Harmonizing Cigar Survey Data Across Tobacco Centers of Regulatory Science, Center for Tobacco Products, and Population Assessment of Tobacco and Health Studies: The Cigar Collaborative Research Group

Howard Fishbein DrPH1, Dan Bauer PhD2, Qilu Yu PhD1, Robin Mermelstein PhD3, Dina Jones MPH4, Amy Miller MPH1, Melissa Harrell PhD5, Alexandra Loukas PhD6, Kymberle Sterling PhD7, Brittany Colip MS1, Beth Mittl BA1

1Public Health and Epidemiology Practice, Westat, Inc, Rockville, MD; 2Curran-Bauer Analytics, Chapel Hill, NC; 3Institute for Health Research and Policy, University of Illinois at Chicago, Chicago, IL; 4Tobacco Center of Regulatory Science, School of Public Health, Georgia State University, Atlanta, GA; 5School of Public Health, University of Texas Health Science Center at Houston, Houston, TX; 6Kinesiology and Health Education, University of Texas, Austin, TX; 7Health Promotion and Behavioral Sciences, University of Texas Health Science Center School of Public Health, Dallas, TX

Corresponding Author: Howard Fishbein, DrPH, Public Health and Epidemiology Practice, Westat, Inc, Rockville, MD 20850, USA. Telephone: (301) 517-8078; Fax: (240) 453-2681; E-mail: HowardFishbein@westat.com

Abstract

Introduction: Cigars are a popular tobacco product of choice for youth and young adults. Despite growing interest in cigar research, there are gaps in the available literature limiting an ability to set evidence-based policies. Too small research samples, the heterogeneity of types of cigars when asking a single question about use, makes analyzing data difficult. Given the Food and Drug Administration’s (FDA’s) authority granted in 2016 to regulate cigars, and its popularity, data to better understand use and preference for cigars will help FDA set appropriate regulatory policies.

Methods: We harmonized cigar survey data previously collected by five independent tobacco regulatory science survey research projects. Data supplying participants included three Tobacco Centers of Regulatory Science, one Center for Tobacco Products grantee, and data from Population Assessment of Tobacco and Health (PATH) study’s public use dataset.

Results: Analyzing 92 data variables from across five studies, and applying a rigorous data harmonization protocol, we report findings on 24 key cigar use variables. The step by step protocol for harmonizing is presented. Selected findings showing strict reproducibility across all five studies reveal youth 17–19 years at highest risk for cigar initiation; relative reproducibility shows males more likely to try cigars than females but with significant differences in magnitude across studies; and areas of inconsistent reproducibility are revealed when evaluating brand preferences.

Conclusion: Harmonizing data from multiple sources fosters a broader view of the robustness and generalizability of survey data than that from a single source. These observations raise awareness to look for the highest degree of reproducibility among and across data sources to inform policy.
Implications: Harmonizing data from discrete datasets provides insights into cigar initiation and use and is presented showing opportunities, challenges, and solutions. Comparing observational data from PATH and four independent research studies provides a best-practices approach and example of data synthesis for the tobacco research community. The dataset of five studies offers a look at the degree of confidence in analyzing harmonized survey results. Variable conclusions raise the need to strive for the highest degree of reproducibility, to best understand the behaviors of cigar users, and allow for the future development of the most effective interventions to alter tobacco use patterns.

Introduction
Cigar smoking is known to be associated with many negative health effects similar to those of cigarette smoking including oral, esophageal, laryngeal, and lung cancer and decreased lung function. Also known is that higher concentrations of nicotine, tobacco-specific nitrosamines, carbon monoxide, tar, and ammonia are found in cigar smoke than cigarette smoke. When compared with cigarette smokers, cigar smokers are found in more modest numbers in the general population. In 2017, an estimated 12.5 million people aged 12 or older in the United States were current cigar smokers; 3.1 million were young adults aged 18–25; as a comparison, 34 million adults were reported to be current cigarette smokers. The relatively lower prevalence of cigar smokers, compared with cigarette smokers, provides a rationale for harmonizing and analyzing data across multiple cigar studies. Benefits of harmonizing data can help ensure sufficient statistical power; more refined subgroup analysis; increased exposure heterogeneity; enhanced generalizability; and a capacity to undertake comparison, cross-validation, or replication across different datasets. Additionally, harmonizing previously collected data is efficient and potentially allows the advance of research beyond the intended scope of any one data collection effort.

This is especially true when harmonizing specific topics not often able to be studied in sufficiently large numbers. Limitations are inherent in using a harmonization approach, most notably, an inability to make changes, substitute, or significantly modify research questions and responses that had been originally asked. Accepting all these limitations, this study was conducted to understand what could be learned regarding cigar use analyzing a harmonized dataset of disparate surveys on cigars. Specific aims of this paper were two-fold: (1) to describe a novel analytic process for harmonizing cigar survey data applied to five diverse cigar survey datasets and (2) to interpret the analytic results of the harmonized cigar use data obtained from responses to the original five cigar surveys.

Methods
The Analysis and Synthesis Core of the Center for Evaluation and Coordination of Training and Research (CECTR) in Tobacco Regulatory Science announced an open call to researchers funded by the Tobacco Centers of Regulatory Science (TCORS) and/or who had a Center for Tobacco Products (CTP) grant for interest in joining a Cigar Collaborative study.

Three researchers joining were funded through the TCORS, one had a Food and Drug Administration (FDA) CTP grant, and data were obtained via the public use dataset of the CTP-funded Population Assessment of Tobacco and Health (PATH) study. The purpose of the Cigar Collaborative study was to leverage existing data collection of survey data on cigar use. The goal was to conduct integrative data analyses (IDA) comparing data collected across five independent studies. IDA allows for the statistical analysis of one dataset created from pooled multiple separate samples. IDA offers advantages including increased statistical power, sample heterogeneity, and support of data sharing and collaboration. A harmonized cigar survey dataset was created by working collaboratively with each individual researcher. A primary aim was to explore general factors that could address cigar use among racially ethnic diverse samples of youth and young adults. This was limited to using (1) the harmonized dataset and (2) responses to specific research questions originally asked by each of the surveys where responses could be harmonized. The research questions in this study were limited to what the original surveys had asked of their survey participants and where the research questions and responses were able to be harmonized across at least three of the five datasets.

Datasets
Supplementary Appendix Table 1A shows the characteristics of the study samples and designs.

Data Harmonization
An IDA requirement is for multiple separate samples to have common measures reflective of similar constructs, which allows for data harmonization. Therefore, we first coded all survey measures into different domains (e.g., demographic characteristics, current use, and flavors). Then, measures within each domain were compared to assess similarity in item wording and the construct assessed. (See Supplementary Appendix Table 2A) Following a review of 92 potential variables that could be pooled, investigators decided to create harmonized variables when three or more of the five questionnaires shared a common question. Consequently, the number of potential harmonized variables was reduced to 24. We developed a harmonized variable based on the lowest common denominator of all available questions. For example, all questionnaires asked a question about sex and asked it differently and coded it differently. We created a variable for sex based on a harmonized question and response option that included all possible responses from the different questionnaires. This approach was followed for all 24 harmonized variables. Table 1 provides the steps in creating the harmonized dataset.

Hypotheses and Research Questions
Once harmonized variables were created, we were able to answer the following research questions/hypotheses:

Question 1(a–e): for cigar demographics, products, and flavoring:
  a. Are racial/ethnic and sexual minorities likely to try cigars?
  b. Is young adulthood the period of greatest risk for trying cigars?
  c. Do racial/ethnic and sexual minorities differ in current/past-30-day cigar use rates?
Table 1. Steps to creating a harmonized dataset

| Step                                                                 | Description                                                                 |
|---------------------------------------------------------------------|----------------------------------------------------------------------------|
| Collect questionnaires from all researchers                          | Separate and categorize variables/questions and types of responses          |
| Count frequencies for each variable/question in a category           | Begin to harmonize the variables/questions by creating a harmonized code    |
| Create a spreadsheet to look at which variables/questions are similar | Create new variables using the harmonized code                              |
| Create a harmonized questionnaire that works for all surveys        | Make available the harmonization code for each investigator to apply to their dataset |
| Create a single harmonized dataset that now includes all study questionnaires | Run analyses on the harmonized dataset                                      |
|                                                                    |                                                                            |
| d. Are those who initiated with flavors more likely to continue to use cigars? |                                                                            |
| e. Among flavored cigar smokers, which cigar brand is most popular?  |                                                                            |

**Analysis Strategy**

We used a model-based inference approach to account for between-study heterogeneity. Each research question was addressed using a statistical model that includes Study as a fixed-factor. The main effect of Study absorbs any between-study differences in outcomes due to sampling design, region, period, and so forth. Interactions between study and other predictors (e.g., age, sex, and race) are tested to evaluate whether the predictor effects are robust across studies or vary significantly between studies.

Sensitivity analyses were conducted in which the PATH sample was thinned to a random subset of cases (comparable in size to the next largest study) to ensure the very large size of PATH did not have an overwhelming influence on final model estimates and conclusions. For certain research questions, some studies were omitted due to missing information on variables or a subset of data meeting a criterion (e.g., ever having tried cigars). Results are shown by the names of collaborating authors.

**Results**

**Research Question 1a: How Does the Likelihood of Ever Using a Cigar Vary Across Demographic Groups?**

Logistic regression models were fitted for the outcome of cigar use with the sociodemographic variable, allowing for the possibility that not only levels of cigar use but also sociodemographic differences in cigar use vary across studies. When this interaction was found significant, pairwise contrasts were used to localize study differences. Supplementary Appendix Table A3 presents results for study analyses by selected sociodemographic characteristics.

**Sex**

For sex, Model 2 was favored over Model 1 ($\chi^2 (3) = 34.75, p < .001$), indicating sex differences varied by study. In all studies, men were significantly more likely to have ever used a cigar than women (i.e., all odds ratios exceeded one and all confidence intervals excluded one). However, the magnitude of this gender difference varied across studies, with odds ratios ranging from 1.40 (Harrell) to 2.45 (PATH). Pairwise comparisons revealed a smaller sex difference in the Harrell and Loukas studies relative to PATH but no other significant differences between studies. Sensitivity analyses excluding Mermelstein and reducing the sample size for PATH did not alter any of these findings.

**Race**

For race, Model 2 was again favored over Model 1 ($\chi^2 (6) = 42.15, p < .001$), indicating race differences varied across studies. Pairwise contrasts indicated that Black versus White differences in trying cigars were consistent across Harrell, Loukas, and Mermelstein (showing no significant differences). In contrast, in the PATH study, African Americans were significantly less likely to have ever used cigars than Caucasians. This difference in effects was significant only between Harrell and PATH. We also compared differences between Multi/Other race/ethnicity versus Whites in trying cigars, and these differences in ever using cigars were consistent across Loukas, Mermelstein, and PATH, with all showing significantly reduced risk for Multi/Other versus White. The degree of reduction, however, differed significantly between Loukas and PATH and Mermelstein and PATH. In contrast, in Harrell, Multi/Other conferred a nonsignificant increase in risk relative to Whites. Pairwise contrasts showed the Multi/Other versus White effect to differ significantly between Harrell and all other studies. Sensitivity analyses excluding Mermelstein did not alter these conclusions. Upon reducing the sample size for PATH, the pairwise contrast for Multi/Other versus White between Mermelstein and PATH became nonsignificant.

**Ethnicity**

Model 2 was again favored over Model 1 ($\chi^2 (3) = 93.64, p < .001$), indicating differential effects across studies. Hispanic individuals were significantly more likely to have ever used cigars in Harrell and Loukas; no more likely to have ever used cigars in Mermelstein; and less likely to have ever used cigars in PATH. Pairwise contrasts showed the effect of being Hispanic in PATH to significantly differ from Harrell and Loukas. Sensitivity analyses excluding Mermelstein and reducing the size of the PATH sample did not alter findings.

**Sexual Orientation**

Neither Harrell nor Mermelstein included data on sex; hence, these studies were excluded from analyses for this variable. When fitting a model to the data, model 2 did not provide a significantly better fit to the data than model 1 when using all available cases ($\chi^2 (1) = 2.73, p = .10$) but was significant when using a random subsample...
of \( N = 4500 \) for the PATH study \( (\chi^2 (1) = 6.40, p = .01) \). Odds ratios indicate that lesbian, gay, and bisexual (LGB) status confers additional risk for ever using a cigar.

**Research Question 1b: Is Young Adulthood the Period of Greatest Risk for Initiation of Cigar Use?**

A survival analysis was conducted to determine the ages of highest risk for initiation of cigar use. The sample excluded individuals from PATH due to the lack of data on age of initiation. Inferential tests of between-study differences were conducted using the log-rank and Wilcoxon tests as implemented in the LIFETEST procedure in SAS. The log-rank test places more weight on later survival/failure times, whereas the Wilcoxon test places more weight on earlier survival/failure times. Pairwise contrasts were used to identify which studies differed from others.

As seen in Figure 1, initiation rates over time significantly differed between studies as indicated by the log-rank \( (\chi^2 (3) = 423.80, p < .001) \) and Wilcoxon \( (\chi^2 (3) = 398.48, p < .001) \) tests of equality. The Mermelstein and Sterling studies produced similar curves with high cumulative risk for cigar initiation. In contrast, the Loukas and Harrell samples displayed lower cumulative risk for initiation. Curves did not significantly differ between Loukas and Harrell but all other pairwise comparisons were significant.

Similar findings were also found when we computed Hazard curves. Figure 2 shows the risk of initiation by differences in current use patterns through adolescence, peaking in most studies between ages 17 and 19. In Harrell, the highest risk interval was 17, 18; in Loukas and Mermelstein, it was 18, 19. In contrast, in Sterling, the highest risk for new initiations occurred in the interval 21, 22; however, the second highest peak was observed in the 18, 19 interval.

**Research Question 1c: Do Racial/Ethnic and Sexual Minorities Differ in Current/Past-30-Day Cigar Rates?**

Past-30-day cigar use frequency was originally harmonized to four categories but given sparseness in the uppermost category (all 30 days); the two highest categories were collapsed. The criterion variable for analyses is a three-level ordinal response of 0, 1–5, and 6–30 days. Using an ordered logit model, data analysis proceeded with model 1 including main effects only and model 2 including the interaction between study and the sociodemographic variable.

Supplementary Appendix Table A4 presents odds ratios and confidence intervals for comparisons across sex, race, ethnic, and sexual orientation in past-30-day cigar use.

**Sex**

Model 2 was favored over model 1 \( (\chi^2 (4) = 19.45, p < .001) \), indicating across-study variation in sex differences. In Harrell, Mermelstein, and PATH, rates did not show a significant difference between men and women. In contrast, Loukas and PATH showed significantly elevated levels of use among men compared to women. The effect observed in Loukas was significantly higher than all studies other than Sterling, and the effect observed in Harrell was significantly lower than either Loukas or PATH. When the PATH sample was thinned to \( N = 2000 \), the comparison between Loukas and PATH became nonsignificant.

**Race**

Model 2 was favored over model 1 \( (\chi^2 (8) = 24.86, p = .002) \), indicating racial differences varied significantly between studies. Among those who had previously tried cigars, current use levels for Blacks significantly exceeded that of Whites in three samples: Mermelstein, PATH, and Sterling. The other two samples, Harrell and Loukas, showed a smaller increase in use among Blacks, not statistically significant. Pairwise comparisons indicated that Black versus White differences were significantly smaller in Loukas than in Mermelstein, PATH, or Sterling. Black versus White differences were significantly smaller in Harrell than Sterling.

Current use levels of Multi/Other individuals did not significantly differ from Whites in four of five studies, with the exception being PATH. In sensitivity analyses with PATH thinned to \( N = 2000 \), the Multi/Other versus White odds ratio for PATH became nonsignificant, as did the pairwise contrast between PATH and Harrell for the same effect. No other findings differed between the full and reduced sample.
Ethnicity
Model 1 was not rejected in favor of model 2, indicating ethnicity effects were comparable across studies. Using all available cases, the odds ratio obtained from model 1 for Hispanic versus non-Hispanic was 1.03, with a 95% confidence interval from 0.92 to 1.15. This result indicates there are no significant differences in current cigar use levels between Hispanic and non-Hispanic individuals (among those who have ever smoked a cigar).

Sexual Orientation
For sex, Harrell and Mermelstein studies were excluded as sexual orientation information was not available. Using three studies, Model 1 was not rejected in favor of model 2 ($\chi^2 (2) = 0.73, p = .69$), indicating consistent effects of LGB status. Pooling over Loukas, PATH, and Sterling, the odds ratio obtained from Model 1 (for LGB vs. Straight) was 1.21, with a 95% confidence interval from 1.05 to 1.41. This result indicates current use rates are significantly higher among LGB individuals.

Research Question 1d: Are Those Who Initiated With Flavors More Likely to Continue to Use Cigars?
The sample was restricted to those individuals who had ever engaged in cigar use. The data analysis strategy for testing this hypothesis was identical to that described for Research Question 1c fitting ordered logit models to evaluate whether past-30-day use rates varied between those initiating with flavored cigars versus not, with tests for study differences in the magnitude of this effect. We again collapsed the upper two categories of past-30-day use rates to create a three-level ordinal outcome.

Model 2 was favored over Model 1 ($\chi^2 (4) = 23.74, p < .001$), indicating inconsistent effects across studies. As shown in Supplementary Appendix Table A5, initiating with a flavored product was associated with increased levels of current use observed in Mermelstein and PATH, lower levels of current use observed in Loukas, and no significant difference in Harrell or Sterling.

Pairwise comparisons identified significant differences in the effect of initiating with flavored products on current use levels between Loukas and Mermelstein, Loukas and PATH, Mermelstein and Sterling, and Sterling and PATH. None of these findings differed in sensitivity analyses conducted with the PATH sample thinned to a random subset of $N = 2000$ individuals.

Research Question 1e: Among Flavored Cigar Smokers, Which Cigar Brand is Most Popular?
The sample was restricted to individuals who indicated they usually flavored cigar products. The sample excluded individuals indicating “Unknown” or “I didn’t smoke in the past 30 days.” Given sparse endorsement of brands besides “Any Black and Milds” and “Any Swisher Sweets,” other brands/responses were collapsed into “Other.” Multinomial logistic regression models were applied to evaluate the relative proportions of Black and Mild users, Swisher Sweets users, and other brand users across studies. Model 1 assumed no study effects, whereas model 2 allowed for study differences in brand preference. Given significant study differences, we conducted pairwise contrasts. In sensitivity analyses, the PATH sample was thinned to a random subset of 160 people to be similar in size to the next largest study (Loukas).

Model 1 was rejected in favor of model 2 ($\chi^2 (8) = 42.87, p < .001$), indicating brand preference among flavored cigar smokers was inconsistent across studies. For instance, Swisher Sweets were more popular than Black and Mild in Harrell, Loukas, and Mermelstein, but the reverse preference was observed in PATH and Sterling. Odds ratios and confidence intervals are shown in Figure 3.

The following pairwise comparisons were significant, indicating differences in brand preference between studies: Harrell versus Loukas, Harrell versus Mermelstein, Harrell versus Sterling, Harrell versus PATH, Loukas versus Sterling, and Loukas versus PATH. Pairwise comparisons not significant were: Loukas versus Mermelstein, Mermelstein versus Sterling, Mermelstein versus PATH, and Sterling versus PATH. The Loukas versus PATH and Sterling versus PATH pairwise comparisons became nonsignificant when conducted with the reduced PATH sample.

Figure 3. Pairwise comparison odds ratios for preferred cigar brand for Cigar Collaborative studies.
Research Question 2: Are Those Cigar Users Who Felt Like They Needed a Cigar More Likely to be Past 30 Days/Current Cigar Smokers?

The sample was restricted to individuals who had ever engaged in cigar use. Additionally, only Harrell, Loukas, and Mermelstein studies provided data on the predictor of ever feeling “like you needed a cigar”; thus, PATH and Sterling studies were excluded from analyses. The data analysis strategy for testing this hypothesis was identical to Research Question 1c and involved fitting cumulative logit models to evaluate whether past-30-day use rates varied between those who indicated they ever felt they “needed” a cigar versus not.

Model 2 provided a better fit to the data than model 1 ($\chi^2 (2) = 44.71, p < .001$), indicating the effect of needing a cigar varied significantly between studies. Odds ratios obtained from model 2 are shown in the Supplementary Appendix Table A6.

Pairwise contrasts showed the effect of ever feeling like one needed a cigar on current cigar use was significantly stronger in Loukas than either Harrell or Mermelstein. Harrell and Mermelstein did not significantly differ from one another for this predictor.

Discussion

A primary goal of this study was to demonstrate the analytic process of creating a harmonized dataset to address research questions that may elude any one individual investigator due to a relatively small sample size or data collection sampling approach. The research questions examined in this study were dependent on the ability to harmonize questions asked by investigators in five different settings and an agreement that, for three of the five studies, questions and responses were similar enough to harmonize. The findings across harmonized studies report: males being more likely to have ever used cigars; a marked variability in past 30-day cigar use by sex, race, ethnicity, and sexual orientation; some consistency for flavors being used during initiation of cigar use; and mixed results as to whether feeling like a cigar was needed. Collectively, these findings point toward the importance of mediating factors in the setting where data were collected and that the characteristics of a single dataset must be looked at carefully when data are to be generalized. Given this cautionary note, the ability to collaborate and combine data across studies may provide new avenues for more timely and cost-effective ways to answer research questions taking advantage of existing data collection efforts. A major strength of this harmonized work is the built-in check for robustness across studies. Analyzing harmonized variables across diverse studies provides insights and a degree of confidence to explore findings from multiple studies all investigating the same relationships. Where we see reproducibility among these studies, the results of analyses are compelling. Where we notice a variation in the findings or in the strength of findings from a single study, the findings may not be as definitive. Where we see consistency across all studies, such as for cigar use initiation, peaking between 17 and 19 years for cigar use initiation, we can be highly confident in this finding.

Where we do not see this type of agreement in analyses of data, this indicates the presence of moderating factors that differ across studies that could account for observed differences. These potential moderators, including, for example, time and place of measurement, variation in measurement tools, and sample selection differences, are areas for further investigation beyond this study.

In summary, findings reported in this study fall into areas of strict reproducibility, relative reproducibility, and inconsistency. Strict reproducibility implies no significant differences between studies as seen with youth 17–19 years being most at increased risk of cigar initiation. Relative reproducibility may include, for example, significant differences in magnitude across studies but with effects all in the same direction as was observed with males more likely to try cigars than females. Inconsistent reproducibility reveals effects not all in the same direction as observed when evaluating brand preference. These observations raise the need for understanding the behaviors of cigar users to be able to tailor effective interventions to change use patterns.

Potential moderators, as noted above, are possible confounders in interpreting findings. Perhaps, in future prospective studies on cigar use, such moderators could be adequately addressed.

In addition to collecting data on potential moderators, when researching cigars, it is a simple but important request to separate questions regarding the use of big cigars, little cigars, and cigarillos when asking participants about these products. When a study does not differentiate between types of cigars, there is no opportunity to harmonize data from that study with other surveys where such data are collected separately. In this study, the five researchers agreed we would harmonize and report out on data where we had a minimum of three of the five surveys being able to be harmonized on a research question and on responses. Likewise, whenever possible, the use of common measures when developing data collection instruments on cigars, as well as on other tobacco products is encouraged. Look to already existing data resources (e.g., PhenX PROMIS) when designing data collection instruments. These are catalogues of high-priority measures for consideration in research studies.

Researchers should continue to search for opportunities to share data that has been collected (but not analyzed) with others who may be in an identical situation or simply have an interest in the same topic. Further comparative analyses of studies, where similar data have been collected from a variety of populations and where data have been harmonized and findings are found to be consistent, will add to the literature and help inform FDA in their development of policies on cigars, with a high level of confidence.

Finally, while standardization has its benefits, we recognize measurement standardization is not always possible or always desirable. For instance, as new tobacco-related products and behaviors change, measures will need to change to account for new activities. A decade ago, people were not asking about e-cigarettes. Measurement of poly use of tobacco products is another example where new approaches are needed to assess the use of multiple products and effects on outcomes. Innovation in measurement should not be discouraged, but investigators should be encouraged to include some standardized measures and link their findings with other studies as appropriate using measures that could be readily harmonized, with more conventional assessments, allowing for greater ease in aggregation and interpretation of data.

Supplementary Material

Supplementary data are available at Nicotine and Tobacco Research online.

Funding

Research reported in this study was supported by the following grants and agencies: grants 1U54CA189222 and 1P50 CA180906 from the National
Declaration of Interest

None declared.

References

1. Nonnemaker J, Rostron B, Hall P, MacMonegle A, Apelberg B. Mortality and economic costs from regular cigar use in the United States, 2010. Am J Public Health. 2014;104(9):e86–e91.
2. National Cancer Institute. Smoking and Tobacco Control, Monograph 9: Cigars Health Effects and Trends. Bethesda, MD: US Department of Health and Human Services, Public Health Service; 1998.
3. Rodriguez J, Jiang R, Johnson WC, MacKenzie BA, Smith LJ, Barr RG. The association of pipe and cigar use with cotinine levels, lung function, and airflow obstruction: A cross-sectional study. Ann Intern Med. 2010;152(4):201–210.
4. Baker F, Ainsworth SR, Dye JT, et al. Health risks associated with cigar smoking. JAMA. 2000;284(6):735–740.
5. Shapiro JA, Jacobs EJ, Thun MJ. Cigar smoking in men and risk of death from tobacco-related cancers. J Natl Cancer Inst. 2000;92(4):333–337.
6. U.S. Department of Health and Human Services. The Health Consequences of Smoking: 50 Years of Progress. A Report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2014.
7. Center for Behavioral Health Statistics and Quality. 2017 National Survey on Drug Use and Health: Detailed Tables. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2018.
8. Fortier I, Raina P, Van den Heuvel ER, et al. Maelstrom research guidelines for rigorous retrospective data harmonization. Int J Epidemiol. 2017;46(1):103–105.
9. Gallacher JE. The case for large scale fungible cohorts. Eur J Public Health. 2007;17(6):548–549.
10. Gallacher J, Hofer SM. Generating large-scale longitudinal data resources for aging research. J Gerontol B Psychol Sci Soc Sci. 2011;66 (Suppl 1):i172–i179.