

THEORY AND METHODS

Low response rate schools in surveys of adolescent risk taking behaviours: possible biases, possible solutions

B C Weitzman, S Guttmacher, S Weinberg, F Kapadia

J Epidemiol Community Health 2003;57:63–67

See end of article for authors' affiliations

Correspondence to:
Dr B C Weitzman, NYU
Wagner Graduate School,
Tisch Hall–Room 600, 40
West 4th Street, New York,
NY 10012, USA;
beth.weitzman@nyu.edu

Accepted for publication
30 April 2002

Study objective: To examine the potential biases introduced when students in low response rate schools are dropped from classroom based surveys of adolescent risk taking behaviour.

Design: Self administered confidential surveys were conducted in classrooms, with follow up visits to each school to survey students absent during the initial survey administration. Data on students in schools that achieved a 70% response rate are compared with data on students in schools that did not achieve this level of response.

Setting: New York City, United States.

Participants: 1854 10th graders in 13 public (state supported) high schools.

Main results: Students in schools with low response rates resulting from high rates of absenteeism have different demographic characteristics and engage in more risk behaviours than students in schools with low absenteeism and high response rates. Excluding schools with low rates of response can have an effect on estimates of risk behaviour, even after data are weighted for individual absences. The potential for bias is greatest when, in sampling schools, the proportion of schools with low response rates is large, and when such schools represent a large share of the students in the area under study.

Conclusions: Excluding schools with poor response rates from survey samples using a classroom based approach does not improve, and may, under some circumstances, underestimate risky behaviour among adolescent populations.

School based research is a common method for estimating and exploring adolescent risk taking behaviour. Sampling youth in schools provides a comparatively inexpensive method for obtaining large samples. Classroom based surveys, whereby stratified samples of schools, classrooms, and students comprise the final sample, have been used in a wide variety of countries to estimate the extent of many different conditions and behaviours. Examples include corticosteroids in Sweden,¹ physical fitness in the United Kingdom,² hypertension in Mexico,³ and tobacco use in Hong Kong.⁴ Perhaps the most ambitious of these studies is the World Health Organisation's cross national Health Behaviour in School-Aged Children survey (HBSC), which is conducted every four years and now includes 32 nations.⁵ Funding for the survey is secured at the national level and effort is made to ensure that sampling procedures, at the school and classroom level, are sufficiently consistent across countries to allow for valid comparison. Teachers, who administer the questionnaire, are encouraged to administer "make-up" surveys to those students absent on the original day of testing (R Smith, personal communication).

In the United States, several national studies of youth risk behaviour select a stratified sample of school districts, and school buildings within them, on the basis of geographical location, demographic characteristics, and willingness to participate.^{6,7} They then survey those students present in designated classes at a specific time on a given day; in some cases, absent students are surveyed in the same class on a second day. Previous research has demonstrated that school based studies provide a biased estimate of adolescent risk taking behaviour because they do not include those students who have dropped out of school.⁸ Similarly, the effectiveness of using a school based strategy to provide estimates of risk taking behaviour among students may be hindered by absenteeism.

In many schools, particularly those serving low income youth, one day or two day survey protocols have the potential

for considerable bias by missing a significant portion of the student sample. For example, the official average *daily* school absenteeism rate (that is, based on homeroom attendance) for New York City (NYC) public high schools ranges from 15% to 30%.⁹ Absenteeism from single classes ("skipping" or "cutting" class) exacerbates this problem. Furthermore, absenteeism can take on a longstanding nature in the form of truancy. As a result of these various forms of absenteeism, estimates of risk behaviour based on classroom samples may not be representative of the urban adolescent population. To the degree that absenteeism and risk behaviour are correlated, as evidenced by previous research, school based samples may underestimate risk taking behaviour.^{7,10–12}

The problem of student absenteeism is not spread randomly across schools or school systems. In fact, there is enormous variation in rates of attendance across schools and school districts.^{13,14} Underlying student characteristics, such as poverty, may contribute to these variations in attendance. For example, poor students may experience more chronic illness or feel more pressure to work for pay. On the other hand, variations in school policies and performance may also contribute to differences in student attendance. Such policies may deal directly with the problem of absenteeism or they may concern other related matters such as school safety or academic enrichment. In using cluster based sampling methods to draw a student sample, researchers must grapple with the question of how to handle absenteeism, and its effect on response rates, at the level of both the individual respondent and the entire school.

Studies vary considerably in their approach to the problem of bias resulting from student absenteeism. Firstly, studies may make repeated attempts to reach students absent during the survey administration. For example, in HBSC as already noted, classroom teachers typically administer the survey and they are encouraged to conduct make ups (R Smith, personal communication), but no adjustments are made beyond that effort. In contrast, other studies weight the data according to

self reported absences.¹⁵ In this approach, adjustments are made at the individual student level where participating students with high rates of absences are weighted to “represent” those students who have not responded.

Another approach has been to eliminate from the sample those schools with low response rates (often less than 70%) and replace them with demographically similar schools where response rates are higher.¹⁶ That is, the researchers aim to obtain a representative sample at the school level, in order to overcome the potential bias of absenteeism. While this method helps to reduce bias at the school level, it may result in further bias for the city or national sample. If risk taking among students in schools with low response rates (and high absenteeism) is different from those in high response rate schools, limiting the sample to those schools with high response rates may exacerbate, rather than reduce, the bias and probably underestimate the extent of risky behaviour.

This study is designed to determine the extent to which risk behaviour among students in low response rate schools differs from that in high response schools, and to consider whether bias is increased, or reduced, by the exclusion of low response rate schools in school based studies. Using data from one large urban school district in the United States, the paper asks:

(1) Do students in low attendance (low response rate) schools differ on a variety of demographic and behavioural variables from students in high response rate schools?

(2) If such differences exist, does exclusion of low response rate schools bias the overall estimates?

METHOD

Sample selection

To provide demographic, socioeconomic, and academic diversity among the study sample, 13 public (state supported) high schools were chosen from a population of 114 “comprehensive” and “vocational” New York City high schools. The schools were stratified by type (vocational or comprehensive) and degree of absenteeism and then randomly selected within these strata. Seven “comprehensive” high schools were included; these schools are typically local high schools without special entrance criteria. Two additional “educational options” schools requiring special selection procedures were included in the sample as were four “vocational high schools”, which train students for careers in particular industries. Half of the schools within each category were selected because they had been identified as having particularly high rates of absenteeism. Tenth graders were selected because, in the United States, the majority are 15 years of age, and thus too young for legal self exemption, or dropping out without parental consent, from school. Human subjects approval was obtained both from New York University and from the New York City Board of Education. A pilot study was conducted in two additional schools in the autumn, 1997, and full field operations were put in place in spring, 1998.

Survey design

Risk behaviours considered in this study include drug and alcohol use, absenteeism, academic achievement, sexual activity, and weapons possession. Questions were drawn from several previous US studies of adolescent risk behaviour, including the CDC’s Youth Risk Behaviour Survey, University of Michigan’s Monitoring the Future Study, the New York State Office of Alcoholism and Substance Abuse Services’ Study of Youth-At-Risk, and the Robert Wood Johnson Foundation funded evaluation of the New York City Public Schools HIV/AIDS Education Program. All selected questions were validated and field tested, and were written for high school students. Wording for the relevant study questions is provided in box 1.

Research design

The data presented in this paper were gathered as part of a larger study examining the impact of student absenteeism on

Box 1 Wording for selected survey questions. New York University Study, Student absenteeism and measurement of risk behaviour

Answer categories are in italics

During the last four weeks, how many whole days of school have you missed because you skipped or “cut” (for no particular reason)?

None, 1 day, 2 days, 3 days, 4 to 5 days, 6 to 10 days, 11 or more

During the last four weeks, how many times did you go to school, but cut a class when you weren’t supposed to?

Not at all, 1 or 2 times, 3-5 times, 6-10 times, 11-20 times, More than 20 times

Which of the following best describes your grades last semester?

Mostly A’s (90-100), Mostly B’s (80-89), Mostly C’s (70-79), Mostly D’s (65-69), Mostly F’s (64 and Below)

During the last four weeks, how many times did you:

... smoke cigarettes?

... drink alcoholic beverages (liquor, beer, wine)?

... smoke marijuana or pot?

Never, Hardly, Sometimes, Often, Very Often

During the past 30 days, on how many days did you carry a weapon such as a knife or box-cutter on school property?

0 days, 1 day, 2 or 3 days, 4 or 5 days, 6 or more days

Have you ever had sexual intercourse?

Yes, No

Please mark the statement that best applies to you:

I often take risks that might give me HIV/AIDS.

I sometimes take risks that might give me HIV/AIDS.

I never take risks that might give me HIV/AIDS.

estimates of youth risk behaviour. The larger study used a four stage data collection strategy, in which each stage represented an increased level of intensity and associated resources. By virtue of the study design, those who were surveyed in the later stages were more likely to be absent from school or to frequently skip classes than those students surveyed in the earlier stages.¹⁵ The four stages were implemented sequentially without overlap; students surveyed in stage 1, for example, were not included in stages 2, 3, or 4. Both stages 1 and 2 were classroom based, and replicate the survey methods used in previous studies on adolescent risk taking behaviour. Stages 3 and 4, took place outside of the classroom (usually an office or library), incorporated financial incentives, and represented a substantial departure from previous research in the field. For maximum comparability to survey approaches typically used in the study of adolescent risk taking behaviour, only data on students surveyed in stages 1 and 2 are included in this paper. Findings concerning the comparison of stages 1 and 2 respondents with those of stages 3 and 4 are to be found in Guttmacher *et al.*¹⁵

Of note, despite the introduction of financial incentives (gift certificates to local stores) and the use of more intensive and costly efforts to reach students who were not in the classroom during the survey administration (for example, letters inviting participation in the survey were sent to the homes of all non-respondents), the increase in the number of student respondents was comparatively small. The initial sample included 2675 students; 1921 were interviewed in the classroom setting, 128 were interviewed as a result of the more intensive outreach efforts, and 626 students (23% of the original sample) were never interviewed. The findings from this earlier study suggest that the increased expense associated with more intensive methods of outreach for classroom based studies may not be justified by improved estimates as comparatively few additional students are surveyed. As classroom based surveys are typically undertaken to reach large numbers of students at comparatively low cost, the level of incentive that

Table 1 Demographic characteristics and risk behaviour for students in high and low response rate schools

| Characteristic | Schools by response rate | | | High response (%) (weighted) Ni=1397 Ns=9 | All schools (%) (weighted) Ni=1854 Ns=13 |
|-------------------------------------|---|---|------------------------------|---|--|
| | High response (%) (unweighted) Ni=1397 Ns=9 | Low response (%) (unweighted) Ni=457 Ns=4 | Significance (two tailed) | | |
| Gender | | | | | |
| Female | 49 | 63 | 0.000 | 49 | 53 |
| Male | 51 | 37 | | 51 | 47 |
| Age | | | | | |
| 16 or younger | 88 | 86 | 0.240 | 87 | 88 |
| 17 or older | 12 | 14 | | 13 | 13 |
| Ethnicity | | | | | |
| African-American | 28 | 35 | 0.000 | 28 | 30 |
| Caribbean-American | 27 | 16 | | 26 | 23 |
| Asian | 8 | 0.2 | | 8 | 6 |
| Hispanic/Latino | 24 | 42 | | 26 | 30 |
| White (including other) | 10 | 5 | | 11 | 9 |
| Household composition | | | | | |
| Both parents | 36 | 18 | 0.000 | 35 | 30 |
| Other relative | 8 | 12 | | 8 | 10 |
| Single parent | 55 | 70 | | 56 | 60 |
| Non-legitimate absences | | | | | |
| 0 | 74 | 63 | 0.000 | 71 | 69 |
| 1–3 | 19 | 26 | | 21 | 22 |
| 4 or more | 7 | 12 | | 9 | 10 |
| Number of classes cut/skipped | | | | | |
| 0 | 42 | 30 | 0.000 | 40 | 37 |
| 1–5 | 45 | 52 | | 46 | 48 |
| 6 or more | 13 | 19 | | 14 | 16 |
| Grades | | | | | |
| As or Bs | 39 | 29 | 0.001 | 37 | 34 |
| Cs | 37 | 44 | | 37 | 39 |
| Ds or Fs | 24 | 27 | | 27 | 27 |
| Cigarette use in past 4 weeks | | | | | |
| % yes | 14 | 19 | 0.020 | 15 | 16 |
| Marijuana use in past 4 weeks | | | | | |
| % yes | 15 | 21 | 0.002 | 16 | 17 |
| Alcohol use in past 4 weeks | | | | | |
| % yes | 36 | 35 | 0.820 | 36 | 36 |
| Sexually active | | | | | |
| % yes | 50 | 53 | 0.288 | 51 | 52 |
| Involved in HIV high risk behaviour | | | | | |
| Sometimes—often | 13 | 17 | 0.025 | 13 | 14 |
| Carried weapons to school | | | | | |
| 1 or more days | 12 | 14 | 0.365 | 13 | 13 |

Ni, number of individuals; Ns, number of schools; High response, students surveyed in stage 1 or 2 in schools achieving a 70% response rate after stage 1 or 2; Low response, students surveyed in stage 1 or 2 in schools not achieving a 70% response rate after stage 2.

may be needed to attract frequently absent students is not probably affordable in most studies.

To differentiate respondents from non-respondents, surveys were kept confidential but not anonymous. Trained data collectors, who administered the survey, reaffirmed the confidential nature of the survey. As almost all research subjects were under the age of 18, both student and parental consent were required before survey administration. About 4% of parents/guardians requested that their child be exempted from the study.

The demographic characteristics of the final study sample closely approximate that of the New York City public high schools.⁹ Slightly more than half of the sample was female and 87% of the sample was between 15 and 16 years of age, which is the appropriate age for 10th grade students. There were approximately equal proportions of African-Americans, Latinos, and Caribbean-Americans, whereas Asians and white groups, combined, accounted for 15% of the students. Ninety four per cent of all respondents surveyed in either stage 1 or 2 (n=1854) are included in this paper.

Analysis

To understand the impact of excluding low response rate schools, schools were categorised as having either high or low response rates according to the proportion of students who were surveyed by the end of stage 2. Schools achieving a 70%

or greater response rate were categorised as “high response schools,” whereas those not achieving a 70% response rate by the second stage of follow up were categorised as “low response schools.” Nine schools met the criterion for high response (n=1397) and four schools met the criterion for low response (n=457). The cut off points were chosen to reflect the inclusion criterion used in some other adolescent health surveys used in the United States. To keep these analyses comparable to those used in other studies, only students from stage 1 were included from schools where a 70% response was achieved after stage 1. Students from stages 1 and 2 were included from the other schools. Difference of proportions tests (χ^2 and *t* tests, as appropriate) were used to test for the statistical significance of the differences between students in low and high response rate schools. Data were then weighted on the basis of student attendance; the weighting protocol assumes that the probability of a student being surveyed is inversely related to the proportion of days absent as reported by the student. Weighted estimates of risk behaviour—for the inclusion and exclusion groups—were compared.

RESULTS

Sample demographics and risk behaviours

In table 1, we see that students surveyed in low response rate schools are more likely to be female ($p<0.001$), African-American ($p<0.001$), and from single parent or non-parent

households ($p < 0.001$). As would be expected, students in high attendance schools are far less likely to report non-legitimate absences or class skipping; for example 74% of students in high attendance schools report never skipping classes, as compared with 30% in low attendance schools ($p < 0.001$). Students in low attendance schools also report far weaker student performance. Whereas 39% of students in high attendance schools report receiving mostly grades A and B, only 29% of students in low attendance schools report such achievement ($p = 0.001$).

There are also variations in risk behaviour across school groups. In particular, reports of cigarette use, marijuana use, and HIV risk behaviour are higher for students in low attendance schools. Only 15% of students in high attendance schools reported marijuana use in the past four weeks, while 21% of students in low attendance school reported such use ($p = 0.002$). Similarly, 19% of youth in low response rate schools had smoked cigarettes in the previous four weeks, as compared with 14% in high response rate schools ($p = 0.020$). Seventeen per cent of students in low response rate schools indicated that they engaged in behaviours that put them at high risk of HIV, as compared with only 13% in high response rate schools ($p = 0.025$).

Despite these differences, limiting the sample to high response rate schools did not substantially change the estimates. When we compare the unweighted data for students in the high response schools, we find a marijuana use rate of 15% as compared with 16.5% among students in all schools. Once the data are weighted for individual student absences, we find very small but consistent increases in the estimates of risk behaviour. The estimate of marijuana use rises to 17% when students from high and low response rate schools are included and the data are weighted for individual absences. While these differences across samples and methods are not statistically significant, the direction is consistent; when low response rate schools are eliminated and when the data are not weighted for absences, there is a risk that estimates of risk behaviour will be too low.

In this particular case, the ratio of high to low response rate schools was 9:4, and the ratio of survey respondents from high to low response rate schools was about 3:1. If, however, the split between low and high response rate schools had been more balanced, or the proportion of respondents from low response rate schools greater, the bias introduced by eliminating low response rate schools would have been greater.

DISCUSSION

Eliminating schools that have not achieved a 70% response rate does not seem to be helpful to estimating risk behaviours among youths. Norms of appropriate behaviour related to school attendance seem to differ from school to school and schools with lower rates of attendance include students who are more likely to engage in a number of risk behaviours. Any method that systematically rejects schools with high rates of absenteeism is also systematically rejecting those schools with highest concentrations of risk behaviour.

In school systems where there is considerable absenteeism, weighting the data on the basis of individual self reported absences seems to improve the overall estimates of risk behaviour and should be routinely used in school based survey research. Currently many studies, including the national YRBS in the United States and the international HBSC, do not follow this protocol and, as a result, they may be underestimating certain behaviours. In addition, the practice of eliminating low response rate schools does not seem to have “pay off” and could result in an underestimate of risk behaviours, especially in those school systems where the differences between low and high attendance schools are large and the proportion of low attendance schools are great. As evidence suggests that students not surveyed in low attendance schools (that is,

Key points

- In classroom based surveys, absenteeism presents a source of bias when estimating adolescent risk behaviour.
- In New York City, cigarette and marijuana use are more prevalent in schools with high rates of absenteeism.
- In New York City, HIV risks are more common among students in high absence schools than students in low absence schools.
- Weighting survey data on the basis of self reports of absenteeism may reduce bias in estimating risk behaviours among adolescents.
- Schools with low response rates should not be excluded from survey samples.

those not “captured” by the survey because of long term absence) are likely to engage in even more risk behaviours than their surveyed counterparts, this bias is compounded.^{8 11 15} While on the surface it might seem that eliminating schools with poor response rates would strengthen the estimates, we find no evidence of a beneficial effect and some suggestion of possible detriment. Most notably, the youth in these schools, who seem to be the ones with the greatest degree of academic problems, are of particular importance to the study of risk behaviour and should not be systematically ignored by the field.

Survey researchers interested in adolescent risk behaviour are faced with a difficult situation. Absenteeism and risk behaviour seem to correlate; schools and communities with higher rates of absenteeism also tend to experience higher rates of risk taking behaviour. Yet, in those localities where rates of absenteeism are greatest, classroom based surveys may be of limited value because they do not include students who are rarely in the classroom. Various attempts to reduce this potential source of bias by intensifying the outreach effort have proved weak or ineffective.¹⁵ Furthermore, these intensive efforts can be expensive and time consuming, thus reducing the cost-saving benefits of classroom based studies. In non-school based studies of adolescent risk behaviour, more intensive efforts to track students have been effective, albeit expensive and time consuming.¹⁷ Of course, the use of more intensive efforts (such as financial incentives) may further bias the study pool if they are of greater appeal to one group than another. As researchers attempt to sort out these competing risks of bias, policy makers and officials in adolescent serving agencies must recognise that current estimates of risk behaviour, typically based on classroom surveys, may underestimate the extent of such behaviour among all adolescents. Programme development should recognise that adolescents who attend school erratically or sporadically may have different, and greater, needs for intervention than those who are included in classroom based surveys. If our primary interest is the study of those adolescents who engage in the most risky behaviours, simple classroom based surveys may not be the most appropriate method.

Authors' affiliations

B C Weitzman, New York University Robert F Wagner Graduate School of Public Service, USA

S Guttmacher, S Weinberg, New York University School of Education, USA

F Kapadia, Center for Urban Epidemiologic Studies, New York Academy of Medicine, USA

Funding: this research was funded by a grant from the United States National Institute of Drug Abuse (1R01DA09549-01A1).

Conflicts of interest: none.

REFERENCES

- 1 **Kindlundh A**, Isacson D, Berglund L, *et al*. Factors associated with adolescent use of doping agents; anabolic-androgenic steroids. [Abstract]. *Addiction* 1994;4:543–53.

- 2 **Bettiol H**, Rona RJ, Chinn S. Variation in physical fitness between ethnic groups in nine year olds. *Int J Epidemiol* 1999;**28**:281–6.
- 3 **Rosenthal J**, Labarthe DB, Selwyn BJ, *et al*. The Blood Pressure Study in Mexican Children (BPMC): I. Distribution and correlates of blood pressure in adolescent Mexican girls. *Int J Epidemiol* 1992;**21**:40–7.
- 4 **Peters J**, Hedley AJ, Lam TH, *et al*. A comprehensive study of smoking in primary school children in Hong Kong: implications for prevention. *J Epidemiol Community Health* 1997;**51**:239–45.
- 5 **Research Unit In Health and Behavioral Change**. *Health behaviour in school-aged children: a WHO cross-national survey research protocol for the 1997–98 Study*. Edinburgh: Research Unit In Health and Behavioral Change, University of Edinburgh Medical School, 1998. (<http://www.ruhbc.ed.ac.uk/hbsc/download/respro98.pdf>)
- 6 **Centers for Disease Control and Prevention (CDC)**. *Handbook for conducting youth risk behavior surveys*. Atlanta, GA: Division of Adolescence and School Health, Center for Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, 1993.
- 7 **Johnston LD**, O'Malley PM. Issues of validity and population coverage in student surveys of drug use. In: Rouse BA, Kozel NJ, Richards LG, eds. *Self-report methods of estimating drug use: meeting current challenges to validity*. NIDA Research Monograph 57. Rockville, MD: Department of Health and Human Services, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, 1985.
- 8 **Swaim RC**, Beauvais F, Chavez EL, *et al*. The effect of school dropout rates on estimates of adolescent substance use among three racial/ethnic groups. *Am J Public Health* 1997;**87**:51–5.
- 9 **New York City Board of Education**. *Annual Reports*. New York: New York City Board of Education, High School Division, 1997.
- 10 **Kandel D**. Reaching the hard-to-reach: illicit drug use among high school absentees. *Addictive Diseases* 1975;**4**:465–80.
- 11 **Michaud PA**, Delbos-Piot I, Narring F. Silent dropouts in health surveys: are nonrespondent absent teenagers different from those who participate in school-based health surveys. *J Adolesc Health* 1998;**22**:326–33.
- 12 **Rainone G**, Marel R, Aryan Z, *et al*. *Youth-at-risk in New York State: alcohol and drug use findings*. New York: New York State Office of Alcoholism and Substance Abuse Services, Bureau of Applied Studies, 1993.
- 13 **National Center for Educational Statistics (June 22, 1996)**. *Student absenteeism and tardiness (indicator of the month-June 1996)* (<http://nces.ed.gov/spider/web spider/96798.shtml>)
- 14 **New York City Board of Education**. *School profiles*. New York: New York City Board of Education, Division of Strategic Planning, 1993.
- 15 **Guttmacher S**, Weitzman BC, Kapadia F, *et al*. Classroom based surveys of adolescent risk-taking behaviors: reducing the risk of absenteeism. *Am J Public Health* 2002;**92**:235–237.
- 16 **O'Malley PM**. *Minutes from meeting of Risk and Absenteeism Project Advisors*. 23 Nov, 1998. Information available from the authors of the article.
- 17 **Morrison TC**, Wahlgren DR, Hovell MF, *et al*. Tracking and follow-up of 16,915 adolescents: minimizing attrition bias. *Control Clin Trials* 1991;**18**:383–96.