

GENERAL ARTICLE

Data Collection Procedures for School-Based Surveys Among Adolescents: The Youth in Europe Study

ALFGEIR LOGI KRISTJANSSON, PhD^{a,b} JON SIGFUSSON, MEd^c INGA DORA SIGFUSDOTTIR, PhD^{d,e} JOHN P. ALLEGRANTE, PhD^{f,g}

ABSTRACT

BACKGROUND: Collection of valid and reliable surveillance data as a basis for school health promotion and education policy and practice for children and adolescence is of great importance. However, numerous methodological and practical problems arise in the planning and collection of such survey data that need to be resolved in order to ensure the validity of the data and to maximize the response rate without being prohibitively costly.

METHOD: This article builds on a 15-yearlong experience of such annual data collections in Iceland and describes the preparation, process, and collection of data that provide a common methodologic framework for the school-based survey, Youth in Europe, a population-based survey of 14- to 16-year-old adolescents, being collected across 18 European cities now participating in the European Cities Against Drugs (ECAD) program.

RESULTS: We identified 11 critical steps for developing and implementing the surveys in light of the recent literature on the preparation and implementation practices in school-based data collection among adolescents.

CONCLUSION: Limiting the disruption of daily operations in schools while at the same time ensuring both quality and clarity of data collection procedures in school-based surveys are of paramount importance for researchers, school personnel, and students.

Keywords: adolescence; children; health surveys; schools; surveillance; survey methodology.

Citation: Kristjansson AL, Sigfusson J, Sigfusdottir ID, Allegrante JP. Data collection procedures for school-based surveys among adolescents: The Youth in Europe Study. *J Sch Health*. 2013; 83: 662-667.

Received on April 4, 2012

Accepted on December 2, 2012

Cross-sectional survey data are commonly collected on children and adolescents in schools for surveillance purposes and other scientific investigations. With regards to collecting sensitive health and behavior surveillance data, cross-sectional surveys are favored over longitudinal designs, as the latter pose the potential problem of compromised anonymity.¹ In addition, pencil-and-paper responses are still favored over computerized responses for the same reason.²

School-based surveys offer a number of benefits to investigators that have been summarized by Smit et al.³ (1) a large number of participants are captive and accessible in a formal setting and participation rates are usually high; (2) costs per respondent are low compared to other methods; (3) respondents from minority groups are accessible in sufficient numbers for analysis with power high enough for generalizing to the wider population of such demographic groups;

^aPostdoctoral Fellow, (kristjansson@tc.columbia.edu), Department of Health and Behavior Studies, Teachers College, Columbia University, 525 West 120th Street, New York, NY 10027.

^bXX, (kristjansson@tc.columbia.edu), Icelandic Centre for Social Research and Analysis, Reykjavik University, 1 Menntavegur, 105 Reykjavik, Iceland.

^cDirector, (jon@rannsoknir.is), Icelandic Centre for Social Research and Analysis, Reykjavik University, 1 Menntavegur, 105 Reykjavik, Iceland.

^dResearch Professor, (ingadora@ru.is), Department of Health and Behavior Studies, Teachers College, Columbia University, 525 West 120th street, New York, NY 10027.

^eXX, (ingadora@ru.is), Icelandic Centre for Social Research and Analysis, Reykjavik University, 1 Menntavegur, 105 Reykjavik, Iceland.

^fProfessor and Deputy Provost, (jpa1@columbia.edu), Department of Health and Behavior Studies, Teachers College, Columbia University, 525 West 120th Street, New York, NY 10027.

^gXX, (jpa1@columbia.edu), Department of Sociomedical Sciences, Mailman School of Public Health, 722 West 168 Street, New York, NY 10032.

Address correspondence to: Alfgeir Logi Kristjansson, Postdoctoral Fellow, (kristjansson@tc.columbia.edu), Department of Health and Behavior Studies, Teachers College, Columbia University, 525 West 120th Street, New York, NY 10027.

and (4) the school and/or class provide higher level contextual information that incorporates opportunities for cluster-based analyses such as nested models.⁴ In addition, schools are also common sites for health interventions (eg, sex and drug prevention education) and school surveys can be used to monitor such interventions.

School-based surveys, however, are subject to numerous methodological and practical problems that may arise in the preparation, planning, and data collection phases of surveys. These problems need to be resolved to maximize response rate without running the risk of introducing sample bias or jeopardizing the validity of data. Studies have shown that it has become increasingly difficult to recruit schools for participation in surveys⁵ and within-school nonresponse also has increased over time.⁶ High nonresponse rates create problems with population measurement validity and increase the likelihood of sample bias because whole school may represent a specific segment of society.⁷ Studies have shown the prevalence of health risk behaviors to be higher among absent than present respondents in school-based surveys where nonparticipants were followed after the initial data collection.^{8,9}

In line with the overview above, this narrative article describes the preparation and procedures of data collection in the school-based survey, Youth in Europe. Our article builds on a 15-yearlong experience in collecting such data in annual cross-sectional survey in Iceland and describes 11 critical steps in the process. We have organized our discussion of the steps in the process in chronological order. Although these steps are what we have found works best for us in our own studies, readers should bear in mind that the time intervals may of course need to be adjusted in other contexts due to institutional complexity, access to registry information, geographical distances, the nature and function of postal services, and political considerations. This article focuses primarily on the practical issues in the preparation and process of data collection in the school setting; we do not discuss questionnaire design and/or sampling techniques that are also of importance.

STEPS IN PLANNING AND COLLECTING DATA AMONG CHILDREN AND ADOLESCENTS IN SCHOOLS

We have followed an 11-step process when planning and collecting data among adolescents in the annual Youth in Iceland surveys that have now been conducted over 15 years.

Step 1: Obtain Institutional Approval for Use of Human Subjects

Institutional review board (IRB)/bioethics committee approval for the proposed study is obtained, if

needed, 2-3 months before planned data collection. This is required now in most countries for anonymous data collection with underaged minors in schools. Obtaining approval requires that all instructions for participation and all parental and subject informed consent protocols should be prepared, pilot-tested, and finalized.

Step 2: Determine Eligible Schools and Potential Sample Sizes

This step is completed about 1-1/2 months prior to data collection. Data are gathered on the sampled school principal (name, address, title, etc) as well as information on the number of students in each class and school. School-based surveys are usually carried out with whole classes or whole cohorts within schools. Before any information or letter of intent is sent to school principals the research team acquires information about the number of potential participants in each class and/or cohort that will be asked to participate in the study. This information is usually kept by a local statistical bureau or a similar institution and is important in both the planning stage and the data collection.

Step 3: Conduct Community Pre-Study Notification

About 1 month before data collection, a short introductory letter about the planned survey is prepared and mailed to school principals. The purpose of this letter is to provide the principal with a brief summary of the aims of the survey (about a half a page), the research team, and reasons for the proposed data collection. It is important to make sure that institutional headers on the letter are included and that the language used is sufficiently appropriate. If official bodies (eg, Ministry of Education, municipal offices, local councils) are participating in the survey or provide support for it, the nature of their respective involvement should be acknowledged in the letter and the appropriate logo included.

Step 4: Solicit Community Study Participation

Approximately 3 weeks prior to data collection, a more detailed letter about the planned survey is mailed to school principals and they are asked to permit the collection of anonymous data in their school. The narrative in this letter is connected to the one in Step 3, so that the principal will know this is the same survey about which he/she had received notification. It is appropriate to use the same header and/or institutional titles as before. The letter should be succinct but underline the most important factors about the study such as who is collecting the data and for what reason and how many participants in that particular school are needed (information gathered in Step 2). Also, it is important to inform the community that the

survey is carried out in accordance with local laws and regulations and has been approved by a local human subjects committee (if required by law). Potential benefits of the survey and its findings for improving the health and welfare of children and adolescents should be underlined, and compensation to the school and principal for allowing the data collection introduced. As a way of offering tangible gratitude to participating schools, researchers often compensate schools with a summary of the school-level findings. This kind of feedback has been shown to be of great value to school principals.¹⁰ Finally, principals are asked to nominate a supervising contact agent (SCA) within their school and include information about how best the research team can communicate with that individual (ie, phone, e-mail address, etc). By offering to compensate each school with a report of the local-level findings, the principal and the SCA will become directly involved in the data collection process and partly responsible for doing everything possible to ensure a high response rate. Such participation in the survey is almost always viewed favorably by the school principals and other authorities.

Step 5: Secure School Principal Support

Within 1 week of completing Step 4, the introductory letter should be followed with a short, courtesy phone call to the principal. This is important not only as a courtesy but also to confirm the contact information about the SCA. The phone call is made regardless of whether a response has been received from the principal or not. Sometimes the principals or school representatives have questions about certain aspects of the study at this point and it becomes important to exhaust any questions they have in a clear-cut manner. Usually these questions involve only minor issues about how to collect the data (because most principals want to do their part in the survey as well as they can) or whether they are allowed to move the time of data collection by a few days in the event of changes in school calendar (ie, school trips or other planned activities). If, on the other hand, the principal or relevant authority has not yet received the letter by the time of this phone call, we ask that individual to take a few minutes to read the letter and then we call again 1-2 days later. It needs to be affirmed that the research team has talked directly to the principal or other authority in each and every school participating in the survey. In our experience, with about 160 schools and annual data collections, it takes between 7 and 10 days from the submission of the introductory letter to confirm the participation of all the schools. Immediately after the nomination of each school SCA, that individual should be contacted for confirmation. Thereafter, all communications and notifications about all steps in the survey process with each school should be through the representative SCA.

Step 6: Identify and Contact School SCA

About 10-12 days prior to data collection, each school SCA is contacted and sent information about the classes that are asked to participate in the survey. The SCA is asked to confirm the class size and label and informed that the research team is about to prepare a box containing questionnaires, envelopes, instructions, and other survey-related materials that will be delivered to the school by the local postal service. In our experience acquiring a confirmation from all the SCA's may take up to a week, and sometimes more, and should be followed up in the same manner as the communication with each principal in Step 5.

Step 7: Prepare Survey Materials for Each Eligible School Class

A list of all participating classes with information about the number of students and class label should now be ready. We now prepare a file for our own in-house use that contains all information about each school, including address, principal, SCA, number of classes and their label, and number of registered students. This file will be important when the completed survey questionnaires begin to arrive back from the schools for the assessment of response rates as well as for data cleaning. From this list, and about 1 week prior to initiating data collection, we prepare a box to each school that contains a letter with instructions for the SCA and 1 large envelope marked for each class. Each class envelope should contain a matching number of questionnaires, pre-sealable envelopes for the return of each questionnaire, introductory letters to parents (in our case, based on passive consent) in line with the number of registered students (information from Step 2), and 2 additional such kits; one for the person supervising the data collection in each class (usually the respective class teacher) and another if additional student might have been added to the class but information thereof that has not been available from the appropriate Statistical Bureau. The box should also contain instructions for the person overseeing the data collection within each class and a sticker with the data collector's address and postcode for the SCA to return the box back once the questionnaires have been filled out by the students. In our experience, we believe it is good practice and ensures quality to train the packing people for about an hour before packing the material for each school and to instruct them to work only with one school at a time to prevent any mixing of materials between the school boxes.

Step 8. Mail Survey Materials to Each School

The prepared boxes are delivered to the mail service approximately 4-5 days before the onset of data collection. Receipts thereof are kept by the research

team for every box in every school in order to be able to trace the provenance of each box in the event that something unexpected comes up with the mail service. This enables the research team to acquire information from the postal service about the delivery of each box to each school. The delivery is followed with a phone call to each SCA for confirmation of receipt. Because the aim is to collect the data on the same day in all schools this becomes an important step so that the research team can trace each box and check if they left the postal service and have been delivered. Sometimes the SCA within a given school contacts us and claims that the box has not yet arrived. This may happen when someone other than the SCA has taken delivery of the box inside the school and put it away without notifying the SCA of its delivery.

Step 9: Distribute Consent Forms to Parents

Three days before data collection begins, each school SCA takes care that the proposed participants are sent home with an introductory letter to parents about the survey and offered the opportunity to withdraw their children from the study or to contact the research team with any questions they might have. This approach to informed consent is consistent with local laws and the principle of passive consent.¹¹ The letter should be written in clear, understandable language, with attention to the appropriate level of literacy required for general understanding, and includes the same relevant information as contained in the letter used in Step 4. Depending on the content, it is likely to differ between studies how many parents might ask for additional information about the survey and how many might wish to withdraw their children from participating. With a health and behavior survey such as that in our case, this has turned out to be rare, with usually fewer than 10 parents of the approximately 7500 children we survey every year withdrawing their children from participation.

Step 10: SCA Reminder

The day before data collection or early during the same day, the SCA is contacted through e-mail with a friendly reminder about the task of the coming day. The task of the SCA is to oversee that the class envelopes are delivered to each class during that same day. Usually a member of school personnel (teacher, teaching assistant, sometimes the SCA) oversees the distribution of questionnaires and the pre-sealed envelopes to each student within a given class room. The survey should be completed by participants within one consecutive class session with tables set up in the same manner as in exams to prevent students from overlooking at each other's responses. This method of administration and data collection was tested

for validity and response distribution in the earliest Youth in Iceland surveys and found to be robust and without distributional differences depending on the position of the individuals overseeing the data collection within the classes.¹² We prefer to work with a pencil-and-paper questionnaire format as previous research has shown it to be more reliable than computerized methods used in the school setting.² Should students need assistance with understanding questions, the overseeing member of the school should bring a blank questionnaire to the respective student to point out the question that needs clarification so as to protect the student's other responses marked to his/her questionnaire from being viewed by the authority. This procedure is also outlined in the instructions to students inside each questionnaire. After the questionnaires have been collected from the students, the SCA then collects and seals each class envelope and puts them into the respective school box, seals it, places the return mail-address sticker on it for the mail service, and delivers it to the relevant party to be posted back to the research team.

Step 11. Distribute Letters of Appreciation for Participation

Finally, once the survey is complete and the questionnaires containing the raw data have been returned, within 2 weeks each school receives a letter of gratitude for their assistance and participation in the survey. The school SCAs and principals (sometimes the same individual) are invited to point out any problems that may have occurred in the planning or data collection process of the survey and to make suggestions for improvements. If compensations are provided, such as the school-level findings, this is also an opportunity to underline that benefit of having participated in the survey. If, however, at this point the school box has not yet been returned from via post mail to the research team, then this can be addressed in a telephone call to the school and the location of the box traced.

When the boxes containing the questionnaires are received back from the schools they need to be handled with great care since the raw data comprises many thousands of pages with tick-marked boxes. In case of the Youth in Europe surveys, we anticipate the use of optical scanning to import and transform the raw data into a digitized format. This greatly decreases the likelihood of misidentification due to misspelling. When the boxes with the questionnaires are handled we realized that it is advisable to work with only one school at a time to prevent mixing between schools. The keel is cut off each questionnaire and a school and individual number stamped on the front of each one in order to distinguish between schools and individuals within the schools before scanning the questionnaires.

Once all the data are in a digitized format, the data cleaning process can commence.

DISCUSSION

We have learned that keeping the school SCAs well informed at all stages and ensuring their involvement is of key importance in the process of acquiring high-quality data. The planning and packing of precisely the right number of questionnaires, pre-sealable envelopes, and information letters to parents in a correctly labeled class envelope in each school box is important to save time and to minimize potential complications when the questionnaires are distributed among students in each class. The principals and SCAs appreciate this effort greatly. This is consistent with Gottfredson and Gottfredson¹³ whose work has suggested that to utilize data from the school setting more effectively, researchers should make an effort to include the school authorities in the planning and implementation of data collection and utilization.

Many studies have been published in recent years regarding potential problems with different elements of data collection in school-based surveys. The most pressing challenges to researchers are to maintain data quality, obtain high response rates, and collect data from unbiased samples. Increasing school refusal is therefore of particular concern.⁵ Further, studies have found a difference between low-response rate schools and high-response rate schools in surveys of risk taking behaviors, which calls into question the issue of excluding schools with low responses as is commonly done.¹⁴ Another issue that has been highlighted and discussed in several studies is the concern about falling response rates and increased risk of sample bias due to passive versus active (written) parental consent.^{11,15,16} Unger et al¹⁷ tested the difference between the 2 consent methods and found a significant difference in the number of boys versus girls, with proportionally more boys, African Americans, students with poor grades, and smokers being in the passive consent part of the sample. Further, O'Donnell et al¹⁸ claim that parental consent facilitates benefits in the form of increased community awareness and involvement in research and evaluation programs but also that it is resource consuming and costly in relation to passive consent requirements.

Another area of methodologic interest is that of computerized methods of data collection, which have become increasingly popular in recent years.¹⁹ They provide an interesting alternative that decreases the time from collection until the data have been digitized and ready for analyses, as well as erasing the issue of misidentification due to misspelling. However, like other modes of data collection these methods are not without potential problems. Yetter et al² studied a Web-based versus paper-and-pencil approach to data

collection and found a substantially lower response rate in the Web-based study; Trapl et al²⁰ proposed to use an audio-enhanced personal digital assistant in their school-based data collection. Their conclusion was that the system was well received by students, decreased time of completion, and reduced missing data. With regard to self-responding (commonly acknowledged as a limitation in self-reported surveys), Nordahl et al²¹ studied the question content in relation to the respondent's age and found that students as young as 11 years were able to provide valid self-reported responses to questions concerning their background and languages.

Data collection in schools using cross-sectional surveys is a common method of social research. Many factors need to be addressed in the process, all with the intention of maximizing response rates while minimizing the likelihood of bias in the data and to maintain cost-effectiveness. This brief overview of the preparation procedures and data collection process that we have used in the Youth in Iceland surveys in the 15-yearlong tradition of such annual data collections in Iceland²² provides a common set of steps by which the Youth in Europe survey can be conducted to ensure uniformity in methodology.

REFERENCES

1. Bjarnason T, Adilbjarnardottir S. Anonymity and confidentiality in school surveys on alcohol, tobacco, and cannabis use. *J Drug Issues*. 2000;30(2):335-343.
2. Yetter G, Capaccioli K. Differences in responses to Web and paper surveys among school professionals. *Behav Res Methods*. 2010;42(1):266-272.
3. Smit F, De Zwart W, Spruit I. Monitoring substance use in adolescents: school survey or household survey? *Drugs*. 2002;9(3):267-274.
4. Snijders TA, Bosker RJ. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage; 1999.
5. Hallfors D, Iritani B. Local and state school-based substance use surveys—availability, content, and quality. *Eval Rev*. 2002;26(4):418-437.
6. White DA, Morris AJ, Hill KB, Bradnock G. Consent and school-based surveys. *Br Dent J*. 2007;202(12):715-717.
7. Thorlindsson T, Bjarnason T, Sigfusdottir ID. Individual and community processes of social closure: a study of adolescent academic achievement and alcohol use. *Acta Sociol*. 2007;50:161-178.
8. Bovey P, Viswanathan B, Faeh D, Warren W. Comparison of smoking, drinking, and marijuana use between students present or absent on the day of a school-based survey. *J Sch Health*. 2006;76(4):133-137.
9. Michaud PA, Delbos-Piott I, Narring F. Silent dropouts in health surveys: are nonrespondent absent teenagers different from those who participate in school-based surveys? *J Adolesc Health*. 1998;22(4):326-333.
10. Cheadle A, Psaty BM, Anderman C, et al. Evaluating the usefulness for school principals of feedback reports from school-based adolescent health survey. *Eval Rev*. 1995;19(6):675-686.
11. Dent CW, Sussman SY, Stacy AW. The impact of written parental consent policy on estimates from a school-based drug use survey. *Eval Rev*. 1997;21(6):698-712.

12. Bjarnason T. Administration mode bias in a school survey on alcohol, tobacco, and illicit drug use. *Addiction*. 1995;90(4):550-559.
13. Gottfredson DC, Gottfredson GD. Quality of school-based prevention programs: results from a national survey. *J Res Crime Delinq*. 2002;39(1):3-35.
14. Weitzman BC, Guttmacher S, Weinberg S, Kapadia F. Low response rate schools in surveys of adolescent risk taking behaviours: possible biases, possible solutions. *J Epidemiol Community Health*. 2003;57(1):63-67.
15. Dyer TA, Marshman Z, Merrick D, Wyborn C, Godson JH. School-based epidemiological surveys and the impact of positive consent requirements. *Br Dent J*. 2008;205(11):589-592.
16. Anderman C, Cheadle A, Curry S, Diehr P, Shultz L, Wagner E. Selection bias related to parental consent in school-based survey research. *Eval Rev*. 1995;19(6):663-674.
17. Unger JB, Gallaher P, Palmer PH, et al. No news is bad news—characteristics of adolescents who provide neither parental consent nor refusal for participation in school-based survey research. *Eval Rev*. 2004;28(1):52-63.
18. O'Donnell LN, Duran RH, SanDoval A, Breslin MJ, Juhn GM, Stueve A. Obtaining written parent permission for school-based health surveys of urban young adolescents. *J Adolesc Health*. 1997;21(6):376-383.
19. Lloyd K. Kids' Life and Times: using an Internet survey to measure children's health-related quality of life. *Qual Life Res*. 2011;20(1):37-44.
20. Trapl ES, Borawski EA, Stork PP, et al. Use of audio-enhanced personal digital assistants for school-based data collection. *J Adolesc Health*. 2005;37(4):296-305.
21. Nordahl H, Krolner R, Pall G, Currie C, Andersen A. Measurement of ethnic background in cross-national school surveys: agreement between students' and parents' responses. *J Adolesc Health*. 2011;49(3):272-277.
22. Sigfusdottir ID, Thorlindsson T, Kristjansson AL, Roe KM, Allegrante JP. Substance use prevention for adolescents: the Icelandic model. *Health Promot Int*. 2009;24(1):16-25.