Federal Committee on Statistical Methodology

Hitting the Target?: The Use of Targeted Samples in Probability-based Samples

Randall K. Thomas, Frances M. Barlas, Megan A. Hendrich, & Kathleen Santos

Ipsos Public Affairs

2022 FCSM Research & Policy Conference
October 25th – October 27th, 2022
Washington D.C.

Abstract

Address-based (ABS) studies often attempt to obtain people with specific characteristics (e.g., black people, 18 to 24 year olds) at higher rates for more precision of group estimates. Households in the sample frame can have commercial-based information appended useful for targeted sample selection. Targeted samples can improve study efficiency and lower study costs. Though many researchers treat individuals with the desired characteristics from targeted sample as equivalent to those from non-targeted sample, individuals from targeted samples might differ. In the study we report, we were interested in obtaining higher rates of cigarette smokers. We first selected a non-targeted general population sample before selecting a targeted smoking sample. We compared the smokers obtained for each sample type and found that targeting was indeed associated with higher the prevalence of smokers, but also found differences in smoker characteristics – smokers from targeted households were older, smoked more, started smoking younger, and were less likely to use other tobacco products. We examined how weighting could offset the effect of targeting, but such adjustments can come at a cost in the unequal weighting effect. Our results show caution is recommended when using targeted samples.

Study Background

When using address-based sample (ABS) for a general population study, the cost for just the materials, postage, and incentive costs alone (excluding labor) often amounts to a minimum of $100 per complete.

However, many times we are NOT interested not in a general population, but our interest is more focused on a particular subgroup, and this is often a group that occurs at low rates in the population (low incidence studies).

For low incidence studies, the cost can rise dramatically to $1,000 or much more per complete, depending on the rarity of the participant.

To obtain more qualified completes with specific characteristics for a lower cost, many studies use targeted ABS sample. These targeted samples often have the desired participants at a higher rate than occurs in a non-targeted sample.

Study Background – Two Type of Targeted Sample

Targeted samples are created using information associated with the addresses in the frame – which includes location information (and associated characteristics of the location, like income levels in a zip code, etc.).

In addition, commercial-based information can be appended to the addresses. This information can include such things as magazine subscriptions, coupons used, shopping scanner data, credit use, etc. that are associated with specific addresses.

From this information, inferential associations can be made that increases the certainty that one or more people in the household have specific characteristics (e.g., age, race-ethnicity) or owns a particular product or has specific interests.

Use of targeted samples can significantly decrease the cost of ABS mailings because it can enhance the incidence rate 2 or 3 times (or more) over the rate of its natural occurrence in an unscreened population.

Study Background – Targeted Sample

With cost constraints on most ABS projects, the use of targeted sample has become commonplace, and many researchers treat low-incidence participants in targeted sample as functionally equivalent to low-incidence participants from a general population sample.

However, there are many reasons why individuals from targeted samples might differ from those in non-targeted samples, especially if using ancillary commercial-based data. For example, to be 'targeted', those with a higher certainty of a given characteristic might also be ones who:

- **Used coupons**
- **Resided at the address for a longer period of time**
- **Purchased the products of interest with a greater frequency**
- **Shopped at stores in higher income areas that tracked purchase behaviors.**

Study Purpose

Ipsos had the opportunity to develop a direct comparison of targeted and non-targeted ABS sample that specifically focused on recruiting cigarette smokers for additional studies. Some of these recruited smokers were needed for a variety of studies looking for probability-based participants to participate in smoking cessation programs.

Our primary goal for the study was to compare the nature of the smokers obtained from a general population sample with those obtained from the targeted smoker stratum using ancillary commercial-based data.

Method

Study Design

General Population Recruitment Survey

- **Invited to participate in online questionnaire (a push-to-web design)**
- **English and Spanish versions**
- **Average response rate was 9.3% (AAPOR – RR3).**
- **Fielded: 10 September to 2 October, 2018.**
- **Number of participants who completed the Recruitment Survey: 40,151**

Study Design – Samples Used

Address-based sample (ABS)

1. **First drew a simple random sample from the DSF frame of all housing addresses, no stratification**
2. **Then the unselected addresses in remaining frame were partitioned into 4 strata:**
    1. **At least 1 smoker adult smoker in HH with high likelihood**
    2. **Hispanic 18-29, smoking unavailable**
    3. **Other 18-29 (non-Hispanic/unavailable race-ethnicity, smoking unavailable)**
    4. **Complement (30+ or unavailable age, smoking unavailable)**

Study Design – ABS Mailing & Incentive Protocol

General Population Recruitment Survey

- **Mailing Protocol**
  - **Mailed invitation - $2 affixed to letter, contained online link and unique password**
  - **2 reminder postcards, with link and password**
- **Incentive Protocol:**
  - **$2 in initial invitation**
  - **$10 to complete a short 5 minute survey**
  - **If an a current smoker, HH is deemed eligible, if more than one, randomly selected within the questionnaire – with an incentive of $20 more to complete a 12 minute survey for selected smoker (if in HH)**

Study Design

ABS Strata Used & Response

Sample and Strata Description	Frame	Sample	Complete	Break-off	Survey Yield	Completion Rate
Sample 1 - General Population	124,486,370	58,758	4,172	732	8.3%	85.1%
Sample 2 - Strata						
1 - At least 1 smoker in HH	2,915,161	172,289	17,146	2,468	11.4%	87.4%
2 - Hispanic 18-29	786,667	19,440	1,117	203	6.8%	84.6%
3 - Other 18-29	4,347,181	70,622	5,691	866	9.3%	86.8%
4 - Complement	116,378,603	174,659	12,025	2,039	8.1%	85.5%
Sample 2 Overall	124,427,612	437,010	35,979	5,576	9.5%	86.6%

Survey yield rate is equal to (completes + break-offs)/sample sent.

Study Design – Survey Design

Survey Structure

1. Establish initial respondent – the household informant – is at least 18
2. Questions on age, gender, ZIP, self-reported Health, ever smoke, number of cigarettes smoked in life, current smoker, Hispanic ethnicity, race, education, HH income, number of adults in HH.
3. Census of household adults – age, gender, and whether current smokers
4. Select 1 adult in HH 21+ who smokes cigarettes. If more than 1 adult 21+ smokes, then randomly select 1 of them for smoking survey.
5. Follow up with selected smoker to ask additional nicotine-related questions and demographics.

Results

Results – Unweighted Household Results

Characteristics of the primary household informant

Sample and Strata Description	Female	18 to 29	College +	Hispanic	HH Income - 100K+
Sample 1 - General Population	58.6%	14.1%	42.2%	10.3%	23.9%
Sample 2 - Strata					
1 - At least 1 smoker in HH	60.3%	7.0%	33.9%	6.5%	23.7%
2 - Hispanic 18-29	66.9%	46.7%	24.1%	76.4%	10.9%
3 - Other 18-29	65.0%	34.2%	38.0%	5.0%	21.3%
4 - Complement	59.2%	11.9%	42.1%	9.5%	23.9%
Sample 2 Overall	60.9%	14.4%	37.0%	9.5%	23.0%

Results – Unweighted Household Results

Characteristics of the primary household informant

Sample and Strata Description	% Excellent/VG health	Single Adult in HH	HH % with at least 1 smoker	# Smokers in Smoking HH
Sample 1 - General Population	59.6%	27.5%	23.5%	1.35
Sample 2 - Strata				
1 - At least 1 smoker in HH	53.5%	17.5%	32.0%	1.44
2 - Hispanic 18-29	53.5%	18.3%	25.0%	1.37
3 - Other 18-29	60.5%	24.0%	27.1%	1.39
4 - Complement	58.7%	27.6%	23.5%	1.31
Sample 2 Overall	56.4%	22.0%	28.1%	1.40

Results – Selected Smoker Results - unweighted

Characteristics of the selected smoker in HH

Sample and Strata Description	Total N	Female	21 to 29	College +	Hispanic	HH Income - 100K+
Sample 1 - General Population	794	47.1%	29.3%	15.0%	10.7%	12.7%
Sample 2 - Strata						
1 - At least 1 smoker in HH	4,469	51.0%	20.8%	12.8%	6.3%	12.9%
2 - Hispanic 18-29	234	45.7%	44.5%	13.7%	71.8%	7.7%
3 - Other 18-29	1,336	47.5%	46.4%	11.2%	4.9%	10.8%
4 - Complement	2,389	49.0%	29.4%	15.6%	8.9%	12.3%
Sample 2 Overall	8,328	49.6%	27.8%	13.4%	8.6%	12.3%

Results – Selected Smoker Results - unweighted

Characteristics of the selected smoker in HH

Sample and Strata Description	Married	% Full-time Employed	Not Moved in Past 5 Years	% Excellent/VG health	% Obese
Sample 1 - General Population	42.2%	50.4%	44.9%	40.3%	31.4%
Sample 2 - Strata					
1 - At least 1 smoker in HH	49.6%	48.3%	50.8%	38.2%	34.5%
2 - Hispanic 18-29	42.2%	58.6%	49.1%	34.6%	38.3%
3 - Other 18-29	36.8%	54.9%	47.2%	39.7%	32.2%
4 - Complement	40.8%	52.8%	43.2%	40.7%	30.3%
Sample 2 Overall	44.9%	50.9%	48.1%	39.0%	33.1%

Results – Demo Weighting Targets for Smokers

Both samples (Gen Pop and Stratified) were weighted overall for each sample. Within Sample 2, smokers in each stratum was also weighted.

The targets used were derived from the PATH Wave 4 survey* for smokers:

Region	%
Northeast	*16.2%*
Midwest	*24.7%*
South	*41.5%*
West	*17.6%*

Education	%
Less Than College	*56.3%*
Some College	*31.6%*
College+	*12.1%*

Gender-age	%
Male 21-24	*2.6%*
Male 25-34	*12.5%*
Male 35-44	*10.8%*
Male 45-54	*9.9%*
Male 55+	*17.3%*
Female 21-24	*2.6%*
Female 25-34	*9.3%*
Female 35-44	*8.8%*
Female 45-54	*10.1%*
Female 55+	*16.1%*

Race-ethnicity	%
White NH	*70.9%*
Black NH	*12.4%*
Other NH	*6.6%*
Hispanic	*10.2%*

* Population Assessment of Tobacco and Health (PATH) Study

Smoking Behaviors – weighted to smoker demos

There were some differences between the smokers from the general population sample and those from the targeted smoker stratum.

Sample and Strata Description	Days smoked in past 30	Number of cigarettes smoked per day	Total cigarettes smoked	Every day smoker	Ever tried e-cig	High interest in quitting
Sample 1 - General Population	25.4	13.7	381.9	73.3%	61.1%	41.2%
Sample 2 - Strata						
1 - At least 1 smoker in HH	25.8	15.5	438.5	76.4%	63.7%	38.2%
2 - Hispanic 18-29	25.9	15.2	426.0	78.3%	66.8%	30.3%
3 - Other 18-29	25.3	14.1	394.8	71.8%	65.7%	37.4%
4 - Complement	25.5	14.3	400.9	73.6%	61.2%	40.5%
Sample 2 Overall	25.6	14.7	413.2	74.7%	63.0%	39.0%

Discussion

Conclusions and Discussion

Is someone from a targeted sample for a specific characteristic different from a person with that same characteristic who comes from a general population sample?

Yes, bigger differences were found when not weighted. However, modest differences in smoking behavior were still found once using common cigarette for smoker demographic targets.

Those smokers from targeted smoker stratum:

- **Did smoke somewhat more cigarettes**
- **Were somewhat more likely to be an every-day smoker**
- **Were somewhat less interested in quitting than those from the general population sample.**

Conclusions and Discussion

The use of other strata (18-29 Hispanic, 18-29 Other, Complement) to form a broader and diverse sample appeared to counteract some of the biasing effects of the targeted stratum - smokers from Sample 1 and the overall Sample 2 were much more similar in their smoking behaviors.

Use of targeted ABS sample can significantly increase the efficiency of recruitment, making the study less costly.

However, when considering the use of targeted sample, we would recommend carefully considering how the targeting information was identified/obtained and how that could correlate with the key variables of interest to the study.

Future Directions

We are currently studying how we could use location information (e.g., Census Block Groups that have higher concentrations of low income, or Hispanic residents, etc.) rather than ancillary information that has been appended to the addresses.

Our initial studies have indicated that, when possible, using location-based information for sample selection may be associated with less bias for samples of interest than using commercial-based ancillary data to select sample.

Thank you!

Randall K. Thomas

Randall.Thomas@ipsos.com