

Recruiting and retaining youth and young adults: challenges and opportunities in survey research for tobacco control

Jennifer Cantrell,^{1,2} Elizabeth C Hair,^{1,2} Alexandria Smith,¹ Morgane Bennett,¹ Jessica Miller Rath,^{1,2} Randall K Thomas,³ Mansour Fahimi,³ J Michael Dennis,⁴ Donna Vallone^{1,5}

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/10.1136/tobaccocontrol-2016-053504>)

¹Evaluation and Science Research, Truth Initiative, Washington DC, USA

²Department of Health, Behavior and Society, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA

³GfK New York, New York, NY, USA

⁴National Opinion Research Center, Chicago, Illinois, USA

⁵Global Institute of Public Health, New York University, New York, NY, USA

Correspondence to

Dr Jennifer Cantrell, Evaluation and Science Research, Truth Initiative, 900 G Street, NW, Fourth Floor, Washington, DC 20001, Columbia, USA; jcantrell@truthinitiative.org

Received 28 October 2016

Revised 28 February 2017

Accepted 10 March 2017

ABSTRACT

Introduction Evaluation studies of population-based tobacco control interventions often rely on large-scale survey data from numerous respondents across many geographic areas to provide evidence of their effectiveness. Significant challenges for survey research have emerged with the evolving communications landscape, particularly for surveying hard-to-reach populations such as youth and young adults. This study combines the comprehensive coverage of an address-based sampling (ABS) frame with the timeliness of online data collection to develop a nationally representative longitudinal cohort of young people aged 15-21.

Methods We constructed an ABS frame, partially supplemented with auxiliary data, to recruit this hard-to-reach sample. Branded and tested mail-based recruitment materials were designed to bring respondents online for screening, consent and surveying. Once enrolled, respondents completed online surveys every 6 months via computer, tablet or smartphone. Numerous strategies were utilized to enhance retention and representativeness

Results Results detail sample performance, representativeness and retention rates as well as device utilization trends for survey completion among youth and young adult respondents. Panel development efforts resulted in a large, nationally representative sample with high retention rates.

Conclusions This study is among the first to employ this hybrid ABS-to-online methodology to recruit and retain youth and young adults in a probability-based online cohort panel. The approach is particularly valuable for conducting research among younger populations as it capitalizes on their increasing access to and comfort with digital communication. We discuss challenges and opportunities of panel recruitment and retention methods in an effort to provide valuable information for tobacco control researchers seeking to obtain representative, population-based samples of youth and young adults in the U.S. as well as across the globe.

Traditionally, many of the largest and most rigorous tobacco-related survey data collections used phone-based methods.⁴⁻⁵ However, significant challenges for conducting phone surveys have emerged with the evolving communications landscape, including reduced coverage of landline sampling frames, declining response rates and increasing costs.⁶ The capacity to fully enumerate units of a population who could be potentially selected via landline has declined dramatically given the rise of cellphone-only households. Current coverage rates of landline frames are estimated to represent less than 50% of the population,⁷ with evidence indicating these frames will reflect less than 10% of households within 10 years.⁸ Studies that combine landline and cellphone frames have not resolved these problems given declining phone response rates.⁷⁻⁹ Such declines can be attributed to numerous factors, including the broad penetration of caller identification and voicemail.¹⁰ These factors serve to significantly reduce the probability of reaching potential respondents, and in turn, increase labour costs associated with data collection.¹¹ Higher costs for cellphone-only surveys⁶⁻¹² and evolving legal restrictions on automated dialing of cellphones for research purposes present other challenges.¹³

Online data collection can be less expensive than phone and is especially valuable when surveying younger populations.⁶ Web-based surveys allow researchers to capitalise on youths' increasing access to and comfort with digital platforms. However, the quality of online survey data for population-based studies depends, in part, on the degree to which the sampling frame from which the sample is drawn represents the target population. Although non-probability samples are typically used for market and commercial research, academic or government studies often require the coverage of representative sampling frames which allow for the estimation of selection probabilities and ensure representative and accurate estimates.¹⁴⁻¹⁶ Such probability-based sampling frames combined with online data collection are growing in popularity in the USA and abroad.¹⁷⁻²⁴ These surveys use phone or address-based sampling (ABS) to recruit individuals for cross-sectional or longitudinal studies. In the USA, ABS frames are based on the US Postal Service (USPS) Computerized Delivery Sequence File which is estimated to cover close to 100% of US households.²⁵

INTRODUCTION

Evaluation studies of population-based tobacco control interventions often rely on large-scale survey data from numerous respondents across many geographical areas to provide evidence of their effectiveness.^{1,2} Large survey data collections also serve as critical surveillance mechanisms to examine tobacco use patterns at a national level.³



CrossMark

To cite: Cantrell J, Hair EC, Smith A, et al. *Tob Control* Published Online First: [please include Day Month Year]. doi:10.1136/tobaccocontrol-2016-053504

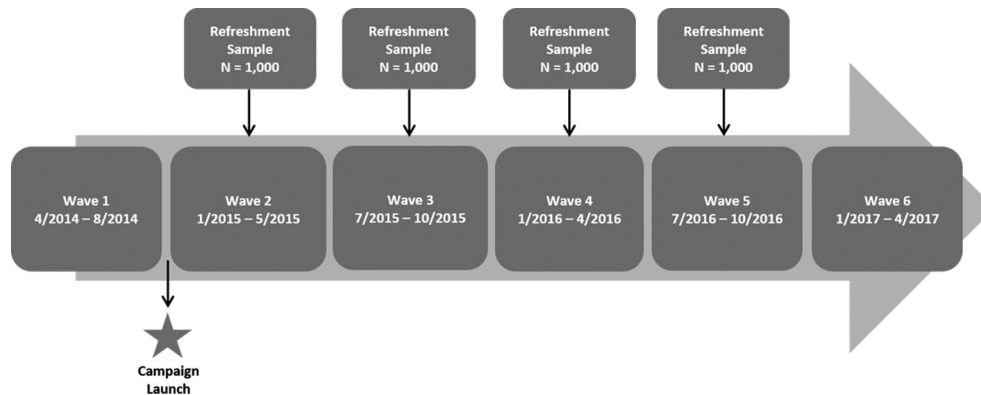


Figure 1 Truth Longitudinal Cohort timeline.

Given the challenges of phone-based survey sampling and data collection, as well as the limitations of online non-probability samples, we employed a hybrid approach to harness both the strengths of ABS sampling with the speed of online data collection in an effort to engage a cohort of youth and young adults. Since the majority of tobacco users begin smoking in adolescence and young adulthood,^{26,27} methods to recruit and retain respondents from this age group are extremely valuable for effective programme planning, implementation and evaluation studies. This paper outlines the specifications related to the study design, sampling methods, recruitment and retention strategies, data collection methods, and sample performance of the Truth Longitudinal Cohort (TLC), a national sample of 15–21-year-olds designed to evaluate the 2014 relaunch of the antitobacco mass media campaign, truth. We present the methodology, results, challenges and opportunities for survey research in an effort to provide valuable information for tobacco control researchers seeking to obtain representative, population-based samples of youth and young adults in the USA as well as across the globe.

METHODS

Study design

The evaluation of the truth campaign required the recruitment and maintenance of a large nationally representative cohort of the target audience: youth and young adults aged 15–21 years. To ensure sufficient power to detect campaign effects, target sample goals specified a prospective cohort of 10 000 respondents at baseline, with follow-up assessments every 6 months for 3 years. Cross-sectional samples of approximately 1000 new respondents of the same age range were added at each data collection period and followed thereafter to address attrition and panel conditioning effects. [Figure 1](#) shows data collection and the timeline for waves 1 to 6. Each data collection period for follow-up and refreshment waves was approximately 3–4 months.

Probability-based sampling

The sample was obtained using the ABS frame described above, which consisted of addresses from the USPS.²⁵ To maximise the ability of locating a member of the ABS-sourced household who fell within our age target, the frame was supplemented with auxiliary information obtained from public and commercial sources for a portion of the sample. Supplementation of the ABS frame with auxiliary information provides data-based insights on the demographic and socioeconomic composition of households in the frame, and allows researchers to adhere to the probability sampling paradigm in designing representative samples of specific hard-to-reach populations if properly designed.^{25–28} Auxiliary data sources are commercially available

from marketing information firms (eg, infoUSA, Acxiom, Experian and Targus).

Of the total sample units invited to the TLC, 14.7% was derived from ‘listed’ households (ie, leveraging auxiliary data) indicated to have an eligible study participant age 15–17 years and 17.9% from a ‘listed’ 18–21-year-old household. The remaining 67.4% of the sample was selected from non-list-assisted households. The sample was not stratified by geography. This protocol led to a highly acceptable design effect of 1.08. Use of the auxiliary information was substantially effective for the intended purpose of increasing the percentage of sample units that qualify for the TLC. In strata 1, 75.7% of households had a 15–17-year-old; for strata 2, 70.1% had an 18–21-year-old; for strata 3, 12.2% had a 15–21-year-old. Online supplementary table 1 provides information on the sample allocation and inclusion probabilities across the strata.

Data collection

In an effort to optimise recruitment and retention, study branding and a project-specific website were used. Branding of health-related goods, processes or services can help establish positive associations between a target audience and the product and can serve to promote commitment and loyalty over time. The TLC branding was designed to present the study as distinctive, trusted and reliable, which is especially important in longitudinal cohort studies where retention is critical.²⁹ Branding included the study name, Connecting Health and Technology (CHAT), paired with an associated logo and logos of study partners (affiliated academic institutions and contract research organisations) on materials to help increase legitimacy (see online supplementary figure A1). The website featured the respondent screener and subsequent surveys, as well as instructions on accessing future surveys and updating one’s contact information, clear and concise information on frequently asked questions, data privacy policies and contact information of the study coordinator through the web portal. Participants were able to return to this website at any time to check for upcoming surveys, review survey completion and the status of participation incentives, update contact information and provide feedback (see online supplementary figure A2).

Testing of all recruitment and retention materials was conducted among a sample of the target audience to refine the content, tone and imagery. Forced exposure tests using online convenience samples were conducted to evaluate receptivity to the study brand name, web content, letters, postcards, envelopes, email messages, logos and imagery for study materials. Cognitive testing of the online data collection process was also conducted with a combined online sample of 15–21-year-olds as well as parents of 15–17-year-olds to examine any potential barriers

related to the comprehension of survey items and skip patterns in the baseline instrument to improve survey comprehension and flow. We also tested usability of the household screener and the parental consent process. Once data instruments were finalised, surveys were optimised for completion on desktop and mobile devices. All recruitment, retention and survey materials were provided in English and Spanish.

Study invitations were sent by mail to all potential respondents in a 9×12 envelope with a letter structured in a simple 'Question and Answer' format, which provided information on the study, study sponsors, survey website, a unique passcode, and a 1800 toll-free telephone number established to field questions from potential respondents. Reminder postcards and letters were sent to all non-respondents. All materials included the CHAT logo and study sponsor information. Descriptions of the study were provided on study sponsors' websites and an online ad, and relevant Google search terms were purchased to direct potential respondents who wanted to learn more about the study.

Once a participant logged into the recruitment website with the unique passcode, they were screened, rostered, selected and consented. First, the household respondent completed the online screening questionnaire, which included a rostering of household members for purposes of selecting a qualified respondent. If multiple eligible 15–21-year-olds were listed in the household, one was randomly chosen. Second, the parent/guardian of an eligible youth age 15–17 years was asked to provide consent for that youth, and then asked to have the youth assent and complete the survey. Those aged 18–21 years could provide their own consent and were asked to complete the survey directly. Finally, both parents/guardians and selected participants were informed of the incentive structure for their participation, and the subsequent requests for completing additional surveys. Parents could choose whether to receive the incentive directly or elect the incentive and surveys be sent directly to their teen. Teens were required to complete the survey themselves and not via parental proxy. If participants started the screening process but did not complete the full screening, consent or survey process, an additional reminder letter and two emails (for those who had provided email addresses) were sent. Reminder letters included

the member portal website and emails included direct links to the portal and the survey.

Each respondent received a base contingent incentive of \$10 for completion of a survey. Respondents defined as hard-to-reach, such as African-Americans and Hispanics, received an additional contingent incentive of \$10, for a total of \$20. Respondents without household internet access were also defined as hard-to-reach and received an additional contingent incentive of \$20. The maximum incentive offered to any single respondent for the baseline survey was \$40.

Numerous pathways for retention were used, including varying contacts at each retention wave if needed, which included an initial invite letter and email, up to seven reminders via email, postcard or letter (some of which provided additional incentives), and an alternative contact and text reminder for those who provided this information. A suspend reminder postcard, letter and email were also sent for those who started a follow-up survey but did not complete it. A large majority of respondents completed the survey within the first three contacts and the additional contacts were used for a small proportion of the sample to boost response. Other strategies were also used to engage respondents, including birthday cards, humorous emails and social media messages. Between-wave postcards and emails, both of which had a link to the member portal, were sent to request updated contact information 2 months before each survey.

RESULTS

Sample performance

Table 1 provides a detailed summary of the total number of invites sent by strata and responses by disposition. At baseline, 1 293 801 participants were sent mailed invites. A total of 1 293 801 reminder postcards and 1 288 572 reminder letters were sent following the initial invite. Across the strata, the listed sample performed better than the unlisted sample: rates of screener eligibility, survey completion and consent were five to six times higher for the listed sample compared with the unlisted sample. Of the total 1 293 5 invites sent, 40 464 were screened and, of those screened, 12 882 participants were determined to be

Table 1 Total numbers of invites and responses by strata and disposition.

	Stratum 1: 15–17 listed (n=1 899 993)			Stratum 2 18–21 listed (n=2 314 033)			Stratum 3 (Not Listed) (n=8 724 055)			Strata 1–3 Combined
	15–17 years	18–21 years	Total	15–17 years	18–21 years	Total	15–17 years	18–21 years	Total	Grand total
No. invited ABS sample units	114 013	75 980	189 993	32 553	198 850	231 403	384 635	487 770	872 405	1 293 801
No. completed screener (1)			5960			7189			27 315	40 464
No. confirmed study eligible (2)	2713	1799	4512	736	4302	5038	1451	1881	3332	12 882
No. consented (3)	2265	1744	4009	623	4180	4803	1266	1842	3, 108	11 920
I=complete interviews (1.1)	2054	1480	3534	570	3500	4070	1142	1511	2653	10 257
P=partial interviews (1.2)	211	264	475	53	680	733	124	331	455	1663
R=refusal and break off (2.1)	448	55	503	113	122	235	185	39	224	962
NC=non-contact (2.2)	0	0	0	0	0	0	0	0	0	0
O=other (2.0, 2.3)	0	0	0	0	0	0	0	0	0	0
Calculating e: e is the estimated proportion of cases of unknown eligibility that are eligible	0.7	0		0.65	0		0.12	0		
UH=unknown household (3.1)	0	0	0	0	0	0	0	0	0	0
UO=unknown other (3.2–3.9) (4)	110 154	73 194	183 348	31 420	192 061	223 481	372 418	472 217	844 635	1 251 464

The reference to ABS refers to 'address-based sample.' The invited sample units consisted of addresses from the USPS Delivery Sequence File.

(1) Completed Screener: Completed the screener and have been determined as ineligible or eligible for the survey. Ineligible households are those in which there are no age-eligible persons. (2) Eligible: Eligible households are those in which an age-eligible child (ages 15–17 years) or adult (ages 18–21 years), was determined and selected to complete the survey. (3) Consented: Selected household member agreed to participate in the survey. For selected survey takers who were 15–17 years, the parent or legal guardian provided consent, and also agreed to have their teen come to the computer/screen, ahead of the teen providing their consent.

Research paper

Table 2 Device used to complete baseline survey for the original sample and each refreshment sample

Device used to complete survey	Per cent of sample				
	Baseline sample	Wave 2 refreshment baseline	Wave 3 refreshment baseline	Wave 4 refreshment baseline	Wave 5 refreshment baseline
Desktop	79.8%	79.8%	75.3%	75.0%	69.1%
Tablet	9.7%	11.0%	11.3%	9.7%	11.0%
Smartphone	9.4%	7.7%	11.8%	13.5%	17.7%
Any mobile device (tablet or smartphone)	19.1%	18.7%	23.1%	23.2%	28.7%

eligible for the baseline survey. Of these, 11 981 were consented. A total of 10 257 respondents completed the baseline. A portion of those who were eligible and consented started but did not complete the survey initially, and thus reminder letters (n=3244) and two emails (n=3244 and 3075) (for those who had provided email addresses) were sent to encourage respondents to complete the survey. Of those who received these reminder letters and emails, 766 eventually completed the survey. For the baseline survey, the incentive structure was more efficient for the listed strata versus the unlisted strata (see online supplementary table 2). Among the sample overall, 66% of respondents received a \$10 incentive, 31% received \$20, 1% received \$30 and 2% received \$40. A greater proportion of respondents in the two listed strata received \$10 compared with those in the unlisted strata. This was primarily due to differences in demographics, discussed further below.

In computing the response rate, each of the three strata had a separate rate of eligibility for the two age groups (15–17 years and 18–21 years). These rates of eligibility were used to estimate the proportion of possible eligible participants in the ‘unlisted’ sample within the strata and group. The 18–21-year-old group had a quota limit; once this quota was achieved, the estimated eligibility for all in the ‘listed’ and ‘unlisted’ samples of 18–21-year-olds were assigned ‘0.’ Any additional completes within this age range were not included as they were considered

‘adults overquota’.²⁵ The response rate was calculated as a weighted RR3 using the American Association of Public Opinion Research (AAPOR) response rate calculator for each of the six sample groups (three strata by two age groups).³⁰ The final weighted survey response rate was 52.4% (AAPOR-Response Rate 3 (RR3) using our quota limit). The refreshment samples were recruited using the same custom ABS methodology and response rates were similar.

Device utilisation

Approximately 20% of the baseline sample at wave 1 completed the survey on a mobile device. In later waves, the proportion of respondents taking the survey on mobile increased, with close to 30% of refreshment respondents taking the survey via a mobile device by wave 5 (see table 2). The baseline survey took a median of 31.8 min, which varied by device: 29.9 min for desktop/laptop; 33.8 min for tablet and 42.6 min for smartphone. The follow-up surveys at wave 2–5 were generally shorter, with a median time taken of approximately 25 min, ranging from 24–25 min for desktop/laptop, 25–27 min for tablet and 27–28 min for smartphone.

Sample representativeness

Table 3 provides demographics by sample strata and for the sample overall in comparison with 2014 census statistics to assess

Table 3 Unweighted demographic characteristics by sample strata

	Listed 15–17 years old n=3518	Listed 18–21 years old n=4066	Unlisted 15–21 years old n=2644	Total n=10 228*	2014 Census†n=29,244,457
Age, years					
15–17	57.8%	13.9%	42.5%	36.4%	44.5%
18–21	42.2%	86.1%	57.5%	63.6%	55.5%
Gender					
Female	48.8%	54.5%	54.3%	52.5%	49.4%
Male	51.2%	45.6%	45.7%	47.5%	50.6%
Race/ethnicity					
White	78.2%	68.9%	51.7%	67.7%	54.9%
Black/African-American	6.1%	9.9%	15.6%	10.1%	14.4%
Hispanic	7.8%	11.5%	19.4%	12.3%	22.3%
Other	7.9%	9.7%	13.2%	10.0%	8.4%
Region					
North-East	24.1%	23.6%	17.0%	22.1%	17.2%
South	31.9%	35.9%	37.6%	35.0%	36.6%
Midwest	29.9%	25.5%	22.3%	26.2%	21.6%
West	14.2%	15.0%	23.1%	16.8%	24.6%
MSA category					
Non-metro	13.1%	13.4%	14.8%	13.6%	15.6%
Metro	87.0%	86.6%	85.3%	86.4%	84.4%

*Twenty-nine respondents were removed from the final baseline sample due to lack of sufficient contact information for recontact at wave 2.

†Source: US Census Bureau, Current Population Survey. <https://www.census.gov/cps/data/>
MSA, metropolitan statistical area.

Table 4 Response rates across waves as a proportion of wave 1 sample

Wave	Sample size overall	Listed 15–17 years old	Listed 18–21 years old	Unlisted 15–21 years old	Overall
Wave 1	10,228*				
Wave 2	7408	74.53%	72.53%	69.48%	72.40%
Wave 3	6973	70.04%	69.36%	63.88%	68.20%
Wave 4	7284	73.45%	73.00%	65.51%	71.22%
Wave 5	6748	68.28%	68.18%	59.53%	65.98%

*Twenty-nine respondents were removed from the final baseline sample due to lack of sufficient contact information for recontact at wave 2.

representativeness. Proportions for age groups were lower for 15–17-year-olds (36.4% vs 44.5%), which may be partially due to the additional parental consent process required for youth participation under age 18 years. Proportions for gender were similar to census numbers and did not vary significantly across strata. The sample overall reflects a somewhat greater proportion of respondents from the North-East and Midwest compared with census data, and this was largely driven by differences in the listed strata. The biggest difference was in race/ethnicity. The total sample reflects a larger proportion of whites, somewhat fewer African-Americans and significantly fewer Hispanics compared with census estimates, and this, again, was largely driven by the listed sample. The proportion of respondents who reside within a city and its suburbs, also known as a metropolitan statistical area (MSA), was similar to census targets with 86.4% of respondents residing in MSAs compared with 84.4% in census estimates.

Sample retention

Retention efforts at follow-up waves were quite successful. Consistent with other longitudinal data collection, the largest proportion of attrition occurred between the baseline survey and first follow-up. Approximately 72% of the original wave 1 cohort was retained at wave 2. Retention rates remained stable through wave 3–5 (see table 4). Respondents who missed one or more waves were recontacted for every subsequent wave. Retention rates were superior for the listed sample compared with the unlisted sample. The retention pattern was similar for the refreshment samples followed over time.

DISCUSSION

This paper provides among the first detailed published descriptions of a probability-based sample that employs a hybrid approach of combining an ABS sampling frame with online data collection to recruit a longitudinal cohort of youth and young adults. These methods leverage the representativeness of an ABS probability-based frame that accurately calculates the selection probability for each respondent while benefiting from the timeliness of online data collection. Given the effectiveness of mass media campaigns and other population-level interventions in reducing youth tobacco use,³¹ survey methods to accurately and efficiently evaluate such efforts are critical to advancing public health.

This methodology provides both strong sample performance and time efficiencies for survey research efforts with youth and young adults, particularly for those seeking to obtain representative, population-based samples. ABS frames are currently considered the gold standard for developing representative household survey samples in the USA.²⁵ Among this often hard-to-reach population, we increased data collection efficiency by supplementing the ABS frame with auxiliary data to create ‘listed’ and

‘unlisted’ strata.²⁵ This method improved rates of eligibility, consent, interview completion and retention while maintaining acceptable sample representativeness on most demographic variables of interest, reducing sampling costs and maintaining an acceptable design effect.²⁵

Despite probability-based coverage, an acceptable design effect and rapid recruitment, other challenges to recruitment remained. Multiple contacts were necessary to achieve sufficient response, which increased costs primarily at baseline due to printed material mailing costs. Follow-up surveys leveraged email as much as possible to improve efficiency and reduce costs. Researchers may want to consider targeted efforts to reduce non-response by using auxiliary data, in-person surveys³² or additional survey mode options³³ during initial recruitment to reduce costs and improve response efficiencies. Further, neither ‘listed’ nor ‘unlisted’ samples adequately represented population estimates for minority populations, especially Hispanic respondents, despite all recruitment and survey materials being available in Spanish. Lower response rates among Hispanic populations is not uncommon. The practical effect of reduced response among these subpopulations is greater variance in survey estimates. Efforts to increase response among Hispanics include using available data to oversample these groups, such as targeting individuals with common Hispanic surnames,³⁴ using consumer data where Hispanic households are ‘flagged’,³⁵ targeting areas with a high Hispanic population density for ‘listed’ and ‘unlisted’ strata development, or conducting in-person interviews in areas with high Hispanic households in connection with area-based targeting.³⁶ This study employed enhanced incentives for engaging harder-to-reach subgroups. Using such procedures can help improve the representativeness of Hispanic and other minority samples, and requires that appropriate correctives are employed to address misclassification and other potential biases these approaches may introduce.²⁸

ABS sampling paired with either in-person or direct mail-based survey data collection requires significant labour and time.^{37 38} Alternatively, this method of linking ABS with web-based data collection has been viewed with scepticism given the gaps in internet access.^{25 38 39} New strategies, however, are being employed to improve access. For example, the advent of mobile devices has led to greater internet penetration in the USA and globally^{40 41} - particularly among young people,⁴² minorities and low-income groups.^{40 43} By providing respondents with home internet access⁴⁴ or access in other locations, supplementing data collection with other modes (phone, mail or in-person)^{21 45 46} and/or structuring incentives to engage those who are less likely to respond online, issues of internet access can be minimised.⁴⁵

Web-based data collection is particularly appropriate when surveying youth and young adults. These populations are more likely to respond to an internet survey than a mailed paper survey compared with older adults.^{6 47 48} Additionally, youth and young adults have the highest rates of internet use,^{40 49} are less likely to use conventional mail, and tend to change residential location, especially as youth transition to young adulthood.⁴⁶ Evidence also demonstrates online data collection almost universally reduces social desirability biases in surveys compared with interviewer-administered modes.^{50–52} This may be most relevant for surveys of tobacco use among underage youth or illegal substance use. Web surveys also allow respondents to complete the survey at their convenience, and can be efficiently optimised to facilitate survey navigation and reduce item non-response.⁵³ Finally, online data collection can allow for the viewing of visual content such as photos, videos or hyperlinks to digital content in real time⁵⁴—platforms that are likely to be familiar and accessible

Research paper

to younger populations, and increasingly important in survey research.⁴⁶ Thus, online surveys are especially valuable for media evaluation studies, as they allow for the easy inclusion of actual digital content assessing exposure to the campaign.

Data from this study demonstrated a 50% increase over a relatively short time period in respondents' use of mobile devices for completing the survey. Given the increased penetration of mobile devices across the globe, including in low-income and middle-income countries (LIMCs), online data collection via mobile devices may be relevant for obtaining representative survey samples.⁵⁵ In the USA, young adults and minorities are more likely to identify their mobile device as their main source to access the internet.⁹ Given the rapid increase in mobile phone usage by these harder-to-reach populations,⁴²⁻⁴⁹ survey instruments should be modified for the mobile platform. Online surveys can also be deployed relatively easily in multiple languages—a clear benefit to global tobacco control evaluation efforts as mobile phone ownership becomes more common in LIMCs.⁴¹

Since consistent contact with participants is a necessary component of longitudinal studies, online platforms can provide a rapid and efficient mode for engaging and retaining participants in a study. Fielding time for online data collection in this study averaged 1–2 months less, at a minimum, than estimated phone or in-person data collection (personal communication, GfK, 2016). Further, survey completion can require less time as compared with other modes.⁴⁶⁻⁵⁶ For example, taking a survey online is estimated to take approximately half the time of taking it orally (personal communication, GfK, 2016). Survey completion averaged 31.8 min for the baseline and 25 min for the follow-ups. It is likely that longer survey completion estimates would have resulted in lower response rates. Rapid, easy and efficient data collection via online surveys can benefit both the researcher and the respondent.

Retaining youth and young adults in longitudinal research is challenging. The longitudinal portion of Monitoring the Future, a national survey of 8th, 10th and 12th graders, yielded 54% participation rates in the first year after high school among seniors initially targeted. The second through sixth follow-ups after high school averaged 49% of the initial target sample.⁵⁷ A representative Massachusetts survey of 12–17-year-olds obtained retention rates of 72.8% after 2 years.⁵⁸ Probability-based longitudinal surveys targeting nationally representative samples of adult smokers in the USA have reported retention rates of 62.8% and 37.4% at waves 2 and 3.⁵ The TLC's retention rates over 2 years are either comparable or significantly higher. Evidence indicates these higher retention rates may reflect enhanced retention efforts used in this study, including listed sample, personalised communications,⁵⁹ the use of humour and varying incentives.^{60–68}

Although this methodology provides an effective and timely approach to obtaining representative survey data, it is not without limitations. Using ABS frames is most applicable for countries that have robust administrative data sources for enumerating all households across a community. Even with ABS sampling frames, recruitment and retention of this age group generally requires multiple contacts, incentives and additional engagement strategies. Studies using non-probability-based samples can use some of the strategies described here for online data collection. For areas with limited internet access, such as some LIMCs, mixed mode studies (online combined with mail and telephone) may be needed to improve representativeness. Finally, the approach described here is relatively novel for this age group, thus further research is needed to evaluate total survey error with respect to sampling frame coverage, online access, non-response and measurement errors. Analyses related

to potential sampling and response bias were conducted in some of the pretesting research and additional analyses are ongoing. Researchers should continue to assess total survey error in the context of the survey mode, costs, timeliness and feasibility.

CONCLUSIONS

Tobacco control programmes rely on rigorous evaluation to determine effectiveness at preventing and decreasing tobacco use. Collecting survey data from nationally representative samples in the age of declining response rates presents many challenges. The methodology described above takes advantage of the coverage of an ABS frame while leveraging the time efficiencies of online data collection. This approach is valuable for obtaining representative survey data among younger populations as it capitalises on their increasing access to and preference for digital communication. Given the rapidly evolving communication and technology landscapes, new approaches are needed to evaluate the effectiveness of tobacco control programmes and policies among younger populations who are most at risk of tobacco use initiation. With the penetration of digital technology worldwide, the sampling and data collection strategies described here can provide key information on challenges and potential solutions for evaluating tobacco control research efforts across the globe.

What this paper adds

- ▶ Evaluation of population-based tobacco control interventions targeting young people often demands accurate and representative survey data.
- ▶ Recruiting and retaining a large, probability-based longitudinal sample can be prohibitive in terms of cost, time and effort.
- ▶ The study detailed here provides a feasible approach that links the rigour of address-based sampling with the time efficiencies of online data collection to effectively evaluate tobacco control interventions among youth and young adults.
- ▶ Despite challenges with recruitment, this methodology provides opportunities for developing robust representative samples that can be retained over time via online platforms.

Acknowledgements The authors thank Edward Mulrow, PhD P Stat and Ned English, both of NORC at the University of Chicago, for reviewing and providing feedback on the study's methods and response rate section.

Contributors JC conceptualised and wrote the paper. DV, ECH, JMR and JMD contributed to the conceptualisation and DV contributed to writing and revisions. AS and MB contributed to writing and revisions. JMD, RKT and MF contributed to the study design, analyses and revisions.

Funding This study was funded by Truth Initiative.

Competing interests None declared.

Ethics approval Chesapeake IRB.

Provenance and peer review Not commissioned; externally peer reviewed.

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- 1 Duke JC, Alexander TN, Zhao X, *et al.* Youth's Awareness of and Reactions to The Real Cost National Tobacco Public Education Campaign. *PLoS One* 2015;10:e0144827.
- 2 Duke JC, Davis KC, Alexander RL, *et al.* Impact of a U.S. antismoking national media campaign on beliefs, cognitions and quit intentions. *Health Educ Res* 2015;30:466–83.

- 3 Hyland A, Ambrose BK, Conway KP, *et al*. Design and methods of the population assessment of tobacco and health (PATH) Study. *Tob Control* 2016.
- 4 Centers for Disease Control and Prevention. 2015. 2012-2013 National Adult Tobacco Survey: Sample Design and Methodology Summary. https://www.cdc.gov/tobacco/data_statistics/surveys/nats/pdfs/2014-methodology-report-tag508.pdf.
- 5 Thompson ME, Fong GT, Hammond D, *et al*. Methods of the international tobacco control (ITC) Four country survey. *Tob Control* 2006;15(suppl_3):iii12–iii18.
- 6 American Association for Public Opinion Research. 2016. Evaluating Survey Quality in Today's Complex Environment. <http://www.aapor.org/Education-Resources/Reports/Evaluating-Survey-Quality.aspx>.
- 7 Dutwin D, Lavrakas P. Trends in telephone outcomes, 2008-2015. *Survey Practice* 2016;9.
- 8 Blumberg SJ, Luke JV. 2015. Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, January-June 2015: National Center for Health Statistics. <https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201512.pdf>.
- 9 Zickuhr K, Smith A. 2012. Digital Differences: Pew Research Center. <http://www.pewinternet.org/2012/04/13/digital-differences/>.
- 10 Qayad M, Pierannunzi C, Chowdhury PP, *et al*. Landline and cell phone response measures in behavioral risk factor surveillance system. *Survey Practice* 2013;6.
- 11 Elkasabi M, Suzer-Gurtekin ZT, Lepkowski JM, *et al*. A comparison of ABS mail and RDD surveys for measuring consumer attitudes. *International Journal of Market Research* 2014;56:737–56.
- 12 Sinclair M, O'Toole J, Malawaraarachchi M, *et al*. Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalised recruitment approaches. *BMC Med Res Methodol* 2012;12:1.
- 13 Telephone Consumer Protection Act 47 U.S.C. § 227. : US Congress, 1991.
- 14 Yeager DS, Krosnick JA, Chang L, *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.
- 15 Callegaro M, Villar A, Krosnick J, *et al*. Online Panel Research: A Data Quality Perspective. *A critical review of studies investigating the quality of data obtained with online panels based on probability and nonprobability samples*. UK: John Wiley & Sons, 2014:23–53.
- 16 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.
- 17 Scherpenzeel CA, Das JWM. 'True' longitudinal probability-based Internet panels: Evidence from the Netherlands. In: Das JWM, Ester P, Kaczmirek L, *Social and behavioral research and the internet*. New York, NY: Taylor & Francis, 2010:77–103.
- 18 InterSurvey. InterSurvey's statement on methodology. <http://www.rickweil.com/s2211/InterSurveyMethods.html>.
- 19 Gallup. Methodology: How Does Gallup Daily Tracking Work? http://www.gallup.com/185468/gallup-daily-tracking-work.aspx?utm_source=METHODOLOGY&utm_medium=topic&utm_campaign=files.
- 20 GfK Knowledge Networks. KnowledgePanel design summary. 2013 [http://www.knowledgenetworks.com/knpanel/docs/KnowledgePanel\(R\)-Design-Summary-Description.pdf](http://www.knowledgenetworks.com/knpanel/docs/KnowledgePanel(R)-Design-Summary-Description.pdf).
- 21 Blom AG, Gathmann C, Krieger U. Setting up an online panel representative of the general population: the German internet panel. *Field methods* 2015;27:391–408.
- 22 Revilla M, Cornilleau A, Cousteaux A-S, *et al*. What is the gain in a probability-based online panel of providing internet access to sampling units who previously had no access? *Soc Sci Comput Rev* 2016;34:479–96.
- 23 Leibniz Institute for the Social Sciences. A probability-based mixed-mode access panel for the social sciences. <http://www.gesis.org/en/services/data-collection/gesis-panel/>.
- 24 Smyth JD, Dillman DA, Christian LM, *et al*. Using the internet to survey small towns and communities: limitations and possibilities in the early 21st century. *American Behavioral Scientist* 2010;53:1423–48.
- 25 Harter R, Battaglia MP, Buskirk TD, *et al*. 2016. Address-Based Sampling: American Association for Public Opinion Research. <http://www.aapor.org/Education-Resources/Reports/Address-based-Sampling.aspx>.
- 26 Centers for Disease Control and Prevention. Youth and tobacco use. 2016 http://www.cdc.gov/tobacco/data_statistics/fact_sheets/youth_data/tobacco_use/.
- 27 US Department of Health and Human Services. 2014. The Health Consequences of Smoking: 50 Yearsof Progress. A Report of the Surgeon General. Atlanta,GA: U.S. Department of Health and HumanServices, Centers for Disease Control and Prevention, National Center for Chronic Disease Preventionand Health Promotion, Office on Smoking and Health.
- 28 Pasek J, Jang SM, Cobb CL, *et al*. Can marketing data aid survey research? Examining accuracy and completeness in Consumer-File data. *Public Opin Q* 2014;78:889–916.
- 29 Taylor RM, Mohain J, Gibson F, *et al*. Novel participatory methods of involving patients in research: naming and branding a longitudinal cohort study, BRIGHTLIGHT. *BMC Med Res Methodol* 2015;15.
- 30 American Association for Public Opinion Research. 2015. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 8th Edition: AAPOR. https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/Standard-Definitions2015_8thEd.pdf.
- 31 U.S. Department of Health and Human Services, US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. *The health consequences of Smoking—50 years of Progress: A Report of the Surgeon General*. Atlanta, GA, 2014.
- 32 Montgomery R, Dennis JM, Ganesh N. 2016. Response Rate Calculation Methodology for Recruitment of a Two-Phase Probability-Based Panel: The Case of AmeriSpeak. http://d33qi0qp55mx5f5.cloudfront.net/amerispeak/ii/research/WhitePaper_ResponseRateCalculation_AmeriSpeak_2016.pdf.
- 33 English N, Ventura I, Bilgen I, *et al*, 2015. Elements of non-response across mode in ABS surveys: the use of GIS-Based modeling to understand the importance of space. American Association for Public Opinion Research, Hollywood, Florida
- 34 Word DL, Perkins RC. 1996. Building a Spanish Surname List for the 1990's— A New Approach to an Old Problem. Washington, DC: Population Division, US Bureau of the Census. <https://www.census.gov/population/documentation/twppo13.pdf>.
- 35 Link MW, Burks AT. Leveraging auxiliary data, differential incentives, and survey mode to target hard-to-reach groups in an address-based sample design. *Public Opin Q* 2013;77:696–713.
- 36 Fiscella K, Fremont AM. Use of geocoding and surname analysis to estimate race and ethnicity. *Health Serv Res* 2006;41(4 Pt 1):1482–500.
- 37 Iannacchione VG. The changing role of Address-Based sampling in survey research. *Public Opin Q* 2011;75:556–75.
- 38 Roster CA, Rogers RD, Albaum G, *et al*. A comparison of response characteristics from web and telephone surveys. *Int J Market Res* 2004;46:359–73.
- 39 McMorris BJ, Petrie RS, Catalano RF, *et al*. Use of web and in-person survey modes to gather data from young adults on sex and drug use: an evaluation of cost, time, and survey error based on a randomized mixed-mode design. *Eval Rev* 2009;33:138–58.
- 40 Perrin A, Duggan M. 2015. Americans' Internet Access: 2000-2015: Pew Research Center. <http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/>.
- 41 Pew Research Center. 2014. Emerging Nations Embrace Internet, Mobile Technology. <http://www.pewglobal.org/2014/02/13/emerging-nations-embrace-internet-mobile-technology/>.
- 42 Lenhart A. 2015. Teens, Social Media & Technology Overview 2015: Pew Research Center. <http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/>.
- 43 Smith A. 2014. African Americans and Technology Use: A Demographic Portrait : Pew Research Center. <http://www.pewinternet.org/2014/01/06/african-americans-and-technology-use/>.
- 44 GfK Knowledge Networks. 2012. KnowledgePanel Fact Sheet. <http://www.knowledgenetworks.com/fact-sheets/KnowledgePanel.pdf>.
- 45 Berzelak N, Vehovar V, Manfreda KL. Web mode as part of Mixed-Mode surveys of the general population: An approach to the evaluation of costs and errors. *Advances in Methodology and Statistics* 2015;12:45–68.
- 46 Stern MJ, Bilgen I, Dillman DA. The state of survey methodology: Challenges, dilemmas, and new frontiers in the era of the tailored design. *Field Methods* 2014;26:284–301.
- 47 Shih T-H, Xitao Fan. Comparing response rates from web and mail surveys: a Meta-Analysis. *Field Methods* 2008;20:249–71;–.
- 48 Kaplowitz MD, Hadlock TD, Levine R. A comparison of web and mail survey response rates. *Public Opin Q* 2004;68:94–101.
- 49 Pew Research Center. 2015. U.S. Smartphone Use In 2015. <http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/>.
- 50 Manfreda KL, Vehovar V. Do mail and web surveys provide the same results? *Advances in Methodology and Statistics* 2002;18:149–69.
- 51 Tourangeau R, Couper MP, Conrad FG. "Up Means Good": The Effect of Screen Position on Evaluative Ratings in Web Surveys. *Public Opin Q* 2013;77(Suppl 1):69–88.
- 52 Kreuter F, Presser S, Tourangeau R. Social desirability Bias in CATI, IVR, and web surveys: the effects of mode and question sensitivity. *Public Opin Q* 2008;72:847–65.
- 53 Callegaro M, Manfreda KL, Vehovar V. *Web survey methodology*. UK: Sage Publications Ltd, 2015.
- 54 Couper MP. *Designing effective web surveys*. New York, NY, US: Cambridge University Press, 2008.
- 55 Blyth B. Mixed mode: the only 'fitness' regime? *Int J Market Res* 2008;50:241–66.
- 56 Greenlaw C, Brown-Welty S. A comparison of web-based and paper-based survey methods: testing assumptions of survey mode and response cost. *Eval Rev* 2009;33:464–80.
- 57 Bachman JG, Johnston LD, O'Malley PM, Schulenberg JE. 2011. The Monitoring the Future Project After Thirty-Seven Years: Design and Procedures. Ann Arbor, MI: Institute for Social Research, The University of Michigan.
- 58 Albers AB, Biener L, Siegel M, *et al*. Impact of parental home smoking policies on policy choices of independently living young adults. *Tob Control* 2009;18:245–8.
- 59 Sánchez-Fernández J, Muñoz-Leiva F, Montoro-Rios FJ. Improving retention rate and response quality in Web-based surveys. *Comput Human Behav* 2012;28:507–14.
- 60 Hanna KM, Scott LL, Schmidt KK. Retention strategies in longitudinal studies with emerging adults. *Clin Nurse Spec* 2014;28:41–5.

Research paper

- 61 Henderson M, Wight D, Nixon C, *et al.* Retaining young people in a longitudinal sexual health survey: a trial of strategies to maintain participation. *BMC Med Res Methodol* 2010;10:1.
- 62 Boys A, Marsden J, Stillwell G, *et al.* Minimizing respondent attrition in longitudinal research: practical implications from a cohort study of adolescent drinking. *J Adolesc* 2003;26:363–73.
- 63 Morrison TC, Wahlgren DR, Hovell MF, *et al.* Tracking and follow-up of 16,915 adolescents: minimizing attrition Bias. *Control Clin Trials* 1997;18:383–96.
- 64 Rath JM, Williams VF, Villanti AC, *et al.* Boosting online response rates among nonresponders: a dose of funny. *Soc Sci Comput Rev* 2016.
- 65 Muñoz-Leiva F, Sánchez-Fernández J, Montoro-Ríos F, *et al.* Improving the response rate and quality in Web-based surveys through the personalization and frequency of reminder mailings. *Qual Quant* 2010;44:1037–52.
- 66 Heerwegh D. Effects of personal salutations in E-mail invitations to participate in a web survey. *Public Opin Q* 2005;69:588–98.
- 67 Gajic A, Cameron D, Hurley J. The cost-effectiveness of cash versus lottery incentives for a web-based, stated-preference community survey. *Eur J Health Econ* 2012;13:789–99.
- 68 Göritz AS. Incentives in web studies: methodological issues and a review. *International Journal of Internet Science* 2006;1:58–70.



Recruiting and retaining youth and young adults: challenges and opportunities in survey research for tobacco control

Jennifer Cantrell, Elizabeth C Hair, Alexandria Smith, Morgane Bennett, Jessica Miller Rath, Randall K Thomas, Mansour Fahimi, J Michael Dennis and Donna Vallone

Tob Control published online April 21, 2017

Updated information and services can be found at:
<http://tobaccocontrol.bmj.com/content/early/2017/04/20/tobaccocontrol-2016-053504>

These include:

References

This article cites 35 articles, 12 of which you can access for free at:
<http://tobaccocontrol.bmj.com/content/early/2017/04/20/tobaccocontrol-2016-053504#BIBL>

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:
<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:
<http://group.bmj.com/subscribe/>