Impact of E-Cigarette Prevention Messages on Adolescents

Statistical Design and Analytic Plan

NCT#05985538

Version Date: 6/25/2020

Statistical Analysis

For all analyses, *p*-values will be 2-sided, and statistical significance will be set at $p < .05$. Analyses will be conducted using *R* Studio version 2024.12.0+467 with *R* version 4.4.3 (R Project for Statistical Computing).

Missing data.

Analyses will use an intent-to-treat approach that includes all data collected from enrolled participants regardless of trial attrition, with multilevel multiple imputation to handle missingness on covariates and outcomes. We will use the Markov chain Monte Carlo-based joint multilevel imputation package in R, *jomo*, to impute 5 multilevel datasets. We will allow 1000 burn-in iterations and will be thinned by selecting every 1000th draw.[1–3]

Outcome models.

We will report covariate-adjusted, unstandardized parameter estimates with 95% confidence intervals. All parameter estimates and standard errors will be computed using Rubin's formulae for aggregating multiply imputed results.[4]

We will use a latent curve modeling (LCM) approach to model change in susceptibility and secondary outcomes.[5] We will use a structural equation modeling (SEM) framework due to its increased modeling flexibility relative to generalized linear mixed models (GLMM); GLMMs are nested within SEM such that any GLMM can be estimated as an SEM.[6,7] In the LCM shown in Equation 1, the random intercept (b_{0i}) and slopes ($b_{1i} - b_{3i}$) are latent variables that affect repeated measures (y_{ti}) through a factor loading design matrix. Study condition (Tx_i) and covariates affect the random intercept; study condition also affects the random slopes. Point estimates and standard errors for the ITT effects of study condition for waves $t = 2$, 3, and 4 will be calculated using Model Constraints in M*plus* by summing $\gamma_{01} + \gamma_{t1}$. If random slope variances

are not significantly different from zero, we will fix the variance of the random slopes to zero for model parsimony and computation efficiency. Likewise, we will constrain the variances of the time-varying residuals to equality: $\sigma_t^2 = \sigma^2$. See Supplementary Figure A1 for the corresponding path diagram.

Next, we will explore whether the effect of the *Breaking Vape* prevention program trial arm (vs attention control) on the primary trial outcome, susceptibility to vaping, vary as a function of gender (male (ref), female, or other), age (13 – 15, 16 – 17 (ref), 18 – 20), race (1 = White; 0 = other race, including individuals identifying as Asian, Black, multiracial, Native Hawaiian or other Pacific Islander, or another race), ethnicity (1 = Hispanic; 0 = not Hispanic), sexual orientation (1 = heterosexual; 0 = lesbian, gay, bisexual, or other orientation), parental education (-2 = < high school; -1 = high school graduate; 0 = some college; 1 = bachelor's degree; 2 = postgraduate degree), whether the adolescent lived with a tobacco user (1 = yes; 0 = no), and adolescent vaping history (0 = never vaped, 1 = vaped but not in past 30 days, 2 = vaped in past 30 days). For these analyses, we will add main effects and interactions between covariates and study condition as predictors of the random intercept and slopes. We will probe statistically significant interactions to examine the nature of moderation effects. Moderators will be tested one-by-one to avoid collinearity.

Finally, we will test whether study condition effects are moderated by baseline susceptibility to vaping. Taking advantage of flexibility provided by the SEM framework, we will regress the random slopes, $b_{1i} – b_{3i}$, on the random intercept, b_{0i}, and we will use Model Constraints to regress the parameters relating study condition to the random slopes on b_{0i}. See Equation 2 and corresponding Supplementary Figure A2.

Reference

1. Enders CK, Mistler SA, Keller BT. Multilevel multiple imputation: A review and evaluation of joint modeling and chained equations imputation. *Psychol Methods*. 2016;21(2):222-240. doi:10.1037/met0000063

2. Matteo Quartagno, James Carpenter. Package "jomo": Multilevel joint modelling multiple imputation. Published online December 10, 2014:2.7-6. Accessed April 8, 2025. https://cran.r-project.org/web/packages/jomo/jomo.pdf.

3. van Buuren S. *Flexible Imputation of Missing Data*. 2nd ed. Chapman and Hall/CRC; 2018. doi:10.1201/9780429492259

4. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Vol 81. John Wiley & Sons.; 2004. Accessed April 8, 2025. https://www.wiley.com/en-us/Multiple+Imputation+for+Nonresponse+in+Surveys-p-9780471655749

5. Bollen KA, Curran PJ. *Latent Curve Models: A Structural Equation Perspective*. 1st edition. John Wiley & Sons.; 2005.

6. Skrondal A, Rabe-Hesketh S. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman and Hall/CRC; 2004. doi:10.1201/9780203489437

7. Bauer DJ. Estimating multilevel linear models as structural equation models. *J Educ Behav Stat*. 2003;28(2):135-167.