Michigan Teen Smoking and Drug Use Survey Sample Design

Kevin Linares
and
Jianing Zou
and
Weishan Jiang
and
Xiaoqing Liu
University of Maryland

April 17, 2025

Abstract

1 Introduction

The State of Michigan Department of Education requires data to monitor teenage smoking and drug use, partly to assess compliance with tobacco industry settlements. This report details the design of a statewide probability sample of Michigan teenagers (enrolled students in grades 7-12) developed to meet the Department's needs. The objective was to create a cost-effective sample design capable of producing estimates for key variables with a coefficient of variation (CV) no greater than 0.05, for both the state overall and for specific regions. Based on cost considerations and a review of alternatives, a two-stage school-based sample design was chosen by the client. This report outlines the overall design, stratification and allocation plan, selection procedures, and estimation methods.

2 Overall Design

The sample design is a two-stage stratified cluster sample.

- **Target Population:** Students enrolled in grades 7 through 12 in public and non-public schools in Michigan during the fall 2025 survey period. The estimated total population size is N = 830,138 students.
- **Sampling Frame:** The primary frame is the Michigan Department of Education's 2024 list of public and non-public schools with student headcounts. This frame is assumed to be complete for the target population, although it excludes homeschooled students and those who have dropped out. At the second stage, student rosters will be obtained from selected schools.
- **Stages of Selection:**

- Stage 1 (Primary Sampling Units, PSUs): Schools selected with probability proportio

- Stage 2 (Secondary Sampling Units, SSUs): Students selected systematically from ros

- **Stratification:** The frame of schools is stratified explicitly by the nine education regions defined by the client to improve precision and allow for regional estimates. Implicit stratification is achieved by sorting schools within each region prior to systematic selection (e.g., by size).
- **Key Variables & Precision Targets:** The design focuses on achieving a coefficient of variation (CV) of 0.05 or less for state-level estimates of:

- Proportion who ever smoked one cigarette (expected p = 0.25).

- Proportion who ever smoked marijuana (expected p = 0.15).

- Mean age when first approached to smoke cigarettes or marijuana (expected mean = 12

- **Initial SRS Sample Sizes:** Ignoring clustering and FPC, the required SRS sample sizes to meet the CV=0.05 target were calculated as: n_srs(cig)=1200, n_srs(mj)=2267, and n_srs(age)=3.
- **Design Parameters (roh, deff):** Using design effects from a similar prior study (deff_cig=2.5, deff_mj=2.0, deff_age=1.7, based on a=150 schools, m=50 students/school), the estimated intraclass correlations (roh) were calculated as: roh_cig=0.0306, roh_mj=0.0204, roh_age=0.0143.
- **Optimal Cluster Size (b*) and Number of Clusters (a):** Considering a budget

of $500,000 and estimated costs of C_a=$3,000 per school and C_b=$50 per student, the optimal subsample size (b*, denoted m_opt in code) and number of schools (a, denoted n_opt in code) were calculated for each variable. To meet the precision target for all variables within budget, the design based on optimizing for marijuana smoking (which required the largest sample size) was chosen. This yields:

  - Optimal average cluster size: **b* = 53 students**
  - Optimal number of clusters (schools): **a = 87 schools** This design has an estimated total cost of $491,550.

- **Non-Response Adjustments:** To account for anticipated non-response rates of 70% among schools (30% RR) and 30% among students within cooperating schools (70% RR), the number of schools and the within-school target take were adjusted:
  - Number of schools to select: a_select = 87 / 0.30 **290 schools**
  - Target students per school (for fieldwork): b_field = 53 / 0.70 **76 students**

Stratification and Allocation

The sample of schools is explicitly stratified by the nine education regions provided by the client. The allocation of the sample across these strata is designed to be proportionate to the total number of students (Measure of Size, MOS) in each stratum, aiming for an equal probability of selection method (epsem) overall.