General Analysis Plan for Binary Outcomes

Refer to the example paper [Owens, V., Ha, T. and Soulakova, J.N. (2019). Widespread Use of Flavored e-Cigarettes and Hookah Tobacco in the United States. Preventive Medicine Reports 11:290-296.](https://www.sciencedirect.com/science/article/pii/S2211335519300397)

**Create the studied cohort.** Limit the sample to respondents with characteristics of interest.

*Example. The sample was limited to current users of e-cigarettes or hookah tobacco.*

**Prepare measures.** Variables may need to be created for measures of interest.

*Example. Respondents who reported they used e-cigarettes “every day” or “some days” were combined into a new “current user” group that formed the sample.*

**Check sample sizes for subgroups**. Delete missing values before checking sample sizes. Sample sizes must be sufficient to conduct tests. As a benchmark, the frequency in each cell should be at least 5. If sample sizes are too small, consult your project supervisor.

*Example. The output below shows insufficient sample size for NH Hawaiian/Pacif Island.*



**Use chi-square tests to assess significance of relationships.** Conduct tests for associations between the main variable of interest and other characteristics.

*Example. Associations were tested between the main measure “use of flavored e-cigarettes” and a variety of sociodemographic characteristics (age, sex, geographic location, etc.).*

**Create a logistic regression model.** Create a model to predict the binary main measure based on other characteristics.

*Example. A logistic regression model was created for both main measures. The model that predicted use of flavored e-cigarettes included age, sex, region of residence, regular cigarette smoking status, race/ethnicity, and marital status.*