



HHS Public Access

Author manuscript

Commun Stat Simul Comput. Author manuscript; available in PMC 2021 March 12.

Published in final edited form as:

Commun Stat Simul Comput. 2021 ; 50(3): 822–831. doi:10.1080/03610918.2019.1568472.

Design-based single-mediator approach for complex survey data

Thanh Pham, Trung Ha, Julia N. Soulakova

Burnett School of Biomedical Sciences, College of Medicine, University of Central Florida,
Orlando, FL, USA

Abstract

We discuss a two-step approach to test for a mediated effect using data gathered via complex sampling. The approach incorporates design-based multiple linear regressions and a generalized Sobel's method to test for significance of a mediated effect. We illustrate the applications to a study of nicotine dependence, race/ethnicity and cigarette purchase price among daily smokers in the U.S. The study goal was to assess significance of cigarette purchase price as a mediator in the association between race/ethnicity (non-Hispanic Black/African American, non-Hispanic White) and nicotine dependence measured in terms of the average number of cigarettes smoked per day. The single-mediator model incorporated 18 covariates as control factors. The results indicated a significant mediated effect of cigarette purchase price on the association. However, the relative effect size of 5% indicated low practical significance of the cigarette purchase price as a mediator in the association between race/ethnicity and nicotine dependence. The approach can be modified to studies where data are gathered via other types of complex sampling.

Keywords

Balanced repeated replications; Multi-stage sampling; Variance estimation in complex surveys

1. Introduction

Many national databases of health outcomes are publicly available for secondary data analysis. For example, data from the Current Population Survey (CPS) and CPS Supplements are commonly used to obtain information on labor force, income, and education in the US (Black, Sanders, and Taylor 2003; Burkhauser, Feng, Jenkins, and Larrimore 2011; U.S. Department of Commerce, Census Bureau 2016). The CPS Tobacco Use Supplement (TUS-CPS) is a survey of use of cigarettes and other tobacco in the U.S. and is administered approximately every 3–4 years.

Because the TUS-CPS utilizes complex sampling, researchers should follow the methodological guidelines when analyzing the data (U.S. Department of Commerce, Census

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

CONTACT Julia N. Soulakova julia.soulakova@ucf.edu Burnett School of Biomedical Sciences, College of Medicine, University of Central Florida, 6900 Lake Nona Blvd, Orlando, FL 32827, USA.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lssp.

Bureau 2016). Specifically, all point estimates should be based on the main weight and variance of the estimates should be computed via the balanced repeated replications (BRR) using replicate weights. The weights are computed and posted with the corresponding data files online; these files are hosted by the U.S. Census Bureau (U.S. Department of Commerce, Census Bureau 2016).

Mediation analysis, commonly utilized in social sciences, allows scientists to test if one variable has an effect on another variable through the third variable (Baron and Kenny 1986; MacKinnon 2008). The traditional mediation analysis was proposed for a simple random sample and is not appropriate for analysis of the TUS-CPS and other complex surveys. We propose a generalization of the mediation methodology that can be used for assessing significance of the mediated effect using the TUS-CPS measures. The remainder of the paper is outlined as follows. In Sec. 2, we review the single-mediator model for a simple random sample. In Sec. 3, we describe the procedure for complex sampling to test for the mediated effect; we use the TUS-CPS data as an example. In Sec. 4, we illustrate the application of the procedure to a nicotine dependence study. We conclude with a discussion presented in Sec. 5.

2. Single-mediator model

Consider a binary or continuous independent variable X , a continuous dependent variable Y , a continuous mediator M , and I binary or continuous covariates Z_i , $i = 1, 2, \dots, I$. Suppose we have a simple random sample of K individuals. Then the single-mediator model (Baron and Kenny 1986; Sobel 1982) can be expressed as follows:

$$\begin{aligned} M_k &= \alpha_1 + \beta_1 X_k + \delta_{11} Z_{1k} + \dots + \delta_{1I} Z_{Ik} + \epsilon_{1k} \\ Y_k | m_k &= \alpha_2 + \beta_2 X_k + \gamma m_k + \delta_{21} Z_{1k} + \dots + \delta_{2I} Z_{Ik} + \epsilon_{2k} \end{aligned} \quad (1)$$

where $k = 1, 2, \dots, K$; α_j ($j = 1, 2$) denotes the regression intercept; β_j , γ , and δ_{ji} ($i = 1, 2, \dots, I$; $j = 1, 2$) represent the regression slopes; ϵ_{1k} and ϵ_{2k} ($k = 1, \dots, K$) are the residuals that are independent, $\epsilon_{jk} \sim N(0, \sigma_j^2)$, where σ_j^2 denotes unknown (constant) variance ($j = 1, 2$).

In the single-mediator model (1), β_1 represents the *effect* of X on M , β_2 represents the *direct effect* of X on Y , γ represents the *effect* of M on Y , and $\beta_1 \cdot \gamma$ represents the *mediated (indirect) effect* of X on Y . The *total effect* of X on Y is represented by the sum of the *direct effect* (β_2) and *mediated effect* ($\beta_1 \cdot \gamma$). Figure 1 illustrates the model (1).

To assess significance of the mediated effect we can use the “product of coefficients” approach (MacKinnon, Lockwood, and Williams 2004). Specifically, the null hypothesis $H_0: \beta_1 \cdot \gamma = 0$ is tested against the alternative hypothesis $H_a: \beta_1 \cdot \gamma \neq 0$ via Sobel’s test (Sobel 1982) based on the test statistic

$$Z = \frac{\hat{\beta}_1 \hat{\gamma}}{SE(\hat{\beta}_1 \hat{\gamma})}, \quad (2)$$

where $\hat{\beta}_1$ and $\hat{\gamma}$ denote the least squares estimates for β_1 and γ , respectively, and the standard error (SE) is given via

$$SE(\hat{\beta}_1\hat{\gamma}) = \sqrt{\hat{\beta}_1^2 SE^2(\hat{\gamma}) + \hat{\gamma}^2 SE^2(\hat{\beta}_1)}. \quad (3)$$

The test rejects H_0 in favor of H_a at significance level α if $Z > z_{\alpha/2}$ or $Z < -z_{\alpha/2}$, where $z_{\alpha/2}$ is such that $P(Z_0 > z_{\alpha/2}) = \frac{\alpha}{2}$ for $Z_0 \sim \mathcal{N}(0, 1)$.

If the estimated indirect effect $\hat{\beta}_1\hat{\gamma}$ and direct effect $\hat{\beta}_2$ are both positive or negative, one can assess the magnitude of the mediated effect using the relative effect size (MacKinnon 2008; Preacher and Kelley 2011):

$$\frac{\hat{\beta}_1\hat{\gamma}}{\hat{\beta}_2 + \hat{\beta}_1\hat{\gamma}}. \quad (4)$$

This descriptive measure represents the practical importance of the mediated effect. Because $\hat{\beta}_1\hat{\gamma}$ estimates the indirect effect and $\hat{\beta}_2 + \hat{\beta}_1\hat{\gamma}$ estimates the total effect, the relative effect size can be interpreted as the proportion (or percentage) of the effect of the independent variable on the dependent variable explained by the mediator (MacKinnon 2008). This is why the relative effect size is also termed the proportion mediated effect (MacKinnon 2008; Preacher and Kelley 2011). However, the relative effect size given in (4) should not be used if the estimated effects, $\hat{\beta}_1\hat{\gamma}$ and $\hat{\beta}_2$, have opposite signs (MacKinnon 2008, 83). In the latter case, analogs of this measure should be used, e.g., the estimated coefficients in (4) are replaced by their absolute values (Alwin and Hauser 1975; MacKinnon 2008).

3. Single-mediator analysis of complex survey data

To incorporate correct adjustments for the survey design used to gather the TUS-CPS data, we propose the following two-step procedure.

In the **first step**, we fit the design-based regression models given in (1) using the survey data. These design-based models should incorporate proper adjustments for the specific design characteristics. Specifically, when analyzing the 2010-11 and 2014-15 TUS-CPS data, we need to use the BRR method with 160 replicate weights to compute the standard errors of estimated model coefficients (U.S. Department of Commerce, Census Bureau 2016; Wolter 2007) as follows.

Suppose θ denotes the parameter of interest, $\hat{\theta}$ is the estimator of θ based on the main weight, and $\hat{\theta}_r$ ($r = 1, \dots, 160$) is the estimator of θ based on the r -th replicate weight.

Then the BRR approach computes the standard error of $\hat{\theta}$ via:

$$SE_{BRR}(\hat{\theta}) = \sqrt{\widehat{Var}_{BRR}(\hat{\theta})}, \text{ where } \widehat{Var}_{BRR}(\hat{\theta}) = \frac{1}{40} \sum_{r=1}^{160} (\hat{\theta}_r - \hat{\theta})^2.$$

The main weight and replicate weights can be used directly in the SAS SURVEYREG procedure in the SAS® 9.4 Survey Package (SAS Institute Inc. 2013) when fitting the model. Upon completing this step, we have the estimated values of the design-based regression coefficients, $\hat{\beta}_1$, $\hat{\gamma}$, and $\hat{\beta}_2$, as well as the standard errors $SE(\hat{\beta}_1)$ and $SE(\hat{\gamma})$.

In the **second step**, we compute the generalized Sobel's test statistic using the estimates derived in step 1 via

$$Z_G = \frac{\hat{\beta}_1 \hat{\gamma}}{SE_{BRR}(\hat{\beta}_1 \hat{\gamma})}, \text{ where } SE_{BRR}(\hat{\beta}_1 \hat{\gamma}) = \sqrt{\hat{\beta}_1^2 SE_{BRR}^2(\hat{\gamma}) + \hat{\gamma}^2 SE_{BRR}^2(\hat{\beta}_1)}.$$

Then we perform testing using a rejection region similar to the one specified in Sec. 2. In addition, (if appropriate) we can compute the relative effect size using the estimates obtained in step 1 and formula (4).

4. Applications to a study of smoking behavior

To illustrate the proposed procedure, we performed a study of nicotine dependence among daily smokers. The goal was to evaluate the significance of cigarette purchase price as a mediator in the association between race/ethnicity and nicotine dependence among U.S. daily smokers (during the period from 2010 to 2015). The dependent variable was the nicotine dependence measured as the average number of cigarettes smoked per day. The cigarette purchase price (per pack) referred to the last self-purchase. We considered two non-Hispanic racial/ethnic groups of daily smokers: White and Black/African American. Thus, we considered the single-mediator model (1) with

M = Cigarette Purchase Price per Pack,

Y = Nicotine Dependence (Average Number of Cigarettes Smoked per Day),

X = Race/Ethnicity (non-Hispanic White, non-Hispanic Black/African American).

Considering race/ethnicity as an independent variable in the model was motivated by the following research findings. First, there are racial/ethnic differences in cigarette purchasing prices. Specifically, among diverse racial/ethnic populations in the U.S., non-Hispanic (NH) American Indian/Alaska Native (AIAN) and NH White adult smokers purchase cigarettes, on average, at lower prices than the other adult smokers (Golden, Kong, and Ribisl 2016). Analyses controlling for additional factors related to consumer behaviors resulted in less pronounced differences in average prices but also indicated that NH AIAN adult smokers, on average, paid \$0.38 more per pack than did NH White adult smokers (Golden et al. 2016). These discrepancies could be explained in part by different consumer behaviors. For example, purchasing cigarettes on Indian reservations is associated with lower purchase prices (DeCicca, Kenkel, and Liu 2015; National Research Council 2015; Wang et al. 2017), and the rate of purchasing cigarettes on Indian reservations is significantly higher for NH AIAN relative to NH White daily smokers, and NH White relative to NH Black/African American daily smokers (Soulakova, Pack, and Ha 2018). Second, the levels of nicotine

dependence differ across race/ethnicity among daily smokers (Soulakova and Danczak 2017). Specifically, heavy smoking (16+ cigarettes per day) was most prevalent in NH White, NH AIAN and NH Multiracial daily smokers. Smoking within 30 minutes from awakening was most prevalent in NH White, NH Black, NH AIAN and NH Multiracial daily smokers, and night-smoking was most prevalent in NH Black, NH AIAN and NH Multiracial daily smokers. NH Hawaiian/Pacific Islander and Hispanic daily smokers had consistently lower rates for all three nicotine dependence measures (Soulakova and Danczak 2017).

Table 1 presents the set of considered covariates. Because some of these covariates are categorical with more than two levels, we fitted the design-based regression models (1) with 18 binary covariates ($I = 18$) using the pooled 2010–2011 and 2014–2015 TUS-CPS data. The sample of daily smokers ($n = 30,777$) was representative of about 20,261,285 daily smokers in the population. The cohort was 89.4% (27,507) non-Hispanic White and 10.6% (3,270) non-Hispanic Black/African American. The daily smokers, on average, smoked 16 cigarettes per day ($SE = 8.2$) and paid \$5.15 per pack of cigarettes during their last cigarette purchase ($SE = \$1.69$). Table 1 presents the summary statistics for the factors included as covariates in the models.

The significance level was 5%. All computing was performed using SAS/STAT®9.4 (SAS Institute Inc. 2017). Specifically, we used PROC SURVEYFREQ, PROC SURVEYMEANS, and PROC SURVEYREG with the BRR option (with Fay correction) and the main and replicate weights. In addition, we constructed the 95% confidence interval based on the standard normal distribution for the mediated effect $\beta_1 \cdot \gamma$.

The model for the mean cigarette purchase price per pack (the mediator) was significant ($R^2 \approx 24\%$, $F(19, 160) \approx 191$, $p < 0.0001$); the intercept and all covariates except for sex and survey mode were significant (p 's < 0.0001). The model for the nicotine dependence (the dependent variable) was also significant ($R^2 \approx 13\%$, $F(20, 160) \approx 164$, $p < 0.0001$); the intercept and all covariates were significant (p 's < 0.0300). Table 2 presents the results for each step of the procedure. As is shown, the generalized Sobel's test statistic had a value of -9.57 (which was in the rejection region), indicating significant mediated effect of cigarette purchase price ($p < 0.0001$). The corresponding 95% confidence interval for the mediated effect was $(-0.28, -0.19)$, which also illustrates that the effect is significantly different from zero. Therefore, the association between daily smoker's race/ethnicity and nicotine dependence is mediated by the cigarette purchase price. Because $\hat{\beta}_1 \hat{\gamma}$ and $\hat{\beta}_2$ were both negative, we also computed the relative effect size. However, the relative effect size was only 0.05, indicating low practical importance of the cigarette purchase price as a mediator in the association between race/ethnicity and nicotine dependence.

We note that it is important to correctly adjust for the TUS-CPS design specifics. Indeed, if one ignored all survey weights and incorrectly treated the sampling strategy as simple random sampling, then the confidence interval for the mediated effect would be $(-0.35, -0.25)$. While both approaches result in a significant finding, the latter interval would (incorrectly) suggest that the mediated effect is larger (in absolute value). In addition, if one used the main survey weight only (ignoring the replicate weights) and estimated variance

using Taylor's linearization, then after rounding to hundredths, the resulting confidence interval would be the same as the one based on the BRR approach, i.e., $(-0.28, -0.19)$. However, this method cannot be recommended in general, because in other cases this method and the correct one (based on the BRR) could result in discrepant findings (Ha and Soulakova 2018).

5. Conclusion

In this paper, we illustrated the applications of a single-mediator model for analysis of the TUS-CPS data. However, the approach can be easily modified to handle other types of designs; these adjustments should be incorporated when computing the design-based model coefficients in step 1 and the standard error in step 2.

The approach has several limitations. The main limitation is that although the analytical results can be used to inform scientists regarding population-wide characteristics and behaviors, and can be used in future research studies, the "observational nature" of the data prohibits making any definite claims. Therefore, no causal inferences can be made. In addition, while independent regressions described in this paper are commonly used to test for a mediated effect (Hayes 2017; Hill, Burdette, and Hale 2009; Parmelee, Harralson, Smith, and Schumacher 2007; Rutchick, Smyth, Lopoo, and Dusek 2009; Yang, Du, Qu, Gong, and Sun 2013), this approach ignores dependence between the mediator and the dependent variable. In addition, when testing for the mediated effect, we assumed that the generalized Sobel's test statistic follows standard normal distribution under the null hypothesis. However, this assumption might be violated in practice. This is a concern especially in studies with a small sample size. Moreover, the probability coverage of confidence intervals (based on the Sobel's standard error) could exceed the nominal confidence level even for large samples, leading to an over-conservative test (MacKinnon, Warsi, and Dwyer 1995). In these instances, alternative methods such as the confidence intervals based on the distribution of the product or resampling have been recommended (MacKinnon et al. 2004).

In the considered study of nicotine dependence among daily smokers, we detected a significant mediated effect of cigarette purchase price on the association between race/ethnicity and nicotine dependence. Specifically, non-Hispanic Black/African American daily smokers, on average, were less nicotine dependent than were non-Hispanic White daily smokers. This association was mediated by cigarette purchase price, but the magnitude of the effect was relatively low. Additional findings were (1) non-Hispanic Black/African American daily smokers pay more, on average, for a pack of cigarettes than do non-Hispanic White daily smokers, and (2) the higher cigarette purchase price is associated with lower nicotine dependence.

The study of nicotine dependence also has some limitations. First, we used TUS-CPS self-reports; thus, daily smokers were also identified using self-reported current smoking status. Therefore, there could be some misrepresentation of the target population (U.S. daily smokers) in the study. Nonetheless, given that TUS-CPS self-reported smoking information is generally accurate (Soulakova and Crockett 2014; Soulakova, Hartman, Liu, Willis, and

Augustine 2012), we anticipate this discrepancy to be negligible. In addition, the surveyed average number of cigarettes smoked per day was truncated at 40 cigarettes for all smokers who indicated smoking more than 40 cigarettes per day. In our cohort, 40 cigarettes per day was observed for 1,076 (3.5%) daily smokers. Therefore, the average number of 16 cigarettes (per day) reported in the study may be a (slight) under-estimate of the true average number of cigarettes smoked per day among daily smokers. The cigarette purchase price used in the study was defined using reports of price paid when last purchased pack or carton of cigarettes. Thus, the measure refers to the average (or actual) price per pack if a carton (or a pack) was purchased. An additional study limitation is that we did not conduct a sensitivity analysis (Imai, Keele, and Yamamoto 2010; Mackinnon 2008, Section 15.7).

Future research can be targeted toward adapting mediation methodology to complex survey data. For example, applications of the causal steps approach (Baron and Kenny 1986), difference of coefficients approach (Freedman and Schatzkin 1992; MacKinnon et al. 2004), and resampling approach based on the empirical distribution (MacKinnon et al. 2004) have not yet been addressed for complex sampling. Moreover, methods for complex survey data with a categorical dependent variable and/or mediator, multi-mediator problems, and problems with a mediator-predictor interaction (VanderWeele 2016; Wang, Nelson, and Albert 2013) have yet to be developed.

Acknowledgments

The authors would like to thank the reviewers for their helpful feedback, and James Holland (Scientific Writer), and Victoria Owens and Richard Pack (Part-time Scientists) for reading the draft and providing editing comments.

Funding

Research of Julia Soulakova and Trung Ha reported in this publication was supported by the National Institute On Minority Health And Health Disparities of the National Institutes of Health under Award Number R01MD009718. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- Alwin DF, and Hauser RM (1975). The decomposition of effects in path analysis. *American Sociological Review* 40 (1):37–47. doi:10.2307/2094445.
- Baron RM, and Kenny DA (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51 (6):1173–82. doi:10.1037/0022-3514.51.6.1173. [PubMed: 3806354]
- Black D, Sanders S, and Taylor L (2003). Measurement of higher education in the census and current population survey. *Source Journal of the American Statistical Association* 98 (463):545–54. doi:10.1198/016214503000000369.
- Burkhauser RV, Feng S, Jenkins SP, and Larrimore J (2011). Estimating trends in US income inequality using the current population survey: The importance of controlling for censoring. *The Journal of Economic Inequality* 9 (3):393–415. doi:10.1007/s10888-010-9131-6.
- DeCicca P, Kenkel D, and Liu F (2015). Reservation prices: An economic analysis of cigarette purchases on Indian Reservations. *National Tax Journal* 68 (1):93–118. doi:10.17310/ntj.2015.1.04.
- Freedman LS, and Schatzkin A (1992). Sample size for studying intermediate endpoints within intervention trails or observational studies. *American Journal of Epidemiology* 136 (9):1148–1159. [PubMed: 1462974]

- Golden SD, Kong AY, and Ribisl KM (2016). Racial and ethnic differences in what smokers report paying for their cigarettes. *Nicotine & Tobacco Research* 18 (7):1649–55. doi:10.1093/ntr/ntw033. [PubMed: 26874329]
- Ha T, and Soulakova JN (2018). Importance of Adjusting for Multi-Stage Design when Analyzing Data from Complex Surveys. In *New frontiers of biostatistics and bioinformatics*, editors Zhao Y and Chen D-G, (pp. 257–268). Switzerland: Springer Nature. doi:10.1007/978-3-319-99389-8.
- Hayes AF (2017). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: Guilford Publications.
- Hill TD, Burdette AM, and Hale L (2009). Neighborhood disorder, sleep quality, and psychological distress: Testing a model of structural amplification. *Health & Place* 15 (4):1006–13. doi:10.1016/j.healthplace.2009.04.001. [PubMed: 19447667]
- Imai K, Keele L, and Yamamoto T (2010). Identification, Inference and sensitivity analysis for causal mediation effects. *Statistical Science* 25 (1):51–71. doi:10.1214/10-STS321.
- MacKinnon DP (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Erlbaum.
- MacKinnon DP, Lockwood CM, and Williams J (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research* 39 (1):37–67. doi:10.1207/s15327906mbr3001_3. [PubMed: 26759934]
- MacKinnon DP, Warsi G, and Dwyer JH (1995). A Simulation Study of Mediated Effect Measures. *Multivariate Behavioral Research* 30 (1):41–62. doi:10.1207/s15327906mbr3001_3 [PubMed: 20157641]
- National Research Council. (2015). *Understanding the U.S. Illicit tobacco market*. Washington, D.C.: National Academies Press. doi:10.17226/19016.
- Parmelee PA, Harralson TL, Smith LA, and Schumacher HR. (2007). Necessary and discretionary activities in knee osteoarthritis: Do they mediate the pain–Depression relationship? *Pain Medicine* 8 (5):449–61. doi:10.1111/j.1526-4637.2007.00310.x. [PubMed: 17661863]
- Preacher KJ, and Kelley K (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods* 16 (2):93–115. doi:10.1037/a0022658. [PubMed: 21500915]
- Rutchick AM, Smyth JM, Lopoo LM, and Dusek JB (2009). Great expectations: The biasing effects of reported child behavior problems on educational expectancies and subsequent academic achievement. *Journal of Social and Clinical Psychology* 28 (3):392–413. doi:10.1521/jscp.2009.28.3.392.
- SAS Institute Inc. (2013). *SAS® 9.4 Product Documentation*.
- Sobel ME (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology* 13:290–312. doi:10.2307/270723.
- Soulakova JN, and Crockett LJ (2014). Consistency and recanting of Ever-Smoking status reported by self and proxy respondents one year Apart. *Journal of Addictive Behaviors Therapy & Rehabilitation* 3 (4):1000113. doi:10.4172/2324-9005.1000114.
- Soulakova JN, and Danczak RR (2017). Impact of menthol smoking on nicotine dependence for diverse racial/Ethnic groups of daily smokers. *Healthcare* 5 (1):2–8. doi:10.3390/healthcare5010002.
- Soulakova JN, Hartman AM, Liu B, Willis GB, and Augustine S (2012). Reliability of adult self-reported smoking history: Data from the tobacco use supplement to the current population survey 2002-2003 cohort. *Nicotine and Tobacco Research* 14 (8):952–60. doi:10.1093/ntr/ntr313. [PubMed: 22318688]
- Soulakova JN, Pack R, and Ha T (2018). Patterns and correlates of purchasing cigarettes on Indian reservations among daily smokers in the United States. *Drug and Alcohol Dependence* 192:88–93. doi:10.1016/j.drugalcdep.2018.07.036. [PubMed: 30243144]
- U.S. Department of Commerce, Census Bureau (2016). National Cancer Institute and Food and Drug Administration co-sponsored Tobacco Use Supplement to the Current Population Survey. 2014-15. (n.d.). Retrieved from https://thedataweb.rm.census.gov/ftp/cps_ftp.html#cpssupps.
- VanderWeele TJ (2016). *Mediation Analysis: A Practitioner’s Guide*. *Annual Review of Public Health* 37 (1):17–32. doi:10.1146/annurev-publhealth-032315-021402.

- Wang W, Nelson S, and Albert JM (2013). Estimation of causal mediation effects for a dichotomous outcome in Multiple-Mediator models using the mediation formula. *Statistics in Medicine* 32 (24):4211–28. doi:10.1002/sim.5830. [PubMed: 23650048]
- Wang X, Xu X, Tynan MA, Gerzoff RB, Caraballo RS, and Promoff GR (2017). Tax avoidance and evasion: Cigarette purchases from Indian reservations among US adult smokers, 2010-2011. *Public Health Reports* 132 (3):304–8. doi:10.1177/0033354917703653. [PubMed: 28395142]
- Wolter K (2007). *Introduction to variance estimation*. New York, NY: Springer.
- Yang J, Du F, Qu W, Gong Z, and Sun X (2013). Effects of personality on risky driving behavior and accident involvement for Chinese drivers. *Traffic Injury Prevention* 14 (6):565–71. doi:10.1080/15389588.2012.748903. [PubMed: 23859184]

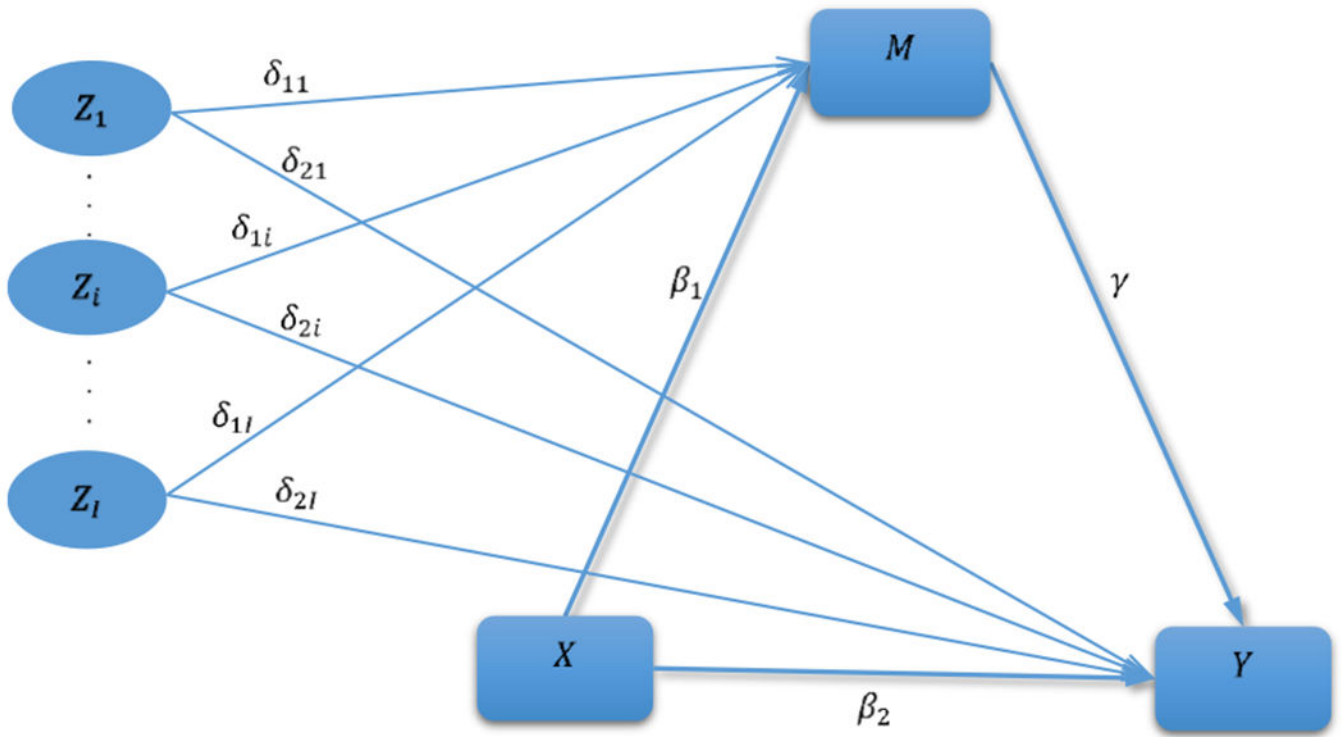


Figure 1.
Single-mediator model with I covariates.

Table 1.

Sample summary statistics for factors considered as covariates; 2010–2011 and 2014–2015 tobacco use supplement to the current population survey.

| Characteristic | Sample count | Percent (%) [*] |
|--|--------------|--------------------------|
| Age group | | |
| 18–24 | 2,149 | 11.6 |
| 25–44 | 11,503 | 37.4 |
| 45–64 | 13,660 | 41.7 |
| 65+ | 3,465 | 9.4 |
| Sex | | |
| Male | 14,905 | 52.2 |
| Female | 15,872 | 48.8 |
| Highest level of education | | |
| Less than high school | 4,756 | 16.0 |
| High school (or equivalent) | 13,070 | 42.3 |
| Some college or a bachelor's degree | 12,203 | 39.4 |
| Graduate degree (or equivalent) | 748 | 2.3 |
| Employment record | | |
| Employed | 17,844 | 58.6 |
| Unemployed | 2,630 | 9.6 |
| Not in labor force | 10,303 | 31.8 |
| Marital status | | |
| Married (spouse is present or absent) | 12,511 | 39.0 |
| Widowed, divorced, or separated | 10,174 | 30.6 |
| Never married | 8,092 | 30.4 |
| Region of residency | | |
| Northeast | 5,420 | 16.5 |
| Midwest | 8,641 | 27.9 |
| South | 11,688 | 42.2 |
| West | 5,028 | 13.4 |
| Area of residency | | |
| Metropolitan | 21,768 | 77.7 |
| Non-metropolitan | 9,009 | 22.3 |
| Residing in a state with an Indian Reservation | | |
| No | 10,433 | 32.2 |
| Yes | 20,344 | 67.8 |
| Survey mode | | |
| Phone | 17,395 | 56.5 |
| In-person | 13,382 | 43.5 |
| Survey year | | |
| 2010–2011 | 16,884 | 54.1 |
| 2014–2015 | 13,893 | 45.9 |

* All percentages except for the survey mode are based on the 2010–2011 and 2014–2015 population counts.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Testing for the Mediated Effect: Results for Each Step of the Procedure.

| Estimated quantity | Description | Estimate | Standard error |
|---|---|----------|----------------|
| Step 1: Estimating the design-based model coefficients | | | |
| $\hat{\beta}_1$ | Estimated effect of race/ethnicity on cigarette purchase price | 0.58 | 0.04 |
| $\hat{\gamma}$ | Estimated effect of cigarette purchase price on the average number of cigarettes smoked per day | -0.41 | 0.03 |
| $\hat{\beta}_2$ | Estimated effect of race/ethnicity (non-Hispanic Black/African American versus non-Hispanic White) on the average number of cigarettes smoked per day | -4.73 | 0.15 |
| Step 2: Testing for the mediated effect | | | |
| $\hat{\beta}_1^* \hat{\gamma}^*$ | Mediated effect | -0.24 | 0.02 |
| Z_G | Value of the generalized Sobel's test statistic | -9.57 | |
| Additional Step: Calculating the relative effect size | | | |
| $\frac{\hat{\beta}_1 \hat{\gamma}}{\hat{\beta}_2 + \hat{\beta}_1 \hat{\gamma}}$ | Relative effect size | 0.05 | |