A dual-frame sampling methodology to address landline replacement in tobacco control research

Robert C McMillen,[1,2] Jonathan P Winickoff,[1,3] Karen Wilson,[1,4] Susanne Tanski,[1,5] Jonathan D Klein[1]

[1]AAP Tobacco Consortium and Julius B. Richmond Center, Elk Grove Village, Illinois, USA
[2]Social Science Research Center and Department of Psychology, Mississippi State University, Starkville, Mississippi, USA
[3]MGH Center for Child and Adolescent Health Policy, Boston, Massachusetts, USA
[4]Children's Hospital Colorado, University of Colorado, Denver, Colorado, USA
[5]Department of Pediatrics and Adolescent Medicine, Dartmouth Medical School, Lebanon, New Hampshire, USA

Correspondence to
Dr Robert McMillen, Social Science Research Center and Department of Psychology, Mississippi State University, One Research Park, Suite 103, Starkville, MS 39759, USA; rcm19@msstate.edu

ABSTRACT

Objectives We assessed the comparability of self-reported smoking prevalence estimates from a dual-frame survey with those from two large-scale, national surveys.

Methods The Social Climate Survey of Tobacco Control (SCS-TC) obtained self-reported current smoking status via a dual-frame methodology in the fall of 2010. One frame used random digit dialling procedures and consisted of households with a landline telephone; the other frame consisted of a population-based probability-based online panel. Current smoking prevalence was compared with national estimates from the 2010 National Health Interview Survey (NHIS) and the 2009–2010 National Health and Nutrition Examination Survey (NHANES).

Results 18.3% (95% CI 17.0% to 19.6%) of SCS-TC respondents reported current smoking. NHIS and NHANES estimates found 19.4% (95% CI 18.8% to 20.1%) and 20.3% (95% CI 18.7% to 22.1%), respectively, reporting current smoking.

Conclusions Prevalence estimates for cigarette smoking obtained from the dual-frame SCS-TC are comparable to those from other national surveys. A mixed-mode approach may be a useful strategy to transition cross-sectional surveys with established trend data to newer dual-frame designs to maintain compatibility with surveys from previous years and to include the growing number of households that do not have landline telephones.

INTRODUCTION

Tobacco use is the leading preventable cause of disease and death in the USA. Approximately 443 000 people die prematurely from smoking or exposure to tobacco smoke each year, while 8.6 million live with a chronic illness caused by smoking.[1] The Centers for Disease Control and Prevention support several telephone and household surveys to monitor cigarette smoking among US adults. We developed the National Social Climate Survey of Tobacco Control (SCS-TC) to complement these large-scale, national surveys by assessing the social and environmental factors related to tobacco use and tobacco counselling from healthcare providers. The flexibility to add survey items that assess current attitudes and practices enhances the pace of scientific discovery around emerging issues in tobacco control.[2] The capacity to adapt this survey annually to address rapidly emerging issues and to report results within 3–6 months of data collection enhances the impact that this research has on tobacco control efforts. Our study of beliefs about 'third-hand smoke'[3]

helped to drive an emerging area of research, while media coverage from *Time*, *The Today Show*, and national public radio introduced this term and concept to the US public. Other studies have used data from the SCS-TC to examine support for banning mentholated cigarettes,[4] use of emerging tobacco products[2] and child healthcare clinicians addressing parental smoking.[5 6]

Since 2000, this survey has used a random-digit-dialling (RDD) sample frame of households with landline telephones. However, wireless substitution of cell phones for landlines continues to increase, and 35.8% of US households are currently wireless only.[7] In addition, wireless substitution is particularly problematic for surveys of tobacco use, as smoking status, and age, region, and several other demographic factors, vary by landline telephone status.[7] For these reasons, wireless substitution has been identified as a major barrier to RDD landline sampling frames.[7–10] Although one early investigation did not detect variations in smoking prevalence across landline-only and cell phone frames,[11] the authors did predict that the rapid growth of wireless substitution, especially among young adults, could become a source of non-coverage. Indeed, significant variations have been detected as wireless substitution rates have increased.[7 9]

To address this increasing source of coverage bias, we added an additional probability-based internet panel frame to the SCS-TC in 2009. We continued to use an RDD frame to maintain compatibility with the SCS-TC from the previous 10 years. As recommended by guidelines for dual frame surveys, weighting adjustments were made for conditions in which these frames overlap and a respondent could be represented in both frames.[8]

To investigate the viability of this dual-frame approach for the SCS-TC, we assessed the comparability of self-reported smoking prevalence estimates from our dual-frame survey with those from two large-scale, national surveys. Specifically, we compared estimates from the SCS-TC to those from the National Health Interview Survey (NHIS) and the National Health and Nutrition Examination Survey (NHANES). We hypothesise that estimates from the dual-frame SCS-TC will be consistent with those from these two large-scale, national surveys.

METHODS

Dual-frame surveys representing national probability samples of non-institutionalised US adults were administered in 2010. The design included an RDD frame and an internet panel frame developed

from a probability sample of the general population to reduce non-coverage issues arising from wireless substitution. We retained the RDD frame to maintain the capacity to examine trends with the RDD frames from previous survey administrations. The Institutional Review Board at Mississippi State University approved this study on 30 July 2010. More detailed methods have been previously published.[2]

The RDD sample frame included households with listed and unlisted landline telephones; five or more attempts were made to contact selected adults who were not at home. Survey Sampling, Inc provided the sample and the Survey Research Laboratory at the Mississippi State University Social Science Research Center administered the survey.

The probability-based internet panel sample frame included an online survey, administered to a randomly selected sample of a nationally representative research panel.[12] This panel is based on a sampling frame which includes listed and unlisted numbers and those without landline telephones. Panel members are recruited using an RDD frame and an address-based sample frame, using the US Postal Service Delivery Sequence File provided by Marketing Solutions Group. The panel does not accept self-selected volunteers,[12] and provides sample coverage for 99% of US households, including low socio-demographic households, households that did not have internet access prior to recruitment, and younger adults.[13] The use of RDD and addressed-based frame recruiting provides this high-level of coverage. Knowledge Networks provided this internet panel and administered the survey to this panel.

Both survey frames were administered in the fall of 2010. Data were weighted to adjust for age, race, gender and region, and for the frame overlap among internet panel respondents who also had a landline telephone and were therefore also eligible for the RDD frame. Overall weights were computed in two steps. First, the two frames were weighted based on 2009 US Census estimates to be representative of the US population. Second, three adjustments to these initial weights were computed to account for the overlap in the two samples. Weights from the RDD frame were multiplied by 0.818 to adjust for the overlap (81.8% of households in the internet panel frame had a landline). Composite adjustments were then computed to combine the two sampling frames. According to the American Association for Public Opinion Research (AAPOR),[14] observations from two sampling frames with overlap may be combined using composite weights. Two compositing factors that sum to one are typically selected. Given that the effective sample sizes of the RDD frame and internet panel frame are similar, the two compositing factors were set to 0.5. The weights of respondents who were represented in both sampling frames (ie, landline owners) were multiplied by the compositing factor. In the final adjustment, a re-standardised weight was computed so that the weighted sample size matched the sum for effective sample size for both independent frames.

Self-reported smoking

Respondents in each of the surveys were asked, 'Have you smoked at least 100 cigarettes in your entire life?' Respondents who reported that they had were then asked, 'Do you now smoke cigarettes every day, some days, or not at all?' Respondents who reported that they have smoked at least 100 cigarettes and now smoke every day or some days were categorised as current smokers. Overall, current smoking prevalence from the SCS-TC was compared with survey data on the 'current smokers' measure from the 2010 NHIS basic adult module household interview, the 2009/10 NHANES sample

person household interview, and the 2010 core telephone interview. The SCS-TC survey samples and the two national surveys used the same protocol to assess 'current smoking'. Estimates from the SCS-TC and NHIS are from adults aged 18 years and older; while the NHANES used this protocol to assess 'current smoking' among adults aged 20 years and older.

Analyses

Descriptive and bivariate analyses examined overall and subpopulation (sex, race and age) prevalence estimates for current smoking from each of the surveys. To explore the possibility of a 'time-on-panel bias', we performed logistic regression analyses to examine the relationship of length of time on panel with self-reported smoking in the SCS-TC panel frame.

RESULTS

In the RDD frame, of 2128 eligible respondents contacted, 1504 (70.7%) completed surveys.[14] For the internet panel frame, 2272 panellists were randomly drawn from the probability panel; 1736 responded to the invitation, yielding a final stage completion rate[15] of 67.5%. The recruitment rate (computed using the AAPOR response rate 3) for this study was 16.4% and the profile rate (at least one member of a recruited household completed a profile survey for the panel) was 65.1%, for a cumulative response rate[15] of 7.2%. Length of time on the panel for the internet panel frame ranged from 0.09 to 11.08 years, with a median length of time on the panel of 2.29 years. Table 1 shows the demographic characteristics of the overall sample.

Table 1 Demographic characteristics of respondents

Demographic variable	Overall N	Overall weighted (%)	RDD frame unweighted (%)	Internet panel frame unweighted (%)
Region				
Northeast	404	12.6	18.7	18.9
Midwest	589	18.4	25.5	22.4
South	1203	37.6	39.5	37.0
West	1007	31.4	16.4	21.7
Race				
White	2346	74.2	87.2	73.8
African American	364	11.5	10.0	8.5
Other	454	14.3	2.7	17.7
Age				
18–24	440	13.7	8.3	8.1
25–44	1241	38.8	18.1	35.3
45–65	1066	33.3	41.9	41.4
65+	455	14.2	31.7	15.2
Education				
Not a high-school graduate	291	9.2	5.6	11.2
High-school graduate	903	28.5	28.6	29.0
Some college	929	29.3	25.9	28.0
College graduate	1044	33.0	40.0	31.7
Sex				
Women	1523	52.3	36.2	46.7
Men	1675	47.6	63.8	53.3

RDD, random digit dialling.

Weighted estimates from the SCS-TC current smoking item found 18.3% (95% CI 17.0% to 19.6%) of respondents reported current smoking (see table 2). Weighted estimates from the NHIS and NHANES for current smoking were 19.4% (95% CI 18.8% to 20.1%) and 20.3% (95% CI 18.7% to 22.1%), respectively. SCS-TC estimates for self-reported current smoking did not statistically differ from those from the NHIS and NHANES.

Estimates by gender, race (non-Hispanic African American and non-Hispanic white) and age are also provided for each survey in table 2, along with CIs. In each survey, men reported higher rates of smoking than women, although only the NHIS detected a statistically significant difference. With the exception of the NHIS, all of the surveys detected higher estimates of self-reported current smoking among African American than white respondents. However, this difference was statistically significant only in the NHANES. Most gender, race and age estimates from the SCS-TC did not statistically differ from those of the NHIS and NHANES. However, the SCS-TC produced estimates for self-reported smoking among white adults that were slightly lower than those from the NHIS for young adults that were slightly lower than those from the NHANES. The SCS-TC and the NHIS produced similar estimates for African Americans, whereas the estimate from the NHANES was substantially higher.

Analyses by length of time on the panel did not detect a time-on-panel bias in the panel frame of the SCS-TC.

CONCLUSIONS

Estimates of self-reported smoking obtained from the SCS-TC are comparable to those from nationally representative household interview surveys among US adults. Overall estimates and those for gender, race and age were similar to those of large-scale, national surveys, demonstrating that the SCS-TC findings are consistent with two government-supported surveys that serve as the principal sources of information about tobacco use in the US population. Previous research using a different online panel has demonstrated that online panel studies can provide estimates of current smokers that closely match those from national household and telephone surveys.[16] This study demonstrates that a dual-frame approach that preserves the

RDD frame of an extant cross-sectional survey and combines it with a probability-based frame can also produce similarly accurate estimates. The internet panel frame alone also produced an estimate for current smoking that closely matched estimates from the NHIS and NHANES, suggesting that perhaps the use of a mixed mode with an RDD frame is not necessary. However, our intention was to assess the viability of using this mixed-mode approach to transition cross-sectional surveys with established trend data to newer dual-frame designs to maintain compatibility with surveys from previous years and include the growing number of households that do not have landline telephones. Moreover, multiple frame surveys are more likely to reduce non-coverage bias by complementing the strengths and limitations of one another.

Although several studies demonstrate that using a dual frame survey of landline and cell phone numbers can provide valid, reliable and representative data,[11 17 18] we selected a probability-based internet panel as a more cost-effective approach to reduce non-coverage bias. Cell phone frame surveys have unique challenges and costs due to the inherent mobility provided by the device.[11] Our costs of conducting an RDD survey frame are comparable to the amount to contract Knowledge Networks for an internet panel survey, whereas cell phone frame surveys typically cost 35% more to conduct.

Although the use of a dual frame substantially reduces concerns about coverage bias dues to wireless substitution, this study is subject to at least three limitations. First, the use of the internet panel raises some concern about the representativeness of the sample. However, several comparison studies have demonstrated that this panel, which was recruited from a population-based frame, yields results comparable to well designed RDD surveys in terms of demographics and outcome variables.[19 20] Internet panels have also demonstrated less evidence of survey satisficing and social desirability than RDD surveys.[19] More recently, Yeager and colleagues conducted a similar sample frame comparison study that also included benchmarks from the NHIS and the Current Population Survey, finding similar comparability for items examining current smoking, in addition to gender, age and education.[20] Second, ongoing engagement may lead to panel conditioning and thereby reduce data reliability if respondents develop a

Table 2 Comparison of prevalence estimates of adult current smoking

	2010 SCS-TC, % (95% CI) (weighted N=3169)	2010 SCS-TC (RDD), % (95% CI) (weighted N=1456)	2010 SCS-TC (KN panel), % (95% CI) (weighted N=11715)	2010 NHIS, % (95% CI) (weighted N=89876)	2009/10 NHANES, % (95% CI) (weighted N=6218)
Current smoking prevalence	18.3 (17.0 to 19.6)	14.8 (13.0 to 16.6)	20.3 (18.4 to 22.2)	19.4 (18.8 to 20.1)	20.3 (18.7 to 22.1)
Gender					
Men	20.2 (18.2 to 22.2)	16.0 (13.2 to 18.8)	22.5 (19.7 to 25.3)	21.6 (20.6 to 22.6)	22.3 (20.2 to 24.5)
Women	16.6 (14.8 to 18.4)	13.6 (11.2 to 16.0)	18.3 (15.8 to 20.8)	17.4 (16.6 to 18.2)	18.5 (16.4 to 20.8)
Race					
White	17.5 (16.0 to 19.0)	14.9 (12.9 to 16.9)	21.4 (15.7 to 27.1)	21.8 (20.9 to 22.7)	19.8 (17.8 to 21.9)
African American	19.7 (15.6 to 23.8)	17.1 (11.2 to 23.0)	26.6 (20.9 to 32.3)	20.3 (18.9 to 21.9)	28.1 (23.3 to 33.5)
Age					
18–44	19.4 (17.5 to 21.3)	16.3 (13.7 to 18.9)	21.2 (18.4 to 24.0)	21.6 (20.5 to 22.6)	25.1 (22.5 to 28.0)
45–64	19.8 (17.4 to 22.2)	16.3 (12.9 to 19.7)	22.1 (18.9 to 25.3)	21.2 (20.2 to 22.3)	20.3 (17.7 to 23.0)
65+	11.0 (8.1 to 13.9)	6.9 (3.6 to 10.2)	13.1 (8.9 to 17.3)	9.5 (8.6 to 10.5)	7.7 (6.4 to 9.2)

KN, Knowledge Networks; NHANES, National Health and Nutrition Examination Survey; NHIS, National Health Interview Survey; RDD, random digit dialling; SCS-TC, Social Climate Survey of Tobacco Control.

'time-on-panel bias' in some variables, due to increased experience with completing surveys. However, our analyses of length of time on panel did not detect any time-on-panel bias in self-reported smoking items. Third, respondents from less educated or lower-income households in the internet panel may have lower levels of computer literacy. However, analyses did not detect higher levels of missing data among respondents with low levels of education or household income. Furthermore, the average length of time on the panel was 2.4 years for low-income adults and 2.3 years for adults with less education than a high-school degree, suggesting that most participants would have developed computer literacy.

Prevalence estimates obtained from the dual-frame SCS-TC are comparable to those from other national surveys. This approach may be a useful means to transition from cross-sectional surveys with established trend data to dual-frame designs that maintain compatibility with surveys from previous years and also include the growing number of households that do not have landline telephones.

What this paper adds

Population-based surveys are a critical component of surveillance and evaluation programmes. Many tobacco control programmes have relied on landline telephone surveys due to the cost efficiency and (formerly) high coverage rate of this approach. However, wireless substitution of cell phones for landlines continues to increase, and 35.8% of US households are currently wireless only. In addition, wireless substitution is particularly problematic for surveys of tobacco use, as smoking status varies by landline telephone status. This study demonstrates that a dual-frame approach that preserves the random digit dialling frame of an extant cross-sectional survey and combines it with a probability-based frame can also produce similarly accurate estimates. This mixed-mode approach can be a useful strategy to transition cross-sectional surveys with established trend data to newer dual-frame designs to both maintain compatibility with surveys from previous years and include the growing number of households that do not have landline telephones.

Ethics approval Mississippi.

Provenance and peer review Not commissioned; internally peer reviewed.

REFERENCES

1 Centers for Disease Control and Prevention. *Tobacco use: targeting the nation's leading killer at a Glance*. 2011. http://www.cdc.gov/chronicdisease/resources/publications/aag/osh.htm (accessed 13 Apr 2012).
2 McMillen R, Maduka J, Winickoff J. Use of emerging tobacco products in the United States. *J Environ Public Health* 2012;2012:1–8.
3 Winickoff JP, Friebely J, Tanski SE, et al. Beliefs about the health effects of 'thirdhand' smoke and home smoking bans. *Pediatrics* 2009;123:e74–9.
4 Winickoff JP, McMillen RC, Vallone DM, et al. US attitudes about banning menthol in cigarettes: results from a nationally representative survey. *Am J Public Health* 2011;101:1234–6.
5 Winickoff JP, Tanski SE, McMillen RC, et al. Child health care clinicians' use of medications to help parents quit smoking: a national parent survey. *Pediatrics* 2005;115:1013–17.
6 Winickoff JP, McMillen RC, Carroll BC, et al. Addressing parental smoking in pediatrics and family practice: a national survey of parents. *Pediatrics* 2003;112:1146–51.
7 Blumberg SJ, Luke JV. *Wireless substitution: early release of estimates from the National Health Interview Survey, January–June 2012*. 2012. http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201212.pdf (accessed 6 Jan 2013).
8 AAPOR. New considerations for survey researchers when planning and conducting RDD telephone surveys in the U.S. with respondents reached via cell phone numbers. 2010. http://aapor.org/AM/Template.cfm?Section=Cell_Phone_Task_Force&Template=/CM/ContentDisplay.cfm&ContentID=2818 (accessed 13 Apr 2012).
9 Delnovo C, Gundersen D, Hagman B. Declining estimated prevalence of alcohol drinking and smoking among young adults nationally: artifacts of sample undercoverage? *Am J Epidemiol* 2008;167:15–19.
10 Delnovo C, Bauer U. Monitoring the tobacco use epidemic III the Host: data sources and methodological challenges. *Prev Med* 2009;48:S16–23.
11 Brick J, Brick P, Dipko S, et al. Cell phone survey feasibility and calling cell numbers versus landline numbers. *Public Opin Q* 2007;71:23–39.
12 Knowledge Networks. *KnowledgePanel: Design summary*. http://www.knowledgenetworks.com/knpanel/docs/KnowledgePanel(R)-Design-Summary-Description.pdf (accessed 13 Apr 2012).
13 Knowledge Networks. KnowledgePanel: Processes & procedures contributing to sample representativeness & tests for self-selection bias. 2010. http://www.knowledgenetworks.com/ganp/docs/knowledgepanelr-statistical-methods-note.pdf (accessed 13 Apr 2012).
14 AAPOR Cell Phone Task Force. New considerations for survey researchers when planning and conducting RDD telephone surveys in the U.S. with respondents reached via cell phone numbers. 2010. http://aapor.org/AM/Template.cfm?Section=Cell_Phone_Task_Force&Template=/CM/ContentDisplay.cfm&ContentID=2818 (accessed Apr 2013).
15 Callegaro M, DiSogra C. Computing response metrics for online panels. *Public Opin Q* 2008;72:1008–32.
16 Klein JD, Thomas RK, Sutter EJ. Self-reported smoking in online surveys: prevalence estimate validity and item format effects. *Med Care* 2007;45:691–5.
17 Keeter S, Kennedy C, Clark A, et al. What's missing from national RDD surveys? The impact of the growing cell-only population. *Public Opin Q* 2007;71:772–92.
18 Hu S, Balluz L, Battaglia M, et al. Improving public health surveillance using a dual-frame survey of landline and cell phone numbers. *Am J Epidemiol* 2011;173:703–11.
19 Chang L, Krosnick JA. National surveys via RDD telephone interviewing versus the internet comparing sample representativeness and response quality. *Public Opin Q* 2009;73:641–78.
20 Yeager DS, Krosnick JA, Chang LC, et al. Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. *Public Opin Q* 2011;75:709–47.