# Sampling Distributions

## Case Study 1: National Youth Tobacco Study

The National Youth Tobacco Survey (NYTS) is a complex survey carried out by the CDC to better understand the rate of e-cigarette usage among teens. It is a school-based, pencil-and-paper questionnaire, self-administered to a nationally representative sample of students in grades 6-12 in the U.S. In 2019, 5,675 students completed the NYTS; the overall survey response rate was 71.6%. Among other questions, students were asked if they had used an e-cigarette in the past 30 days.

1. Describe the sample. How large is it ($n$) and what is the unit of observation?

2. Describe the population. Approximately how large is it ($N$)?

3. Select a population parameter that researchers would be interested in estimating and the corresponding statistic that will serve as the estimate.

4. Identify possible sources of statistical bias and whether they are selection, measurement, and non-response bias. How do you expect this will effect the estimate of the parameter?

5. Identify possible sources of variation and whether they are sampling or measurement variability. How do you expect this will effect the estimate of the parameter?

6. On the back of this page make two sketches: 1. an empirical distribution of the data and 2. what you think the sampling distribution of the statistic might look like given what you know about the data collection. Be sure to label the axes of both plots. Illustrate any anticipation of bias you have by adding a vertical line on the sampling distribution that hits the x-axis where you expect the population parameter to be.