

**Automated Content Analysis to Quantify Valence, Themes, and Norms Mentioned in
Tobacco and E-cigarette Coverage across Four Media Sources**

L.A. Gibson¹, S.E. Williams², L.N. Siegel², C.R. Lee², A.C. Volinsky², and R.C. Hornik²

This file contains headers that may not be able to be expanded in some versions of Word (e.g., for Word 2016 for Mac). If unable to expand headers, please refer to the version of this document with the headers already expanded.

Please note: The media examples in this document were randomly selected and their content may be offensive, explicit and/or not safe for work.

This research was supported by the National Cancer Institute (NCI) of the National Institutes of Health (NIH) and FDA Center for Tobacco Products (CTP) under Award Number P50CA179546 (the University of Pennsylvania Tobacco Center of Regulatory Science (UPenn TCORS)). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the Food and Drug Administration (FDA).

¹ Medical Ethics and Health Policy, Perelman School of Medicine at the University of Pennsylvania

² Annenberg School for Communication, University of Pennsylvania

Automated Content Analysis to Quantify Valence, Themes, and Norms Mentioned in Tobacco and E-cigarette Coverage across Four Media Sources

In this study, we used supervised machine learning to measure the coverage of valence, tobacco control themes, and norms in (1) passing mentions (PM), as well as more than passing mentions (MPM) of (2) tobacco and e-cigarette products in (3) long-form media sources (Broadcast news transcripts, the AP, Newspapers, and Websites with RSS feeds). Texts were collected from May 18, 2014 – December 31, 2017. The following example texts were pulled from the 135,764 relevant texts collected through June 30, 2017.

Corpus

We first searched the Lexis-Nexis and MIT MediaCloud databases for texts that included a range of tobacco-related terms (e.g., “smoking”, “tobacco”, “smokes”, “e-cigarette”, “vaping”, “nicotine”, “hookah”, “cigar”). This was a broad search meant to capture the majority of tobacco-related content. We then used automated coding methods with validation to clean the database more precisely to tobacco-related content. Texts came from:

- (1) Broadcast TV and radio news transcripts (**BTN**)
ABC, CBS, NBC, PBS, FOX News Network, CNN, MSNBC, and NPR
n=4,275, 8% e-cigarettes
- (2) AP Newswire (**AP**)
national, state and local
n=8,522, 13% e-cigarettes
- (3) 50 U.S. English-language newspapers with the highest circulation (**Newspaper**)
n=52,561, 7% e-cigarettes
- (4) 50 most popular websites each quarter among 12-17 and 18-24 year olds (**Web**)
Popularity assessed by Nielsen, collected using the MIT MediaCloud database, limited to websites that had RSS feeds and primarily consisted of text
n=70,406, 8% e-cigarettes

Passing Mentions (PM) versus More than Passing Mentions (MPM)

Prior hand-coded content analysis of similar sources have only included MPM texts. They typically required a certain amount of coverage of the topic (e.g., text focused on the topic in the 1st paragraph plus one other, or in 50% of paragraphs) or pre-selected texts that would be potentially relevant (e.g., TV health segments). We include both PM and MPM texts in this analysis. MPM texts had at least three tobacco or e-cigarette-related keywords within 100 words (i.e., the low-end of an abstract paragraph length, Hacker & Sommers, 2010). This is a rough approximation of a paragraph focused on the topic. We also required that the three keywords could not be immediately next to one another, operationalized as having at least 20 words between the first and last keyword (i.e., the upper-end of sentence length, Cutts, 2013). Finally, since titles summarize content, one keyword in the title of a text also counted as an MPM text. On average across sources, e-cigarette coverage was 67% MPM, while coverage of other tobacco products was only 16% MPM.

Tobacco (no ecig) versus E-cigarettes (ecig)

We define e-cigarettes as electronic devices typically meant for nicotine delivery though they could be used with or without nicotine. These are relatively new products relative to other tobacco products. Any texts which contain a mention of e-cigarettes are coded as (ecig). Our definition of

tobacco (no ecig) products includes cigarettes, cigars, cigarillos, little cigars, smokeless tobacco, and hookah – all non-ecig products used without marijuana. It is possible that there are some mentions of these other tobacco products in the texts which contain e-cigarette mentions. However, there are relatively few mentions of e-cigarettes in the corpus, therefore, we decided to count both texts only about e-cigarettes and texts about e-cigarettes and other products as “ecig”.

Valence, Themes, and Norms coded

We used supervised machine learning to measure the coverage of valence, tobacco control themes, and norms in this corpus of texts. Mturk workers coded samples of ~2400 texts which were split into training and test sets for developing automatic algorithms. Reliability was $>.75$ for themes, $>.65$ for norms, and $>.53$ for valence.

Valence was coded separately for e-cigarettes and other tobacco products. If texts contained both types of products, they were coded for both valences. Texts that on the whole were mostly supportive of using the product or the industry making the product were coded “Pro”, while texts that were mostly against the use of the product or the industry making the product were coded “Anti”. The correlations between the predicted probabilities yielding from valence automated coding and the Mturker ratings in held-aside test sets were good, ranging .79 to .88.

Four tobacco control themes were chosen *a priori* based on prior tobacco content analyses. We expected these themes to be present in the texts in fairly high numbers, and that they would likely have an impact on the general public’s beliefs about tobacco products. (1) “Health”: effects of product use on the user’s physical health (specifically excluding health effects of use for non-users (e.g., secondhand smoke)), (2) “Policy”: mandatory policy/law/regulation by a government, company, or institution, (3) “Addiction”: explicit references to products being addictive or users being addicted, and (4) “Youth”: use, access, or purchase of products by anyone up to 21 years old. The correlations between the predicted probabilities yielding from theme automated coding and the Mturker ratings in held-aside test sets were high, ranging .87 to .90.

Like valence, normative information was also coded separately for e-cigarettes and other tobacco products. In contrast to valence, if texts contained both e-cigarettes and other tobacco products, they were only coded for e-cigarette norms. Normative information was divided into two types: “Population” and “Individual”. Population-level summary information (“population norms”) conveys behavior prevalence in an explicit and straightforward way by providing numbers, percentages or trends from census data, survey results, reports, or educational campaigns. Population norms were coded for whether or not they were explicitly increasing, or described as high (“high”) or decreasing, or described as low (“low”). There were not enough examples in our hand-coded samples to build classifiers for Ecig Population Norms - Low, so the database was not coded for that. Individual-level behavior cues (“individual norms”), on the other hand, facilitate formation of descriptive norm perceptions more subtly by providing knowledge about and exposure to instances of others’ performing (or not performing) the behavior. Individual norms were coded for whether or not they described “use” (as opposed to a reduction in use, or non-use). There were not enough examples in our hand-coded samples to build classifiers for non-use, so the database was not coded for that. The same text could be coded for multiple types of norms (i.e., population-high, population-low, individual). The correlations between the predicted probabilities yielding from theme automated coding and the Mturker ratings in held-aside test sets were fairly good, ranging .63 to .83.

Current Set of Examples

In all there are 13 different codes. In order to have some illustrative examples of the texts that were coded, we pulled 10 examples for each valence, theme, and norm. These sets of 10 were pulled separately for MPM/PM texts and tobacco (no ecig)/ecig texts if they existed for that code. Also, in each set of 10, we tried to choose 2 examples each of AP, BTN, and Newspaper texts and 4 examples from the Web. Some combinations were not available. In all there are 320 examples included here. Individual texts were randomly selected from all texts matching those criteria with predicted probabilities between 0.8 and 0.9. If no texts met those requirements, we relaxed the probabilities range to be from 0.6 to 1.0. If no texts met the requirements for a particular source, we randomly selected from the other sources matching the parameters.

Hand-coding of the examples does not perfectly match the machine coding. This is expected because the test set correlations checking the validity of the automated coding were not perfect (see above). Also, some sets don't match for themes because the validity was tested across products (e.g., ecig addiction and other tobacco products youth have several mistakes because the theme refers to a different product than the one indicated). **Rather than the definitions of each code, the examples provided here are meant to give the reader a sense of the actual texts included for each code.** We have included the team's hand-coding of whether or not these examples match the code (Yes or No) before each example. Effects analyses aggregate the data to at least the day-level ensuring that the amount of valence, theme, and norm coverage is estimated from many individual texts, thus we are less concerned about individual texts being misclassified.

Table of Contents

How to use these examples	7
Tobacco (no ecig) Pro.....	8
Tobacco (no ecig) texts	8
More than Passing Mentions	8
E-cigarette texts	8
More than Passing Mentions	8
Tobacco (no ecig) Anti	8
Tobacco (no ecig) texts	8
More than Passing Mentions	8
E-cigarette texts	9
More than Passing Mentions	9
Ecig Pro.....	9
E-cigarette texts	9
More than Passing Mentions	9
Ecig Anti	9
E-cigarette texts	9
More than Passing Mentions	9

Health.....	10
Tobacco (no ecig) texts	10
More than Passing Mentions	10
Passing Mentions	10
E-cigarette texts	11
More than Passing Mentions	11
Passing Mentions	11
Policy.....	11
Tobacco (no ecig) texts	11
More than Passing Mentions	11
Passing Mentions	12
E-cigarette texts	12
More than Passing Mentions	12
Passing Mentions	12
Addiction.....	13
Tobacco (no ecig) texts	13
More than Passing Mentions	13
Passing Mentions	13
E-cigarette texts	13
More than Passing Mentions	13
Passing Mentions	14
Youth.....	14
Tobacco (no ecig) texts	14
More than Passing Mentions	14
Passing Mentions	14
E-cigarette texts	15
More than Passing Mentions	15
Passing Mentions	15
Tobacco (no ecig) Population Norms – Low	15
Tobacco (no ecig) texts	15
More than Passing Mentions	15
Passing Mentions	16
Tobacco (no ecig) Population Norms – High.....	16

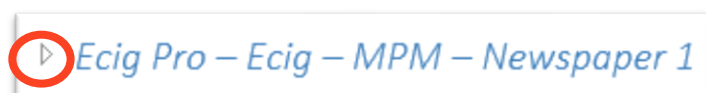
Tobacco (no ecig) texts	16
More than Passing Mentions	16
Passing Mentions	16
Ecig Population Norms – High.....	17
E-cigarette texts	17
More than Passing Mentions	17
Passing Mentions	17
Tobacco (no ecig) Individual Norms.....	17
Tobacco (no ecig) texts	18
More than Passing Mentions	18
Passing Mentions	18
Ecig Individual Norms.....	18
E-cigarette Texts	18
More than Passing Mentions	18
Passing Mentions	19

How to use these examples

This document uses Word's heading styles to make the examples easier to navigate. Each heading style can be opened or closed to reveal content lower down. All examples are closed by default when opening the document. No matter which examples you open, they will be closed the next time you open Word. Instructions for expanding all headings are available here: <https://support.office.com/en-us/article/collapse-or-expand-parts-of-a-document-701786e0-95e2-40bf-bfe5-f0233cd3520c>

To open an example:

- Hover your cursor over the blue labels and an open triangle will appear to the left of each label.



- Click the open triangle to open the text. Each text has key tobacco-related words highlighted in yellow. (Note “smoke” and “smoked” were not used to determine MPM vs. PM texts because there were too many false positives for burning buildings and other burning materials.)

Ecig Pro – Ecig – MPM – Newspaper 1

Predicted Probability: 0.81

ArticleID: 21850

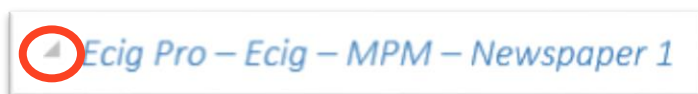
Article Title: Study finds e-cigarettes helped smokers to quit"

Article Content: Study finds e-cigarettes helped smokers to quit NEW YORK - A large British study involving 6,000 smokers has found that those trying to quit were substantially more likely to succeed if they used electronic cigarettes rather than over-the-counter therapies such as nicotine patches or gum. Story on A7."

Source Title: news

To close an example:

- Hover your cursor over the blue labels again and a closed triangle will appear to the left of each label. Click the closed triangle to close the example. You can also close entire sections using the same method.



Tobacco (no ecig) Pro

Note: For analysts of the dataset, the code for Tobacco (no ecig) Pro within the database is “cprotob2”.

Tobacco (no ecig) texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “tobnoecigmpm_SOURCE_tpro”.

Tobacco (no ecig) Pro – Tobacco (no ecig) – MPM – AP 1
Tobacco (no ecig) Pro – Tobacco (no ecig) – MPM – AP 2
Tobacco (no ecig) Pro – Tobacco (no ecig) – MPM – BTN 1
Tobacco (no ecig) Pro – Tobacco (no ecig) – MPM – BTN 2
Tobacco (no ecig) Pro – Tobacco (no ecig) – MPM – Newspaper 1
Tobacco (no ecig) Pro – Tobacco (no ecig) – MPM – Newspaper 2
Tobacco (no ecig) Pro – Tobacco (no ecig) – MPM – Web 1
Tobacco (no ecig) Pro – Tobacco (no ecig) – MPM – Web 2
Tobacco (no ecig) Pro – Tobacco (no ecig) – MPM – Web 3
Tobacco (no ecig) Pro – Tobacco (no ecig) – MPM – Web 4

E-cigarette texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “ecigmpm_SOURCE_tpro”.

Tobacco (no ecig) Pro – Ecig – MPM – AP 1
Tobacco (no ecig) Pro – Ecig – MPM – BTN 1
Tobacco (no ecig) Pro – Ecig – MPM – Newspaper 1
Tobacco (no ecig) Pro – Ecig – MPM – Newspaper 2
Tobacco (no ecig) Pro – Ecig – MPM – Newspaper 3
Tobacco (no ecig) Pro – Ecig – MPM – Web 1
Tobacco (no ecig) Pro – Ecig – MPM – Web 2
Tobacco (no ecig) Pro – Ecig – MPM – Web 3
Tobacco (no ecig) Pro – Ecig – MPM – Web 4
Tobacco (no ecig) Pro – Ecig – MPM – Web 5

Tobacco (no ecig) Anti

Note: For analysts of the dataset, the code for Tobacco (no ecig) Anti within the database is “cantitob2”.

Tobacco (no ecig) texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “tobnoecigmpm_SOURCE_tanti”.

Tobacco (no ecig) Anti – Tobacco (no ecig) – MPM – AP 1
Tobacco (no ecig) Anti – Tobacco (no ecig) – MPM – AP 2
Tobacco (no ecig) Anti – Tobacco (no ecig) – MPM – BTN 1
Tobacco (no ecig) Anti – Tobacco (no ecig) – MPM – BTN 2

Tobacco (no ecig) Anti – Tobacco (no ecig) – MPM – Newspaper 1

Tobacco (no ecig) Anti – Tobacco (no ecig) – MPM – Newspaper 2

Tobacco (no ecig) Anti – Tobacco (no ecig) – MPM – Web 1

Tobacco (no ecig) Anti – Tobacco (no ecig) – MPM – Web 2

Tobacco (no ecig) Anti – Tobacco (no ecig) – MPM – Web 3

Tobacco (no ecig) Anti – Tobacco (no ecig) – MPM – Web 4

E-cigarette texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “ecigmpm_SOURCE_tanti”.

Tobacco (no ecig) Anti – Ecig – MPM – AP 1

Tobacco (no ecig) Anti – Ecig – MPM – AP 2

Tobacco (no ecig) Anti – Ecig – MPM – BTN 1

Tobacco (no ecig) Anti – Ecig – MPM – BTN 2

Tobacco (no ecig) Anti – Ecig – MPM – Newspaper 1

Tobacco (no ecig) Anti – Ecig – MPM – Newspaper 2

Tobacco (no ecig) Anti – Ecig – MPM – Web 1

Tobacco (no ecig) Anti – Ecig – MPM – Web 2

Tobacco (no ecig) Anti – Ecig – MPM – Web 3

Tobacco (no ecig) Anti – Ecig – MPM – Web 4

Ecig Pro

Note: For analysts of the dataset, the code for Ecig Pro within the database is “cproecig2”.

E-cigarette texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “ecigmpm_SOURCE_epro”.

Ecig Pro – Ecig – MPM – AP 1

Ecig Pro – Ecig – MPM – AP 2

Ecig Pro – Ecig – MPM – BTN 1

Ecig Pro – Ecig – MPM – BTN 2

Ecig Pro – Ecig – MPM – Newspaper 1

Ecig Pro – Ecig – MPM – Newspaper 2

Ecig Pro – Ecig – MPM – Web 1

Ecig Pro – Ecig – MPM – Web 2

Ecig Pro – Ecig – MPM – Web 3

Ecig Pro – Ecig – MPM – Web 4

Ecig Anti

Note: For analysts of the dataset, the code for Ecig Anti within the database is “cantiecg2”.

E-cigarette texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “ecigmpm_SOURCE_eanti”.

Ecig Anti – Ecig – MPM – AP 1
Ecig Anti – Ecig – MPM – AP 2
Ecig Anti – Ecig – MPM – BTN 1
Ecig Anti – Ecig – MPM – BTN 2
Ecig Anti – Ecig – MPM – Newspaper 1
Ecig Anti – Ecig – MPM – Newspaper 2
Ecig Anti – Ecig – MPM – Web 1
Ecig Anti – Ecig – MPM – Web 2
Ecig Anti – Ecig – MPM – Web 3
Ecig Anti – Ecig – MPM – Web 4

Health

Note: For analysts of the dataset, the code for Health within the database is “chealth8”.

Tobacco (no ecig) texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “tobnoecigmpm_SOURCE_heal”.

Health – Tobacco (no ecig) – MPM – AP 1
Health – Tobacco (no ecig) – MPM – AP 2
Health – Tobacco (no ecig) – MPM – BTN 1
Health – Tobacco (no ecig) – MPM – BTN 2
Health – Tobacco (no ecig) – MPM – Newspaper 1
Health – Tobacco (no ecig) – MPM – Newspaper 2
Health – Tobacco (no ecig) – MPM – Web 1
Health – Tobacco (no ecig) – MPM – Web 2
Health – Tobacco (no ecig) – MPM – Web 3
Health – Tobacco (no ecig) – MPM – Web 4

Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is (“tobnoecig_SOURCE_heal” – “tobnoecigmpm_SOURCE_heal”).

Health – Tobacco (no ecig) – PM – AP 1
Health – Tobacco (no ecig) – PM – AP 2
Health – Tobacco (no ecig) – PM – BTN 1
Health – Tobacco (no ecig) – PM – BTN 2
Health – Tobacco (no ecig) – PM – Newspaper 1
Health – Tobacco (no ecig) – PM – Newspaper 2
Health – Tobacco (no ecig) – PM – Web 1
Health – Tobacco (no ecig) – PM – Web 2
Health – Tobacco (no ecig) – PM – Web 3
Health – Tobacco (no ecig) – PM – Web 4

E-cigarette texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “ecigmpm_SOURCE_he”.

Health – Ecig – MPM – AP 1

Health – Ecig – MPM – AP 2

Health – Ecig – MPM – BTN 1

Health – Ecig – MPM – BTN 2

Health – Ecig – MPM – Newspaper 1

Health – Ecig – MPM – Newspaper 2

Health – Ecig – MPM – Web 1

Health – Ecig – MPM – Web 2

Health – Ecig – MPM – Web 3

Health – Ecig – MPM – Web 4

Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is (“ecig_SOURCE_he” – “ecigmpm_SOURCE_he”).

Health – Ecig – PM – AP 1

Health – Ecig – PM – BTN 1

Health – Ecig – PM – BTN 2

Health – Ecig – PM – Newspaper 1

Health – Ecig – PM – Newspaper 2

Health – Ecig – PM – Web 1

Health – Ecig – PM – Web 2

Health – Ecig – PM – Web 3

Health – Ecig – PM – Web 4

Health – Ecig – PM – Web 5

Policy

Note: For analysts of the dataset, the code for Policy within the database is “cpolicy8”.

Tobacco (no ecig) texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “tobnoecigmpm_SOURCE_pol”.

Policy – Tobacco (no ecig) – MPM – AP 1

Policy – Tobacco (no ecig) – MPM – AP 2

Policy – Tobacco (no ecig) – MPM – BTN 1

Policy – Tobacco (no ecig) – MPM – BTN 2

Policy – Tobacco (no ecig) – MPM – Newspaper 1

Policy – Tobacco (no ecig) – MPM – Newspaper 2

Policy – Tobacco (no ecig) – MPM – Web 1

Policy – Tobacco (no ecig) – MPM – Web 2

Policy – Tobacco (no ecig) – MPM – Web 3

Policy – Tobacco (no ecig) – MPM – Web 4

Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is (“tobnoecig_SOURCE_pol” – “tobnoecigmpm_SOURCE_pol”).

Policy – Tobacco (no ecig) – PM – AP 1

Policy – Tobacco (no ecig) – PM – AP 2

Policy – Tobacco (no ecig) – PM – BTN 1

Policy – Tobacco (no ecig) – PM – BTN 2

Policy – Tobacco (no ecig) – PM – Newspaper 1

Policy – Tobacco (no ecig) – PM – Newspaper 2

Policy – Tobacco (no ecig) – PM – Web 1

Policy – Tobacco (no ecig) – PM – Web 2

Policy – Tobacco (no ecig) – PM – Web 3

Policy – Tobacco (no ecig) – PM – Web 4

E-cigarette texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “ecigmpm_SOURCE_pol”.

Policy – Ecig – MPM – AP 1

Policy – Ecig – MPM – AP 2

Policy – Ecig – MPM – BTN 1

Policy – Ecig – MPM – BTN 2

Policy – Ecig – MPM – Newspaper 1

Policy – Ecig – MPM – Newspaper 2

Policy – Ecig – MPM – Web 1

Policy – Ecig – MPM – Web 2

Policy – Ecig – MPM – Web 3

Policy – Ecig – MPM – Web 4

Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is (“ecig_SOURCE_pol” – “ecigmpm_SOURCE_pol”).

Policy – Ecig – PM – AP 1

Policy – Ecig – PM – AP 2

Policy – Ecig – PM – BTN 1

Policy – Ecig – PM – BTN 2

Policy – Ecig – PM – Newspaper 1

Policy – Ecig – PM – Newspaper 2

Policy – Ecig – PM – Web 1

Policy – Ecig – PM – Web 2

Policy – Ecig – PM – Web 3

Policy – Ecig – PM – Web 4

Addiction

Note: For analysts of the dataset, the code for Addiction within the database is “caddict8”.

Tobacco (no ecig) texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “tobnoecigmpm_SOURCE_add”.

Addiction – Tobacco (no ecig) – MPM – AP 1

Addiction – Tobacco (no ecig) – MPM – AP 2

Addiction – Tobacco (no ecig) – MPM – BTN 1

Addiction – Tobacco (no ecig) – MPM – BTN 2

Addiction – Tobacco (no ecig) – MPM – Newspaper 1

Addiction – Tobacco (no ecig) – MPM – Newspaper 2

Addiction – Tobacco (no ecig) – MPM – Web 1

Addiction – Tobacco (no ecig) – MPM – Web 2

Addiction – Tobacco (no ecig) – MPM – Web 3

Addiction – Tobacco (no ecig) – MPM – Web 4

Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is (“tobnoecig_SOURCE_add” – “tobnoecigmpm_SOURCE_add”).

Addiction – Tobacco (no ecig) – PM – AP 1

Addiction – Tobacco (no ecig) – PM – AP 2

Addiction – Tobacco (no ecig) – PM – BTN 1

Addiction – Tobacco (no ecig) – PM – BTN 2

Addiction – Tobacco (no ecig) – PM – Newspaper 1

Addiction – Tobacco (no ecig) – PM – Newspaper 2

Addiction – Tobacco (no ecig) – PM – Web 1

Addiction – Tobacco (no ecig) – PM – Web 2

Addiction – Tobacco (no ecig) – PM – Web 3

Addiction – Tobacco (no ecig) – PM – Web 4

E-cigarette texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “ecigmpm_SOURCE_add”.

Addiction – Ecig – MPM – AP 1

Addiction – Ecig – MPM – AP 2

Addiction – Ecig – MPM – BTN 1

Addiction – Ecig – MPM – BTN 2

Addiction – Ecig – MPM – Newspaper 1

Addiction – Ecig – MPM – Newspaper 2

Addiction – Ecig – MPM – Web 1

Addiction – Ecig – MPM – Web 2

Addiction – Ecig – MPM – Web 3

Addiction – Ecig – MPM – Web 4

Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is (“ecig_SOURCE_add” - “ecigmpm_SOURCE_add”).

Addiction – Ecig – PM – AP 1

Addiction – Ecig – PM – AP 2

Addiction – Ecig – PM – BTN 1

Addiction – Ecig – PM – BTN 2

Addiction – Ecig – PM – Newspaper 1

Addiction – Ecig – PM – Newspaper 2

Addiction – Ecig – PM – Web 1

Addiction – Ecig – PM – Web 2

Addiction – Ecig – PM – Web 3

Addiction – Ecig – PM – Web 4

Youth

Note: For analysts of the dataset, the code for Youth within the database is “cyouth8”.

Tobacco (no ecig) texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “tobnoecigmpm_SOURCE_you”.

Youth – Tobacco (no ecig) – MPM – AP 1

Youth – Tobacco (no ecig) – MPM – AP 2

Youth – Tobacco (no ecig) – MPM – BTN 1

Youth – Tobacco (no ecig) – MPM – BTN 2

Youth – Tobacco (no ecig) – MPM – Newspaper 1

Youth – Tobacco (no ecig) – MPM – Newspaper 2

Youth – Tobacco (no ecig) – MPM – Web 1

Youth – Tobacco (no ecig) – MPM – Web 2

Youth – Tobacco (no ecig) – MPM – Web 3

Youth – Tobacco (no ecig) – MPM – Web 4

Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is (“tobnoecig_SOURCE_you” – “tobnoecigmpm_SOURCE_you”).

Youth – Tobacco (no ecig) – PM – AP 1

Youth – Tobacco (no ecig) – PM – AP 2

Youth – Tobacco (no ecig) – PM – BTN 1

Youth – Tobacco (no ecig) – PM – BTN 2

Youth – Tobacco (no ecig) – PM – Newspaper 1

Youth – Tobacco (no ecig) – PM – Newspaper 2

Youth – Tobacco (no ecig) – PM – Web 1

Youth – Tobacco (no ecig) – PM – Web 2

Youth – Tobacco (no ecig) – PM – Web 3

Youth – Tobacco (no ecig) – PM – Web 4

E-cigarette texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “ecigmpm_SOURCE_you”.

Youth – Ecig – MPM – AP 1

Youth – Ecig – MPM – AP 2

Youth – Ecig – MPM – BTN 1

Youth – Ecig – MPM – BTN 2

Youth – Ecig – MPM – Newspaper 1

Youth – Ecig – MPM – Newspaper 2

Youth – Ecig – MPM – Web 1

Youth – Ecig – MPM – Web 2

Youth – Ecig – MPM – Web 3

Youth – Ecig – MPM – Web 4

Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is (“ecig_SOURCE_you” - “ecigmpm_SOURCE_you”).

Youth – Ecig – PM – AP 1

Youth – Ecig – PM – AP 2

Youth – Ecig – PM – BTN 1

Youth – Ecig – PM – BTN 2

Youth – Ecig – PM – Newspaper 1

Youth – Ecig – PM – Newspaper 2

Youth – Ecig – PM – Web 1

Youth – Ecig – PM – Web 2

Youth – Ecig – PM – Web 3

Youth – Ecig – PM – Web 4

Tobacco (no ecig) Population Norms – Low

Note: For analysts of the dataset, the code for Tobacco (no ecig) Population Norms – Low within the database is “cpopnormstob9”.

Tobacco (no ecig) texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “tobnoecigmpm_SOURCE_tpoplo”.

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – MPM – AP 1

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – MPM – AP 2

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – MPM – BTN 1

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – MPM – BTN 2

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – MPM – Newspaper 1

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – MPM – Newspaper 2

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – MPM – Web 1

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – MPM – Web 2

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – MPM – Web 3

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – MPM – Web 4

Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is (“tobnoecig_SOURCE_tpoplo” - “tobnoecigmpm_SOURCE_tpoplo”).

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – PM – AP 1

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – PM – AP 2

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – PM – BTN 1

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – PM – BTN 2

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – PM – Newspaper 1

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – PM – Newspaper 2

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – PM – Web 1

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – PM – Web 2

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – PM – Web 3

Tobacco (no ecig) Population Norms – Low – Tobacco (no ecig) – PM – Web 4

Tobacco (no ecig) Population Norms – High

Note: For analysts of the dataset, the code for Tobacco (no ecig) Population Norms – High within the database is “cpopnormstob8”.

Tobacco (no ecig) texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “tobnoecigmpm_SOURCE_tpophi”.

Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – MPM – AP 1

Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – MPM – AP 2

Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – MPM – BTN 1

Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – MPM – BTN 2

Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – MPM – Newspaper 1

Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – MPM – Newspaper 2

Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – MPM – Web 1

Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – MPM – Web 2

Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – MPM – Web 3

Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – MPM – Web 4

Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is (“tobnoecig_SOURCE_tpophi” - “tobnoecigmpm_SOURCE_tpophi”).

Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – PM – AP 1

Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – PM – AP 2

Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – PM – BTN 1

Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – PM – BTN 2
Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – PM – Newspaper 1
Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – PM – Newspaper 2
Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – PM – Web 1
Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – PM – Web 2
Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – PM – Web 3
Tobacco (no ecig) Population Norms – High – Tobacco (no ecig) – PM – Web 4

Ecig Population Norms – High

Note: For analysts of the dataset, the code for Ecig Population Norms – High within the database is “cpopnormsecig7”.

E-cigarette texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “ecigmpm_SOURCE_epophi”.

Ecig Population Norms – High – Ecig – MPM – AP 1
Ecig Population Norms – High – Ecig – MPM – AP 2
Ecig Population Norms – High – Ecig – MPM – BTN 1
Ecig Population Norms – High – Ecig – MPM – BTN 2
Ecig Population Norms – High – Ecig – MPM – Newspaper 1
Ecig Population Norms – High – Ecig – MPM – Newspaper 2
Ecig Population Norms – High – Ecig – MPM – Web 1
Ecig Population Norms – High – Ecig – MPM – Web 2
Ecig Population Norms – High – Ecig – MPM – Web 3
Ecig Population Norms – High – Ecig – MPM – Web 4

Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is (“ecig_SOURCE_epophi” - “ecigmpm_SOURCE_epophi”).

Ecig Population Norms – High – Ecig – PM – AP 1
Ecig Population Norms – High – Ecig – PM – AP 2
Ecig Population Norms – High – Ecig – PM – BTN 1
Ecig Population Norms – High – Ecig – PM – BTN 2
Ecig Population Norms – High – Ecig – PM – Newspaper 1
Ecig Population Norms – High – Ecig – PM – Newspaper 2
Ecig Population Norms – High – Ecig – PM – Web 1
Ecig Population Norms – High – Ecig – PM – Web 2
Ecig Population Norms – High – Ecig – PM – Web 3
Ecig Population Norms – High – Ecig – PM – Web 4

Tobacco (no ecig) Individual Norms

Note: For analysts of the dataset, the code for Tobacco (no ecig) Individual Norms within the database is “cindnormstob1”.

Tobacco (no ecig) texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “tobnoecigmpm_SOURCE_tindnorm”.

Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – MPM – AP 1
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – MPM – AP 2
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – MPM – BTN 1
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – MPM – BTN 2
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – MPM – Newspaper 1
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – MPM – Newspaper 2
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – MPM – Web 1
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – MPM – Web 2
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – MPM – Web 3
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – MPM – Web 4

Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “tobnoecig_SOURCE_tindnorm” - “tobnoecigmpm_SOURCE_tindnorm”).

Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – PM – AP 1
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – PM – AP 2
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – PM – BTN 1
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – PM – BTN 2
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – PM – Newspaper 1
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – PM – Newspaper 2
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – PM – Web 1
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – PM – Web 2
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – PM – Web 3
Tobacco (no ecig) Individual Norms – Tobacco (no ecig) – PM – Web 4

Ecig Individual Norms

Note: For analysts of the dataset, the code for Ecig Individual Norms within the database is “cindnormsecig1”.

E-cigarette Texts

More than Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is “ecigmpm_SOURCE_eindnorm”.

Ecig Individual Norms – Ecig – MPM – AP 1
Ecig Individual Norms – Ecig – MPM – AP 2
Ecig Individual Norms – Ecig – MPM – BTN 1
Ecig Individual Norms – Ecig – MPM – BTN 2
Ecig Individual Norms – Ecig – MPM – Newspaper 1
Ecig Individual Norms – Ecig – MPM – Newspaper 2

Ecig Individual Norms – Ecig – MPM – Web 1

Ecig Individual Norms – Ecig – MPM – Web 2

Ecig Individual Norms – Ecig – MPM – Web 3

Ecig Individual Norms – Ecig – MPM – Web 4

Passing Mentions

Note: For analysts of the dataset, the variable name within the dataset is (“ecig_SOURCE_eindnorm” - “ecigmpm_SOURCE_eindnorm”).

Ecig Individual Norms – Ecig – PM – AP 1

Ecig Individual Norms – Ecig – PM – AP 2

Ecig Individual Norms – Ecig – PM – BTN 1

Ecig Individual Norms – Ecig – PM – BTN 2

Ecig Individual Norms – Ecig – PM – Newspaper 1

Ecig Individual Norms – Ecig – PM – Newspaper 2

Ecig Individual Norms – Ecig – PM – Web 1

Ecig Individual Norms – Ecig – PM – Web 2

Ecig Individual Norms – Ecig – PM – Web 3

Ecig Individual Norms – Ecig – PM – Web 4