Short communication

Identifying misclassification in youth self-reported smoking status: Testing different consent processes of biological sample collection to capture misclassification

Noella A. Dietz [a,*], Kristopher L. Arheart [b], David J. Lee [a], David F. Sly [c], Laura A. McClure [d]

[a] University of Miami Miller School of Medicine, Sylvester Comprehensive Cancer Center, Department of Public Health Sciences, 1120 NW 14th Street, 9th Floor C202, Miami, FL 33136, USA
[b] University of Miami Miller School of Medicine, Department of Public Health Sciences, 1120 NW 14th Street, 10th Floor, Miami, FL 33136, USA
[c] Florida State University, College of Social Sciences, 543 Old Cove Road North, Jasper, GA 30143, USA
[d] University of Miami Miller School of Medicine, Sylvester Comprehensive Cancer Center, 1120 NW 14th Street, 9th Floor, Miami, FL 33136, USA

ARTICLE INFO

ABSTRACT

Introduction: In Florida, since 1998, identical survey items have been used to measure youth smoking status for the CDC sponsored state school-based survey and the tobacco control program evaluation telephone survey. The two surveys should parallel one another to track tobacco use. Tobacco items collected in the two surveys closely paralleled one another until recently. Since 2008, data show dramatically divergent youth smoking estimates (e.g., relative differences as high as 50%), which cannot be explained by differences in survey and sampling design. As a first step in detecting misclassification of smoking status, we examined the feasibility of asking youth to self-report their smoking behavior and collect a biological sample, with the expectation that some youth will misreport their smoking status.
Methods: Using a cross-sectional population level telephone survey, youth were randomly assigned to one of three groups to test mode effects of collecting biological data with self-reported survey data ($n = 303$).
Results: It showed two groups of youth (those who are not asked for a biological and those asked for a biological with an indirect explanation of its use) had similar response rates and self-reported smoking status, while the third group (biological request with a direct explanation of its use) had a substantially lower response rate and self-reported smoking status.
Conclusions: The data show youth who are given an indirect explanation of how biological data are to be used were as likely to self-report their smoking status as youth who were not asked to provide a biological sample.

© 2015 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Eighty percent of adult smokers had their first cigarette before age 18, and half of adult smokers became daily users prior to this age (Mowery et al., 2000; US Department of Health and Human Services, 2000a). To eliminate the morbidity and mortality associated with tobacco use, the US Surgeon General calls for a reduction in smoking uptake among youth, increases in cessation among young adults and adults, and a reduction in secondhand smoke (SHS) exposures through tobacco control programs

(American Cancer Society, 2013; US Department of Health and Human Services, 2000b, 2012). The Centers for Disease Control and Prevention (CDC) funds state school-based tobacco surveys as part of their statistical data collection efforts to track tobacco use initiation, prevalence, cessation, and SHS exposure (Centers for Disease Control and Prevention, 2011); however, these CDC sponsored state school-based surveys do not contain the necessary items needed for tobacco control program media evaluation.

Comprehensive evaluation is an essential component of tobacco control programs and includes targeted surveillance on the reach of anti-tobacco advertising and its impact on tobacco use attitudes and behaviors, particularly among youth (Caraballo et al., 2004; Centers for Disease Control and Prevention, 2003; Messer and Pierce, 2010). The CDC recommended evaluation plan, or logic model, for tobacco control programs postulates that knowledge, beliefs, and attitudes directly influence susceptibility to tobacco

use, that knowledge and beliefs also directly and indirectly influence tobacco use, and that susceptibility directly influences tobacco use behaviors (Centers for Disease Control and Prevention, 2003). Evidence to support such programs often is obtained through the surveillance and evaluation mechanisms that are established as part of the tobacco control program (National Cancer Institute, 2008).

Data from the CDC sponsored state school-based tobacco surveys can be used to corroborate findings from anti-tobacco program evaluation telephone surveys. The CDC sponsored state school-based tobacco surveys of youth use self-administered questionnaires to collect smoking prevalence data. As states develop anti-tobacco programs, they collect additional information via telephone surveys for program evaluation purposes. In Florida, since 1998, identical survey items have been used to measure youth smoking status for the CDC sponsored state school-based survey and the tobacco control program evaluation telephone survey. In general, the two surveys should parallel one another to track tobacco use among youth. Questions of over- or under-reporting of tobacco use behaviors have been investigated, with an emphasis on the modality of questionnaire administration. In general, most methods used to collect data have their own biases, due to social desirability, exaggeration, inaccurate recall, and so forth (Adams et al., 2008; Bowling, 2005). However, verification studies on smoking prevalence among youth have shown self-reported smoking behavior and biochemical measurement results to be similar (Caraballo et al., 2001, 2004; Post et al., 2005).

As shown in studies by Caraballo et al. (2001, 2004), the main discrepancy between self-report and biochemical verification of smoking among youth using a national sample was in the self-reported smoker category; that is, youth who self-reported they smoked were considered nonsmokers by their serum cotinine levels. Smoking patterns of youth are most likely the explanation for this discrepancy, since most youth are sporadic smokers rather than daily users and cotinine has an 18 h half-life (Caraballo et al., 2001, 2004; Post et al., 2005; Stevens and Munoz, 2004). In any case, these findings are contradictory to the assumption that misclassification will be among youth who identify as nonsmokers due to their unwillingness to admit to being a smoker because of social desirability or not wanting their parents to know they smoke. To date, self-reported smoking prevalence, either by household interviews, telephone surveys, or self-administered, has been a dependable method. In fact, the tobacco items collected in Florida in the two surveys (i.e., the CDC sponsored state school-based surveys and the program evaluation telephone surveys) closely paralleled one another until recently. Since 2008, the tobacco program evaluation telephone data show dramatically divergent youth smoking estimates (e.g., relative differences as high as 50%), which cannot be explained by differences in survey and sampling design (Dietz et al., 2010a).

This discrepancy questions the validity of the data for each survey. Additionally, evidence suggests this trend has happened in other states with tobacco control programs (e.g., California and Massachusetts; Biglan et al., 2004; Giovino, 2002; Kann et al., 2002). In fact, youth smoking prevalence estimates at the national level also show a discrepancy in self-reported smoking behavior between school-based surveys, telephone surveys, and household interview (Biglan et al., 2004; Giovino, 2002; Kann et al., 2002). This discrepancy highlights the need to determine which data are really tracking trends in tobacco use among youth. Research on this topic is vital for comprehensive tobacco control programs as they attempt to implement optimal strategies to reduce the morbidity and mortality associated with tobacco use (Giovino, 2002).

As a first step toward determining the degree of youth smoking misclassification, we tested the feasibility of collecting telephone-based self-reported smoking behavior survey data, while collecting biological samples, with the expectation that youth will misreport their smoking status. Specifically, we examined if the consent process used for confidential telephone surveys could be used to mask, but not deceive, youth when collecting biological samples of hair/nail and self-reported smoking status. We hypothesized that youth responses to an indirect consent explanation of how the biological sample will be used will be similar to responses where no biological data are collected and response differences will be greatest when biological data are collected with a direct consent explanation.

2. Methods

2.1. Data collection/sample

State level cross-sectional data were collected via confidential telephone interview in January 2013 ($n = 303$). Telephone interviews were conducted by the University of Florida Survey Research Center (UFSRC). We employed the same CDC recommended study design and methodology used for state anti-tobacco media program evaluations (Dietz et al., 2010; Sly and Heald, 1999; Sly et al., 2001, 2002; Zucker et al., 2000). The sampling frame was comprised of vender generated targeted lists (Genesys, Inc.), and a random sample of youth (12–17 years) was taken. Detailed discussions of the sampling procedures, interview protocols, survey content, and representativeness of the samples have been described previously (Dietz et al., 2010; Sly et al., 2001). Youth were randomly assigned to one of three groups to test the mode of consent for collecting biological data with self-reported telephone survey data. The three groups were: (1) youth not asked for a biological sample; (2) youth asked for a biological sample using an indirect explanation (but not deception) of how the sample will be used; and (3) youth asked for a biological sample using a direct explanation of how the sample will be used. Parents were consented first, then the youth.

To collect biological data, twice a week the UFSRC would send the research team a secure Internet link with participant names to receive the biological collection kits. Collection kits were mailed within 24 h to respondents' home addresses. Youth collected the biological sample of hair/nail and returned them to the university in prepaid envelopes (83% return rate with no follow up prompts).

2.2. Measures

The CDC recommended item used to measure "ever smoke" was derived from the item, "Have you ever tried cigarettes smoking, even one or two puffs?" Any cigarette use defined a youth as having ever smoked. Youth also were asked a susceptibility item to determine if they were susceptible to smoke. Youth were asked, "How many of your four best friends smoke?" This measure was coded none versus one or more (0 = none; 1 = one or more). This is a validated CDC recommended item shown to predict susceptibility to uptake (Sly et al., 2005).

3. Results

Youth who did not submit a biological sample and those who did but had an indirect explanation of its use had a similar number of interviews, while the third group (biological request using a direct explanation of its use) had a substantially fewer interviews (Table 1). It took 17.0 enrollment calls to complete an interview for youth who did not submit a biological sample versus 26.7 calls for youth receiving an indirect explanation of how the biological will be used, and 35.4 calls when the direct explanation was used. The latter group had twice as many attempted calls as the group not asked for a biological sample. These results show the direct explanation created a selection bias in the consent process; we hypothesize that this modality yielded a false percentage of youth who self-reported smoking because smokers were more likely to refuse to participate. The relevance of this cannot be underestimated when attempting to measure the degree of misclassification when asking for a biological sample while recording youth self-reported smoking status.

We found minor differences for self-reported ever smoking between the group with no biological request and the group with an indirect explanation of how the biological will be used, and considerable differences between the third group where we used the direct explanation for how the biological sample will be used. Eleven percent of youth who were not asked for a biological sample admitted to ever smoking a cigarette, with 9% of youth admitting