Michigan Teen Smoking and Drug Use Survey
Sample Design

Team Cochran
April 16, 2025

Project Overview & Objectives

- **Key Variables and Precision Requirement:**
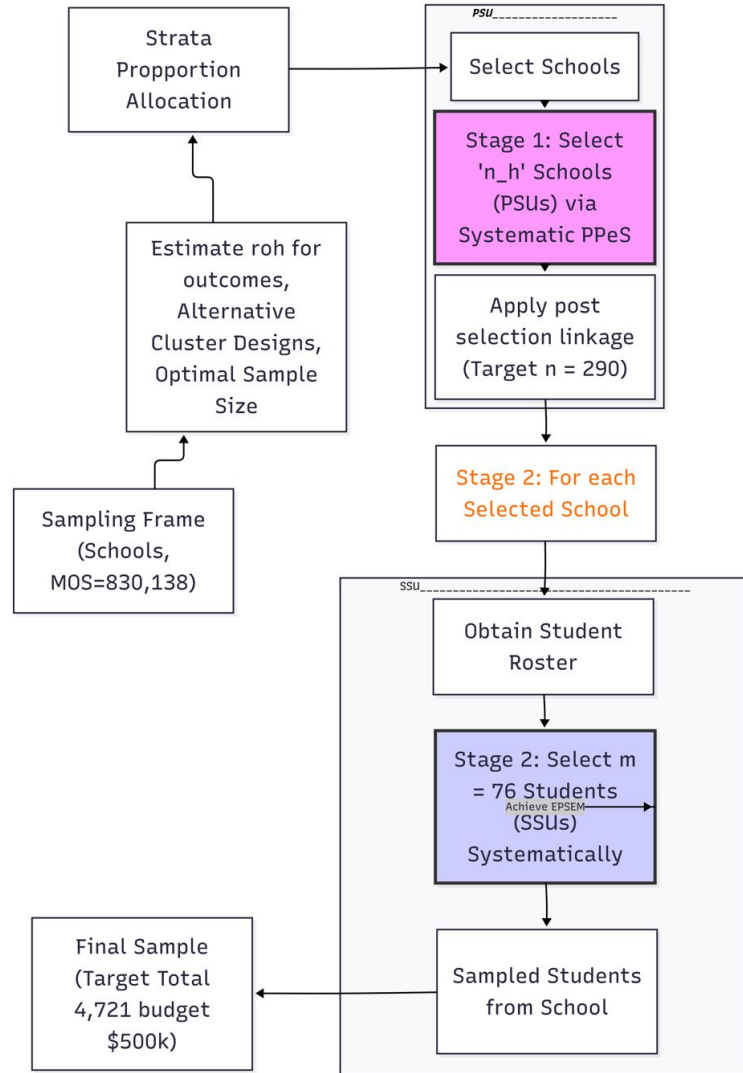
Outcome	type	desire_cv	expect_mean
smoked_cig	prop	0.05	0.25
smoked_mj	prop	0.05	0.15
age_approached_to_smoke	mean	0.05	12.00

- **Sampling Frame:** 2024 MDE list of (non)public schools with head counts. Total Students (N) ≈ 830,138.

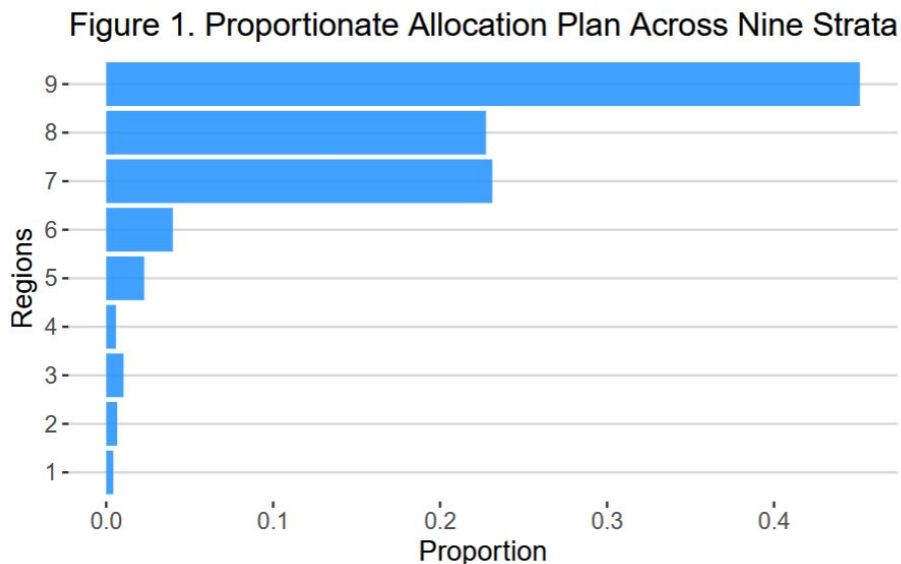- **Frame Limitations:** Excludes homeschooled students and dropouts. Assumes provided list is exhaustive.

Stratification & Allocation

- Nine Education regions
- Increase precision, separate regional estimates

Figure 1. Proportionate Allocation Plan Across Nine Strata

UNIVERSITY OF MARYLAND

Determining Sample Size Components (n & m))

- **Estimating roh:**

$$\hat{roh} = \frac{deff - 1}{m - 1}$$

Outcome	desire_deff	roh
smoked_cig	2.5	0.0306
smoked_mj	2.0	0.0204
age_approached_to_smoke	1.7	0.0143

- **Budget & Costs:** c_n= \$3, 000, c_m= \$50, total C = \$500, 000

$$m_{opt} = \sqrt{\frac{c_n}{c_m} \frac{1 - roh}{roh}}$$

Outcome	m_opt	n_opt	deff_new	total_nm	cost
smoked_cig	43	96	2.3035	4208	\$494,400
smoked_mj	53	87	2.0746	4721	\$491,550
age_smoke	64	80	1.9053	5173	\$496,000

$`Option 1`

Outcome	expect_mean	se	lower	upper	var_ck
smoked_cig	0.25	0.011	0.228	0.272	yes
smoked_mj	0.15	0.008	0.135	0.165	no
age_approached_to_smoke	12.00	0.059	11.885	12.115	yes

$`Option 2`

Outcome	expect_mean	se	lower	upper	var_ck
smoked_cig	0.25	0.012	0.227	0.273	yes
smoked_mj	0.15	0.007	0.135	0.165	yes
age_approached_to_smoke	12.00	0.058	11.887	12.113	yes

$`Option 3`

Outcome	expect_mean	se	lower	upper	var_ck
smoked_cig	0.25	0.012	0.227	0.273	yes
smoked_mj	0.15	0.008	0.135	0.165	no
age_approached_to_smoke	12.00	0.058	11.887	12.113	yes

UNIVERSITY OF MARYLAND

FEARLESSLY FORWARD

Stage 1, School Selection (PPeS)

- Non-response Adjustment
  - n = 290 (n_opt = 87)
  - m = 76 (m_opt = 53)
  - Cluster_allocation, n_h
- Selection Method
  - Systematic PPeS, within stratum
  - Implicit stratification
    - Sorted by student size
  - School allocation

$$n_{opt} \times prop\_allocation$$

- Sampling Interval
  - Calculate sample interval
- Linking
  - Post selection linkage

Region	prop_allocation	n_h	k_h	RN
1	0.004	1	3561	3168
2	0.007	2	2737	2310
3	0.010	3	2877	1321
4	0.006	2	2428	131
5	0.023	7	2701	2122
6	0.040	12	2761	2114
7	0.231	67	2866	374
8	0.227	66	2861	380
9	0.451	130	2883	1673

UNIVERSITY OF MARYLAND FEARLESSLY FORWARD

Stage 2, Student Selection & Linking

- Goal
  - Maintain EPSEM across strata
- Selection Technique
  - Calculate random start, 1 to k_{hi}
  - Select student at random start, $\epsilon k'_{hi} th$
  - Minimum MOS check, if fails link w/ next school until MOS is achieved

- Student Selection
  - Systematic random sampling from student roster within selected cluster of schools
- Probability of student inclusion
  - Target students, 4, 721

$$P_{hi} = \frac{n_h \times MOS_{hi}}{MOS_h} \times \frac{m_h}{MOS_{hi}} = \frac{n_h \times m_h}{MOS_h}$$

UNIVERSITY OF MARYLAND FEARLESSLY FORWARD

6

Variance Estimation

Method
- Paired Difference Replication Method (Taylor Series Linearization compatible). Accounts for stratification and clustering.

Variance Strata (Pseudo-Strata)
- Formed by sequentially pairing the a_select = 290 selected schools based on selection order within original strata. Total V = 145 variance/pseudo strata.

Variance Formula Concept
- Var(estimate) = Sum[(Estimate_PairMember1 - Estimate_PairMember2)^2] summed over all V=145 pairs (with appropriate scaling).

Degrees of Freedom
- df = Number of Pairs = 145.

Subclass Estimation (e.g., 20% low-income)
- Use same variance estimation method and df. Expected ~11 subclass members per school appears adequate.

Questions?