ASSESSING THE EFFECTIVENESS OF USING PROPENSITY SCORES TO TARGET HARD-TO-REACH POPULATIONS

MARKET DECISIONS RESEARCH

Better Data
Better Insights
Better Outcomes

AAPOR 2023

Brian Robertson, PhD, VP of Research
John Charles, MS, Senior Analyst
Anna Driscoll, Research Analyst

Acknowledgement

Market Decisions Research would like to thank the Vermont Department of Health for allowing us to use their survey data in this presentation.

VERMONT
Department of Health

Our Presentation

- Project Background
- The Challenge
- Our Methodology and Approach
- Propensity Scoring
- Adventures in Data Collection
- Analysis and Results
- Conclusions and Future Research

Project Background

The Vermont Adult Tobacco Survey

The Vermont Adult Tobacco Survey is a survey of Vermont adults.

Data are used to evaluate the effectiveness of Vermont Tobacco Control Program efforts to reduce smoking and increase awareness and knowledge of smoking-related issues.

Summary

Vermont has conducted the Adult Tobacco Survey annually from 2001 to 2008, every other year from 2010 to 2016, and in 2022.

The sample includes 1,600 to 2,000 respondents each year. Half are Vermonters who smoke, and half are Vermonters who do not smoke.

The Challenge

A primary focus is to gather data from those using or who have recently used tobacco products: How do you find and survey current smokers and e-cigarette user and recent quitters?

First, these groups represent a small percentage of the population.

Group	% of Vermont Residents (2020 BRFSS)
Current Smokers	12.6%
Current E-Cigarette Users	3.7%
Recent Quitters (Quit <5 years ago)	7.2%

Second, members of these groups are generally less likely to respond to surveys on tobacco use, especially one administered by **a Department of Health**

The Challenge

- In 2022, the goal was to complete 50% of surveys among current smokers and e-cigarette users and recent quitters (current users and recent quitters)
  - 800 of the 1,600 total surveys
- Prior to 2022 the survey was administered by telephone (RDD)
- **It simply became cost prohibitive to conduct the survey by telephone**

Methodology and Approach

Our New Approach — The Sample

Data collection will combine probability and non-probability samples

- A random address-based sample (ABS) stratified into four regions by county
- ABS supplemented with a targeted oversample where one or more persons in the household was identified as a likely smoker
- The sample was provided by Marketing Systems Group
- Both the random ABS and likely smoker targeted sample included appended individual and household level demographics…**if available**

Methodology and Approach

Our New Approach — Data Collection

- Mailing of push-to-web survey invitation letter (random ABS) with reminder
- Mailing of survey booklet targeting areas with low access to high-speed internet (Random ABS) with reminder

Review response!

- Draw sample from targeted smoking sample file
  - **Identified as having a higher probability of being a current tobacco user based on propensity score analysis**
  - Push to web invitation with reminder card.
- Final push-to-web survey invitation letter if needed to achieve goal of 1,600 total surveys
- $10 incentive to all respondents completing the survey

Propensity Scoring

- Propensity scoring is a statistical technique used in observational studies to balance the distribution of covariates between treatment and control groups
- The goal of propensity scoring is to match everyone in the treatment group with one or more individuals in the control group who have similar propensity scores
- The propensity score is typically estimated using logistic regression
- **Propensity scoring can be used for sampling when the goal is to select a representative sample of individuals who are like a population in terms of belonging to a group or exhibiting a behavior**
  - **For example: People who smoke or use e-cigarettes**

Propensity Scoring

Calculating Propensity Score

2020 Vermont Behavioral Risk Factor Surveillance Survey (BRFSS) data was used to estimate the propensity of being a tobacco user based on geodemographics

- Age
- Gender
- Income
- Education Level
- Rurality
- Housing Status
- Marital Status
- Number of Children in Household

Variable	Category
Gender	Female/Male
Age Categories	18 to 34 35 to 44 45 to 54 55 to 64 65 or older
Income	Less than $10K $10K-$15K $15K-$20K $20K-$25K $25K-$35K $35K-$50K $50K-$75K $75K+
Education Level	Less than high school High school diploma Some college College degree
Rurality	Urban/Rural
Housing Status	Renter/Owner
Marital Status	Not married/Married
Number of Children in Household	None 1 child 2 children 3 or more children

Propensity Scoring

Calculating Propensity Score

- The regression model computed a propensity score (predicted probability) of being a tobacco user
- Higher scores were associated with a higher likelihood that person was a current tobacco user
- Once calculated from the BRFSS, the scores were appended to records in the sample file for the Vermont ATS by matching geodemographics

Propensity Score Table

Gender	Age Categories	Income	Education Level	Rurality	Housing Status	Marital Status	Number of children in household	Predicted Probability of Smoking
Male	35 to 44	Less than $10K	Less than high school	Urban	Renter	Not married	2 children	**0.689**
Female	18 to 34	$20K-$25K	Less than high school	Rural	Renter	Not married	1 child	**0.566**
Male	35 to 44	$50K-$75K	High school diploma	Rural	Renter	Not married	2 children	**0.482**
Male	35 to 44	$75K	High school diploma	Rural	Owner	Not married	1 child	**0.315**
Male	45 to 54	$35K-$50K	Less than high school	Urban	Owner	Married	2 children	**0.315**
Male	45 to 54	$35K-$50K	Some college	Rural	Renter	Married	None	**0.259**
Female	18 to 34	$50K-$75K	Some college	Rural	Renter	Not married	1 child	**0.259**
Female	65 or older	$75K	College degree	Urban	Owner	Married	2 children	**0.013**
Male	65 or older	$75K	College degree	Urban	Owner	Married	None	**0.012**
Female	65 or older	$75K	College degree	Urban	Owner	Married	None	**0.011**

Adventures in Data Collection

Things went so right and so wrong!

- The survey response was much higher than expected
  - 2,255 completed surveys through the initial push to web and mail surveys
  - This exceeded our overall target of 1,600 surveys
- While response was higher than expected, the number of current users and recent quitters was less than half we expected
  - Only 182
  - Given the overall response we would have expected about 450

So, our data collection strategy went out the window!

Adventures in Data Collection

All subsequent data collection efforts screened for current users and recent quitters

- All others were screened out

Data collection became the ultimate kitchen sink approach

- Mailings to any sample record with a high propensity score
- Working with Vermont Tobacco Control Network to reach out to identify smokers to complete the survey
- Taking out ads
- Contacting FQHCs and putting up posters
- Increasing incentives to $25
- Using online panels

After all this we finally ended up with 696 current smokers/e-cigarette users and recent quitters

Analysis and Results

But could we learn *anything?*

We lost the ability to fully evaluate the effectiveness of propensity scores as a tool to target current users and recent quitters:

- At each stage of data collection, and
- To use information from earlier stages to refine our targeting in later stages

But because we appended propensity scores to the ABS probability sample, we could at least evaluate the effectiveness of propensity scores during the initial push-to-web and mail surveys

We hope we learn something that will help us to adjust methods in the future!

Analysis and Results

How well do propensity scores work in identifying smokers/e-cigarette users and recent quitters?

Again, looking only at the surveys completed during the initial push to web and mail surveys:

- We ran a simple crosstabulation of completed surveys
  - Smoking/e-cigarette use status by propensity score category

And it did show that using propensity scores could be an effective method to identify current tobacco users and recent quitters

Analysis and Results

Summary of Results Broken out by Propensity Score

Smoking and E-Cigarette Use Status					
Propensity Score	Never Used	Formerly Used	Recently Quit (<5 yrs. ago)	Currently Use	% Current Users and Recent Quitters
Less than .1	**63%**	**30%**	**4%**	**3%**	**7%**
.1-.19	**60%**	**30%**	**4%**	**6%**	**10%**
.2-.29	**49%**	**35%**	**10%**	**7%**	**16%**
.3 or greater	**58%**	**23%**	**4%**	**15%**	**19%**
Total	**60%**	**30%**	**4%**	**5%**	**9%**

The percentage of current smokers/e-cigarette users increased with increasing propensity scores

Conclusions and Future Research

- Those with higher propensity score were more likely:
  - Current smokers/e-cigarette users, or
  - Recent quitters (quit with the last 4 years)

But our work had serious limitations

- We were only able to use a limited set of geodemographic variables to score the sample
  - Only 30% of the sample records had all geodemographic variables
    - The rest were scored based upon the variables that were available
  - Only 6% of probability sample records were scored as 0.2 or higher
- Many known correlates with tobacco use were not available

Conclusions and Future Research

How can we improve the process of identifying and surveying current smokers/e-cigarette users and recent quitters?

- **Identify a sample source with complete (and additional) information for propensity scoring**
  - More complete demographic information for all sample records
  - Additional correlates with tobacco use
    - Easier: Veteran's status, insurance coverage
    - Harder: Physical and mental health status, sexual orientation, substance use

Conclusions and Future Research

- **Swap the order of data collection**
  - First focus on current tobacco users and recent quitters
  - Utilize the targeted sample of likely smokers and records, especially those with high propensity scores
  - Also, include any records from the probability sample with high propensity scores
- Then use the probability sample to obtain the necessary responses for a data set "representative of the population"
  - To which the non-probability sample responses can then be added and weighted appropriately

For More Information

Brian Robertson, Ph.D.	brianr@marketdecisions.com
John Charles, MS	jcharles@marketdecisions.com
Anna Driscoll	adriscoll@marketdecisions.com

Thank you!