JOURNAL OF ADOLESCENT HEALTH 2003;33:436–457

REVIEW ARTICLE

Assessment of Factors Affecting the Validity of Self-Reported Health-Risk Behavior Among Adolescents: Evidence From the Scientific Literature

NANCY D. BRENER, Ph.D., JOHN O. G. BILLY, Ph.D., AND WILLIAM R. GRADY, M.A.

Abstract: **We reviewed the existing empirical literature to assess cognitive and situational factors that may affect the validity of adolescents' self-reports of alcohol and other drug use, tobacco use, behaviors related to unintentional injuries and violence, dietary behaviors, physical activity, and sexual behavior. Specifically, we searched for peer-reviewed journal articles published in 1980 or later that examined the factors affecting self-report of the six categories of behavior listed above. We also searched for studies describing objective measures for each behavior. Self-reports of each of six types of health-risk behaviors are affected by both cognitive and situational factors. These factors, however, do not threaten the validity of self-reports of each type of behavior equally. The importance of assessing health-risk behaviors as part of research activities involving adolescents necessitates the use of self-report measures. Researchers should familiarize themselves with the threats to validity inherent in this type of assessment and design research that minimizes these threats as much as possible.**

KEY WORDS:
Adolescence
Alcohol
Dietary behavior
Drug use
Health risk behavior
Injuries
Physical activity
Psychometrics
Questionnaires
Self assessment
Sexual behavior
Tobacco use
Violence

Health-risk behaviors such as cigarette smoking, weapon-carrying, and unprotected sexual intercourse contribute to the leading causes of morbidity, mortality, and social problems among adolescents. Consequently, many reasons exist for collecting data on these and other health-risk behaviors. For example, policymakers and program directors use data on the prevalence of these behaviors to monitor trends, set program goals, identify target populations, seek funding, and advocate for support. Assessment of these behaviors also is a critical component of research that examines associations between health-risk behaviors and other factors, builds theories of behavioral change, develops policies and programs designed to prevent these behaviors, and evaluates these policies and programs.

Health-risk behaviors usually are measured among adolescents by administering questionnaires that require retrospective self-reports about engaging in these behaviors. The truthfulness and accuracy of these self-reports may be compromised because some health-risk behaviors are difficult to recall and some are so sensitive that respondents may not want to report them. In addition, adolescents may purposely underreport or overreport some health-risk behaviors because they believe engaging in these behaviors is socially undesirable or desirable, respectively.

From the Division of Adolescent and School Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, Georgia (N.D.B.); and Battelle, Centers for Public Health Research and Evaluation, Seattle, Washington (J.O.G.B., W.R.G).

Address correspondence and reprint requests to: Dr. Nancy D. Brener, Division of Adolescent and School Health, CDC, Mailstop K-33, 4770 Buford Highway NE, Atlanta, GA 30341. E-mail: nad1@cdc.gov.

Manuscript accepted November 25, 2002.

1054-139X/03/$–see front matter
doi:10.1016/S1054-139X(03)00052-1

Published by Elsevier Inc., 360 Park Avenue South, New York, NY 10010

Most of the data provided by self-reports cannot be verified independently in a cost-effective, feasible, and ethical manner. Furthermore, for a researcher or policymaker, determining the accuracy of self-reported data is not sufficient. One also must know both the magnitude of the inaccuracy and its likely sources. Such information makes the data more useful in that likely biases can be taken into account when survey results are interpreted and applied.

Although some studies have examined the validity of self-reported data for particular behaviors among adolescents [1–3], no study has done so across a wide range of behaviors. Consequently, our goal is to review the existing empirical literature to better assess those factors that may affect the validity of adolescents' reports of several health-risk behaviors.

Factors Affecting Validity

Two major theoretical perspectives have been advanced to explain the source of validity problems that may emerge with some self-reported data. The cognitive perspective focuses on the mental processes underlying self-reported data and attributes validity problems to inaccuracies arising from comprehension, recall, and other cognitive operations. The situational perspective focuses on validity problems that arise from factors related to social desirability and interviewing conditions. These two perspectives are not mutually exclusive. For example, basic cognitive models of the question-answering process have been expanded in an attempt to account for situational factors [4]. Nevertheless, for the purpose of discussion, it is useful to consider each perspective separately.

The cognitive perspective. To synthesize what is known about cognitive processes for answering questions, several basic models of the process have been proposed [4–7]. These models postulate that four basic cognitive processes influence the question-answering process: (a) comprehension; (b) retrieval; (c) decision-making; and (d) response generation. Respondents first engage in comprehension processes that determine how a question is interpreted and encoded in memory. Retrieval cues are generated on the basis of output from this process and then used to search memory in the subsequent retrieval stage. The adequacy of any retrieved information is evaluated during the decision-making stage of processing. If the retrieved information is deemed adequate for the purposes of answering the question, then response generation will ensue. If, on the other hand, this information is deemed inadequate, additional retrieval attempts will be made or strategies involving estimation or heuristics will be initiated. More complex models [6] include a second decision-making stage in which the adequacy of the retrieved information is evaluated according to other criteria such as consistency with beliefs and values.

It has been hypothesized that error potentially arises at each of these stages, which in turn contributes to validity problems. In addition, because the specific cognitive operations employed in responding to a question may differ depending on such factors as the length of the reference period and the type of response required (e.g., frequency of a behavior vs. simply whether the behavior occurred), validity can vary from question to question. The context of the question, including other questions on the questionnaire and the environment in which the questionnaire is administered, also can influence how various cognitive processes are executed. This process, in turn, can lead to systematic biases in responding.

The situational perspective. The situational perspective focuses on validity problems that arise from characteristics of the external environment instead of internal processing. Factors presumed to be especially influential include the presence of others while responding to questions and respondents' perceptions of the level of privacy or confidentiality that responses are afforded. Social desirability, which is the desire to provide others with a favorable impression of oneself [8], is one construct used to explain situational biases. Questions that are most likely to be influenced by a social-desirability bias have response options that "involve attributes considered desirable to have, activities considered desirable to engage in, or objects considered desirable to possess" [9].

Another construct related to social desirability that might account for response biases is the desire for attention. This factor is particularly likely to lead to response biases among adolescents, for whom some behaviors, such as alcohol use, drug use, and sexual behavior, are associated with status in certain settings [10,11].

A perceived lack of confidentiality, anonymity, or privacy within the situational context also could cause response biases because of a fear of reprisal. In particular, behaviors that are illegal, stigmatized, or laden with moral implications may be underreported because of this concern.

To assess the cognitive and situational factors that affect the measurement of health-risk behaviors among adolescents, we performed a literature search on this topic. We organized the review according to six categories of behavior: alcohol and other drug use, tobacco use, behaviors leading to unintentional injuries and violence, dietary behaviors, physical activity, and sexual behaviors. In this paper, we synthesize the results of this literature search by first describing, for each of the six categories of behavior, the cognitive and situational factors that affect reporting. We also examine available evidence that these factors actually have an impact on self-reported behavior by assessing whether approaches specifically designed to mitigate the impact of these cognitive and situational factors on the behavioral reports have the hypothesized effect on those reports. Finally, on the basis of the results of the review, we determine whether an objective measure exists for each behavior and assess the extent to which self-reports of the behavior approach that measure.

Methods

To find relevant articles, we searched Medline, ERIC, Sociological Abstracts, and PsycINFO using the following keywords: "validity," "reliability," "self-report," "self-assessment," "alcohol use," "drug use," "tobacco use," "injury," "violence," "suicide," "diet," "nutrition," "physical activity," and "sexual behavior." We limited the searches to adolescent populations. We also cross-referenced the reference sections of relevant articles already in our possession and obtained through the literature search. Specifically, we searched for peer-reviewed journal articles published in 1980 or later that examined the factors affecting self-report of these six categories of behavior. We also searched for studies that described objective measures for each behavior, which we defined as assessment methods other than self-report, such as biochemical tests or official records. We use the term "gold standard" to describe those objective measures that are widely accepted as being least subject to error or bias, such as certain biochemical tests.

Table 1 provides a list of the studies we reviewed, information about each study population, and a breakdown of the methodological approaches used in each study into cognitive factors, situational factors, and objective measures. Studies are listed by category of risk behavior. Those studies examining behaviors in more than one of these categories are listed separately in each relevant category.

Results

Alcohol and Other Drug Use

Cognitive factors. Reports of alcohol and other drug use usually are obtained for one or more reference periods (e.g., 1 month, 1 year) as well as for the individual's lifetime. Problems in the retrieval of the required information can occur because behaviors have to be both recalled and placed within the appropriate time period. The difficulty of this task is increased because respondents may be unable to remember events that occurred while they were under the influence of abused substances.

To the extent that the retrieval of information about past use of alcohol and other drugs is problematic for survey respondents, one would expect less accurate reporting of alcohol and other drug use when reference periods are long. Working under the assumption that higher prevalence rates are more accurate than lower prevalence rates in reports of substance use [12,13], evidence that shorter recall periods lead to more accurate reporting can be seen in studies that found proportionally higher prevalence rates of alcohol and other drug use for shorter periods [14]. For example, Bachman and O'Malley [14] found that reported 30-day use rates multiplied by 12 exceeds reported 12-month use rates. Similarly, questions assessing age of initiation of alcohol and other drug use tend to elicit inaccurate responses among adolescents, which is, at least in part, a function of forgetting over time [15–18].

Sources of error also include comprehension problems from unfamiliar terms and difficulty defining and using reference periods. For example, Johnston and O'Malley [19] found that "recanting" (i.e., denial of having ever used a drug after previously reporting use of that drug) was more common for tranquilizers and barbiturates than for marijuana and cocaine. The authors note that definitions of the former are probably less clear to respondents. In longitudinal studies, reports of ever having used a substance tend to be more reliable than reports of frequency of use during particular time periods [15,20]. In general, the more complex the recall task, the less reliable the reporting. This suggests that the quality of responses can be improved by using strategies designed to enhance recall, such as relatively short reference periods and simple language.

Situational factors. The use of alcohol by adolescents is not only illegal, but also is subject to social disapproval. The formal and informal sanctions associated with illegal drug use are even greater. Thus,

we would expect that self-reports of alcohol and illegal drug use would be subject to biases related to both social desirability and fear of reprisal.

One way to assess whether self-reports of alcohol and other drug use are affected by social desirability or fear of reprisal is to compare reports obtained under different modes of administration that have varying levels of privacy and anonymity. One such comparison is between estimates derived from interviewer-administered questionnaires (IAQs) and those obtained from self-administered questionnaires (SAQs). In several methodological studies, the greater privacy provided by the SAQ format produced higher reported rates of alcohol and other drug use [21,22]. Similarly, studies using computer-assisted self-interviewing (CASI) in households produced even higher reported rates of alcohol and other drug use than paper-and-pencil SAQs [13,23], although similar studies conducted in schools showed no such effect [24,25].

These mode effects may also differ according to the substance considered. For example, Turner et al [22] found the greatest mode effects for cocaine use, which carries the greatest legal sanctions, and the smallest mode effects for alcohol use, which is legal among adults. That study also found that adolescents' responses were more sensitive to differences in data collection method than those of adults. For example, although the overall difference in reported alcohol use by interview mode was negligible among adults, among adolescents, 36% more reported use during the past 30 days when the SAQ was used. Similarly, Wright et al [23] found that adolescents reported higher rates of alcohol and other drug use when using CASI than when using paper-and-pencil SAQs, whereas adults did not show such differences. Another study found gender differences in mode effects. Webb et al [26] found that adolescent girls attending a community health clinic reported a greater frequency of alcohol and marijuana use with CASI than with the SAQ, while the opposite was true for boys.

Additional evidence that adolescent self-reports of alcohol and other drug use are affected by privacy and confidentiality can be found in studies which report significantly higher substance use from surveys conducted in schools than in households [27–29]. It is assumed that higher prevalence estimates are more accurate, an assumption that suggests less privacy and confidentiality in household-based surveys lead to underreporting.

Another way to assess whether self-reports of alcohol and other drug use are affected by situational factors is to examine the effect of adopting a "bogus pipeline" approach, in which respondents are led to believe their true behavior can be detected even though it cannot [30,31]. If reports of alcohol and other drug use are higher when this approach is used, it indicates that some respondents are deliberately misreporting their behaviors under standard interviewing conditions.

Attempts to use a bogus pipeline approach to obtain more valid reports of alcohol use have not met with success. Several studies have examined the impact of using a saliva test as a bogus pipeline among adolescents, and none found a significant difference between rates of those who knew they were going to provide saliva samples and rates of those who did not [30,31]. This suggests that respondents were either providing valid responses or did not believe the biomarkers actually could capture past alcohol use.

Assessment of the validity of self-reports. Test-retest reliability, or the extent to which a question yields the same response when asked more than once, is a necessary precondition for validity. Reliability has been investigated in studies of both alcohol use [11,18,32,33] and other drug use [11,33,34]. In general, reliability levels are high for all self-reported measures considered. Although these studies provide partial indication that self-reported data also are valid, they do not provide an objective measure against which the self-reported data can be compared.

Biochemical measures often are considered the gold standard in validation studies because they are believed to be more objective and less susceptible to bias than other available techniques [35]. Laboratory measures of alcohol use, however, have severe limitations. At best, breath tests can capture alcohol use only within the 24 hours preceding the test, and blood tests are best for identifying very heavy use [36].

Obtaining an objective measure of illegal drug use is somewhat easier because of the greater availability of biochemical markers. Most studies of the validity of self-reported drug use employ urinalysis. Saliva, sweat, and hair also can be used to detect drug use, but these methods are not as technologically advanced as urine-based methods [37]. Even urinalysis is not fool-proof in assessing marijuana use among adolescents. At least two studies found a moderate percentage of false negatives for urinalysis when compared with self-reported marijuana use [38,39].

Table 1. Methodological Studies of Self-reported Adolescent Health-risk Behaviors

Study (Authors and Year)	Study Population	Methodological Approaches		
		Cognitive	Situational	Objective Measures
Alcohol and other drug use				
Akinci, Tarter, & Kirisci, 2001 [38]	n = 200 Ages 15–18 years			Biochemical validation (urinalysis)
Bachman & O'Malley, 1981 [14]	n = 16,654 Grade 12	Reference periods		
Bailey, Flewelling, & Rachal, 1992 [15]	n = 5770 Grades 6–8 at baseline	Consistency of responses in longitudinal study		
Beebe, Harrison, McCrae, et al., 1998 [24]	n = 368 Alternative school students		Mode effects (paper and pencil vs. CASI)	
Brener, Kann, McManus, et al., 2002 [33]	n = 4619 Grades 9–12			Test-retest reliability
Campanelli, Dielman, & Shope, 1987 [30]	n = 291 Grades 7–9		Bogus pipeline	
Engels, Knibbe, & Drop, 1997 [16]	n = 1063 Mean age at baseline = 12.4 years Students in the Netherlands	Consistency of responses in longitudinal study		
Fisher, Kupferman, & Lesser, 1992 [41]	n = 133 Grades 9–12 Students attending a school-based clinic			Random response technique
Gfroerer, Wright, & Kopstein, 1997 [27]	n = 4533, ages 12–17 years (household) n = 8843, grades 8, 10, and 12 (school)		School vs. household setting	
Hallfors, Khatapoush, Kadushin, et al., 2000 [25]	n = 2296 Grades 7, 9, and 11		Mode effects (paper and pencil vs. CASI)	
Johnson & Mott, 2001 [17]	n = 10,042 Mean age at assessment = 14 years (National Longitudinal Study of Youth)	Consistency of responses in longitudinal study		
Johnston & O'Malley, 1997 [19]	Follow-up with Grade 12 students n = 16,300, age 18 years n = 8,900, ages 19–32 years	Definition of terms Consistency of responses in longitudinal study		
Kann, Brener, Warren, et al., 2002 [28]	n = 3201 (household) n = 13,420 (school) Grades 9–12		School vs. household setting	
Murphy, Durako, Muenz, & Wilsen, 2000 [39]	n = 182 Ages 13–20 years			Biochemical validation (urinalysis)
Needle, McCubbin, Lorence, & Hochhauser, 1983 [32]	n = 155 Grades 7–11		School vs. household setting	Test-retest reliability
Needle, Jou, & Su, 1989 [34]	n = 439 Ages 15–17 years		Mode effects (SAQ with interviewer present vs. absent)	Test-retest reliability
O'Malley, Bachman, & Johnston, 1983 [20]	Two samples: n = 1294 and n = 1215 Grade 12 at baseline	Reference periods Consistency of responses in longitudinal study		

Table 1. Continued

Study (Authors and Year)	Study Population	Methodological Approaches		
		Cognitive	Situational	Objective Measures
Patrick, Cheadle, Thompson, et al., 1994 [35]	Meta-analysis Includes adolescent and adult studies			Biochemical validation
Poulin, MacNeil, & Mitic, 1993 [42]	n = 3491 Grades 7, 9, 10, and 12 Nova Scotia, Canada			Fictitious drug
Rootman & Smart, 1985 [29]	n = 4737 (school) n = 577 (household) Ages 12–19 years		School vs. household setting	
Schober, Fe Caces, Pergamit, & Branden, 1992 [21]	n = 11,607 Ages 14–21 years at baseline		Mode effects (IAQ vs. SAQ)	
Shillington & Clapp, 2000 [18]	n = not provided (children of females participating in National Longitudinal Survey of Youth) Ages 14–23 years	Consistency of responses in longitudinal study		
Turner, Lessler, & Devore, 1992 [22]	n = 4000 Ages 12 years and older		Mode effects (IAQ vs. SAQ)	
Webb, Zimet, Fortenberry, & Blythe, 1999 [26]	n = 388 Ages 13–20 years Students attending community adolescent health clinics		Mode effects (paper and pencil vs. CASI)	
Werch, Gorman, Marty, et al., 1987 [31]	n = 191 Grades 5–10		Bogus pipeline	
Winters, Stinchfield, Henly, & Schwartz, 1991 [11]	6 samples (drug clinic and school-based) n's = 48, 61, 151, 432, 454, and 682 Ages 12–18 years			Test-retest reliability
Wright, Aquilino, & Supple, 1998 [23]	n = 3169 Ages 12–34 years		Mode effects (paper and pencil vs. CASI)	
Tobacco use				
Aguinis, Pierce, & Quigley, 1993 [47]	Meta-analysis Includes adolescent and adult studies		Bogus pipeline	
Akers, Massey, Clarke, & Lauer, 1983 [49]	n = 2150 Grades 7–12		Bogus pipeline	Random response technique Biochemical validation
Bauman & Ennett, 1994 [52]	n = 1823 Ages 12–14 years			Biochemical validation (CO and cotinine)
Bauman, Koch, Bryan, et al., 1989 [50]	n = 1854 Ages 12–14 years			Biochemical validation (cotinine)
Biglan & Ary, 1985 [51]	n = 1123 Grades 7–12			Biochemical validation (CO)
Brener, Kann, McManus, et al., 2002 [33]	n = 4619 Grades 9–12			Test-retest reliability
Brittingham, Tourangeau, & Kay, 1998 [46]	n = 22000 Ages 12 years and older		Mode effects (SAQ vs. IAQ)	
Caraballo, Giovino, & Pechacek, 2004 [53]	n = 2107 Ages 12–17 years			Biochemical validation (cotinine)

Table 1. Continued

Study (Authors and Year)	Study Population	Methodological Approaches		
		Cognitive	Situational	Objective Measures
Engels, Knibbe, & Drop, 1997 [16]	n = 1063 Mean age at baseline = 12.4 years, Students in the Netherlands	Consistency of responses in longitudinal study		
Fisher, Kupferman, & Lesser, 1992 [41]	n = 133 Grades 9–12 Students attending a school-based clinic			Random response technique
Gfroerer, Wright, & Kopstein, 1997 [27]	n = 4533, ages 12–17 years (household) n = 8843, grades 8, 10, and 12 (school)		School vs. household setting	
Hedges & Jarvis, 1998 [45]	n = 573 (household) n = 417 (school) Ages 11–15 years		School vs. household setting	Biochemical validation (cotinine)
Johnson & Mott, 2001 [17]	n = 10,042 Mean age at assessment = 14 years (National Longitudinal Study of Youth)	Consistency of responses in longitudinal study		
Kann, Brener, Warren, et al., 2002 [28]	n = 3201 (household) n = 13,420 (school) Grades 9–12		School vs. household setting	
Martin & Newman, 1988 [3]	n = 1160 Grade 9			Random response technique Biochemical validation (CO)
Murray, O'Connell, Schmid, & Perry, 1987 [48]	n = 770 Grade 10		Bogus pipeline	Biochemical validation (CO)
Needle, McCubbin, Lorence, & Hochhauser, 1983 [32]	n = 155 Grades 7–11		School vs. household setting	Test-retest reliability
Needle, Jou, & Su, 1989 [34]	n = 439 Ages 15–17		Mode effects (SAQ with interviewer present vs. absent)	Test-retest reliability
O'Malley, Bachman, & Johnston, 1983 [20]	Two samples: n = 1294 and n = 1215 Grade 12 at baseline	Reference periods Consistency of responses in longitudinal study		
Patrick, Cheadle, Thompson, et al. 1994 [35]	Meta-analysis Includes adolescent and adult studies			Biochemical validation
Pokorski, Chen, & Bertholf, 1994 [44]	n = 496 Mean age = 19.3 years Navy recruits			Biochemical validation (urine cotinine)
Poulin, MacNeil, & Mitic, 1993 [42]	n = 3491 Grades 7, 9, 10, and 12 Nova Scotia, Canada			Fictitious drug
Rootman & Smart, 1985 [29]	n = 4737 (school) n = 577 (household) Ages 12–19 years		School vs. household setting	
Shillington & Clapp, 2000 [18]	n = not provided (children of females participating in National Longitudinal Survey of Youth) Ages 14–23 years	Consistency of responses in longitudinal study		

Table 1. Continued

Study (Authors and Year)	Study Population	Methodological Approaches		
		Cognitive	Situational	Objective Measures
Stanton, McClelland, Elwood, et al., 1996 [43]	n = 937 Ages 15 and 18 years	Consistency of responses in longitudinal study		Biochemical validation (cotinine)
Turner, Lessler, & Devore, 1992 [22]	n = 4000 Ages 12 years and older		Mode effects (IAQ vs. SAQ)	
Walsh, Ellison, Hilton, et al., 2000 [54]	n = 1226 High school baseball players			Biochemical validation (cotinine)
Webb, Zimet, Fortenberry, & Blythe, 1999 [26]	n = 388 Ages 13–20 years Students attending community adolescent health clinics		Mode effects (paper and pencil vs. CASI)	
Werch, Gorman, Marty, et al., 1987 [31]	n = 191 Grades 5–10		Bogus pipeline	
Wright, Aquilino, & Supple, 1998 [23]	n = 3169 Ages 12–34 years		Mode effects (paper and pencil vs. CASI)	
Injury-related behaviors				
Beebe, Harrison, McCrae, et al., 1998 [24]	n = 368 Alternative school students		Mode effects (paper and pencil vs. CASI)	
Brener, Kann, McManus, et al., 2002 [33]	n = 4619 Grades 9–12			Test-retest reliability
De Man & LeDuc, 1994 [58]	n = 111 Ages 12–18 years High school students in Canada			Personality measures
Hilton, Harris, & Rice, 1998 [2]	n = 182 Mean age = 15.3 years	Reference periods		
Kann, Brener, Warren, et al., 2002 [28]	n = 3201 (household) n = 13,420 (school) Grades 9–12		School vs. household setting	
Klimes-Dougan, 1998 [56]	n = 196 2 cohorts, mean ages = 13.9 and 17.7 years at assessment	Reference periods	Mode effects (paper and pencil vs. structured interview)	
Osman, Barrios, Panak, et al., 1994 [57]	n = 215 Ages 15–18 years		Social desirability	
Turner, Ku, Rogers, et al., 1998 [13]	n = 1690 Men ages 15–19 years		Mode effects (paper and pencil vs. audio-CASI)	
Velting, Rathus, & Asnis, 1998 [55]	n = 48 Ages 12–20 years		Mode effects (SAQ vs. semi-structured interview)	
Webb, Zimet, Fortenberry, & Blythe, 1999 [26]	n = 388 Ages 13–20 years Students attending community adolescent health clinics		Mode effects (paper and pencil vs. CASI)	
Nutrition				
Livingstone, Prentice, Coward, et al., 1992 [70]	n = 22 Ages 15 and 18 years			Biochemical validation (doubly labeled water technique)

Table 1. Continued

Study (Authors and Year)	Study Population	Methodological Approaches		
		Cognitive	Situational	Objective Measures
Brener, Kann, McManus, et al., 2002 [33]	n = 4619 Grades 9–12			Test-retest reliability
Cavadini, Decarli, Dirren, et al., 1999 [61]	n = 20 Ages 16–19 years	FFQ vs. diet history		
Field, Colditz, Fox, et al., 1998 [62]	n = 102 High school students	FFQ vs. dietary recall		
Frank, Nicklas, Webber, et al., 1992 [63]	n = 1108 Ages 12–17 years	FFQ vs. dietary recall		Test-retest reliability
French, Peterson, Story, et al., 1998 [66]	n = 43 Ages 13–17 years		Mode effects (SAQ vs. clinical interview)	
Kann, Brener, Warren, et al., 2002 [28]	n = 3201 (household) n = 13,420 (school) Grades 9–12		School vs. household setting	
Rockett, Wolf, & Colditz, 1995 [67]	n = 179 Ages 9–18 years			Test-retest reliability
Rockett, Breitenbach, Frazier, et al., 1997 [64]	n = 275 Ages 9–18 years	FFQ vs. dietary recall (interviewer-obtained)		
Rosen & Poplawski, 1987 [65]	n = 211 Grades 9–12	Weight control questionnaire vs. diary reports	Weight control questionnaire and diary reports vs. peer and parent reports	
Webb, Zimet, Fortenberry, & Blythe, 1999 [26]	n = 388 Ages 13–20 years Students attending community adolescent health clinics		Mode effects (paper and pencil vs. CASI)	
Physical activity				
Aaron, Kriska, Dearwater, et al., 1995 [1]	n = 100 Age 15–18 years			Test-retest reliability SAQ vs. fitness test scores Sports team rosters
Rogers, Reybrouck, Weymans, et al., 1994 [85]	n = 69 Ages 6–18 years Patients post-surgery			SAQ vs. physiologic measure (ventilatory threshold)
Sallis, Buono, Roby, et al., 1993 [72]	n = 102 Grades 5, 8, and 11			Test-retest reliability SAQ vs. physiologic measure (heart rate monitor)
Webb, Zimet, Fortenberry, & Blythe, 1999 [26]	n = 388 Ages 13–20 years Students attending community adolescent health clinics		Mode effects (paper and pencil vs. CASI)	
Weston, Petosa, & Pate, 1997 [75]	n = 119 Grades 7–12			Test-retest reliability SAQ vs. pedometer, accelerometer, and heart rate monitor
Sexual behavior				
Alexander, Somerfield, Ensminger, et al., 1993 [10]	n = 758 Grade 8 at baseline	Consistency of responses in longitudinal study		

Table 1. Continued

Study (Authors and Year)	Study Population	Methodological Approaches		
		Cognitive	Situational	Objective Measures
Brener, Kann, McManus, et al., 2002 [33]	n = 4619 Grades 9–12			Test-retest reliability
Clark, Brasseux, Richmond, et al., 1997 [99]	n = 149 Ages 12–21 years Adolescent medicine clinic patients			Medical records
Davoli, Perucci, Sangalli, et al., 1992 [91]	n = 383 Ages 13–21 years High school students in Rome, Italy		Mode effects (SAQ vs. IAQ)	
Fisher, Kupferman, & Lesser, 1992 [41]	n = 133 Grades 9–12 Students attending a school-based clinic			Random response technique
Ford & Norris, 1991 [89]	n = 64 Ages 15–21 years	Definition of terms		
Kann, Brener, Warren, et al., 2002 [28]	n = 3201 (household) n = 13,420 (school) Grades 9–12		School vs. household setting	
McFarlane & St. Lawrence, 1999 [88]	n = 296 Ages 12–19 years	Reference periods		
Millstein & Irwin, 1983 [92]	n = 108 Ages 14–20 years		Mode effects (SAQ vs. IAQ)	
Newcomer & Udry, 1988 [93]	n = 1152 Grades 7–9	Consistency of responses in longitudinal study	Self-reported honesty	
Orr, Fortenberry, & Blythe, 1997 [97]	n = 255 Ages 15–19 years Clinic patients			Biological measures (genitourinary cultures)
Rodgers, Billy, & Udry, 1982 [94]	n = 408 Grades 7–9	Consistency of responses		
Shew, Remafedi, Bearinger, et al., 1997 [98]	n = 540 Ages 13–21 years			Biological measures (diagnosis of STD)
Siegel, Aten, & Roghmann, 1998 [90]	n = 3144 Grades 7–12		Self-reported honesty	
Smith, McGraw, Crawford, et al., 1993 [100]	n = 586 Ages 15–19 years			Indirect evidence of behavior
Turner, Ku, Rogers, et al., 1998 [13]	n = 1690 Males aged 15–19 years		Mode effects (paper and pencil vs. audio-CASI)	
Webb, Zimet, Fortenberry, & Blythe, 1999 [26]	n = 388 Ages 13–20 years Students attending community adolescent health clinics		Mode effects (paper and pencil vs. CASI)	
Zelnik, Kantner, & Ford, 1981 [95]	n = 4392 Ages 15–19 years			Random response technique

CASI = Computer-assisted self-interviewing;
FFQ = food frequency questionnaires;
IAQ = interviewer-administered questionnaires;
SAQ = self-administered questionnaires.

Another method used to assess the validity of self-reported drug use is the random response technique (RRT). In studies using the RRT, respondents are presented with sets of two "yes/no" questions, one that is non-sensitive and will be answered with known probabilities, and one that is sensitive.

Through a random process, subjects determine which question in each set will be answered. An interviewer or others present never know which question respondents are answering, but responses to the sensitive questions can be derived mathematically. The RRT, however, is difficult to administer in large-scale studies, and analyses are limited to aggregate-level estimation rather than individual-level analysis [40]. A recent study that compared RRT responses with those from a standard questionnaire found that RRT responses were marginally higher for alcohol use but not for marijuana and cocaine use, which provides support for the validity of the standard questionnaire used in that study [41].

A final method used to assess validity is the inclusion of a question on the use of a fictitious drug. One study found that few students reported using a fictitious drug, but students who did so were more likely to report maximum frequency of other "real" drug use, a finding that suggests they were generally overreporting [42].

Tobacco Use

Cognitive factors. Like alcohol and other drug use, reports of tobacco use often are obtained for one or more reference periods. Although questions about current use or "ever" use of a tobacco product are relatively easy to answer, longitudinal studies of adolescents have found inconsistent responses over time [18,43]. Adolescents also tend to be inconsistent when asked to recall the age at which they initiated tobacco use. Recall beyond a 1-year period tends to be inaccurate [16,17,43]. In addition, asking respondents to estimate the frequency of smoking and the number of cigarettes they smoked on specific recent days is cognitively challenging, in part "because [smoking] tends to be habitual, repetitious, and almost unconscious" [7].

In addition to the potential for recall error, comprehension and decision-making processes also may affect the accuracy of self-reports of tobacco use. Pokorski et al [44] suggest adolescents may underreport their smoking behavior because infrequent and episodic smoking may make it difficult for them to describe their usual pattern of smoking, and many may not define themselves as smokers.

Situational factors. Tobacco use meets with more social disapproval today than in the past. For adolescents under 18 years old, laws prohibit the sale of tobacco products. Given these formal and informal sanctions, researchers generally expect self-reported tobacco use to be underreported because of concerns of social desirability and fear of reprisal.

One way to assess whether self-reported tobacco use is affected by social desirability and fear of reprisal is to compare the rates generated in different interview settings. Studies comparing school-based and household-based data collection have found that the prevalence of tobacco use is higher in surveys conducted in schools [27,28,32,45]. Further, Hedges and Jarvis [45] compared smoking prevalence estimates obtained in a school-based and a household-based setting with those obtained from a biochemical measure, and found the higher prevalence rate reported in the school-based setting was closer to the biochemical measure than was the household-based prevalence rate. Together, these studies suggest self-reported tobacco use is affected by perceptions of privacy and confidentiality.

An alternative way to assess whether self-reported tobacco use is affected by social desirability and fear of reprisal is to vary the mode of administration. Since a perceived lack of confidentiality and a desire to give socially appropriate responses may lead to underreporting, we would expect an IAQ to produce less reporting of tobacco use than an SAQ. This hypothesis generally has found support. For example, in testing alternative designs for the National Household Survey of Drug Abuse (NHSDA), Turner et al [22] found that among adolescents aged 12 to 17 years, IAQs compared to SAQs reduced reported use of cigarettes during the past year and the past 30 days. Similarly, Brittingham et al [46] found that SAQs produced higher prevalence estimates of smoking than did IAQs, but only among adolescents. In a panel study, Needle et al [34] examined the effect of using SAQs in the home vs. using mailed questionnaires on adolescents' self-reports of lifetime and 1-year prevalence of cigarette use. They conjectured that less anonymity and privacy in the home, simply because an interviewer was present, would result in lower self-reported rates under this condition. However, no difference in reported levels of cigarette use was found. Coupled with other study findings, this suggests that it is the direct interaction with the interviewer during the interviewing process that may lead to underreporting. In addition, a study comparing a paper-and-pencil SAQ with CASI found no effects by mode of administration [26].

Under the assumption that subjects will be less likely to intentionally distort their responses when they believe their responses will be validated, the pipeline and bogus pipeline approaches have been

used to assess whether self-reports of tobacco use are affected by situational factors. The difference between a "pipeline" and "bogus pipeline" approach is that, in the former, respondents initially are told the veracity of their responses will be checked by an objective measure, and they subsequently are. In the latter, respondents are also initially told their responses will be checked, but the objective measure is not a legitimate measure and serves no purpose other than to motivate respondents to respond truthfully. In the tobacco use literature, the pipeline approach frequently is used to determine whether individuals are deliberately misreporting their behaviors under standard interviewing conditions. The objective measure by which to subsequently "validate" the self-reports is typically one of several available biochemical measures.

A meta-analysis of 15 studies found that, overall, a higher prevalence of smoking was detected when a pipeline approach was used [47]. However, some studies failed to find any pipeline effect, while others found very large effects. Other than poor study design, several reasons for the mixed results can be identified. First, a precondition for a pipeline effect is for the behavior to be regarded as socially undesirable. For some subgroups of adolescents, tobacco use may not be viewed this way. Second, the pipeline message must be credible. Adolescents must believe their use of tobacco can be detected with the biochemical measure explained and used in the study. Third, with or without a pipeline protocol, the most important factor in obtaining accurate reports may be sufficient assurance that responses will be kept confidential or anonymous. Finally, adolescents may actually provide valid responses. The response validity for tobacco use may be at such a high level that a pipeline approach can produce no significant increment to validity [47,48].

Assessment of the validity of self-reports. The test-retest reliability of self-reported behaviors related to tobacco use among adolescents is very high, higher than for other health-risk behaviors [33].

The RRT approach has been used to validate self-reports of tobacco use. Martin and Newman [3] found that the RRT produced slightly higher estimates of adolescent cigarette smoking than those obtained with a self-administered questionnaire. They concluded that the magnitude of this difference did not warrant the added difficulties of administering and interpreting results from the RRT. Similarly, in a school-based study of the smoking behavior of adolescents, Akers et al [49] found the percentage of students reporting smoking at least once on an SAQ was in close agreement with the percentage derived from the RRT. Although the RRT is not a completely objective measure by which to assess validity, these studies provide some evidence that adolescents provide accurate self-reports of their smoking behavior when an SAQ is used.

The most common method of validating self-reports of tobacco use has been to compare them to biochemical tests, such as measurement of cotinine in plasma, saliva, or urine, thiocyanate in plasma or saliva, and carbon monoxide (CO) in expired air. Biochemical measures, however, are not fool-proof tests for assessing response validity of tobacco use. Levels of thiocyanate can be elevated by certain foods, and CO can be elevated by environmental pollutants [50]. Also, thiocyanate and CO cannot always distinguish between tobacco and marijuana use [3,50]. A particular problem in using these tests is that they are relatively insensitive to the low levels of smoking and experimental smoking characteristic of adolescents [51]. Other concerns with these measures include their invasiveness, which increases refusal rates, and their relatively high cost, which limits their feasibility in many studies [35].

Despite these limitations, biochemical testing has become the method of choice, and the most credible means, for validating self-reported tobacco use. A meta-analysis of 26 studies that validated self-reported current and recent smoking behavior with biochemical measures found relatively high levels of sensitivity (the percentage of respondents who reported smoking and who tested positive with biochemical measures) and specificity (the percentage of respondents who reported absence of smoking and who tested negative with biochemical measures) [35]. These studies also revealed, however, that the sensitivity of adolescents' self-reports tended to be lower than that of the general population. Patrick et al attributed this difference to issues related to legality and the fact that many adolescent tobacco users may not define themselves as smokers [35].

Several studies not included in the meta-analysis described above also have used biochemical measures to assess the validity of self-reported tobacco use, including smokeless tobacco use, among adolescents [3,49,52–54]. These studies generally have shown strong agreement between self-reported and biochemical measures of tobacco use. At least two of these studies, however, showed a discrepancy between self-reported and biochemical measures associated with race/ethnicity; adolescents in minority

populations tended to underreport smoking more than white adolescents [52,53].

Behaviors Related to Unintentional Injuries and Violence

Cognitive factors. One threat to the validity of self-reported behaviors related to injuries and violence is recall error. Adolescents are often asked to provide information that quantifies behaviors such as seatbelt nonuse, drinking and driving, and fighting during specific time periods. While studies of this issue are scarce, one study found that high school students' absolute estimates of interpersonal violence-related behaviors did not vary significantly when different reference periods (i.e., 1 month, 6 months, 12 months) were used [2]. Such responses suggest students may have some difficulty in reporting behavior for specific time periods, although reporting for specific time periods was better for more severe behaviors.

Even salient behaviors such as suicide attempts may not be reported accurately. For example, in one study, adolescent outpatients were asked to explain discrepancies between self-reported suicidal behavior on an SAQ and a semi-structured interview [55]. For many, these discrepancies were caused by a lack of clarity in their definition of suicidal behavior. For example, on the SAQ, some respondents reported incidents of self-mutilation, but claimed during the interview that such mutilation was not intended as a suicide attempt [55]. In another study, 19% of respondents reporting suicidal ideation during an assessment covering a short interval failed to report suicidal ideation during their lifetime in a separate assessment [56].

Situational factors. The impact of situational factors on behavioral self-reports also is likely to differ by type of outcome. Some behaviors, such as nonuse of seatbelts, are not very sensitive, carry relatively minor sanctions, and would not be expected to be subject to large biases related to either social desirability or fear of reprisal. Indeed, Webb et al [26] found no differences between CASI and a paper-and-pencil SAQ in self-reported seatbelt use. Drinking while driving, in contrast, is a more sensitive behavior, and one would therefore expect a greater potential for bias owing to situational factors. These differences explain why we cannot make any global statements about the impact of specific situational factors on self-reported behaviors related to unintentional injuries and violence.

Several studies have examined the impact of different modes of administration on reports of behaviors related to suicide and violence. As described above, one study found a higher prevalence of self-reported suicidal behavior in an SAQ than in a face-to-face interview [55]. Similarly, Klimes-Dougan [56] found a higher prevalence of reported suicidal ideation in a paper-and-pencil survey than in a structured interview. These studies suggest that a lack of privacy can lead to underreporting of suicidal behavior.

Turner et al [13] found CASI produced statistically significantly higher prevalence rates than did the paper-and-pencil SAQ for weapon-carrying, acts of violence, and threatened violence. The effects were, however, weaker than those for other, more sensitive behaviors, such as male-male sexual contact, suggesting that reports of violence-related behaviors are affected less by privacy than are other, more sensitive behaviors.

One study that tried to directly assess the impact of social desirability on responses to questions measuring adolescents' suicide ideation found social desirability was significantly associated with two of four scales constructed by the researchers, suggesting that respondents were providing socially desirable responses [57].

Assessment of the validity of self-reports. Brener et al [33] found that questions assessing unintentional injuries and violence among adolescents exhibited moderate test-retest reliability. Other studies have demonstrated high reliability for questions on suicide ideation [58].

In our review of the literature, we did not identify any attempt to validate self-reports of violence-related behaviors or suicide attempts. Studies concerned with suicide ideation have restricted their attention to establishing the reliability and construct validity of instruments assessing suicidal behavior. A review of such instruments designed for use with adolescents concluded that research using these instruments has not paid sufficient attention to their validity, and that well-validated instruments are needed to assess suicidal ideation and behavior among adolescents [59].

Dietary Behaviors

Cognitive factors. Much research has employed food frequency questionnaires (FFQs) to gain insights into eating habits of populations and to analyze food intake at a level corresponding to energy

and nutrients [60]. This type of questionnaire asks respondents to indicate their "usual" food intake over a weekly, monthly, or yearly reference period. The validity of this method has been explored extensively because of anticipated cognitive difficulty for respondents.

For example, the potential for recall bias in FFQs can be assessed by diary methods. The FFQ is compared with "diet records" that respondents make when they eat or shortly thereafter. A study of 16- to 19-year-olds found good agreement between an FFQ and diet records, although the level of agreement varied widely by type of food [61].

Other studies of adolescents have compared FFQ results to those of 24-hour recalls, obtained via interview with a dietician [62–64]. Rockett et al found an average correlation of .54 between the measures, a value similar to that found for adults [64]. Another study was restricted to fruit and vegetable questions. They found that, although FFQ assessments are useful for ranking subjects' consumption of fruits and vegetables, as compared to dietary recall, FFQs tend to underestimate the prevalence of fruit and vegetable consumption among high school students [62]. Similarly, Frank et al found that their FFQ provided underestimates of consumption of many types of food relative to a 24-hour recall [63].

Assessment of dietary behaviors also includes questions about weight control practices. The validity of adolescent self-reported data on behaviors related to weight control has been assessed using diary methods. Rosen and Poplawski [65] administered a 14-item weight control questionnaire to high school students. Before completing the questionnaire, subjects had been instructed to keep a daily, 7-day record of all exercise and food and liquid consumed, as well as use of diet pills, laxatives, and vomiting. Those who reported that they were engaged in weight control practices had significantly more diary reports of exercising, skipping meals, and using diet pills than did others. Those reporting that they were trying to lose weight also had lower caloric intake as calculated from the diet records. However, self-reported fasters, vomiters, and laxative users did not exhibit such behaviors in their diary records. For relatively non-sensitive weight control practices, then, adolescents had little difficulty in recalling or reporting these efforts, but for unhealthy weight control methods, substantial discordance between diary reports and self-reports obtained from a questionnaire was found. This latter finding may be less a function of recall error and more a function of social desirability bias against reporting sensitive behaviors that may be symptomatic of eating disorders.

Situational factors. Because researchers generally regard dietary intake as a relatively non-sensitive behavior, the literature has tended to focus on cognitive factors, especially recall problems, that threaten the validity of self-reports of food consumption. This particular focus is reflected in the numerous studies described above. Situational factors are, however, likely to affect reports of weight control practices among adolescents. For example, Rosen and Poplawski's [65] finding that unhealthy weight control methods were not consistently reported in diet records and subsequent self-administered questionnaires lends support to contentions that eating disorders are subject to social disapprobation and consequent denial on the part of patients diagnosed with these disorders. In addition, a study that assessed behaviors related to dieting, purging, and binge eating used an SAQ followed by a clinical interview, and found significantly lower prevalence rates of all behaviors when assessed via interview [66]. This difference suggests reporting of such behaviors can be affected by privacy and confidentiality. Webb et al [26], however, showed that less sensitive behaviors, such as limiting salt and fat and eating fresh vegetables, did not vary by mode of administration (CASI vs. paper-and-pencil SAQ).

Assessment of the validity of self-reports. Many of the studies discussed above that have compared diet records with FFQs also have administered the FFQ twice to assess the test-retest reliability of the questionnaire. In one such study, Rockett et al [67] found reliability coefficients for specific foods ranging from .39 to .57. In another, Frank et al [63] found agreement ranging from .33 to 1.0.

Recent reviews of the validity of dietary assessment methods among school-age children found that correlations between self-reported data and objective measures generally were higher for diet records and recalls than for FFQs, with underreporting of food intake among older children and adolescents [68,69]. In addition, studies have found diet histories provide more valid reports of food intake than diet records. Many of these studies use the doubly labeled water technique, a biochemical method that measures energy expenditure by having subjects drink isotope-labeled water and then provide a urine sample after a specified time. For example, Livingstone et al [70]

found adolescents aged 15 and 18 years underreported their intake as assessed by diet records, but their reports of intake derived from interviewer-administered diet histories were in good agreement with the energy expenditure measures. It is somewhat disconcerting that diet records are not in good agreement with an objective measure such as doubly labeled water since, as discussed above, many studies have drawn conclusions about the validity of self-reports derived from FFQs on the basis of comparison with diet records. Indeed, studies reported by Schoeller [71] that compared results from the doubly labeled water technique with self-reports from diet records found substantial underreporting of food intake in diet records among obese subjects, female endurance athletes, and adolescents. These studies suggest that diet records themselves should not be used as independent methods of validation of food consumption.

With respect to self-reports of behaviors related to weight control, Brener et al [33] obtained moderate reliability estimates for questions assessing these behaviors among adolescents. In the study by Rosen and Poplawski [65] described above, high school students' reports from a self-administered 14-item weight control questionnaire also were compared with external ratings of the respondents' behavior by parents and peers (defined as siblings or friends). Relatively high levels of consistency were obtained for reports of trying to gain or lose weight (82% for parents and 76% for peers). However, agreement between parents and respondents on the use of unhealthy weight control methods (fasting, using diet pills, and vomiting) was very poor, although higher levels of agreement were obtained between the peer and respondent reports. One possible conclusion from these findings is that, given the lower agreement on reports of unhealthy weight control methods, measuring the use of these methods through self-report is invalid. An alternative and more likely interpretation, however, is that because such behaviors tend to be stigmatized and are therefore more likely to be carried out in isolation, peers and parents simply are not aware of them. Furthermore, peers may be more aware of the sensitive eating behaviors of adolescents than are parents. This points to a general problem of using others' reports to validate self-reported behavior, namely, that the others must have sufficient contact and rapport with the subject to observe and report on her or his behavior accurately.

Physical Activity

Cognitive factors. Reporting on levels of physical activity is a complex and challenging endeavor for several reasons. First, many separate activities such as walking or bicycling must be separately reported. Second, as with the other behaviors, reporting is usually requested for some reference period, such as a day or a week. Respondents therefore must accurately recall many separate events, some of which (e.g., climbing stairs) may not be particularly salient or memorable. Third, some surveys request information about both duration and intensity of each activity, greatly increasing the difficulty of the recall task. Fourth, the categories of physical activities defined in surveys (e.g., moderate or vigorous physical activity) require respondents to make judgments about which specific activities fall within each category.

Several studies have demonstrated that cognitive factors have an impact on self-reported physical activity. In one study, students were twice administered a Seven Day Activity Recall questionnaire that asked about time spent in sleep and in moderate, hard, and very hard activities [72]. The time between interviews varied between 2 and 6 days, so some days were reported on in both interviews. Further, the varying time period between interviews meant that lag time from the first to the second report on those overlapping days also varied. The researchers found that repeated reports from longer intervals (4–6 days between interviews) were less reliable than those from shorter intervals. They suggest this relationship is owing to decay in a subject's ability to remember specific physical activities. In a related study, Rifas-Shiman et al [73] found a seasonal format questionnaire led to more accurate reporting of physical activity than an annual format questionnaire.

When a survey attempts to obtain retrospective reports of physical activity, the nature of those reports is influenced by the way questions are asked. Evidence for this can be found in studies showing weak correlations among indices of physical activity obtained from different types of questionnaires [72], as well as from studies finding that the correlations between these indices and physiologic or mechanical measures of activity vary widely [74,75].

Situational factors. Behaviors related to physical activity are not sensitive. Therefore, we would expect that situational factors would have only a small impact on self-reported data about these activities. However, exercise does tend to be positively valued

and athletes often are held in high esteem. Based on these considerations, reports of exercise and strenuous activity, especially in a sports context, may be subject to some degree of social desirability bias. However, Webb et al [26] found no significant differences between CASI and paper-and-pencil SAQ on self-reported measures of physical fitness and ability to play at active sports.

Assessment of the validity of self-reports. Studies assessing the reliability of questions assessing self-reported physical activity among adolescents find the questions to be moderately to substantially reliable [1,33,72,74,76,77]. A recent review also found moderate to high reliability among several measures of self-reported physical activity [78].

Attempts to objectively assess the validity of self-reports of physical activity among children and adolescents have compared them to one or more other types of measures: mechanical or electronic monitors, including accelerometers and heart rate monitors; energy expenditure, including doubly labeled water and calorimetry; measures of fitness; and direct observations [79]. Sallis and Saelens [78] describe this as "relative validity" as opposed to "absolute validity." These types of measures have not been shown to correlate strongly with self-report, in part because they do not measure the same thing. For example, the doubly labeled water technique measures total energy expenditure, of which physical activity is just one component.

Similarly, although monitors provide an objective measure expected to be related highly to self-reports of physical activity, studies have not found strong correlations between self-reported data and readings from accelerometers [79] or activity monitors [80]. One study, however, found a correlation of .88 between self-reported physical activity and accelerometer readings [75]. While the overall weak correlations could indicate invalid self-reports, it is far more likely they are owing to limitations of the available accelerometers, which do not accurately assess activities such as bicycling in which there is minimal vertical acceleration and deceleration of the body [81]. Accelerometers also may miss many light and moderate physical activities. Some support for this supposition is found in the fact that Janz et al [74] found stronger correlations for vigorous exercise than for other kinds of activities. In addition, many kinds of activities that are not well-monitored (e.g., bicycling), or cannot be monitored at all (e.g., swimming), are the kinds of activities in which children and adolescents engage. Thus, an accelerometer may be particularly unsuited to validate the self-reports of children and adolescents. It clearly does not provide a gold standard against which self-reported data can be compared.

Sallis et al [72] used heart-rate monitors to assess the validity of physical activity reports among 5th, 8th, and 11th grade students. For 5th and 8th grade students, only moderate correlations (r = .33 to .45) were found between hours of recalled very hard activity and minutes with heart rates of 140 or above or 160 and above. For 11th grade students, the correlations were somewhat higher, particularly when minutes with heart rates of 160 and above were considered. In this latter case the correlation reached .72. The researchers concluded that self-reported data on very hard activities are well-validated. Other studies, however, have found lower correlations between self-reported physical activity and heart rate [82].

A few studies have compared measures of fitness with various self-reported measures of physical activities in children and adolescents. The measures used include body mass index [1,83], oxygen uptake or aerobic capacity [83–85], and fitness test scores [1,76]. Although some studies found significant associations between self-reports of physical activity and fitness measures, their usefulness as a validation tool is limited. These measures represent relatively long-term outcomes of the behaviors being reported and are influenced by genetic and environmental factors [79]. Therefore, they cannot assess the validity of self-reports of specific activities over a short time. They are useful for establishing a kind of construct validity but, contrary to some claims [84], may not represent a gold standard.

Other studies have compared self-reported physical activity with observations of this activity. For example, Grunbaum et al [86] observed physical education classes and found that students overreported the time they spent in moderate-to-vigorous activity in a paper-and-pencil questionnaire.

Sports team participation is one area of physical activity for which a gold standard does exist. One study that examined sports team participation showed that adolescents accurately reported their participation on school sports teams; a high degree of correspondence between self-reported sports team participation and school team rosters was found [1].

Sexual Behavior

Cognitive factors. Although attention has mainly been given to the effects of situational factors on

self-reports of sexual behavior among adolescents, some evidence in the literature suggests that cognitive factors also play a role. As with most behaviors, the accuracy with which people can recall sexual behaviors is influenced by such factors as the length of the recall period and the vividness of the events [87]. For example, McFarlane and St. Lawrence [88] found that among African-American adolescents, estimates of yearly sexual behavior based on 2-week, 2-month, and 12-month reference periods provided discrepant results. Regardless of the length of the reference period, studies generally have shown that high-frequency sexual behaviors are reported less consistently than low-frequency sexual behaviors, because respondents who engage in a behavior frequently are less likely to remember specific instances [88].

Another cognitive factor is the terminology used in questions about sexual behavior. Although different populations might use different words for describing the same behaviors, the majority of respondents comprehend standard terminology [87]. For example, Ford and Norris [89] found that both African-American and Hispanic adolescents were able to answer questions about their sexual behavior when anatomical, rather than slang, words were used.

Another study indicated that 96% of students provided the same answer to a question about ever having had sexual intercourse regardless of whether the question included alternate terms for sexual intercourse (e.g., "making love," "going all the way"). That same study also showed that 93% of students defined vaginal sex as sexual intercourse, 62% of students defined anal sex as sexual intercourse, and 22% of students defined oral sex as sexual intercourse (CDC, unpublished data, 2000).

Situational factors. Sexual behavior generally is regarded as a private matter. People are reluctant to divulge information about their sexual practices because of potential embarrassment and concern about confidentiality and anonymity. Because unprotected intercourse is a leading cause of HIV infection, it is possible that people's responses also are influenced by fear of disapproval and informal social sanctions. Given these situational factors, researchers generally expect an underreporting of adults' sexual behaviors. For some population subgroups, however, overreporting is a distinct possibility. For example, as a sign of maturity and a means to attain adult status, some adolescents may perceive the need to exaggerate their sexual involvement. It is likely, therefore, that socially desirable answers change with age and developmental stage [10]. Given differences in cultural norms and expectations, differences in accuracy of reporting according to a person's gender and race/ethnicity are possible. Some evidence for gender differences can be found in a study of self-reported sexual behavior among middle school students [90]. After completing SAQs, students were asked whether they had responded honestly. While the majority of students said their responses were very honest or completely honest, male students were more likely to report that they overstated their actual behavior, while female students were more likely to say they underreported their actual behavior. Of course, no gold standard exists for self-reported honesty; it is possible respondents do not answer such questions any more truthfully than questions about behavior.

Social desirability and the need to present oneself in a positive light enormously complicates the task of validating self-reports of sexual behavior. Since both underreporting and overreporting may occur, we cannot assume that higher reported rates under certain test or interviewing conditions mean truer rates. Nevertheless, to address the potential for any social desirability bias, numerous studies have been conducted to ascertain the conditions under which respondents report differently about their sexual behavior.

Perhaps the most widely investigated method of assessing bias in self-reported sexual behavior data attributable to social desirability is through mode of administration. Typically, this approach has compared the use of paper-and-pencil SAQs with face-to-face IAQs. With few exceptions, researchers have found these two modes of administration yield different levels of self-reported behavior. For example, Davoli et al [91] found that more adolescents reported engaging in sexual intercourse and fewer reported using condoms with a self-administered version of the questionnaire. Similarly, in a study of adolescent girls, Millstein and Irwin [92] found the percentage of respondents who reported having ever engaged in eight sexual behaviors was higher under an SAQ than an IAQ mode of administration.

Differences in responses by SAQ vs. IAQ mode of administration have been so well-researched that investigators have moved more recently to comparisons of new modes of administration, especially CASI. Among adolescent males, Turner et al [13] found that the percentage of respondents who reported engaging in sexual behaviors such as intercourse with a prostitute, five or more sexual partners, and male-male sexual contact was significantly

higher in an audio-CASI interview than in the standard SAQ. For less sensitive sexual behaviors such as intercourse with a female, however, no such mode effects were found. Similarly, Webb et al [26] found no difference between paper-and-pencil SAQ and CASI on measures of sexual intercourse and condom use.

Assessment of the validity of self-reports. Numerous studies have examined the test-retest reliability of adolescents' self-reports of sexual behavior. In one study [33], kappa values for questions assessing sexual behavior among adolescents ranged from 40% to 90%. Davoli et al [91] also examined the consistency in adolescents' responses to questions about sexual experience and condom use across two administrations of a questionnaire and found high levels of agreement (kappas $>$ 70%) on reports of various sexual behaviors. Reports of condom use were slightly less consistent (kappas $>$ 60%).

A study by Alexander et al [10] focused on whether differences in the consistency of responses to questions about sexual behavior varied by age, gender, and race. They administered a questionnaire to junior high and high school students annually for 3 years, focusing on reports of whether an individual had ever had sexual intercourse, lifetime frequency of sexual intercourse, and age at first intercourse. Reports of sexual intercourse and frequency of intercourse were considered inconsistent if they involved recanting (e.g., if respondents reported in grade 8 that they ever had sexual intercourse but reported in grade 9 that they had not). More inconsistencies were obtained between the first two tests (during grades 8 and 9) than between the second two tests (during grades 9 and 10). They claimed the social environment of high school minimizes the need to exaggerate one's involvement in socially unacceptable behaviors to achieve status. They also found levels of consistency varied by gender and race. For example, white females had the lowest level of inconsistencies in their reports of lifetime sexual intercourse, and black males had the highest level. Inconsistencies obtained for white males and black females were between these two extremes and approximately equal. Finally, these investigators found the level of consistency varied among the measures of sexual behavior that were obtained. Whereas levels of inconsistency were relatively low for both engaging in sexual intercourse and lifetime frequency of sexual intercourse, levels were dramatically higher for reports of age at first intercourse.

Newcomer and Udry [93] conducted a variation of the test-retest method. In their study, adolescents were asked to repeat what they had reported about their sexual experience during an interview 2 years earlier, and whether they had told the truth at that time. Of the adolescents interviewed in both 1980 and 1982, 83% reported in 1982 that they had told the truth about their sexual experience during the first interview; 61% said they had been honest in reporting that they never had sexual intercourse; and 22% said they had been honest in reporting they had engaged in sexual intercourse at the first interview. Seven percent of adolescents said they had been dishonest about their earlier reports. The remaining 10% said they had told the truth, but their answers about their sexual experience at Time 1 were not concordant between survey rounds. Results indicated that males, Blacks, and adolescents with sexual experiences other than intercourse were less likely to report being honest. Newcomer and Udry concluded that adolescents will admit to not telling the truth about earlier reports and that most are able to recall their earlier responses. Nevertheless, the researchers cautioned that an adolescent's ability to precisely date her or his age of sexual initiation is suspect.

Rodgers et al [94] assessed the reliability of self-reported sexual behavior by examining the consistency of reports within one questionnaire administered to junior high school students. In general, these investigators obtained low levels of inconsistency, although rates varied according to the presumed sensitivity of the behavior. For behaviors that ranged from holding hands to intercourse, the researchers found a correlated increase in inconsistency.

As with some other behaviors discussed in this paper, the RRT has been used to validate self-reports of sexual behavior. For example, Zelnik et al [95] found the RRT produced slightly higher levels of reported sexual behavior among respondents in the National Survey of Young Women.

Researchers agree the best way to validate self-reported sexual behavior is through comparison with biochemical measures or official records. Each of these techniques, however, has some rather severe drawbacks that make its use in sexual behavior research problematic. Although biochemical measures often have been used in validity studies of tobacco-, alcohol-, and other drug-use behavior, they are less practical in validating self-reports of sexual behavior because no biological or chemical tests can fully detect sexual behavior. Sperm can be detected in urine, but this test is only applicable for assessing the sexual behavior of women, can only detect an

underreporting bias in self-reports, and only measures very recent sexual activity. Other biological markers such as pregnancy rates, seroconversion rates for HIV, and rates of other STD infection can corroborate self-reported sexual behavior, but risk behaviors that do not result in pregnancy or infection cannot be captured [96]. In addition, biologic measures are invasive, costly to analyze, and provide a limited amount of information only on recent or current sexual behavior. For these reasons, Catania et al [96] conclude that this method generally has proved unfeasible for the validation of most sexual behavior research.

Despite the imperfect nature of using STD infection to validate self-reported sexual behaviors, it has been used in at least two studies. In the first, adolescents who had been treated for STDs were asked about their sexual behavior when they returned for follow-up medical appointments. The researchers found a good correspondence between the adolescents' reports of sexual intercourse and number of sexual partners and their subsequent STD infection [97]. In a similar study, adolescents were given a physical examination including laboratory tests of specimens after completing an SAQ that asked about sexual behavior. The researchers found self-reported condom use with the last two partners was associated with the absence of an acute STD [98].

As with biochemical measures, the use of official records for validation purposes also has drawbacks. Only very indirect and limited information reflecting a person's sexual behavior is available, such as records of pregnancies and STD infections. Further, these outcomes reflect the sexual experiences of only some individuals, typically those who have had unprotected sex. Many of these sexual outcomes also are likely to have a very low prevalence and incidence in the adolescent population, thereby making statistical comparisons difficult.

Given these limitations, validation of self-reported sexual behavior *per se* through the use of records is rare. Our review of the literature discovered only two such attempts. In the first, adolescent clinic patients provided self-reports of STDs and pregnancies during structured interviews. When these reports were compared with the patients' medical records, the researchers found approximately half the adolescents did not provide accurate reports of STD infections. Though reports of pregnancies were more accurate, they still were far from perfect [99].

In the second study, Smith et al [100] took a rather unusual approach in examining "official records" to verify the accuracy of self-reported condom use. After being interviewed about their sexual and condom use behavior, a sample of Latino adolescents was asked whether they had any condoms in their possession and, if so, to show them to the interviewer. Those who reported they had purchased condoms were 3.0 times as likely as others to show a condom to an interviewer; those who said they had used condoms recently were 2.3 times as likely as others to have a condom in their possession at the time of interview. For this sample of adolescents, then, self-reports of condom use were related strongly to condom possession. This type of validation technique, in which a respondent is required to provide some evidence of a behavior, is rather indirect and reflects behavioral intentions rather than actual behavior. Nevertheless, it is innovative and sheds some light on the validity of these adolescents' responses.

Discussion

As our review of the literature has shown, self-reports of each of six types of health-risk behaviors are affected by both cognitive and situational factors. These factors, however, do not threaten the validity of self-reports of each type of behavior equally. Further, each type of behavior differs in terms of the extent to which it can be validated by an objective measure.

This review has several limitations. First, although we attempted to be systematic and thorough in our literature search, it is possible we missed articles meeting our inclusion criteria. Second, as with any review of published literature, this review is subject to bias in that studies not finding significant effects are less likely to be published [101]. This bias, however, is less of a problem in the types of studies reviewed in this paper because nonsignificant findings are more likely to be published as part of methodological studies than in intervention studies. Third, some of the six risk behavior categories reviewed in this paper include a wide range of behaviors. This is especially true for the category of injury-related behaviors, which includes behaviors related to suicide, unintentional injuries, perpetration of violence, and violent victimization. These behaviors are disparate and the issues surrounding the validity of the self-report of these behaviors also vary widely, but only selected issues could be covered in this paper. Relatedly, while some of the risk behavior categories are widely researched, others have few studies assessing the validity of self-reported behav-

ior. In those cases, such as injury-related behaviors, the few studies reviewed in this paper take on a disproportionate weight in the conclusions we draw. Further research in these areas will help alleviate this imbalance.

The importance of assessing the prevalence of health-risk behaviors as part of research activities involving adolescents often necessitates the use of self-report measures. This review has demonstrated that self-reports of these types of behaviors are indeed affected by both cognitive and situational factors in varying degrees. Researchers should familiarize themselves with these threats to validity and design studies that minimize these threats as much as possible.

References

1. Aaron DJ, Kriska AM, Dearwater SR, et al. Reproducibility and validity of an epidemiologic questionnaire to assess past year physical activity. Am J Epidemiol 1995;142:191–201.
2. Hilton NZ, Harris GT, Rice ME. On the validity of self-reported rates of interpersonal violence. J Interpers Violence 1998;13:58–72.
3. Martin GL, Newman IM. Assessing the validity of self-reported adolescent cigarette smoking. J Drug Educ 1988;18: 275–84.
4. Cannell CF, Miller PV, Oksenberg L. Research on interviewing techniques. In: Leinhardt S (ed). Sociological Measurement. San Francisco: Jossey-Bass, 1981:389–437.
5. Eisenhower D, Mathiowetz NA, Morganstein D. Recall error: Sources and bias reduction techniques. In: Biemer PP, Groves RM, Lyberg LE, et al (eds). Measurement Errors in Surveys. New York: John Wiley & Sons, 1991:127–44.
6. Lessler JT, Tourangeau R, Salter W. Questionnaire design in the cognitive research laboratory. Vital and Health Statistics, Series 6, No. 1. Washington, DC: Government Printing Office; 1989. DHHS Pub. No. PHS 89-1076.
7. Means B, Habina K, Swan GE, et al. Cognitive research on response error in survey questions on smoking. Vital and Health Statistics, Series 6, No. 2. Washington, DC: Government Printing Office, 1992. DHHS Pub. No. PHS 92-1080.
8. DeMaio T. Social desirability and survey measurement: A review. In: Turner CF, Martin E (eds). Surveying Subjective Phenomena. New York, NY: Russell Sage Foundation, 1984: 257–82.
9. Sudman S, Bradburn NM. Response Effects in Surveys. Chicago: Aldine Publishing Company, 1974.
10. Alexander CS, Somerfield MR, Ensminger ME, et al. Consistency of adolescents' self-report of sexual behavior in a longitudinal study. J Youth Adolesc 1993;22:455–71.
11. Winters KC, Stinchfield RD, Henly GA, Schwartz RH. Validity of adolescent self-report of alcohol and other drug involvement. Int J Addict 1991;25:1379–95.
12. Gans JE, Brindis CD. Choice of research setting in understanding adolescent health problems. J Adolesc Health 1995; 17:306–13.
13. Turner CF, Ku L, Rogers SM, et al. Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. Science 1998;280:867–73.
14. Bachman JG, O'Malley PM. When four months equal a year: Inconsistencies in student reports of drug use. Public Opin Q 1981;45:536–48.
15. Bailey SL, Flewelling RL, Rachal JV. The characterization of inconsistencies in self-reports of alcohol and marijuana use in a longitudinal study of adolescents. J Stud Alcohol 1992;53: 636–47.
16. Engels R, Knibbe RA, Drop MJ. Inconsistencies in adolescents' self-reports of initiation of alcohol and tobacco use. Addict Behav 1997;22:613–23.
17. Johnson TP, Mott JA. The reliability of self-reported age of onset of tobacco, alcohol, and illicit drug use. Addiction 2001;96:1187–98.
18. Shillington AM, Clapp JD. Self-report stability of adolescent substance use. Drug Alcohol Depend 2000;60:19–27.
19. Johnston LD, O'Malley PM. The recanting of earlier reported drug use by young adults. The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates. NIDA Res Monogr 1997;167:59–80.
20. O'Malley PM, Bachman JG, Johnston LD. Reliability and consistency in self-reports of drug use. Int J Addict 1983;18: 805–24.
21. Schober SE, Fe Caces M, Pergamit MR, Branden L. Effect of mode of administration on reporting of drug use in the National Longitudinal Survey. In: Turner CF, Lessler JT, Gfroerer JC (eds). Survey Measurement of Drug Use. Washington, DC: Government Printing Office, 1992:267–76.
22. Turner CF, Lessler JT, Devore JW. Effects of mode of administration and wording on reporting of drug use. In: Turner CF, Lessler JF, Gfroerer JC (eds). Survey Measurement of Drug Use: Methodological Studies. Washington, DC: Government Printing Office, 1992:177–220.
23. Wright DL, Aquilino WS, Supple AJ. A comparison of computer-assisted and paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use. Public Opin Q 1998;62:331–53.
24. Beebe TJ, Harrison PA, McCrae JA, Jr, et al. An evaluation of computer-assisted self-interviews in a school setting. Public Opin Q 1998;62:623–32.
25. Hallfors D, Khatapoush S, Kadushin C, et al. A comparison of paper vs. computer-assisted self interview for school alcohol, tobacco, and other drug surveys. Eval Program Plann 2000; 23:149–55.
26. Webb PM, Zimet GD, Fortenberry JD, Blythe MJ. Comparability of a computer-assisted versus written method for collecting health behavior information from adolescent patients. J Adolesc Health 1999;24:383–8.
27. Gfroerer J, Wright D, Kopstein A. Prevalence of youth substance use: The impact of methodological differences between two national surveys. Drug Alcohol Depend 1997; 47:19–30.
28. Kann L, Brener ND, Warren CW, et al. An assessment of the effect of data collection setting on the prevalence of health-risk behaviors among adolescents. J Adolesc Health 2002;31: 327–35.
29. Rootman I, Smart RG. A comparison of alcohol, tobacco and drug use as determined from household and school surveys. Drug Alcohol Depend 1985;16:89–94.
30. Campanelli PC, Dielman TE, Shope JT. Validity of adolescents' self-reports of alcohol use and misuse using a bogus pipeline procedure. Adolescence 1987;85:7–22.
31. Werch CE, Gorman DR, Marty PJ, et al. Effects of the bogus-pipeline on enhancing validity of self-reported adolescent drug use measures. J Sch Health 1987;57:232–6.

32. Needle R, McCubbin H, Lorence J, Hochhauser M. Reliability and validity of adolescent self-reported drug use in a family-based study: A methodological report. Int J Addict 1983;18: 901–12.
33. Brener ND, Kann L, McManus T, et al. Reliability of the 1999 Youth Risk Behavior Survey questionnaire. J Adolesc Health 2002;31:336–42.
34. Needle RH, Jou S, Su SS. The impact of changing methods of data collection on the reliability of self-reported drug use of adolescents. Am J Drug Alcohol Abuse 1989;15:275–89.
35. Patrick DL, Cheadle A, Thompson DC, et al. The validity of self-reported smoking: A review and meta-analysis. Am J Public Health 1994;84:1086–93.
36. Midanik LT. Validity of self-reported alcohol use: A literature review and assessment. Brit J Addict 1988;83:1019–30.
37. Cone EJ. New developments in biological measures of drug prevalence. The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates. NIDA Res Monogr 1997;167:108–30.
38. Akinci IH, Tarter RE, Kirisci L. Concordance between verbal report and urine screen of recent marijuana use in adolescents. Addict Behav 2001;26:613–9.
39. Murphy DA, Durako S, Muenz LR, et al. Marijuana use among HIV-positive and high-risk adolescents: A comparison of self-report through audio computer-assisted self-administered interviewing and urinalysis. Am J Epidemiol 2000;152:805–13.
40. Tourangeau R, Jobe JB, Pratt WF, et al. Design and results of the women's health study. Proceedings of the Annual Meeting of the American Statistical Association, Section on Survey Research Methods, Volume I; 1994:49–58.
41. Fisher M, Kupferman LB, Lesser M. Substance use in a school-based clinic population use of the randomized response technique to estimate prevalence. J Adolesc Health 1992;13:281–5.
42. Poulin C, MacNeil P, Mitic W. The validity of a province-wide student drug survey: Lessons in design. Can J Public Health 1993;84:259–64.
43. Stanton WR, McClelland M, Elwood C, et al. Prevalence, reliability, and bias of adolescents' reports of smoking and quitting. Addiction 1996;91:1705–14.
44. Pokorski TL, Chen WW, Bertholf RL. Use of urine cotinine to validate smoking self-reports in US Navy recruits. Addict Behav 1994;19:451–4.
45. Hedges B, Jarvis M. Cigarette smoking. In: Prescott-Clarke P, Primatesta P (eds). Health Survey for England: The Health of Young People, 1995–1997. London: The Stationery Office, 1998:191–221.
46. Brittingham A, Tourangeau R, Kay W. Reports of smoking in a national survey: Data from screening and detailed interviews, and from self- and interviewer-administered questions. Ann Epidemiol 1998;8:393–401.
47. Aguinis H, Pierce CA, Quigley BM. Conditions under which a bogus pipeline procedure enhances the validity of self-reported cigarette smoking. J Appl Soc Psychol 1993;23:352–73.
48. Murray DM, O'Connell CM, Schmid LA, Perry CL. The validity of smoking self-reports by adolescents: A reexamination of the bogus pipeline procedure. Addict Behav 1987; 12:7–15.
49. Akers RL, Massey J, Clarke W, Lauer RM. Are self-reports of adolescent deviance valid? Biochemical measures, randomized response, and the bogus pipeline in smoking behavior. Soc Forces 1983;62:234–51.
50. Bauman KE, Koch GG, Bryan ES, et al. On the measurement of tobacco use by adolescents: Validity of self-reports of smokeless tobacco use and validity of cotinine as an indicator of cigarette smoking. Am J Epidemiol 1989;130:327–37.
51. Biglan A, Ary DV. Methodological issues in research on smoking prevention. NIDA Res Monogr 1985;63:170–95.
52. Bauman KE, Ennett SE. Tobacco use by black and white adolescents: The validity of self-reports. Am J Public Health 1994;84:394–8.
53. Caraballo RS, Giovino GA, Pechacek TF. Self-reported cigarette smoking versus serum cotinine among U.S. adolescents. Nicotine Tob Res 2004 (in press).
54. Walsh MM, Ellison J, Hilton JF, et al. Spit (smokeless) tobacco use by high school baseball athletes in California. Tob Control 2000;9(Suppl II):ii32–9.
55. Velting DM, Rathus JH, Asnis GM. Asking adolescents to explain discrepancies in self-reported suicidality. Suicide Life Threat Behav 1998;28:187–96.
56. Klimes-Dougan B. Screening for suicidal ideation in children and adolescents: Methodological considerations. J Adolesc 1998;21:435–44.
57. Osman A, Barrios FX, Panak WF, et al. Validation of the Multi-Attitude Suicide Tendency Scale in adolescent samples. J Clin Psychol 1994;50:847–55.
58. De Man AF, Leduc CP. Validity and reliability of a self-report suicide ideation scale for use with adolescents. Soc Behav Pers 1994;22:261–6.
59. Garrison CZ, Lewinsohn PM, Marsteller F, et al. The assessment of suicidal behavior in adolescents. Suicide Life Threat Behav 1991;21:217–29.
60. Zulkifli SN, Yu SM. The food frequency method for dietary assessment. J Am Diet Assoc 1992;92:681–5.
61. Cavadini C, Decarli B, Dirren H, et al. Assessment of adolescent food habits in Switzerland. Appetite 1999;32:97–106.
62. Field AE, Colditz GA, Fox MK, et al. Comparison of 4 questionnaires for assessment of fruit and vegetable intake. Am J Public Health 1998;88:1216–8.
63. Frank GC, Nicklas TA, Webber LS, et al. A food frequency questionnaire for adolescents: Defining eating patterns. J Am Diet Assoc 1992;92:313–8.
64. Rockett HRH, Breitenbach M, Frazier AL, et al. Validation of a youth/adolescent food frequency questionnaire. Prev Med 1997;26:808–16.
65. Rosen JC, Poplawski D. The validity of self-reported weight loss and weight gain efforts in adolescents. Int J Eat Disord 1987;6:515–23.
66. French SA, Peterson CB, Story M, et al. Agreement between survey and interview measures of weight control practices in adolescents. Int J Eat Disord 1998;23:45–56.
67. Rockett HRH, Wolf AM, Colditz GA. Development and reproducibility of a food frequency questionnaire to assess diets of older children and adolescents. J Am Diet Assoc 1995;95:336–40.
68. Hill RJ, Davies PSW. The validity of self-reported energy intake as determined using the doubly labelled water technique. Br J Nutr 2001;85:415–30.
69. McPherson RS, Hoelscher DM, Alexander M, et al. Dietary assessment methods among school-aged children: Validity and reliability. Prev Med 2000;31:S11–33.
70. Livingstone MBE, Prentice AM, Coward WA, et al. Validation of estimates of energy intake by weighed dietary record and diet history in children and adolescents. Am J Clin Nutr 1992;56:29–35.

71. Schoeller DA. Limitations in the assessment of dietary energy intake by self-report. Metabolism 1995;44(Suppl 2):18–22.
72. Sallis JF, Buono MJ, Roby JJ, et al. Seven-day recall and other physical activity self-reports in children and adolescents. Med Sci Sports Exerc 1993;25:99–108.
73. Rifas-Shiman SL, Gillman MW, Field AE, et al. Comparing physical activity questionnaires for youth: Seasonal vs annual format. Am J Prev Med 2001;20:282–5.
74. Janz KF, Witt J, Mahoney LT. The stability of children's physical activity as measured by accelerometry and self-report. Med Sci Sports Exerc 1995;27:1326–32.
75. Weston AT, Petosa R, Pate RR. Validation of an instrument for measurement of physical activity in youth. Med Sci Sports Exerc 1997;29:138–43.
76. Booth ML, Okely AD, Chey T, Bauman A. The reliability and validity of the physical activity questions in the WHO health behaviour in schoolchildren (HSBC) survey: A population study. Br J Sports Med 2001;35:263–7.
77. Godin G, Shephard RJ. Normative beliefs of school children concerning regular exercise. J Sch Health 1984;54:443–5.
78. Sallis JF, Saelens BE. Assessment of physical activity by self-report: Status, limitations, and future directions. Res Q Exerc Sport 2000;71:1–14.
79. Kohl HW, Fulton JE, Caspersen CJ. Assessment of physical activity among children and adolescents: A review and synthesis. Prev Med 2000;31:S54–76.
80. LaPorte RE, Cauley JA, Kinsey CM, et al. The epidemiology of physical activity in children, college students, middle-aged men, menopausal females and monkeys. J Chronic Dis 1982; 35:787–95.
81. Richardson MT, Leon AS, Jacobs DR, Jr, et al. Ability of Caltrac accelerometer to assess daily physical activity levels. J Cardiopulm Rehabil 1995;15:107–13.
82. Janz KF, Golden JC, Hansen JR, Mahoney LT. Heart rate monitoring of physical activity in children and adolescents: The Muscatine study. Pediatrics 1992;89:256–61.
83. Bouchard C, Tremblay A, Leblanc C, et al. A method to assess energy expenditure in children and adults. Am J Clin Nutr 1983;37:461–7.
84. Murphy JK, Alpert BS, Dupaul LM, et al. The validity of children's self-reports of physical activity: A preliminary study. J Hum Hypertens 1990;4:130–2.
85. Rogers R, Reybrouck T, Weymans M, et al. Reliability of subjective estimates of exercise capacity after total repair of tetralogy of Fallot. Acta Paediatr 1994;83:866–9.
86. Grunbaum JA, Gingiss P, Orpinas P, et al. A comprehensive approach to school health program needs assessments. J Sch Health 1995;65:54–9.
87. Catania, JA, Turner H, Pierce RC, et al. Response bias in surveys of AIDS-related sexual behavior. In: Ostrow DG, Kessler RC (eds). Methodological Issues in AIDS Behavioral Research. New York: Plenum Press, 1993:133–162.
88. McFarlane M, St. Lawrence JS. Adolescents' recall of sexual behavior: Consistency of self-report and effect of variations in recall duration. J Adolesc Health 1999;25:199–206.
89. Ford K, Norris K. Methodological considerations for survey research on sexual behavior: Urban African American and Hispanic youth. J Sex Res 1991;28:539–55.
90. Siegel DM, Aten MJ, Roghmann KJ. Self-reported honesty among middle and high school students responding to a sexual behavior questionnaire. J Adolesc Health 1998;23: 20–8.
91. Davoli M, Perucci CA, Sangalli M, et al. Reliability of sexual behavior data among high school students in Rome. Epidemiol 1992;3:531–5.
92. Millstein SG, Irwin CE. Acceptability of computer-acquired sexual histories in adolescent girls. J Pediatr 1983;103:815–9.
93. Newcomer S, Udry JR. Adolescents' honesty in a survey of sexual behavior. J Adolesc Res 1988;3:419–23.
94. Rodgers JL, Billy JOG, Udry JR. The rescission of behaviors: Inconsistent responses in adolescent sexuality data. Soc Sci Res 1982;11:280–96.
95. Zelnik M, Kantner J, Ford K. Sex and Pregnancy in Adolescence. Beverly Hills, CA: Sage, 1981.
96. Catania JA, Gibson DR, Chitwood DD, et al. Methodological problems in AIDS behavioral research: Influences on measurement error and participation bias in studies of sexual behavior. Psychol Bull 1990;108:339–62.
97. Orr DP, Fortenberry JD, Blythe MJ. Validity of self-reported sexual behaviors in adolescent women using biomarker outcomes. Sex Transm Dis 1997;24:261–6.
98. Shew ML, Remafedi GJ, Bearinger LH, et al. The validity of self-reported condom use among adolescents. Sex Transm Dis 1997;24:503–10.
99. Clark LR, Brasseux C, Richmond D, et al. Are adolescents accurate in self-report of frequencies of sexually transmitted diseases and pregnancies? J Adolesc Health 1997;21:91–6.
100. Smith KW, McGraw SA, Crawford SL, et al. HIV risk among Latino adolescents in two New England cities. Am J Public Health 1993;83:1395–9.
101. Rosenthal R. The "file drawer" problem and tolerance for null results. Psychol Bull 1979;86:638–41.