



# HHS Public Access

Author manuscript

*Prev Med.* Author manuscript; available in PMC 2019 December 01.

Published in final edited form as:

*Prev Med.* 2018 December ; 117: 38–42. doi:10.1016/j.ypmed.2018.09.006.

## Understanding Individual Differences in Vulnerability to Cigarette Smoking is Enhanced by Attention to the Intersection of Common Risk Factors

Diann E. Gaalema<sup>1,2,3</sup>, Adam M. Leventhal<sup>4</sup>, Jeffrey S. Priest<sup>5</sup>, and Stephen T. Higgins<sup>1,2,3</sup>

<sup>1</sup>Vermont Center on Tobacco Regulatory Science, University of Vermont, Burlington, VT

<sup>2</sup>Departments of Psychiatry, University of Vermont, Burlington, VT

<sup>3</sup>Departments of Psychology, University of Vermont, Burlington, VT

<sup>4</sup>Department of Preventive Medicine, University of Southern California Keck School of Medicine, Los Angeles, CA

<sup>5</sup>Medical Biostatistics, University of Vermont, Burlington, VT

### Abstract

While smoking prevalence in the U.S. and other industrialized countries has decreased substantially, this change has been unevenly distributed, with dramatic decreases in certain subpopulations but little change or even increases in others. Accordingly, considerable attention has been fruitfully devoted to identifying important risk factors for smoking (e.g., mental illness, other substance use disorders). However, there has been little research on the intersection of these risk factors. As risk factors rarely occur in isolation, it is important to examine risk-factor profiles as is commonly done in studying other chronic conditions (e.g., cardiovascular disease). The purpose of this Commentary is to encourage greater interest in the intersection of multiple risk factors using cigarette smoking as an exemplar. We focus on the intersection of eight well-established risk factors for smoking (age, gender, race/ethnicity, educational attainment, poverty, drug abuse/dependence, alcohol abuse/dependence, mental illness). Studying the intersection of risk factors is likely to require use of innovative data-analytic methods. We illustrate, using years 2011–2016 of the US National Household Survey on Drug Use and Health, how Classification and Regression Tree (CART) analysis can be an effective tool for identifying risk profiles for smoking. Examination of the intersection of these risk factors elucidates a series of risk profiles with associated, orderly gradations in vulnerability to current smoking, including the striking and reliable strength of a college education as a stand-alone profile predicting low risk for current smoking, and illustrating the potentially increasing importance of drug abuse/dependence as a risk factor.

---

**Conflict of Interest:** The authors declare there is no conflict of interest.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Keywords

cigarette smoking; tobacco use; vulnerability; vulnerable populations; risk factors; risk profiles; Classification and Regression Tree analysis

Cigarette smoking continues to represent a major U.S. and global public health problem, with almost half a million smoking-attributable premature deaths reported yearly in the U.S. and 6.4 million globally (USDHHS, 2014; Reistma et al., 2017). Substantial progress has been made in reducing U.S. smoking rates (USDHHS, 2014). However, this decrease has been realized mainly among those who are more affluent and without other vulnerabilities, with certain subpopulations experiencing little or no decrease (Stanton et al., 2016; Dickerson et al., 2018) and still others experiencing increases (Fiore et al., 2008; Higgins & Chilcoat, 2009; Schroeder, 2016). Given these disparities, considerable effort is being devoted to identifying and understanding risk factors for smoking. While not an exhaustive list, age, gender, race, educational attainment, poverty, mental illness, and drug or alcohol use disorders have welldocumented associations with cigarette smoking in the U.S. and other industrialized countries (Higgins et al., 2015; 2016; Schroeder, 2016; USDHHS, 2014).

While the study of individual risk factors can begin to account for disparities in smoking risk, a more complete scientific understanding necessitates examining the intersection of risk factors. That is, many of the major risk factors for smoking are independently associated with risk and inevitably co-occur in what can be considered as risk profiles (e.g., mental illness inevitably occurs in the presence of chronological age, gender, race/ethnicity, and educational attainment; see Higgins et al., 2015; 2016). Investigation of risk profiles is common in the study of other chronic conditions (e.g., cardiovascular disease, diabetes). For example, in cardiology, the contribution of intersecting risk factors (e.g. cholesterol levels, smoking, age) to the probability of developing heart disease or having a stroke is well characterized (e.g. Goff et al., 2014). Unlike these other chronic conditions, however, relatively less scientific attention has been devoted to understanding how smoking risk varies in correspondence to the presence of co-occurring risk factors. The purpose of this Commentary is twofold: First, we suggest that risk profiles can be developed for understanding smoking vulnerability as they are for other chronic conditions through greater attention to characterizing how risk factors intersect. Second, we suggest examining smoking risk in this manner may have practical utility such as helping policy makers with decisions on how to better target tobacco control and regulatory efforts to reduce disparities in cigarette smoking and use of other tobacco products. For example, decisions on how to target (a) state-level tobacco control funds for smoking cessation, (b) federal public health education campaigns on the dangers of tobacco use, or (c) new tobacco control or regulatory research initiatives within the National Institutes of Health may be enhanced by considering the type of risk profiling illustrated below. We are not suggesting that this is going to qualitatively alter such decision-making, but perhaps enhance it through a more complete scientific understanding of how smoking risk varies and is distributed in the population of interest.

## Examining Risk Factor Intersections

Studying the intersection of risk factors is optimized by use of innovative data-analytic methods that incorporate the multitude of complex combinations of designations that individuals may have across multiple risk factors to calculate a parsimonious risk likelihood score. While certainly not a silver bullet in that regard, below we illustrate Classification and Regression Tree (CART) analysis as an exemplar method that can be effective in identifying risk profiles for smoking. CART analysis has been used to develop risk profiles across a variety of health-related outcomes from likelihood of recovery from cardiac arrest (Kaji et al., 2014) to estimating risk for illicit drug use (Kurti et al., 2016).

CART analysis is a nonparametric procedure for dividing a population of interest into mutually exclusive subgroups or nodes based on a dependent variable of interest such as current smoking status (Breiman et al., 1984). During this process the observed independent variables with the most explanatory power in accounting for that dependent variable are identified. These observed variables can be used repeatedly across branches, depending on their relative importance in splitting groups. Beginning with the entire sample, an algorithm identifies a single independent variable, where splitting the sample (parent node) on that variable will maximize the distinction between the two resulting subsamples (child nodes) on the dependent variable. Nodes continue to be split into subsamples in this fashion, based on which independent variable will continue to maximize distinction between the resulting nodes until the subsamples reach a minimum size that was either predetermined (i.e. there is no standard size; we set minimum node size at 1000 in the following examples so that each node would represent at least 0.5% of the sample) or where further splits do not significantly improve classification within the model (terminal nodes). CART also has the advantage of not being prone to multicollinearity. The same independent variable can appear multiple times across splits and nodes, as a “primary” or “surrogate” splitter. The less correlated the independent variables are, the greater their probable differences in roles as primary and surrogate splitters. The number of times an independent variable serves as a primary and surrogate splitter will determine its relative strength as a predictor in the model. Additional details of the CART process can be found elsewhere (Breiman et al., 1984; Lemon et al., 2003; Lei et al., 2015) and guides to using this method are also available (Therneau et al., 2018a,b).

## Risk Profiles Using the US National Survey on Drug Use and Health

Higgins et al. (2016) conducted an initial examination of the intersection of common risk factors for current smoking among U.S. adults by using CART to analyze data from the annual U.S. National Survey on Drug Use and Health (NSDUH) collapsed across the years 2011–2013. The estimated rate of current smoking for that period was 21.6% (N = 114,426) and the risk factors examined were age, gender, race/ethnicity, educational attainment, poverty, drug abuse/dependence, alcohol abuse/dependence, and mental illness. The relative strength of the eight risk factors can be seen in Table 1. The analysis resulted in the 13 terminal nodes or risk profiles displayed in the bottom row of Figure 1, indicating for each profile proportion of the overall population, the smoking rate, and the share of the current smokers overall contained in that profile are listed.

Educational attainment was the strongest predictor of smoking risk, with the first split of the entire population into nodes being based on whether someone was a college graduate. This initial classification, represented by the leftmost terminal node, had the lowest smoking prevalence (11%), but represented 30% of the U.S. adult non-institutionalized population and thus 15% of adult current smokers. Further splits were based on the presence or absence of additional risk factors. Smoking prevalence varied in an orderly manner across the 13 risk profiles from a low of 11% (those with a college education) to a high of 74% (< college education, past year drug abuse/dependence, aged 26–64 years), almost a sevenfold difference. However, note that the highest risk profile was in a node representing only 1% of the population and thus only 3% of U.S. adult smokers. A final risk profile worth underscoring is the fourth node (at least high school or some college, 18–64 years, racial/ethnic makeup excluding Asian or Latino). This node contains the largest proportion of U.S. adult smokers (43%) yet lacks risk factors commonly associated with smoking disparities (substance use disorders, poverty, mental illness).

In the interest of looking at the stability of these risk-profiles, we conducted an identical CART analysis using the years 2014–2016 of the NSDUH (Figure 2). The overall current adult smoking rate for this time-period was 19.7% (N = 127,857). These updated analyses show a generally stable pattern of risk profiles, with a few notable features. Again, being a college graduate was associated with the lowest risk and was the only single-predictor risk profile. The reliability of this finding across the two data sets, along with the fact that some level of educational attainment is represented in each of the risk profiles, underscores the striking strength of the relationship between educational attainment and smoking risk (Graham et al., 2006; Higgins and Chilcoat, 2009; Higgins et al., 2009; Hiscock et al., 2012; Kandel et al., 2009; Schroeder, 2016). The relative strength of the other risk factors was also relatively stable across analyses (Table 1). Any one risk factor changed at most by one level, with the exception of past year drug abuse/dependence which moved from 4th to 2nd place. The highest risk profile represented those without a college degree and past year drug/abuse dependence with a smoking rate of 60%, again a seven-fold increase over the lowest risk profile. The largest proportion of the smoking population is once again found in an intermediate terminal node representing those without any college, no past year alcohol or drug abuse/dependence, between the ages of 18–64 years, an annual income above the federal poverty level, and non-Hispanic White, Native American or Other race/ethnicity with a smoking rate of 32% and accounting for 21% of the U.S. adult smoking population. Indeed, the majority of US smokers (72%) in this recent analysis are located in profiles representing less than a college education in combination with age and race/ethnicity, without the presence of any risk factors suggesting obvious social or individual instability.

## Conclusions

These analyses have several scientific and practical implications worth underscoring. First, they illustrate the overarching point of this Commentary, which is that smoking risk typically varies in an orderly, graded manner in association with the presence of multiple, co-occurring, independent risk factors (Higgins et al., 2015; 2016). While it is common to speak of smoking prevalence as being particularly high in certain vulnerable populations (e.g., those with mental illness or other substance use disorders) these findings illustrate how

those rates inevitably involve the presence of other co-occurring risk factors. The present research focuses exclusively on current smoking among adults but recent research demonstrates a similar pattern for smoking initiation among adolescents (Wellman et al. 2018). Indeed, while in this commentary we focus on current smoking, this approach would be complimented by examinations of uptake and cessation and use of longitudinal data.

Second, the one exception to this pervasive pattern of co-occurring risk factors in both CART analyses warrants underscoring. A college education was the only single-predictor risk profile in each and the profile associated with the lowest risk. Educational attainment has a striking, inverse association with smoking risk, as well as other health behaviors (Cutler & Lleras-Muney, 2010; Cutler et al., 2015; Gaalema et al., 2017), with evidence from natural experiments suggesting a causal influence (e.g., Currie & Moretti, 2003). While low educational attainment is more typically the point of focus regarding smoking risk, the present results put the focus on higher educational attainment perhaps relating to the growing importance of a college education to economic success in industrialized economies (Moretti, 2012). Recent research on educational attainment as a protective factor against smoking and other health behaviors suggests an influence of macroeconomic factors on that relationship (Cutler et al., 2015).

Third, the results across the two CART analyses suggest that the relative strength of the connection between drug abuse/dependence and smoking in the U.S. may be increasing, perhaps related to the ongoing opioid epidemic. It is important to note that initiation of smoking typically precedes use of illicit drugs and a recent CART analysis (Kurti et al., 2016) identified alcohol use and cigarette smoking as the two strongest predictors for illicit drug use among U.S. adults. While substantial resources are appropriately being directed towards curtailing the opioid epidemic, policy makers may want to be mindful of these associations with alcohol use and cigarette smoking.

Lastly, as was noted above, examining risk profiles in combination with the proportion of the overall and smoking populations represented (as demonstrated above) provides potentially useful information to consider in discussions on how best to target limited tobacco control and regulatory resources in efforts to reduce smoking disparities. Up to seven-fold differences in smoking prevalence were demonstrated across risk profiles illustrating tremendous disparities in smoking vulnerability and differential need for assistance. However, it was also the case that the risk profiles with the highest smoking prevalence rates included relatively small proportions of the overall smoking population, while those with more modest smoking prevalence rates included the majority of the smoking population. Those contrasts would seem to contain the grist for some tough policy decisions. Indeed, if we had to identify a single observation revealed by the risk-profile models discussed in this Commentary that might be overlooked with a more conventional single-factor approach it would be those intermediate risk profiles in Figures 1 and 2. Those profiles involved what in many ways are unremarkable combinations of cooccurring educational attainment, age, and race/ethnicity that were nevertheless associated with an above-average smoking prevalence. Moreover, because the profiles also represented a relatively large swath of the overall U.S. population they also included a larger proportion of the smoker population than any other

profile in both models and as such warrant considerable attention in efforts to reduce smoking disparities.

While our purpose in this Commentary is not to make specific recommendations about how best to target efforts to reduce smoking disparities, we suggest that such decision making would be enhanced by greater attention to understanding how risk varies in correspondence to the intersection of independent risk factors. By leveraging analytic tools like CART to identify and quantify the smoking risk associated with those intersections and the proportions of the population impacted by them, policy could be backed by a more complete and stronger empirical foundation.

## Acknowledgments

**Funding:** Support for this project came from TCORS award P50DA036114 from the National Institute on Drug Abuse (NIDA) and FDA, Centers of Biomedical Research Excellence P20GM103644 award from the National Institute on General Medical Sciences, and TCORS award P50 CA180905. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the FDA.

## References

- Breiman L; Friedman JH Olshen RA; Stone CJ Classification and regression trees. Belmont, CA: Wadsworth; 1984.
- Currie J, & Moretti E (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *The Quarterly Journal of Economics*, 118(4), 1495-1532.
- Cutler DM, Huang W, & Lleras-Muney A (2015). When does education matter? The protective effect of education for cohorts graduating in bad times. *Social Science & Medicine*, 127, 63–73. [PubMed: 25113567]
- Cutler DM, & Lleras-Muney A (2010). Understanding differences in health behaviors by education. *Journal of health economics*, 29(1), 1–28. [PubMed: 19963292]
- Dickerson F, Schroeder J, Katsafanas E, Khushalani S, Origoni AE, Savage C, . & Yolken RH. (2018). Cigarette Smoking by Patients With Serious Mental Illness, 1999–2016: An Increasing Disparity. *Psychiat Serv*, 69, 147–153
- Fiore MC, Jaen CR, Baker T, Bailey WC, Benowitz NL, Curry SEEA, & Henderson PN (2008). *Treating tobacco use and dependence: 2008 update*. Rockville, MD: US Department of Health and Human Services.
- Gaalema DE, Elliott RJ, Morford ZH, Higgins ST, & Ades PA (2017). Effect of socioeconomic status on propensity to change risk behaviors following myocardial infarction: implications for healthy lifestyle medicine. *Progress in Cardiovascular Diseases*, 60(1), 159–168. [PubMed: 28063785]
- Goff DC, Jr, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Sr, Gibbons R, Greenland P, Lackland DT, Levy D, O'Donnell CJ, Robinson JG, Schwartz JS, Shero ST, Smith SC, Jr, Sorlie P, Stone NJ, Wilson PWF. (2014). 2013 ACC/ AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*, 63, 2935–59. [PubMed: 24239921]
- Graham H, Inskip HM, Francis B, & Harman J (2006). Pathways of disadvantage and smoking careers: evidence and policy implications. *J Epidemiol Commun H*, 60(suppl 2), ii7–ii12.
- Higgins ST, & Chilcoat HD (2009). Women and smoking: an interdisciplinary examination of socioeconomic influences. *Drug Alcohol Depen*, 104, S1–S5.
- Higgins ST, Heil SH, Badger GJ, Skelly JM, Solomon LJ, & Bernstein IM (2009). Educational disadvantage and cigarette smoking during pregnancy. *Drug Alcohol Depen*, 104, S100–S105.
- Higgins ST, Kurti AN, Redner R, White TJ, Gaalema DE, Roberts ME, & Henningfield JE (2015). A literature review on prevalence of gender differences and intersections with other vulnerabilities to tobacco use in the United States, 2004–2014. *Prev Med*, 80, 89–100. [PubMed: 26123717]

- Higgins ST, Kurti AN, Redner R, White TJ, Keith DR, Gaalema DE, . & Priest JS. (2016). Co-occurring risk factors for current cigarette smoking in a US nationally representative sample. *Prev Med*, 92, 110–117. [PubMed: 26902875]
- Hiscock R, Bauld L, Amos A, Fidler JA, & Munafò M (2012). Socioeconomic status and smoking: a review. *Ann of NY Acad Sci*, 1248(1), 107–123. [PubMed: 22092035]
- Kaji AH, Hanif AM, Bosson N, Ostermayer D, & Niemann JT (2014). Predictors of neurologic outcome in patients resuscitated from out-of-hospital cardiac arrest using classification and regression tree analysis. *The American Journal of Cardiology*, 114(7), 1024–1028. [PubMed: 25118118]
- Kandel DB, Griesler PC, & Schaffran C (2009). Educational attainment and smoking among women: risk factors and consequences for offspring. *Drug Alcohol Depen*, 104, S24–S33.
- Kurti AN, Keith DR, Noble A, Priest JS, Sprague B, & Higgins ST (2016). Characterizing the intersection of Co-occurring risk factors for illicit drug abuse and dependence in a US nationally representative sample. *Preventive medicine*, 92, 118–125. [PubMed: 27687534]
- Lei Y, Nollen N, Ahluwalia JS, Yu Q, Mayo MS (2015). An application in identifying high-risk populations in alternative tobacco product use utilizing logistic regression and CART: A heuristic comparison. *BMC Public Health*; 15:341 [PubMed: 25879872]
- Lemon SC, Roy J, Clark MA, Friedmann PD, Rakowski W (2003). Classification and regression tree analysis in public health: Methodological review and comparison with logistic regression. *Ann Behav Med*; 26(3):172–181. [PubMed: 14644693]
- Moretti E (2012). *The new geography of jobs*. Boston: Mariner Books, Houghton Mifflin Harcourt.
- Reitsma MB, Fullman N, Ng M, Salama JS, Abajobir A, Abate KH, . & Adebisi AO (2017). Smoking prevalence and attributable disease burden in 195 countries and territories, 1990–2015: a systematic analysis from the Global Burden of Disease Study 2015. *The Lancet*, 389(10082), 1885–1906.
- Schroeder SA (2016). American health improvement depends upon addressing class disparities. *Prev Med*, 92, 6–15. [PubMed: 27018943]
- Stanton CA, Keith DR, Gaalema DE, Bunn JY, Doogan NJ, Redner R, . & Higgins ST (2016). Trends in tobacco use among US adults with chronic health conditions: National Survey on Drug Use and Health 2005–2013. *Preventive medicine*, 92, 160–168. [PubMed: 27090919]
- Therneau TM, and Atkinson EJ (2018a). An Introduction to Recursive Partitioning Using the RPART Routines. <https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf>
- Therneau T, Atkinson E, Ripley B (2018b). Package ‘rpart’. <https://cran.rproject.org/web/packages/rpart/rpart.pdf>.
- U.S. Department of Health and Human Services. *The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General* Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2014.
- Wellman RJ, Sylvestre MP, O’Loughlin EK, Dutczak H, Montreuil A, Datta GD, & O’Loughlin J (2018). Socioeconomic status is associated with the prevalence and co-occurrence of risk factors for cigarette smoking initiation during adolescence. *International journal of public health*, 63(1), 125–136. [PubMed: 29116338]
- Williams JM, Steinberg ML, Griffiths KG, & Cooperman N (2013). Smokers with behavioral health comorbidity should be designated a tobacco use disparity group. *American journal of public health*, 103(9), 1549–1555. [PubMed: 23865661]

### Highlights

Risk factors for current smoking rarely occur in isolation

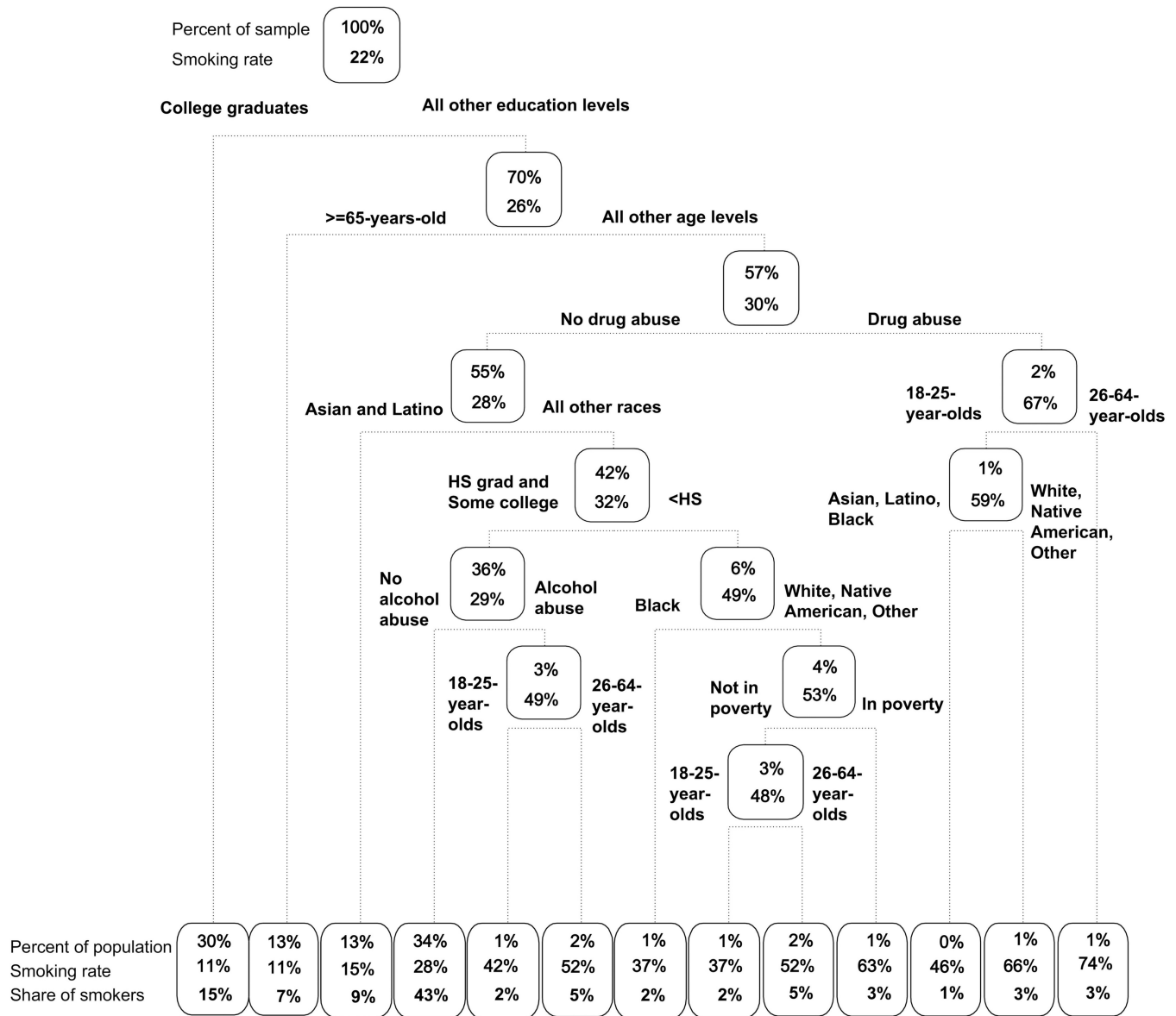
Classification and Regression Tree (CART) analysis is one way to examine risk profiles

Data from a national survey demonstrate how combinations of factors associate with current smoking

Education is a strong protective factor and the importance of drug use may be increasing

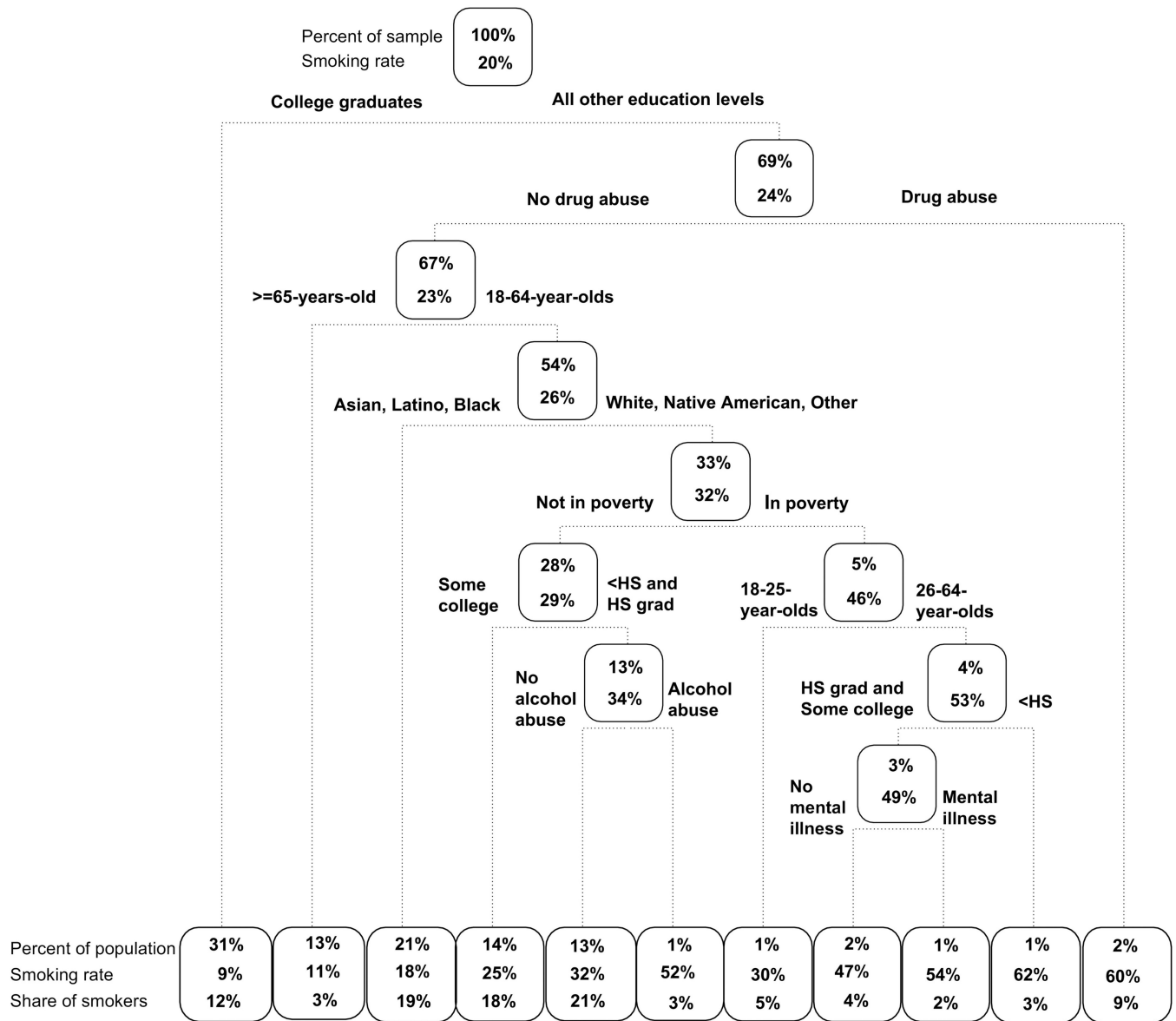
Policy decision making would be enhanced by examining risk profiles





**Figure 1:** A pruned, weighted classification and regression tree (CART) model of associations between current (past 30 days) smoking status and the following eight risk factors in the U.S. adult ( 18 years of age) population: educational attainment, age, race/ethnicity, past year drug abuse/dependence, past year alcohol abuse/dependence, annual income below federal poverty level, and past year mental illness in years 2011–2013 of the National Survey on Drug Use and Health (N = 114,246). Rectangles (nodes) represent the entire population (top-most node) or population subgroups (all other nodes). Within each node the top line lists the percent of the overall adult population represented within that node and the second line represents the smoking rate for that node. Using the root node as an example, this node represents 100% of the U.S. non-institutionalized adult population and 22% of them are smokers. Lines below nodes represent the binary branching around particular risk factors and risk-factor levels into subgroup nodes with further potential partitioning based on

additional risk factors/levels. The bottom row comprises terminal nodes (i.e., final partitioning for a particular subgroup, minimal terminal node size set to 1000 individuals). Terminal nodes contain the same information as the other nodes plus an additional line showing percent of all adult current smokers represented by that node.



**Figure 2:** A pruned, weighted classification and regression tree (CART) model of associations between current (past 30 days) smoking status and the following eight risk factors in the U.S. adult (18 years of age) population: educational attainment, age, race/ethnicity, past year drug abuse/dependence, past year alcohol abuse/dependence, annual income below federal poverty level, and past year mental illness in years 2014–2016 of the National Survey on Drug Use and Health (N = 127,857). Rectangles (nodes) represent the entire population (top-most node) or population subgroups (all other nodes). Within each node the top line lists the percent of the overall adult population represented within that node and the second line represents the smoking rate for that node. Using the root node as an example, this node represents 100% of the U.S. non-institutionalized adult population and 20% of them are smokers. Lines below nodes represent the binary branching around particular risk factors and risk-factor levels into subgroup nodes with further potential partitioning based on

additional risk factors/levels. The bottom row comprises terminal nodes (i.e., final partitioning for a particular subgroup, minimal terminal node size set to 1000 individuals). Terminal nodes contain the same information as the other nodes plus an additional line showing percent of all adult current smokers represented by that node.

**Table 1:**

Ranking of Eight Risk Factors Affecting Smoking from Greatest to Least in the US National Survey on Drug Use and Health

<u>2011–2013</u>	<u>2014–2016</u>
Education	Education
Age	Drug
Race	Age
Drug	Race
Alcohol	Poverty
Poverty	Alcohol
Sex	Sex
Mental Illness	Mental Illness

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript