Cover page for Data Analysis Plan

NCT03860077

Grant number: R01DA047356

Grant Title: Impact of Nicotine Reduction on Adolescent Cigarette Use, Alternative Tobacco Use, and Harm From Tobacco

Document date: 9/1/23

Data Analytic Approach and Power.

Overview and justification. Mixed effects models evaluated change in total cigarette use, alternative tobacco product use, study cigarette use, and biomarkers of nicotine exposure from baseline to the final study week. All analyses were with SAS V9.4 software (SAS Institute Inc., Cary, NC, USA, 2016). Due to data-collection disruptions related to the COVID-19 pandemic and intermittent data-collection challenges with the ecological momentary assessment (EMA) program, we elected to utilize the most complete data stream for primary analyses, which was Time Line Follow Back (TLFB) of outcomes collected at Week 4 and Baseline. Consistent with our prior work (e.g., [1]), aggregated timeline follow-back (TLFB) data from the final week of study cigarette use (i.e., Week 4) was compared to a one-week baseline period during which participants used usual brand cigarettes (i.e., Baseline). The mixed effect model using TLFB data was chosen over other possible analytic approaches (e.g., ANCOVA with baseline outcome covariates) to reduce bias stemming from (1) pre-randomization group differences in baseline outcome levels, and (2) missing outcome data.

Analyses to address project aims. To address primary aims, 2-group (VLNC & NNC) x 2 repeated measures (Baseline & Week 4) mixed effects models tested the effect of the randomization group (VLNC vs. NNC) on change in cigarette use and alternative tobacco product (ATP) use from Baseline to Week 4. Similarly, to address secondary aims related to biomarkers of nicotine exposure, 2 x 2 mixed models tested the effects of VLNC vs. NNC cigarettes on change in biomarkers of nicotine exposure (i.e., total nicotine equivalents [TNEs] and cotinine), as well as average number study cigarettes smoked per day relative to average number of usual brand cigarettes smoked per day. In all models, Group (VLNC = 1; NNC = 0) was a between-subjects predictor, and Time (Baseline vs. Week 4) was the repeated measure. Subject was treated as a random effect to account for subject-level differences in baseline outcome levels. The focal tests evaluated the difference in change from Baseline to Week 4 (i.e., Time) in the VLNC vs. NNC condition (i.e., Group), which is a cross-level interactive effect of Time and Group. A p-value < 0.05 was interpreted as meeting statistical significance criteria. For all estimation

parameters, we report effect estimates, standard errors, and 95% confidence intervals in addition to *p* values.

Consistent with the original statistical analysis plan, demographics and baseline data on participant tobacco use history and individual characteristics were summarized using univariate descriptive statistics (e.g., means, standard deviations, frequencies). Data visualization and formal statistical evaluation of means, standard deviations, skewness, and kurtosis, assessed the distributional characteristics of each outcome. Where the assumptions of models implementing a normal (Gaussian) outcome distribution were not met, the statistical approach was revised accordingly. The following analytic procedures were required to meet model assumptions and account for skewed outcome distributions and low frequency events for certain outcomes.

1. Average cigarettes per day was positively skewed, with a non-normal residual distribution in a mixed model analysis specifying a normal (Gaussian) outcome distribution. Therefore, this outcome was recoded to integers and modeled as a count with a negative binomial distribution and log link.
2. Frequencies of alternative product use (ATP) were low, particularly for combustible ATP use. Therefore, for the primary ATP outcome, combustible and non-combustible ATP use were modeled jointly and recoded as a dichotomous outcome (0 = <7 combined days of ATP use; 1 = 7+ combined days of ATP use), with a binary distribution and log link.
3. As secondary outcomes and to evaluate sensitivity, combustible and non-combustible ATP use were also considered independently as separate outcomes. Frequencies of combustible ATP use at Week 4 were too low to allow for model convergence. Frequencies of non-combustible alternative product use (ATP) were bimodal with peaks at 0 (no use) and 7 (daily use). Therefore, this outcome was recoded as a dichotomous outcome (0 = non-daily non-combustible ATP use; 1 = daily non-combustible ATP use) with a binary distribution and log link.
4. The study cigarette secondary outcome was modeled with a normal, Gaussian distribution. Baseline (Time = 0) in this analysis was average number of usual brand cigarettes at baseline,

prior to randomization, and Week 4 (Time = 1) was the average number of study cigarettes at Week 4.

5. Biomarker outcomes (i.e., total nicotine equivalents [TNEs] and cotinine) were modeled with normal, Gaussian distributions.

Missing data and sample size considerations. Missing data were assumed missing-at-random[2] (Whittaker, Pituch, & McDougall, 2014). Under the missing-at-random assumption, missing data for each focal variable is not expected to be related to that variable itself, e.g., a missed outcome assessment was not related to what that value would have been if it were assessed[3] (Enders, 2006). The mixed model approach utilizing SAS software implemented restricted maximum likelihood estimation (continuous outcomes) and pseudolikelihood estimation (count and binary outcomes) to account for missing data. The mixed model reduces bias relative to other common approaches, particularly in cases where Group size is unbalanced due to attrition and/or baseline outcome levels differ between randomization groups.

Original power analyses suggested a target sample size of 120 participants (60 per Group) yielded sufficient power to detect effects of Group (VLNC vc. NNC) on change in cigarettes smoked and alternative products used, as measured via ecological momentary assessment (EMA). Specified population parameter estimates were based on prior literature [4,5] and previous studies[1], and power was calculated to reflect a small to medium effect of VLNC vs. NNC on continuous (i.e., normally distributed) outcomes. For the direct effect of Group on primary outcomes, power to detect a small to medium direct effect (estimate = .15) with 120 participants (60 per Group) was .74 with 95% coverage, where power to detect a slightly larger effect (estimate = .20) was .96, with 95% coverage. Due to unforeseeable effects of the COVID-19 pandemic on data collection and unanticipated, intermittent challenges with the EMA software, the total number of participants was reduced, as was the degree of complete EMA data. As such, TLFB were utilized for an intent-to-treat analysis, thereby retaining all participants who provided Baseline outcome data.

1. Cassidy RN, Tidey JW, Jackson KM, et al. The Impact of Reducing Nicotine Content on Adolescent Cigarette Smoking and Nicotine Exposure: Results from a Randomized Controlled Trial. *Nicotine Tob Res Off J Soc Res Nicotine Tob*. Published online December 9, 2022:ntac279. doi:10.1093/ntr/ntac279

2. Whittaker TA, Pituch KA, McDougall GJ. Latent growth modeling with domain-specific outcomes comprised of mixed response types in intervention studies. *J Consult Clin Psychol*. 2014;82(5). doi:10.1037/a0036664

3. Enders CK. A primer on the use of modern missing-data methods in psychosomatic medicine research. *Psychosom Med*. 2006;68(3):427-436. doi:10.1097/01.psy.0000221275.75056.d8

4. Donny EC, Denlinger RL, Tidey JW, et al. Randomized Trial of Reduced-Nicotine Standards for Cigarettes. *N Engl J Med*. 2015;373(14):1340-1349. doi:10.1056/NEJMsa1502403

5. Hatsukami DK, Luo X, Jensen JA, et al. Effect of Immediate vs Gradual Reduction in Nicotine Content of Cigarettes on Biomarkers of Smoke Exposure: A Randomized Clinical Trial. *JAMA*. 2018;320(9):880-891. doi:10.1001/jama.2018.11473