Preventive Medicine 102 (2017) 93–99

Contents lists available at ScienceDirect

Preventive Medicine

journal homepage: www.elsevier.com/locate/ypmed

Crowdsourced data collection for public health: A comparison with nationally representative, population tobacco use data

John D. Kraemer [a], Andrew A. Strasser [b], Eric N. Lindblom [c], Raymond S. Niaura [d,e,f], Darren Mays [f,*]

[a] Department of Health Systems Administration, School of Nursing & Health Studies, Georgetown University, Washington, DC, United States
[b] Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
[c] O'Neill Institute for National and Global Health Law, Georgetown University Law Center, Washington, DC, United States
[d] Schroeder Institute for Tobacco Research and Policy Studies, Truth Initiative, Washington, DC, United States
[e] Department of Health, Behavior, and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States
[f] Department of Oncology, Georgetown University Medical Center, Cancer Prevention & Control Program, Lombardi Comprehensive Cancer Center, Washington, DC, United States

ARTICLE INFO

Article history:
Received 28 February 2017
Received in revised form 30 June 2017
Accepted 5 July 2017
Available online 08 July 2017

Keywords:
Tobacco control
Young adult
Crowdsourcing

ABSTRACT

Introduction. Internet-based crowdsourcing is increasingly used for social and behavioral research in public health, however the potential generalizability of crowdsourced data remains unclear. This study assessed the population representativeness of Internet-based crowdsourced data.

Methods. A total of 3999 U.S. young adults ages 18 to 30 years were recruited in 2016 through Internet-based crowdsourcing to complete measures taken from the 2012–2013 National Adult Tobacco Survey (NATS). Post-hoc sampling weights were created using procedures similar to the NATS. Weighted analyses were conducted in 2016 to compare crowdsourced and publicly-available 2012–2013 NATS data on demographics, tobacco use, and measures of tobacco perceptions and product warning label exposure.

Results. Those in the crowdsourced sample were less likely to report an annual household income of $50,000 or greater, and e-cigarette, waterpipe, and cigar use were more prevalent in the crowdsourced sample. High proportions of both samples indicated cigarette smoking is very harmful and very addictive. Comparable proportions of non-smokers and smokers reported cigarette warning label exposure, however the likelihood of reporting that smoking is very harmful by frequency of warning label exposure was lower among smokers in the crowdsourced sample.

Conclusions. Our findings indicate that crowdsourced samples may differ demographically and may not produce generalizable estimates of tobacco use prevalence relative to population data after post-hoc sample weighting. However, correlational analyses in crowdsourced samples may reasonably approximate population data. Future studies can build from this work by testing additional methodological strategies to improve crowdsourced sampling strategies.

1. Introduction

Internet crowdsourcing, defined as a "distributed problem-solving and production model that leverages the collective intelligence of online communities," is a tool with potential to address public health challenges (Brabham et al., 2014). Crowdsourcing offers an efficient way to obtain information from online respondents more quickly than some conventional data collection tools. Crowdsourcing applications entail varying involvement from participants, including answering questions (e.g., surveys); providing feedback on concepts (e.g., policies, programs); coding data (e.g., images); and creating user-generated content (e.g., communication messages) (Brabham et al., 2014). Researchers are increasingly using crowdsourcing data collection, particularly in social and behavioral sciences (Bohannon, 2016).

Crowdsourcing has value because data collection is efficient and relatively low cost and participants are readily available without geographic constraints (Brabham et al., 2014; Gosling and Mason, 2015; Mason and Suri, 2012). Examples of research using crowdsourcing include tobacco control (Mays et al., 2016a; Mays et al., 2016b; Leas et al., 2016; Brewer et al., 2016), skin cancer prevention (Mays and Tercyak, 2015), and sexual behavior (Syme et al., 2017). Research using crowdsourcing includes observational studies to characterize specific constructs such as health beliefs, and correlational investigations of how exposures such as health messaging relate to outcomes such as beliefs or behavior (Mays et al., 2016a; Mays et al., 2016b; Leas et al., 2016; Brewer et al., 2016; Mays and Tercyak, 2015; Syme et al., 2017). Peer-reviewed papers using a single crowdsourcing platform (Amazon Mechanical Turk) increased from 61 in 2011 to 1120 in 2015 (Bohannon, 2016). Increasing interest has spurred development of methodological tools for researchers (Litman et

* Corresponding author at: Department of Oncology, Georgetown University Medical Center, Cancer Prevention & Control Program, Lombardi Comprehensive Cancer Center, 3300 Whitehaven Street NW, Suite 4100, Washington, DC 20007, United States.
E-mail address: dmm239@georgetown.edu (D. Mays).

http://dx.doi.org/10.1016/j.ypmed.2017.07.006
0091-7435/

al., 2016), and crowdsourcing is appearing in funding opportunities from agencies such as the National Institutes of Health.

Crowdsourcing platforms can replicate data collected using validated behavioral measures and tasks (Briones and Benham, 2016; Crump et al., 2013), and crowdsourcing samples may provide greater demographic diversity than traditional convenience samples (e.g., college students) (Briones and Benham, 2016; Berinsky et al., 2012). There are concerns about generalizability since crowdsourced participants are those with technology access who are motivated to engage in research, and because there may be relatively small numbers of individuals in crowdsourced participant pools who meet specific eligibility criteria at any given time (Brabham et al., 2014; Chandler et al., 2014). There is also some evidence indicating that crowdsourced samples may differ from the population on measures relevant to public health research, such as political beliefs (Chandler and Shapiro, 2016).

Several recent studies have used crowdsourcing to examine questions aimed at informing Food and Drug Administration (FDA) tobacco regulations (Mays et al., 2016a; Mays et al., 2016b; Leas et al., 2016; Brewer et al., 2016; Pearson et al., 2016). Such an efficient data collection approach has potential value for tobacco regulatory science because FDA is charged with supporting regulations using population data and often such data need to be generated within a short timeframe to inform regulations (Husten and Deyton, 2013). Crowdsourcing also provides the capability to reach priority groups to inform tobacco regulations, such as tobacco users and nonusers and specific demographic groups (Husten and Deyton, 2013). Although crowdsourcing is increasingly used for tobacco research, as noted above prior studies have used crowdsourcing for topics ranging from skin cancer prevention (Mays and Tercyak, 2015) to sexual risk behavior (Syme et al., 2017), and its use is increasing overall (Bohannon, 2016). Empirical evidence on how crowdsourcing can be used to inform tobacco regulation as a case study can guide research in other public health domains as well.

This study empirically examined the potential contributions of crowdsourced data in public health research by comparing crowdsourced data to nationally representative U.S. survey data. The study focuses on tobacco use as an example because deidentified, nationally representative data from the National Adult Tobacco Survey (NATS) are publicly available from the U.S. Centers for Disease Control and Prevention (CDC) (Centers for Disease Control and Prevention, 2016). This study additionally focused on young adults ages 18 to 30 years because they are defined as a priority population for tobacco research (National Cancer Institute, 2016) and can be accessed through crowdsourcing platforms where participation is limited to adults ages 18 and older. Our aim was to determine if crowdsourced data provides similar estimates of demographics, tobacco use behavior, and tobacco risk perceptions by comparing data collected through Amazon Mechanical Turk to the NATS. Additionally, we aimed to determine if correlational analyses in crowdsourced data are similar to population data, drawing from research indicating exposure to tobacco warning labels can affect risk perceptions (Mays et al., 2016a; Mays et al., 2016b; Leas et al., 2016; Brewer et al., 2016). Our study focused on these specific measures because monitoring population-level trends in tobacco use behavior, perceptions, and exposure to interventions such as warning labels is critical to tobacco regulatory decision-making (Husten and Deyton, 2013). We tested the study aims by comparing data collected through Amazon Mechanical Turk to the NATS on demographics, tobacco use behaviors and perceptions, and exposure to tobacco warnings using parallel measures and sample weighting strategies for crowdsourced data.

2. Methods

2.1. Sampling

2.1.1. National Adult Tobacco Survey

The most recent population-based tobacco use dataset at the time of the study was the 2012–2013 NATS (Center for Disease Control and Prevention, 2015). The NATS is a stratified, random-digit dialed landline and cellular telephone survey of non-institutionalized adults ≥ 18 years old residing in the 50 U.S. states and the District of Columbia (Agaku et al., 2014). From October 2012 to July 2013, 60,192 interviews were conducted (44.9% response rate) including 6682 young adults ages 18 to 30 in the analytic sample (Center for Disease Control and Prevention, 2015; Agaku et al., 2014). NATS data are weighted by inverse probability of selection, adjusted for nonresponse and household characteristics, and raked to population totals on state, age, gender, race/ethnicity, marital status, education, and phone type (Centers for Disease Control and Prevention, 2014).

2.2. Crowdsourced data

Crowdsourced data were collected in April 2016 through Amazon Mechanical Turk, an Internet marketplace where researchers can post “human intelligence tasks” including surveys or other data collection (Crump et al., 2013). After reviewing a brief study description inviting them to take a survey with questions about tobacco use, Mechanical Turk members interested in participating reviewed a more comprehensive description with a link to a consent form and eligibility screener. Young adults ages 18 to 30 were eligible to participate, access to the task was limited to those with accounts registered in the U.S. All study procedures were implemented in English without translation to other languages. To ensure representation of cigarette smokers, smoking status was assessed at screening and smokers were oversampled to reflect 40% of the sample. Post-hoc weighting (described below) was used to adjust to the national smoking prevalence in the NATS. Eligible, consenting individuals proceeded to an online survey consisting of measures described below. The target sample was 4000 respondents—the maximum practical sample that could be achieved with study resources and a target that meets or exceeds those of similar crowdsourced studies to date (Mays et al., 2016a; Mays et al., 2016b; Leas et al., 2016; Brewer et al., 2016). Participants completing procedures were given a $1 monetary credit through Mechanical Turk. The data collection protocol was reviewed and determined to be exempt by Georgetown University's IRB.

2.3. Measures

For comparison to the NATS dataset, the crowdsourced data collection used measures directly from the NATS wherever possible. Any differences are noted below. We selected a subset of measures from the NATS for crowdsourced data collection due to practical constraints on the number of items that could be implemented using crowdsourcing (Mason and Suri, 2012) and based on priority topics for tobacco regulatory science described above (Husten and Deyton, 2013).

2.4. Demographics

Demographics were assessed using NATS items including age, gender, race/ethnicity, education, and marital status (Agaku et al., 2014). NATS measures household income using multi-question probing identifying general income levels (e.g., less than $50,000) and determining specific levels through follow-up questions (e.g., $30,000 to $40,000, $40,000 to $50,000) (Center for Disease Control and Prevention, 2015). This type of measure could not practically be administered online, so a single item asking “What was your household income before taxes last year? Please report the income that is most important to you, whether that is your own income or your parents” was used (Mays et al., 2016a; Mays et al., 2016b). Respondents were grouped into similar income categories across the samples.

2.5. Tobacco use

Measures captured use of cigarettes, electronic cigarettes, hookah (waterpipe tobacco), cigars/cigarillos/filtered little cigars, and smokeless tobacco (Agaku et al., 2014). Consistent with the NATS, current cigarette smokers were defined as those who had smoked ≥100 lifetime cigarettes and now smoked cigarettes every day or some days. The following thresholds from the NATS were used for other tobacco products: electronic cigarettes (ever use), hookah (ever use), cigars/cigarillos/filtered little cigars (≥50 times), and smokeless tobacco (≥20 times) (Agaku et al., 2014). Those meeting thresholds were defined as current users if they also reported using the product every day or some days (Agaku et al., 2014).

2.6. Perceptions of cigarette smoking

Three items from the NATS assessed perceived harm and addictiveness of smoking (Center for Disease Control and Prevention, 2015). One assessed "How harmful do you think cigarette smoking is to a person's health?" with response options for "Not at all harmful," "Moderately harmful," and "Very harmful." A second assessed perceived addictiveness using similar response options ("Not at all addictive," "Moderately addictive," "Very addictive"). We analyzed these variables by response category and by comparing those reporting very harmful or very addictive to any other response.

The third item assessed how much "people harm themselves when they smoke some days but not every day" with response options for "Not at all," "A little," "Somewhat," and "A lot." We analyzed this variable by response category and by comparing those reporting "A lot" to all other responses. NATS presents this third item to respondents ages 18 to 29, so comparisons across the samples were limited to this age group.

2.7. Exposure to tobacco warning labels

Two items assessed frequency of exposure to warning labels on cigarette and smokeless tobacco packaging (Agaku et al., 2016): "How often, if at all, have you seen a health warning on cigarette (smokeless tobacco) packages in the past 30 days?" Response options included "Very often," "Often," "Sometimes," "Rarely," and "Never."

2.8. Weighting and sample merging

The crowdsourced sample was weighted to be equivalent to the weighted NATS sample of 18–30 year-olds on four variables: gender, race/ethnicity, educational attainment, and age. We chose these variables because they are among the characteristics used in NATS weighting procedures that are associated with tobacco use in the population (Agaku et al., 2016; Hu et al., 2016). We conducted weighting separately for smokers and non-smokers by raking using Stata's *ipfweight* command. A composite weight was then created to account for oversampling of current smokers in the crowdsourced sample (Deville et al., 1993). After creating sampling weights for the crowdsourced data, the NATS and crowdsourced samples were merged into a single file with sampling strata kept separate to preserve the sampling design. Weights were rescaled so both surveys contributed an equal number of weighted observations. Variables were fully accorded between samples.

2.9. Statistical analyses

Weighted estimates of demographics, tobacco use, and perceptions of smoking were created for the NATS and crowdsourced samples. Differences between weighted estimates from the samples and their 95% confidence intervals (CIs) were produced as linear combinations of coefficients. Differences between the NATS and crowdsourced samples in the frequency of reported exposure to cigarette and smokeless tobacco warning labels was estimated separately for users and non-users of these products in the same manner, as were differences between the NATS and crowdsourced samples in the associations between perceptions of smoking and warning label exposure. Taylor linearization was used to adjust standard errors for survey designs. Missing data were minimal (<1% for each variable) therefore observations with missing data were excluded from analyses without imputation. Stata version 14.1 was used for analyses.

3. Results

3.1. Sample characteristics

In the crowdsourced data collection, 9918 individuals were screened for eligibility, and 3999 (40.3%) met eligibility criteria and completed study procedures. Table 1 displays demographics of both samples. Weighted samples were similar on variables used in weighting (gender, race/ethnicity, education, cigarette smoking) indicating successful application of weighting procedures to crowdsourced data. The primary difference was in household income: a smaller proportion of the crowdsourced sample reported an annual income >$50,000 (25.1% [95% CI 23.1–27.2] vs. 45.4% [95% CI 43.9–47.0]), and a larger proportion of the NATS sample refused to report income (6.4% [95% CI 5.6–7.2] vs. 2.7% [95% CI 2.0–3.7]). Supplemental Tables S1 and S2 in the Appendix display similar patterns in the demographic differences between the samples when stratified by cigarette smoking status.

3.2. Non-cigarette tobacco use and perceptions of smoking

Table 2 provides a comparison of weighted NATS and crowdsourced samples on non-cigarette tobacco use and perceptions of smoking. Compared with the NATS, a larger proportion of the crowdsourced sample reported using electronic cigarettes (13.9% [95% CI 12.4–15.6] vs. 2.6% [95% CI 2.1–3.0]), waterpipe (6.4% [95% CI 5.2–7.8] vs. 1.9% [95% CI 1.5–2.3]), and cigars (8.4% [95% CI 7.1–9.9] vs. 3.2% [95% CI 2.7–3.8]). Differences between the samples for non-cigarette tobacco use were larger among current smokers than non-smokers (Supplemental Tables S1 and S2).

Although the proportion of participants reporting that smoking is very harmful was high in the NATS (88.8%) and crowdsourced (81.8%) samples, it was lower in the crowdsourced sample (−7.1% [95% CI difference −9.2, −4.9]) (Table 2). There were no significant differences between the samples in the proportion of participants indicating smoking some days is harmful or smoking is very addictive (Table 2).

3.3. Exposure to tobacco warning labels

Fig. 1 displays frequency of exposure to cigarette warning labels by smoking status; underlying data are in Supplemental Table S3. Non-smokers in the crowdsourced sample were less likely than NATS non-smokers to report exposure to cigarette warnings very often (16.9% [95% CI 14.4, 19.7] vs. 21.7% [95% CI 20.4, 23.1]), more likely to report exposure sometimes (20.5% [95% CI 18.0, 23.4] vs. 10.1% [95% CI 9.1–11.2]) and rarely (20.5% [95% CI 18.2, 23.0] vs. 11.6% [95% CI 10.5, 12.7]), and less likely to report no exposure (29.8% [95% CI 27.2, 32.6] vs. 44.9% [95% CI 43.3, 46.6]). Smokers in the crowdsourced sample were less likely than NATS smokers to report warning label exposure very often (48.4% [95% CI 44.9, 52.0] vs. 68.3% [95% CI 65.1, 71.3]) and more likely to report exposure sometimes (15.9% [95% CI 13.6, 18.6] vs. 7.2% [95% CI 5.5, 9.3]) and rarely (7.5% [95% CI 5.8, 9.7] vs. 4.6% [95% CI 3.4, 6.2]).

Supplemental Fig. S1 displays similar data for smokeless tobacco warnings among users and non-users, with the underlying data in Supplemental Table S3. Results for smokeless tobacco warning labels are similar to those for cigarette warning labels.

Fig. 2 displays the proportion of participants reporting that cigarette smoking is very harmful by exposure to cigarette warning labels, with underlying data in Supplemental Table S4. A majority of non-smokers in both samples indicated smoking is harmful with relatively little variation by cigarette warning label exposure. Crowdsourced smokers were less likely than NATS smokers to report smoking is very harmful across almost all levels of warning label exposure.

Table 1
Comparison of demographics and cigarette smoking between 2012 and 2013 National Adult Tobacco Survey (NATS) and crowdsourced samples.

	Weighted NATS	Weighted crowdsourced	Difference	Unweighted crowdsourced	Difference
Age (mean, CI)	23.8 (23.7, 24.0)	24.2 (24.0, 24.3)	0.3 (0.1, 0.5)	24.9 (24.8, 25.0)	1.0 (0.9, 1.2)
Female gender	48.5 (47.0, 50.0)	48.4 (45.8, 51.0)	−0.1 (−3.1, 2.9)	47.0 (45.5, 48.6)	−1.5 (−3.6, 0.7)
Race/ethnicity					
Non-Hispanic White	55.4 (53.9, 56.9)	55.3 (52.6, 58.0)	−0.1 (−3.2, 3.0)	70.6 (69.1, 72.0)	15.2 (13.1, 17.2)
Non-Hispanic Black	10.6 (9.6, 11.6)	10.6 (9.0, 12.4)	0.0 (−1.9, 2.0)	8.1 (7.3, 9.0)	−2.5 (−3.8, −1.2)
Non-Hispanic, other	12.1 (11.2, 13.1)	12.1 (10.5, 13.9)	0.0 (−1.9, 2.0)	10.2 (9.3, 11.1)	−1.9 (−3.3, −0.6)
Hispanic	21.9 (20.6, 23.3)	22.0 (19.4, 24.8)	0.0 (−3.0, 3.0)	11.2 (10.2, 12.2)	−10.7 (−12.4, −9.1)
Education					
High school or less	46.0 (44.5, 47.5)	46.0 (43.3, 48.8)	0.1 (−3.1, 3.2)	13.9 (12.9, 15.0)	−32.1 (−33.9, −30.2)
Some college	34.2 (32.8, 35.6)	34.2 (32.2, 36.3)	−0.1 (−2.4, 2.5)	42.8 (41.2, 44.3)	8.6 (6.5, 10.7)
4-year degree+	19.8 (18.9, 20.8)	19.7 (18.4, 21.1)	−0.1 (−1.8, 1.5)	43.3 (41.8, 44.8)	23.5 (21.7, 25.3)
Income					
<20,000	12.0 (11.0, 13.1)	24.2 (21.9, 26.7)	12.2 (9.6, 14.8)	18.8 (17.7, 20.1)	6.8 (5.2, 8.4)
20,000–49,999	36.2 (34.6, 37.7)	47.9 (45.3, 50.5)	11.8 (8.8, 14.8)	45.8 (44.2, 47.3)	9.6 (7.4, 11.8)
50,000+	45.4 (43.9, 47.0)	25.1 (23.1, 27.2)	−20.3 (−23.0, −17.7)	32.8 (31.3, 34.3)	−12.7 (−14.8, −10.5)
Prefer not to say	6.4 (5.6, 7.2)	2.7 (2.0, 3.7)	−3.6 (−4.7, −2.5)	2.6 (2.2, 3.1)	−3.8 (−4.7, −2.8)
Marital status					
Single	62.3 (60.9, 63.7)	56.2 (53.6, 58.7)	−6.1 (−9.1, −3.2)	54.0 (52.4, 55.5)	−8.3 (−10.4, −6.2)
Living w/ partner	14.8 (13.8, 15.9)	20.9 (18.9, 23.1)	6.1 (3.7, 8.4)	21.3 (20.0, 22.6)	6.5 (4.8, 8.1)
Married	18.7 (17.7, 19.7)	19.0 (17.1, 21.1)	0.3 (−1.9, 2.6)	21.0 (19.8, 22.3)	2.3 (0.7, 4.0)
Divorced, widowed, separated, or other	4.2 (3.6, 4.8)	3.9 (3.0, 5.1)	−0.3 (−1.5, 1.0)	3.7 (3.2, 4.3)	−0.5 (−1.3, 0.4)
Current smoker	21.0 (19.8, 22.2)	21.0 (19.8, 22.2)	0.0 (−1.7, 1.7)	40.2 (40.2, 40.2)	19.2 (18.0, 20.4)

Note: Data displayed are percent of the samples and 95% confidence intervals unless otherwise indicated.

Supplemental Figs. S2 and S3 display similar patterns across the two samples for perceived harm from smoking cigarettes some days but not all days and perceived smoking addictiveness.

4. Discussion

This study compared crowdsourced data on young adult tobacco use to a population survey using parallel measures and weighting procedures for crowdsourced data. The findings highlight strengths and limitations of crowdsourced data that can be considered when conducting and interpreting crowdsourced research and areas of future research.

We recruited a crowdsourced sample of nearly 4000 young adults in less than one week. After weighting, the crowdsourced sample was comparable to the NATS data on demographics involved in the weighting process, as expected. The samples were weighted to be comparable on smoking prevalence, and both reflected the general finding of prior studies that smokers are more likely than non-smokers to endorse non-cigarette tobacco use (Popova and Ling, 2013; Weaver et al., 2016). We found comparably high proportions of young adults perceived smoking to be harmful and addictive and similar patterns where more frequent exposure to tobacco warnings was reported among current users than non-users in both samples. Finally, we observed similar trends overall that participants were more likely to perceive smoking is harmful with greater reported exposure to cigarette warning labels.

There were also notable differences between the samples. Household income diverged between the samples and the proportion of respondents reporting use of most non-cigarette tobacco products was greater in the crowdsourced sample. The NATS employs a national sampling frame and methods to ensure correct sampling of cellular phone-only households (Center for Disease Control and Prevention, 2015). Compared to U.S. Census Bureau estimates, NATS' young adult income distribution appears representative (United States Census Bureau, 2016). It is possible income distributions differed between samples because individuals with lower household incomes are overrepresented in crowdsourcing platforms or due to differing measurement approaches used in the NATS and crowdsourced data. A greater proportion of NATS respondents declined reporting household income, which may reflect different comfort in responding to an anonymous online survey versus a governmental telephone interview (Kreuter et al., 2009). Weighted income distribution in the crowdsourced sample was similar to research in an online nationally representative young adult sample, further supporting this possibility (Williams et al., 2016). Prior studies have also shown that crowdsourced samples tend to over-represent respondents with lower incomes, a finding consistent with our data (Berinsky et al., 2012; Levay et al., 2016).

Research comparing telephone probability samples, Internet probability samples, and Internet convenience samples suggests telephone and Internet probability samples are more demographically

Table 2
Comparison of current tobacco use and perceptions of cigarette smoking between 2012 and 2013 National Adult Tobacco Survey (NATS) and crowdsourced samples.

	Weighted NATS	Weighted crowdsourced	Difference	Unweighted crowdsourced	Difference
Current use of…					
Electronic cigarettes	2.6 (2.1, 3.0)	13.9 (12.4, 15.6)	11.4 (9.7, 13.1)	19.4 (18.3, 20.5)	16.8 (15.6, 18.0)
Waterpipe tobacco	1.9 (1.5, 2.3)	6.4 (5.2, 7.8)	4.5 (3.1, 5.8)	7.8 (7.0, 8.6)	5.9 (5.0, 6.8)
Cigar products	3.2 (2.7, 3.8)	8.4 (7.1, 9.9)	5.2 (3.7, 6.7)	11.0 (10.1, 11.9)	7.8 (6.7, 8.8)
Smokeless tobacco	3.4 (3.0, 3.9)	4.2 (3.2, 5.6)	0.8 (−0.5, 2.1)	4.3 (3.7, 5.0)	0.9 (0.1, 1.7)
Perceptions of cigarette smoking					
Smoking is very harmful	88.8 (87.9, 89.7)	81.8 (79.7, 83.7)	−7.1 (−9.2, −4.9)	79.1 (77.9, 80.3)	−9.7 (−11.2, −8.2)
Smoking some days is harmful[a]					
Not at all/a little	13.7 (12.7, 14.9)	13.3 (11.5, 15.4)	−0.4 (−2.6, 1.8)	14.1 (13.0, 15.2)	0.3 (−1.2, 1.9)
Somewhat	38.5 (36.9, 40.0)	40.2 (37.7, 42.9)	1.8 (−1.3, 4.8)	43.8 (42.2, 45.4)	5.3 (3.1, 7.6)
A lot	47.8 (46.2, 49.4)	46.4 (43.8, 49.1)	−1.4 (−4.5, 1.8)	42.1 (40.6, 43.7)	−5.7 (−7.9, −3.4)
Smoking is very addictive	71.9 (70.5, 73.3)	73.7 (71.2, 75.9)	1.8 (−1.0, 4.5)	73.3 (72.0, 74.7)	1.4 (−0.5, 3.4)

Note: Data displayed are percent of the samples and 95% confidence intervals. Smokeless tobacco includes chew, snuff, or dip. Cigar products include cigars, cigarillos, and little filtered cigars.

[a] Comparison is restricted to 18 to 29 year old participants based on National Adult Tobacco Survey age-based skip patterns.

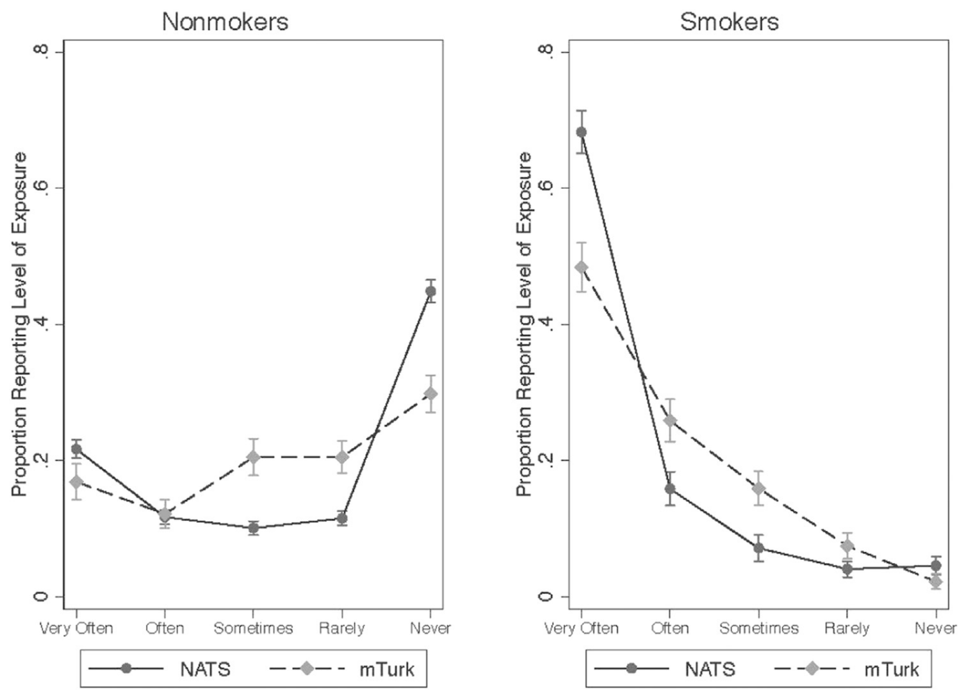

Fig. 1. Cigarette warning label exposure by cigarette smoking status in the 2012–2013 National Adult Tobacco Survey and crowdsourced samples.

representative of the U.S. population even with post-hoc weighting of non-probability data (Chang and Krosnick, 2009). However, non-probability Internet sampling was also more likely to draw participants familiar with the survey subject (Chang and Krosnick, 2009). This is important with respect to our finding that non-cigarette tobacco use was more prevalent in the crowdsourced sample, and this difference that was even more pronounced among current smokers. This suggests framing of the crowdsourced survey (a survey on tobacco) may have contributed to the higher proportion of crowdsourced respondents endorsing non-cigarette tobacco use by making it more likely that tobacco users would participate. This could be addressed by framing descriptions for crowdsourced data collection more generally (e.g., a survey on health) to reduce the likelihood that those familiar with the survey subject will be more likely to respond. Research examining effects of different framing of crowdsourced studies is an important avenue for further study.

Although we observed similar patterns in perceived harms of smoking, exposure to tobacco warnings, and correlations between these two variables among smokers and non-smokers in the samples, some estimates differed significantly. Generally, in the crowdsourced sample reported exposure to tobacco product warnings was lower among tobacco users and the correlation between exposure frequency and perceived harms of smoking was attenuated. This finding is difficult to interpret given evidence that text-only tobacco warning labels used

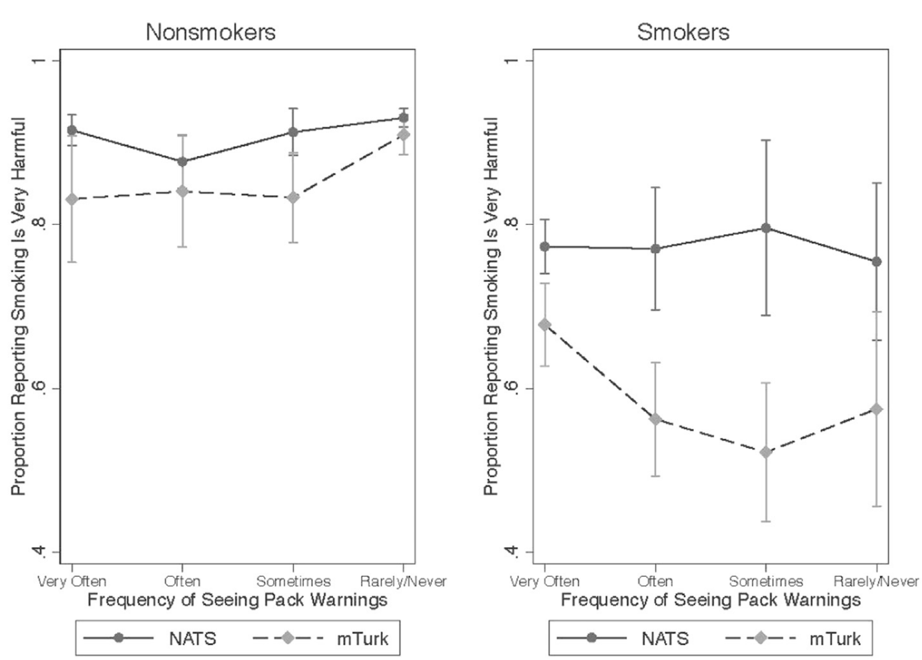

Fig. 2. Perceived harm of cigarette smoking by exposure to cigarette warning labels and smoking status in the 2012–2013 National Adult Tobacco Survey (NATS) and crowdsourced data.

in the U.S. are infrequently attended to and have limited effects on tobacco perceptions (Noar et al., 2015). Differences in mode of survey administration across the studies may have also affected participants' responses—for example a telephone survey may lead to more socially desirable reporting of warning exposure and perceptions than an anonymous, online survey. This finding is also consistent with research demonstrating variance in perceptions of smoking across probability and crowdsourced samples (Brewer et al., 2016) and may be a feature of greater measurement error in non-probability samples (Yeager et al., 2011).

Our findings should be interpreted in light of limitations. We focused on young adult tobacco use illustratively; additional research should determine the strengths and limitations of crowdsourced data collection in other populations and for other public health issues. Data collection relied on self-reported measures administered in English only. Although all measures used have been validated they are subject to potential reporting biases. As noted above, it is also possible the study findings were influenced by different modes of survey administration. Post-hoc weighting of crowdsourced data was informed by the NATS' approach but relied on a limited set of variables. Weighting based on other variables (e.g., geographic location) may improve crowdsourced data's representativeness. Population-level trends in tobacco use may have contributed to differences observed between the data sets. NATS data were collected 3 years earlier; the prevalence of young adult non-cigarette tobacco product use, including electronic cigarettes and waterpipe, increased modestly during this period (Agaku et al., 2014; Hu et al., 2016). The prevalence of non-cigarette tobacco use in another recent population survey of U.S. adults was similar to the NATS (Kasza et al., 2017), and increases in non-cigarette tobacco use prevalence were observed among U.S. youth during this time, some of whom may have reached young adulthood and contributed to observed differences (Arrazola et al., 2015). Given the time lag between federal agencies' collection of population survey data and its release to the public this limitation is difficult to directly address. However, the magnitude of population increases in prevalence is insufficient to account for observed differences between the NATS and crowdsourced samples. This suggests the possibility that tobacco users may be overrepresented in crowdsourced data. Finally, our approach does not account for all potential unmeasured confounders that may have affected observed differences between the samples. As evidence on the characteristics of crowdsourced samples grows, this will help to better understand the potential influence of such confounding variables across studies (Chandler and Shapiro, 2016). It will be important to address these methodological limitations through continued comparisons of crowdsourced and population-based data with careful attention to the timing of data collection, implementing additional measures of confounding variables, and expanding analyses to other population data sources to better understand if the findings observed were specific to the NATS.

Despite these limitations, our study provides insights into the strengths and limitations of crowdsourced data related to tobacco and potentially other public health domains. Even after post-hoc sample weighting, crowdsourced samples may differ demographically and may not provide generalizable estimates of behaviors such as tobacco use relative to population data. These findings suggest crowdsourced data collection likely provides an efficient means of gathering data at the stages of idea development, designing public health interventions, and ascertaining feedback on alternative approaches, as is illustrated in other recent studies as well (Mays et al., 2016a; Mays et al., 2016b; Leas et al., 2016; Pearson et al., 2016). Our data and other recent work (Leas et al., 2016) also suggest crowdsourcing data collection may be valuable when targeted samples are desired, given for example the relatively higher prevalence of tobacco users in the crowdsourced data. However, for the purposes of generalizing to larger populations the adequacy of this sampling method is in need of further testing. This includes research to understand the potential value of crowdsourced data for purposes that require greater generalizability, such as monitoring of population trends in behaviors and other outcomes. Continued comparisons of crowdsourced data to population data sources including the NATS and other nationally representative surveys, such as the Behavioral Risk factor Surveillance Surveys or National Health Interview Surveys, will be informative. Additional research on methodological approaches to improve sampling (e.g., using quotas) and post-hoc adjustments through weighting and other means will advance this research area as well.

Conflicts of interest

All authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Georgetown University Center of Excellence in Regulatory Science and Innovation (CERSI; U01FD004319), a collaborative effort between the university and the Food and Drug Administration (FDA) to promote regulatory science through innovative research and education. This research was also supported in part by the National Institutes of Health (NIH) and the FDA Center for Tobacco Products under NIH grant number K07CA172217 and by the Georgetown Lombardi Comprehensive Cancer Center Support under NIH grant number P30CA051008. The study sponsors had no role in the study design; in the collection, analysis and interpretation data; in the writing of the report; and in the decision to submit the paper for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the FDA. The authors thank Charlene Kuo and Kathryn Rehberg for their assistance with manuscript preparation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ypmed.2017.07.006.

References

Agaku, I.T., King, B.A., Husten, C.G., et al., 2014. Tobacco product use among adults—United States, 2012–2013. MMWR Morb. Mortal. Wkly Rep. 63 (25), 542–547 (doi: mm6325a3 [pii]).

Agaku, I.T., Singh, T., Rolle, I.V., Ayo-Yusuf, O.A., 2016. Exposure and response to current text-only smokeless tobacco health warnings among smokeless tobacco users aged ≥18years, United States, 2012–2013. Prev. Med. 87:200–206. http://dx.doi.org/10.1016/j.ypmed.2016.02.014.

Arrazola, R.A., Singh, T., Corey, C.G., et al., 2015. Tobacco use among middle and high school students - United States, 2011–2014. MMWR Morb. Mortal. Wkly Rep. 64 (14), 381–385 (doi: mm6414).

Berinsky, A.J., Huber, G.A., Lenz, G.S., 2012. Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. Polit. Anal. 20 (3):351–368. http://dx.doi.org/10.1093/pan/mpr057.

Bohannon, J., 2016. Mechanical Turk upends social sciences. Science 352 (6291): 1263–1264. http://dx.doi.org/10.1126/science.352.6291.1263.

Brabham, D.C., Ribisl, K.M., Kirchner, T.R., Bernhardt, J.M., 2014. Crowdsourcing applications for public health. Am. J. Prev. Med. 46 (2):179–187. http://dx.doi.org/10.1016/j.amepre.2013.10.016.

Brewer, N.T., Morgan, J.C., Baig, S.A., et al., 2016. Public understanding of cigarette smoke constituents: three US surveys. Tob. Control. (doi: tobaccocontrol-2015-052897 [pii]).

Briones, E.M., Benham, G., 2016. An examination of the equivalency of self-report measures obtained from crowdsourced versus undergraduate student samples. Behav. Res. Methods http://dx.doi.org/10.3758/s13428-016-0710-8.

Center for Disease Control and Prevention. National Adult Tobacco Survey: 2012–2013 national adult tobacco survey sample design and methodology summary. https://www.cdc.gov/tobacco/data_statistics/surveys/nats/pdfs/2012-2013-nats-methodology-final.pdf. Updated 2015. Accessed February 2, 2017.

Centers for Disease Control and Prevention. National Adult Tobacco Survey: 2012–2013 national adult tobacco survey weighting specifications. https://www.cdc.gov/tobacco/data_statistics/surveys/nats/pdfs/2012_13_weighting_methodology.pdf. Updated 2014. Accessed February 2, 2017.

Centers for Disease Control and Prevention. National Adult Tobacco Survey (NATS). https://www.cdc.gov/tobacco/data_statistics/surveys/nats/. Updated 2016. Accessed February 2, 2017.

Chandler, J., Shapiro, D., 2016. Conducting clinical research using crowdsourced convenience samples. Annu. Rev. Clin. Psychol. 12:53–81. http://dx.doi.org/10.1146/annurev-clinpsy-021815-093623.

Chandler, J., Mueller, P., Paolacci, G., 2014. Nonnaivete among Amazon Mechanical Turk workers: consequences and solutions for behavioral researchers. Behav. Res. Methods 46 (1):112–130. http://dx.doi.org/10.3758/s13428-013-0365-7.

Chang, L., Krosnick, J.A., 2009. National surveys via RDD telephone interviewing versus the internet: comparing sample representativeness and response quality. Public Opin. Q. 73 (4):641–678. http://dx.doi.org/10.1093/poq/nfp075.

Crump, M.J., McDonnell, J.V., Gureckis, T.M., 2013. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PLoS One 8 (3):e57410. http://dx.doi.org/10.1371/journal.pone.0057410.

Deville, J., Särndal, C., Sautory, O., 1993. Generalized raking procedures in survey sampling. J. Am. Stat. Assoc. 88 (423):1013–1020. http://dx.doi.org/10.1080/01621459.1993.10476369.

Gosling, S.D., Mason, W., 2015. Internet research in psychology. Annu. Rev. Psychol. 66: 877–902. http://dx.doi.org/10.1146/annurev-psych-010814-015321.

Hu, S.S., Neff, L., Agaku, I.T., et al., 2016. Tobacco product use among adults - United States, 2013–2014. MMWR Morb. Mortal. Wkly Rep. 65 (27):685–691. http://dx.doi.org/10.15585/mmwr.mm6527a1.

Husten, C.G., Deyton, L.R., 2013. Understanding the tobacco control act: efforts by the US Food and Drug Administration to make tobacco-related morbidity and mortality part of the USA's past, not its future. Lancet 381 (9877):1570–1580. http://dx.doi.org/10.1016/S0140-6736(13)60735-7.

Kasza, K.A., Ambrose, B.K., Conway, K.P., et al., 2017. Tobacco-product use by adults and youths in the United States in 2013 and 2014. N. Engl. J. Med. 376 (4):342–353. http://dx.doi.org/10.1056/NEJMsa1607538.

Kreuter, F., Presser, S., Tourangeau, R., 2009. Social desirability bias in CATI, IVR, and web surveys: the effects of mode and question sensitivity. Public Opin. Q. 72 (5):847–865. http://dx.doi.org/10.1093/poq/nfn063.

Leas, E.C., Pierce, J.P., Dimofte, C.V., Villasenor, A., Strong, D.R., 2016. US adult smokers' perceptions of Australia's cigarette warning labels: variance by warning content and consistency across socio-demographic sub-segments. Tob. Control. (doi: tobaccocontrol-2016-053006 [pii]).

Levay, K.E., Freese, J., Druckman, J.N., 2016. The demographic and political composition of Mechanical Turk samples. SAGE Open. 6 (1). http://dx.doi.org/10.1177/2158244016636433.

Litman, L., Robinson, J., Abberbock, T., 2016. TurkPrime.com: a versatile crowdsourcing data acquisition platform for the behavioral sciences. Behav. Res. Methods http://dx.doi.org/10.3758/s13428-016-0727-z.

Mason, W., Suri, S., 2012. Conducting behavioral research on Amazon's Mechanical Turk. Behav. Res. Methods 44 (1):1–23. http://dx.doi.org/10.3758/s13428-011-0124-6.

Mays, D., Tercyak, K.P., 2015. Framing indoor tanning warning messages to reduce skin cancer risks among young women: implications for research and policy. Am. J. Public Health:e1–e7 http://dx.doi.org/10.2105/AJPH.2015.302665.

Mays, D., Smith, C., Johnson, A.C., Tercyak, K.P., Niaura, R.S., 2016a. An experimental study of the effects of electronic cigarette warnings on young adult nonsmokers' perceptions and behavioral intentions. Tob. Induc. Dis. 14. http://dx.doi.org/10.1186/s12971-016-0083-x (17-016-0083-x. eCollection 2016).

Mays, D., Tercyak, K.P., Lipkus, I.M., 2016b. The effects of brief waterpipe tobacco use harm and addiction education messages among young adult waterpipe tobacco users. Nicotine Tob. Res. 18 (5):777–784. http://dx.doi.org/10.1093/ntr/ntv223.

National Cancer Institute. Tobacco control research priorities for the next decade: working group recommendations for 2016–2025. https://cancercontrol.cancer.gov/pdf/nci-tobacco-control-research-priorities-rpt-feb-2016.pdf. Updated 2016. Accessed February 2, 2017.

Noar, S.M., Hall, M.G., Francis, D.B., Ribisl, K.M., Pepper, J.K., Brewer, N.T., 2015. Pictorial cigarette pack warnings: a meta-analysis of experimental studies. Tob. Control. http://dx.doi.org/10.1136/tobaccocontrol-2014-051978.

Pearson, J.L., Richardson, A., Feirman, S.P., et al., 2016. American spirit pack descriptors and perceptions of harm: a crowdsourced comparison of modified packs. Nicotine Tob. Res. 18 (8):1749–1756. http://dx.doi.org/10.1093/ntr/ntw097.

Popova, L., Ling, P.M., 2013. Alternative tobacco product use and smoking cessation: a national study. Am. J. Public Health 103 (5):923–930. http://dx.doi.org/10.2105/AJPH.2012.301070.

Syme, M.L., Cohn, T.J., Barnack-Tavlaris, J., 2017. A comparison of actual and perceived sexual risk among older adults. J. Sex Res. 54 (2):149–160. http://dx.doi.org/10.1080/00224499.2015.1124379.

United States Census Bureau. HINC-03. People in households-households, by total money income, age, race and Hispanic origin of householder. http://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hinc/hinc-03.2013.html. Updated 2016. Accessed February 2, 2017.

Weaver, S.R., Majeed, B.A., Pechacek, T.F., Nyman, A.L., Gregory, K.R., Eriksen, M.P., 2016. Use of electronic nicotine delivery systems and other tobacco products among USA adults, 2014: results from a national survey. Int. J. Public Health 61 (2):177–188. http://dx.doi.org/10.1007/s00038-015-0761-0.

Williams, V.F., Smith, A.A., Villanti, A.C., et al., 2016. Validity of a subjective financial situation measure to assess socioeconomic status in US young adults. J. Public Health Manag. Pract. http://dx.doi.org/10.1097/PHH.0000000000000468.

Yeager, D.S., Krosnick, J.A., Chang, L., et al., 2011. Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. Public Opin. Q. 75 (4):709–747. http://dx.doi.org/10.1093/poq/nfr020.