The Effects of a Standardized Research E-Cigarette on the Human Lung: A Clinical Trial with Bronchoscopic Biomarkers

NCT03691350

Statistical Analysis Plan, 09/19/22

SREC SAP

The Specific Aim for this two year project is:

To assess inflammatory changes over 10 weeks (8 weeks experimental condition) for lung and urine biomarkers in smokers in a clinical trial where smokers are randomized to: continued use (n=32), complete switching to the nicotine standardized research electronic cigarette (SREC) (n=32), complete switching to the placebo SREC (n=32), or complete switching to nicotine replacement therapy (NRT) (n=32).

1) We will enroll 128 subjects (32 per group). For power calculations, we will use a conservatively planned 66% retention rate at 5 weeks, knowing that we have a 100% retention for never smokers receiving serial bronchoscopies. Thus, the final sample size will be 21 subjects per group (n=84). Although we expect a higher retention rate, we assume 66% to ensure that we have the minimal numbers required to reach meaningful conclusions. Since follow-up samples would not be obtained from subjects that do not finish the treatment protocol, and these samples (inflammation/omics) are our primary outcomes, we cannot use an 'intent to treat' strategy since dropouts will have missing outcome data for any analyses including follow-up measurements (the primary Aim). In light of this, we will compare dropout rates between arms and with demographic/clinical to ensure that the analyses results are not biased by unbalanced dropouts. We will also test for differences in baseline molecular measurements between subjects that dropout after baseline samples are taken vs. subjects that complete the protocol. While we will make every effort to minimize the missing data for this study, missing data can arise due to various reasons. Violation of missing completely at random (MCAR) will be checked by evaluating whether any covariates are associated with missing data. If so, these covariates will be subsequently included and controlled for in the GL(M)M models. In addition, sensitivity analyses will also be conducted using results from multiple imputation.

2) Descriptive statistics and clustering (e.g., principal components analysis [PCA]), will be performed for all biomarker data measured at the different visits. Baseline data (1st bronchoscopy) will then be compared between the four groups (3 conditions and control) using a one-way analysis of variance (ANOVA). The non-parametric Kruskal-Wallis test (continuous or ordinal variables) or the chi-square test (categorical variables) will be applied when the normality assumption of the data is not met. We will compare controls to complete substitution with the SREC (with and without nicotine) or NRT. Generalized linear models (GLM) will be employed with measurement (cell count, gene expression, etc.) as the dependent variable, a covariable for baseline measure, and a main effect of arm. Other relevant variables that are associated with the biomarker results, e.g., baseline smoking, Fagerstrom Test for Nicotine Dependence (FTND), or gender will also be included in our models as they are potential confounders. We also will assess batch effects and include batch as a co-variable in the model (all samples will be analyzed at the same time, except for cell counts). We plan to use baseline molecular measurement as a covariable rather than

modeling the change between baseline and follow-up because the former is more powerful [1]. Given the large dynamic range of the cell count and the 'omics data, we will initially consider a normal distribution. If the data are determined to deviate from normality, we will log-transform the data or assume a non-normal distribution of residuals (such as Poisson) for the GLM. Urine propylene glycol (PG) will be measured at the 2nd bronchoscopy visit and compared to baseline to verify e-cig use. As an exploratory analysis, urinary PG will be used as the exposure predictor variable assessing the biomarker results as dependent variables. Urine anatabine, urine anabasine and nicotelline will be used to confirm complete substitution or a predictor variable for smoking reduction. The cotinine level will be used as a surrogate for smoking reduction.

The 'omics' (miRNA, mRNA and metabolomics) analysis and visualization of the data will be performed in the R statistical language. Data will be log2-transformed and normalized using either quantile normalization for gene expression (miRNA and mRNA) or TIC/MSTUS normalization for metabolomics. Unsupervised clustering analysis, including PCA and hierarchical clustering will be performed to visualize natural clusters in the dataset and evaluate data quality. Multiple testing corrections using the Benjamini and Hochberg False Discovery Rate (FDR) will be performed. Significantly altered genes and metabolites will be defined as those with FDR corrected p-value<0.05 and fold changes >2 or >1.5, respectively. Identified genes or metabolites will be used to identify relevant networks and pathways using Ingenuity Pathways Analysis. For gene expression, we will further study associations between miRNA-mRNA pairs based on publically available target prediction databases such as TargetScanHuman, miRDB, miRWalk2.0 and others. Global integrative analysis, through correlations, will be performed using Partek software.

Understanding the epigenomic/genomic/proteomic/environment context and regulation of metabolic phenotypes will expand our knowledge of the effects of e-cig on lung biology, and could contribute to finding successful interventions, including accurate predictions of e-cig related adverse effects. Typical omics integration methods include logistic regression or correlations among omics measurements (when measured in the same samples), intersection of gene/protein/metabolite lists (when measured in different samples), or further downstream integration by assessing alterations in pathways. Because our omics measurements will be made in the same sample, we will be able to directly assess relationships between our molecular data through regression or correlation analyses. Of note, typical regression/correlation approaches are not appropriate for low sample numbers, and they do not readily adjust for differences in sample size between phenotype groups or for potential confounders. Another approach, which we propose here, is to directly test whether associations between miRNA or gene or cytokine and metabolite levels are specific to a group assignment. To test this, we will apply a novel linear model approach recently developed by Dr. Mathe for these types of analysis: m = g + t + g:t, where m are metabolite abundances, g are miRNA or gene or cytokine levels, t is treatment group (e.g., SREC vs. continued smoking), and g:t is the interaction between miRNA or gene or cytokine level and treatment group. A statistically significant g:t interaction p-value would indicate that gene (or miRNA or cytokine):metabolite relationships are present in one treatment group and not the

other. Pathway enrichment analysis of associated miRNA, genes, cytokines, and metabolites (resulting from the linear models) will be performed in Ingenuity Pathway Analysis to pinpoint altered pathways and better understand how the epigenomic/genomic/proteomic/environment affects metabolic phenotypes associated with e-cig usage. Furthermore, miRNA, genes, cytokines, and metabolites involved in treatment-specific associations will be used to predict treatment group, in an effort to produce an e-cig inflammation signature. Ten-fold cross-validations of various linear-based and non-linear based machine learning methods will be tried (e.g., Random forests, naïve Bayes, Lasso) and the resulting specificity and sensitivity will be assessed to select the best performing model. Of note, the cross-validation will help ensure that the predictive models are not overfit and could be predictive on new validation data. The contribution of input molecular measurements (gene, cytokines, miRNAs, metabolites) to the model will be assessed to define the minimal set of features that yield the optimal cross-validated models.

For methylation, raw intensity data (idat files) will be normalized using the Subset-quantile Within Array Normalization (SWAN). Any probes with detection P > 0.05 and probes in Y-chromosome to avoid sex-specific methylation bias will be filtered out before further analysis. To identify differentially methylated CpGs, methylation β-values will be converted to M-values by logit-transformation for the statistical modeling purpose. For preliminary identification of patterns in DNA methylation, unsupervised hierarchical clustering among the groups of samples will be performed. The Euclidian distance among the groups of samples will be calculated by the average linkage. In order to assess variance among samples, Principal Component Analysis (PCA) will be done. To characterize the methylation patterns, differentially methylated CpGs will be classified by enhancer, CpG island or its neighbors [2 kb regions upstream and downstream of the CpG islands (shores) or 2 kb regions upstream and downstream of the CpG island shores (shelves)] and functional promoters [within 1500 bp of a transcription start site (TSS) (TSS1500), within 200 bp of a TSS (TSS200), 5’ untranslated regions (5’UTR), first exon (1stExon)], and other regions (body, 3’UTR, or intergenic). Separately, we will also analyze for DNA methylation levels of miRNAs promoter regions to investigate if miRNAs methylation appears to be altered between the groups. When we have a full sample, multiple testing corrections will be performed using the Benjamini and Hochberg False Discovery Rate (FDR) with significantly differential methylation levels defined at corrected P <0.05. Genes corresponding to differentially methylated CpGs will be used for gene networks and biological pathways using the Ingenuity Pathways Analysis. In addition, we will examine differences in methylation for candidate genes that have been previously shown to be associated with lung cancer and/or with smoking using the same methods. To understand the potential biological contribution of differentially methylated CpGs to coding gene or miRNA expression (from on-going pilot study), we will correlate between the differentially methylated CpGs (M-value) from the EPIC methylation array and matched coding gene or miRNA expression (log2 transformed intensity) from the Affymetrix Human Transcriptome array or Affymetrix GeneChip miRNA array using Spearman correlations. To determine if differences in DNA methylation are associated with cytokines levels (as measured for IL-6, IL- 10, TNF-alpha, and TNF-alpha in BAL in the current on-going study), we will select CpGs corresponding to IL-6, IL- 10, and TNF-alpha from the EPIC methylation array (targeted analysis) and correlate between

methylation levels and corresponded cytokines. Correlations controlling for possible confounders (e.g., age) will be explored using multiple linear regression models. We will also explore if the correlation between methylation and cytokine levels may be mediated by gene expression of cytokines using interaction analysis.

3) To compare treatment groups at follow-up for inflammatory markers and 'omics data, we will use generalized linear models (pairwise) with a covariable for baseline measure to account for baseline differences:

Model 1: Followup_Measure = Baseline_Measure + arm + e.

Power is provided for a t-test [2]. Expression/metabolomics/proteomics data tend to follow Gaussian distributions, although there can be violations within each platform. With the large number of variables present in 'omics data, it is intractable to evaluate the appropriateness of the model for each individual variable. Therefore, we use the overall distribution across all variables within a platform to assume a distribution. While individual variables may randomly deviate from the model assumptions (normality, in most cases), if the overall distribution is Gaussian we assume that each of the features was sampled from the overall Gaussian. However, to check if assumptions are causing spurious positive results, we will test each associated feature, post hoc, for departures from the model assumptions. Cell counts were also considered Gaussian for these power calculations. While counts are inherently discrete, because the expected cell counts are very large (>2E+07), Gaussian should be a good approximation for this distribution. The parameters included in the power calculation are the CVs, the p-values, and the effect sizes. The CVs used were estimated from preliminary data (and similar data from other studies). The effect sizes assumed are reasonable (smaller) compared to the effect sizes we see in the preliminary data. The p-values used were stringent to adjust for multiple comparisons. Because, especially with 'omics data, we do not know a priori the number of tests, we use significance thresholds that correspond to an expected number of false discoveries per number of tests to control for multiple comparisons without defining the number of tests. For the primary analyses, we accept that one expected false discovery per 10,000 tests is a reasonable error rate. In practice, we will use FDR to determine the expected fraction of false discoveries for assuming a particular significance threshold. For a final sample size of 21 in the control group and 21 in each use condition, for a two-sided analysis, we have 80% power to detect ≥1.5-fold difference in cell counts/cytokine levels between each condition and controls for a two-sample t-test. For gene expression/metabolomics data, with the same considerations except a CV=0.5 (approximate from many gene expression data sets with which we have worked), we have 80% power to detect ≥2.1-fold differences.

4) For mtDNA features, Model 1 is also used. CN is expected to be Gaussian, while mutation rate is expected to be low counts; therefore, a t-test was powered for CN and Poisson [3] for mutation rate. Because there will only be several comparisons here, the significance threshold used for power corresponds to one expected false discovery per 100 tests. We believe that this threshold is a good balance between potential Type I and Type II errors, given the number of tests. The coefficient of variation was calculated from results in the

literature, as cited in the main body of the protocol. For CN, with 21 samples per group, considering a significance threshold of 0.01, a two-sided, two-sample t-test achieves 80% power to detect an 1.4 fold difference between arms considering a coefficient of variation of 0.32 [4]. For mutation, Poisson regression achieves 80% power to detect a mutation rate ratio of 2.7 between arms considering a significance threshold of 0.01 and 21 samples per group [5].

5) For associations between mtDNA and other 'omics features, we assume the regression model [6]:

$$\text{Followup\_Expression} = \text{Baseline\_expression} + \text{arm} + \Delta\text{mtDNA feature} + e,$$

Where ΔmtDNA =Followup_ mtDNA – Baseline_ mtDNA. We consider this model across all arms. The effect of arm is included as a covariable to ensure that associations are not simply due to both features being associated with arm; rather, we are looking for a more direct relationship. Again, we consider a significance threshold corresponding to one expected false discovery per 10,000 tests, as described above. To generalize, we assume unit standard deviations of the dependent and independent variables (each feature will have a different standard deviation, so results are calculated in terms of standard deviations). With 84 samples, considering a significance threshold of p=0.0001 and unit standard deviations for dependent variable residuals and ΔmtDNA features, linear regression achieves 80% power to detect a slope of 0.48.

6) For comparing mtDNA features between nasal and bronchial samples, mixed Poisson or mixed Gaussian will be employed with a random effect for subject, and a main effect for location (bronch vs. nasal):

$$\text{Feature} = \text{sample\&rand} + \text{location} + e.$$

Poisson regression achieves at least 80% power to detect a mutation rate ratio of 2.1 between the sample origins considering a significance threshold of 0.01 and 84 paired samples. For mtDNA CN, a two-sided paired t-test achieves 80% power to detect a mean of paired differences of 0.4 standard deviations of the paired differences considering 84 paired samples and a significance threshold of 0.01. Correlations (CN) and Cohen's kappa (mutational burden) will also be calculated between the two tissues for each of the assays.

7) As a secondary aim, we will test if treatment is associated with differences in mtDNA genetic features between tissues. Generalized linear mixed models will be employed with measurement (mtDNA genetic features) as the dependent variable, a random effect for patient, covariables for baseline measure and sample origin, a main effect of arm, and an interaction effect of origin* arm:

$$\text{Feature} = \text{sample\&rand} + \text{baseline} + \text{origin} + \text{arm} + \text{arm*origin} + e.$$

This analysis will investigate whether treatment is associated with differences between tissues. The interaction term will be comprised of four groups of 21 for each pair of arms.

To conservatively estimate the power to detect this interaction effect, the power will be calculated while adjusting the significance threshold for the number of pairwise tests (n=6; p=0.01/6=0.0017). A two-sided, two-sample t-test achieves 80% power to detect an 1.5 fold difference between any two groups considering a significance threshold of 0.0017, 21 samples per group, and a coefficient of variation of 0.32 [4].

8) As an additional exploratory analysis, we will study dose-response associations of cigarette smoking with biomarkers at baseline bronchial and nasal samples separately. Generalized linear models (again, Poisson or Gaussian) will be fit with biomarkers as the dependent variable and covariables (e.g., cigarettes per day [cigs/day] and urinary cotinine levels, separately), adjusting for confounders, such as sex and age. THC use will also be examined.

References

1. Zhang, S., et al., *Empirical comparison of four baseline covariate adjustment methods in analysis of continuous outcomes in randomized controlled trials.* Clin Epidemiol, 2014. **6**: p. 227-35.
2. Julious, S.A., *Sample sizes for clinical trials with normal data.* Stat Med, 2004. **23**(12): p. 1921-86.
3. Signorini, D.F., *Sample size for Poisson regression.* Biometrika, 1991. **78**(2): p. 446-450.
4. Lee, J.Y., et al., *Mitochondrial DNA copy number in peripheral blood is independently associated with visceral fat accumulation in healthy young adults.* Int J Endocrinol, 2014. **2014**: p. 586017.
5. Tan, D., et al., *Associations between cigarette smoking and mitochondrial DNA abnormalities in buccal cells.* Carcinogenesis, 2008. **29**(6): p. 1170-7.
6. NCSS.com, *Linear Regression* PASS Sample Size Software. **Chapter 855**: p. 855-1-855-8.