



HHS Public Access

Author manuscript

Health Commun. Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

Health Commun. 2022 January ; 37(1): 29–38. doi:10.1080/10410236.2020.1816282.

Do Longitudinal Trends in Tobacco 21-Related Media Coverage Correlate with Policy Support? An Exploratory Analysis Using Supervised and Unsupervised Machine Learning Methods

Leeann N. Siegel¹, Allyson Volinsky Levin², Elissa C. Kranzler³, Laura A. Gibson⁴

¹Annenberg School for Communication, University of Pennsylvania

²Department of Communication, Villanova University

³Research and Training Institute, Cancer Support Community

⁴Perelman School of Medicine, University of Pennsylvania

Abstract

Media coverage can impact support for health policies and, ultimately, compliance with those policies. Prior research found consistent, high support for Tobacco 21 policies, which raise the minimum legal age of tobacco purchase to 21, among adults and nonsmoking youth. However, a recent study found support (i.e., agreement with the statement: “The legal age to buy tobacco cigarettes should be increased from 18 to 21”) among 13–20-year-old smokers increased from 2014 until mid-2016 and then declined steadily through mid-2017. To assess whether media coverage could be related to young smokers’ changing support, we conducted an exploratory content analysis to identify texts about Tobacco 21 in a large corpus of tobacco texts (N=135,691) published in four popular media sources from 2014–2017. For this content analysis, we developed a novel methodological approach that combined supervised and unsupervised machine learning methods and could be useful in other areas of communication research. We found that the prevalence of Tobacco 21 media coverage and Tobacco 21 support among young smokers exhibited similar temporal patterns for much of the study period. These findings highlight the need for continued research into the effects of media coverage on Tobacco 21 support among young smokers, a group that must comply with Tobacco 21 policies in order to ensure maximum effectiveness. This research is of particular utility following the 2019 passage of a federal Tobacco 21 regulation, as the public health impact of this regulation could be limited by low public support, and thus low rates of policy compliance.

Despite consistent declines in tobacco cigarette smoking rates among youth and adults over the past several decades (Johnson et al., 2018; Wang et al., 2018), tobacco use remains the leading cause of preventable death in the United States. (U.S. Department of Health and Human Services, 2014). Tobacco use is typically established during adolescence; almost 90% of adult daily smokers initiated before their 19th birthdays (Institute of Medicine, 2015). Thus, developing strategies to prevent youth from starting to smoke is important for reducing tobacco-related morbidity and mortality. One promising strategy is introducing

Tobacco 21 policies that raise the minimum purchasing age of tobacco products to 21 years. Between June 2014 and June 2017, several city- and county-level Tobacco 21 policies were introduced and the first two state-level Tobacco 21 policies were implemented. In December 2019, Tobacco 21 became federal law, making the purchase of any tobacco product by anyone under 21 years illegal in the U.S.

Prior to the 2019 passage of this federal law, a substantial proportion of adults in the U.S. supported Tobacco 21 policies, including both smokers and nonsmokers (King et al., 2015; Morain et al., 2016; Winickoff et al., 2016). However, recent research showed that support was lower among youth, with particularly low support among young smokers (Dai, 2017; Volinsky et al., 2018). This lack of support for Tobacco 21 among young smokers is cause for concern. Young people are more likely to comply with and encourage their peers to comply with anti-smoking policies that they support (Glover-Kudon et al., 2019; Record, 2017; Unger et al., 1999). If young people do not support anti-smoking policies, perhaps because they perceive them to be too repressive, they may be more likely to try to circumvent those policies in order to assert their autonomy (Brehm & Brehm, 2014; Jeffery et al., 1990; Unger et al., 1999). Widespread compliance with the federal Tobacco 21 regulation among young smokers is crucial to maximizing the regulation's impact on smoking rates (Crawford et al., 2002; Levy et al., 2001).

Past research demonstrates that exposure to policy-specific content on traditional and social media platforms can influence public opinion about policies, including tobacco control policies, ultimately affecting rates of compliance as well as policy adoption and maintenance (Asbridge, 2004; Borland, 2006; Burstein, 2003; Niederdeppe et al., 2007). Thus, media coverage of Tobacco 21 policies may change young smokers' policy opinions. One way to assess the relationship between Tobacco 21 news coverage and policy support among young smokers is to first measure the prevalence of Tobacco 21 news coverage over recent years. To date, the only content analysis on this topic examined approximately 100 newspaper articles reporting on Tobacco 21 policies from 2006 to 2016, but only assessed the scientific quality of articles (Huey & Apollonio, 2018). No prior study has measured the prevalence of Tobacco 21 coverage, longitudinal trends in coverage, online source coverage, or coverage after mid-2016.

To address this gap, we used an innovative methodology that combined unsupervised and supervised machine learning with hand-coding. This methodology was particularly useful in overcoming a common problem communication researchers face when using machine learning in content analysis work: identifying rare categories of texts (Weiss, 2004). Using this approach, we identified and coded popular news media texts that discussed Tobacco 21 policies and assessed whether there was a discernible longitudinal pattern in Tobacco 21 media coverage from 2014–2017. We then tested whether longitudinal trends in media coverage were associated with corresponding trends in policy support among young smokers to better understand the interplay between Tobacco 21 media coverage and patterns of Tobacco 21 support.

Tobacco 21 Policy

In this study, we focus on media coverage and policy support from mid-2014 to mid-2017 to correspond with the available media content and survey data described below. During this time period, the first two state-level Tobacco 21 policies were introduced. While Needham, Massachusetts, was the first town to raise the minimum age of legal access from 18 to 21 years in April 2005 (Kessel Schneider et al., 2016), Hawaii became the first state to enact a Tobacco 21 policy in June 2015. The most populous state in the nation, California, followed suit by passing its own Tobacco 21 policy in May 2016. Since then, many other Tobacco 21 laws have been introduced across the country; by July 2019, over half of the U.S. population lived in a state or locality covered by Tobacco 21 laws (American Lung Association, 2019). Then, in December 2019, a federal Tobacco 21 policy was signed into law and immediately took effect.

Tobacco 21 Policy Support Among Youths

Despite broad support for Tobacco 21 policies among adults, youth demonstrate considerably lower rates of support for such policies. A recent study based on the 2015 National Youth Tobacco Survey demonstrated that 67.3% of nonsmoking youth (aged 11–18) expressed support for Tobacco 21 policies, while only 17.1% of current smoking youth indicated policy support (Dai, 2017). Similarly, a study conducted with a nationally representative sample of youth and young adults, surveyed from 2014–2017, demonstrated evidence of differences in Tobacco 21 policy support (defined as agreement with the statement: “The legal age to buy tobacco cigarettes should be increased from 18 to 21”) between smokers and nonsmokers (Volinsky et al., 2018). While 75.8% of nonsmokers between the ages of 13 and 20 were in favor of raising the legal age to buy tobacco products to 21, the level of support was only 40.1% among smokers in this age range, the group most directly impacted by a nationwide Tobacco 21 policy. Furthermore, longitudinal trends in Tobacco 21 support differed between these two groups; despite relatively constant support among nonsmokers throughout the study period, support among smokers increased over the initial two-year period from 2014 until mid-2016, then peaked at approximately 50% support before steadily decreasing thereafter (Volinsky et al., 2018).

The Influence of Media Coverage on Tobacco 21 Policy Support

While a substantial body of research demonstrates that media can impact young people’s smoking-related attitudes and behavior (e.g. Paek & Gunther, 2007; Sargent et al., 2009; K. C. Smith et al., 2008), little research has explored the impact of media on young people’s tobacco policy opinions. Among the few studies that have been done, a recent observational study found that young people who had viewed or engaged with tobacco-related content on social media reported lower support for e-cigarette regulatory policies, even when controlling for e-cigarette use (Majmundar et al., 2019). Hersey et al. (2003) found that exposure to media campaigns with messaging tactics that disparaged tobacco industry practices led to higher endorsement of anti-industry beliefs, including endorsement of increased government oversight of the tobacco industry, among youths. Studies showing that media coverage can affect tobacco-relevant policy opinions among adults (e.g. Niederdeppe

et al., 2017; Rennan et al., 2014; Tan et al., 2015) also suggest that Tobacco 21 media coverage might impact young smokers' policy support.

One mechanism that may explain the relationship between volume of news coverage and tobacco control policy support is the valence of such coverage. Valence reflects the opinion slant or overall attitude of an article towards a topic in question.¹ When an article has a predominant valence, this valence can shape how news coverage of a policy influences policy support (Blake et al., 2010a, 2010b). For example, articles that are mostly in favor of a particular tobacco control policy often differ from articles that are mostly opposed to that policy in terms of the facts and arguments presented and the frames used (Eckler et al., 2016; Huey & Apollonio, 2018). Because of these differences, these two groups of articles might be expected to have opposite effects on public support for the policy in question (Myers et al., 2017). Indeed, past research has demonstrated this pattern of effects; for example, Nagelhout et al. (2012) found that newspaper coverage that was negative towards smoking bans led to a negative effect on support for smoke-free bars and restaurants, while newspaper coverage that was positive towards such bans had a positive effect on support among some individuals.

Media can impact support for specific tobacco control policies through multiple pathways, including raising awareness of or increasing knowledge about these policies, and changing the salience and perceived importance of arguments for and against them (Harris et al., 2010; Menashe, 1998; Thrasher et al., 2014). Regardless of an author's stance on a given policy, increases in the volume of policy-relevant coverage may influence audience perceptions of the policy's importance, thereby increasing policy-specific support (Long, Slater, & Lysengen, 2006). According to the Influence of Presumed Influence hypothesis (Gunther et al., 2006), which argues that individuals are impacted by their presumptions about how media have influenced others, young people exposed to Tobacco 21 media coverage may assume their peers have seen and affected by the same coverage, and change their opinions for better alignment with expected opinions of their peers. If any of these mechanisms were at play, we would expect a relationship between Tobacco 21-related media coverage and support among young smokers.

Past research on agenda-setting has identified perceived relevance of a topic as a predictor of how likely people are to seek out, be exposed to and be affected by media coverage of that topic (McCombs and Stroud, 2014; Kim, 2009). Individuals' perceptions of the relevance of a topic are in turn shaped by their self-interests (McCombs, 1999, 2014; Evatt and Ghanem, 2001; McCombs and Stroud, 2014). If young smokers perceive Tobacco 21-related media coverage to be more personally relevant to their interests than do young non-smokers, they may be more affected by such coverage simply because they are more likely to attend to it. Indeed, survey data from a nationally representative sample of young people support this argument; among those younger than 21, 14.5% of smokers reported actively looking for

¹It is worth noting that an article's valence towards tobacco use in general may not reflect its valence towards a particular tobacco control policy. For example, while prior studies investigating the valence of tobacco-related news media content have found that such coverage is primarily against the use of tobacco, media coverage of a specific tobacco control policy often presents mixed valence towards that particular policy, including arguments both for and against the policy in question (Long, 2006; Myers et al., 2017; Myers et al., 2019).

information about cigarettes or other tobacco products within the past month, while only 5% of non-smokers reported doing so (Authors, unpublished data). Among smokers, those younger than 21 may be even more apt to attend to, and thus be affected by, Tobacco 21 coverage relative to their older counterparts because Tobacco 21 policies would directly impact their ability to purchase tobacco products and thus might be seen as more related to their personal interests and more relevant..

Of note, media coverage of Tobacco 21 could also impact young smokers' policy support even if they were not directly exposed to it because individual exposure represents only one pathway through which media effects occur (Hornik, 2002). Alternatively, it is plausible that Tobacco 21 media coverage affected young smokers by first impacting the opinions held by others in their social networks (Hornik, 2002). Past research has shown that the tobacco-related opinions and behaviors of young people's parents and friends impact the manner in which they interpret smoking-related media content (McCool et al., 2011; Setodji et al., 2013) and their support for tobacco control policies (Bernat et al., 2009; Glover-Kudon et al., 2019).

Methods

Tobacco 21-related media texts were identified through a multi-stage process that utilized both supervised and unsupervised machine learning methods. Supervised machine learning (SML) methods require a sample of texts that have been separated into a predetermined set of categories through human coding. This hand-coded sample is used to train an algorithm to identify crucial features that can differentiate between texts in these different categories, then use these features to categorize future documents (Grimmer & Stewart, 2013). Conversely, unsupervised machine learning (UML) methods do not require a sample of texts to be hand-coded; rather, they use modeling assumptions and textual features to estimate different clusters of features, or "topics," into which texts are grouped (Grimmer & Stewart, 2013).

In this study, we developed an SML classifier to identify tobacco-related media texts that contained mentions of a particular theme we refer to as *policy*. This classifier was applied to a corpus of over 130,000 tobacco-related texts from popular media sources. We then used UML to classify texts as containing Tobacco 21 language. We iteratively fit a series of Latent Dirichlet Allocation (LDA) models to the subset of texts that were classified as containing mentions of the policy theme. LDA models are generative, probabilistic models that use a corpus of texts (or other classes of observations) to identify an underlying set of topics and model each text as a finite mixture of each topic (Blei et al., 2003). Each text is assigned a percentage contribution from each topic. A final LDA model was selected, and this model was used to identify a distinct cluster of "Tobacco 21" texts. Finally, trends in the prevalence of these Tobacco 21 texts over time were examined and compared to those observed in Tobacco 21 support among young smokers and nonsmokers. Each step of this process is explicated further below.

Tobacco 21 Policy Support

Support for Tobacco 21 policies among young smokers and young non-smokers was measured using a single survey item. This item was included in a series of items about support for different tobacco control policies (the order of these items was randomized for each survey participant) and used the following wording: “I’m going to read a few more statements. Please tell me whether you strongly disagree, disagree, agree, or strongly agree with them ... The legal age to buy tobacco cigarettes should be increased from 18 to 21.” Participants who responded “agree” or “strongly agree” to this item were counted as expressing support for Tobacco 21. Thus, our Tobacco 21 support variable was defined as support for changing the legal age to purchase cigarettes in general rather than support of a particular Tobacco 21 regulation at the local, state or federal level. The wording used in this item was very similar to that used in other studies that have measured Tobacco 21 support (e.g. Winickoff et al., 2016; Dai, 2017).

Corpus of Media Texts

The texts used in this study were collected as part of a larger research project examining how media coverage of tobacco products, including e-cigarettes, impacts youth and young adults’ cognitions, behavioral intentions, and behaviors. Texts that were published between May 2014 and June 2017 were drawn from four major media sources: the AP Newswire (“AP”), the 50 U.S. English-language newspapers with the highest circulation (“News”), broadcast TV and radio news transcripts from eight sources (“BTN”), and the websites most popular among members of the population of interest in this study (youth and young adults) according to Nielsen ratings (“Web”). Texts were classified as relevant to tobacco and e-cigarettes through a multi-step coding process (Gibson et al., 2019). The final sample of texts that was used in this study consisted of 135,691 texts, of which 10,598 (8%) mentioned e-cigarettes. The majority of these texts came from News ($n = 52,561$) and Web ($n = 70,406$), with fewer from AP ($n = 8,522$) and BTN ($n = 4,202$). Gibson et al. (2019) describe this corpus in greater detail.

Development of Supervised Machine Learning Classifier

The process of developing the SML classifier used to identify texts that mention policy, which was defined as mandatory tobacco-related policy, law, or regulation by a government, company, or institution, is described in greater detail in Gibson et al., 2019. To develop this classifier, a random sample of 2,400 texts, stratified by source, was pulled from our media corpus and hand-coded by crowdworkers through the Amazon Mechanical Turk (MTurk) platform. The use of crowdworkers in text labeling has become increasingly popular in social science research because it enables large corpora of texts to be labeled more quickly, at a lower cost, and by a more diverse group of coders than would be possible if, for example, undergraduate research assistants were used as coders (Budak et al., 2016; Mason & Suri, 2012). We took additional steps to enhance the accuracy of crowdworkers’ coding including requiring MTurkers to pass a theme-specific qualification exercise to be eligible to code texts for policy relevance and providing them with a codebook for the policy theme that contained definitions and examples of policy-relevant texts. Additionally, individual

MTurkers who participated in policy coding but had low individual reliability with the rest of the coders were dropped.

Each text in the 2,400-text sample was coded by a minimum of 7 coders who determined the presence or absence of the policy theme. Only texts for which at least 75% of MTurkers agreed on policy classification were used to train the SML classifier. The hand-coded sample was then divided into a training sample (80%) and a held-aside test sample (20%), which was not used in the development of the classifier. SML using logistic regression classifiers was conducted through several modules of the scikit learn package for Python version 3.6 (Pedregosa et al., 2011). Texts were pre-processed through: vectorizing; grouping features into unigrams, bigrams and trigrams; and relative pruning, in which words that appeared in fewer than 1% or more than 99% of texts were removed (Maier et al., 2018). We used the chi-squared distribution of each feature with the outcome (i.e. being a “policy” text) to select the 5,000 most predictive features. Next, the optimal number of features for the classifier was determined using recursive feature elimination, a process through which the least predictive features are repeatedly pruned from sets of features that grow smaller as the process continues, and 5-fold cross-validation. The classifier was fit with the optimal number of features from the training samples and summary metrics were computed. The final classifier developed to identify policy texts used 652 features, of which the 5 strongest predictors were *ban*, *tax*, *taxes*, *banned*, and *tobacco products*.

Two main criteria were used to evaluate the classifier’s performance, both of which were assessed in the test sample: the classifier’s F1 score, and the correlation between the classifier predicted probabilities for texts and the MTurk workers’ coding. F1 scores represent a harmonic mean between precision (the conditional probability that a text is relevant, given that it is retrieved) and recall (the conditional probability that a text will be retrieved, given that it is relevant) (Stryker et al., 2006). Within its test sample ($n = 356$ texts), the policy classifier had an F1 score of .93 and a correlation with the MTurk workers’ coding of .90, indicating it was sufficiently accurate. The policy classifier was then used to code our entire corpus.

Latent Dirichlet Allocation (LDA)

After applying the policy classifier to code our corpus of texts, we compiled a new sample of texts consisting of all the texts that had been assigned a greater than .50 probability of containing the policy theme by the classifier. This new sample (total $n = 17,477$) contained 3,138 AP texts, 256 BTN texts, 7,609 News texts, and 6,475 Web texts. Then, LDA analysis was performed using the Gensim (Rehder & Sojka, 2010) and NLTK (Wagner, 2010) libraries in Python version 3.6. Texts were pre-processed by: (1) tokenizing, (2) removing string punctuation and special characters, (3) removing stop words, and (4) stemming words to remove plural or verb conjugation endings (Maier et al., 2018). In the third step, we applied a custom list of stop words comprised of all the words in NLTK’s default set of English stop words, along with a set of tobacco-related words. We filtered out these tobacco-related words because they were used to categorize texts as tobacco-relevant in the initial coding stage, and therefore appeared with disproportionate frequency in our corpus.

We then applied relative pruning, removing words that appeared in fewer than 25 texts or more than 99% of texts (Maier et al., 2018).

After pre-processing, we produced a dictionary of all words appearing in the sample of texts. Each text was turned into a bag-of-words array, in which it was represented only by the number of times it used each word in the dictionary (D. A. Smith & McManis, 2015). We then fit several online LDA models to the resulting matrix (Hoffman et al., 2010). While fitting these models, the number of topics was varied from 3 to 10, the chunk size was set to 256, and 1,000 total passes were made through the entire corpus (Hoffman et al., 2010).

The final LDA model was selected on the basis of its semantic validity, or the extent to which the topics identified contained coherent groups of texts that were similar to one another and different from texts in other clusters, as well as the utility of the topics identified (Grimmer & Stewart, 2013; Quinn et al., 2010). To assess models' semantic validity, we examined the most discriminating words (those that best distinguish texts in one topic cluster from texts in others) for each topic in the model, as well as the texts assigned the highest percentage contribution from that topic. We were mainly interested in models that contained a single topic related to Tobacco 21 policies, and thus focused on such models. After selecting a final model that contained a coherent Tobacco 21 topic, we assessed the topic distributions for each text and grouped texts according to the topics for which they had the highest percentage contribution. We subsequently focused on texts assigned to the Tobacco 21 topic cluster, which will be referred to as Tobacco 21 texts for the remainder of this paper.

Prior to conducting analyses, we first created a dataset of quarterly measures of Tobacco 21 media coverage and Tobacco 21 policy support among smokers and nonsmokers between the ages of 13 and 20.² To produce the media coverage variable, we summed the number of Tobacco 21 texts published in each quarter of the study period, with each text weighted by its percentage contribution from the Tobacco 21 topic.³ The Tobacco 21 policy support measure came from a rolling cross-sectional phone (cell and landline) survey of individuals between the ages of 13 and 25 (total $n = 11,847$, including 8,361 respondents aged 13–20 years), conducted over the same time period during which our media measures were collected (mid-2014 to mid-2017) (Volinsky et al., 2018). The sample was weighted to be representative of the U.S. population of individuals aged 13–25 years old. For analysis purposes, we focused on participants between the ages of 13 and 20 (i.e., younger than 21). Tobacco 21 policy support was the quarterly weighted average level of policy support expressed by smokers and nonsmokers.

While prior studies have found that the valence of tobacco-related news media content is primarily against the use of tobacco, media coverage of specific tobacco control policies

²We chose to conduct analyses at the quarterly level rather than at smaller time intervals (e.g. monthly), as there were very few Tobacco 21 articles published in each month of the study period, and very few smokers aged 13–20 who were surveyed during each month. Because of this, our monthly measures of both Tobacco 21 media coverage and Tobacco 21 policy support among young smokers were unstable; our quarterly measures of these variables were more stable.

³When collapsing to the quarterly level, data collected before July 1, 2014 were omitted because we only began collecting media data in mid-May 2014 and survey data in mid-June 2014. Because we did not have complete data for the second quarter of 2014, we did not include it in our analyses.

may present mixed valence content (both pro- and anti-policy support) (Myers et al., 2017; Myers et al., 2019), suggesting the importance of examining the valence of media content. To complement our main analyses, we selected a stratified random subsample of 250 Tobacco 21 texts and hand-coded these texts to estimate the valence of coverage for the population of texts. Half of these texts were drawn randomly from all texts that had been assigned the highest percentage contribution from the Tobacco 21 topic, while the other half were drawn from the subset of such texts that had a percentage contribution from the Tobacco 21 topic greater than or equal to .50. Texts were coded by two expert coders as being *pro* (mostly supportive of the Tobacco 21 policy under discussion or Tobacco 21 policies in general), *anti* (mostly against the Tobacco 21 policy under discussion or Tobacco 21 policies in general), *mixed* (mentioning arguments for and against the Tobacco 21 policy under discussion or Tobacco 21 policies in general) or *neutral* (not having a discernible opinion slant towards Tobacco 21 policies). Inter-coder reliability was found to be acceptable (Krippendorff's Alpha = .80), and disagreements between coders were resolved through discussion in order to determine which valence code would ultimately be assigned to each text. Results from this analysis were used to inform our discussion of the most plausible mechanisms through which Tobacco 21 media coverage impacted policy support.

To assess the extent to which Tobacco 21 media coverage correlated with the average level of policy support among 13–20-year-old smokers and nonsmokers, we calculated the Pearson's product-moment correlations between these variables. Using the ggplot2 package in R (Wickham, 2009) we plotted quarterly measures of Tobacco 21 coverage and levels of support for Tobacco 21 policies among smokers and nonsmokers between the ages of 13 and 20, to visually compare these longitudinal patterns.

Results

The final LDA model we selected included 9 topics. For each topic, the 10 most discriminating words that distinguish that topic from other topics are included in Table 1. By examining the words most predictive of a text belonging to each topic, we labeled Topic 5 as a cluster of texts about Tobacco 21 policies. To support this conclusion, we read a random stratified sample of 50 texts whose dominant percentage contributions were from Topic 5, the Tobacco 21 topic (i.e., their Tobacco 21 topic contributions were larger than contributions from any other topic). One stratum was a random sample of 30 texts from all texts assigned to this topic, and the other was a random sample of 20 texts with Tobacco 21 topic contribution percentages greater than .50. These texts primarily discussed the proposal, passage or implementation of specific Tobacco 21 laws. For example, the text with the highest percentage contribution from the Tobacco 21 topic was an AP article describing the progression of a New Jersey law to raise the minimum legal age of tobacco purchase from 19 to 21 through the state legislature. Other texts with high contributions from the Tobacco 21 topic discussed California's Tobacco 21 law and Tobacco 21 policies being considered in Massachusetts and Minnesota.

We followed the same process to understand the content of texts assigned to each other LDA topic. While detailing the types of texts assigned to each cluster is beyond the scope of this

paper, this analysis revealed, for example, that texts assigned to Topic 1 tended to discuss bans on smoking in public places, and texts assigned to Topic 6 tended to discuss e-cigarette regulations. After assigning texts to their dominant topic, we found that the clusters of texts for each topic differed notably in their size. Topic clusters ranged in size from 658 texts (Topic 2) – 3,983 texts (Topic 8). A total of 1,453 texts were assigned to Topic 5, the Tobacco 21 topic.

Figure 1 shows the quarterly prevalence of Tobacco 21 media coverage, measured as the sum of Tobacco 21 texts published during that quarter, with each text weighted by its percentage contribution from the Tobacco 21 topic. The quarterly prevalence of Tobacco 21 coverage varied substantially from roughly 17 in the fourth quarter of 2016 to approximately 108 in the first quarter of 2016, during which California’s Tobacco 21 law was passed by the state legislature. As can be seen in Figure 1, Tobacco 21 media coverage and the Tobacco 21 support among young smokers showed roughly similar longitudinal patterns from the beginning of 2015 until the end of 2016, although the two trends did not appear to align outside of this time period. Both trends tended to increase until mid-2016, and then decline through the end of 2016. In contrast, the Tobacco 21 support among young nonsmokers remained flat over the entire study period.

Our valence analysis revealed that Tobacco 21 coverage was primarily *neutral*, with 47% of articles about Tobacco 21 merely describing Tobacco 21 policies or providing updates about the progress of such policies through the legislative process. Smaller numbers of articles were *pro* (28%) or *mixed* (21%), with very few articles (4%) coded as *anti*-Tobacco 21.

The Pearson’s correlation between Tobacco 21 media coverage and support for Tobacco 21 policies among young smokers was .44. This correlation was not statistically significant ($p = .16$), potentially due to the very small sample size ($n = 12$ quarters) and the divergence observed between these two trends at the beginning and end of the study period. However, the Pearson’s correlation between Tobacco 21 media coverage and Tobacco 21 support among young nonsmokers was half that at .20 and was also not statistically significant ($p = .53$).

Discussion

Available research suggests that the recently enacted national Tobacco 21 policy could substantially contribute to the prevention or delay of youth tobacco initiation and prevent hundreds of thousands of premature deaths caused by tobacco (Winickoff, 2018; Dai, 2017). However, low support for this policy among young smokers may inhibit the effectiveness of this regulation by reducing policy compliance (Crawford et al., 2002; Glover-Kudon et al., 2019; Unger et al., 1999). In this study, we examined whether media coverage of Tobacco 21 policies between 2014–2017 might have contributed to the concerning trend in Tobacco 21 support observed among young smokers during this time period.

In service of this goal, we used an innovative, multi-step approach to automated content analysis, combining both supervised and unsupervised machine learning methods, to identify texts from four popular media sources that discussed Tobacco 21 policies and

examine the prevalence of Tobacco 21 coverage over time. We also assessed associations between measures of Tobacco 21-related media coverage and support for Tobacco 21 policies among young smokers and nonsmokers over a three-year period. While visual examination revealed similarities between the trends observed in Tobacco 21 media coverage and Tobacco 21 support among young smokers for much of the study period, the Pearson's correlation between these two trends was not statistically significant. However, this lack of statistical significance may be attributed to the small sample size included in these analyses, as both variables were measured at the quarterly level and there were only 12 quarters in the study period. Despite this issue, we still opted to conduct analyses at the quarterly level, rather than at the monthly level, because of the instability of our measures of Tobacco 21 media coverage and Tobacco 21 policy support among young smokers at the monthly level (see Footnote 2).

Our analysis of articles' valence toward Tobacco 21 policies highlighted some mechanisms through which Tobacco 21 media coverage could have impacted young smokers' opinions about Tobacco 21 policies. *Neutral* articles, the most commonly occurring category of Tobacco 21 coverage, may have increased awareness of and knowledge about Tobacco 21 policies, as well as shifted perceptions about the importance given to Tobacco 21 policies on the public agenda and the likelihood of such policies being passed. As *neutral* coverage increased or decreased, respectively, policy support might have increased or decreased concomitantly through any or all of these pathways. Additionally, the substantial number of *pro*-Tobacco 21 articles might have increased the salience of arguments for Tobacco 21 policies in individuals' minds; thus, those interviewed in times of higher *pro*-Tobacco 21 coverage might have been more likely to recall such arguments and express support for Tobacco 21 policies. The current analyses do not allow us to definitively identify which mechanisms were at work. Future research examining precisely how Tobacco 21 coverage impacts policy support should seek to test these hypotheses.

Of note, we did not observe a discernible relationship between Tobacco 21 media coverage and Tobacco 21 support among young non-smokers. As mentioned, one possible explanation for this finding is that young non-smokers do not attend to tobacco-related news coverage to the same extent as young smokers, potentially due to a lack of interest or perceived relevance. Our finding (described earlier) that young smokers are much more likely than young non-smokers to seek out tobacco-related information is consistent with this explanation. If young non-smokers were not paying attention to or did not come across Tobacco 21 media coverage, it would have been unlikely to affect their policy support through individual exposure. However, as we were not able to directly test whether low exposure or low attention to tobacco-related news coverage could explain the lack of a relationship between Tobacco 21 media coverage and policy support among non-smokers, future research on this topic is warranted.

As previously mentioned, our methodological approach was particularly useful in overcoming a common problem in machine learning: identifying a category of texts that is both rare in comparison to other categories of texts (relatively rarity) and rare in terms of the absolute number of texts that compose it (absolute rarity) (Weiss, 2004). We ultimately identified fewer than 1,500 Tobacco 21 texts from a corpus of 135,691 tobacco-related

texts; thus, such texts were rare in comparison to texts about other topics. Our approach helped mitigate this issue by first identifying a more prevalent category of texts (policy texts) through the use of SML classification, and then limiting UML analyses to these policy texts, in which Tobacco 21 texts were expected to be less rare relative to other texts. Applying UML in the second step of our approach also helped us deal with the absolute rarity of Tobacco 21 texts, because a category of texts that is small in number may still emerge as a topic cluster in UML analyses if those texts are sufficiently distinct from others in the corpus (Weiss, 2004). Further, using UML rather than SML in this step prevented us from needing to collect and hand-code a text sample with adequate numbers of Tobacco 21 texts to be able to effectively train and assess an SML classifier. Doing so would have been difficult, given the absolute rarity of Tobacco 21 texts in our corpus, and would have involved additional costs for MTurkers to produce the hand-coded sample.

Another automated coding technique, the dictionary or keyword classifier approach, can be applied in certain research contexts to identify rare categories of texts. However, that approach would have been less useful in this study because of the diverse language used by texts in our corpus to discuss Tobacco 21 policies. For example, a simple keyword search for the phrase “Tobacco 21” turned up only 74 texts including this phrase; thus, using this search would have caused us to miss 95% of the 1,453 Tobacco 21 texts we ultimately identified. To use more complex dictionary methods in automated content analysis, researchers either need to identify a pre-existing relevant dictionary (which did not exist in this case) or develop their own. Given the diversity of language used in the Tobacco 21 texts in our corpus, developing a dictionary would have been time-consuming and arduous (Barberá et al. 2016). Our methodological innovation required fewer resources to successfully identify a comprehensive set of Tobacco 21 texts.

There were additional benefits to our use of both UML and SML. UML methods enabled us to discover new sub-categories among the texts in our corpus that had been coded for presence of the policy theme (Grimmer & Stewart, 2013). While we began our LDA analysis expecting, and indeed hoping, to identify a cluster of Tobacco 21 texts, we did not have any *a priori* assumptions about the other clusters that would emerge. Thus, our methodological approach provided a richer understanding of the range of subjects discussed in tobacco policy texts.

It warrants mention that, although we found similarities between trends in popular media coverage about Tobacco 21 policies and support for such policies among smokers between the ages of 13 and 20, this, by itself, does not constitute evidence of a causal relationship. Media exposure has the capacity, and has been demonstrated, to impact public support for certain policies (Foster et al., 2012; Kim, 2015; Yanovitzky, 2002). However, it is also possible that the prevalence of Tobacco 21-related media coverage and support for Tobacco 21 policies among young smokers appeared to move in tandem for part of the study period because they were both being driven by external factors, such as the proposal or passage of new Tobacco 21 policies.

Limitations

There are a number of limitations to this study. First, we only considered trends in the absolute number of texts published each quarter, rather than trends in other characteristics of Tobacco 21 texts, such as their valence. Although we hand-coded the valence of a sample of texts, coding all texts for valence is beyond the scope of this study. Further, given the relatively small number of Tobacco 21 texts in our dataset, it would be difficult to generate stable estimates of the prevalence of Tobacco 21 texts in each valence category, even at the quarterly level.

Additionally, because of the data we had available, we were only able to include a single-item measure of expressed support for Tobacco 21. While other Tobacco 21-related studies have also used single-item measures of policy support that included similar wording (e.g. Winickoff et al., 2016; Dai, 2017; Glover-Kudon et al., 2019), it is possible that using a single-item measure may have negatively impacted measurement reliability and that our results would have differed had we used a different measure of policy support. Further, we were not able to examine the relationship between Tobacco 21 media coverage and any other Tobacco 21-related attitudinal or knowledge variables. Future research should assess whether such variables play an explanatory role in the relationship between Tobacco 21 media coverage and policy support.

Our corpus was limited to texts published in four media sources – the AP Newswire, broadcast TV and radio news shows, major newspapers and popular websites during a three-year period (2014 – 2017). The websites included in this study were specifically chosen to reflect the most popular sites among individuals between the ages of 13 and 20, the population of interest in this study, but the other media sources may not represent the sources used most frequently by this age group. These sources may differ in important ways from social media sources, for example. Additionally, this three-year time period does not include more recent media coverage or youth/young adult policy endorsement following Tobacco 21 enactment first in additional states and municipalities and then at the national level. To address these limitations, future work should incorporate additional media sources, including social media sources like Twitter, and consider effects over time given temporal trends in Tobacco 21 policy implementation.

Conclusions

Achieving high levels of support for the nationwide Tobacco 21 policy introduced in 2019 among young smokers is critical to ensuring compliance with the policy so that it will reduce smoking rates (Glover-Kudon et al., 2019; Record, 2017; Unger et al., 1999). This paper provides some evidence of a link between Tobacco 21 media coverage and Tobacco 21 support among young smokers. Although this evidence is not causal, these findings are still valuable given limited scholarship investigating effects of media coverage on tobacco policy support among young people and highlight the need for additional research in this area. Future research on Tobacco 21 that harnesses our innovative content coding methods, enlists a larger sample of young smokers and includes additional knowledge-based or attitudinal measures related to policy support would enhance our understanding of the effects of Tobacco 21 media coverage on policy support and compliance among young smokers. More

broadly, future health communication work will benefit from the unique methodological paradigm we developed that combines SML and UML methods to accurately identify rare categories of texts within large corpora. In light of the growing accessibility of corpora of previously unimaginable sizes (Shah et al., 2015), this new methodological approach provides an avenue for communication researchers studying rare topics to engage in work that might otherwise not be possible.

References

- American Lung Association. (2019, July 18). Tobacco 21 laws: Tracking progress toward raising the minimum sales age for all tobacco products to 21. Retrieved from <https://www.lung.org/our-initiatives/tobacco/cessation-and-prevention/tobacco-21-laws.html>
- Apollonio DE, & Glantz SA (2016). Minimum ages of legal access for tobacco in the United States From 1863 to 2015. *American Journal of Public Health*, 106(7), 1200–1207. 10.2105/AJPH.2016.303172 [PubMed: 27196658]
- Asbridge M (2004). Public place restrictions on smoking in Canada: Assessing the role of the state, media, science and public health advocacy. *Social Science & Medicine*, 58(1), 13–24. 10.1016/S0277-9536(03)00154-0 [PubMed: 14572918]
- Barberá P, Boydston A, Linn S, McMahon R, & Nagler J (2016, 9). Methodological challenges in estimating tone: Application to news coverage of the U.S. economy. Annual meeting of the American Political Science Association, Philadelphia, PA.
- Bernat DH, Klein EG, Fabian LEA, & Forster JL (2009). Young adult support for clean indoor air laws in restaurants and bars. *Journal of Adolescent Health*, 45(1), 102–104. 10.1016/j.jadohealth.2008.12.023
- Blake KD, Viswanath K, Blendon RJ, & Vallone D (2010a). The role of tobacco-specific media exposure, knowledge, and smoking status on selected attitudes toward tobacco control. *Nicotine & Tobacco Research*, 12(2), 117–126. 10.1093/ntr/ntp184 [PubMed: 20018944]
- Blake KD, Viswanath K, Blendon RJ, & Vallone D (2010b). The role of reported tobacco-specific media exposure on adult attitudes towards proposed policies to limit the portrayal of smoking in movies. *Tobacco Control*, 19(3), 191–196. 10.1136/tc.2009.031260 [PubMed: 20008152]
- Blei DM, Ng AY, & Jordan MI (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Borland R (2006). Support for and reported compliance with smoke-free restaurants and bars by smokers in four countries: Findings from the International Tobacco Control (ITC) Four Country Survey. *Tobacco Control*, 15(suppl_3), iii34–iii41. 10.1136/tc.2004.008748 [PubMed: 16754945]
- Brehm SS, & Brehm JW (2014). Psychological reactance: A theory of freedom and control. Elsevier Science. <http://qut.eblib.com.au/patron/FullRecord.aspx?p=1839510>
- Budak C, Goel S, & Rao JM (2016). Fair and balanced? Quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80(S1), 250–271. 10.1093/poq/nfw007
- Burstein P (2003). The impact of public opinion on public policy: A review and an agenda. *Political Research Quarterly*, 56(1), 29–40. 10.1177/106591290305600103
- Crawford M, Balch G, Mermelstein R, & Tobacco Control Network Writing Group. (2002). Responses to tobacco control policies among youth. *Tobacco Control*, 11(1), 14–19. 10.1136/tc.11.1.14 [PubMed: 11891362]
- Dai H (2017). Attitudes toward Tobacco 21 among US youth. *Pediatrics*, 140(1), e20170570. 10.1542/peds.2017-0570 [PubMed: 28630116]
- Eckler P, Rodgers S, & Everett K (2016). Characteristics of community newspaper coverage of tobacco control and its relationship to the passage of tobacco ordinances. *Journal of Community Health*, 41(5), 953–961. 10.1007/s10900-016-0176-8 [PubMed: 27000041]
- Evatt D, & Ghanem SI (2001, 9). Building a scale to measure salience. Paper presented at the World Association of Public Opinion Research annual conference, Rome, Italy.

- Foster C, Thrasher J, Kim S-H, Rose I, & Besley J (2012). Agenda-building influences on the news media's coverage of the U.S. Food and Drug Administration's push to regulate tobacco, 1993–2009. *Journal of Health and Human Services Administration*, 35(3), 303–330. [PubMed: 23293806]
- Gibson LA, Siegel L, Kranzler E, Volinsky A, O'Donnell MB, Williams S, Yang Q, Kim Y, Binns S, Tran H, Maidel Epstein V, Leffel T, Jeong M, Liu J, Lee S, Emery S, & Hornik RC (2019). Combining crowd-sourcing and automated content methods to improve estimates of overall media coverage: Theme mentions in e-cigarette and other tobacco coverage. *Journal of Health Communication*, 24(12), 889–899. 10.1080/10810730.2019.1682724 [PubMed: 31718524]
- Glover-Kudon R, Plunkett E, Lavinghouze R, Trivers KF, Wang X, Hu S, & Homa DM (2019). Association of peer influence and access to tobacco products with U.S. youths' support of Tobacco 21 laws, 2015. *Journal of Adolescent Health*, 65(2), 202–209. 10.1016/j.jadohealth.2018.11.020
- Grimmer J, & Stewart B (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 1–31.
- Gunther AC, Bolt D, Borzekowski DLG, Liebhart JL, & Dillard JP (2006). Presumed influence on peer norms: How mass media indirectly affect adolescent smoking. *Journal of Communication*, 56(1), 52–68. 10.1111/j.1460-2466.2006.00002.x
- Harris JK, Shelton SC, Moreland-Russell S, & Luke DA (2010). Tobacco coverage in print media: The use of timing and themes by tobacco control supporters and opposition before a failed tobacco tax initiative. *Tobacco Control*, 19(1), 37–43. 10.1136/tc.2009.032516 [PubMed: 19965799]
- Hersey JC, Niederdeppe J, Evans WD, Nonnemaker J, Blahut S, Farrelly MC, Holden D, Messeri P, & Haviland ML (2003). The effects of state counterindustry media campaigns on beliefs, attitudes, and smoking status among teens and young adults. *Preventive Medicine*, 37(6), 544–552. 10.1016/j.ypmed.2003.07.002 [PubMed: 14636787]
- Hoffman MD, Blei DM, & Bach F (2010). Online learning for Latent Dirichlet Allocation. *Advances in Neural Information Processing Systems*, 24.
- Hornik RC (Ed.). (2002). *Public health communication: Evidence for behavior change*. L. Erlbaum Associates.
- Huey J, & Apollonio DE (2018). A content analysis of popular media reporting regarding increases in minimum ages of legal access for tobacco. *BMC Public Health*, 18(1), 1129–1135. 10.1186/s12889-018-6020-6 [PubMed: 30223799]
- Institute of Medicine. (2015). *Public health implications of raising the minimum age of legal access to tobacco products* (Bonnie RJ, Stratton K, & Kwan LY, Eds.). National Academies Press. 10.17226/18997
- Jeffery RW, Forster JL, Schmid TL, McBride CM, Rooney BL, & Pirie PL (1990). Community attitudes toward public policies to control alcohol, tobacco, and high-fat food consumption. *American Journal of Preventive Medicine*, 6(1), 12–19. 10.1016/S0749-3797(18)31039-0 [PubMed: 2340187]
- Johnson AL, Collins LK, Villanti AC, Pearson JL, & Niaura RS (2018). Patterns of nicotine and tobacco product use in youth and young adults in the United States, 2011–2015. *Nicotine & Tobacco Research*, 20(suppl_1), S48–S54. 10.1093/ntr/nty018 [PubMed: 30125012]
- Kessel Schneider S, Buka SL, Dash K, Winickoff JP, & O'Donnell L (2016). Community reductions in youth smoking after raising the minimum tobacco sales age to 21. *Tobacco Control*, 25(3), 355. 10.1136/tobaccocontrol-2014-052207 [PubMed: 26071428]
- Kim SH (2015). Who is responsible for a social problem? News framing and attribution of responsibility. *Journalism & Mass Communication Quarterly*, 92(3), 554–558. 10.1177/1077699015591956
- Kim YM (2009). Issue publics in the new information environment: Selectivity, domain specificity, and extremity. *Communication Research*, 36(2), 254–284. 10.1177/0093650208330253
- King BA, Jama AO, Marynak KL, & Promoff GR (2015). Attitudes toward raising the minimum age of sale for tobacco among U.S. adults. *American Journal of Preventive Medicine*, 49(4), 583–588. 10.1016/j.amepre.2015.05.012 [PubMed: 26163165]

- Levy DT, Friend K, Holder H, & Carmona M (2001). Effects of policies directed at youth access to smoking: Results from the SimSmoke computer simulation model. *Tobacco Control*, 10, 108–116. [PubMed: 11387530]
- Long M, Slater MD, & Lysengen L (2006). US news media coverage of tobacco control issues. *Tobacco Control*, 15(5), 367–372. 10.1136/tc.2005.014456 [PubMed: 16998170]
- Maier D, Waldherr A, Miltner P, Wiedemann G, Niekler A, Keinert A, Pfetsch B, Heyer G, Reber U, Häussler T, Schmid-Petri H, & Adam S (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2–3), 93–118. 10.1080/19312458.2018.1430754
- Majmundar A, Chou C-P, Cruz TB, & Unger JB (2019). Relationship between social media engagement and e-cigarette policy support. *Addictive Behaviors Reports*, 9, 100155. 10.1016/j.abrep.2018.100155 [PubMed: 31193757]
- Mason W, & Suri S (2012). Conducting behavioral research on Amazon’s Mechanical Turk. *Behavior Research Methods*, 44(1), 1–23. 10.3758/s13428-011-0124-6 [PubMed: 21717266]
- McCombs ME (1999). Personal involvement with issues on the public agenda. *International Journal of Public Opinion Research*, 11, 152–168.
- McCombs ME (2014). Why agenda-setting occurs. In *Setting the agenda: Mass media and public opinion* (Second edition, pp. 63–78). Polity Press.
- McCombs ME, & Stroud NJ (2014). Psychology of agenda-setting effects. Mapping the paths of information processing. *Review of Communication Research*, 2(1), 68–93. 10.12840/issn.2255-4165.2014.02.01.003
- McCool J, Cameron LD, & Robinson E (2011). Do parents have any influence over how young people appraise tobacco images in the media? *The Journal of Adolescent Health*, 48(2), 170–175. 10.1016/j.jadohealth.2010.06.012 [PubMed: 21257116]
- Menashe CL (1998). The power of a frame: An analysis of newspaper coverage of tobacco issues-United States, 1985–1996. *Journal of Health Communication*, 3(4), 307–325. 10.1080/108107398127139 [PubMed: 10977260]
- Morain SR, Winickoff JP, & Mello MM (2016). Have Tobacco 21 laws come of age? *New England Journal of Medicine*, 374(17), 1601–1604. 10.1056/NEJMp1603294
- Myers AE, Southwell BG, Ribisl KM, Moreland-Russell S, Bowling JM, & Lytle LA (2019). State-level point-of-sale tobacco news coverage and policy progression over a 2-year period. *Health Promotion Practice*, 20(1), 135–145. 10.1177/1524839917752108 [PubMed: 29338430]
- Myers AE, Southwell BG, Ribisl KM, Moreland-Russell S, & Lytle LA (2017). Setting the agenda for a healthy retail environment: Content analysis of US newspaper coverage of tobacco control policies affecting the point of sale, 2007–2014. *Tobacco Control*, 26(4), 406–414. 10.1136/tobaccocontrol-2016-052998 [PubMed: 27413061]
- Nagelhout GE, van den Putte B, de Vries H, Crone M, Fong GT, & Willemsen MC (2012). The influence of newspaper coverage and a media campaign on smokers’ support for smoke-free bars and restaurants and on secondhand smoke harm awareness: Findings from the International Tobacco Control (ITC) Netherlands Survey. *Tobacco Control*, 21(1), 24–29. 10.1136/tc.2010.040477 [PubMed: 21586760]
- Niederdeppe J, Farrelly MC, & Wenter D (2007). Media advocacy, tobacco control policy change and teen smoking in Florida. *Tobacco Control*, 16(1), 47–52. 10.1136/tc.2005.015289 [PubMed: 17297073]
- Niederdeppe J, Kellogg M, Skurka C, & Avery RJ (2017). Market-level exposure to state antismoking media campaigns and public support for tobacco control policy in the United States, 2001–2002. *Tobacco Control*. 10.1136/tobaccocontrol-2016-053506
- Paek H, & Gunther AC (2007). How peer proximity moderates indirect media influence on adolescent smoking. *Communication Research*, 34(4), 407–432. 10.1177/0093650207302785
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, & Thirion B (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Quinn KM, Monroe BL, Colaresi M, Crespin MH, & Radev DR (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54(1), 209–228.

- Record RA (2017). Tobacco-free policy compliance behaviors among college students: A Theory of Planned Behavior perspective. *Journal of Health Communication*, 22(7), 562–567. 10.1080/10810730.2017.1318984 [PubMed: 28494210]
- Rehder H, & Sojka P (2010). Software framework for topic modelling with large corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50.
- Rennen E, Nagelhout GE, van den Putte B, Janssen E, Mons U, Guignard R, Beck F, de Vries H, Thrasher JF, & Willemsen MC (2014). Associations between tobacco control policy awareness, social acceptability of smoking and smoking cessation. Findings from the International Tobacco Control (ITC) Europe Surveys. *Health Education Research*, 29(1), 72–82. 10.1093/her/cyt073 [PubMed: 23861478]
- Sargent JD, Gibson J, & Heatherton TF (2009). Comparing the effects of entertainment media and tobacco marketing on youth smoking. *Tobacco Control*, 18(1), 47–53. 10.1136/tc.2008.026153 [PubMed: 18948391]
- Setodji CM, Martino SC, Scharf DM, & Shadel WG (2013). Friends moderate the effects of pro-smoking media on college students' intentions to smoke. *Psychology of Addictive Behaviors*, 27(1), 256–261. 10.1037/a0028895 [PubMed: 22686961]
- Shah DV, Cappella JN, & Neuman WR (2015). Big data, digital media, and computational social science: Possibilities and perils. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 6–13. 10.1177/0002716215572084
- Smith DA, & McManis C (2015). Classification of text to subject using LDA. 131–135. 10.1109/ICOSC.2015.7050791
- Smith KC, Wakefield MA, Terry-McElrath Y, Chaloupka FJ, Flay B, Johnston L, Saba A, & Siebel C (2008). Relation between newspaper coverage of tobacco issues and smoking attitudes and behaviour among American teens. *Tobacco Control*, 17(1), 17–24. 10.1136/tc.2007.020495 [PubMed: 18218802]
- Stryker JE, Wray RJ, Hornik RC, & Yanovitzky I (2006). Validation of database search terms for content analysis: The case of cancer news coverage. *Journalism and Mass Communication Quarterly*; Columbia, 83(2), 413–426, 428–430.
- Tan ASL, Lee C, & Bigman CA (2015). Public support for selected e-cigarette regulations and associations with overall information exposure and contradictory information exposure about e-cigarettes: Findings from a national survey of U.S. adults. *Preventive Medicine*, 81, 268–274. 10.1016/j.ypmed.2015.09.009 [PubMed: 26400638]
- Thrasher JF, Kim S-H, Rose I, Navarro A, Craft M-K, Davis KJ, & Biggers S (2014). Print media coverage around failed and successful tobacco tax initiatives: The South Carolina experience. *American Journal of Health Promotion*, 29(1), 29–36. 10.4278/ajhp.130104-QUAN-11 [PubMed: 24555442]
- Unger JB, Rohrbach LA, Howard KA, Cruz TB, Johnson CA, & Chen X (1999). Attitudes toward anti-tobacco policy among California youth: Associations with smoking status, psychosocial variables and advocacy actions. *Health Education Research*, 14(6), 751–763. 10.1093/her/14.6.751 [PubMed: 10585383]
- U.S. Department of Health and Human Services. (2014). *The health consequences of smoking—50 years of progress: A report of the Surgeon General*. 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.
- Volinsky AC, Kranzler EC, Gibson LA, & Hornik RC (2018). Tobacco 21 policy support by U.S. individuals aged 13–25 years: Evidence from a rolling cross-sectional study (2014–2017). *American Journal of Preventive Medicine*, 55(1), 129–131. 10.1016/j.amepre.2018.03.008 [PubMed: 29776779]
- Wagner W (2010). Steven Bird, Ewan Klein and Edward Loper: *Natural Language Processing with Python, Analyzing Text with the Natural Language Toolkit*. *Language Resources and Evaluation*, 44(4), 421–424. 10.1007/s10579-010-9124-x
- Wang TW, Asman K, Gentzke AS, Cullen KA, Holder-Hayes E, Reyes-Guzman C, Jamal A, Neff L, & King BA (2018). Tobacco product use among adults—United States, 2017. *MMWR. Morbidity and Mortality Weekly Report*, 67(44), 1225–1232. 10.15585/mmwr.mm6744a2 [PubMed: 30408019]

- Weiss GM (2004). Mining with rarity: A unifying framework. *ACM SIGKDD Explorations Newsletter*, 6(1), 7. 10.1145/1007730.1007734
- Wickham H (2009). *Ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer Publishing Company, Incorporated.
- Winickoff JP (2018). Maximizing the impact of Tobacco 21 laws across the United States. *American Journal of Public Health*, 108(5), 594–595. 10.2105/AJPH.2018.304376 [PubMed: 29617615]
- Winickoff JP, McMillen R, Tanski S, Wilson K, Gottlieb M, & Crane R (2016). Public support for raising the age of sale for tobacco to 21 in the United States. *Tobacco Control*, 25(3), 284–288. 10.1136/tobaccocontrol-2014-052126 [PubMed: 25701856]
- Yanovitzky I (2002). Effects of news coverage on policy attention and actions: A closer look into the media-policy connection. *Communication Research*, 29(4), 422–451.

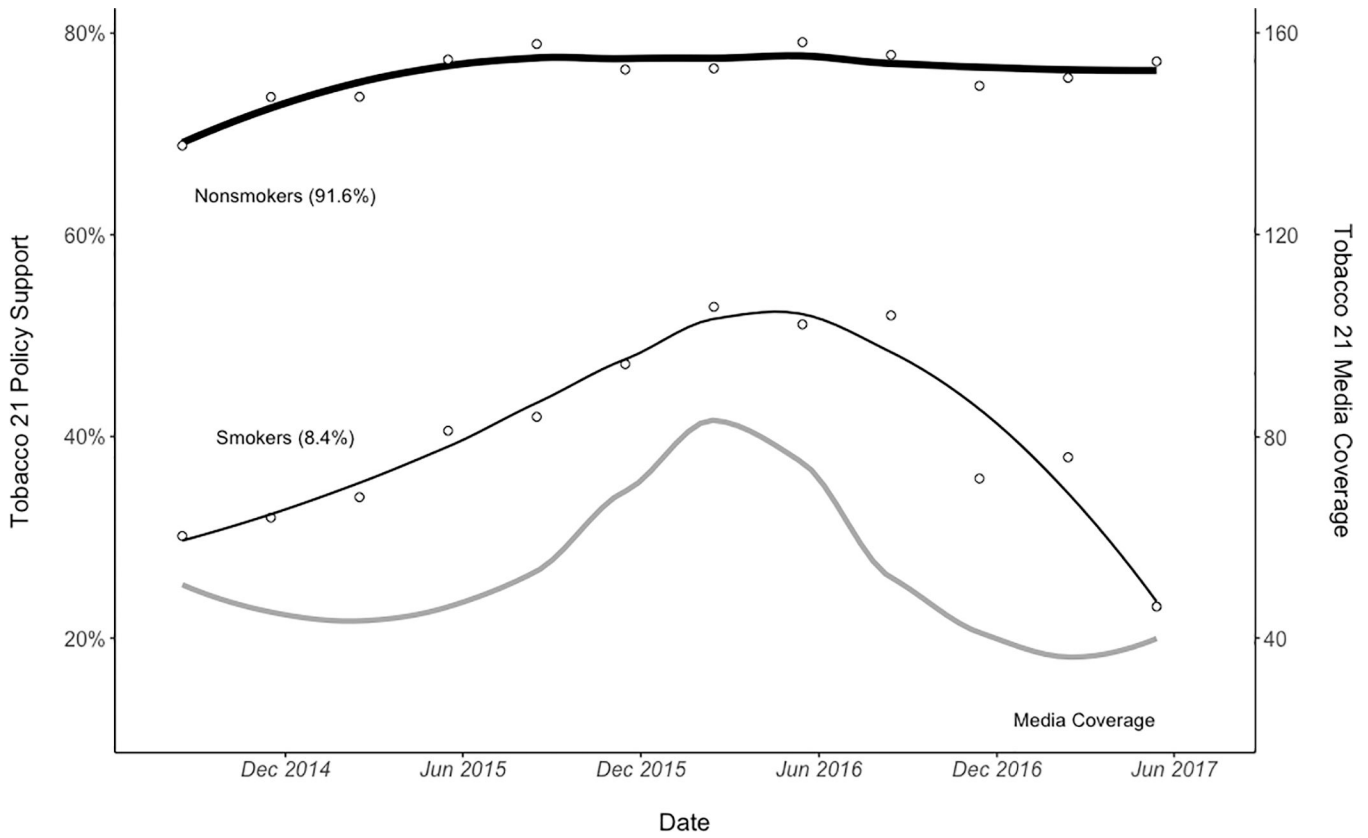


Figure 1. Longitudinal trends in Tobacco 21 media coverage and policy support among young smokers and nonsmokers (by quarter)
Notes: The data included were collected between July 2014 and June 2017. The gray line is a Loess curve showing temporal trends in Tobacco 21 media coverage, measured as the sum of Tobacco 21 texts (identified using LDA) published each quarter, with each text weighted by its percentage contribution from the Tobacco 21 topic (right y-axis). The black lines are Loess curves illustrating temporal trends in weighted mean Tobacco 21 policy support (left y-axis) by quarter among smokers aged 13–20 years (thin line) and nonsmokers aged 13–20 years (thick line).

Table 1.

Ten most predictive words by topic (k=9) in Latent Dirichlet Allocation model and number of texts assigned to each topic

(Topic 1)	(Topic 2)	(Topic 3)	(Topic 4)	(Topic 5)	(Topic 6)	(Topic 7)	(Topic 8)	(Topic 9)
3,308 texts	658 texts	2,479 texts	998 texts	1,453 texts	2,147 texts	899 texts	3,983 texts	1,553 texts
ban	state	health	like	bill	product	said	tax	company
city	cost	percent	one	age	use	new	would	industry
public	company	people	people	would	FDA	product	increase	court
said	money	year	say	21	said	sale	state	said
would	million	public	get	state	study	2016	pack	morris
place	care	cancer	time	california	device	store	million	philip
law	settlement	said	want	said	used	york	revenue	reynolds
area	health	rate	know	measure	new	state	per	smokeless
bar	fund	quit	think	law	health	selling	raise	american
smokefree	program	american	would	18	teen	business	price	market

Notes: Topic 5 was determined to be a topic about Tobacco 21 policies based on an examination of the most predictive words for that topic and a stratified random sample of texts assigned to that topic. The 1,453 texts assigned a higher percentage contribution from this topic than from any other were aggregated to create the measure of Tobacco 21 media coverage used in this study.