

How Study Design can Influence Substantive Survey Results

Dhuly Chowdhury, Karol Krotki, Lily Trofimovich
RTI International

Abstract

Study design plays a vital role in the analysis results. Effective study design can avoid bias introduced by the experimenter and is the first step to yield unbiased estimates. In this paper, we present how age and gender distribution of the youth inspectors can influence the estimated retailer violation rate in Synar survey. Weighted survey logistic model in SAS showed that age, gender, and interaction between age and gender of the youth inspectors has a significant effect on the estimated retailer violation rate.

KEY WORDS: measurement bias, field procedures

1. Introduction

This paper presents how study design can influence substantive survey result using Synar survey data. The objective of the Synar survey is to determine the retailer violation rate (RVR) for each state based on annual, random, unannounced inspections of a probability-based sample of tobacco outlets accessible to youth. The RVR is also referred as retailer non-compliance rate or buy rate. The buy rate is defined as the proportion of tobacco outlets that sell cigarettes to minors. Since 1999, States receiving SAMHSA Block Grant funds need to demonstrate an 80% or better compliance rate with tobacco youth access laws.

To conduct the Synar survey, each state hires youth inspectors who attempts to buy cigarettes from randomly selected outlets. Some researchers claim that states can change the buy rate by manipulating the age and gender distribution of the youth inspectors. According to CDC, smoking rate among girls is higher than the boys. In middle school 8% of boys smoke cigarettes compared with 9% of girls. In high school the smoking rate increases to 22.1% for boys and 22.4% for girls². It is assumed that older female inspectors tend to increase the buy rate and younger male inspectors tend to depress it and thus changing the gender and age distribution of the inspectors can play a vital role in the validity of the estimated buy rate. This paper describes the statistical data analysis performed to explore whether and to what extent the age and gender distribution of the youth inspectors impacts the estimated retailer violation rate.

Our data analysis focused on identifying and measuring the effect of inspectors' gender and age on the buy rate. Different models were fitted to identify influential variables that affect the buy rate. Independent variables such as state, age of inspectors, gender of inspectors, presence of law enforcement during inspection etc. were considered in the model. The interactions between the age and gender variables as well as the influence of covariates such as state and presence of law enforcement during inspection were also considered in the analysis. Based on the findings, this paper includes procedures for the control of bias introduced by the study design.

1.1 Synar Amendment and Regulation

In July 1992 Congress enacted the Synar Amendment aimed at decreasing minors' access to tobacco products. It was sponsored by late Representative (D) of Oklahoma Mike Synar (1950-1996). Synar amendment requires states to pass and enforce laws that prohibit the sales of tobacco to individuals 18 years of age or younger.

In order to be eligible for SAPT Block Grant funds, all States must have a law prohibiting the sale or distribution of tobacco products to minors. The Secretary is required by statute to withhold funds from States that have not enacted the required prohibitions and to decrease the annual Substance Abuse Prevention and Treatment (SAPT) Block Grant award for States that do not comply with the enforcement and reporting requirements.⁵

To determine compliance, the Amendment requires each State to enact and enforce State youth tobacco access control law prohibiting any manufacturer, retailer or distributor of tobacco products from selling or distributing such products to any individual under the age of 18. In addition to enacting and enforcing a law, states must implement annual random, unannounced compliance inspections (RUIs) to determine their buy rates of tobacco products sold to youth under the age of 18; and report RVR to the HHS Secretary. If a state's buy rate exceeds 20%, it may lose 40% of its Federal Substance Abuse Prevention & Treatment (SAPT) Block Grant funds. These funds are used to support alcohol, drug and mental health treatment programs and services.

1.2 The Synar Survey

The Synar survey is an annual, statewide undercover youth tobacco survey conducted to ascertain the State's compliance with laws that prohibit sales of tobacco products to underage. The sample of outlets inspected must be representative of the geographic distribution of tobacco outlets in the State. States that do not choose to participate in this survey lose their eligibility for the SAMHSA Block Grant. It is also important to conduct this survey and conduct tobacco enforcement as a means to decrease youth access to tobacco. The Retailer Violation Rate (RVR) obtained over a number of years will be used to assess the State's progress towards achieving the overall Synar goal of 80% or more compliance rate.

An over-the-counter or vending machine is eligible for the Synar Survey if it sells tobacco products and is accessible to youth under age 18. Ineligible tobacco outlets are bars, taverns, or other adult-only clubs that have enforced minimum age restriction for entry of 18 or older.

To conduct the survey, guidance currently provided to states about how they should hire inspectors is:

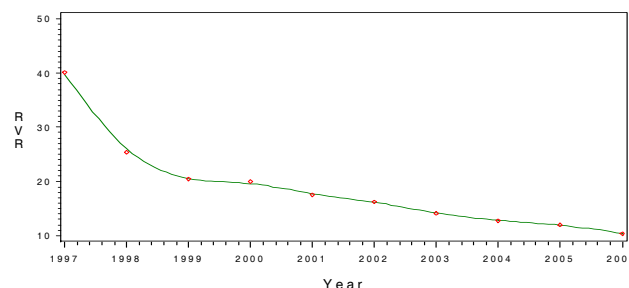
1. Inspectors should be 15 or 16 years of age and should look their age.
2. An equal number of male and female youth should be selected.
3. Racial or ethnic characteristics of inspectors should be approximately the same as those in the region where the inspection will be carried out.
4. Inspectors should dress as they would normally.

However, every state faces unique challenges to conduct the survey and implementing the proposed design may not be practical in most states. As a result, age and gender distribution of youth inspectors are skewed in many states. This raises the question whether estimated RVR depends on age and gender of the inspectors.

2. Analysis

National weighted average RVR was around 40% in 1997 when Synar survey started; over the past 10 years RVR went down rapidly. Some researchers question how reliable this estimated retailer violation rate. This paper aims to answer the research question: how age and gender distribution of youth inspectors can influence substantive survey result?

Figure 1: National Weighted* Average Retailer Violation Rate, FY 97–06.



2.1 Hypothesis

The following primary hypotheses were tested –

Hypothesis 1: Female inspectors are more likely to successfully buy cigarettes than male inspectors.

Hypothesis 2: Older inspectors are more likely to buy cigarettes than younger inspectors.

Hypothesis 3: There is interaction effect between age and gender of the youth inspectors.

2.2 Data Compiling

Synar survey data from 42 states in fiscal year 2007 were considered in this analysis. Synar Survey Estimation System (SSES) is an optional software tool developed by the Center for Substance Abuse Prevention (CSAP) to assist states in estimating and reporting of their annual Synar survey results. Raw data was available in excel format for all the states that used SSES. Out of 59 states and US territories 48 used SSES and provided raw data. Data from 42 states and territories was available in usable form and was included in this analysis. The following 11 state and territories was excluded from analysis as they did not use SSES in 2007: American Samoa (AS), Federated States of Micronesia (FM), Gum (GU), Iowa (IA), Marshal Islands (MH), Pennsylvania (PA), Puerto Rico (PR), Republic of Palau (PW), Utah (UT), Vermont (VT), and Virgin Islands (VI). The master dataset for analysis was at the outlet level and contained the following variables: state, region, census division, fiscal year, presence of law enforcement officer required by law, outlet ID, sampling weights, violation flag, outlet type, youth inspector ID, youth inspector Gender, youth inspector age.

2.3 Sampling Technique and Weights

The survey weights were calculated by taking the inverse of the selection probability. Eleven states used cluster sampling and weights for these states were available in the SSES input data. All other states used stratified sampling. For these states weights were calculated by taking proportion of estimated outlet population size and number of eligible outlets in the sample for each stratum.

2.4 Weighted Survey Logistic Model

We analyzed the data using weighted survey logistic regression model in SAS 9.1.3. Each state was considered as a stratum. The response variable violation flag was a dichotomous variable (1 and 0) where 1 referred to retailer violation. The model considered the following predictor variables: gender, age, region, outlet type, and presence of law enforcement. We also included the interaction between age and gender of the inspector.

$$\text{Logit (Violation)} = \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_3) + \beta_4(x_4) + \beta_5(x_5) + \beta_6(x_6) + e$$

Where:

- x_1 = Age
- x_2 = Gender
- x_3 = Interaction between Age and Gender
- x_4 = Region
- x_5 = Outlet Type (OTC or VM)
- x_6 = Presence of Law Enforcement Officer
- e = Random Error

The fitted survey logistic model showed that all the independent variables and interaction between age and gender were significant at 5% level of significance. Since the interaction between age and gender was significant, we tested the difference between female and male buy rate at different age group.

Figure 2: Contrast Rows Estimation and Testing Results difference between Female vs. Male

Contrast	Estimate	Standard Error	95% Confidence Limits	
F vs. M at 14	-0.43	0.08	-0.59	-0.27
F vs. M at 15	-0.23	0.05	-0.33	-0.14
F vs. M at 16	-0.04	0.03	-0.09	0.02
F vs. M at 17	0.16	0.04	0.08	0.24
F vs. M at 18	0.36	0.07	0.21	0.50

All age groups showed significant difference between males and females except age group 16. Figure 2 shows the estimates and testing results for difference between female and male at different age.

The direction of the difference is opposite for age group below and above 16. Estimated buy rate among females are significantly less than that of male for age group 14 and 15. Whereas, estimated female buy rate is significantly higher than male buy rate for age group 17 and 18. Buy rates are not statistically significantly different between male and female for age group 16.

The table below (Figure 3) shows the average RVR by region; Northeast region had highest average RVR and South region had lowest RVR.

Figure 3: Average RVR by Region

Region	Average RVR
Northeast	16.30
West	11.82
Midwest	11.53
South	9.18

The overall “Region” effect was significant (p-value = 0.0027). Only “Northeast” region was found marginally statistically significantly different than “South” region with a p-value of 0.05. The odds ratio indicated that the “Northeast” region is 1.3 times more likely than the “South” region. The “South” region was considered as the reference group. Others regions were not significantly different than the reference region. In fiscal year 2007, RVR for Massachusetts was 22.7. This was higher than the acceptable RVR of 20%. As a result average RVR for “Northeast” region was higher than other regions.

2.5 Analysis by State

Study design is different for each state to address both the unique circumstances of the state and requirement of the Synar Regulation. To show illustrative examples of what happens at the state level we analyzed few states independently. We analyzed Kansas, Tennessee, Idaho, Indiana, and Maryland.

In many states number of attempted buy are unequally distributed over age and gender of the youth inspectors. Same inspector may inspect a large number of outlets if assigned to a more populated area; whereas inspectors assigned to rural areas may inspect only a few outlets. Figure 4 shows the distribution of age and gender of

youth inspectors and attempted buy for the 5 states that we analyzed independently.

Figure 4: Age, Gender, and Attempted Buy by State

Age	Gender	Attempted Buy
Kansas		
15	F	69
15	M	61
16	F	149
16	M	59
17	F	80
17	M	29
Tennessee		
15	F	151
15	M	168
16	F	140
16	M	324
17	F	162
17	M	280
Idaho		
15	F	18
15	M	27
16	F	259
16	M	274
17	F	66
17	M	99
Indiana		
15	F	294
15	M	203
16	F	1123
16	M	550
17	F	485
17	M	811
Maryland		
16	F	100
16	M	355
17	F	234
17	M	82

Weighted survey logistic result for different states was different for the states we analyzed independently. The analysis result for Kansas was same as the overall

analysis. Kansas showed significant interaction between age and gender of inspectors at 5% level of significance. Tennessee showed both age and gender as significant factor but not the interaction between them. Only gender effect was significant for Idaho; and only age effect was significant for Indiana and Maryland. Figure 5 shows the weighted RVR and whether age, gender and their interaction were significant for the five states that was analyzed independently.

Figure 5: Average RVR and Significant Variables by State

State	RVR	Significant at 5%		
		Gender	Age	Age* Gender
Kansas	19.9			Yes
Tennessee	14.8	Yes	Yes	
Idaho	13.2	Yes		
Indiana	11.3		Yes	
Maryland	9.5		Yes	

2.5.1 Kansas

The interaction between age and gender of inspectors was significant in Kansas. Kansas showed that age and gender was significantly different for age group 14, 15, and 17. For age group 16 there was no difference between female and male. The direction of the difference is different for age group above and below 16. For age group 14 and 15 male are more likely to successfully buy cigarettes. And for age group 17 females are more likely to make a purchase based on the data from Kansas in federal fiscal year 2007. Analysis results for Kansas are summarized in table below.

Figure 6: Contrast Estimation and Testing Results difference between Female vs. Male for Kansas

Contrast	Estimate	Standard Error	95% Confidence Limits	
Kansas				
F vs. M at age 14	-1.23	0.38	-1.98	-0.48
F vs. M at age 15	-0.68	0.23	-1.14	-0.23
F vs. M at age 16	-0.13	0.14	-0.40	0.14
F vs. M at age 17	0.42	0.20	0.02	0.82

2.5.2 Tennessee, Idaho, Indiana, and Maryland

Analysis results for Tennessee, Idaho, Indiana, and Maryland are summarized in Figure 7.

Figure 7: Wald Confidence Interval for Adjusted Odds Ratios (Idaho, Indiana, and Maryland)

Contrast	Estimate	95% Confidence Limits	
Tennessee			
Gender F vs. M	2.0	1.24	3.227
Age	1.8	1.21	2.53
Idaho			
Gender F vs. M	2.5	1.58	3.97
Age	1.2	0.73	1.85
Indiana			
Gender F vs. M	1.4	1.17	1.7
Age	1.2	1.08	1.4
Maryland			
Gender F vs. M	0.8	0.41	1.69
Age	6.6	2.99	14.63

Both age and gender was significant for Tennessee at 5% level of significance. The odds ratio indicates that female inspectors are 2.0 times more likely to successfully buy cigarettes than male inspectors. And older inspectors are 1.8 times more likely to buy cigarette.

Idaho State used a single-stage cluster sample in fiscal year 2007 and weighted RVR was 13.2%. The buy rate for females was much higher than the buy rate for the males. According to the annual Synar report, RVR was 8.7% for the males and 18.7% for the females. The survey logistic model showed that only gender effect is significant at 5% level of significance. And the estimated odds ratio indicates that female inspectors are 2.5 times more likely to successfully buy cigarettes than male inspectors.

Indiana State used a census in fiscal year 2007 and RVR was 11.3%. The survey logistic model showed that age and gender of the youth inspectors were significant at 5% level of significance. The estimated odds ratio indicated that female inspectors are 1.4 times more likely to successfully buy cigarette than the male inspectors. Older inspectors are 1.2 times more likely to successfully buy cigarette.

A stratified simple random sample was used in Maryland and average RVR was 9.5%. Annual Synar Report (ASR) shows that male RVR was 6.8% and female RVR was 12.8%. Only gender effect was significant for Maryland and the odds ratio shows that female inspectors are 6.6 times more likely to buy cigarette than male inspectors.

3.1 Limitations

The biggest challenge was to get the data ready in a useable form. Raw data from all the states were not available so we considered only 42 states and territories. Also, there were a lot of missing for both response and predictor variables. We assumed that the missing values were random and no imputation was done to impute the missing. All records with missing were ignored in the survey logistic model.

Race and ethnicity of the youth inspectors may have an important role on the RVR. According to CDC, smoking rate differs by race. Twenty percent (20%) of White women, 17% of African American women, 11% of Hispanic women, 6% of Asian American women and 27% American Indian/Alaska Native women are smokers.³ However, race was not included in the study as we did not have available data.

3.2 Conclusion

The overall data analysis for all the states showed significant interaction between age and gender at 5% level of significance. Age and/or gender of the inspector have significant effect on RVR for most of the states we analyzed independently.

3.3 Recommendations

Equal distribution of age and gender will standardize the bias introduced by the study design. We strongly recommend that states should not change age and gender distribution from year to year. Most importantly, we recommend that attempted buy should be equally distributed among all inspectors to minimize the personal bias introduced by the inspectors.

Acknowledgements

The authors would like to thank Grant Hills from CSAP/SAMHSA and Jennifer Wagner, and Patty Martin from JBS International for their support and collaboration on this project. RTI International funded this research.

References

1. http://www.americanlegacy.org/Files/Women_and_Smoking_Fact_Sheet.pdf

2. CDC. Tobacco Use, Access, and Exposure to Tobacco in Media Among Middle and High School Students- - United States, 2004, MMWR 2005; 54(12): 297-301.
3. CDC. Cigarette Smoking Among Adults— United States, 2005. MMWR 2006; 55(42): 1145-1148
4. <http://www.girlpower.gov/press/research/synar1.htm>
5. <http://prevention.samhsa.gov/tobacco/fctsheet.aspx>
6. <http://prevention.samhsa.gov/tobacco/require.aspx>
7. <http://www.prevnet.org/synar/synar.asp>
8. <http://www.hss.state.ak.us/press/pdfs/pr110403synarfactsheet.pdf>
9. U.S. Department of Health and Human Services. Substance Abuse & Mental Health Services Administration. Center for Substance Abuse and Prevention. *Guide for A Synar Sampling Frame Coverage Study*. Available: <http://download.ncadi.samhsa.gov/csap/CSAPCoverageStudyGuide-2006.pdf> [January 2006]