

The Behavioral Risk Factor Surveillance System Questionnaire: Its Reliability in a Statewide Sample

ABSTRACT

The reliability of the Behavioral Risk Factor Surveillance System questionnaire was assessed in a random sample of adults ($n = 122$) and a separate sample of Black and Hispanic adults ($n = 200$) in Massachusetts. The questionnaire was administered twice, 21 to 44 days apart, by telephone (210 completed reinterviews, 65% response rate for second administration). There were no statistically significant differences in the distribution of demographic or risk factor variables across administrations. Individual-level reliability (kappa for categorical variables, correlation for continuous variables) for demographic characteristics was more than 0.80 for White respondents and more than 0.60 for Black and Hispanic respondents. Employment and income were reported less consistently than other variables. Reliability coefficients for behavioral risk factors were generally above 0.70. Exceptions were variables with extreme distributions. These data support the use of the Behavioral Risk Factor Surveillance System questionnaire for surveillance and research. (*Am J Public Health*. 1993;83:1768-1772)

Aryeh D. Stein, PhD, Ruth I. Lederman, MPH, and Steven Shea, MD

Introduction

The Behavioral Risk Factor Surveillance System (BRFSS), an ongoing surveillance program developed and funded by the Centers for Disease Control and Prevention (CDC), is designed to estimate the prevalence of risk factors for the major causes of death in the United States.¹ Data from the BRFSS are a central component of federal and state activities designed to monitor progress toward achieving the health objectives for the year 2000.² We assessed the test-retest reliability of the BRFSS questionnaire in a representative sample of Massachusetts residents.

Methods

The BRFSS questionnaire has been described in detail elsewhere.¹ Demographic and risk factor variables are shown in Tables 1 and 2.

Two separate populations were recruited. The first consisted of respondents to the regular March 1992 sample of the Massachusetts BRFSS. The sampling methods were consistent with published CDC procedures. All interviews were conducted in English. Inability of the selected individual to participate in a telephone interview or speak English was considered a cause of ineligibility. With known eligible households as the denominator, the response rate for this survey was 65%.

The second population consisted of a separately recruited sample of Black and Hispanic residents of Massachusetts. The sampling frame for this study consisted of telephone exchanges in towns identified in the 1990 census as containing high concentrations of Blacks and Hispanics. Households were eligible if at least one household member was identified by the initial respondent as being Black or Hispanic. A random member of the household, irrespective of ethnicity, was interviewed. If necessary, a translation prepared by the CDC was used to conduct the interview in Spanish. With known eligible households as the denominator, the response rate for this survey was 39%.

A computer-assisted telephone interview system was used to collect data. In the first interview, respondents were not told that they might be contacted again. Callbacks for the second interview began 21 days after the first interview. Efforts were made to reinterview all of the original respondents, and these efforts were of the same intensity as those used to obtain the first interview. Eighty-six of 122 respondents to the regular survey (70%) and 124 of 200 respondents to the minority survey (62%) were successfully reinterviewed. Ninety-five percent of the 210 successful second interviews were completed within 44 days, and all were completed within 64 days. Univariate distributions of demographic variables, as recorded during the first interview, did not differ significantly ($P > .05$ for each demographic variable) between the 210 respondents and the 112 nonrespondents to the repeat interview.

Subjects were asked to identify race and whether they were of Hispanic origin. Subjects were classified as White non-Hispanic, Black non-Hispanic, or Hispanic according to the categories used by the CDC in reporting BRFSS data. Of 86 participants in the first sample, 11 were classified as non-White based on their responses to the first interview. Of 124 participants in the second sample, 15 were classified at the first interview as neither Black non-Hispanic nor Hispanic. Data from these 26 individuals were included in pooled analyses but were excluded from race/ethnicity-specific analyses.

At the time of the study, Aryeh D. Stein and Ruth I. Lederman were with the Bureau of Health Statistics, Research and Evaluation, Massachusetts Department of Public Health, Boston. Aryeh D. Stein was also with the Division of Epidemiology, Columbia University School of Public Health, New York, NY. Steven Shea was with the Division of Epidemiology, Columbia University School of Public Health, and the Division of Medicine, College of Physicians and Surgeons, Columbia University.

Requests for reprints should be sent to Aryeh D. Stein, PhD, Program in Epidemiology, A214 E Fee Hall, Michigan State University, East Lansing, MI 48824-1316.

This paper was accepted May 11, 1993.

Algorithms developed at the CDC (Dr Emma Frazier, written communication, July 1992) were used in computations of behavioral risk status from questionnaire items. Individuals with unknown values for demographic characteristics on either administration of the questionnaire were excluded from determination of group distributions for that administration. An exception was income; more than 5% of respondents refused to state or reported not knowing their annual household income and were considered a stratum for analysis. Individuals of unknown risk factor status were considered not to be at risk. This approach consistently underestimates prevalence because unknown responses are retained in the denominator.

The *z* test (for dichotomous variables and proportions, such as gender and risk status), the chi-square test (for polychotomous variables, such as educational attainment), and the *t* test (for means of continuous variables, such as cholesterol level) were used in assessing the consistency of distributions. The kappa statistic³ (for dichotomous variables), unweighted kappa (for polychotomous variables), and Pearson's correlation coefficient (for continuous variables) were used in examining the consistency of individual responses. Individuals with unknown values on ei-

TABLE 1—Demographic Characteristics Available for Analysis in the 1992 Behavioral Risk Factor Surveillance System Questionnaire

Demographic Characteristic	Categories Used in Analysis
Age	Recorded in exact y
Sex	Male Female
Education	No high school/some high school High school graduate/some technical school/technical school graduate Some college but did not graduate College graduate/postgraduate or professional degree
Marital status	Married/living as a couple Divorced/separated Widowed Never married
Employment	Employed Self-employed Unemployed less than 1 y/unemployed more than 1 y Homemaker Retired Student
Income	Under \$10 000/\$10 000–\$15 000 \$15 000–\$20 000/\$20 000–\$25 000/\$25 000–\$35 000 \$35 000–\$50 000/over \$50 000 Unknown/refused
Race/ethnicity ^a	White but not of Hispanic origin Black but not of Hispanic origin Hispanic of any race Asian, Pacific Islander, or Native American/other/ unknown/refused

Note. Categories separated by a slash were differentiated during the interview but combined for this report.
^aCategories were defined on the basis of responses to the questions "What is your race?" and "Are you of Hispanic origin?"

TABLE 2—Risk Behaviors and Characteristics Available for Analysis in the Core Modules of the 1992 Behavioral Risk Factor Surveillance System Questionnaire

Module	Risk Behavior/Characteristic	Definition
Safety belts	Safety belt nonuse	Does not always use safety belt when driving or riding in a car
Hypertension	Recent blood pressure check Diagnosed hypertensive	Has had blood pressure measured in past 2 y Has been told on more than one occasion that blood pressure was high or is taking blood pressure medication
Cholesterol	Recent cholesterol check Cholesterol awareness Serum cholesterol Hypercholesterolemia	Has had cholesterol checked in past 5 y Can state own cholesterol level as a number Cholesterol level stated as a number Cholesterol level 200 mg/dL or higher
Exercise	Sedentary life-style Regular aerobic exercise	Did not exercise in past mo Performed an aerobic activity at least three times per wk for at least 20 min per occasion in past mo
Smoking	Current cigarette use Cigarette consumption Former cigarette use	Smoked 100 or more cigarettes over lifetime and currently smokes Number of cigarettes smoked daily Smoked 100 or more cigarettes over lifetime but does not currently smoke
Alcohol	Any consumption Total consumption Chronic consumption Acute consumption Drives after alcohol consumption	Any alcoholic drinks consumed in past mo Number of alcoholic drinks consumed in past mo Drank 60 or more alcoholic drinks in past mo Drank five or more alcoholic drinks on an occasion at least once in past mo Drove a car after consuming too much alcohol at least once in past mo
Obesity	Body mass index Obese Weight control	Weight in kilograms divided by square of height in meters Body mass index exceeding 27.8 kg/m ² (men) or 27.3 kg/m ² (women) Currently trying to lose weight
Diabetes	Diagnosed diabetes	Doctor-diagnosed diabetes

TABLE 3—Consistency of Distribution of Demographic Characteristics across Two Administrations of the Behavioral Risk Factor Surveillance System Questionnaire among 210 Respondents: Massachusetts, 1992

Demographic Characteristic	All Respondents (n = 210)		White Non-Hispanic (n = 75)		Black Non-Hispanic (n = 64)		Hispanic (n = 45)	
	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
Mean age, y (SD)	40.7 (16.7)	40.4 (16.7)	46.7 (17.5)	46.4 (17.4)	40.7 (17.4)	40.7 (17.4)	34.0 (11.8)	33.4 (12.2)
Male sex, %	41.4	41.4	56.0	56.0	35.9	35.9	31.1	31.1
Education, %								
Less than high school	23.9	20.9	14.7	14.7	20.6	15.9	44.4	39.5
High school graduate	32.1	33.5	30.7	28.0	33.3	42.9	33.3	34.9
Some college	17.2	19.9	14.7	16.0	27.0	27.0	6.7	11.6
College graduate/professional	26.8	25.7	40.0	41.3	19.0	14.3	15.6	14.0
Marital status, %								
Married/live as couple	45.2	45.2	54.7	56.0	32.8	32.8	53.3	48.9
Divorced/separated	15.7	15.2	10.7	12.0	17.2	18.7	17.8	15.6
Widowed	5.2	5.2	8.0	6.7	6.3	6.3	2.2	4.4
Never married	33.8	34.3	26.7	25.3	43.7	42.2	26.7	31.1
Employment, %								
Employed	56.7	56.7	60.0	58.7	51.6	54.7	60.0	55.6
Self-employed	5.2	5.2	6.7	6.7	3.1	1.6	0.0	2.2
Unemployed	14.8	15.7	6.7	9.3	23.4	18.7	17.8	20.0
Homemaker	7.6	7.1	4.0	5.3	7.8	7.8	15.6	13.3
Retired	11.4	11.4	20.0	17.3	12.5	15.6	0.0	0.0
Student	4.3	3.8	2.7	2.7	1.6	1.6	6.7	8.9
Income, %								
<\$15 000	30.9	32.9	20.0	20.0	26.6	37.5	48.9	40.0
\$15 000–\$35 000	32.9	30.9	30.7	29.3	39.1	31.3	26.7	31.1
>\$35 000	29.5	30.5	45.3	48.0	23.4	25.0	17.8	17.8
Unknown/refused	6.7	5.7	4.0	2.7	10.9	6.3	6.7	11.1
Race/ethnicity, %								
White	39.1	37.6
Black	31.4	32.4
Hispanic	24.3	24.8
Other	5.2	5.2

Note. The total for all respondents is greater than the sum of White non-Hispanic, Black non-Hispanic, and Hispanic respondents as a result of exclusion of 26 subjects from ethnicity-specific analyses (see text for details). None of the differences between time 1 and time 2 values were significant ($P > .05$) within the whole sample or each ethnicity separately. The consistency of distribution of race/ethnicity was not compared for ethnicity-specific strata; it should be noted that the distribution of race/ethnicity in this sample does not reflect the ethnic composition of Massachusetts because of oversampling of Black and Hispanic residents.

ther administration of the interview for demographic characteristics or continuous variables and those for whom the continuous variable was not asked (e.g., number of cigarettes smoked among nonsmokers) were excluded from tests of individual consistency.

Results

Differences across administrations for demographic characteristics and risk factors were generally small (Tables 3 and 4). At the individual level, employment and income were reported less consistently than the other demographic variables (Table 5). Black and Hispanic respondents tended to report educational attainment, employment status, and income less consistently than did White respondents. Reliability was generally high for behavioral risk factors (Table 6), with

most coefficients above .70 and many above .80. Consistency tended to be lower among Hispanic respondents than among White or Black respondents, although the differences were not marked.

Discussion

This paper suggests good test-retest reliability of the BRFSS questionnaire. Our data are consistent with the results of an earlier, more restricted study,⁴ and extend these results to an analysis of the complete core BRFSS questionnaire in a statewide sample. Two studies have shown fair to good consistency of group mean prevalence of selected risk factors,^{5,6} but no published studies have addressed the validity of individual responses. Our study should be relevant to all agencies participating in the BRFSS, because we adhered to BRFSS subject re-

cruitment and data collection protocols. Our results concerning individual-level reliability are of potential interest to epidemiologic studies involving telephone-based surveys.⁷

True change between administrations of the questionnaire leads to lower estimates of reliability. Conversely, if the time period between administrations is too short, the questionnaires cease to be independent. In the present study, the time interval between the two administrations was 21 to 44 days. For most behaviors and characteristics, this interval was probably adequate. Questions concerning alcohol consumption and exercise, which addressed behaviors occurring in the previous month, showed the lowest consistency. For these variables, under the conditions of our study, it is not possible to separate questionnaire effects from true change. The sample size was too small to

TABLE 4—Consistency of Distribution of Risk Behaviors and Characteristics across Two Administrations of the Behavioral Risk Factor Surveillance System Questionnaire among 210 Respondents: Massachusetts, 1992

Module	Risk Behavior/Characteristic	All Respondents (n = 210)		White Non-Hispanic (n = 75)		Black Non-Hispanic (n = 64)		Hispanic (n = 45)	
		Time 1	Time 2	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
Safety belts	Safety belt nonuse, %	60.5	61.4	52.0	58.7	62.5	57.8	66.7	68.9
Hypertension	Recent blood pressure check, %	96.7	95.7	96.0	94.7	96.9	96.9	95.6	93.3
	Diagnosed hypertension, %	16.7	13.8	16.0	16.0	26.6	21.9	13.3	6.7
Cholesterol	Recent cholesterol check, %	69.1	68.6	73.3	74.7	60.9	65.6	68.9	57.8
	Cholesterol awareness, no. (%)	50 (23.8)	49 (23.3)	26 (34.7)	30 (40.0)	5 (7.8)	4 (6.3)	9 (20.0)	6 (13.3)
	Cholesterol level (among those aware of level), mean (SD)	192.5 (41.5)	192.0 (59.3)	194.5 (42.6)	206.0 (65.0)	194.2 (60.7)	143.5 (58.3)	198.9 (41.9)	170.2 (41.3)
	Hypercholesterolemia, %	15.7	16.7	21.3	22.7	12.5	12.5	8.9	8.9
Exercise	Sedentary, %	30.5	28.1	25.3	21.3	34.4	31.3	40.0	40.0
	Regular aerobic exercise, %	10.5	8.6	14.7	13.3	7.8	6.3	8.9	4.4
Smoking	Current smoker, no. (%)	54 (28.6)	58 (29.1)	19 (28.0)	18 (24.0)	13 (21.9)	17 (26.6)	13 (35.6)	12 (33.3)
	Number of cigarettes per d (current smokers only), mean (SD)	17.7 (10.9)	14.8 (11.3)	21.0 (10.5)	18.3 (10.1)	9.9 (6.4)	8.4 (6.0)	19.3 (14.0)	16.0 (14.5)
	Former smoker, %	29.5	28.1	38.7	45.3	28.1	21.9	22.2	17.8
Alcohol	Alcohol consumption, no. (%)	116 (55.2)	109 (51.9)	56 (74.7)	54 (72.0)	24 (37.5)	24 (37.5)	22 (48.9)	18 (40.0)
	Number of drinks per mo (drinkers only), mean (SD)	17.4 (20.0)	16.2 (19.3)	20.4 (22.8)	17.2 (20.3)	14.1 (17.1)	15.3 (20.3)	11.9 (12.4)	13.3 (11.7)
	Chronic consumption, %	2.9	2.4	5.3	4.0	0.0	1.6	0.0	0.0
	Acute consumption, %	14.8	16.7	20.0	20.0	9.4	10.9	8.9	13.3
	Drink and drive, %	1.9	0.5	2.7	0.0	0.0	1.6	2.2	0.0
Obesity	Body mass index, no. reporting	200	204	74	73	61	63	40	42
	Body mass index, mean (SD)	25.9 (4.8)	26.2 (4.9)	25.6 (4.0)	25.6 (3.8)	27.9 (6.1)	27.9 (6.0)	24.7 (3.7)	26.2 (4.6)
	Obese, %	27.6	30.5	26.7	24.0	42.2	40.6	20.0	33.3
	Trying to lose weight, %	47.6	39.5	48.0	42.7	51.6	40.6	44.4	37.8
Diabetes	Diagnosed diabetes, %	6.2	5.2	5.3	4.0	10.9	10.9	4.4	2.2

Note. The total for all respondents is greater than the sum of White non-Hispanic, Black non-Hispanic, and Hispanic respondents as a result of exclusion of 26 subjects from ethnicity-specific analyses (see text for details). None of the differences between time 1 and time 2 values were significant ($P > .05$) within the whole sample or each ethnicity separately.

assess the effect on estimates of reliability of varying intervals between the two administrations.

Participation in the first interview may have prompted some respondents to modify their life-style, but this did not appear to be the case, inasmuch as there was no overall improvement in the health profile of the sample. A related issue is that if the population is homogenous with respect to the characteristic being studied, estimates of reliability become highly sensitive to small departures from perfect concordance.⁸ This may explain the lower estimates of reliability for recent blood pressure check (prevalence at first administration of 96.7%), chronic alcohol consumption (prevalence of 2.9%), and driving after alcohol consumption (prevalence of 1.9%). For the race/ethnicity-specific analyses, instability due to extreme mar-

TABLE 5—Consistency of Individual Responses to Two Administrations of the Demographics Module of the Behavioral Risk Factor Surveillance System Questionnaire among 210 Respondents: Massachusetts, 1992

Demographic Characteristic	Reliability			
	All Respondents (n = 210)	White Non-Hispanic (n = 75)	Black Non-Hispanic (n = 64)	Hispanic (n = 45)
Age	1.00	1.00	1.00	.97
Sex	1.00	1.00	1.00	1.00
Education	0.80	0.89	0.76	0.62
Marital status	0.93	0.95	0.98	0.86
Employment	0.77	0.93	0.71	0.63
Income	0.75	0.82	0.65	0.67
Race/ethnicity	0.87

Note. Age was retained as a continuous variable. Pearson's correlation coefficient was used to estimate reliability of age. Kappa was used to estimate reliability of categorical variables. In instances in which a demographic characteristic included more than two categories, unweighted kappa was computed. See Table 1 for the number of categories used in computing kappa. All reliability coefficients were highly significant ($P < .001$). All analyses were based on the complete sample, except for age among all respondents (n = 208) and White respondents (n = 74) and education among all respondents (n = 206), Black respondents (n = 63), and Hispanic respondents (n = 43). Kappa was not computed for consistency of race/ethnicity classification within ethnicity-specific subgroups.

TABLE 6—Consistency of Individual Responses to Two Administrations of the Behavioral Risk Modules of the Behavioral Risk Factor Surveillance System Questionnaire among 210 Respondents: Massachusetts, 1992

Module	Risk Behavior/Characteristic	Reliability			
		All Respondents (n = 210)	White Non-Hispanic (n = 75)	Black Non-Hispanic (n = 64)	Hispanic (n = 45)
Safety belt	Safety belt nonuse	0.76	0.81	0.77	0.75
Hypertension	Recent blood pressure check	0.61	0.85	-0.03	0.79
	Diagnosed hypertension	0.89	1.00	0.87	0.63
Cholesterol	Recent cholesterol check	0.77	0.83	0.83	0.57
	Hypercholesterolemia	0.86	0.88	0.86	0.73
	Cholesterol awareness	0.80	0.83	0.64	0.60
	Actual cholesterol level	0.88 (n = 42)	0.85 (n = 25)	1.00 (n = 3)	0.96 (n = 5)
Exercise	Sedentary life-style	0.57	0.59	0.50	0.63
	Regular aerobic exercise	0.45	0.61	-0.07	0.64
Smoking	Current smoker	0.83	0.90	0.79	0.85
	Number of cigarettes per d	0.73 (n = 49)	0.60 (n = 18)	0.83 (n = 13)	0.70 (n = 10)
	Former smoker	0.73	0.86	0.67	0.58
Alcohol	Any alcohol consumption	0.82	0.93	0.73	0.73
	Number of drinks per mo	0.72 (n = 103)	0.79 (n = 54)	0.57 (n = 20)	0.60 (n = 17)
	Chronic consumption	0.53	0.55	...a	...a
	Binge consumption	0.64	0.67	0.57	0.55
	Drink and drive	-0.01	...a	...a	...a
Obesity	Body mass index	0.90 (n = 198)	0.97 (n = 73)	0.91 (n = 61)	0.59 (n = 39)
	Obese	0.77	0.86	0.84	0.55
	Weight control	0.72	0.73	0.72	0.68
Diabetes	Diagnosed diabetes	0.82	0.85	1.00	-0.03

Note. Kappa was used to estimate reliability for dichotomous variables. Pearson's correlation coefficient was used to estimate reliability for continuous variables (actual cholesterol level, number of cigarettes per day, number of drinks per month, body mass index). Analyses were performed on data from all respondents except where specified. For continuous variables, reliability was computed only for respondents with nonzero, nonmissing values at both administrations. All kappa coefficients were highly significant ($P < .001$), except for recent blood pressure check and regular aerobic exercise among Black respondents, driving after alcohol consumption in the whole sample, and diagnosed diabetes among Hispanic respondents ($P > .05$ for each). It should be noted that the low values of kappa reflect extreme marginal prevalence of the risk behaviors rather than a large proportion of discordant cells. All correlation coefficients were significant ($P < .001$), except cholesterol level among Black ($P = .014$) and Hispanic ($P < .01$) respondents, number of cigarettes among White ($P < .01$) and Hispanic ($P = .024$) respondents, and number of drinks per month among Black ($P < .01$) and Hispanic ($P < .01$) respondents.

^aPrevalence of zero at either time 1 or time 2, kappa not computed.

ginal prevalence was compounded by the smaller sample sizes, leading in some instances to uninterpretable negative values for the kappa statistic.

A limitation of this study is the low response rate to the initial interviews. The response rate for the main sample was consistent with that obtained in Massachusetts over the previous 3 years,⁹ al-

though it was lower than that of most other states participating in the BRFSS (C. Leutzinger, written communication, June 1992). For the minority study, the response rate to the initial interview was 39%; therefore, estimates of prevalence derived from this study may be subject to sampling bias. The response rate for the second interview was higher: 62% for the

minority sample and 70% for the main sample. Thus, we are able to state with fair confidence that, among responders, consistent responses may be obtained for most parameters. We cannot assess whether reliability differs between responders and nonresponders.

In conclusion, the BRFSS questionnaire yields highly consistent group mean estimates of prevalence when administered repeatedly to the same individuals. Individual reliability is also high. This study, together with previous studies of reliability and validity,⁴⁻⁶ supports the use of the BRFSS questionnaire in surveillance activities. □

Acknowledgments

This study was funded through a cooperative agreement between the Centers for Disease Control and Prevention, the Massachusetts Department of Health, and the Massachusetts Health Research Institute.

References

- Remington PL, Smith MY, Williamson DF, Anda RF, Gentry EM, Hogelin GC. Design, characteristics and usefulness of state-based behavioral risk factor surveillance. *Public Health Rep.* 1988;103:366-375.
- Healthy People 2000: National Health Promotion and Disease Prevention Objectives.* Washington, DC: US Dept of Health and Human Services; 1990. DHHS publication PHS 91-50212.
- Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960; 20:37-46.
- Shea S, Stein AD, Lantigua R, Basch CE. Reliability of the Behavioral Risk Factor Survey in a triethnic population. *Am J Epidemiol.* 1991;133:489-500.
- Anda RF, Dodson DL, Williamson DF, Remington PL. Health promotion data for state health departments: telephone versus in-person estimates of smoking and alcohol use. *Am J Health Promotion.* 1989;4:32-36.
- Jackson C, Jatulis DE, Fortmann SP. The Behavioral Risk Factor Survey and the Stanford Five-City Project Survey: a comparison of cardiovascular risk behavior estimates. *Am J Public Health.* 1992;82:412-416.
- Marcus AC, Crane LA. Telephone surveys in public health research. *Med Care.* 1986; 24:97-112.
- Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol.* 1987;126:161-169.
- Stein AD. *Behavioral risk factors and preventive health measures—Massachusetts Behavioral Risk Factor Surveys 1986-90.* Boston, Mass: Massachusetts Dept of Public Health; 1992.