant after adjustment. The odds ratio for Recidivists, the group in the worst health, versus Never Smokers, the healthiest group, changes little and remains significant. The contrasts of Smokers to Nonsmokers and Ever to Never Smokers after adjustment also remain significant.

The mean number of days exercised changes by a fraction of a day for most groups. Results that were not significant before adjustment remain non significant. The mean difference estimates between Never Quitters and Recidivists

and Recidivists versus Former Smokers are increased.

5.2 *Smokers' opinions.* After adjustment, odds ratios, which are reported for the three questions in table 8, for comparing Never Quitters to Recidivists are more extreme. Never Quitters appear even less likely relative to Recidivists to think smoking already has had a health impact, wish they had not started smoking, and want to quit. That is, Never Quitters relative to Recidivists appear to be more resistant to quitting smoking after adjustment. The odds ratio for addiction for these two groups, however, is closer to one.

There is not much change when comparing Never Quitters and Former Smokers. The odds ratio for Recidivists versus Former Smokers on the question of starting smoking, while still significant, decreases a lot. Changes in results for Current versus Former Smokers and Never Quitters versus Smokers who have tried to quit reflect the pair-wise changes in comparisons versus Recidivists.

Comparing pairs of smoking groups within propensity score classes provides some interesting descriptions of the sample. For example, Never Quitters report less risk of disease than do Recidivists. As the predicted probability of being a Never Quitter increases, the reported likelihood of disease also increases, especially for the Never Quitters. That is, the higher the estimated probability of being a Never Quitter, the more Never Quitters and Recidivists report that they are incurring risk from smoking. The biggest difference in reported risk occurs in the propensity class predicted to be Recidivists.

An important predictor for the status of Never Quitter versus Recidivist is amount smoked. Thus it appears that adjusted for other factors respondents smoking a lot recognize their risk of disease, but Never Quitters smoking less do not feel as much at risk. Similarly, as the estimated probability of being a Never Quitter versus a Recidivist increases, Never Quitters increasingly report that their health has been affected, they would not start smoking again, and they want to quit.

6. **Conclusions**

Differences in responses among smoking groups is due possibly in part to differences in observable covariates. Sample size is inadequate to poststratify on all important variables and their interactions. Estimated propensity scores are an univariate summary of the differences between groups. In the example, subclassification on propensity scores, which were estimated using logistic regression, simultaneously balanced the covariate distributions of several variables. Although some effects are reduced when adjusting via propensity scores for covariate differences, others are increased and there still appears to be real, large differences between pairs of smoking groups. In particular, Smokers respond that they are putting themselves at higher risk for heart, lung, and mouth and throat disease and want to quit, but keep on smoking. Results such as those presented here could have implications for anti-smoking initiatives.

Acknowledgments. The author would like to thank members of the Gallup Organization: Elaine Christiansen for initially involving him in Gallup's smoking study and discussing the project, Ness Sandoval for help accessing the data, and Manas Chattopadhyay for explaining Gallup's weighting system and the smoking project in general. Also thanks to Don Rubin of Harvard University for advice and instruction about propensity scores and causal analysis and a referee for very helpful comments.

References

Chattopadhyay, M., Srinivasan, R., Christiansen, E., and Gaertner, G. (1995). A practical application of two telephone sampling designs, *Proc. of the ASA, Section of Survey Research Methods*, 945-950.

Cochran, W.G. (1953). Matching in analytical studies. *Amer. Jour. Public Health*, **43**, 684-691.

Cochran, W.G., and Rubin, D.B. (1973). Controlling bias in observational studies: a review. *Sankhya-A*, **35**, 417-446.

Czajka, J.L., Hirabayashi, S.M., Little, R.J.A., and Rubin, D.B. (1992). Projecting from advance data using propensity modeling: an application to income and tax statistics. *Jour. Business and Economic Statistics*, **10**, 117-131.

Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap.* Chapman and Hall.

Gallup Poll Monthly. Despite growing concerns, most nonsmokers oppose stringent bans, 21-26, March 1991.

– – –– Despite increasing hostility, one in four Americans still smokes, 2-10, December 1991.

– – –– Majorities now support ban on public smoking, cigarette advertising, although one-third of Americans continue to smoke, 37-40, September 1988.

Gallup Report. Cigarette smoking at 45-year low, 23-26, July 1989.

Greenberg, B.G. (1953). The use of analysis of covariance and balancing in analytical surveys. *Amer. Jour. Public Health*, **43**, 692-699.

Met Life. (1992). Cigarette smoking among U.S. adults, 1985-1990. *Statistical Bulletin*, 2-19.

Rao, J.N.K., and Scott, A.J. (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *Jour. Amer. Statist. Assoc.*, **76**, 221-230.

Rao, J.N.K., and Thomas, D.R. (1989). Chi-squared tests for contingency tables, in *Analysis of Complex Surveys*, C.J. Skinner, D. Holt, and T.M.F. Smith eds., 89-114, John Wiley and Sons: Chichester.

Rosenbaum, P.R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Jour. Royal Statist. Soc.,* Series A, **147**, 656-666.

Rosenbaum, P.R., and Rubin, D.B. (1983a). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41-55.

– – –– (1983b). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Jour. Royal Statist. Soc.,* Series B, **45**, 212-218.

Rosenbaum, P.R., and Rubin, D.B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Jour. Amer. Statist. Assoc.*, **79**, 516-524.

– – –– (1985). The bias due to incomplete matching. *Biometrics*, **41**, 103-116.

RUBIN, D.B. (1976). Inference and missing data. *Biometrika*, **63**, 581-590.
– – ––(1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Jour. Amer. Statist. Assoc.*, **74**, 318-328.
––––(1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*, **127**, 757-763.
VITAL AND HEALTH STATISTICS (1989). *Smoking and Other Tobacco use: United States, 1987*. Series 10: Data from the National Health Survey No. 169. Hyattsville, MD; U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Center for Health Statistics, 1989. DHHS Publication No. (PHS) 89-1597.

MICHAEL D. LARSEN
DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY
1 OXFORD STREET
CAMBRIDGE, MA 02138 U.S.A.
617-495-1600
e-mail : larsen@stat.harvard.edu