Title: "Natural" and "Organic" Cigarette Descriptors: Association With Expectancies, Subjective Effects, Topography, and Biomarkers of Exposure Among Daily Smokers

NCT number: NCT05468333

Date of document: 7/19/2024

INTRODUCTION

"Natural" and "organic" (subsequently natural/organic) consumer products have sustained substantial growth in recent decades, with sales breaking $50 billion for the first time in 2018.[1] Perceived health benefits are a central reason why people buy natural/organic products; consumers commonly ascribe them positive characteristics such as lower calorie density, lower fat, and greater health.[2-5] Like prior research on "light" cigarettes,[6,7] research has established that natural/organic descriptors on cigarette packaging and advertising mislead current, former, and never smokers: consumers believe that cigarettes with these or similar descriptors are significantly more appealing, healthier, or less harmful than cigarettes without these descriptors.[8-17] Despite these perceptions, there is no evidence that natural/organic cigarettes are less harmful to a person's health than conventional cigarettes. Natural American Spirit (NAS) is the most prominent US brand using these descriptors, controlling roughly 3-4% of the cigarette market in 2017.[18] In comparison, Newport, the #2 brand in the US, controlled only 14% of the market in the same year.[19] In 2019, NAS was the only cigarette brand owned by British American Tobacco (which owns Camel and Newport) that increased both market share and sales volume in the US[20]; NAS's sales volume grew by another 6% in 2020.[21] Given NAS's success, it is unsurprising that other US tobacco brands (e.g., Winston, Nat Sherman, Seneca) also use these descriptors. While less common, cigarette brands also employ these terms outside the US: an evaluation of cigarette packs in 14 low and middle income countries found that cigarette brands used "natural" in 10/14 countries.[22]

The passage of the 2009 Family Smoking Prevention and Tobacco Control Act (FSPTCA) prohibited tobacco companies from making explicit or implicit reduced-risk claims without a permissive order from the Food and Drug Administration (FDA).[23] In August 2015, FDA issued a warning letter to Santa Fe Natural Tobacco Company, the makers of NAS cigarettes, and two other tobacco manufacturers, stating that their labeling and advertising of products as "natural" and "additive-free" violated section 911 of the FSPTCA because their labeling implies that "the tobacco products present a lower risk of tobacco-related disease or is less harmful" than other tobacco products.[24,25] In 2017, RJ Reynolds (parent company of SFNTC) agreed to remove "additive-free" and some instances of "natural" from NAS labeling and marketing, but the brand retained "natural" as part of its brand name. It is within FDA CTP's purview to demand changes to misleading brand names, and the agency could act on this issue in the future.

Unlike "natural," FDA CTP has not publicly initiated regulatory action against the use of "organic." NAS cigarettes use the phrase "made with organic tobacco" as allowed by the United States Department of Agriculture (USDA) National Organic Program, meaning that at least 70% of its ingredients come from a certified organic source. If the use of the "organic" descriptor is allowed by USDA, but in violation of section 911 of the FSPTCA, then FDA's and USDA's regulatory authorities are in conflict. Resolving this conflict will require science to guide regulatory decision making.

The proposed research will build upon prior work examining descriptor effects on expectancies only by additionally examining how organic/natural expectancies affect subjective effects (e.g. harshness, strength), puffing behavior, and acute exposure to harmful chemicals (e.g., TSNAs, etc.), potentially linking descriptors with actual tobacco use behavior. We will use response expectancy theory to guide our investigation of how tobacco users interpret the experience of using a natural/organic tobacco product. Response expectancy theory asserts that expectancies can shape a person's interpretation of a physical experience (e.g., subjective effects of smoking, such as taste) and affect behavior.[26] For example, smokers commonly

understand “strong” or “harsh” cigarette smoke as an indicator of increased harm[6]; smokers interpret “smoothness” as an indicator of reduced harm.[27] Similarly, a smoker may engage their understanding of “organic” from other consumer product labeling, and interpret the subjective experience of smoking an “organic” cigarette differently than if that same cigarette pack did not bear the “organic” label. Descriptors may prime consumers to interpret the subject effects of using a cigarette differently, leading to unexpected responses to regulatory actions meant to reduce abuse liability and/or appeal.[6,26] Thus, this work will provide the FDA with critical knowledge regarding how descriptors affect not only perceptions but also the subjective interpretation of using natural/organic cigarettes, which may guide regulatory decisions related to cigarette marketing and design.

We will also examine whether organic/natural expectancy effects are moderated by gender. While men and women report occasional organic product purchasing at similar rate, women are more likely than men to report consistent organic purchases[28] and to hold more positive views of organic products.[28,29] The feminization of products framed as “healthy” or “lower harm” is evident in the history of tobacco companies’ attempts to lessen the fears of health concerned smokers. Women were more likely than men to adopt several of the major “harm reducing” tactics promoted by the tobacco industry, including filtered and “low yield” cigarettes.[30] Given these facts and results from our pilot study,[31] we hypothesize that women will react more favorably than men to the “natural” and “organic” cigarette conditions in the proposed study.

METHODS

Study overview

The overall purpose of this study is to examine the relationship between the “organic” and “natural” descriptors and health risk expectancies, subjective effects, topography, and resulting chemical exposures. This is a within-person human laboratory study with assessments of psychological, behavioral, and biological exposure data. The procedures described here were tested in a pilot study conducted in late 2019.[32] Data will be collected at the University of Nevada, Reno (Reno, Nevada, USA), the Desert Research Institute (Reno, Nevada, USA), and the University of Pennsylvania (Philadelphia, Pennsylvania, USA). The study will consist of 5 total lab visits: the baseline visit, and four experimental visits (own brand comparator condition, “natural” advertising condition, “organic” advertising condition, and “conventional” advertising condition) (Figure 1). Conditions will be randomized and counterbalanced to account for order effects.

This is a deception study; participants will believe that they are participating in market testing cigarettes and will not know the real purpose of the study. We created a faux tobacco brand and associated cigarette mailer for the organic, natural, and conventional cigarette varieties (Figure 2). While participants will believe they are testing three varieties of a new brand of cigarette, in reality they will use the same product (Nat’s “Rich Smooth Taste” cigarettes, the blinded experimental cigarette) three times. None of the labs involved in data collection are in medical or public health-focused buildings, which aids in the study deception. Results from our pilot study demonstrate that this study design will deceive participants.[32] At the last pilot study visit, we asked our 22 pilot study participants to restate the purpose of the study as they saw it. Though a handful of participants inferred that things weren’t *exactly* as they seemed, none of the participants described the actual study purpose. We will include the same question at the end of Session 5 to identify participants who were not deceived by the study (see Missing Data).

Aims & Hypotheses

Aim 1 hypotheses are primary; Aims 2 and 3 hypotheses are secondary. The specific aims and associated hypotheses are:

Aim 1: Examine the association between exposure to "natural" or "organic" cigarette descriptors and: 1) health risk expectancies (e.g., chemical content); 2) subjective effects of smoking (e.g., chemical taste); 3) smoking topography (e.g., total puff volume); and 4) biomarkers of exposure (e.g., salivary aldehydes).

$H_{1.1}$: More participants will rate cigarettes in the "natural" and "organic" conditions as having fewer chemicals relative to cigarettes in the "conventional" condition.

$H_{1.2}$: Participants' average ratings of the "chemical aftertaste" of the cigarettes in the "natural" and "organic" conditions will be higher (e.g., closer towards the "less chemical aftertaste" side of the scale) relative to their average ratings of the cigarette in the "conventional" condition.

$H_{1.3}$: Total puff volume will be higher in the "natural" and "organic" conditions than the "conventional" condition.

$H_{1.4}$: Change in salivary aldehyde concentrations pre/post administration of the study cigarette(s) will be higher in the "natural" and "organic" conditions than the "conventional" condition.

Aim 2: Examine whether the relationships in Aim 1 differ by preference for a "natural/organic" cigarette brand (NAS smokers vs. non-NAS smokers).

$H_{2.1}$: Compared to non-NAS smokers, more NAS smokers will rate the "natural" and "organic" conditions as having fewer chemicals relative to their ratings in the "conventional" condition, but not relative to their own brand condition.

$H_{2.2}$: Compared to non-NAS smokers, NAS smokers' average ratings of the "chemical aftertaste" of the cigarettes in the "natural" and "organic" conditions will be higher (e.g., closer towards the "less chemical aftertaste" side of the scale) relative to their average ratings of the cigarette in the "conventional" condition, but not relative to their own brand condition

$H_{2.3}$: NAS smokers' total puff volume will be lower in the "conventional" condition than the "natural," "organic," and own brand conditions.

$H_{2.4}$: NAS smokers' change in salivary aldehyde concentrations pre/post study cigarette(s) administration will be lower in the "conventional" condition than the "natural," "organic," and own brand conditions.

Aim 3: Examine if relationships in Aim 1 differ by gender.

$H_{3.1}$: A greater proportion of women than men will rate the cigarettes in the "natural" and "organic" conditions as having fewer chemicals relative to cigarettes in the "conventional" condition.

$H_{3.2}$: Compared to men, women's average ratings of the "chemical aftertaste" of the cigarettes in the "natural" and "organic" conditions will be higher (e.g., closer towards the "less chemical aftertaste" side of the scale) relative to their average ratings of the cigarette in the "conventional" condition.

$H_{3.3}$: Compared to men, the difference in women's total puff volume will be greater in the "natural" and "organic" conditions relative to their own brand than the difference in the "conventional" condition relative to their own brand.

$H_{3.4}$: Compared to men, women's change in salivary aldehyde concentrations pre/post study cigarette(s) administration will be greater in the "natural" and "organic" conditions relative to their own brand than in the "conventional" condition relative to their own brand.

Recruitment. We will recruit 125 adult daily NAS smokers and 125 adult daily non-NAS smokers (250 total). Our goal is to recruit 50% women for each group. Participants will be recruited via public online postings (e.g., Craigslist), participant referrals, physical flyers, and other means as necessary. Recruitment documents will direct potential participants to answer questions in an online screening survey. The screener will be programed to identify eligible individuals and inform others that they are not eligible, and will collect IP address so that the study team may identify individuals who attempt to qualify for the study by taking the screener multiple times. Study personnel will contact eligible individuals to confirm eligibility and schedule their baseline session.

Eligibility. Eligible subjects will be daily cigarette smokers who smoke at least 5 cigarettes/ day and have smoked that amount for at least the past six months; are between 21 and 65 years old; are not currently undergoing smoking cessation treatment or tying to quit; are able to communicate fluently in English (speaking, writing, and reading); and are capable of providing written informed consent, which includes compliance with the requirements and restrictions listed in human participant protections.

Additional eligibility criteria by participants' preferred brand are presented in Table 1. We define smokers' preferred brand as a brand they have used daily for at least the past 30 days. We will further restrict the participants' preferred cigarette brand characteristics to restrain variability in our sample's preferred subjective effects (e.g., we will exclude menthol tobacco users because study-supplied products will be non-menthol). We will confirm smoking status at the baseline visit using an exhaled CO test of >= 10ppm. Smokers reporting having used non-combusted forms of nicotine/tobacco in the past 30 days at baseline will not be eligible, as they are at greater risk of using these products to evade nicotine abstinence requirements before visits 2-5. We will also exclude participants who have smoked marijuana in the past week because frequent marijuana use could affect their CO or biomarker values. Ineligible participants will be those who report the blinded experimental cigarette brand (Nat's "Rich Smooth Taste" – see *Experimental Tobacco Product*) as their preferred brand (as they are likely to recognize their brand); plan to quit smoking in the next 30 days; are pregnant and/or lactating; have any impairment including, but not limited to, visual, physical, and/or neurological impairments preventing the completion of procedures included within this protocol; and have had a serious or unstable disease within the past 12 months (e.g., heart disease, cancer, schizophrenia).

Compensation. Participants will receive monetary incentives totaling up to $200 if they complete all study sessions. Participants will receive $20 for attending each session. Additionally, they will receive an abstinence bonus of $10 for the second session (total $30), $20 for the third session (total $40), $30 for the fourth session (total $50), and $40 for the fifth session (total $60). Compensation amounts increase at each session to encourage continued engagement in the protocol; compensation amounts are similar to the amounts used in a pilot study at UNR which utilized similar procedures to evaluate participant reactions to "conventional" and "organic" cigarette descriptor conditions only. [31]

Study procedures

At the beginning of Session 1, study personnel will explain the study, obtain written informed consent, and confirm eligibility. Next, participants will complete questionnaire measures of demographic information, tobacco use history, and nicotine dependence. They will also confirm their preferred cigarette brand (as reported on the screener), record sub brand information, and provide an exhaled CO test to verify tobacco use status. We will also familiarize the participant with the topography equipment and other data collection processes. At the end of the visit, staff will instruct the participants to abstain from smoking overnight prior to their next visit and will compensate them for their time.

To remove the influence of any prior cigarette consumption on smoking the study-provided cigarettes (e.g., participants who just smoked their brand before arriving at the laboratory may take few puffs due to already being satiated by their brand), participants will smoke the blinded experimental cigarette in the lab in Sessions 2-5 after 12 hours of abstinence. All participants must provide an exhaled CO test of $<$ 10ppm before starting Sessions 2-5. If participants fail to arrive abstinent, they will have one more opportunity to start the condition before removal from the study. Participants will receive the $20 session payment at their first failure as a good-faith measure but will not receive the $20 session payment if they fail again. We anticipate that we will need to recruit up to 50 additional participants to account for smokers who are unable to remain abstinent before Sessions 2-5. All sessions will begin at the same time of day (congruent within ~4 hours) to minimize diurnal variation in smoking behaviors. All study sessions will occur no less than 1 day apart and must occur within 3 months.

Session 2 is the participant's own brand assessment. Study personnel will collect a CO reading (to confirm abstinence), assess harm perceptions pre-administration of the product, collect a saliva sample, allow the participant to smoke their cigarettes (ad lib) for 15 minutes (consuming as much as they like) with the topography equipment attached, assess subjective effects, collect a saliva sample and CO reading directly after self-administration, and administer the purchase task. The purpose of this session is to understand how reinforcing participants find their own brand and is useful for disentangling the effect of the exposure from the effect of smoking a novel cigarette. For example, if topography results are null, we can use the OB condition topography data to evaluate to what degree are our results due to participants not liking the experimental cigarette. We will also use data from Session 2 to describe NAS smokers' OB evaluations and to compare them to non-NAS smokers' OB evaluations. Collecting expectancies and subjective effects will help us understand the motivations NAS smokers have for using their preferred brand.

Sessions 3-5 are the experimental exposure sessions and will have identical protocols other than the introduction of different advertising exposures. Study personnel will collect a CO reading (to confirm abstinence) and a saliva sample, introduce the cigarette to be used that day (see "Description of the exposure"), assess participants' expectancies, measure their smoking topography for 15 minutes of ad lib use, collect their subjective impressions of the cigarette, and

collect a second CO reading and saliva sample. During the final session, we will debrief the participant about the study (including clarification that "organic" and "natural" cigarettes are not reduced-harm products) and assess whether participants were deceived by the protocol by asking them to re-state their understanding of the purpose of the study.

Description of the exposure. We will tell participants that our labs have a contract to test a new line of cigarettes. We will expose participants to three direct mailers and brand descriptions for a "new" line of cigarette products by "Capital Tobacco." The mailer designs will not include American Indian or eco-friendly imagery, which could convey reduced-harm messages.[11,12,33,34] Mailers will also use the same color (orange), as varying colors could convey reduced harm messages.[35] We chose a direct mailer as the exposure for three reasons: 1) tobacco companies use direct mailers to sustain smoking and encourage switching[36,37]; 2) direct mailers are a dominant form of tobacco advertising in the USA[38]; and 3) the pilot study used a direct mailer that participants found convincing.

The mailers were produced by a professional graphic designer and are presented in Figure 2. In the "natural" condition, we will position "natural" as part of the brand name (Natural Capital Cigarettes) so that our use is similar to how NAS and other brands employ **"**natural." Similarly, we will position "organic" as a descriptor and not part of the brand name to be parallel with this descriptor's use in the NAS brand.

As participants view the mailer, study personnel will also read a script introducing the sub brand to ensure that each participant has an adequate dose of the exposure. The study sub brand script is based on existing cigarette brand descriptions taken from tobacco product websites. (See Supplemental Materials for brand descriptions.)

Experimental cigarette. After exposure to the direct mailer in Sessions 3-5, participants will smoke the faux "Capital" cigarette ad lib for 15 minutes, smoking as many cigarettes as they like. Though study personnel will present the cigarette as three different varieties of the same brand, the cigarettes will be Nat's "Rich Smooth Taste" (yellow box) in all conditions; participants will be blinded to this fact. We chose Nat's because it is a relatively unknown brand, so participants are unlikely to recognize the product's taste or appearance. We chose the "Rich Smooth Taste" sub brand because it is Nat's version of a medium-bodied cigarette.

Measurement instruments and measures

Outcomes. Our primary outcomes are differences by condition in: 1) belief that the cigarette has fewer chemicals (expectancies); 2) chemical aftertaste (subjective effects); 3) total puff volume (topography); and, 4) salivary aldehydes concentration (biomarkers of exposure).

Measures. Table 2 presents a summary of our proposed measures and schedule of assessments. We include descriptions of plans for additional measures described in Table 2 in the appropriate sections below.

At Session 1 we will collect: demographics, tobacco use history and current behavior, the Fagerstrom Test for Nicotine Dependence (FTND), the Ethically-Minded Consumer Behavior Scale (Factor 1), and the Organic Product Scale. We will assess health concern about participants' own smoking using an item from the Population Assessment of Tobacco and Health (PATH) Study, which was previously associated with NAS preference in our team's descriptive study of NAS smokers in 2016.[22]

Expectancies: We will assess expectancies directly after exposing participants to the experimental condition in Sessions 3-5. Items are rated on a 5-point Likert scale (1 = "definitely untrue", 5 = "definitely true") and begin with the statement, "Compared to your own cigarettes that you usually smoke, the cigarettes you will smoke today …" Expectancy items are listed in Table 2. The primary expectancy outcome measures is whether the experimental cigarette has

fewer chemicals than the participants' usual cigarette; this choice was informed by our pilot study data, which suggested that the perceived presence/absence of chemicals plays a significant role in smokers' understanding of the source of the harm of smoking.[32] Data from additional measures in this section will be used to describe participants health risk expectancies connected to their own brand of cigarettes, and to evaluate the validity of these items as a composite score describing the health risk expectancies for natural/organic cigarettes.

Subjective ratings: We will assess subjective ratings after each ad lib session using items developed by the tobacco industry and the UPenn laboratory and adapted given pilot study results. [31] Subjective rating items are listed in Table 2. The "chemical aftertaste," "clean tobacco taste," and "overall feeling" items are new items that reflect how smokers discussed the subjective effects of smoking an organic cigarette in our pilot study[31]; the "chemical aftertaste" is the main outcome for the subjective ratings. Items are rated on a 100 mm visual analog scale with item-specific anchors (e.g., strength: 0 = "no chemical aftertaste," 100 = "strong chemical aftertaste"). As with the additional expectancies items, the additional subjective effects items will be used to characterize participants' subjective effects from their own brand cigarettes (comparing NAS and non-NAS smokers will be especially informative), and to evaluate the validity of these items as a composite score describing subjective effects of natural/organic cigarette smoking.

As exploratory behavioral measures, we will also include a cigarette purchase task (CPT), a behavioral economic measure, to assess demand in each cigarette condition. The CPT will ask participants how many of the cigarette types (e.g., natural vs organic vs own vs conventional) they would be willing to consume a day at escalating price points (i.e., starting at free then increasing by 5 cent increments (i.e., $1.00 per pack) to $1.45 per cigarette or $29 per pack). These responses are used to calculate five demand indices: breakpoint (the first price that consumption is suppressed to zero), Omax (the maximum expenditure for the product), intensity (consumption at the lowest price, i.e., zero), Pmax (price at Omax), and elasticity (sensitivity to change in price); the first four are observed while the last is computed based on a combination of these items. Breakpoint will be the primary CPT-related outcome for these exploratory analyses. In addition to capturing actual smoking behavior with the topography device, assessing both behavioral intentions and demand indices obtained by the CPT will provide us with multiple behavioral measures to comprehensively understand how organic/natural expectancies influence behavior.

Total puff volume: We will collect participants' puffing behavior using a topography instrument developed and manufactured at the American University of Beirut and calibrated across labs to determine participants' total inhalation volume. To use the topography device, the cigarette is placed in a sterilized mouthpiece, and the internal pressure transducer measures pressure changes that occur during inhalation. The pressure changes are amplified, digitized and sampled at 1000 Hz, then software converts the signal to airflow (ml/s) in real time (s), and from this provides number of puffs, puff volume, puff duration, maximum flow, and interpuff interval (time between puffs). The primary topography outcome is total puff volume due to its well-characterized responsiveness to changes in cigarette type. [39-41]

Biomarkers and saliva sample collection procedures

The primary purpose of the biomarker analysis is to assess whether exposure to aldehydes and nicotine in cigarette smoke varies by descriptor condition. These markers could reflect smoking behavior changes (i.e., mouth holding and inhalation) which could not be captured by topography measurement (i.e., puffing). Saliva samples will be collected from the participants before and after each ad lib tobacco product use bout in Sessions 2-5. The research

staff will ask subjects to wash their mouth with clean water when collecting the 'before session' saliva sample. The "after session" samples will be collected just after the participant finishes their last puff. Saliva samples (approx. 3 mL) will be collected using a Salimetrics saliva collection kit (Salimetrics, Carlsbad, CA). Saliva samples will be shipped to DRI Reno campus and be stored at -40°C prior to analysis. Salivary aldehydes (formaldehyde and acetaldehyde) and nicotine are exposure biomarkers. We will measure the concentration of these biomarkers in the mouth, before they are metabolized by the body, and make within-person comparisons in concentrations by study descriptor conditions. In addition, nicotine metabolites (cotinine and trans-3'-hydroxycotinine) will be measured to confirm smoking abstinence.

Missing data

Participants who complete at least three study visits will be retained for analyses; those with three or fewer visits will be excluded. Additionally, we will exclude people from the study if they correctly describe the purpose of the study and the nature of the study deception at the last visit. Self-reported measures: We do not anticipate missing data within person for the baseline, expectancies, subjective measures, or purchase task because the survey software does not allow participants to complete surveys with missing data. Research assistants will be nearby to help with troubleshooting during data collection. If for some reason there is missing self-report data, we will listwise delete that observation from analyses. Topography: It is possible that the topography devices will produce out-of-range values due to malfunction. To minimize the potential malfunction, we will inspect and calibrate the topography device before each session. Using guidance from the device manufacturer, we will merge puffs with an inter-puff interval (IPI) of less than 400ms and count them as 1 puff. Similar to prior work, we will also exclude first puff data as the first puff may be affected the topography mouthpiece.[42] We will then identify topography datapoint outliers, defined as puff duration > 10 seconds and/or puff volume < 10 ml. Topography data from sessions where more than 20% of puffs are defined as topography datapoint outliers will be excluded from analyses due to device malfunction. Biosamples: Saliva sample data with less than 0.5 mL of fluid will be excluded because we are unable to analyze samples that small.

Sample size calculations

Sample size calculations were performed using GLIMMPSE Version 3.[43] Though a Bayesian model averaging approach will be used, sample size calculations using a non-Bayesian approach are applicable here because they correspond to the model that is believed to be most likely true *a priori*, and also because we plan to use uninformative priors on the parameters in the model, which makes our analyses comparable to the non-Bayesian approach that is used for sample size calculations. As we do not have pilot data to inform our sample size calculations for the NAS/non-NAS group, our sample size calculations are based on the group means by gender that were observed in the pilot study on "total puffs" and a standard deviation of 5.5.[31] Given NAS smokers' preexisting misperceptions about natural/organic cigarettes, we anticipate that group means will be more extreme in the NAS/non-NAS group.[44] Thus, the minimum sample size required for Aim 3 analyses should be sufficient to address the Aim 2 analyses. Assuming an intraclass correlation between repeated measurements of 0.5, 40 total subjects are needed to detect a significant interaction between gender and condition (natural/organic vs conventional/own brand) at 90% power and a significance level of 5%, while 99% power requires 68 subjects. Since these assumptions use information based on only 22 subjects in the pilot study, sample size calculations were also performed with an effect size 50% smaller and a

covariance matrix 50% larger than assumed: under these conditions, 90% power is achieved with 228 total subjects. Fewer subjects are needed to detect differences hypothesized in Aim 1 at each given power; for example, under the smallest effect size observed in the topography pilot data (which assumes a mean difference of 120ml with standard deviation 165ml), 120 subjects are needed to achieve 90% power under an effect size that is 50% smaller and covariance matrix that is 50% larger than assumed. Thus, we will enroll 125 participants per cigarette type (250 total, though we have budgeted for up to 300 participants to account for loss to follow up) to ensure at least 228 participants complete the full study. Our proposed sample size is adequate to detect a similar topography effect size to our pilot study, examine moderation by NAS preference and gender, and allow for adjustments to the significance level due to multiple comparisons.

Statistical analyses

The study will examine the associations between condition ("organic," "natural," or "conventional") and four classes of primary outcomes: 1) expectancies (binary outcome); 2) subjective ratings (continuous outcome); 3) topography (continuous outcome); and 4) immediate biomarkers of exposure (continuous outcome). Within these classes of outcomes, our primary comparisons are average within-person differences by condition in: 1) belief that the cigarette has fewer chemicals; 2) chemical aftertaste; 3) total puff volume; and, 4) salivary aldehydes concentration. We will treat analyses of all other outcomes as exploratory.

In a crossover design, where each subject experiences all conditions (such as this design), a differential carryover effect may bias the estimated effect of the condition. [45] A differential carryover effect would mean that the effect of the condition depends on the period (visit number), which is a carryover effect, and that this carryover effect is larger for certain sequences of visits or types of subjects (e.g. NAS-smoking type or gender). The traditional approach of Grizzle (1965) [46] uses only independent data if there is evidence of a carryover effect, which would mean using only data from visits 1-3 (out of 5) for all subjects in this study; on other hand, if there is no evidence of a carryover effect, then all data can be used. With a counterbalanced design, using only data from visits 1-3 would allow us to test hypotheses H(1.1)-H(3.4), but hypothesis testing would have lower statistical power compared to using data from all 5 visits. Even though counterbalancing reduces carryover effects, it is not guaranteed to prevent them. Differences in carryover effects can be incorporated within a Bayesian framework without resorting to the 'all-or-nothing' approach that follows a preliminary test [47]; we will therefore use a Bayesian approach for these analyses.

In particular, a Bayesian linear and logistic mixed effects model will be used to test hypotheses H(1.1) - H(3.4). Since we do not anticipate a carryover effect, main analyses will use a prior distribution that assumes there is no carryover effect. However, sensitivity to this choice of prior will be assessed by repeating analyses under different choices of prior that allow for a carryover effect using an approach similar to Grieve (1985).[48] For example, one particular choice for this prior assumes a carryover effect is absolutely certain (more details of this prior are below).

We will consider models that: 1) allow the effect of the condition (conventional, organic, and natural) to differ based on NAS-smoking status and gender (as hypothesized); 2) allow the effect of the condition to differ based only on NAS-smoking status; 3) allow the effect of the condition to differ based only on gender; and 4) do not allow the effect of the condition to differ based on NAS-smoking status or gender. We will also consider models that combine the effects of the natural and organic conditions, as we believe the effects of these two conditions could be similar for our outcomes. We will also consider in the model an effect for the own brand

condition (which may or not depend on NAS-smoking status and gender), period effects, random effects, and effects for other covariates that may be predictive of the outcome. Bayes factors will be used to determine if the effect of the condition depends on NAS-smoking status or gender, and if a combined effect (one effect vs two) should be used in the model for the natural and organic conditions.

Baseline covariates that may be significant predictors of the outcome will be considered in the model because including predictors that reduce noise may increase the efficiency of the analyses in H(1.1) – H(3.4). We will use Bayesian model averaging to control for other covariates and estimate the parameters in each considered model.[49] Bayesian model averaging incorporates model uncertainty into inference and has several other advantages over the approach that relies on choosing a single best model among many candidate models.[50] In general, Bayesian model averaging uses all possible models that can be considered given a set of covariates (there are 2^k possible models when there are k covariates) and estimates parameters by using a weighted average of the estimates obtained from each model, where better fitting models are given larger weights (see Raftery et al. (1997)[49] and Hinne et al. (2020)[50] for more details). The following baseline measures in particular will be considered (which are *a priori* expected to be related to the outcomes that are used to test hypotheses H(1.1) – H(3.4)): score on the organic products scale, age, education, heaviness of smoking, perceived health, and own brand total puff volume (for all outcomes except the topography outcome, which uses total puff volume as the response); however, other covariates collected at baseline (e.g. race, ethnicity) will also be considered in the model. Using Bayes factors to first choose the best representation for the effects of the condition in the model and using Bayesian model averaging to estimate the parameters in the model will be repeated separately for each outcome, so that we have one model for each of the four outcomes: expectancies, subjective ratings, topography, and biomarkers of exposure.

A continuous response is a reasonable assumption for all outcomes except those in H(1.1), H(2.1), H(3.1), and H(4.1); in these hypotheses, the outcome will be modeled as a binary response (0/1) and an alternative model formulation (e.g. logistic mixed model) consisting of the same model parameters (e.g. same linear predictor as the other outcomes but using a logit link function) will instead be used to model the probability that the outcome is equal to 1. Other outcomes may also be modeled as a binary response depending on their distribution (more details are provided below).

Preliminary analyses assume there are no carryover effects. However, sensitivity to this assumption will be examined by considering models that instead assume there are carryover effects (as described above); if there are carryover effects, then the effect of the condition also depends on the period (in addition to possibly gender and NAS-smoking status). For the main analyses that assume there are no carryover effects, we will use uninformative priors for all model parameters. Markov Chain Monte Carlo (MCMC) sampling will be used to estimate the posterior distribution of the model parameters; if there is a convergence issue for a parameter (e.g. if variance hyperparameters are on the boundary) or if model fit is a concern, we will instead consider weakly informative priors. [51] When uninformative priors are used, results using a single candidate model with a Bayesian model averaging approach are equivalent to non-Bayesian, frequentist analyses (i.e. maximum likelihood estimation techniques).

To assess sensitivity of our choice of prior for the carryover effect, we will use a process similar to Grieve (1985).[48] To allow for a carryover effect, we will first use a Normal distribution with mean 0 and variance σ_{co}^2 as the prior for the carryover effects, and use an uninformative prior for σ_{co}^2 that ranges between 0 and ∞. Next, we will let π be the prior odds of M_0 to M_1, where M_0 is the model above without carryover effect and M_1 is the model with carryover

effect. We can then assess the impact of a range of prior belief in a carryover effect by repeating analyses for different values of π. The primary analyses assume there is no carryover effect, which corresponds to π equal to 1, while sensitivity analyses will assume there is a carryover effect, which corresponds to $0 \leq \pi < 1$.

Since MCMC methods will be used, the entire posterior distribution corresponding to the parameters of interest will be estimated and can be produced as figures. Figures may be overlaid to show how results are sensitive to our choice of π. Even though we will obtain the entire posterior distribution of our parameters, we will also calculate 95% credible intervals, which is an alternative way of presenting effect sizes corresponding to hypotheses H(1.1) – H(3.4). Credible intervals will also be used to conduct hypothesis testing (details on hypothesis testing are below). Below are the model outcomes that will be used to test each hypothesis:

H(1.1), H(2.1), H(3.1): Expectancies (Have fewer chemicals)

We will create a binary variable from a 5-point Likert scale (Definitely true, Somewhat true, Neither true nor false, Probably false, Definitely false) for this measurement to indicate if the participant believes the cigarettes smoked during the visit have fewer chemicals (if answered Definitely true or Somewhat true) or not (if answered Neither true nor false, Probably false, or Definitely false). We will use an approach that models the probability of belief that cigarettes smoked during the visit have fewer chemicals, and we will use the logit link function so that the form of the best model will resemble a mixed effects logistic regression model; other link functions (e.g. probit) will be considered if model fit is a concern.

H(1.2), H(2.2), H(3.2): Subjective effects (chemical aftertaste)

Chemical aftertaste will be used as the outcome, which will have items rated on a 100 mm visual analog scale with item-specific anchors (e.g., strength: 0 = "strong chemical aftertaste," 100 = "no chemical aftertaste"). We do not anticipate many measurements on the boundaries (0 or 100), which makes Normality an appropriate assumption for the residuals. However, if there are many measurements on the boundaries (i.e. if Normality does not seem like an appropriate assumption for the residuals), then a binary outcome that indicates if the measurement is greater than 50 will be used instead, and we will use the logit link function so that the form of the best model will resemble a mixed effects logistic regression model (like above for the expectancies outcome); other link functions (e.g. probit) will be considered if model fit is a concern.

H(1.3), H(2.3), H(3.3): Topography (total puff volume)

Normality is expected to be an appropriate assumption for total puff volume (a quantitative variable) based on pilot data.[32] However, if Normality does not seem like an appropriate assumption for the residuals in this sample, then a transformed outcome (e.g. log-transformation) will be considered (particularly if the outcome is skewed). If normality is still violated, we will consider a binary outcome using a median cutoff and use the logit link function so that the form of the best model will resemble a mixed effects logistic regression model (like above for the expectancies outcome); other link functions (e.g. probit) will be considered if model fit is a concern.

H(1.4), H(2.4), H(3.4): Biomarkers of exposure (Salivary aldehydes)

Concentrations of formaldehyde and acetaldehyde will be used as the outcome (difference between pre- and post-visit), where Normality is expected to be an appropriate assumption for the error term. However, just as with the other outcomes, if Normality does not

seem like an appropriate assumption, a transformed outcome (e.g. log-transformation) will first be explored, and if normality still does not hold, then we will consider a binary outcome using a median cutoff and use the logit link function so that the form of the best model will resemble a mixed effects logistic regression model (like above for the expectancies outcome); other link functions (e.g. probit) will be considered if model fit is a concern.

Hypothesis testing

H(1.1 - 1.4) are primary, and under the null hypotheses in H(1.1 - 1.4), the natural and organic condition means are equal to the conventional condition means. To test the hypotheses in Aim 1, we will first determine if the effects of the condition depend on gender and NAS-smoking status (as hypothesized in H(2.1) - H(3.4)). If the effect of the condition does not depend on gender or NAS-smoking status, then we will estimate the following effects in our model:

1. effect of the conventional condition
2. effect of the natural condition
3. effect of the organic condition

The null hypotheses in H(1.1 - 1.4) state that the conventional condition is equivalent to the natural and organic conditions. We can therefore test H(1.1 - 1.4) by estimating 95% credible intervals for the difference between the effects of the organic and natural conditions and the conventional condition. This will yield two credible intervals (one for the difference between natural and conventional effects, and another for the difference between organic and conventional effects) for each outcome (four outcomes) therefore making 8 total credible intervals of interest in Aim 1. If the effects of the condition depend on NAS-smoking status but do not depend on gender, then we will estimate the following effects in our model:

1. effect of the own brand condition for Non-NAS smokers
2. effect of the conventional condition for Non-NAS smokers
3. effect of the natural condition for Non-NAS smokers
4. effect of the organic condition for Non-NAS smokers
5. effect of the own brand condition for NAS smokers
6. effect of the conventional condition for NAS smokers
7. effect of the natural condition for NAS smokers
8. effect of the organic condition for NAS smokers

The effect of the condition will then be tested separately by NAS-status. That is, we will calculate 95% credible intervals for the differences between the conventional condition and the natural and organic conditions (organic - conventional and natural - conventional) separately for non-NAS smokers and NAS smokers. This will yield four credible intervals (organic - conventional for NAS smokers, organic - conventional for non-NAS smokers, natural - conventional for NAS smokers, and natural - conventional for non-NAS smokers) for each outcome (four outcomes) therefore making 16 total credible intervals of interest in Aim 1.

If the effects of the condition depend on gender but not NAS-smoking status, then we will estimate the following effects in our model:

1. effect of the own brand condition for men
2. effect of the conventional condition for men
3. effect of the natural condition for men

4. effect of the organic condition for men
5. effect of the own brand condition for women
6. effect of the conventional condition for women
7. effect of the natural condition for women
8. effect of the organic condition for women

The effect of the condition will then be tested separately by gender. That is, we will calculate 95% credible intervals for the differences between the conventional condition and the natural and organic conditions (organic - conventional and natural - conventional) separately for men and women. This will yield four credible intervals (organic - conventional men, organic - conventional women, natural - conventional men, and natural - conventional women) for each outcome (four outcomes) therefore making 16 total credible intervals of interest in Aim 1.

Lastly, if the effects of the condition depend on both gender and NAS-smoking status (as hypothesized), then we will estimate 95% credible intervals for the differences between the conventional condition and the natural and organic conditions separately for non-NAS men, non-NAS women, NAS men, and NAS women. This will yield eight credible intervals for each outcome, so that there are 32 credible intervals of interest (four outcomes each with eight credible intervals) in Aim 1. If it is determined that the natural and organic condition effects can be combined into a single effect (e.g. if the Bayes factor provides evidence for the model that combines the effect over the model that includes separate effects for the natural and organic conditions), then only a single comparison will be made (rather than two comparisons) for the organic/natural conditions in above.

The multilevel modeling approach of Gelman et al. (2012) [52] will be used to account for multiple comparisons within each aim. The number of comparisons (or contrasts of interest) in Aim 1 (which are based on the following contrasts: organic - conventional and natural - conventional) are determined according to whether the effect of the condition depends on NAS-smoking status or gender, and whether the effects for the natural and organic conditions can be combined (as outlined above). It is relatively straightforward to account for multiplicity within each outcome following the multilevel modeling approach outlined in Gelman et al. (2012). [52] However, incorporating this approach across the four different outcomes in each aim may not be as trivial; the approach of Westfall (1997) [53] may instead be used to account for multiple outcomes.

To test H(2.1) and H(2.2), we will estimate effects for the following subgroups:

1. effect of NAS smokers under the own brand condition
2. effect of NAS smokers under the conventional condition
3. effect of NAS smokers under the natural condition
4. effect of NAS smokers under the organic condition
5. effect of non-NAS smokers under the own brand condition
6. effect non-NAS smokers under the conventional condition
7. effect non-NAS smokers under the natural condition
8. effect of non-NAS smokers under the organic condition

95% credible intervals will be estimated for the difference in effects between 1) 4. - 2. and 8. - 6., 2) 3. - 2. and 7. - 6, 3) 4. - 1. and 8. - 5., and 4) 3. - 1. and 7. - 5. If the credible interval in 1) does not contain 0, then there is evidence (against the null hypothesis) that NAS and non-NAS smokers differ in their mean organic condition outcomes relative to their conventional condition outcomes (otherwise there is evidence for the hull hypothesis and therefore no

difference in these two groups). If 2) does not contain 0, then there is evidence (against the null hypothesis) that NAS and non-NAS smokers differ in their mean natural condition outcomes relative to the conventional condition outcomes (otherwise there is evidence for the hull hypothesis and therefore no difference in these two groups). If 3) does not contain 0, then there is evidence (against the null hypothesis) that NAS smokers and non-NAS smokers differ in their mean organic condition outcomes relative to their own brand condition outcomes (otherwise there is evidence for the hull hypothesis and therefore no difference in these two groups). If 4) does not contain 0, then there is evidence (against the null hypothesis) that NAS smokers and non-NAS smokers differ in their mean natural condition outcomes relative to their own brand condition outcomes (otherwise there is evidence for the hull hypothesis and therefore no difference in these two groups). H(2.1) and H(2.2) hypothesizes that there will be evidence against the null hypothesis in 1) and 2) but evidence for the null hypothesis in 3) and 4).

To test H(2.3) and H(2.4), we will estimate the 95% credible intervals corresponding to the difference in means between 6) 4. and 2. above, 7) 3. and 2, 8) 4. and 1., and 9) 3. and 1. If the credible interval in 6) / 7) / 8) / 9) does not contain 0, then there is evidence that NAS smokers mean outcome under the organic / natural / organic / natural condition is significantly different than NAS smokers mean outcome under the conventional / conventional / own brand / own brand condition (evidence against null), otherwise there is evidence that NAS smokers mean outcome is no different in the organic / natural / organic / natural condition compared to the convention / conventional / own brand / own brand condition (evidence for null). We hypothesize that the organic and natural conditions will differ from the conventional condition (i.e. evidence against null in 6) and 7)) but not the own brand condition (i.e. evidence for null in 7) and 8)).

To test H(3.1)-H(3.4), we will estimate effects for the following subgroups:

1. effect of men smokers under the own brand condition
2. effect of men smokers under the conventional condition
3. effect of men smokers under the natural condition
4. effect of men smokers under the organic condition
5. effect of women smokers under the own brand condition
6. effect of women smokers under the conventional condition
7. effect of women smokers under the natural condition
8. effect of women smokers under the organic condition

95% credible intervals will be estimated for the difference in effects between 9) 4. - 2. and 8. - 6., 10) 3. - 2. and 7. - 6., 11) 4. - 1. - (2. - 1.) and 8. - 5. - (6. - 5.), and 12) 3. - 1. - (2. - 1.) and 7. - 5. - (6. - 5.). If the credible interval in 9) / 10) does not contain 0, then there is evidence that men smokers mean outcome under the organic / natural condition relative to the convention condition differs from the women smokers mean outcome under the organic / natural condition relative to the convention condition, which would provide evidence against the null hypothesis in H(3.1) and H(3.2). If the credible interval in 11) / 12) does not contain 0, then there is evidence that men smokers mean outcome difference between the organic / natural condition relative to the own brand condition and convention condition relative to own brand condition differs from the women smokers mean outcome difference between the organic / natural condition relative to the own brand condition and convention condition relative to own brand condition, which would provide evidence against the null hypothesis in H(3.3) and H(3.4) (assuming the direction of the difference is not opposite the hypothesized direction).

Biomarker analyses

Previous work, including our preliminary biosample analyses, demonstrate that short exposures similar to our 15 min ad lib bout increase exposure biomarkers like saliva nicotine (>200-fold) and aldehydes (2-10-fold).[54,55] Participants' saliva samples (before and after tobacco use sessions) will be characterized for nicotine and aldehyde concentrations and compared across descriptor conditions.

Salivary aldehyde (e.g., acetaldehyde, and formaldehyde) concentrations will be analyzed using O-(2,3,4,5,6-Pentafluorobenzyl)-hydroxylamine (PFBHA, Supelco, Bellefonte, PA) derivatization method. We focus formaldehyde and acetaldehyde as they may make the greatest contribution to cancer risk of all major harmful constituents in mainstream cigarette smoke and because we have demonstrated feasibility of this collection technique to characterize aldehydes in saliva after a 15 min smoking session in our pilot study.[31] In brief, 1 ml of saliva samples will be added into a 10 ml glass tube. 500 µl of HPLC grade water, 50 µl of deuterated internal standard mixture, and 100 µl of PFBHA solution (250 mg/ml in HPLC grade water) will be added into the glass tube, pH will be adjusted to 3.9 using sulfuric acid, and the sample will be incubated for an hour under room temperature in the dark, and then 200 µl of sulfuric acid will be added to stop the reaction. 1 ml of dichloromethane (DCM) will be added into the glass tube and shaken well for 1 minute, then the DCM part will be transferred into a new glass tube. This procedure will be repeated three times to ensure all target analytes are extracted. 100 mg of sodium sulfate will be added into the DCM extract and shaken well to remove remaining water. The DCM fraction will be transferred to toluene (final volume=200 µl) under gentle stream of nitrogen. The sample will be injected into gas chromatography (GC)-mass spectrometer (MS) for aldehyde-PFBHA quantification (PerkinElmer Clarus 680 GC and Clarus SQ 8T MS, PerkinElmer, Waltham, MA). Injection volume and injector temperature will be 1 µl and 250 °C, respectively. The target analytes will be separated using PerkinElmer Elite-5MS column (L=30 m, ID=0.25 µm, DF=0.25 µm) with column oven gradients starting at 40 °C, hold for 2 minutes, increase to 230 °C at 5 °C/minute gradient, hold for 2 minutes, then increase to 325 °C at 20 °C/minute gradient, and hold for 5 minutes. Limit of detection (LOD) and limit of quantification (LOQ) for formaldehyde-PFBHA are 0.34 ng/ml and 1.13 ng/ml, respectively, and acetaldehyde-PFBHA are 0.41 ng/ml and 1.37 ng/ml, respectively.

To measure nicotine (NIC), cotinine (COT), and trans-3'-hydroxycotinine (3OHC), 200 µl of saliva will be transferred into a micro-centrifuge tube. Known amounts of deuterated internal standards (NICd3, COTd3, and 3OHCd3, Cayman chemical, Ann Arbor, MI) will be added into the tube as an internal standard. The tube will be stored for 10 minutes in 4 °C refrigerator after adding 750 µl of glacial acetone, followed by centrifuging at 3000 g for 30 minutes. The supernatants will be collected in a new micro-centrifuge tube, 150 µl of methanol will be added into the tube and concentrated until 100 µl under gentle stream of nitrogen. The sample will then be transferred into an amber autosampler vial and 10 µl of the sample will be injected into an Ultra Performance Liquid Chromatography (UPLC)-tandem mass spectrometer (MS/MS) (Waters Aquity UPLC and Xeto TQ-S MS/MS, Waters, MA, USA) equipped with a Waters BEH C18 column (1.7 µm, 2.1×50 mm, Waters, MA, USA). The column will be heated at 35 °C. Methanol with 0.1% Formic acid and HPLC-grade water with 0.1% Formic acid will be used as phase A and B. Mobile phase gradient (0.3 ml/minute) will be 21% phase A at 0 minute and hold for 0.8 minutes, increase to 90% phase A in 0.4 minutes and hold for 0.9 minutes, then decrease to 21% phase A in 0.2 minutes and hold for 0.7 minutes. Target analytes will be identified using 25 V cone voltage and 35 V collision voltage. Parent and daughter m/z are 163 and 132 for NIC (166 and 132 for NICd3), respectively, 177 and 80 for COT (180 and 80 for COTd3), respectively, and

193 and 80 for 3OHC (196 and 80 for 3OHCd3), respectively. LOD for NIC, COT, and 3OHC are 1.20 ng/ml, 5.70 ng/ml, and 6.22 ng/ml, respectively, and LOQ for the compounds are 4.01 ng/ml, 19.0 ng/ml, and 20.7 ng/ml, respectively.

REFERENCES

1. U.S. organic sales break through $50 billion mark in 2018 17 May 2019, 2019. Accessed 13 July 2020. https://ota.com/news/press-releases/20699
2. Lee W-cJ, Shimizu M, Kniffin KM, Wansink B. You taste what you see: Do organic labels bias taste perceptions? *Food Qual Prefer*. 2013;29(1):33-39. doi:10.1016/j.foodqual.2013.01.010
3. Gaylord K, Chen D, Nair SS, Andersond MJ. The Role of Healthy Diet Belief in Mediating the Organic Label Effect on Increased Food Consumption. *Journal of Food Distribution Research*. 2020;51(1):50-56.
4. Rana J, Paul J. Consumer behavior and purchase intention for organic food: A review and research agenda. *Journal of Retailing and Consumer Services*. 2017;38:157-165. doi:10.1016/j.jretconser.2017.06.004
5. Bublitz MG, Peracchio LA, Block LG. Why did I eat that? Perspectives on food decision making and dietary restraint. *J Consum Psychol*. July 2010;20(3):239-258.
6. Shiffman S, Pillitteri JL, Burton S, Rohay J, Gitchell JG. Smokers' beliefs about "Light" and "Ultra Light" cigarettes. *Tob Control*. 2001;10(1):i17-i23.
7. O'Connor RJ, Caruso RV, Borland R, et al. Relationship of cigarette-related perceptions to cigarette design features: findings from the 2009 ITC U.S. Survey. *Nicotine Tob Res*. Nov 2013;15(11):1943-7. doi:10.1093/ntr/ntt075
8. Arnett JJ. Winston's "No Additives" campaign: "straight up"? "no bull"? *Public Health Rep*. Nov-Dec 1999;114(6):522-7. doi:10.1093/phr/114.6.522
9. Moran MB, Brown J, Lindblom E, et al. Beyond 'Natural': Cigarette Ad Tactics that Mislead about Relative Risk. *Tobacco Regulatory Science*. Sep 2018;4(5):3-19. doi:10.18001/trs.4.5.1
10. Moran MB, Pearson JL. Real. Simple. Deadly. A Pilot Test of Consumer Harm Perceptions in Response to Natural American Spirit Advertising. *Tobacco Regulatory Science*. 2019;5(4):360-368. doi:10.18001/trs.5.4.6
11. Iles IA, Pearson JL, Lindblom E, Moran MB. "Tobacco and Water": Testing the Health Halo Effect of Natural American Spirit Cigarette Ads and Its Relationship with Perceived Absolute Harm and Use Intentions. *Health Commun*. Jun 2021;36(7):804-815. doi:10.1080/10410236.2020.1712526
12. Gratale SK, Maloney EK, Cappella JN. Regulating language, not inference: an examination of the potential effectiveness of Natural American Spirit advertising restrictions. *Tob Control*. Aug 2019;28(e1):e43-e48. doi:10.1136/tobaccocontrol-2018-054707
13. Gratale SK, Maloney EK, Sangalang A, Cappella JN. Influence of Natural American Spirit advertising on current and former smokers' perceptions and intentions. *Tob Control*. Sep 2018;27(5):498-504. doi:10.1136/tobaccocontrol-2017-053881
14. Czoli CD, Hammond D. Cigarette packaging: Youth perceptions of "natural" cigarettes, filter references, and contraband tobacco. *J Adolesc Health*. Jan 2014;54(1):33-9. doi:10.1016/j.jadohealth.2013.07.016

15. Byron M, Baig S, Moracco K, Brewer N. Adolescents' and adults' perceptions of 'natural', 'organic' and 'additive-free' cigarettes, and the required disclaimers. *Tob Control*. 2016;25(5):517-20. doi:doi:10.1136/tobaccocontrol-2015-05256
16. Pearson JL, Richardson A, Feirman SP, et al. American Spirit Pack Descriptors and Perceptions of Harm: A Crowdsourced Comparison of Modified Packs. *Nicotine Tob Res*. Aug 2016;18(8):1749-56. doi:10.1093/ntr/ntw097
17. Baig SA, Byron MJ, Lazard AJ, Brewer NT. "Organic," "Natural," and "Additive-Free" Cigarettes: Comparing the Effects of Advertising Claims and Disclaimers on Perceptions of Harm. *Nicotine Tob Res*. Jun 21 2019;21(7):933-939. doi:10.1093/ntr/nty036
18. Pearson J, Ganz O, Ohman-Strickland P, Wackowski OA. Shifts in preference for Natural American Spirit and associated belief that one's own cigarette brand might be less harmful than other brands: results from Waves 1-4 of the Population Assessment of Tobacco and Health (PATH) Study (2013-2018). *Tob Control*. Dec 24 2021;doi:10.1136/tobaccocontrol-2021-056985
19. US Centers for Disease Control and Prevention. Tobacco Brand Preferences. Accessed 12 February, 2021. https://www.cdc.gov/tobacco/data_statistics/fact_sheets/tobacco_industry/brand_preference/index.htm
20. British American Tobacco. *Annual Report and Form 20-F*. 2019. Accessed 14 July 2020. https://www.bat.com/group/sites/UK__9D9KCY.nsf/vwPagesWebLive/DOAWWGJT/$file/BAT_Annual_Report_and_Form_20-F_2019.pdf
21. British American Tobacco. BAT Annual Report and Form 20-F 2020: Our combustible products. Accessed 3 January, 2022. https://www.bat.com/ar/2020/pdfs/BAT_AR20-F_2020_Our_Combustible_Products.pdf
22. Welding K, Kennedy R, Moran M, Cohen J, Smith K. Natural, organic, additive-free and pure on cigarette packs in 14 Countries. *Tobacco Regulatory Science*. 2019;5(4):352-359.
23. Family Smoking Prevention and Tobacco Control Act (section 911), U.S.C. §§ 387k(g) (2009).
24. U.S. Food and Drug Administration. Warning letter to Santa Fe Natural Tobacco Company, Inc. 8/27/15. Accessed October 6, 2015. https://www.truthinadvertising.org/wp-content/uploads/2015/09/FDA-warning-letter-Santa-Fe-Natural-Tobacco-Co.pdf
25. U.S. Food and Drug Administration. FDA takes action against three tobacco manufacturers for making “additive-free” and/or “natural” claims on cigarette labeling. Updated Aug 27. Accessed October 6, 2015. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm459840.htm
26. Harrell PT, Juliano LM. A direct test of the influence of nicotine response expectancies on the subjective and cognitive effects of smoking. *Exp Clin Psychopharmacol*. Aug 2012;20(4):278-86. doi:10.1037/a0028652
27. Elton-Marshall T, Fong GT, Yong HH, et al. Smokers' sensory beliefs mediate the relation between smoking a light/low tar cigarette and perceptions of harm. *Tob Control*. Nov 2015;24 Suppl 4:iv21-7. doi:10.1136/tobaccocontrol-2014-051977
28. Formanski K. *The Natural/Organic Food Shopper*. 2019. July 2019. Accessed 13 July 2020.
29. Ureña F, Bernabéu R, Olmeda M. Women, men and organic food: differences in their attitudes and willingness to pay. A Spanish case study. *Int J Consumer Stud*. 2007;32(1):071203213649007-??? doi:10.1111/j.1470-6431.2007.00637.x
30. Pollay RW, Dewhirst T. The dark side of marketing seemingly "Light" cigarettes: successful images and failed fact. *Tob Control*. Mar 2002;11 Suppl 1(Suppl 1):I18-31. doi:10.1136/tc.11.suppl_1.i18

31. Pearson JL, Watanabe M, Sanchez J, Mann S, Drake C, Mercincavage M. The "organic" descriptor and its association with commercial cigarette health risk expectancies, subjective effects, and smoking topography: a pilot human laboratory study. *Nicotine Tob Res*. Jul 21 2021;doi:10.1093/ntr/ntab151
32. Pearson JL, Watanabe M, Sanchez J, Mann S, Drake C, Mercincavage M. The "Organic" Descriptor and Its Association With Commercial Cigarette Health Risk Expectancies, Subjective Effects, and Smoking Topography: A Pilot Human Laboratory Study. *Nicotine Tob Res*. Jan 1 2022;24(1):69-76. doi:10.1093/ntr/ntab151
33. Epperson AE, Averett PE, Blanchflower T, Gregory KR, Lee JGL. "The Packaging Is Very Inviting and Makes Smokers Feel Like They're More Safe": The Meanings of Natural American Spirit Cigarette Pack Design to Adult Smokers. *Health Educ Behav*. Apr 2019;46(2):260-266. doi:10.1177/1090198118820099
34. Epperson AE, Henriksen L, Lambin EF, Flora JA, Prochaska JJ. Health Beliefs of American Indian Imagery on Natural American Spirit Packs. *Tobacco Regulatory Science*. 2019;5(4):369-380. doi:10.18001/trs.5.4.7
35. Hammond D, Parkinson C. The impact of cigarette package design on perceptions of risk. *J Public Health (Oxf)*. Sep 2009;31(3):345-53. doi:10.1093/pubmed/fdp066
36. Brock B, Schillo BA, Moilanen M. Tobacco industry marketing: an analysis of direct mail coupons and giveaways. *Tob Control*. Sep 2015;24(5):505-8. doi:10.1136/tobaccocontrol-2014-051602
37. Choi K, Chen JC, Tan ASL, Soneji S, Moran MB. Receipt of tobacco direct mail/email discount coupons and trajectories of cigarette smoking behaviours in a nationally representative longitudinal cohort of US adults. *Tob Control*. Jun 19 2018;doi:10.1136/tobaccocontrol-2018-054363
38. Federal Trade Commission. *Cigarette Report for 2018*. 2019. https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-cigarette-report-2018-smokeless-tobacco-report-2018/p114508cigarettereport2018.pdf
39. Mercincavage M, Smyth JM, Strasser AA, Branstetter SA. Reduced Nicotine Content Expectancies Affect Initial Responses to Smoking. *Tob Regul Sci*. Oct 2016;2(4):309-316. doi:10.18001/TRS.2.4.3
40. Mercincavage M, Souprountchouk V, Tang KZ, et al. A Randomized Controlled Trial of Progressively Reduced Nicotine Content Cigarettes on Smoking Behaviors, Biomarkers of Exposure, and Subjective Ratings. *Cancer Epidemiol Biomarkers Prev*. Jul 2016;25(7):1125-33. doi:10.1158/1055-9965.EPI-15-1088
41. Mercincavage M, Lochbuehler K, Wileyto EP, et al. Association of Reduced Nicotine Content Cigarettes With Smoking Behaviors and Biomarkers of Exposure Among Slow and Fast Nicotine Metabolizers: A Nonrandomized Clinical Trial. *JAMA Netw Open*. Aug 2018;1(4):e181346. doi:10.1001/jamanetworkopen.2018.1346
42. Lee EM, Malson JL, Waters AJ, Moolchan ET, Pickworth WB. Smoking topography: reliability and validity in dependent smokers. *Nicotine Tob Res*. Oct 2003;5(5):673-9. doi:10.1080/1462220031000158645
43. Kreidler SM, Muller KE, Grunwald GK, et al. GLIMMPSE: Online Power Computation for Linear Models with and without a Baseline Covariate. *J Stat Softw*. Sep 2013;54(10)doi:10.18637/jss.v054.i10
44. Pearson JL, Johnson A, Villanti A, et al. Misperceptions of harm among Natural American Spirit smokers: results from wave 1 of the Population Assessment of Tobacco and Health (PATH) Study (2013-2014). *Tob Control*. Mar 2017;26(e1):e61-e67. doi:10.1136/tobaccocontrol-2016-053265

45. Hills M, Armitage P. The two-period cross-over clinical trial. *Br J Clin Pharmacol*. Jul 1979;8(1):7-20. doi:10.1111/j.1365-2125.1979.tb05903.x
46. Grizzle JE. The Two-Period Change-Over Design and Its Use in Clinical Trials. *Biometrics*. 1965;21doi:doi:10.2307/2528104
47. Selwyn MR, Dempster AP, Hall NR. A Bayesian approach to bioequivalence for the 2 x 2 changeover design. *Biometrics*. 1981;37:11-21.
48. Grieve AP. A Bayesian analysis of the two-period crossover design for clinical trials. *Biometrics*. Dec 1985;41(4):979-90.
49. Raftery AE, Madigan D, Hoeting JA. Bayesian Model Averaging for Linear Regression Models. *J Amer Statistical Assoc*. 92(437):179-191.
50. Hinne M, Gronau QF, Bergh Dvd, Wagenmakers E-J. A Conceptual Introduction to Bayesian
Model Averaging. *Advances in Methods and
Practices in Psychological Science*. 2020;3(2):200-215.
51. Chung Y, Gelman A, Rabe-Hesketh S, Liu J, Dorie V. Weakly Informative Prior for Point Estimation of Covariance Matrices in Hierarchical Models. *Journal of Educational and Behavioral Statistics*. 2015;40(2):136-157. doi:10.3102/1076998615570945
52. Gelman A, Hill J, Yajima M. Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*. 2012;5:189-211.
53. Westfall PH. A Bayesian Perspective on the Bonferroni Adjustment. *Biometrika*. 1997;84(419 - 427)
54. Salaspuro V, Salaspuro M. Synergistic effect of alcohol drinking and smoking on in vivo acetaldehyde concentration in saliva. *Int J Cancer*. Sep 10 2004;111(4):480-3. doi:10.1002/ijc.20293
55. Shafagoj YA, Mohammed FI, Hadidi KA. Hubble-bubble (water pipe) smoking: levels of nicotine and cotinine in plasma, saliva and urine. *Int J Clin Pharmacol Ther*. Jun 2002;40(6):249-55. doi:10.5414/cpp40249