

Measuring injury risk factors: question reliability in a statewide sample

Jane Koziol-McLain, David Brand, Daniel Morgan, Marilyn Leff, Steven R Lowenstein

Abstract

Background—Recently (1996–98), Colorado added 15 questions pertaining to injury related risks and behaviors to the behavioral risk factor surveillance system (BRFSS). Questions addressed bicycle helmet use, traffic crashes, exposure to violence, suicidal behavior, and gun storage.

Objective—To measure the test-retest reliability of these injury related questions. **Methods**—Of 330 BRFSS participants, 229 (69%) were called a second time and reasked nine selected injury questions. Retests were completed 7–28 days after the original interview.

Results—Test-retest agreement was very high ($\kappa > 0.80$) for bicycle helmet use, domestic police visits, and gun ownership. All other injury risk questions had substantial agreement ($\kappa > 0.60$).

Conclusions—The injury related questions added to the Colorado BRFSS have high test-retest reliability.

(*Injury Prevention* 2000;6:148–150)

Keywords: risk factors; reproducibility of results; health surveys; data collection/methods

Personal habits and lifestyles play an important part in causing injury, disability, and premature death. Yet, with a few exceptions, injury related risk factors and behaviors are omitted from surveillance systems. To remedy this, from 1996 through 1998 Colorado added questions pertaining to injury related risks and behaviors to its behavioral risk factor surveillance system (BRFSS). The injury module included 15 questions about bicycle helmet use, traffic crashes, exposure to violence, suicidal behavior, and gun storage.

The BRFSS, sponsored by the Centers for Disease Control and Prevention, is a population based random digit dial telephone survey that has been conducted by 50 states since 1993. Approximately 150 Colorado residents aged 18 years and over are surveyed each month throughout the year. Although reliability testing has been performed on core questions of the BRFSS,^{1–4} state added questions, including those that address injury risk factors, have not typically been subjected to the same rigor.

This study was conducted to measure the test-retest reliability of the state added questions. Test-retest reliability is an assessment of the stability of a measure over time—that is, the extent to which people answer questions consistently at different times.^{5–8} Reliability is a necessary survey characteristic to ensure that

the data are useful for surveillance, monitoring trends, and intervening to prevent injuries.

Methods

During April and May of 1998, BRFSS respondents were called a second time and reasked selected injury questions. Trained interviewers conducted both initial and recall interviews. Standard BRFSS survey methods were employed, including computer assisted telephone interviewing to facilitate direct data entry and coding, interviewer monitoring, and quality control. Recalls were completed one to four weeks after the initial BRFSS interview. Up to 15 calls in three different calling periods were placed to speak with the original respondent. Calls were conducted in either English or Spanish.

The injury control module included 15 questions adapted from published surveys that have not previously been evaluated for reliability. Seven questions were asked of all respondents and eight were asked of only a subset of respondents based on their previous answers. The seven questions that were asked of all respondents were included in the retest module. In addition, two questions that were asked of only a subset of respondents were included in the retest module because the previous month's BRFSS data demonstrated a prevalence of at least 3%.

Data analysis proceeded by first assessing differences between respondents who were successfully recalled for retest and those who could not be recalled. The χ^2 (for age, sex, marital status, and Hispanic origin) and Kruskal-Wallis (for age) test statistics were used to test for differences. Then, the reliability of injury question responses was calculated using the κ and weighted κ statistics. Measuring κ is preferable to measuring "per cent agreement," as κ measures the agreement that occurs beyond what would be expected by chance alone.^{9–12} Landis and Koch provide the following benchmarks for the interpretation of κ : 0.4–0.6 = moderate; 0.61–0.80 = substantial; and 0.81–1.0 = almost perfect.¹⁰ Data were analyzed using the SAS statistical package (SAS, North Carolina). Simple κ statistics were used to measure agreement for variables with dichotomous response sets, and the weighted κ was used for variables with ordinal response sets. A two month cohort study sample was chosen to allow precise estimates of the κ statistic (± 0.13). Ninety five per cent confidence intervals were calculated for the simple and weighted κ statistics.

School of Nursing,
Johns Hopkins
University, 525 N Wolfe
Street, Room 306,
Baltimore, MD
21205–2110, USA
J Koziol-McLain

Colorado Department
of Public Health and
Environment, Denver,
Colorado
D Brand
D Morgan
M Leff

Departments of
Emergency Medicine,
Preventive Medicine
and Biometrics,
School of Medicine,
University of Colorado
Health Sciences
Center, Denver,
Colorado
S R Lowenstein

Correspondence to:
Dr Koziol-McLain
(e-mail:
jkoziol-mclain@son.jhmi.edu)

Table 1 Demographic characteristics

	Recall interview		p Value
	No (%) contacted	No (%) unable to contact	
Total	229 (69)	101 (31)	
Sex (male)	100 (44)	52 (51)	0.189
Marital status			0.235
Married	126 (55)	52 (51)	
Divorced	31 (14)	12 (12)	
Widowed	26 (11)	6 (6)	
Separated	6 (3)	2 (2)	
Never married	35 (15)	25 (25)	
Member of unmarried couple	5 (2)	4 (4)	
Hispanic origin (yes)	27 (12)	18 (18)	0.132
Age (mean, years)	45.2	42.2	0.139

Results

Of the initial 330 BRFSS interviews conducted during the two month study period, 229 (69%) were successfully contacted for retesting. The time between the initial and second call varied: 34% (n=78) were called the second week, 45% (n=104) the third week, and 21% (n=47) the fourth week. Persons who were recontacted were similar to those who could not be recontacted with respect to sex, age, marital status, and Hispanic origin (see table 1). Initial interview injury risk factor prevalence rates did not differ between those recontacted and those not contacted.

Despite varied injury risk factor prevalence rates, test-retest reliability was high for all injury questions (see table 2). Three questions (bicycle helmet use, domestic police visits, and gun ownership) had κ values that exceeded 0.80, considered "almost perfect" agreement by some authors.¹⁰ Test-retest agreement was also high ($\kappa > 0.60$) for the remaining six injury questions.

Discussion

Most large behavioral risk factor surveys, including the national BRFSS, focus on chronic diseases. Injury prone behaviors and risk factors should receive greater emphasis.

Our findings demonstrate that the injury related questions added to the Colorado BRFSS are highly reliable.

One important limitation of this study is that it is more difficult to assess agreement beyond chance when prevalence is low.¹³ Thus, the precision in this study varied among the questions.

There are two additional important limitations to consider when testing reliability by the test-retest method.⁵⁻⁸ First, real change could have occurred between the testing occasions that caused participants to answer differently. For example, improving weather conditions may have made riding a bicycle more prevalent in the recall interview. Perhaps it is not surprising that the question with the lowest κ statistic (0.66) asked women about "feeling unsafe now"—a state that could easily change over short time intervals.

Second, persons may respond on the retest based on their memory of how they responded on the first test, leading to an over estimate of retest reliability. In this occasion, testing "memory" was thought to be less likely, given that the initial BRFSS interview included over 150 questions. However, by administering only a portion of the original 150 question survey, test-retest reliability may have been affected in other ways. Finally, although this study supports the test-retest reliability of the injury questions, tests of validity are still needed.

Reducing high risk behaviors is a priority of the national health objectives for the year 2010 and a cornerstone of state injury control strategic plans.¹⁴ Therefore, information about the epidemiology of common injury prone behaviors is needed. This study supports the reliability of questions to measure and monitor the prevalence of injury prone behaviors.

This study was supported by a grant from the Centers for Disease Control and Prevention (R49/CCR811509). During the project period Dr Koziol-McLain was supported by a National Research Service Award from the National Institute of Mental Health (F31 MH11716).

Table 2 Agreement between test and retest administrations

Question	Response	Prevalence (%)			95% CI
		Initial	Retest	Agreement	
During the past month, how many times did you ride a bicycle? (n=229)	None	80.5	75.7	$\kappa=0.67$	0.56 to 0.78
	1-10	16.8	22.6		
	11-20	1.8	1.3		
	≥ 21	0.9	0.4		
When you rode a bicycle during the past month, how often did you wear a helmet? (among those who rode, n=39)	Always	33.3	35.9	$\kappa=0.89$	0.77 to 1.0
	Nearly always	7.7	7.7		
	Sometimes	0	2.6		
	Seldom	7.7	0		
	Never	51.3	53.8		
During the past year, how many times were you in a traffic accident in which you were injured and had to be treated by a doctor, nurse, chiropractor, dentist, or other medical person? (n=228)	One or more	2.2	2.2	$\kappa=0.80$	0.52 to 1.0
In the past year, have you ever seriously thought about trying to hurt yourself in a way that might have resulted in your death? (n=228)	Yes	5.3	3.5	$\kappa=0.79$	0.59 to 0.99
Within the past year, have you been hit, slapped, kicked, raped, or otherwise physically hurt by someone? (n=227)	Yes	4.4	4.4	$\kappa=0.79$	0.59 to 0.99
Considering your current partners or friends, or any past partners or friends, is there anyone who is making you feel unsafe now? (n=225)	Yes	1.8	2.2	$\kappa=0.66$	0.3 to 1.0
In the past year, have the police ever been called to your home because of a fight or argument, no matter who was fighting or who was at fault? (n=226)	Yes	4.4	5.8	$\kappa=0.86$	0.71 to 1.0
Are there any working revolvers, pistols, or handguns in your home? (n=223)	Yes	25.1	24.7	$\kappa=0.89$	0.82 to 0.96
Are any of them loaded? (among those with guns, n=51)	Yes	25.5	29.4	$\kappa=0.80$	0.62 to 0.99

CI=confidence interval; w=weighted.

- 1 Brownson RC, Jackson-Thompson J, Wilerson JC, et al. Reliability of information on chronic disease risk factors collected in the Missouri behavioral risk factor surveillance system. *Epidemiology* 1994;5:561-71.
- 2 Shea S, Stein AD, Lantigua R, et al. Reliability of the behavioral risk factor survey in a triethnic population. *Am J Epidemiol* 1991;133:489-500.
- 3 Stein AD, Courval JM, Lederman RI, et al. Reproducibility of responses to telephone interview: demographic predictors of discordance in risk factor status. *Am J Epidemiol* 1995;141:1097-105.
- 4 Stein AD, Lederman RI, Shea S. The behavioral risk factor surveillance system questionnaire: its reliability in a statewide sample. *Am J Public Health* 1993;83:1768-72.
- 5 Aday LA. Designing and conducting health surveys: a comprehensive guide. San Francisco: Lossey-Bass, 1989: 41-4.
- 6 Carmines EG, Zeller RA. *Reliability and validity assessment*. Newbury Park: Sage, 1979: 37-51. (Publication series/ number 07-017.)
- 7 Cook TD, Campbell DT. *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNally College Publishing, 1979: 37-94.
- 8 Nunnally JC. *Psychometric theory*. 2nd Ed. New York: McGraw Hill, 1978: 233-34.
- 9 Fleiss JL. *Statistical methods for rates and proportions*. New York: John Wiley, 1973: 143-53.
- 10 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- 11 Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol* 1987;126:161-9.
- 12 Rosner B. *Fundamentals of biostatistics*. 4th Ed. Belmont, CA: Duxbury Press, 1995: 423-6.
- 13 Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990;43:543-9.
- 14 US Department of Health and Human Services. *Developing objectives for healthy people 2010*. Washington, DC: US Department of Health and Human Services, 1997.

Intentional or unintentional?

Although our policy is to avoid papers dealing with child abuse, it is important to acknowledge that a large proportion of injuries in young children is intentionally inflicted, usually by parents. Two papers in the January 2000 issue of *Archives of Pediatric and Adolescent Medicine* provide interesting estimates. In one, by Reece and Sege, 19% of 287 children ages 1 week to 6.5 years were classified as definite abuse. Many presented with head injuries. The second study, based on nearly 2000 records in the National Pediatric Trauma Registry in the US found that abuse accounted for 10.6% of cases of blunt trauma (*Arch Pediatr Adolesc Med* 2000;154:9-22).

Short children are bullied more often

“This report suggests that short children are more likely to be bullied than their taller peers. More short pupils also report a degree of social isolation—the result, or possibly even the cause, of their victimization”. So concludes an impressive study conducted in Southampton involving 92 short normal adolescents and 117 controls. The finding is hardly surprising and is part of popular wisdom. But it is valuable to add a scientific element to a problem that is likely to be a major component of violent activities among school children (*BMJ* 2000;320:612-3).

US trends in gun deaths

A report in the *Boston Globe* notes that between 1993 and 1997 there has been a 21% decline in gun deaths in the US, reaching the lowest level in more than 30 years. Factors responsible for this welcome news include tougher gun control laws, the economy, better policing, and gun safety courses. The author of the report J Lee Annett, a statistician with CDC, stated, “This progress is really encouraging and really says that joint prevention efforts of public health officials, legislators and law enforcement should continue”. Needless to say, the National Rifle Association said the numbers prove that more gun laws are not needed. Another commentator calls attention to enforcement: “Police were not treating guns in a preventive sense prior to 1993 and now they are”. The economy is credited with allowing governments to spend more on services that prevent gun violence such as domestic violence shelters and youth recreation programs. Still, on average 265 people a day were shot in 1997. A long way to go (*Associated Press, Boston Globe*, 19 November 1999).

Man watches as train kills two granddaughters

In Ingersoll, Ontario, a man watched a train slam into a car carrying his two granddaughters on Christmas Day, killing them and a friend. He claims town officials have failed to install proper warning signals at the rail crossing. The grandparents said they repeatedly complained to the town council about the lack of warning lights on gates at the crossing. “But they said it wouldn’t happen for another 10 years, if it happened at all, because it would increase the tax burden”, said the grandfather (*Canadian Press*, 27 December 1999).