**Research Strategy**

**Significance**

The concurrent use of multiple tobacco and/or e-cigarette products (MTEPs) poses a considerable barrier to cessation and increases both toxicant exposure and disease risk (1-5). Younger age (<18) has been identified as a significant risk factor for MTEP use, with 10% of US adolescents (ages 9-17) using MTEPs according to the 2021 National Youth Tobacco Survey (NYTS) compared with 5%–6% of younger adults (18–44 years) and 1%–2% of older adults (≥ 45 years) according to the 2020 National Health Interview Survey (6, 7). Moreover, among current tobacco users, approximately 28% of high school students and 33% of middle school students are MTEP users, compared with 18% of adults (6, 7). From 2011-2020, MTEP use and exclusive e-cigarette use among adolescents increased significantly, while exclusive use of other tobacco products decreased (8). Adolescent MTEP users have greater symptoms of nicotine dependence than single product users (10-12), making cessation more difficult. A significant number of adolescents who have never used tobacco or e-cigarette products are susceptible to MTEP use (i.e., lack a firm commitment to abstain) (9). MTEP use is significantly associated with tobacco-related morbidity and mortality, and makes the regulation of tobacco products more complex without addressing the risk factors specific to MTEP use.

MTEP use is difficult to study due to highly complex MTEP use patterns. The NYTS, a nationally representative survey of middle and high school students in the US, assesses the use of 11 different nicotine and tobacco products currently available in the US marketplace: cigarettes, e-cigarettes, cigars (including large cigars, cigarillos, and little, filtered cigars), chewing tobacco/snuff/dip, hookah/waterpipe, roll-your-own cigarettes, pipe, snus, dissolvables, bidis, heated tobacco products, and nicotine pouches. MTEP use includes any combination of two or more of these 11 products. Examining every possible combination is complex and often not feasible due to small sample sizes associated with specific combinations, which can prevent further analysis according to sociodemographic factors (e.g., age groups, racial/ethnic minoritized groups). Secondary analyses of nationally representative data sets often simplify MTEP use into dual-use (two products) and polyuse (three or more products) groups (12) or conduct latent class analyses to identify patterns (9, 13). However, these approaches can mask nuances in MTEP use, including which products are used and the frequency of product use. However, the health risks associated with the use of multiple non-combustible products (e.g., e-cigarettes, smokeless tobacco) are likely different from the risks associated with the use of multiple combustible products (e.g., cigarettes, cigars).

Significant differences in the risk factors associated with different MTEP use patterns can inform the development of programs and regulations intended to reduce adolescent tobacco use. In general, risk factors associated with adolescent MTEP use include male sex; non-Hispanic white and Hispanic racial/ethnic identification; receptivity to tobacco marketing; perceived peer prevalence of tobacco use; and tobacco harm perceptions. However, risk factors likely differ across age groups (middle school vs. high school), sex, and patterns of MTEP use (6, 10, 12, 14, 15). Approaches that oversimplify MTEP use can mask critical differences associated with the use of specific products or the frequency of use. In one study of 2017 NYTS data (14), the dual-use pattern consisting of e-cigarettes used together with any other products was more prevalent among non-Hispanic white adolescents than among non-Hispanic African American adolescents, but no difference in the dual-use pattern consisting of cigarettes used together with other combustible products was observed between demographic groups. Studies examining nationally representative data often combine several racial/ethnic groups into a single “non-Hispanic Other” category due to the limitations of common statistical methods when handling small sample sizes; however, research suggests substantial variations exist in MTEP use patterns between racial/ethnic subgroups (16). Therefore, research examining the risk factors associated with MTEP use requires the development of new statistical methods that remain robust when analyzing small sample sizes while also allowing for the examination of complex MTEP use patterns, including the use of specific product combinations and different use frequencies.

Canonical correlation analysis (CCA) has the potential to overcome statistical challenges to MTEP use analysis because it allows for the examination of associations between multiple risk factors (independent variables) and the frequency of specific tobacco product use (dependent variable) among MTEP users. Although in general in CCA it is customary not to explicitly referred to one set of variables as independent and the other as the dependent set of variables, in this project their cause-effect relationship is evident and hence we considered fair and illuminating to use these terms. CCA can be considered a generalization of linear regression with multiple outcomes that does not require the separate analysis of each exposure–outcome relationship, allowing the simultaneous analysis of multiple outcomes for the same tobacco-use exposures, which is one of its main advantages over other techniques that are commonly applied to cross-sectional analyses of large numbers of variables (e.g., generalized additive mixed models). CCA provides a holistic view of the complex relationships that exist among all variables at the same time.

Most existing data sets examining tobacco use and health were collected using complex survey designs rather than simple random samples. The accurate analysis of these data requires the development of statistical methods able to account for complications introduced by complex study designs, and newly developed methods must be easily integrated into statistical programming software to allow for use by researchers.

Although the most common methods for summarizing and analyzing complex survey design data, such as linear regression, are readily available, tobacco and nicotine researchers currently lack access to all of the statistical tools they might require testing their hypotheses. For example, the application of CCA to tobacco research is limited, even though the mathematical characteristics make this method especially well-suited for exploring relationships among tobacco product marketing, packaging, and labeling of new tobacco products and dual-use patterns among adolescents and both younger and older adults. The multidimensional nature of these relationships requires the application of complex statistical methods, more over some of these advanced statistical methods have not been adapted to accommodate complex survey data.

The proposed research will ***(1) develop and test a new statistical method for applying CCA to complex survey designs, which can be applied to future analyses of complex survey data sets; and (2) contribute to the current understanding of the risk factors (including marketing influences) associated with MTEP use patterns among adolescents, which can guide adolescent-targeted educational efforts.*** Successful completion of this proposal will inform the development of tobacco and e-cigarette marketing regulations and educational campaigns designed to discourage MTEP use among adolescents. Increasing the rate of tobacco-use cessation among adolescents is likely to significantly decrease their future morbidity and mortality risks. Given the versatility of CCA to analyze multidimensional problems, the statistical packages that are developed will be beneficial for future researchers using complex survey data sets to study the risk factors associated with various complex behaviors.

***This research will address the following Center for Tobacco Product priorities: (a) examining the impact of marketing exposure (including digital) on youth tobacco use and (b) examining how the understanding of risk relates to tobacco product use.***

RAUL – pretty sure this is where you are supposed to put preliminary studies. I’ve added mine here, but I believe this is where you should add the preliminary analysis you’ve run:

***Preliminary Studies.*** MTEP use in young adults: Dr. Mead-Morse has conducted several studies examining the dual use of cigarettes and cigars among young adults (18-34). In one study using ecological momentary assessments in a population of African American young adult dual users (17-19), she found that cigar use, especially flavored cigar use, was highly prevalent, with gender-specific differences in preferred flavors. Cigars were also smoked additively with cigarettes rather than as cigarette substitutes. At the intrapersonal level, risk factors for cigar smoking included feeling relaxed and smoking with others, whereas risk factors for cigarette smoking included feeling stressed and being alone. In a recently completed study of young adult dual users (20), cigars were found to be similarly addictive as cigarettes, and risk factors for increased cigar addiction included flavored cigar use, socioeconomic status, and gender, whereas more frequent cigar use was not a risk factor (results unpublished). In addition, cigar use in place of cigarette use when cigarettes were unavailable was relatively limited. ***Overall, these results indicate that: (1) cigars are addictive and smoked for reasons in addition to just cigarette substitution; (2) flavors are popular and associated with increased addiction; and (3) affect and peer use may be significant risk factors for dual use.***

**Innovation**

When considering limited sample sizes, such as the population of adolescent dual MTEP users, many existing data analysis methods designed for complex survey data are not able to correctly identify statistically significant differences. Although CCA is robust for smaller sample sizes as long as the number of observations exceeds the number of variables, classical CCA is not designed to accommodate the complex survey designs used to generate most existing tobacco and nicotine use data sets, including those obtained from the Population Assessment of Tobacco and Health Study, the Tobacco Use Supplement to the Current Population Survey, the Behavior Risk Factor Surveillance System, the National Survey on Drug Use and Health, and the National Health and Nutrition Examination Survey. Various essential complex survey factors, such as replicate weights, strata, clusters, and survey weights, must be considered when analyzing these data sets. We propose to extend existing CCA procedures to accommodate these types of information.

We propose to develop software necessary to implement CCA for complex survey design and apply it to priority research questions that fall within the scope of this FOA. We will also promote the theory behind this method. Our software will be designed to be easily accessible and freely available for implementation in the most used statistical programs applied to tobacco and nicotine use research (SAS, Stata, and R).

**Approach**

The overall objective of this proposal is to apply our novel Complex Survey Design CCA to the nationally representative 2021 NYTS data set to determine the risk factors associated with MTEP use patterns among adolescents in the US. The NYTS is a school-based survey of middle school (grades 6–8) and high school (grades 9–12) students that uses a three-stage, cluster-sampling design to generate a cross-sectional, nationally representative sample of tobacco use data from all 50 states and the District of Columbia. Data are adjusted for nonresponse and weighted to produce national prevalence estimates. The 2021 survey was conducted online from January 18, 2021, to May 21, 2021, and the response rate was 81.2% (6) (n=20,413).

NYTS Data

***Study sample.*** Adolescents who reported at least 1 day of use for each of two or more tobacco and/or e-cigarette products during the past 30 days. (Asian Sample size=1150, Pacific Islanders and Hawaiians Sample size=439)

***Dependent variables****.* During the past 30 days, the number of days they used e-cigarettes, cigarettes, cigars, chewing tobacco/snuff/dip, hookah/waterpipe roll-your-own cigarettes, pipe, snus, dissolvables, heated tobacco products, nicotine pouch, or any other tobacco product.

***Independent variables.*** Tobacco marketing: how often do you see e-cigarette internet ads, print ads, store ads, TV ads, or use in movies; how often do you see e-cigarette posts or content on social media, post self or others using e-cigarettes, or liked e-cigarette posts; number of social media sites where exposed to e-cigarette content; how often do you see cigarette and other tobacco internet ads, print ads, store ads, TV ads, or use in movies. Anti-tobacco messaging: how often did you see a warning label on an e-cigarette/cigarette/cigar/chewing tobacco/hookah package number of campaigns seen. Perceived harm: how much do you think people harm themselves if they smoke/use cigarettes/cigars/chewing tobacco/e-cigarettes/hookah some days but not every day; how strongly do you agree that all tobacco products are dangerous; perceived harm from secondhand smoke. Perceived addiction relative to cigarettes: perception of how addictive cigars/chewing tobacco/e-cigarettes/hookah are relative to cigarettes. Social norms: percentage of peers who smoke cigarettes/vape e-cigarettes; perceived friends’ acceptability of e-cigarettes; number of places at school you saw anyone using e-cigarettes; number of products used by anyone who lives with them.

Although some of these variables are ordinal, we used them as numeric variables for the purposes of the preliminary experiments presented in this proposal, which has been shown to be a valid approach. (Robitzsch, 2020).

Classic Canonical Correlation Analysis

CCA is a popular exploratory statistical method that allows for the analysis of relationships between two sets of variables, offering several advantages over other statistical techniques, including the ability to limit the probability of committing Type I errors when more than two variables are being examined ([Hair](#_bookmark45) et al., [2009](#_bookmark45)). CCA reflects the reality of relationships among factors in a manner consistent with the reality of the problem and has been applied to diverse fields, ranging from ecological studies ([Gittins,](#_bookmark41) [1985](#_bookmark41)) to human geography ([Clark,](#_bookmark35) [1975](#_bookmark35)). In the area of tobacco research, it has been used to predicting attempts and sustained cessation of smoking after the introduction of workplace smoking bans (Borland et al., 1991), change tobacco use among veterans (Morris et al., 2018), examine the effect of tobacco smoke uptake in urine, saliva and hair (Sastre Toraño and van Kan, 2003), links between links between problematic internet behavior with a propensity to use tobacco and illicit drugs (De Leo & Wulfert, 2013) and assessing gender differences in college cigarette smoking (Page and Gold, 1983).

Recently, the use of CCA has been challenged by data sets collected using complex survey design methods rather than through simple random samples of the objective population. To overcome this issue, we will employ our recently developed, innovative technique.

CCA studies the extent and nature of the correlation between two sets of variables, *X* = {*X*1, *X*2,..., *X*p} and *Y* = {*Y*1, *Y*2,..., *Y*q} (Hotelling, 1936). As in Clark (1975), without loss of generality, it is assumed that *p* ≥ *q* and that each and every variable in *X* and *Y* has been standardized to have a mean zero and a variance equal to one. CCA explores sample correlations between *X* and *Y* observed on the same experimental units by analyzing the coefficients, *A*1 = (a1, a2,..., a*p* )T and *B*1 = (b1, b2,..., b*q*)T, which maximize the correlation between linear combinations of the variables *X* and *Y* subject to *Variance* [*XA*1 ] = *Variance*[*YB*1 ] = 1.

The “classic” canonical correlation can be defined as (Fukumizu, 2012):

(1)

subject to , as seen in (Hardoon, 2011). These linear combinations *U*1 = *XA*1 and *V*1 = *YB*1 are called the first canonical variates. Their correlation, *ρ*1, is the first canonical correlation, and the coefficients *A*1 and *B*1 are the first canonical coefficients.

The first canonical coefficients in each set will help to identify variables from *X* and *Y* that predict each other well. Variables with a strong positive relationship will have coefficients with large magnitudes and similar signs (positive or negative), whereas negative relationships will have coefficients with opposing signs. The coefficient magnitudes are commonly applied to determine which variables are the most relevant. The process can be duplicated to identify the coefficients, *A*2 and *B*2, that maximize the correlation, *ρ*2, subject to *Variance* [*XA*2 ]= *Variance* [*YB*2 ] = 1, with the additional constraint that the new canonical variates, *U*2 = *XA*2 and *V*2 = *YB*2, are uncorrelated with the first pair of canonical variates. This process can be repeated *q* times. These secondary canonical correlations are important because not all the significant relations that exist among variables can be expressed in the first canonical correlation.

Because researchers base their conclusions regarding the variables included in the CCA on canonical factor loadings and canonical coefficients, limiting further exploration to those elements of the canonical structure that are associated with statistically significant canonical correlations, tools able to accurately estimate these values are necessary. For example, Travis and Lagrosen (2014) limited their conclusion to the information inferred from the first set of canonical coefficients because only the first canonical correlation had a significant *p*-value. In Stowe et al. (1980), three canonical correlations have statistically significant *p*-values, and all three sets of canonical coefficients are presented.

For the sample cross-correlation matrix ,canonical correlations can be calculated by finding the descending-ordered square roots of the matrix eigenvalues .The canonical coefficients for *Y* are the corresponding *q* eigenvectors, and the elements of eigenvector *i* are the canonical coefficients *Bi*. The correlations between the canonical variates and the original variables are called canonical factor loadings, which can be used to analyze the intra- and inter-set relationships. They offer some information that is complementary to canonical coefficients and, hence, is preferred by some researchers (Meredith, 1964).

Weighted CCA

Broadly speaking, when data sets have been collected using complex survey design methods, two main factors are used to account for the fact that the sample is not a simple random sample:

* Survey weights are used to create a synthetic sample for which the distribution of covariates is similar to the population under study, leading to correct population parameter estimates (Lewis, 2017).
* Repetition weights, clusters, and strata are complex survey design elements used during the calculation of standard errors to calculate accurate confidence intervals and *p*-values for any hypothesis that is tested.

Standard statistical methods that ignore structures such as clustering or stratification, can give seriously misleading results (Holt et al., 1980). Analyzing a stratified sample as if it were a simple random sample tends to overestimate the standard errors, whereas analyzing a clustered sample as if it were a simple random sample tends to underestimate the standard errors, and analyzing the data without taking into consideration the survey weight will lead to incorrect point estimates (Lumley, 2004). Examples of how intracluster correlations can severely affect the effective sample size can be found in (Killip et al., 2004).

From (Winship, 1994), we know that the formulas for weighted quantities of covariance and variance for two variables, *X1* and *X2*, are:

and (2)

, (3)

where *W* is the vector that stores the survey weight for each of the *n* subjects in the sample.

If we choose *Aj* and *Bj*, such that , then substituting (4) and (5) into (1) reduces the weighted canonical correlation to

. (4)

It can be proven that the *q* canonical coefficients are the *q* eigenvectors of:

. (5)

Similarly, the canonical coefficients B, can be found similarly.

Moreover, the *q* eigenvalues λ are the weighted canonical correlations; hence, parameters *A* and *B* are not needed to calculate the weighted canonical variates to find the weighted canonical correlations, as this calculation will derive the same values.

Modern statistical software, including SAS’s *PROC CANCORR* (SAS Institute Inc. 2008), Stata’s *canon* command, and R’s *cca* function, apply these formulas and are able to consider survey weights in their calculations of weighted canonical correlations and corresponding *p*-values. However, these programs are unable to account for other factors, such as clusters and strata. When applied to weighted data, Stata’s *canon* command is limited to the calculation of the statistical significance of the first canonical correlation. As mentioned previously, ignoring these complex survey design elements will lead to incorrect estimates of the standard errors and the incorrect estimation of *p*-values for weighted canonical correlations.

Complex Survey Design CCA

The proposed Complex Survey Design CCA method (**Figure 1**) is based on three well-known ideas related to data collected using complex survey designs: (1) the survey weights are sufficient to calculate correct parameter estimates (i.e., canonical coefficients); (2) the linear combination of variables in a complex survey data set follows the same structure as the original variables; and (3) the *p*-value for a simple linear regression coefficient and a correlation using the same two variables is the same.

Existing CCA methods that consider survey weights can be used to identify canonical coefficients, which, in turn, can be used to calculate canonical variates. Once canonical variates have been identified, *p*-values for canonical correlations can be estimated using a set of complex survey linear regression procedures. Linear regression survey procedures in modern statistical software can include not only survey weights but also information regarding repetition weights, clusters, and strata.

1. Select sets of variables *X* and *Y* from existing datasets of tobacco use and health.
2. Include survey weights in existing CCA procedure to find accurate estimates of the weighted canonical coefficients and correlations.
3. Use weighted canonical coefficients to calculate the *q* weighted canonical variates *Uj* and *Vj*, where 1≤j≤ *q.*
4. For each pair of *Uj* and *Vj* weighted canonical variates perform a simple linear regression analysis that includes all complex survey design elements, e.g., survey weights, cluster, and strata, to find the correct *p*-values for their corresponding *j* canonical correlation, where 1≤j≤ *q.*

**Figure 1**. Proposed Complex Survey Design Canonical Correlation Algorithm

For example, to implement Complex Survey Design CCA in SAS, we would first use PROC CANCOR, which allows the use of the WEIGHT statement to find appropriate point estimates for the canonical coefficients (as cluster and strata information are only essential to calculate standard errors and *p*-values). These coefficients can be used to find canonical variables for inclusion into a simple PROC SURVEYREG, with WEIGHT, CLUSTER, and STRATA statements, to find *p*-values.

A byproduct of the proposed algorithm is that we will obtain *p*-values for each individual *j* canonical correlation, i.e., while current software tests the hypothesis . A further advantage of our enhanced CCA algorithm is that it will allow the statistical significance of each canonical correlation (and the conclusions that might be drawn from the associated canonical structure) to be evaluated separately. Because the effective sample size might change for each canonical correlation depending on the canonical coefficients, a larger correlation magnitude might not imply a smaller *p*-value.

Preliminary Results for the 2021 National Youth Tobacco Survey

The proposed Complex Survey CCA method was applied to the 2021 NYTS data set to evaluate the relationships of variables associated with polytobacco use with measures of exposure to e-cigarette internet marketing and with variables related to perceived addiction relative to cigarettes. To demonstrate the usefulness of our proposed method for data sets with small sample sizes, we only examined two subsets: (1) Asian Americans and (2) Pacific Islanders and Hawaiians.

The canonical coefficients are shown in Table 1. These values can be obtained by considering only survey weights. The importance of our proposed Complex Survey CCA method is shown in Table 2, which compares the *p*-values obtained using existing statistical software that considers only survey weights with those obtained using our newly developed method, which incorporates the remaining complex survey design elements, resulting in different conclusions regarding the significance of the first canonical correlation and the relationships that can be inferred from the examination of canonical coefficients. The weights in red are the dominant weights and hence the focus of our conclusions.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1**. Canonical coefficients (dominant coefficients are in red font). | | | |
| **2021 NYTS Variables** | | Canonical coefficients | |
| Asian Americans (n = 1150) | Pacific Islanders and Hawaiians  (n = 439) |
| ***X*: Polytobacco Use Variables** | | | |
| QN9 | During the past 30 days, on how many days did you use e-cigarettes? | −0.13 | 0.32 |
| QN38 | During the past 30 days, on how many days did you smoke cigarettes? | 3.03 | 0.50 |
| QN40 | During the past 30 days, on the days you smoked, about how many cigarettes did you smoke per day? | −4.49 | −1.25 |
| QN53 | During the past 30 days, on how many days did you smoke cigars, cigarillos, or little cigars? | 3.36 | −0.02 |
| QN54 | During the past 30 days, on the days that you smoked, about how many cigars, cigarillos, or little cigars did you smoke per day? | −19.39 | −1.51 |
| QN64 | During the past 30 days, on how many days did you use chewing tobacco, snuff, or dip? | −0.12 | −0.22 |
| QN69 | During the past 30 days, on how many days did you smoke tobacco in a hookah or waterpipe? | 0.31 | −3.75 |
| QN74 | During the past 30 days, on how many days did you smoke roll-your-own cigarettes? | 5.52 | 0.35 |
| QN76 | During the past 30 days, on how many days did you smoke pipes filled with tobacco? | -19.07 | −0.61 |
| QN78 | During the past 30 days, on how many days did you use snus? | 1.41 | 4.07 |
| QN80 | During the past 30 days, on how many days did you use dissolvable tobacco products? | 26.78 | −0.14 |
| QN82 | During the past 30 days, on how many days did you smoke bidis? | -15.55 | −1.22 |
| QN85 | During the past 30 days, on how many days did you use a “heated tobacco product”? | 0.14 | 0.66 |
| QN88 | During the past 30 days, on how many days did you use a “nicotine pouch”? | –1.14 | 0.91 |
| QN89 | During the past 30 days, on how many days did you use any tobacco product(s)? | 0.12 | −0.34 |
| ***Y*: Perceived Addiction Relative to Cigarettes Variables** | | | |
| QN113 | Do you believe that cigars, cigarillos, or little cigars are more addictive than cigarettes? | – | −0.57 |
| QN115 | Do you believe that chewing tobacco, snuff, dip, snus, or dissolvable tobacco products are more addictive than cigarettes? | – | −0.59 |
| QN117 | Do you believe that e-cigarettes are more addictive than cigarettes? | – | 0.60 |
| QN119 | Do you believe that smoking tobacco in a hookah or waterpipe is more addictive than cigarettes? | – | 0.65 |
| ***Y*: E-cigarette Internet Marketing Variables** | | | |
| QN128 | When you are using the Internet, how often do you see ads or promotions for e-cigarettes? | 0.79 | – |
| QN129 | When you read newspapers or magazines, how often do you see ads or promotions for e-cigarettes? | −0.10 | – |
| QN130 | When you go to a convenience store, supermarket, gas station, kiosk/storefront, or shopping center, how often do you see ads or promotions for e-cigarettes? | 0.25 | – |
| QN131 | When you watch TV or streaming services (such as Netflix, Hulu, or Amazon Prime), or go to the movies, how often do you see ads or promotions for e-cigarettes? | −0.73 | – |
| QN132 | When you watch TV or streaming services (such as Netflix, Hulu, or Amazon Prime), or go to the movies, how often do you see people or characters using e-cigarettes? | –0.44 | – |
| QN134 | When you use social media, how often do you see posts or content (pictures, videos, or text) related to e-cigarettes? | 0.03 | – |
| Canonical coefficients for Asian Americans were assessed using X: Polytobacco Use Variables vs. Y: E-Cigarette Marketing Variables. Canonical coefficients for Pacific Islanders and Hawaiians were assessed using X: Polytobacco Use Variables vs. Y: Perceived Addiction Relative to Cigarettes Variables. Red indicates significant correlations. | | | |

Based on these canonical coefficients, an increase in exposure to e-cigarette ads on the internet and a decrease in exposure to e-cigarette ads on streaming services is associated with decreases in the number of cigars smoked, decreased pipe and bidis use, and increased dissolvables use among Asian Americans. Among Pacific Islanders and Hawaiians, the perception that e-cigarettes and hookah/waterpipe use are more addictive than cigarette use but the use of other products is not more addictive than cigarettes is associated with reduced hookah/waterpipe use and increased snus use.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 2**: First canonical correlation for preliminary experiments between Polytobacco Use and a second set of variables | | | |
| **Analysis** | **Canonical Correlation Magnitude** | **Complex Survey Design *p*-value** | **Weighted p-value** |
| Polytobacco Use vs. E-Cig Marketing in Asian Americans | 0.19 | 0.14 | 0.00 |
| Polytobacco Use vs. Perceived Addiction in Pacific Islanders/Hawaiians | 0.37 | 0.05 | 0.00 |

The importance of our proposed Complex Survey CCA method is shown in Table 2, which compares the *p*-values obtained using existing statistical software that considers only survey weights with those obtained using our newly developed method, which incorporates the remaining complex survey design elements, resulting in different conclusions regarding the significance of the first canonical correlation and the relationships that can be inferred from the examination of canonical coefficients. The proposed Complex Survey Design CCA allows us to recognize that neither of the conclusions drawn from Table 1 should be not considered valid because they are associated with canonical correlations that are statistically insignificant, as shown in Table 2. Current statistical methods would result in erroneous conclusions that these canonical correlations are significant and that examining these relationships is worthwhile.

Preliminary Results relevance for Specific Aim 1

The fact that the thirds canonical correlation is statistically significant…. TBD

Preliminary Results relevance for Specific Aim 2

It is remarkable that in the case of….TBD

Application of Complex Survey Design CCA

Applying our new Complex Survey Design CCA would generate significant new insights regarding the complex web of relationships that exist among the variables assessed by the NYTS. Successful completion of this proposal could contribute to the development of more comprehensive and holistic treatment and prevention guidelines for adolescents who engage in MTEP use.

**Proposed Timeline**. We request funding for a 2-year period, from September 1, 2023, to August 31, 2025. The first year will be primarily focused on transforming Dr. Cruz-Cano’s loose Stata programs into a command that will be submitted to the SSC Archive, allowing it to be easily installed by anybody with a Stata license. The application of this Stata command to the data set described in Aim 2 will proceed immediately following the acceptance of the command into the SSC Archive, and the results will be submitted to *The* *Stata Journal*.

Preliminary results will be presented at the 2024 Society for Research on Nicotine & Tobacco to receive feedback from our fellow tobacco and nicotine researchers attending this conference. We will incorporate feedback into the analysis to improve quality and add value to the tobacco and nicotine research community.

The first half of the second year will be spent developing an R package and an SAS Macro to implement the Complex Survey Design CCA, in addition to preparing a manuscript for submission. A preliminary version of these results will be presented at the 2025 Society for Research on Nicotine & Tobacco and the 2024 SAS Users Group International. By the end of the second year, Dr. Cruz-Cano will have prepared and submitted two manuscripts for publication. These analyses will also form the basis for a larger grant application.

At all times, Dr. Mead-Morse will interpret the results obtained from the statistical analysis and help drawing the appropriate conclusions. She will also play an essential part in the writing of all results in journal and conference papers and future proposals. The graduate student included in our budget will help to apply Dr. Cruz-Cano’s statistical software to the to the tobacco data set and be co-author in the publications listed above, enhancing his/her graduate student experience and careers prospects.

**Table 3:** Timeline of the proposed activities



1. Teo KK, Ounpuu S, Hawken S, Pandey MR, Valentin V, Hunt D, Diaz R, Rashed W, Freeman R, Jiang L, Zhang X, Yusuf S. Tobacco use and risk of myocardial infarction in 52 countries in the interheart study: A case-control study. Lancet. 2006;368(9536):647-58. Epub 2006/08/22. doi: 10.1016/s0140-6736(06)69249-0. PubMed PMID: 16920470.

2. Chao A, Thun MJ, Henley SJ, Jacobs EJ, McCullough ML, Calle EE. Cigarette smoking, use of other tobacco products and stomach cancer mortality in us adults: The cancer prevention study ii. International Journal of Cancer. 2002;101(4):380-9. Epub 2002/09/05. doi: 10.1002/ijc.10614. PubMed PMID: 12209964.

3. Wetter DW, McClure JB, de Moor C, Cofta-Gunn L, Cummings S, Cinciripini PM, Gritz ER. Concomitant use of cigarettes and smokeless tobacco: Prevalence, correlates, and predictors of tobacco cessation. Preventive medicine. 2002;34(6):638-48. Epub 2002/06/08. doi: 10.1006/pmed.2002.1032. PubMed PMID: 12052025.

4. Post A, Gilljam H, Rosendahl I, Bremberg S, Galanti MR. Symptoms of nicotine dependence in a cohort of swedish youths: A comparison between smokers, smokeless tobacco users and dual tobacco users. Addiction (Abingdon, England). 2010;105(4):740-6. Epub 2010/02/13. doi: 10.1111/j.1360-0443.2009.02852.x. PubMed PMID: 20148785.

5. Messer K, Vijayaraghavan M, White MM, Shi Y, Chang C, Conway KP, Hartman A, Schroeder MJ, Compton WM, Pierce JP. Cigarette smoking cessation attempts among current us smokers who also use smokeless tobacco. Addictive behaviors. 2015;51:113-9. Epub 2015/08/09. doi: 10.1016/j.addbeh.2015.06.045. PubMed PMID: 26253939; PMCID: PMC4558255.

6. Gentzke AS, Wang TW, Cornelius M, Park-Lee E, Ren C, Sawdey MD, Cullen KA, Loretan C, Jamal A, Homa DM. Tobacco product use and associated factors among middle and high school students - national youth tobacco survey, united states, 2021. MMWR Surveill Summ. 2022;71(5):1-29. doi: 10.15585/mmwr.ss7105a1. PubMed PMID: 35271557.

7. Cornelius ME, Loretan CG, Wang TW, Jamal A, Homa DM. Tobacco product use among adults - united states, 2020. MMWR Morb Mortal Wkly Rep. 2022;71(11):397-405. Epub 2022/03/18. doi: 10.15585/mmwr.mm7111a1. PubMed PMID: 35298455; PMCID: PMC8942309 Journal Editors form for disclosure of potential conflicts of interest. No potential conflicts of interest were disclosed.

8. Jebai R, Osibogun O, Li W, Gautam P, Bursac Z, Ward KD, Maziak W. Temporal trends in tobacco product use among us middle and high school students: National youth tobacco survey, 2011-2020. Public Health Reports. 2022:00333549221103812. doi: 10.1177/00333549221103812.

9. Bold KW, Buta E, Simon P, Kong G, Morean M, Camenga D, Krishnan-Sarin S. Using latent class analysis to examine susceptibility to various tobacco products among adolescents. Nicotine & Tobacco Research. 2020;22(11):2059-65. doi: 10.1093/ntr/ntz216.

10. Parms TA, Valverde R, Park-Lee E, Graham-Glover BS, Cunningham CS, Sawdey MD, Cullen KA. Symptoms of tobacco dependence among middle and high school tobacco users – data from the 2019–2020 national youth tobacco survey. Addictive behaviors. 2023;137:107537. doi: <https://doi.org/10.1016/j.addbeh.2022.107537>.

11. Gomez Y, Creamer M, Trivers KF, Anic G, Morse AL, Reissig C, Agaku I. Patterns of tobacco use and nicotine dependence among youth, united states, 2017–2018. Preventive medicine. 2020;141:106284. doi: <https://doi.org/10.1016/j.ypmed.2020.106284>.

12. Lee YO, Hebert CJ, Nonnemaker JM, Kim AE. Youth tobacco product use in the united states. Pediatrics. 2015;135(3):409. Epub 2015/02/04. doi: 10.1542/peds.2014-3202. PubMed PMID: 25647680.

13. Sun T, Lim CCW, Rutherford BN, Johnson B, Connor J, Gartner CE, Hall WD, Leung J, Chan G. Changes in patterns of youth multiple tobacco and/or e-cigarette product use in the us between 2014 and 2020: A multiple-group latent class analysis. Tobacco Control. 2022:tobaccocontrol-2022-057266. doi: 10.1136/tobaccocontrol-2022-057266.

14. Cho B, Hirschtick JL, Usidame B, Meza R, Mistry R, Land SR, Levy DT, Holford T, Fleischer NL. Sociodemographic patterns of exclusive, dual, and polytobacco use among u.S. High school students: A comparison of three nationally representative surveys. Journal of Adolescent Health. 2021;68(4):750-7. doi: <https://doi.org/10.1016/j.jadohealth.2020.11.019>.

15. Osman A, Kowitt SD, Ranney LM, Heck C, Goldstein AO. Risk factors for multiple tobacco product use among high school youth. Addictive behaviors. 2019;99:106068. doi: <https://doi.org/10.1016/j.addbeh.2019.106068>.

16. Mattingly DT, Hirschtick JL, Fleischer NL. Unpacking the non-hispanic other category: Differences in patterns of tobacco product use among youth and adults in the united states, 2009–2018. Journal of Immigrant and Minority Health. 2020;22(6):1368-72. doi: 10.1007/s10903-020-01089-0.

17. Mead EL, Lindstrom Johnson S, Siddiqui J, Butler J, Kirchner T, Feldman RH. Beyond blunts: Reasons for cigarette and cigar use among african american young adult dual users. Addiction Research & Theory. 2017;26(5):1-11. Epub 09/01/2017. doi: 10.1080/16066359.2017.1366456, NIHMSID: 958329.

18. Mead E, Chen JC, Kirchner T, Butler III J, Feldman R. An ecological momentary assessment of cigarette and cigar dual use among american american young adults. Nicotine & Tobacco Research. 2017;Accepted.

19. Chen-Sankey JC, Choi K, Kirchner TR, Feldman RH, Butler J, 3rd, Mead EL. Flavored cigar smoking among african american young adult dual users: An ecological momentary assessment. Drug and alcohol dependence. 2019;196:79-85. Epub 2019/02/13. doi: 10.1016/j.drugalcdep.2018.12.020. PubMed PMID: 30754022; PMCID: PMC6377319.

20. Mead-Morse EL, Cassidy RN, Oncken C, Tidey JW, Delnevo CD, Litt M. Validity of a little cigars/cigarillos purchase task in dual users of cigars and cigarettes. Addictive behaviors. 2022;130:107285. Epub 2022/03/08. doi: 10.1016/j.addbeh.2022.107285. PubMed PMID: 35255241; PMCID: PMC9254358.