



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Master's Thesis of Naeun Kang

Content Analysis of Korean Online
Newspaper Articles on New
Tobacco Products:
Topic Modeling approach

신종담배 관련 국내 온라인 신문 기사 내용 분석:
토픽 모델링 (Topic Modeling)의 적용

August 2023

Graduate School of Public Health
Seoul National
University
Public Health Major

Naeun Kang

Content Analysis of Korean Online Newspaper Articles on New Tobacco Products: Topic Modeling approach

Sung-il Cho

Submitting a master's thesis of Public Health

May 2023

Graduate School of Public Health
Seoul National University
Public Health Major

Naeun Kang

Confirming the master's thesis written by
Naeun Kang

June 2023

Chair Sungho Won (Seal)

Vice Chair Wonkwang Jo (Seal)

Examiner Sung-il Cho (Seal)

Abstract

Objectives: The tobacco industry has used the news media as a way of promoting tobacco products, however, the monitoring of news coverage on tobacco in South Korea has received insufficient attention. This study aims to investigate and comprehend the discourse surrounding New Tobacco Products in Korean online news articles, employing quantitative text mining techniques.

Methods: A text analysis of Korean online news articles on E-cigarettes(EC) and Heated Tobacco Products(HTPs) was conducted. The article samples were sourced from Big Kinds(www.bigkinds.or.kr), utilizing the search terms “New and emerging tobacco products” , “e-cigarettes” , and “Heated tobacco products” . The analysis involved comparing words’ log odds ratio based on the type of news articles(economic and general) and words’ TF-IDF based on the publication company to discern differences in word usage. Furthermore, latent Dirichlet allocation (LDA) and structural topic modeling (STM) were employed to identify the inherent topics within the articles.

Results: A total of 5,162 articles were published between January 1st, 2019, and December 31st, 2022. The number of articles exhibited a declining trend during this period, and the frequently used words varied depending on the type of article and the publication company. With a total of 14,001 unique words, this study discovered a total of 11 topics. Shared topics between LDA and STM included ‘New Product Launching’, ‘Harm of e-cigarettes and Discontinuation in Korea’, ‘Health Risks and Use Behaviors of Electronic Cigarettes in Korea’, ‘Industry Business Performance’, ‘Illegal Use of Tobacco Products’, ‘Health Risks of e-cigarettes revealed in U.S.’, and ‘Tobacco Control Policy (P, W) in Korea’ mostly related to P and W in MPOWER measures, ‘Taxation Plan

focused on E-cigarettes (R)', and 'Comprehensive Taxation Plan in Korea'. 'New Product Launching' was the most frequent in both topic models. Moreover, the topics in the news articles on EC and HTPs exhibited a conflict between a marketing-oriented position and a policy-oriented position on electronic cigarettes.

Conclusions: Analyzing media coverage on diseases and health information holds significant importance in promoting public health and establishing proper reporting guidelines. By exploring the discourse on EC and HTPs in Korean online news articles, this study observed frequent TAPS activity in the news coverage, which varied based on the type of newspaper and consistently utilized specific and changing rhetoric. To achieve the tobacco endgame, it is crucial to implement real-time and systematic monitoring of tobacco-related media contents, with a focus on the source of information and media type used by individuals, to prevent people from using tobacco products.

Keyword: E-cigarettes, Heated Tobacco Products, newspaper, topic modeling, LDA, STM

Tables

<Table 1.> Summary Table of Previous Studies	6
<Table 2.> Classification of Tobacco Products in South Korea	15
<Table 3.> Topic Classification in Previous Studies	22
<Table 4.> Descriptive Statistics of the articles on E- cigarettes and HTPs	27
<Table 5.> Frequent Words in the News Articles on E- cigarettes and HTPs	29
<Table 6.> More Frequent Words in Economic Daily News Articles than General Daily News Articles	30
<Table 7.> More Frequent Words in General Daily News Articles than Economic Daily News Articles	31
<Table 8.> Comparison Table of TF-IDF by the Newspaper Company	33
<Table 9.> Distribution Topics and Words from LDA.....	49
<Table 10.> Distribution Topics and Words from STM.....	51

Figures

<Figure 1.> Tobacco Sales Trend in South Korea (2014~2022)	2
<Figure 2. > Plate Notation representing the LDA and STM18	
<Figure 3.> Work flow of the Analysis Process	26
<Figure 4.> Comparison Chart of Words' Log Odds Ratio by the Type of newspaper	32
<Figure 5.> Comparison Chart of TF-IDF by the newspaper company	37
<Figure 6.> Diagnostic Values by the Number of Topics in LDA	41
<Figure 7.> Diagnostic Values by the Number of topics in STM	41
<Figure 8.> MPOWER Measures for Tobacco Control.....	46
<Figure 9.> Comparison Chart of Topics by Topic Model...	57
<Figure 10.> Comparison of Topic Proportion from STM by the Type of news articles.....	58
<Figure 11.> Trends of Topics over time (2019–2022)	60
<Figure 12.> Conflict between Positions on Electronic Cigarettes.....	76
<Figure 13.> Harm Reduction Strategy by Tobacco Industry (BAT).....	77

Table of Contents

Chapter 1. Introduction

1.1 Backgroud.....	1
1.2 Literature Review	4
1.2.1 Tobacco in Media.....	4
1.2.2 Tobacco Products in Newspaper Articles	9
1.2.3 The Influence of Newspaper Types on News Articles: General vs. Economic(business)	11
1.3 Objectives.....	13

Chapter 2. Methods

2.1 Operational Definition	14
2.2 Data Collection	15
2.3 Statistical Analysis	16
2.3.1 Pre-processing	16
2.3.2 Descriptive Analysis.....	16
2.3.3 Words' Log Odds Ratio and TF-IDF	17
2.3.4 Topic modelling: LDA, STM	18
2.4 Validity.....	24
2.4.1 Setting the number of Topics	24
2.4.2 Cross-checking for Topic Labelling	25

Chapter 3. Results

3.1 Descriptive Characteristics.....	26
3.2 Words Comparison by the Type and Company	30
3.3 Topics from LDA and STM	41
3.4 Trends of Topic Proportion	59

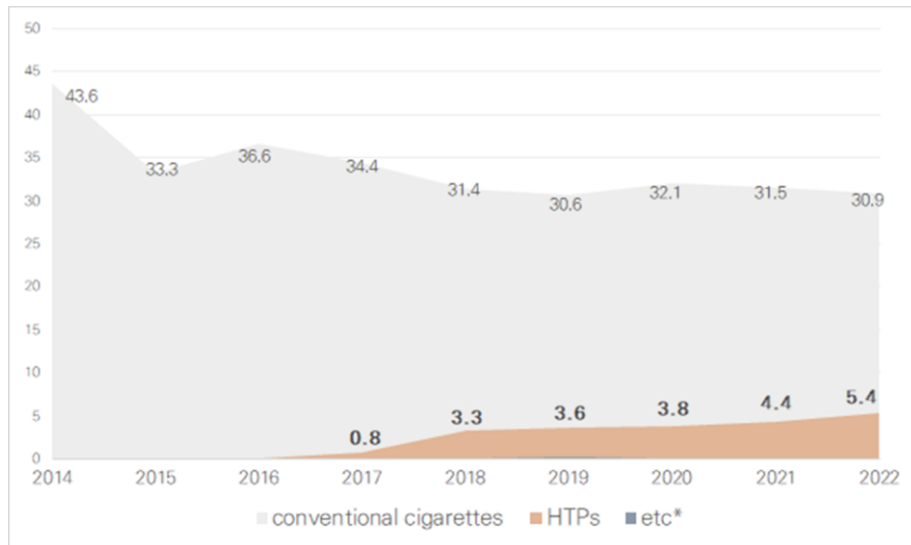
Chapter 4. Discussion	75
-----------------------------	----

Chapter 5. Conclusion	82
Bibliography	83
Abstract in Korean.....	94

Chapter 1. Introduction

1.1. Background

As a signatory to the World Health Organization (WHO) Framework Convention on Tobacco Control (FCTC) in 2005, South Korea has implemented various tobacco control policies aimed at enhancing public health. These measures encompass initiatives such as increasing cigarette tax rates and expanding non-smoking areas [1]. Consequently, there has been a consistent decrease in the prevalence of smoking among adult men over the past two decades in South Korea. However, despite this progress, approximately 30% of adult men persist in using conventional cigarettes, and the smoking rate among women remains stagnant [2]. Furthermore, with the ongoing surge in sales and market share of Heated Tobacco Products (HTPs) in Korea, the overall usage rate of tobacco products, including what are commonly referred to as "New and Emerging Tobacco Products," is projected to surpass that of conventional cigarette smoking (Figure 1) [3]. Since the introduction of Electronic Nicotine Delivery Systems (ENDS or E-cigarettes, EC) in 2008 and HTPs in 2017 to the Korean tobacco market, there has been a notable increase in their adoption among both adolescents and adults [4]. This surge in usage extends to novel tobacco products, particularly among individuals who use multiple tobacco products [5].



*etc: E-cigarettes which heat e-liquids or solid substances

Figure 1. Tobacco Sales Trend in South Korea (2014~2022)

WHO has highlighted the proliferation of "New and Emerging nicotine and tobacco products" across various markets in recent years, while the tobacco industry suggests that ENDS are comparatively safer than conventional cigarettes[6]. The tobacco industry actively promotes HTPs and e-cigarettes as being "less harmful" to the human body. These products are marketed as aids for smoking cessation and are aggressively targeted towards young individuals through Social Network Services[4]. As a result, many individuals perceive HTPs as being less detrimental to health or as potential alternatives for quitting conventional cigarette smoking[7–8]. This marketing strategy employed by the industry has led to a continual expansion of its customer base, including the attraction of younger users.

Nonetheless, previous studies have demonstrated that the use of EC or HTPs can cause health risks. WHO noted that E-cigarettes don't contain tobacco but are harmful to health, and HTPs are inherently toxic and contain carcinogen[9]. Of particular concern regarding EC is the increased risk of users inhaling

excessive amounts of nicotine due to the dual use of e-cigarettes and conventional cigarettes. Additionally, the aerosols discharged during the utilization of HTPs contain nicotine and heavy metals[10]. Above all, there is no clear evidence yet that the use of EC and HTPs has reduced the health risk of the population, so it is necessary to be careful. Therefore, it is essential to exercise caution when approaching the use of new tobacco products due to these health risks, and attention should be paid to the escalating purchase of these novel cigarettes.

Article 13 of WHO FCTC mandates the implementation of a comprehensive ban on ‘Tobacco Advertising, Promotion, and Sponsorship (TAPS)’, which recommends restrictions on the use of direct and indirect incentives to promote the purchase of tobacco products by the public, through both traditional media (print, television and radio) and all media platforms including Internet, mobile telephones and other new technologies as well as films[11]. Advertising and promotion of tobacco in Korea are currently possible only within a limited range, including exhibitions of advertisements in the retailer's business place, advertisements on magazines or newspapers published periodically once or less a week within 10 times a year, and advertisements in an airplane and a passenger ship operating on an international line[12–13]. In particular, newspapers published regularly once a week or less under the Newspaper Promotion Act are legal only when advertising cigarettes 10 times a year if they are not intended for women or adolescents.

However, a previous study revealed that news articles pertaining to tobacco industry activities extensively cover detailed information regarding HTPs, including release dates, pricing information, product features, and specific retail locations where these products are sold[14]. And it found that a close correlation between the surge in online news articles focusing on HTPs and people's search behavior on the Internet.

1.2. Literature Review

1.2.1. Tobacco in Media

Media can have significant impacts on individual and population health[15]. Exposure to media, especially among youth, may affect health behaviors such as substance use, sexual activity, and eating habits[16]. Because well-executed health mass media campaigns can have effects on health knowledge, beliefs, attitudes, and even behaviors as well, therefore, Mass media campaigns have long been a tool for promoting public health.

Since the reporting frame or tone has an important influence on the process of individuals' recognition of the importance of health problems or the formation of a willingness to solve health problems, news media has a great influence on the public's perception and attitude toward health problems[17–18]. This is because the public sees and listens to various information on health-related issues through the media, and is greatly influenced by their perceptions and attitudes. The influence of the media on health information and behavior has both positive and negative aspects.

Newspapers, one of the media outlets, are also effective in spreading health-related information and have a kind of educational nature that delivers health-related information along with disease prevention and treatment information[19]. On the other hand, there are many studies that report the negative effects of the media. Previous studies focusing on cigarettes that have appeared in the media have pointed out the risk that frames, tone, and even sales information include factors that beautify the image of cigarettes and induce the use of cigarettes through various media channels[14,20].

Prior research examining media coverage of tobacco-related issues underscored the significant impact of media coverage on public knowledge, attitudes, and behaviors concerning tobacco use and tobacco control policies[21–34] (Table 1). And a considerable body of literature emphasized the pivotal role played by the tobacco

industry in shaping media coverage with the aim of promoting their products or exerting influence over public opinion. In addition, several studies discussed the need for effective tobacco control strategies, encompassing initiatives such as risk communication, public awareness campaigns, and the enactment of more stringent tobacco control laws.

Table 1. Summary Table of Previous Studies

Title	Region	Subject	Medium	Method	Main findings
Durrant, R. et al., 2003	Australia	tobacco–related issues	newspaper articles (n=1,188)	content analysis	The most frequent themes were secondhand smoke and education/prevention/cessation programs. The majority of articles were positive for tobacco control objectives.
Smith, K. C. et al., 2005	U.S., Australia	tobacco issues	newspaper articles (n=1,188 / n=1,317), 2001	comparative analysis	The expression of hostile opinion toward tobacco control efforts differed. (U.S., 4%; Australia, 7.1%) Secondhand smoke and education, cessation, and prevention efforts dominated coverage in Australia (29.2%) more than in the United States (17.6%).
Løchen, M. L. et al., 2007	Norway	tobacco promotion, tobacco–related health topics	magazines, newspaper articles	content analysis (kappa statistic)	Coverage promoting smoking was more common than coverage of tobacco and health, especially in men's magazines.
Haddock, C. K. et al., 2008	U.S.	tobacco messages and advertisements	Military Times magazines	content analysis	While no advertisements for cigarettes or smoked tobacco were found, there were frequent advertisements for smokeless tobacco products. The ads aimed to associate smokeless tobacco use with positive emotions and enjoyable activities.
Harris, J. K. et al., 2010	U.S. Missouri	tobacco issues	newspaper articles (n= 1,263)	content analysis	The majority of articles were news stories and pro–tobacco control, but there was an increase in anti–tobacco control articles, particularly economically themed ones, in the month before the election.

Braun, S. et al., 2011	Argentina	tobacco–related content	newspaper articles (n=336)	time–trend analysis	Advertisements were predominantly found in specific sections, while non–advertisement images appeared in sports and entertainment sections. 39% focused on health topics and 55% emphasized the negative effects of tobacco on health.
Wackowski, O. A. et al., 2013	UK	smokeless tobacco	newspaper articles (n=677) opinions (n=176), 2006–2010	content analysis	Health risk references were present in 40% of the articles, and while the majority of opinion articles had an anti–SLT slant, a significant portion (25.6%) were pro–SLT.
He, S. et al., 2014	China	tobacco issues	newspaper articles (n=4,821), 2008–2011	content analysis	Education, prevention, and cessation programs were the most frequent theme. The majority of articles were positive for tobacco control.
Rooke, C. et al., 2014	U.K., Scotland	Electronic Nicotine Delivery Systems (ENDS)	newspaper (n= 119), 2007–2012	thematic analysis	Newspaper coverage of ENDS increased substantially over this period.
Nelson, D. E. et al., 2015	U.S.	tobacco	newspapers, newswires, television	content analysis	From 2004 to 2010, TV stories were more focused on cessation, addiction, and health effects, while newspapers and newswires had more variation in the choice of tobacco themes covered.
Sun, S., Veltri, G. A. et al., 2018	China	e–cigarettes	newspaper articles (n=476), 2004. 3. –	content analysis, supervised automatic	E–cigarettes have not received enough attention in terms of its salience in the media, though the amount of coverage has been growing.

			2016. 3.	text analysis	
Hwang, Ji-eun., 2020	South Korea	HTPs	news coverage (n=533), 2017–2020	content analysis	Article and photographic news, excluding etc., accounted for 37.1% of tobacco company activities and comprised news about policies (29.7%), health-related issues (11.6%), and trends (11.0%).
Joseph, N. et al., 2021	India	tobacco appearances, pro or anti-tobacco messages	newspapers, television, YouTube	content analysis	In newspapers, tobacco appearances were found in 0.3% of advertisements, all of which were pro-tobacco messages. On YouTube, 11.44% of videos had tobacco appearances, and 98.5% of the incidents were pro-tobacco messages.
Janmohamed, K. et al., 2022	Nigeria	shisha use	national newspapers	content analysis	Shisha smoking was gaining popularity among youths, with reports of misconceptions about its safety and addiction. There were subtle advertisement of shisha lounges, which violates tobacco control laws.
Gunnar, S. et al., 2022	Norway	Snus	newspaper articles (n= 943), 2002–2011	quantitative content analysis	The number of articles about snus increased over time, with the greatest attention focused on "snus use," followed by themes related to tobacco policy, economy/markets, and health. The majority of articles were neutral/mixed or negative in tone, suggesting that the newspaper coverage did not glamorize snus, but the growing amount of coverage may have contributed to the normalization of snus use.

1.2.2. Tobacco Products in Newspaper Articles

In South Korea, studies focusing on cigarettes in the media have largely dealt with conventional cigarettes. One research analyzed the conflicts between non-smoking and smoking discourses, which showed that there are oppositions between price escalation and budget expansion, between expansion of non-smoking area and infringement on a right and between public welfare lawsuit and lawsuit useless[35]. Another study, which analyzed news coverage on tobacco control policy from 5 Korean daily newspapers and 3 broadcasting stations between 2011 and 2016, showed that a conflict between pro-policies and anti-policies[36]. Especially, there was mostly negative tone on 'Cigarette tax' but mostly positive on 'Smokefree zone'. Other previous studies mainly dealt with how positions and opinions were formed on conventional cigarettes, focusing on social issues such as legal litigation[37].

Furthermore, there was a study that analyzed what discourse was formed in newspaper articles on HTPs. A previous study, which analyzed a total of 571 news representations of the regulation of HTPs in South Korea, demonstrated taxation and warning labels were two regulation topics covered the most[38]. And a content analysis, which examined a total of 766 news articles and TV news stories about HTPs in South Korea, showed the news stories on benefits of HTPs focused on reduced harm, greater social acceptability, and convenience, while discussions of drawbacks included potentially being equally or more harmful than cigarettes, the possibility of extensive future regulations, and HTPs' unknown health effects[39]. From these studies, the government and the tobacco industry were the source cited often across newspaper. From a study, which investigated the marketing strategies of tobacco companies on news media, by reviewing the previous research analyzing the tobacco companies' documents and the news coverage of tobacco, showed that tobacco companies

have created close relationships with news media, a quid pro quo, by providing advertising and diverse forms of sponsorships and obtaining favored treatment on news coverage and have built networks with journalists by supporting journalist organizations [40]. Furthermore, from the study, news coverage on tobacco showed that news media tended to deal with tobacco issues lukewarmly, without a clear position and perspective and deliver naturally tobacco industry's perspective and message that smoking is controversial rather than totally undesirable and unacceptable. Also, the news coverages have often reflected tobacco industry's publicity strategies identified in tobacco industry documents [36].

In review, previous studies focusing on tobacco products that appeared in the media shared the following two similarities. First, the studies mainly focused on conventional cigarettes. Even if some studies have explored tobacco in news coverage, few have specifically focused on articles pertaining to EC and HTPs in South Korea. Tobacco products have diversified with various type including even mixed products that take the form of EC and HTPs [41], moreover, EC is expanding its users, especially for teenagers. And the usage rate of electronic cigarettes plunged in 2020 when COVID-19 began to spread in earnest and increased again the following year [42]. Taking these circumstances, therefore, it is needed to pay attention to not only conventional cigarettes but also the New Tobacco Products, which includes EC and HTPs.

Second, qualitative methods, in which researchers directly analyze a small number of articles through coding, was predominantly utilized in most studies on newspaper articles with cigarette depictions. However, amid the recent rapid production and distribution of text, the need to analyze large amounts of text is increasing, and newspaper articles are no exception. In particular, when analyzing text, the method called Topic Modeling has been used as a tool to understand text in the media in various studies so far. Topic Modeling is one of the text mining techniques and has

been used as an auxiliary means of accessing discourse. A study analyzed a total of 745 news articles on thirdhand smoke by using topic modelling, showing that topic modeling can offer insights into understanding news reports related to thirdhand smoke [29]. In Korea, it is quite rare to conduct a topic modeling study on newspaper articles on tobacco products, and there is no research on the news articles on comprehensive new tobacco products.^①

1.2.3. The Influence of Newspaper Types on News Articles: General vs. Economic(business)

The news production process affects various factors, and for this reason, there are news differences according to media despite the same issues [44]. In the case of newspapers, there are differences by the type of newspapers, which means general and economic(business), in the concept, article contents and reporting methods [45]. In particular, Economic newspapers have been criticized that rely on corporate information than General newspapers. In particular, Economic newspapers have a pro-business and pro-industrial reporting attitude compared to General newspapers, so they are criticized for relying more on corporate information sources than general newspapers and representing corporate positions. Until now, most of the content analysis of domestic newspaper articles, especially health-related articles, has been conducted for general newspapers. Considering the activities

^① DiMaggio, P., Nag, M., & Blei, D. presented the five mechanisms [43], for the reasons why we should focus on newspapers, through which measurements of media content may tap factors that shape individual and collective perceptions and understandings:

- (1) Priming of existing schematic representations
- (2) Development of new representations
- (3) Integration with broader schemata
- (4) Indirect influence through selective re-telling
- (5) Proxy value

of tobacco companies that have been striving to build a positive image of the tobacco industry through various media such as newspapers, it is necessary to consider the types of newspapers such as general and economic newspapers when analyzing newspaper contents.

1.3. Objectives

This study aims to investigate the content of online news articles in Korea, with the keywords on E-cigarettes and HTPs, given the necessity to monitor aggressive marketing practices in the background of growing sales volume of ‘New Tobacco Products.’ The specific aims of this study are as follows:

First, to investigate the descriptive characteristics of Korean online daily news articles that incorporate keywords related to ‘New Tobacco Products’

Second, to explore the frequently used words by the type of newspapers and the respective newspaper company.

Third, to gain insight into the discourse surrounding ‘New Tobacco Products’ in Korea by examining the inherent topics within the news articles.

Chapter 2. Methods

2.1. Operational Definition

WHO has categorized “Novel and emerging nicotine and tobacco products” into two categories: Heated Tobacco Products (HTPs) ^② and E-cigarettes ^③ [46]. Based on this classification, this study adopts the definition of new tobacco products as e-cigarettes and HTPs. In South Korea, e-cigarettes are commonly referred to as “(액상형) 전자담배” and HTPs are known as “가열담배”, “찜담배”, and “궐련형 전자담배”, based on the legal classification of tobacco in Korea (Table 2) [47]. Both types of products share the common name “전자담배 (electronic cigarettes)”. Therefore, this study employed ‘전자담배’ as the primary search term, supplemented by additional related terms such as ‘신종담배’, ‘가열담배’, ‘찜담배’, ‘액상담배’. ^④ All Korean online news articles containing any of these five search terms were selected as the subjects of analysis. ^⑤ Only the body text of the articles was utilized for this study.

The newspapers were classified into two types: national general daily newspapers and economic daily newspapers according to the classification criteria provided by Bigkinds ^⑥.

^② Heated tobacco products (HTPs), a hybrid between conventional and electronic cigarettes heat, rather than combust, processed tobacco in a controlled fashion. Tobacco companies actively promoted HTPs as being odorless and less harmful to health.

^③ There are many different types of e-cigarettes in use, also known as Electronic Nicotine Delivery Systems (ENDS) and sometimes Electronic Non-nicotine Delivery Systems (ENNDS). These systems heat a liquid to create aerosols that are inhaled by the user. Electronic cigarettes (or e-cigarettes) are the most common form of ENDS and ENNDS. But ENDS contain varying amounts of nicotine and harmful emissions.

^④ "신종담배" OR "전자담배" OR "가열담배" OR "찜담배" OR "액상담배"

^⑤ The words "궐련형" and "액상형" were excluded from the search terms because the same search result could be obtained with the common word "전자담배" of the two words in Bigkinds, without searching for "궐련형" and "액상형" respectively.

^⑥ <https://www.bigkinds.or.kr/>

Table 2. Classification of Tobacco Products in South Korea

1. Cigarettes: Tobacco made by shredding leaf tobacco of a certain width after adding flavoring, etc. and rolling it with cigarette paper for easy smoking with a cigarette manufacturing machine, or similar type of tobacco that can be used for smoking;
2. Electronic cigarettes: Tobacco classified as follows, which is made to cause the same effect as smoking by inhaling nicotine solution or shredded tobacco and a tobacco stick into the body through respiratory organ with an electronic device: <ul style="list-style-type: none">(a) Electric cigarettes using nicotine solution(b) Electric cigarettes using shredded tobacco and a tobacco stick:<ul style="list-style-type: none">(i) Cigarette types;(ii) Other types;

2.2. Data Collection

This study examines online news articles obtained from a combined total of 11 national and 8 economic daily newspapers, spanning the period from January 1, 2019 to December 31, 2022. Since previous studies on tobacco products in Korea conducted so far targeted newspaper articles until 2018, the period from 2019 to 2022 was set as the analysis target period. The comprehensive metadata of the articles selected for analysis was collected using BigKinds. The collected information includes the publication date, newspaper company, journalist, article titles, integrated classification (content category), and a compilation of nouns from the body text of each article, all recorded in an Excel file. In order to account for the characteristics of framing and to align with the study objective, duplicate articles were also included as subjects of analysis.^⑦

^⑦ To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation,

2.3. Statistical Analysis

2.3.1. Pre-processing

In the text preprocessing stage, punctuation marks, Chinese characters, and one-letter words, which are irrelevant for analysis purposes, were removed. Furthermore, multiple letter words were standardized to their singular forms. In addition, only words that appeared more than five times were considered for the final analysis.^⑧ Moreover, articles containing fewer than three words were excluded from the analysis. The words within each article were transformed into a bag of words (BoW) representation to perform topic modeling, and a Document-Term Matrix (DTM) was constructed based on this representation.

2.3.2. Descriptive Analysis

Descriptive analysis was conducted to understand the overall reporting trends of the study subject, including the frequency and percentage of the articles and the usage of words over time, categorized by newspaper company, newspaper type.

and/or treatment recommendation for the item described [48]. In short, frames highlight some bits of information about an item that is the subject of a communication, thereby elevating them in salience. To consider the impact of the newspaper type on the texts of news articles, this study included all duplicate articles.

^⑧ When words were sorted in descending order based on the frequency of appearance and then the cumulative sum was calculated, words that appeared more than five times accounted for 97.4% of the total text to be analyzed. In other words, these words are considered to represent the majority of the document, and words with less than five appearances were excluded from the analysis for the ease of topic modeling analysis.

2.3.3. Words' Log Odds Ratio and TF-IDF

According to the classification of daily newspapers, the word Log Odds Ratio was calculated and compared to identify relatively frequent words specific to each newspaper type, namely 'economic' and 'general'. The odds is the ratio of the probability that the event of interest occurs to the probability that it does not[49]. The log odds ratio can take any value and has an approximately Normal distribution[50]. It has the useful property that if researchers reverse the order of the categories for one of the variables. This study used log odds ratio to visualize text differences clearly by representing bars in the opposite direction depending on which text the word appeared more prominently. In addition, even if the main words were compared in consideration of the type of newspaper, each newspaper company has different characteristics such as political ideology and pro-business orientation, so this study additionally checked the frequently used words according to the newspaper company by calculating TF-IDF.

Term Frequency Inverse Document Frequency (TF-IDF) is the multiplication of term frequency (TF) and inverse document frequency (IDF), which determines how relevant a given word is in a particular document[51–53]. TF is how many times a term is present in a document. And the inverse document frequency IDF assigns lower weight to frequent words and assigns greater weight for the words that are infrequent. TF-IDF is evolved from IDF with heuristic intuition that a query term which occurs in many documents is not a good discriminator, and should be given less weight than one which occurs in few documents[54]. The classical formula of TF-IDF used for term weighting is the following equation:

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

where $w_{i,j}$ is the weight for term i in document j , N is the number of documents in the collection, $tf_{i,j}$ is the term frequency of

term i in document j and df_i is the document frequency of term i in the collection.

2.3.4. Topic modeling: LDA, STM

Text mining is one type of data mining technique for extracting or mining knowledge from the text document[55]. It discovers the previously unknown information extracting it automatically from different source[56]. Topic modelling can be described as a method for finding a group of words from a collection of documents that best represents the information in the collection[57]. It can also be thought of as a form of text mining – a way to obtain recurring patterns of words in textual material.

This study used Latent Dirichlet Allocation (LDA) and Structural Topic Modeling (STM) among text mining techniques to extract topics and analyze the contents of target articles. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words[58]. As such, the sum of the topic proportions across all topics for a document is one, and the sum of the word probabilities for a given topic is one. Total process of analysis is represented on Figure 2.

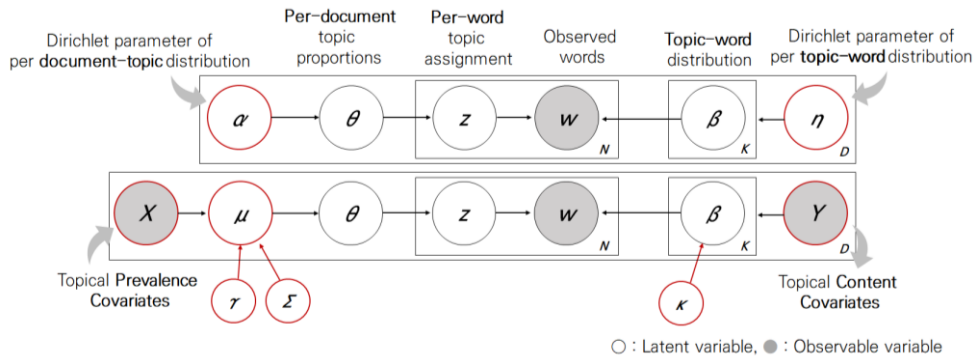


Figure 2. Plate Notation representing the LDA and STM

LDA is the most widely used topic modeling technique, which is a generative probabilistic model for collections of discrete data such as text corpora using a Bayesian model[59].

The number of topics k has to be fixed a-priori for LDA [60]. The LDA model assumes the following generative process for a document $\omega = (\omega_1, \dots, \omega_N)$ of a corpus D containing N words from a vocabulary consisting of V different terms, $\omega_i \in \{1, \dots, V\}$ for all $i = 1, \dots, N$.

For LDA the generative model consists of the following steps.

The term distribution β is determined for each topic by

$$\beta \sim \text{Dirichlet}(\delta).$$

The proportions θ of the topic distribution for the document ω are determined by

$$\theta \sim \text{Dirichlet}(\alpha).$$

For each of the N words ω_i

(a) Choose a topic $z_i \sim \text{Multinomial}(\theta)$.

(b) Choose a word ω_i from a multinomial probability distribution conditioned on the topic z_i : $p(\omega_i | z_i, \beta)$.

β is the term distribution of topics and contains the probability of a word occurring. The full probability formula that generates a document is in below formula[61]:

$$P(\mathbf{W}, \mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\varphi}; \alpha, \beta) = \prod_{i=1}^K P(\varphi_i; \beta) \prod_{j=1}^M P(\theta_j; \alpha) \prod_{t=1}^N P(Z_{j,t} | \theta_j) P(W_{j,t} | \varphi_{Z_{j,t}})$$

Since the development of LDA, models with various conditions have been developed to increase the efficiency and accuracy of topic estimation. STM, one of the representative methods, is also a generative model of word counts. In this model, topical prevalence and topical content are specified as a simple generalized linear model on an arbitrary number of document-level covariates, such as news source and time of release, enabling researchers to introduce elements of the experimental design that informed document collection into the model, within a generally applicable framework [62].

The generative process for each document (indexed by d) with vocabulary of size V for a STM model with K topics can be summarized as [63]:

Draw the document-level attention to each topic from a logistic-normal generalized linear model based on a vector of document covariates X_d .

$$\vec{\theta}_d | X_d \gamma, \Sigma \sim \text{LogisticNormal}(\mu = X_d \gamma, \Sigma)$$

where X_d is a 1 -by- p vector, γ is a p -by- $(K - 1)$ matrix of coefficients and Σ is a $(K - 1)$ -by- $(K - 1)$ covariance matrix.

Given a document-level content covariate y_d , form the document-specific distribution over words representing each topic (k) using the baseline word distribution (m), the topic specific deviation $k_k^{(t)}$, the covariate group deviation $k_{y_d}^{(c)}$ and the interaction between the two $k_{y_d, k}^{(i)}$.

$$\beta_{d, k} \propto \exp(m + \kappa_k^{(t)} + \kappa_{y_d}^{(c)} + \kappa_{y_d, k}^{(i)})$$

m , and $k_k^{(t)}$, $k_{yd}^{(c)}$ and $k^{(i)}_{yd,k}$ are V -length vectors containing one entry per word in the vocabulary.

For each word in the document, ($n \in \{1, \dots, N_d\}$):

Draw word's topic assignment based on the document-specific distribution over topics.

$$z_{d,n} | \vec{\theta}_d \sim \text{Multinomial}(\vec{\theta}_d)$$

Conditional on the topic chosen, draw an observed word from that topic.

$$w_{d,n} | z_{d,n}, \beta_{d,k=z_{d,n}} \sim \text{Multinomial}(\beta_{d,k=z_{d,n}})$$

$$\beta_{d,k,v} = \frac{\exp(m_v + \kappa_{k,v}^{(t)} + \kappa_{y_d,v}^{(c)} + \kappa_{y_d,k,v}^{(i)})}{\sum_v \exp(m_v + \kappa_{k,v}^{(t)} + \kappa_{y_d,v}^{(c)} + \kappa_{y_d,k,v}^{(i)})}, \quad \text{for } v = 1 \dots V \text{ and } k = 1 \dots K,$$

In this study, meta information is intended to largely reflect the two types of covariates: newspaper type and publication date(time). From STM, the top 20 words based on appearance probability and FREX, and five top documents with high relevance by topic were derived. FREX is a metric to summarize topics that combines term frequency and exclusivity to that topic into a univariate summary statistic[63–64]. This statistic calculates the harmonic mean of the empirical CDF of a term's frequency under a topic with the empirical CDF of exclusivity to that topic. FREX statistic is defined as:

$$\text{FREX}_{k,v} = \left(\frac{\omega}{\text{ECDF}(\beta_{k,v} / \sum_{j=1}^K \beta_{j,v})} + \frac{1 - \omega}{\text{ECDF}(\beta_{k,v})} \right)^{-1}$$

Following the topic modeling process, this study assigned topic labels to each group of words generated by the LDA and STM algorithms. During this stage, the topic classifications used in previous studies were taken into account (Table 3). The

classifications were categorized into three main themes, namely health effects, industry, and regulation, with variations based on the level of granularity in the classification. R studio was used for statistical analysis and result visualization.

Table 3. Topic Classification in Previous Studies

Title	Classification
Durrant, R. et al., 2003	secondhand smoke issues, education and prevention, health effects, tobacco Industry, advertising and promotion, consumption, economic issues, product issues, addiction, youth access, other, environmental damage, farming
Løchen, M. L. et al., 2007	tobacco and health, indirect tobacco advertisements, pictures with individuals smoking
Harris, J. K. et al., 2010	youth prevention, adult cessation, smoke-free policies, tobacco taxes, tobacco science
Braun, S. et al., 2011	health effects, legislation, advertising, promotions or sponsorship, cigars and economic issues
Wackowski, O. A. et al., 2013	business, product regulation, harm reduction, prevention/cessation
He, S. et al., 2014	education and prevention, health effects, second-hand smoke, tobacco-related crimes, economics, tobacco industry, farming, products and regulations, unintended damage, tobacco consumption, youth access, other, female smokers, advertising
Rooke, C. et al., 2014	getting around smokefree legislation, risk and uncertainty, healthier choice, celebrity use, price
Sun, S., Veltri, G. A. et al., 2018	nicotine/constituents/features, tobacco control/regulation, children's use of E-cigarettes, the tobacco market/industry
Hwang, Ji-eun., 2020	policy and regulation, health related issue, trend, industry activities, etc.
Gunnar, S. et al., 2022	Snus use, tobacco policy, economy/market, health

The prominent improvements of STM are that the document-level structure information is introduced to influence topical prevalence (i.e., per-document topic proportions) and topic content (i.e., topic-word distributions), thereby emphasizing the suitability of investigating how covariates affect text content. Therefore, both LDA and STM analysis were performed in the judgment that it was necessary to compare the performance and result derivation of each model by referring to the previous studies[65–66]. To find the impact of the meta-information on topic extraction, topics in two models were compared in two stages:

First, if the topics, a mixture of words, in LDA and STM were identical or similar, they were considered to be the potential similar topics.

Second, if the same articles in each topic matched most, those topics were considered as similar topics.

Moreover, this study further investigated the consistency of topics related to TAPS by examining their changes over time. Among various topic models, Dynamic Topic Models (DTM) are commonly used for capturing topic evolution over time. DTM a generative model is an extension of LDA, which captures the evolution of topics in a sequentially organized corpus of documents[67]. However, DTM assumes that the entire documents should be divided into equal time sections, and it assumes that the topics' evolution remains constant over time. In the case of news articles, which are the focus of this study, they tend to concentrate on specific topics during specific time periods. Considering this characteristic of news articles and the study objectives, considering the impact of meta-information such as the newspaper type, this study further employed STM on a yearly basis to examine whether the topics exhibit consistent patterns over time.

2.4. Validity

Topic modeling analysis still entails the limitation of researcher subjectivity. However, this study made efforts to mitigate these limitations through several strategies:

Firstly, the number of topics is chosen by hyperparameter tuning to ensure optimal coverage and coherence.

Secondly, the process of labeling topics is cross-checked to enhance accuracy and consistency.

Lastly, supplementary analyses are conducted to validate and refine the obtained results, further bolstering the robustness of the findings.

2.4.1. Setting the number of Topics

Topic modeling is to automatically discover the topics from a collection of documents, so in a data set, the K topic distribution must be learned through statistical inference [68]. In this context, the algorithm defines a generative process as a joint probability distribution over both observed and hidden variables. Both LDA and STM necessitate the selection of the number of topics to be identified. For LDA, the optimal number of topics is determined through hyperparameter tuning to specify a variety of topics from at least 2 to up to 30, and then their fitness indices are calculated to identify the most suitable number of topics for analysis. For STM, R function is employed, which generates a series of plots that aid in evaluating the optimal topic numbers via visual representation of diagnostic functions, which are exclusivity, semantic coherence, heldout likelihood, bound, lbound, and residual dispersion [69]. Once the appropriate number of topics is determined through this process and a random start seed number is assigned, both topic models are executed.

2.4.2. Cross-checking for Topic Labelling

In the topic labeling stage, the titles and body text of the top 5 articles assigned to each topic with high proportion of the article which is made up of words in LDA and in STM were also taken into consideration. Meanwhile, most topic models require the additional step of attaching meaningful labels to estimated topics, a process that is not scalable, suffers from human bias, and is difficult to replicate[70]. In order to address these issues, the topic labeling stage of this study involved the participation of three other researchers with expertise in the domain of smoking. These researchers were randomly presented with bundles of words associated with each topic and tasked with labeling each topic using a noun phrase in a single sentence. The labeling results from all researchers were collected and utilized for topic labeling (refer to Appendix 1 for details). A visual representation of the entire analysis process is depicted in Figure 3.

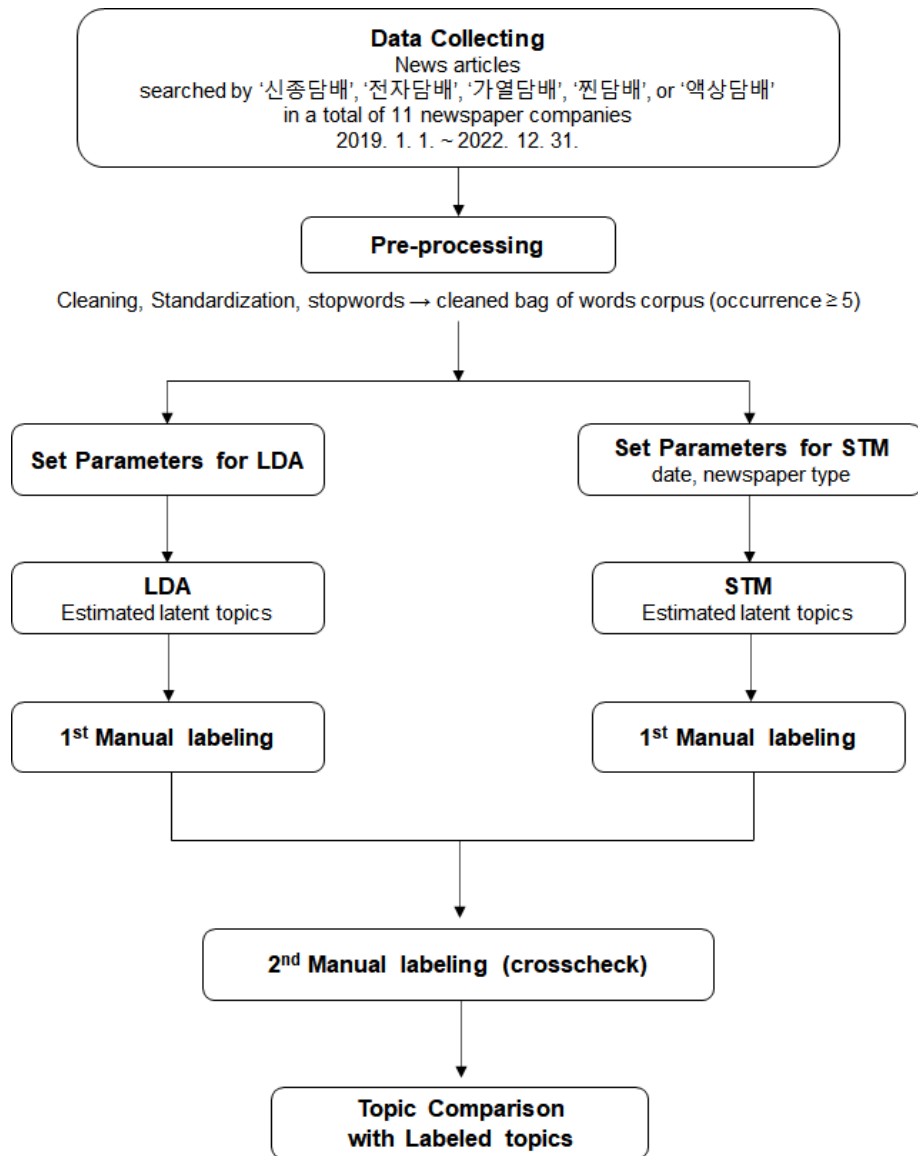


Figure 3. Work Flow of the Analysis Process

Chapter 3. Results

3.1. Descriptive Characteristics

Among the articles reported through a total of 19 Korean online news articles from daily newspapers during the period, a total of 5,162 articles including the search terms were identified (Table 4). The highest number of articles was observed in 2019 but a declining trend was observed in subsequent years. It showed an increasing trend again at the end of 2022. When examining the distribution by newspaper type, it was found that economic daily newspapers accounted for 3,145 articles, while general daily newspapers accounted for 2,017 articles. This indicates that a larger proportion of the articles were reported by economic daily newspapers, with four out of the top five newspapers falling into this category.

Table 4. Descriptive Statistics of the articles on E–cigarettes and HTPs

Category	N	%
Sample articles	5,162	100
Year of Publication		
2019	2329	45.1
2020	1033	20.0
2021	860	16.7
2022	940	18.2
Type of the articles		
General	2,017	39.1
Economic	3,145	61.0
Newspaper Company		
E1	553	10.7
E2	537	10.4
E3	509	9.9
E4	425	8.2
G1	408	7.9
E5	360	7.0
E6	334	6.5

E7	282	5.5
G2	236	4.6
G3	234	4.5
G4	231	4.5
G5	199	3.9
G6	173	3.4
G7	150	2.9
G8	147	2.9
E8	145	2.8
G9	118	2.29
G10	71	1.38
G11	50	0.97

Out of a corpus consisting of 997,622 words, a total of 14,001 unique words were identified as appearing more than five times in the articles subjected to analysis. Among these words, the top 2,000 words based on frequency accounted for 78.4% of the total words. This implies that a minority of 2,000 words encompassed the majority of the content in the articles, suggesting the potential for generalizing the overall content from this minority set of words. Table 5 displays the words with the highest frequencies, ranking "tobacco" (39,806), "electronic" (21,320), "liquid" (9,923), and "e-cigarettes" (8,663) as the most prevalent. Excluding the search terms, other frequently occurring words included "products" (8,371), "sales" (7,158), "cigarettes" (7,034), "smoking" (6,211), "KT&G" (6,042), "use" (6,010), "market" (5,416), "U.S." (4,274), and "government" (4,284).

Table 5. Frequent Words in the News Articles on E-cigarettes and HTPs

No.	Words	Frequency	
		N	%
1	tobacco	39,806	4.22
2	electronic	21,320	2.26
3	liquid	9,923	1.05
4	e-cigarettes	8,663	0.92
5	products	8,371	0.89
6	sales	7,158	0.76
7	cigarettes	7,034	0.75
8	smoking	6,211	0.66
9	KT&G	6,042	0.64
10	use	6,010	0.64
11	market	5,416	0.57
12	U.S.	5,290	0.56
13	launch	4,327	0.46
14	government	4,284	0.45
15	smoking cessation	3,946	0.42
16	nicotine	3,730	0.4
17	smoker	3,418	0.36
18	general	3,297	0.35
19	adolescent	3,250	0.34
20	result	2,865	0.3
21	increase	2,718	0.29
22	health	2,644	0.28
23	consumer	2,567	0.27
24	discontinuity	2,483	0.26
25	harm	2,421	0.26
26	IQOS	2,402	0.25
27	investigation	2,402	0.25
28	expansion	2,286	0.24
29	regulation	2,191	0.23
30	decrease	2,138	0.23

*Some grammatical edits were made to improve comprehension

3.2. Words Comparison by the Type and Company

The frequently used words were compared based on the types of newspapers using log odds ratios.^⑨ Log odds ratios were calculated using the top 2000 words selected according to their overall frequency. Table 6, Table 7, and Figure 4 represent the results of comparing the top 71 words based on the order of small absolute values of the log odds ratio.

In economic daily newspapers, more frequent words appeared in the following order: 'KT&G (0.96)', 'Market (0.65)', 'launching (0.50)', 'expansion (0.47)', and 'increase (0.45)'. On the other hand, in general daily newspapers, the more frequent appeared in the following order: " adolescents " (−0.55), " Human " (−0.51), "smoking" (−0.50), " Investigation " (−0.50), "Smoking cessation" (−0.47), "health" (−0.46), "nicotine" (−0.45), " verification " (−0.38) "material" (−0.36), and " ingredient " (−0.36).

Table 6. More Frequent Words in Economic Daily News Articles than General Daily News Articles

No.	Words	N		Log OR (E/G)
		E	G	
1	KT&G	4,805	1,237	0.96
2	market	4,011	1,405	0.65
3	launching	3,069	1,258	0.50
4	expansion	1,608	678	0.47
5	increase	1,901	817	0.45
6	business	1,307	576	0.42
7	device	1,349	629	0.37
8	PMI	1,262	612	0.33
9	sales volume	1,249	609	0.32
10	IQOS	1,612	790	0.32

^⑨ When the log OR is negative, it means that the words appear relatively more in the general daily newspapers than economic. And the positive OR means the words relatively more in the economic daily newspapers than general.

Table 7. More Frequent Words in General Daily News Articles than Economic Daily News Articles

No .	Words	N		Log OR (E/G)
		E	G	
1	adolescent	1,499	1,751	-0.55
2	human	833	939	-0.51
3	smoking	2,946	3,265	-0.50
4	investigation	1,140	1,262	-0.50
5	smoking cessation	1,899	2,047	-0.47
6	health	1,282	1,362	-0.46
7	nicotine	1,812	1,918	-0.45
8	verification	777	764	-0.38
9	materials	904	875	-0.36
10	ingredient	1,004	968	-0.36

Figure 4. Comparison Chart of Words' Log Odds Ratio by the Type of newspaper

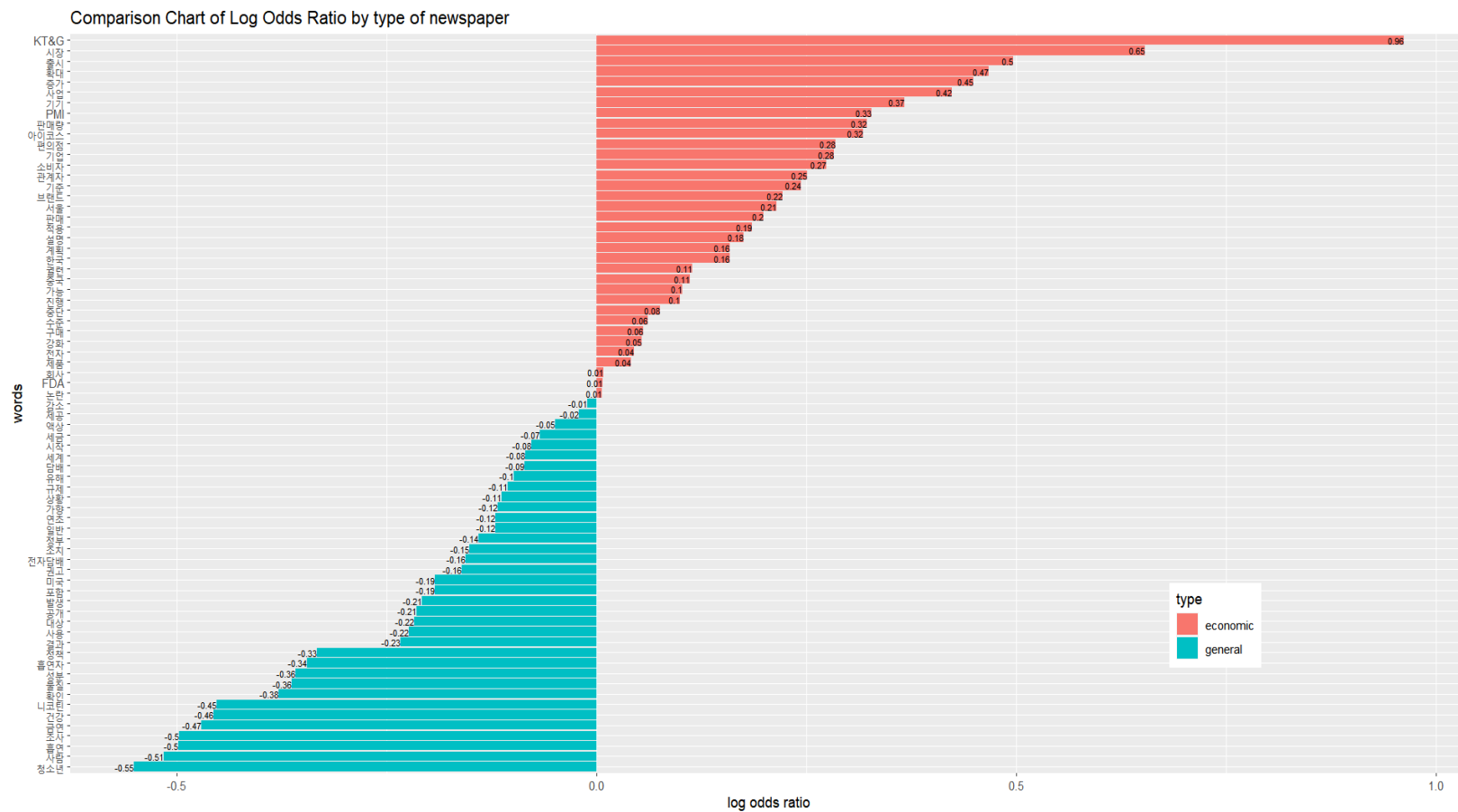


Table 8 and Figure 5 represent the TF–IDF values of each publishing newspaper company. It was observed that each company had distinct words that prominently appeared compared to others. Among a total of 19 companies, it was found that specific products, brand names, or names of tobacco companies were remarkably used in four economic daily newspapers (E2, E3, E4, E8) and three general daily newspapers (G1, G3, G5).

Table 8. Comparison Table of TF–IDF by the Newspaper Company

company	Words	N	tf_idf
E1	***	58	0.00146
	***	41	0.00135
	***	83	0.00124
	stock price	155	0.00111
	***	97	0.00108
	purchase	95	0.00106
	original text	32	0.00105
	operating profit	126	0.00105
	***	31	0.00102
	electromagnetic waves	84	0.00094
E2	jouz	137	0.00100
	UAE	34	0.00087
	operating profit	86	0.00073
	sales figures	139	0.00060
	federation	59	0.00058
	***	17	0.00057
	sale counters	27	0.00057
	Uruguay	27	0.00057
	***	29	0.00051

	5 types	24	0.00050
E3	***	100	0.00157
	***	51	0.00157
	***	80	0.00146
	***	35	0.00141
	prices	83	0.00131
	***	29	0.00117
	***	71	0.00112
	Ploom TECH	93	0.00110
	unicorn	57	0.00104
	interlocking system	24	0.00096
E4	***	59	0.00278
	***	130	0.00208
	stock price	141	0.00145
	***	64	0.00137
	HAKA Korea	52	0.00130
	HAKA Signature	39	0.00115
	law school	24	0.00113
	HNB	38	0.00112
	jouz	99	0.00102
	group purchase	21	0.00099
E5	top pick (stock)	30	0.00185
	dual user	28	0.00173
	recruitment	28	0.00132
	leather	25	0.00118
	carrier	18	0.00111
	***	18	0.00111

	razor	16	0.00099
	***	40	0.00097
	8%	15	0.00093
	click	14	0.00086
E6	NQ	120	0.00711
	***	31	0.00115
	NBCI	18	0.00107
	metabolome	18	0.00107
	the Vatican	17	0.00101
	set exam questions	42	0.00097
	KT	31	0.00097
	standardized goods	16	0.00095
	whisky	21	0.00078
	***	13	0.00077
E7	Ministry of Industry	50	0.00252
	Ministry of Environment and Oceans and Fisheries	39	0.00196
	Ministry of Agriculture, Food and Rural Affairs	31	0.00156
	yuan	40	0.00139
	Statistics Korea	26	0.00131
	streaming	19	0.00125
	Fair Trade Commission	24	0.00121
	AR	18	0.00118
	***	17	0.00112
	Ministry of Economic Affairs	22	0.00111
E8	vitamin E acetate	27	0.00186

	Busan Main Customs	11	0.00167
	HAKA Signature	17	0.00162
	guidebook	10	0.00152
	cake & confectionary	9	0.00137
	group lawsuit	9	0.00137
	last month	9	0.00137
	southeast asia	11	0.00128
	U.S. Securities and Exchange Commission	8	0.00121
	***	8	0.00121
G1	Ploom TECH	108	0.00143
	***	40	0.00095
	JTI	82	0.00094
	2.0	225	0.00081
	rest area	17	0.00077
	service area	17	0.00077
	garbage	38	0.00067
	forties	18	0.00062
	weight	13	0.00059
	smoking scene	13	0.00059
G2	stock company	57	0.00277
	***	33	0.00209
	***	47	0.00158
	lactobacillus	23	0.00146
	***	20	0.00127
	basic income	26	0.00126
	honor	19	0.00121

	Jeju	24	0.00116
	(military) operation	18	0.00114
	dispute	23	0.00112
G3	traffic warden	18	0.00158
	Ploom TECH	56	0.00145
	***	12	0.00105
	airline ticket	11	0.00097
	stamp	10	0.00088
	JTI Korea	34	0.00088
	Vaping room	22	0.00088
	sponsorship	12	0.00081
	BAT Korea	84	0.00077
	***	19	0.00076
G4	secondhand goods	60	0.00258
	meat	29	0.00235
	social media	28	0.00143
	***	16	0.00130
	unicorn	33	0.00121
	dismissal	21	0.00107
	***	13	0.00106
	posts	37	0.00102
	studying abroad	16	0.00099
	stand-in	12	0.00097
G5	stationery store	21	0.00233
	bats (animal)	16	0.00178
	dark web	18	0.00153
	ice flavour	12	0.00133

	laboratory	22	0.00129
	lie	15	0.00128
	HAKA Signature	16	0.00112
	***	10	0.00111
	firefighting	10	0.00111
	Illinois	10	0.00111
G6	***	29	0.00313
	***	25	0.00270
	vote	18	0.00195
	electromagnetic waves	52	0.00191
	Astypalaia	17	0.00184
	Uber	30	0.00172
	pilot	20	0.00165
	election	15	0.00162
	boot	28	0.00160
	lift	23	0.00156
G7	plop	24	0.00284
	National Cancer Center	30	0.00222
	battalion commander	18	0.00213
	prostitution	18	0.00213
	lover	18	0.00213
	rape	17	0.00201
	communication expense	15	0.00177
	lever	14	0.00166
	publication	13	0.00154
	sexual relations	17	0.00154
G8	***	20	0.00281

	sweeteners	14	0.00197
	***	27	0.00172
	aspartame	11	0.00154
	bias	11	0.00154
	tenderness	24	0.00153
	***	24	0.00153
	companion dog	13	0.00140
	calorie	9	0.00126
	vaping	67	0.00121
G9	***	32	0.00611
	***	31	0.00592
	Worldcup	29	0.00553
	***	27	0.00515
	***	25	0.00477
	***	39	0.00394
	Qatar	18	0.00343
	competition	18	0.00263
	stadium	13	0.00248
	***	13	0.00248
G10	Xiaomi	25	0.00597
	***	19	0.00593
	ecology	17	0.00531
	mock test	26	0.00509
	KT&G	20	0.00477
	president	15	0.00468
	battery	25	0.00354
	fire identification team	11	0.00343

	lithium	20	0.00330
	***	10	0.00312
G11	Busan	14	0.00448
	Korean *** Institute	10	0.00320
	sentence pattern	9	0.00288
	future career	9	0.00288
	wrongdoer	10	0.00245
	Nodam(No smoking)	12	0.00241
	set exam questions	19	0.00238
	beer	33	0.00230
	male student	21	0.00228
	health promotion program	7	0.00224

3.3. Topics from LDA and STM

Based on the criteria of selecting a model with the highest independence between topics and the highest relevance between keywords of topics, a total of 11 topics for LDA and 15 topics for STM were derived and analyzed (Figure 6, 7).

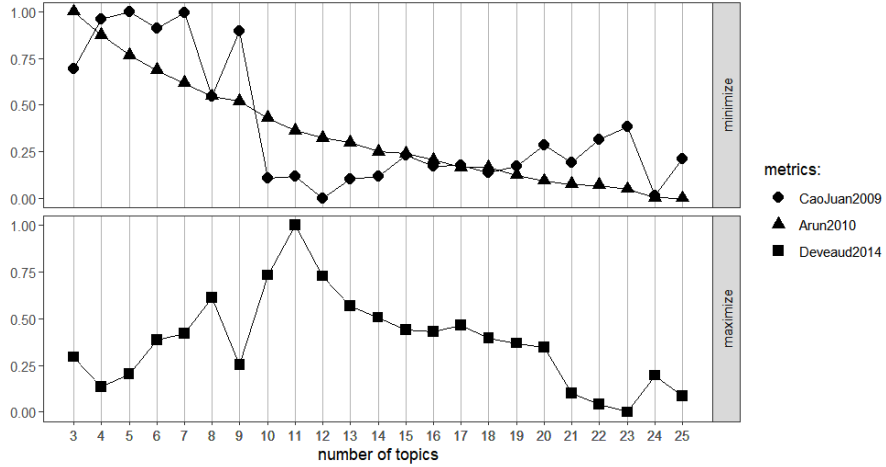


Figure 6. Diagnostic Values by number of topics in LDA ^⑩

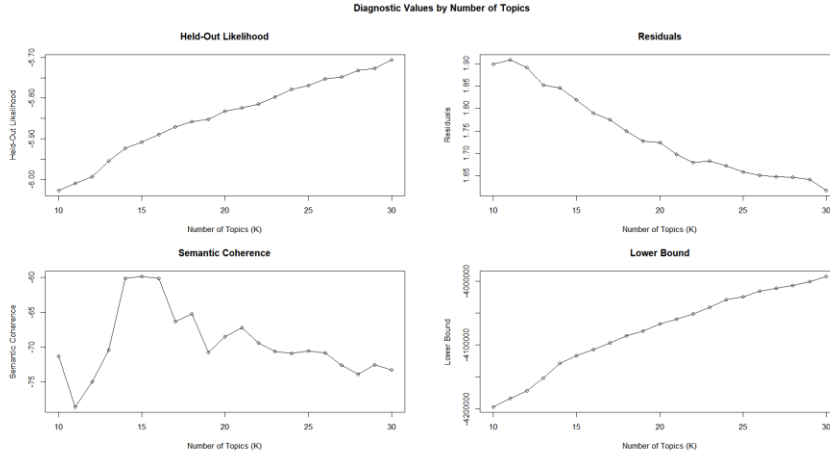


Figure 7. Diagnostic Values by number of topics in STM ^⑪

^⑩ Arun et al., 2010 and Cao et al., 2009 aim to find the number of topics that minimize correlation or overlap of terms configuring each code [71–72]. Deveaud et al., 2014 seek to maximize the differences among codes [73].

A total of 11 topics in LDA are followings: ‘New Product Launching’, ‘Harm of E-cigarette Ingredients and Discontinuation’, ‘Health Risks and Use Behaviors of E-cigarettes and HTPs in Korea’, ‘Industry Growth Initiatives’, ‘Industry Business Performance’, ‘Drugs-related Crime’, ‘E-cigarettes Control Policy in other countries’, ‘Tobacco Control Policy(P, W) in Korea’, ‘Taxation Plan focused on E-cigarettes (R)’, ‘Comprehensive Taxation Plan in South Korea’, and ‘Korean CSAT Examination Protocols and Prohibited Items’. The prevalence of each topic and the top 20 words based on probability are presented in Table 9.

A total of 15 topics in STM are followings: ‘New Product Launching (HTPs)’, ‘Industry Business Performance’, ‘Harm of e-cigarettes and Discontinuation in Korea’, ‘Health Risks of e-cigarettes revealed in U.S.’, ‘Illegal Use of Tobacco Products’, ‘Tobacco Control Policy(P, W) in Korea’, ‘New Product Launching (e-cigarettes and Flavored cigarettes)’, ‘Health Risks and Use Behaviors of E-cigarettes and HTPs in Korea’, ‘Proposal of Revision Bill for Raising Taxes on E-cigarettes’, ‘Stance Expression on HTPs as Cigarette Substitutes’, ‘E-cigarettes in Media’, ‘Taxation Issues related to E-cigarettes’, ‘Activity for Smoke-free Environment Construction’, ‘Safety of Electronic Devices of Tobacco Products’, and ‘Defensive Action for HTPs as Cigarette Substitutes’. The prevalence of each topic and the top 20 words based on probability and FREX are presented in Table 10.

In line with previous studies, all topics in both models were categorized into three main themes: tobacco control policy-oriented, marketing-oriented, and others.

^⑪ This study used Held-Out likelihood, semantic coherence, residual provided by STM[69]. The procedure was: first, changing the number of potential topics from 10 to 30; second, select the number of topics at the point of maximum of semantic coherence, relatively high Held-Out Likelihood, and relatively low Residuals.

- **Marketing-oriented Topics**

In relation to tobacco industry activities, LDA model identified three topics: ‘New Product Launching’, ‘Industry Growth Initiatives’, ‘Industry Business Performance’. The topic ‘New Product Launching’ encompasses words that pertain to specific tobacco brands or product names, along with information related to purchasing or encouraging the purchase of those products, such as news about product launches, device performance, retail outlets, and so on. The topic ‘Industry Business Performance’ includes words that reflect financial indicators of the industry's success, such as ‘sales’, ‘increase’, ‘performance’, ‘export’, ‘growth’, ‘record’, and so on. Lastly, the topic ‘Industry Growth Initiatives’ encompasses a broader range of the industry’s activities and efforts to expand their market presence, including initiatives for market expansion, Environmental, social, and governance (ESG) practices, technological innovation, and more.

Regarding tobacco industry activities, STM also identified topics related to tobacco industry, two of which were similar to the results of the LDA model: ‘New Product Launching’, and ‘Industry Business Performance’. In the LDA model, ‘New Product Launching’ covered all kinds of electronic cigarettes, brands, and products. However, in the STM model, the topic 'New Product Launching' was further divided into two distinct topics based on the type of electronic cigarettes: ‘New Product Launching (HTPs)’ and ‘New Product Launching (e-cigarettes and Flavored cigarettes)’.

Additionally, the STM model revealed three additional topics that were not present in the LDA model. These topics were ‘Stance Expression on HTPs as Cigarette Substitutes’, ‘Activity for Smoke-free Environment Construction’, and ‘Defensive Action for HTPs as Cigarette Substitutes’, each representing an independent theme or aspect related to the tobacco industry.

The topic ‘Stance Expression on HTPs as Cigarette Substitutes’ consists of the words related to official announcements or statements from the tobacco industry. This topic emphasized smokers’ rights to use ‘less harmful’ cigarettes and advocated for the transition from conventional cigarettes to electronic cigarettes. This topic has partially similar words to ‘Health Risks and Use Behaviors of Electronic Cigarettes in Korea’ topic in LDA model, including the words addressing the approval of e–cigarettes as Modified Risk Tobacco Product (MRTP) by the U.S. Food and Drug Administration (FDA) [74].¹² It reflected the industry’s direct response to or dissatisfaction with government regulations on e–cigarettes, drawing on the MRTP approval by the U.S. FDA and the campaign of the UK Public Health England (PHE).

The topic ‘Activity for Smoke–free Environment Construction’ included words associated with efforts to create a smoke–free environment by establishing designated areas for vaping only. On the other hand, the topic ‘Defensive Action for HTPs as Cigarette Substitutes’ comprised words related to the tobacco industry lawsuits against the Korean government¹³ [75] and the organization

¹² On 7 July 2020, the FDA authorized the marketing of the IQOS tobacco heating system as a MRTP with reduced exposure information. The FDA denied PMI a ‘risk modification’ order, for which PMI originally applied. A ‘risk modification’ order requires the applicant to demonstrate that products “(1) Significantly reduce harm and the risk of tobacco–related disease to individual tobacco users; and (2) Benefit the health of the population as a whole taking into account both users of tobacco products and persons who do not currently use tobacco products” . PMI “did not demonstrate” that IQOS met these standards. The exposure modification standard that the FDA authorized “establishes a lower standard” than that of risk modification.

¹³ The Ministry of Food and Drug Safety found five cancer–causing substances in Heated Tobacco Products sold in the local market in 2018, with the level of tar detected in some of them exceeding that of conventional cigarettes. Philip Morris Korea Inc. filed a lawsuit against the ministry, demanding the disclosure of information on a test concluded electronic cigarettes contain harmful substances. In 2020, the court decided partially in favor of Philip Morris Korea Inc., saying “Except for

of seminars or forums advocating for the introduction of policies based on ‘Harm Reduction’ perspective.

● Policy-oriented Topics

Regarding tobacco control policy, the LDA model identified five relevant topics: ‘Harm of e-cigarettes and Discontinuation in Korea’, ‘Health Risks and Use Behaviors of Electronic Cigarettes in Korea’, ‘E-cigarettes Control Policy in other countries’, ‘Tobacco Control Policy (P, W)’, and ‘Taxation Plan focused on E-cigarettes (R)’.^⑭

The topic ‘Harm of e-cigarettes and Discontinuation in Korea’ included words representing harmful liquid ingredients in e-cigarettes, which led to cases of lung damage and death in 2019. It also addressed recommended measures to discontinue the use of e-cigarettes in Korea. The topic ‘Health Risks and Use Behaviors of Electronic Cigarettes in Korea’ encompassed words related to investigations into the use behaviors of electronic cigarettes and the associated health risks. This included the health risks of dual use, involving both conventional cigarettes and e-cigarettes, as well as potential risks when transitioning from conventional cigarettes to e-cigarettes. The topic ‘E-cigarettes Control Policy in other countries’ consisted of the words related to tobacco control policies in foreign countries, with a particular focus on flavored e-cigarettes, primarily policies implemented in the United States. The topic ‘Tobacco Control Policy’ encompassed words related to non-price regulatory policies, specifically those addressing ‘Protecting

information that is unlikely to be kept by the Ministry of Food and Drug Safety, some private information should be disclosed.”

^⑭ LDA model classified the price regulation policy into two categories: 2019 when the e-cigarette tax increase amendment was proposed over time, and 2020 when the tax increase on e-cigarettes was delayed. Related topic in STM model covered all of them into one topic, focusing on the revision of the tax law in 2020.

people from tobacco smoke(P)’ and ‘Warning about the dangers of tobacco(W)’ as defined in the MPOWER measures. This included measures such as designated non-smoking area and the inclusion of warning signs and graphic images on tobacco packaging(Figure 8) [76]. The topic ‘Taxation Plan focused on E-cigarettes (R)’ consisted of words associated with revisions in tax laws specifically targeting e-cigarette liquids.



Figure 8. MPOWER Measures for Tobacco Control

STM also identified topics related to tobacco control policy, and all five of them align with the results obtained from the LDA model. Those topics are as follows: ‘Harm of e-cigarettes and Discontinuation in Korea’, ‘Health Risks of e-cigarettes revealed in U.S.’, ‘Tobacco Control Policy (P, W) in Korea’, ‘Health Risks and Use Behaviors of Electronic Cigarettes in Korea’ and ‘Tobacco Control Policy (P, W)’.

- **Topics in the articles on other issues**

Both LDA and STM models extracted topics related to other issues as well. The topics derived from LDA include ‘Drugs–related Crime’, ‘Comprehensive Taxation Plan in South Korea’, and ‘Korean CSAT Examination Protocols and Prohibited Items’. The Topic ‘Drugs–related Crime’ topic consisted of words that describe social issues involving e–cigarette devices used in drug–related crimes or sex crimes. The topic ‘Comprehensive Taxation Plan in South Korea’ included words related to the revision of the tax laws in 2020.¹⁵ The topic ‘Korean CSAT Examination Protocols and Prohibited Items’ comprises words associated with the CSAT, highlighting e–cigarettes as restricted items not permitted during the examination.

Topics on other issues extracted from STM were similar to those identified by LDA: ‘Illegal Use of Tobacco Products’, ‘Electronic Cigarettes in Media’, ‘Taxation Issues related to E–cigarettes’, and ‘Safety of Electronic Devices of Tobacco Products’. The topic ‘Illegal Use Behaviors of Tobacco Products’ included both drugs–related crimes and other social issues related to the illicit use of e–cigarettes, such as the problem of adolescents’ proxy purchasing e–cigarettes and prohibition during the CSAT. Furthermore, the topic ‘Electronic Cigarettes in Media’ included words associated with celebrities using e–cigarettes indoors and the portrayal of tobacco products in media, such as dramas or entertainment programs. However, there was a distinct topic in

¹⁵ On 2nd December 2020, the National Assembly approved the government’s bill to amend sixteen tax laws including the Individual Income Tax Law (IITL), the Corporate Income Tax Law (CITL), the VAT Law (VATL) and other laws. In the bill, there were two following proposals on e–cigarettes: 1) Expanding the scope of taxable cigarettes to the cases of roots, stems, and fruits of the tobacco are raw materials, and 2) raising the individual consumption tax rate for liquid of e–cigarettes from 370 won to 740 won per ml of nicotine solution. First proposal was adopted but second proposal was not adopted. It was decided to maintain the current tax rate (370 won/ml) on liquid.

STM model, not present in the LDA model: ‘Safety of Electronic Devices of Tobacco Products’. This topic consisted of words related to the unsafety aspects of batteries in electronic devices, including e-cigarettes, which have been associated with incidents of fires and explosions.

In summary, the findings from both LDA and STM models exhibited a considerable degree of similarity. However, the results obtained from STM, which incorporated time and newspaper types as covariates, provided additional granularity and specificity in depicting the activities of the tobacco industry. Furthermore, the prominence of e-cigarettes as a social issue portrayed in the media was evident, with sufficient frequency to warrant its identification as an independent topic.

Table 9. Distribution Topics and Words from LDA

Topics	N(%)	Words by Probability
New Product Launching	693(13.43)	tobacco, electronic, products, launch, device, consumer, IQOS, sale, use, cigarette, convenience store, exclusive, Philip Morris Korea, Pro, offer, nationwide, customer, application, possibility, Brand
Harm of E–cigarette Ingredients and Discontinuation	626(12.13)	tobacco, electronic, liquid, use, sales, product, discontinuation, U.S., government, ingredients, electronic cigarettes, recommendation, harm, flavor, substance, vitamin, result, action, severe, detection
Health Risks and Use Behaviors of Electronic Cigarettes in Korea	620(12.01)	tobacco, smoking cessation, electronic, smoker, smoking, electronic cigarettes, cigarette, use, general, health, nicotine, result, picture, warning, person, conventional cigarette, professor, risk, policy, occurrence
Tobacco Industry Growth Initiatives: Market Expansion, ESG, technological innovation	558(10.81)	tobacco, products, KT&G, electronic, market, cigarette, PMI (Philip Morris International), launch, sale, consumer, Philip Morris, IQOS, 2.0, BAT, global, Korea, Hybrid, brand, Japan, world
Industry Business Performance	482(9.34)	KT&G, tobacco, market, sales, increase, performance, electronic, contrast, previous year, export, growth, forecast, record, cigarette, business, investment, overseas, expansion, researcher, turnover

Drugs-related Crime	470(9.1)	drugs, accusation, cannabis, e-cigarette, police, assault, crime detection, the prosecution, Mr. A, tobacco, investigation, chairman, facts, case, sentence, Russia, person, claim, victim, penal servitude
E-cigarettes Control Policy in other countries	463(8.97)	U.S., China, tobacco, sales, business, electronic cigarettes, dollar, electronic, market, FDA, company, enterprise, regulation, Korea, president, Juul Labs, Trump, coverage, duty free shop, smoking
Tobacco Control Policy(P, W) in Korea	413(8)	tobacco, smoking, adolescent, electronic cigarettes, electronic, advertisement, smoking cessation, sales, use, crackdown, behavior, video, Lim Young-woong, nicotine, fine (penalty), non-smoking area, liquid, zone, target, health
Taxation Plan focused on E-cigarettes (R)	389(7.54)	tobacco, electronic, liquid, cigarette, tax, sales volume, nicotine, increase, government, electronic cigarettes, general, sales, taxation, imposition, tax rate, Nicotiana tabacum L, Ministry of Strategy and Finance, reduction, standard, level
Comprehensive Taxation Plan in Korea	295(5.71)	government, target, health, support, income, increase, standard, COVID-19, nation(population), taxation, reduction, expansion, planning, strengthening, administration, investigation, level, policy, propel, enterprise
Korean CSAT Examination Protocols and Prohibited Items	153(2.96)	CSAT, test, examinees, test site, test area, taking an exam, subject, carrying-in, goods, selection, mask, taking a test, check, cheating, portable, exam supervisor, Siid, fraudulent act, grade, watch

Table 10. Distribution Topics and Words from STM

Topics	N(%)	Words by Probability	Words by FREX
New Product Launching (HTPs)	560(10.9)	launch, tobacco, electronic, KT&G, product, cigarette, 2.0, sale, Hybrid, IQOS, consumer, device, Pro, market, nationwide, expansion, Stick, ILUMA, use, application	Able, Boost, Edition, Fiit, HYBRID, MIIX, 2.0, ILUMA, Miix, Jouz Korea, Induction, Solid, SMARTCORE, Mini, Prime, HTPs-only Stick, Aiim, Pro, 1.0, Blade
Industry Business Performance	545(10.6)	KT&G, tobacco, market, electronic, sale, increase, performance, overseas, export, contrast, growth, previous year, cigarette, outlook, Global, record, PMI, expansion, business, occupy	operation profit, EM-Tech, Elentec, Baek Bok-in, FCP, target stock prices, client company, Ginseng Corporation, Korea Tobacco and Ginseng Corporation (KT&G), ITM Semiconductor CO., Ltd, performance, dividend stock, stock price, good performance, KGC (Korea Ginseng Corp), the Middle East, good condition, overseas corporation, Consensus, buying
Harm of E-cigarette and Discontinuation in Korea	514(10.0)	tobacco, liquid, electronic, use, discontinuation, product, ingredient, government, sale, recommendation, U.S., harmful (maleficence), vitamin,	acetate, lung damage, vitamin, severe, Tropical, THC, recommendation, ingredient, doubt, glycerin, diacetyl, detection, case of death, self-control, discontinuation,

		substance, nicotine, severe, result, detection, doubt, electronic cigarettes	acetoin, propylene glycol, ppm, Vitamin E Acetate, suspected case
Health Risks of e-cigarettes revealed in U.S.	423(8.2)	tobacco, electronic, electronic cigarettes, U.S., sale, smoking, adolescent, FDA, flavor, regulation, product, Trump, local time, market, the authorities, enterprise, report, nicotine, action, company	San Francisco, marijuana, New York State, Burns, Altria, Trump, Lawrence, Donald, menthol, Walmart, WSJ, federation, flavored tobacco, local time, hospital room, staff reduction, New Zealand, flight attendant, Carbon, Mint flavor
Illegal Use of Tobacco Products	410(8.0)	CSAT, drugs, test, suspicion, test takers, crime detection, cannabis, test site, area, carrying in, taking an exam, assault, subject, the prosecution, behavior, choice, police, electronic cigarettes, Mr. A, confirm	CSAT, examinee, takingn an exam, subject, examination, cheating, exam supervisor, school year, prosecution, test identification slip, exam paper, Korean history, dosing, Kim Chang-hwan, nullity, the accused, Lee Seok-cheol, entrance, methamphetamine (meth), Lee Seung-hyun
Tobacco Control Policy (P, W) in Korea	390(7.6)	tobacco, smoking, smoking cessation, adolescent, picture, warning, advertisement, health, electronic, cigarette package, phrase, crackdown, smoker, electronic cigarettes, Ministry of	cigarette package, warning picture, warning phrase, antismoking advertisements, phrase, tobacco advertisements, picture, non-smoking area, Korea Health Promotion Institute, smoking

		Health and Welfare, behavior, non-smoking area, area, policy, warning picture	prevention, promotion, zone, warning, advertisements, penalty (fine), webtoon, Healthy life expectancy, crackdown, antismoking campaigns, surveillant
New Product Launching (e-cigarettes and Flavored cigarettes)	371 (7.2)	tobacco, electronic, launch, liquid, product, market, sale, smell, Korea, KT&G, consumer, cigarette, convenience store, electronic cigarette, Cartridge, Vapor, Ploom TECH, nicotine, U.S., device	SiiD, Vitabon, Ploom TECH, SENSE, MEVIUS, Vapor, LBS (Less Breath Smell), SiiD, JTI Korea, USB, Slide, JTI, Raison, HAKA signature, landing, Himalaya, throat hit, smell, Juul Labs Korea, hyvä
Health Risks and Use Behaviors of Electronic Cigarettes in Korea	333 (6.5)	tobacco, smoking cessation, smoking, use, electronic cigarettes, nicotine, smoker, electronic, human, cigarettes, investigation, health, result, female, increase, professor, indoor, Lim Young-woong, male, general	Lim Young-woong, cardio-cerebro, depression, physical activity, metabolic syndrome, withdrawal symptom, cotinine, vessel, dressing room, research team, receptor, NEW ERA PROJECT, dust, Mapo-gu Office, NEW ERA, admonition, cardio-cerebrovascular disease, Mulberry School (TV program), male student, depressive disorder
Proposal of Revision Bill for Raising Taxes on E-cigarettes	329 (6.4)	tobacco, electronic, cigarettes, liquid, sales volume, tax, increase, government, electronic	Charges for National Health Promotion, tax on cigarettes, taxed and public charges, charge, Public

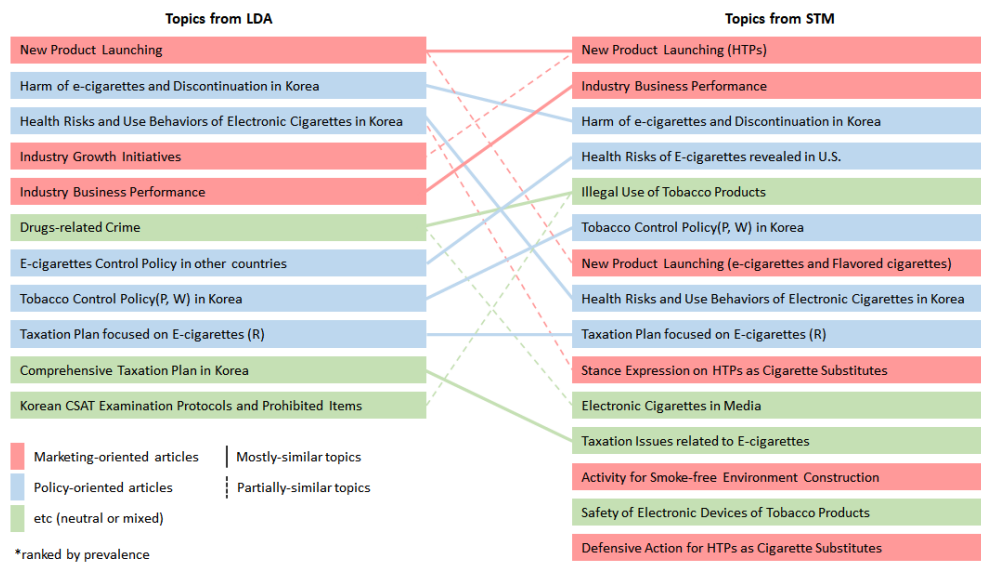
		cigarettes, general, imposition, taxation, tax rate, Ministry of Strategy and Finance, decrease, nicotine, sale, tobacco, criteria, level	Charge, local education tax, tobacco consumption tax, Individual Consumption Tax, 0.7, specific duties, 1 pack, increase, price, external cost, Ministry of Economy and Finance, labor service, tax, equity, sales volume, tax rate
Stance Expression on HTPs as a cigarettes substitutes	271 (5.3)	tobacco, product, electronic, IQOS, general, smoker, cigarettes, electronic cigarettes, harm, Philip Morris, PMI, switch, result, smoking, conventional cigarettes, scientific, UK, substance, regulation, Philip Morris Korea	MRTTP, authorization, substitute, switch, (Jacek) Olczak, science, André (Calantzopoulos), Philip Morris, non- combustible product, UK, PHE, Glo, scientific, toxic substance, Jacek (Olczak), public, Baek Young-jae, potential, Risk, 95%
Electronic Cigarettes in Media	231 (4.5)	video, human, appearance, controversy, Russia, thought, photo, child, entertainment company, friend, electronic cigarettes, situation, broadcast, U.S., apology, filming, self, actor, tobacco, Kim Dong-hee	Kim Dong-hee, Jung Joon-won, North Korea, school violence, Da-jeong, Yoon Yeo-jung, Ahn Hyun-mo, Woo-young, Griner (1), Rhymer, Dae-young, Griner (2), Oscar, disimprisonment, Swag, Oh Jun-seok, Boot, detention, Kim Jong-un, viewer
Taxation Issues related to E-cigarettes	230 (4.5)	enterprise, China, government, target, income, taxation, tax,	delinquent taxpayer, tax delinquency, income tax, amount in arrears, large

		investment, Dollar, release, stock listing, local taxes, support, criteria, deduction, list, delinquent taxpayer, application, tax increase, next year	amount of money, YJM Games, comprehensive real estate tax, local administration restriction, vibration motors, single householder, transfer income tax, RLX, transaction tax, WeWork, Ministry of Agriculture Food and Rural Affairs, DiDiChuXing, security transaction tax, Ministry of Oceans and Fisheries, other incomes, list
Activity for Smoke-free Environment Construction	220(4.3)	enterprise, production, product, tobacco, brand, environment, representative, CEO, business, company, BAT, factory, market, Global, innovation, Philip Morris Korea, BAT Korea, consumer, Korea, society	Yangsang factory, factory, ESG, Vaping Room, Sacheon, management, Yangsang, talented person, production, Gyeong-nam (Gyeongsangnam-do Province), Republic of Korea, workplace, Vaping, Coffee Bean, CEO, Project, environment, ownership structure, Sacheon factory, executives and staff members
Safety of Electronic Devices of Tobacco Products	201(3.9)	product, duty free shop, battery, Service, electronic cigarettes, investigation, customer, purchase, consumer, electromagnetic waves, sale,	electromagnetic waves, arrival hall, Brooks, used (article), KIA, club, \$600, electric scooter, pitcher, KBO, Pulmuone, leather, luxury, ignition, supplementary battery (portable

		Brand, safety, transaction, possibility, Online, use, confirmation, arrival hall, fire	charger), Chicken, Worldcup, eye drops, end cell, baseball
Defensive Action for HTPs as a cigarettes substitutes	134 (2.6)	policy, U.S., professor, argument, regulation, government, member (of assembly), COVID-19, health(건강), nation(country), Ministry of Food and Drug Safety, China, comment, information, health(보건), release, nation(people), lawsuit, reduction, Korea	NQ(Not Quantitative), reduction, Wuhan, Virus, origin, winning of a lawsuit, government management, concept, laboratory, INVOSSA -K, member(of assembly), lawsuit, Korea Party(한국당), prodding, research institute, Ministry of Food and Drug Safety, commissioner, association, therapy product, Seminar

The topics that exhibited strong presence in both LDA and STM models were as follows: ‘New Product Launching’, ‘Harm of e–cigarettes and Discontinuation in Korea’, ‘Health Risks and Use Behaviors of Electronic Cigarettes in Korea’, ‘Industry Business Performance’, ‘Illegal Use of Tobacco Products’, ‘Health Risks of e–cigarettes revealed in U.S.’, and ‘Tobacco Control Policy (P, W) in Korea’ mostly related to P and W in MPOWER measures, ‘Taxation Plan focused on E–cigarettes (R)’, and ‘Comprehensive Taxation Plan in Korea’. Among those topics, ‘New Product Launching’ emerged as the most prevalent in the analyzed articles, as indicated by the findings from both models (Figure 9).

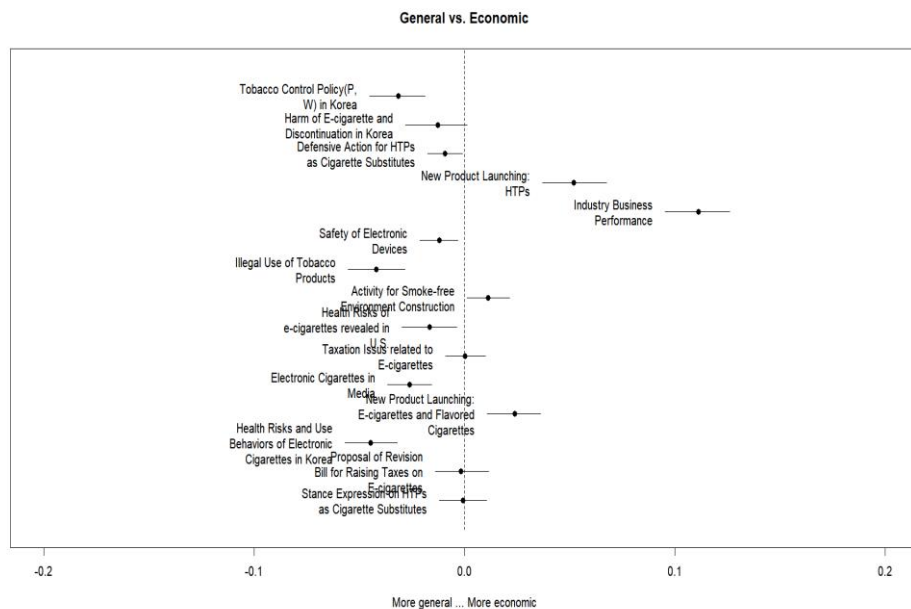
Figure 9. Comparison Chart of Topics by Topic Model¹⁶



¹⁶ ‘Mostly–similar topics’ are defined as the topics sharing same articles each other over 50% of the total of their articles based on a distribution of topics over the articles. And ‘Partially–similar topics’ are defined as the topics sharing same articles either of them over 20% of the total of their articles based on a distribution of topics over the articles.

Based on the proportions of topics by newspaper type, both LDA and STM models revealed a higher prevalence of topics associated with tobacco industry activities in economic daily newspaper compared to general daily newspapers. These topics include ‘Industry Business Performance’, ‘New Product Launching (HTPs)’, and ‘New Product Launching (E–cigarettes and Flavored Cigarettes)’ (Figure 10). The dominance of these topics in economic daily newspapers suggests that the tobacco industry is actively engaged in developing its activities and establishing a strong presence within economic domains.

Figure 10. Comparison of Topic Proportion from STM by the Type of news articles



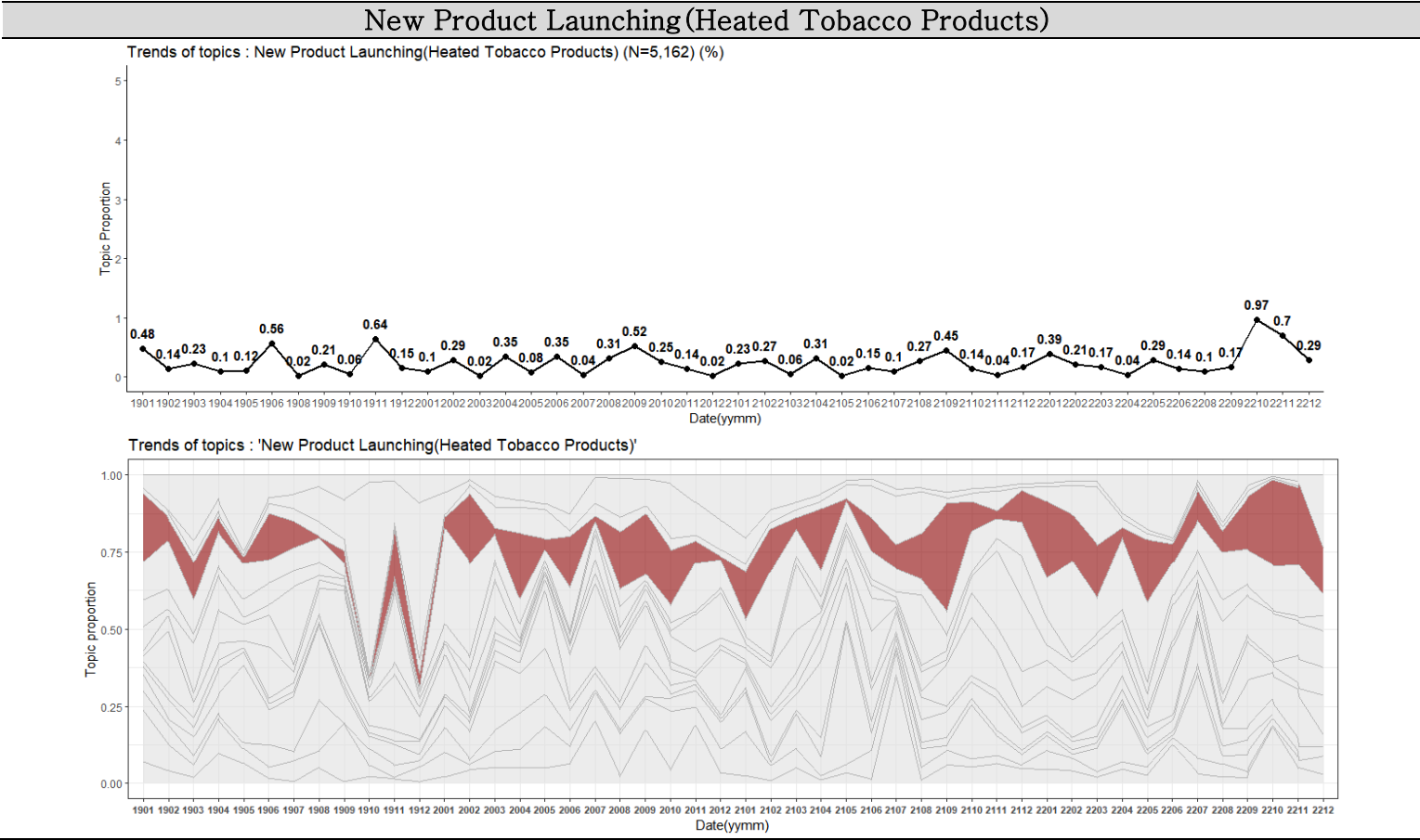
3.4. Trends of Topic Proportion

As depicted in Figure 11, the main findings of the trends of topics from STM analysis are as follows:

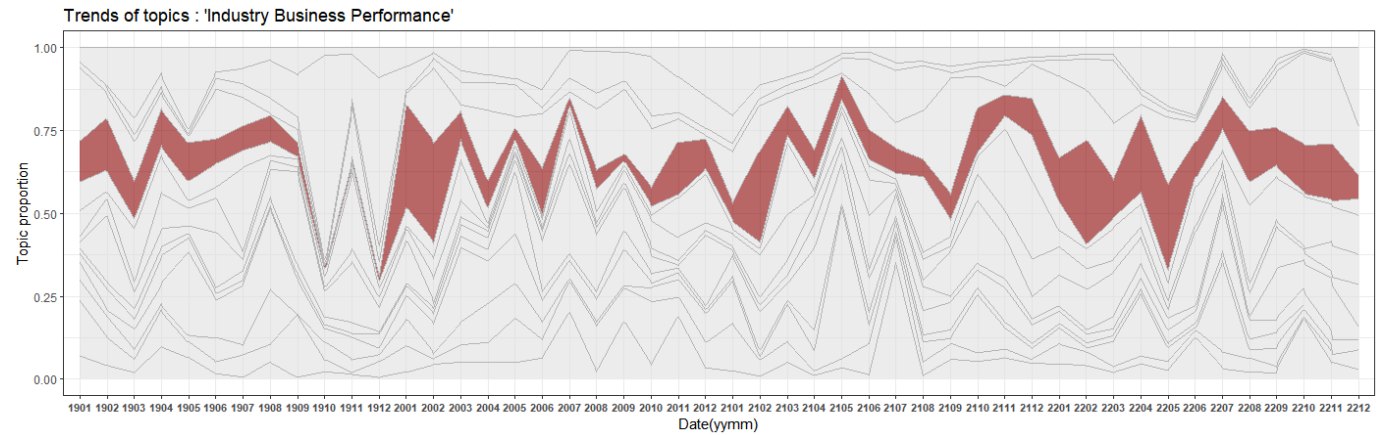
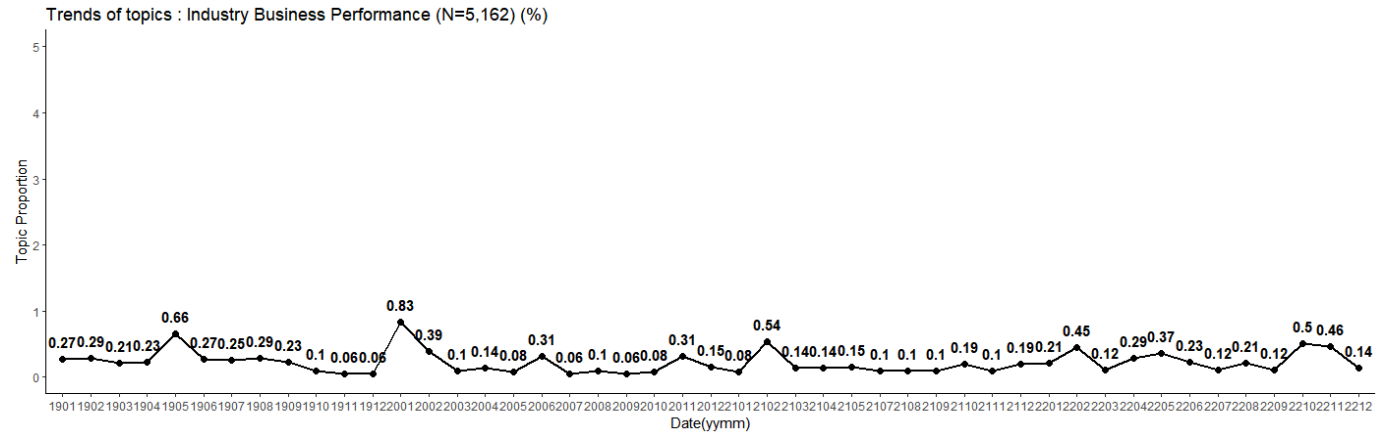
First, the articles in some topics were reported explosively at the time of a social issue. For example, the articles with a high probability of belonging to ‘Harm of E-cigarette and Discontinuation in Korea’ and ‘Health Risks of E-cigarettes revealed in U.S.’ were reported in the second half of 2019. And the articles related to taxation issue were also reported at the time of the proposal of taxation bill on E-cigarettes and the results of enactment of it.

Secondly, the number of reported articles with a high probability of belonging to topic ‘New Products Launching’ has changed over time by the type of tobacco products. The articles on ‘New Product Launching (E-cigarettes and Flavored Cigarettes)’ have steadily declined since the second half of 2019 when ‘JUUL’ issues occurred, and since the second half of 2020, they have accounted for quite a small portion of coverage. On the other hand, the articles on ‘New Product Launching (Heated Tobacco Products)’ showed a certain level of continuous reporting even after the rapid decline in the second half of 2019 when the ‘JUUL’ issues occurred. This pattern was similar to the tendency of the articles, which are highly likely to belong to ‘Industry Business Performance’.

Figure 11. Trends of Topics over time (2019–2022)

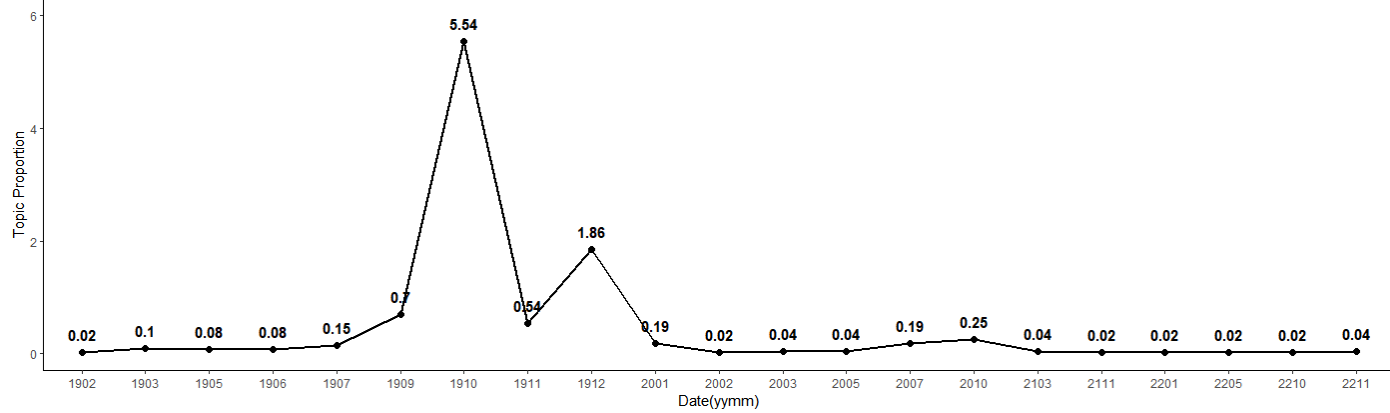


Industry Business Performance

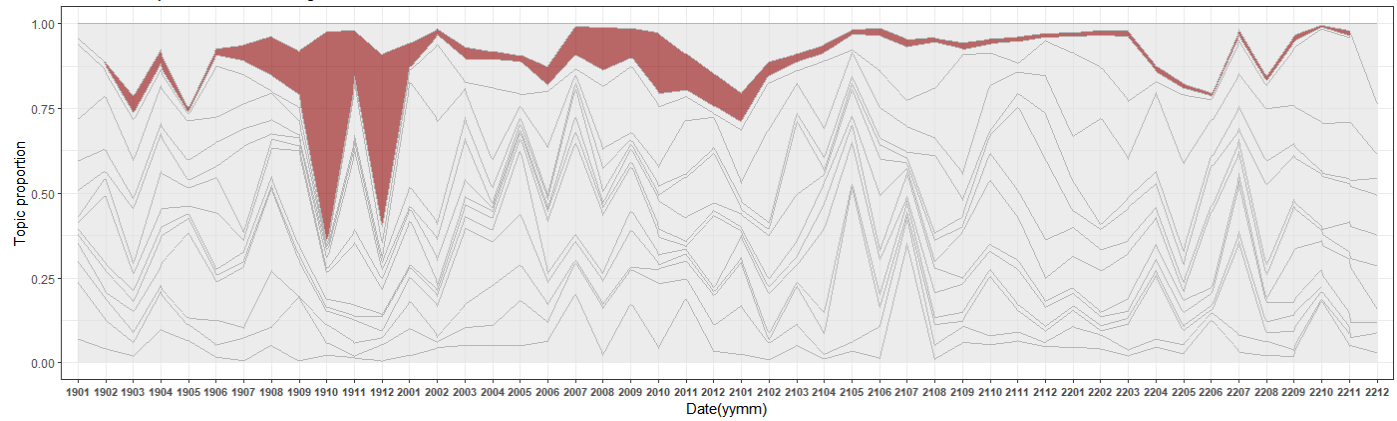


Harm of E-cigarette and Discontinuation in Korea

Trends of topics : Harm of E-cigarette and Discontinuation in Korea (N=5,162) (%)

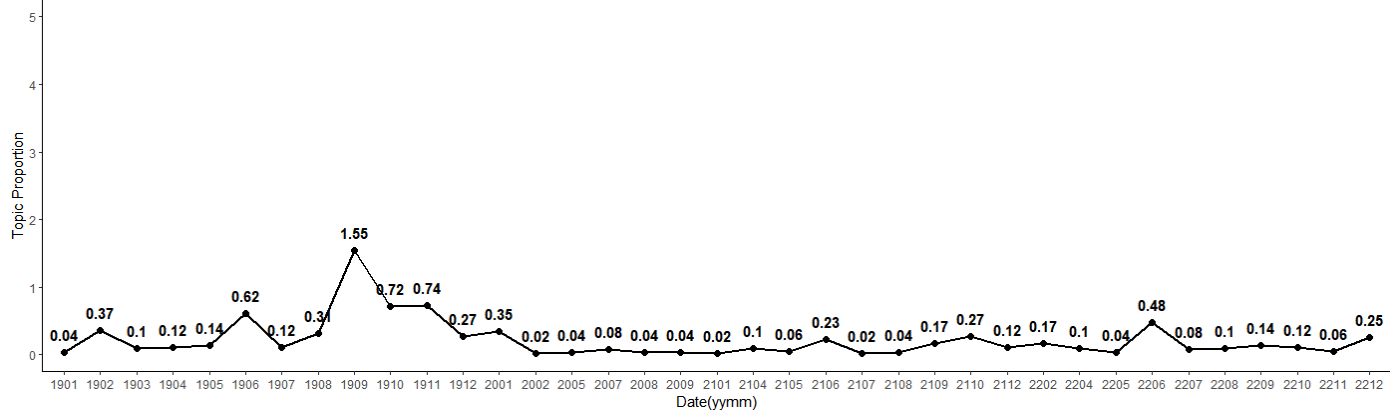


Trends of topics : 'Harm of E-cigarette and Discontinuation in Korea'

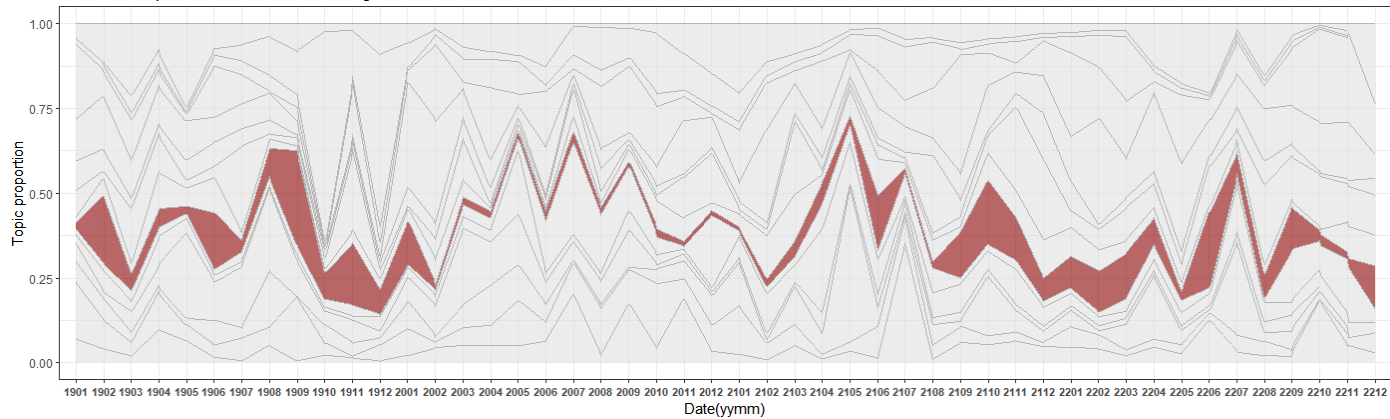


Health Risks of E-cigarettes revealed in U.S.

Trends of topics : Health Risks of E-cigarettes revealed in U.S. (N=5,162) (%)

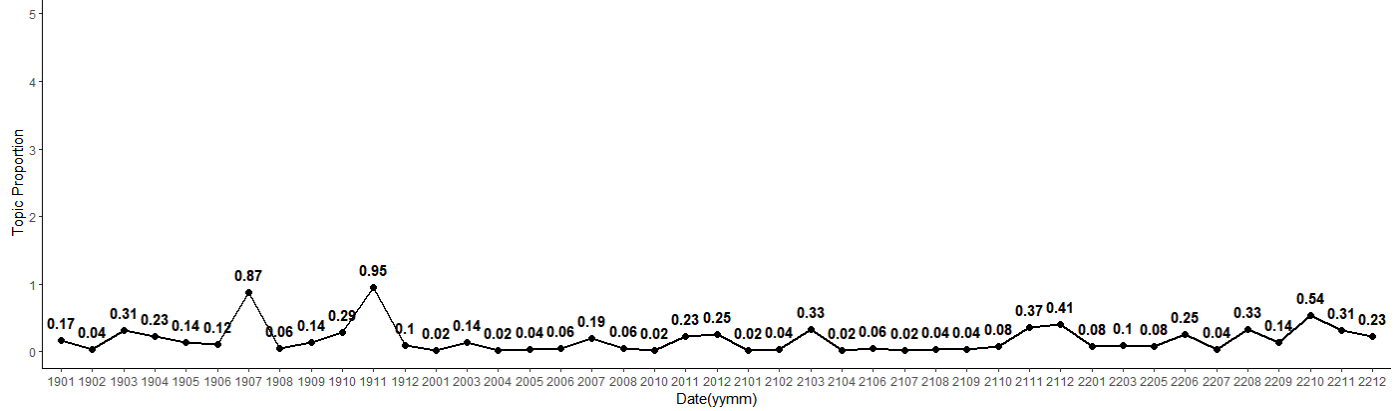


Trends of topics : 'Health Risks of E-cigarettes revealed in U.S.'

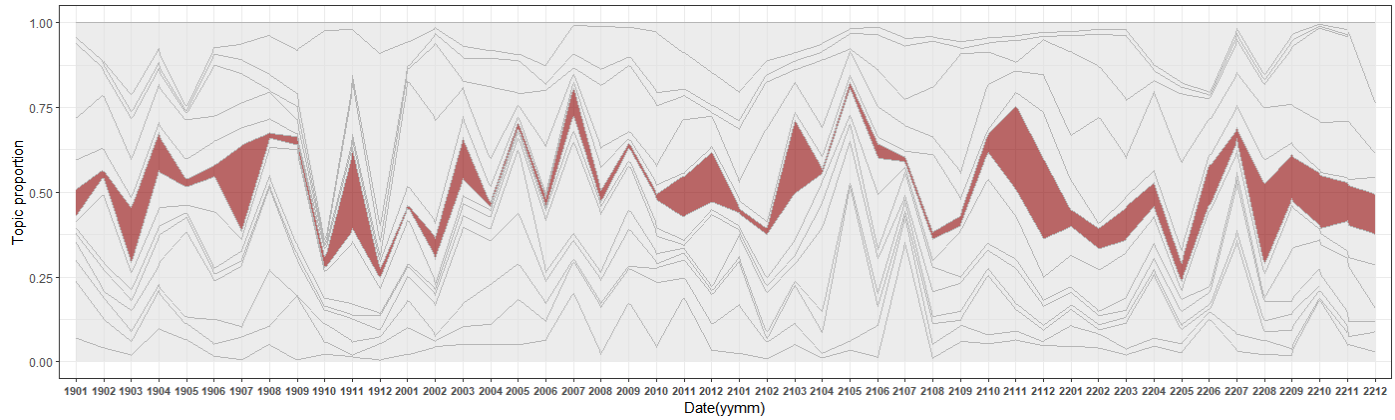


Illegal Use of Tobacco Products

Trends of topics : Illegal Use of Tobacco Products (N=5,162) (%)

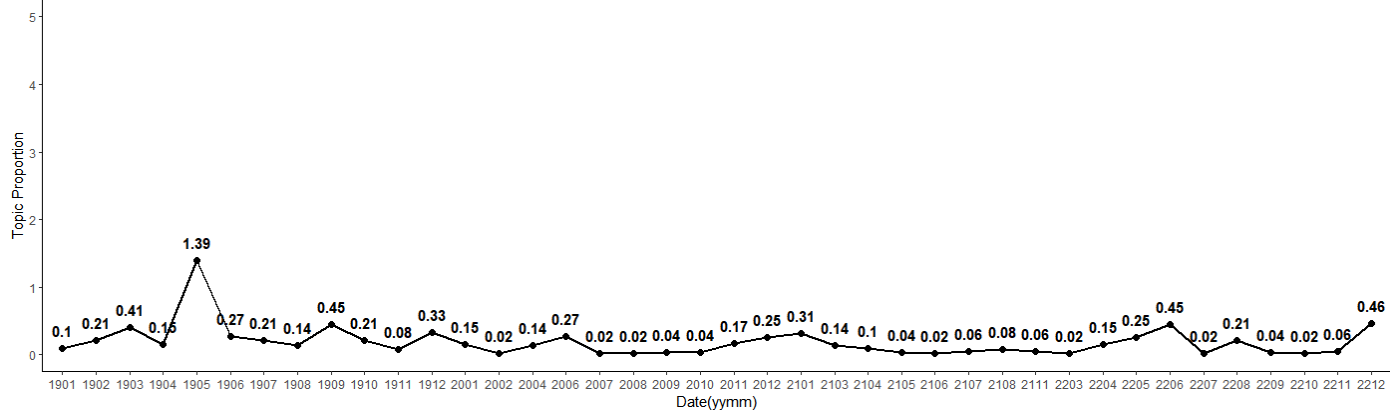


Trends of topics : 'Illegal Use of Tobacco Products'

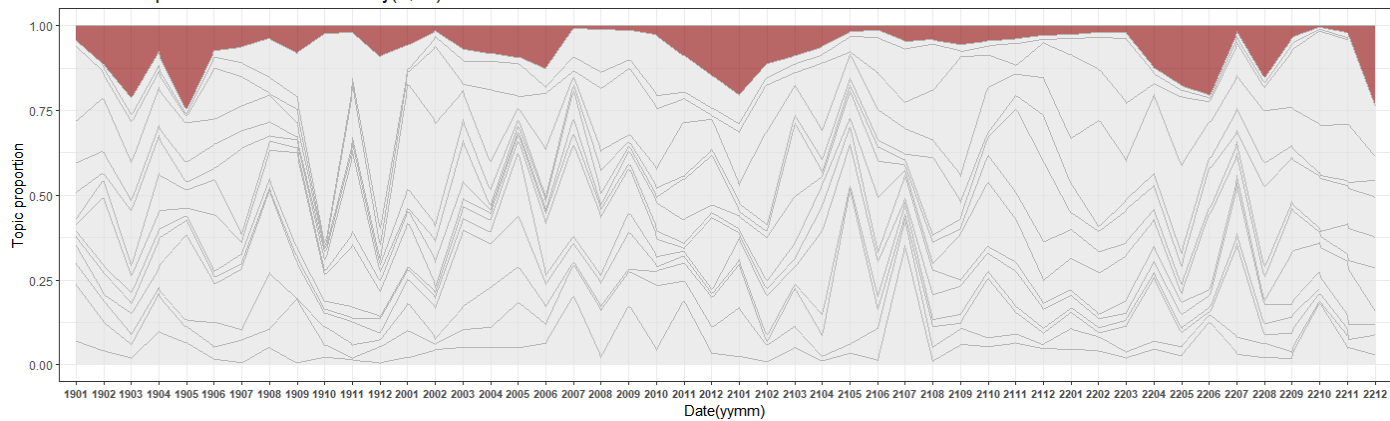


Tobacco Control Policy (P, W) in Korea

Trends of topics : Tobacco Control Policy(P, W) in Korea (N=5,162) (%)

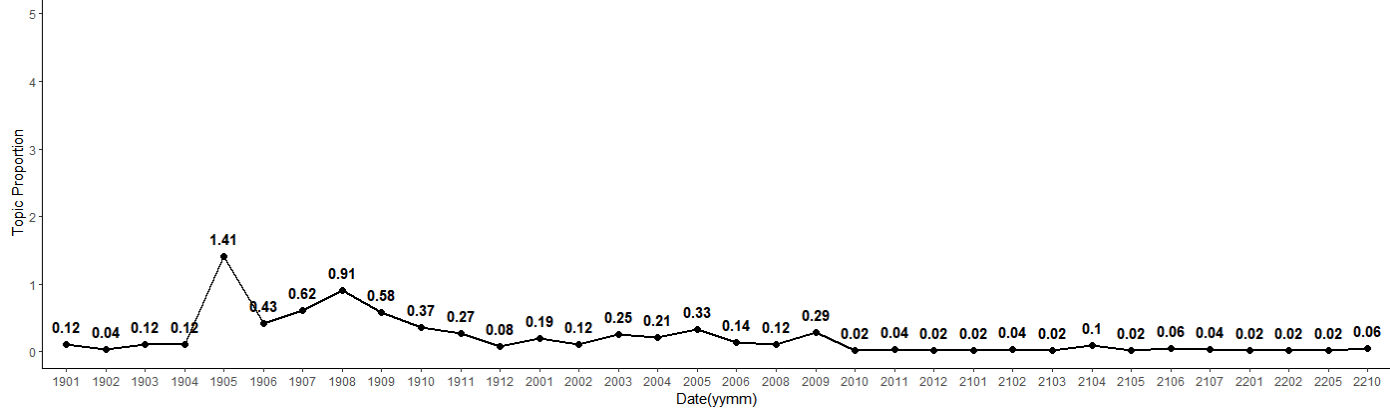


Trends of topics : 'Tobacco Control Policy(P, W) in Korea'

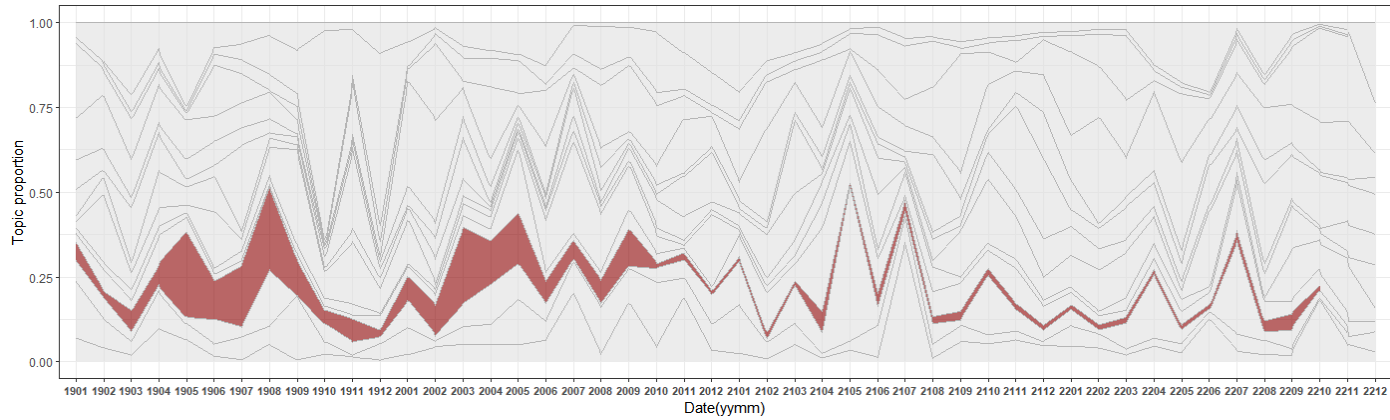


New Product Launching(E-cigarettes and Flavored Cigarettes)

Trends of topics : New Product Launching(E-cigarettes and Flavored Cigarettes) (N=5,162) (%)

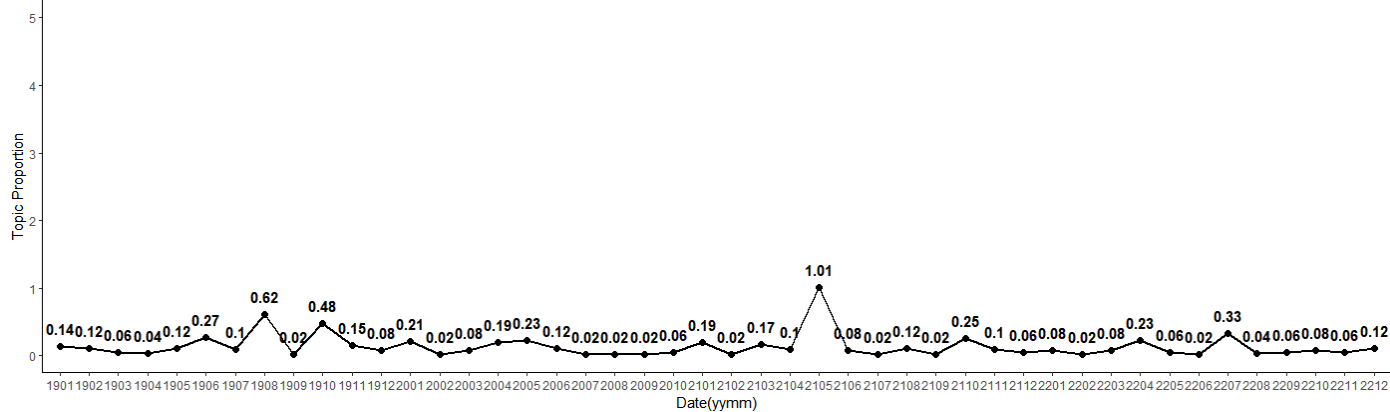


Trends of topics : 'New Product Launching(E-cigarettes and Flavored Cigarettes)'

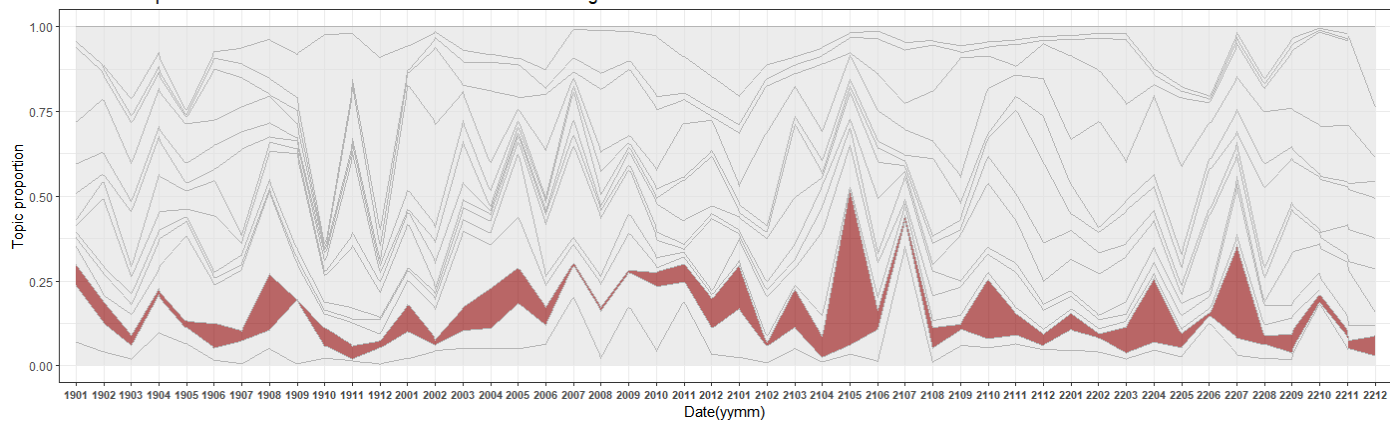


Health Risks and Use Behaviors of Electronic Cigarettes in Korea

Trends of topics : Health Risks and Use Behaviors of Electronic Cigarettes in Korea (N=5,162) (%)

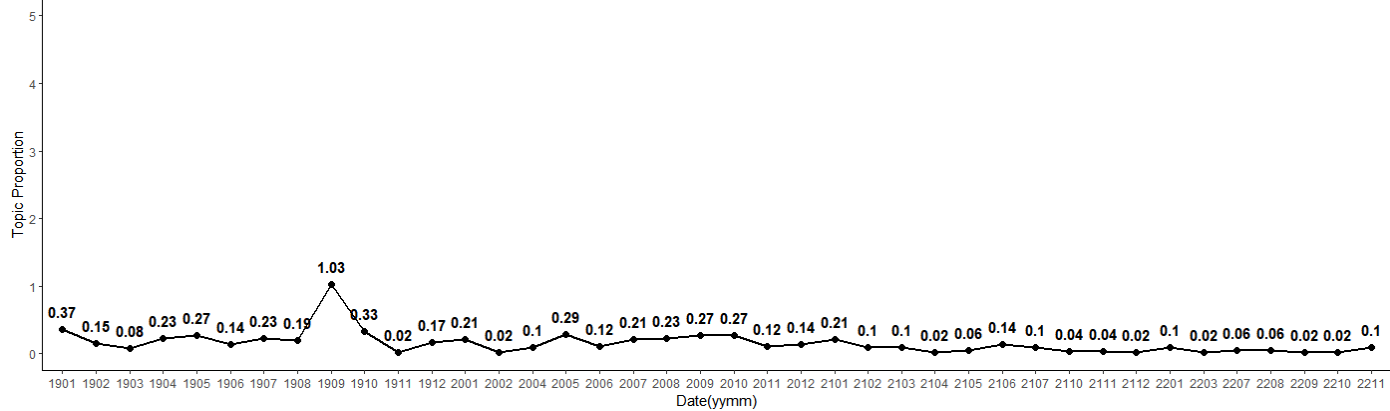


Trends of topics : 'Health Risks and Use Behaviors of Electronic Cigarettes in Korea'

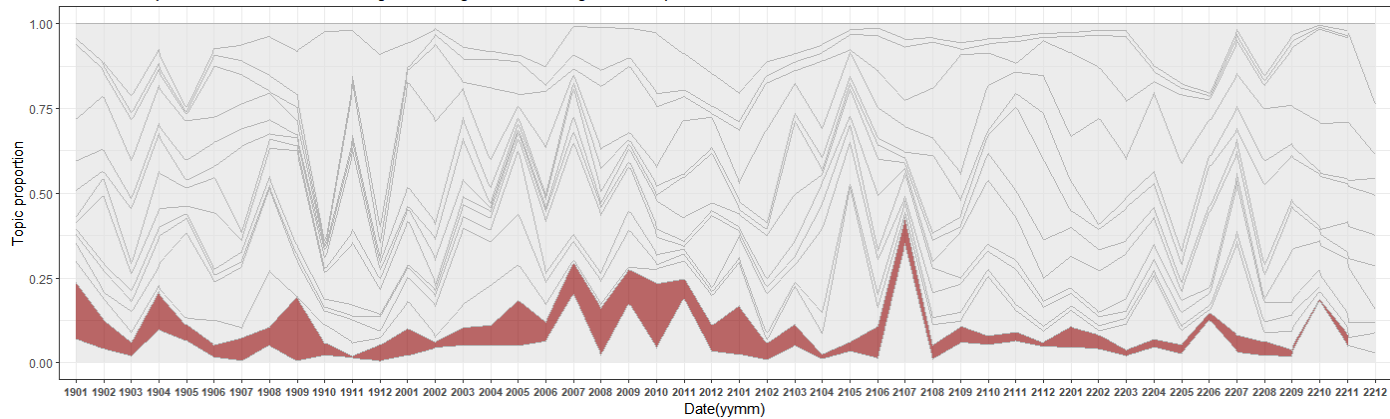


Taxation Plan focusing on raising taxes on E-cigarettes' liquid

Trends of topics : Taxation Plan focusing on raising taxes on E-cigarettes' liquid (N=5,162) (%)

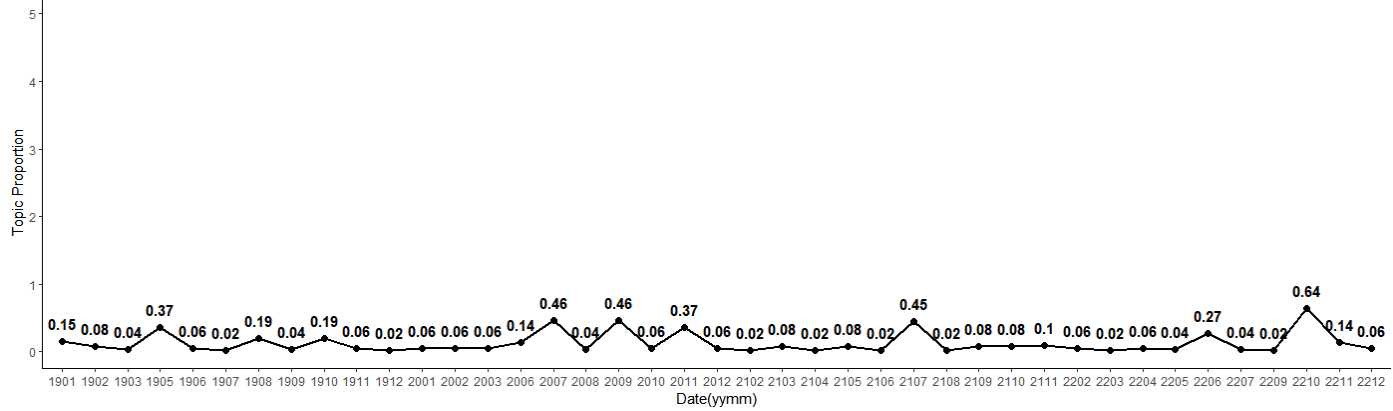


Trends of topics : 'Taxation Plan focusing on raising taxes on E-cigarettes' liquid'

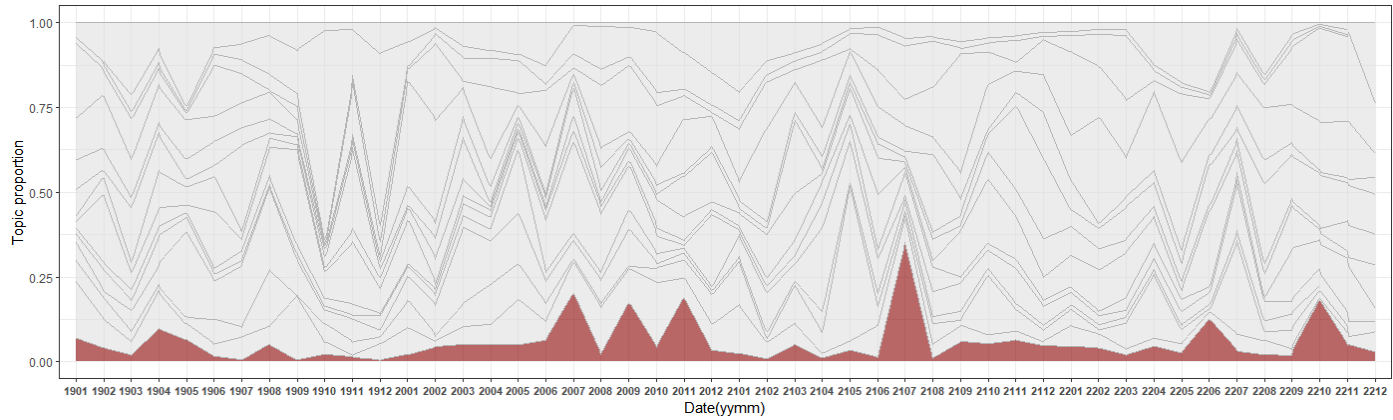


Stance Expression on HTPs as Cigarette Substitutes

Trends of topics : Stance Expression on HTPs as Cigarette Substitutes (N=5,162) (%)

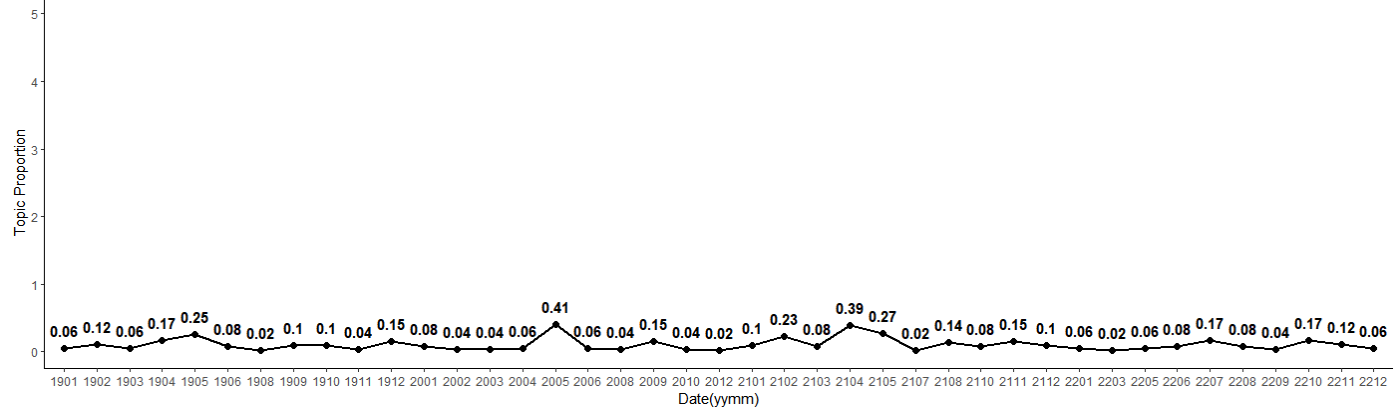


Trends of topics : 'Stance Expression on HTPs as Cigarette Substitutes'

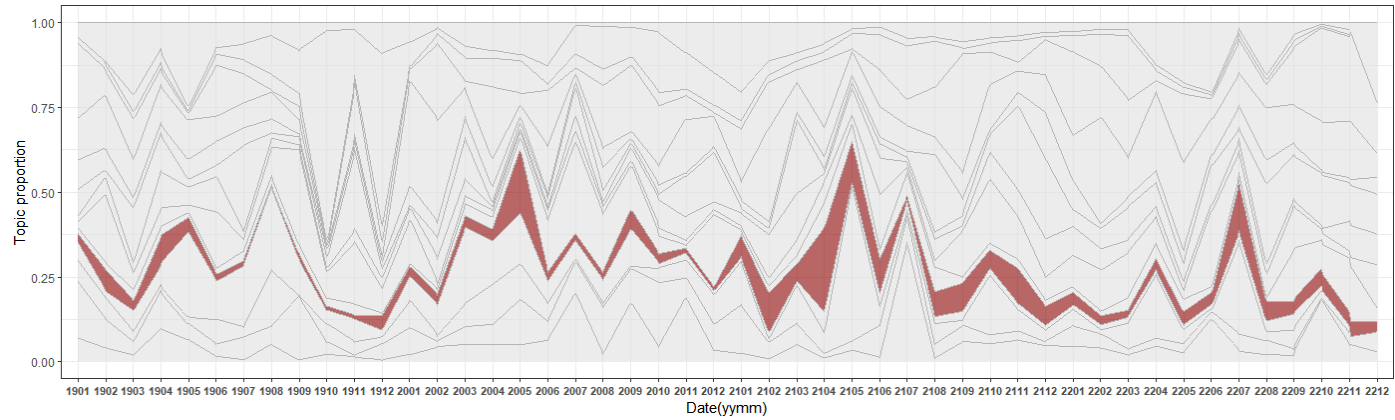


Electronic Cigarettes in Media

Trends of topics : Electronic Cigarettes in Media (N=5,162) (%)

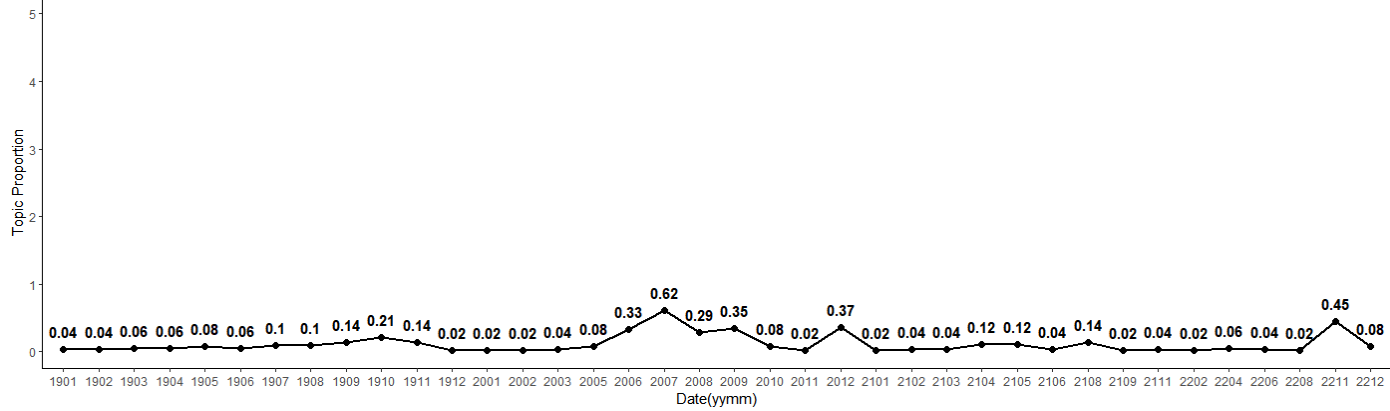


Trends of topics : 'Electronic Cigarettes in Media'

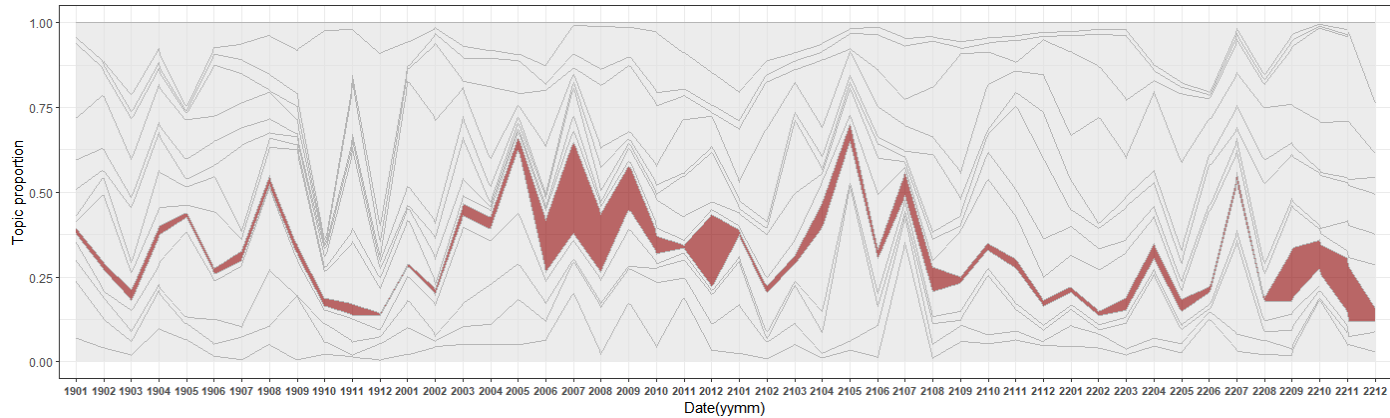


Comprehensive Taxation Issues including tax increase on E-cigarettes' liquid

Trends of topics : Comprehensive Taxation Plan including tax increase on E-cigarettes' liquid (N=5,162) (%)

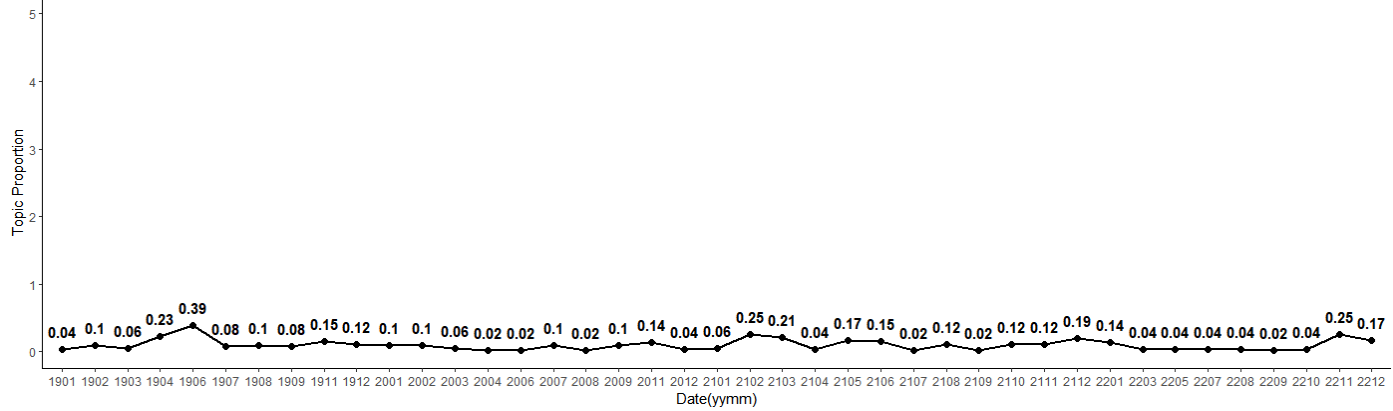


Trends of topics : 'Comprehensive Taxation Issues including tax increase on E-cigarettes' liquid'

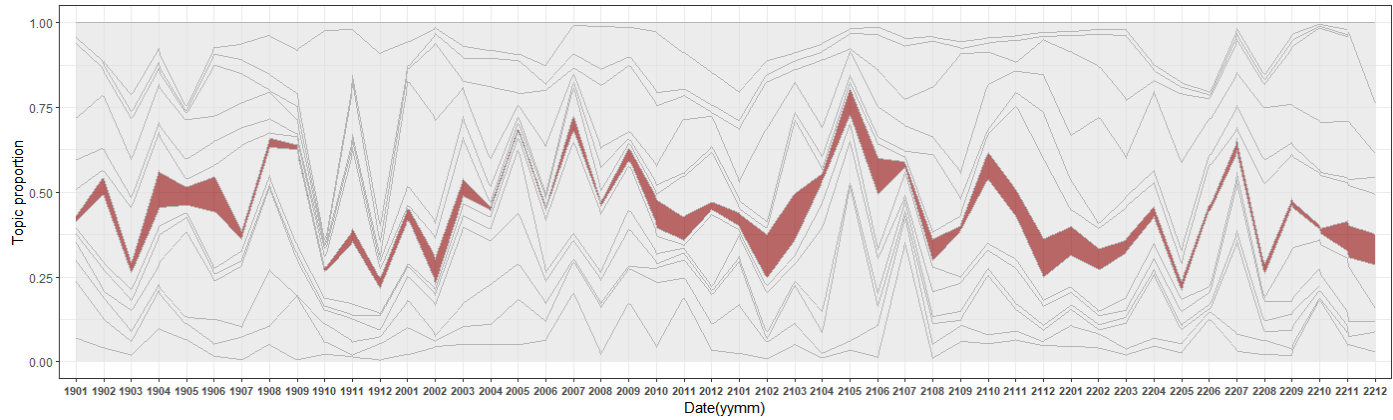


Activity for Smoke-free Environment Construction

Trends of topics : Activity for Smoke-free Environment Construction (N=5,162) (%)

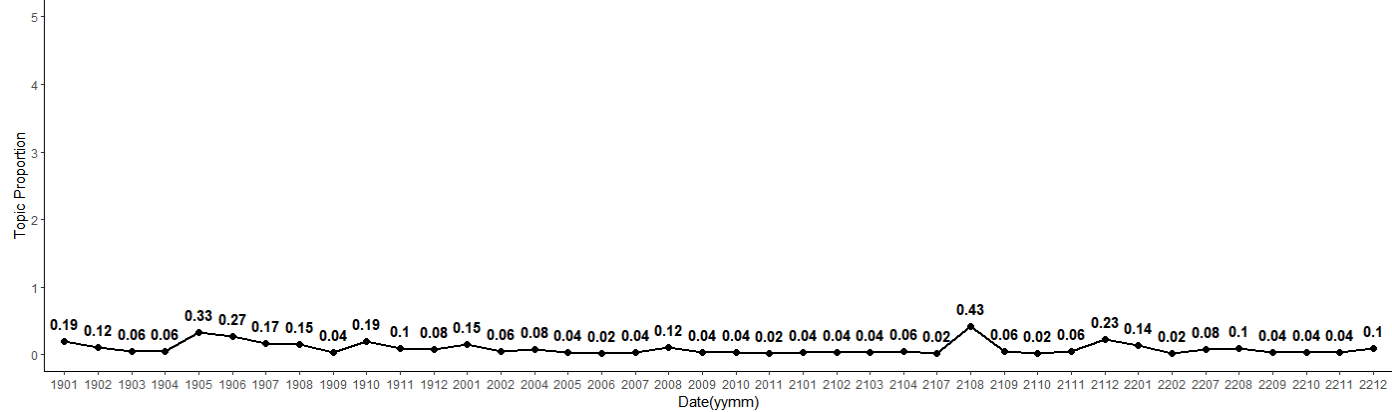


Trends of topics : 'Activity for Smoke-free Environment Construction'

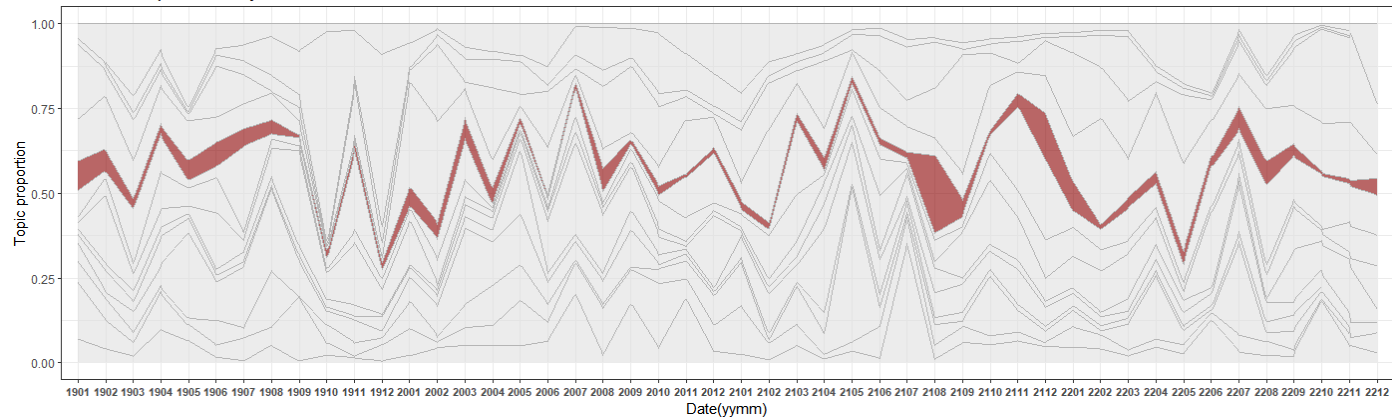


Safety of Electronic Devices

Trends of topics : Safety of Electronic Devices (N=5,162) (%)

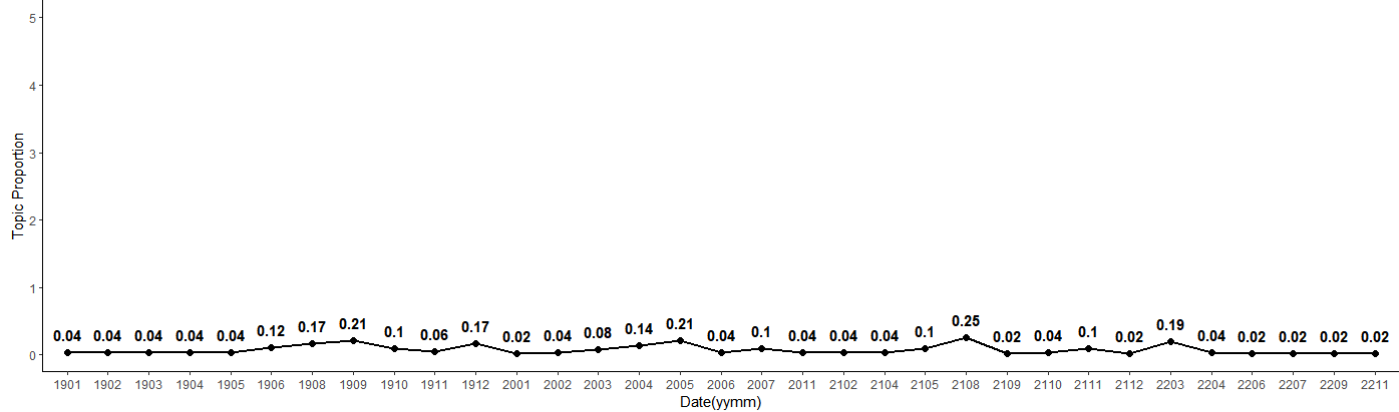


Trends of topics : 'Safety of Electronic Devices'

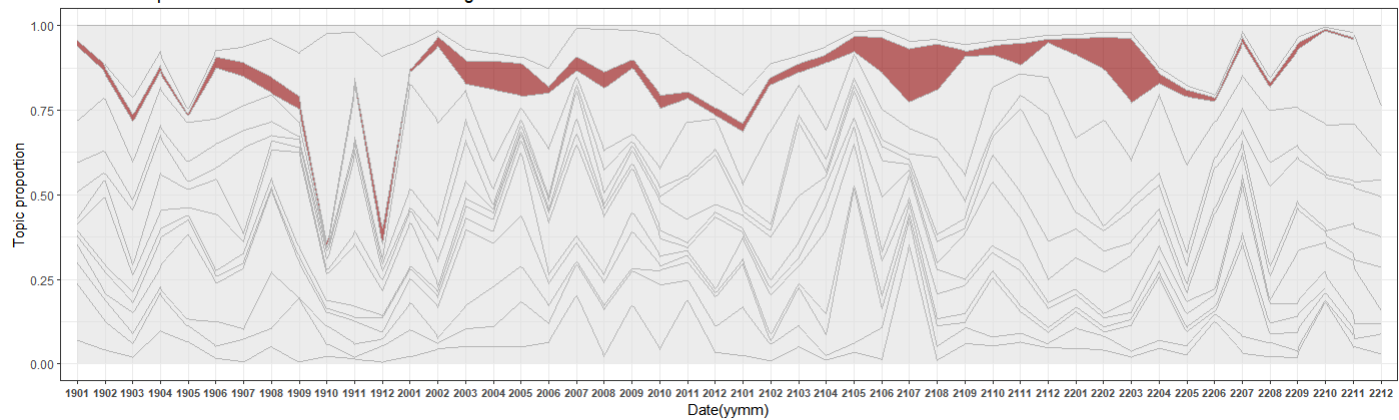


Defensive Action for HTPs as Cigarette Substitutes

Trends of topics : Defensive Action for HTPs as Cigarette Substitutes (N=5,162) (%)



Trends of topics : 'Defensive Action for HTPs as Cigarette Substitutes'



Chapter 4. Discussion

In order to achieve ‘tobacco-free future’ , it is important to consider the journalism ecosystem as one of the factors capable of shaping the beliefs of policy recipient[77]. Numerous studies have highlighted the potential influence of mass media on the dissemination of e-cigarettes and HTPs; however, the role of newspapers in this context has not received adequate attention. This study carries implications as it emphasizes the necessity of monitoring and analyzing the discourse surrounding tobacco products and tobacco control policies. By employing topic modelling as an analytical approach, this study provides the grounds for identifying the marketing strategies employed by the tobacco industry to promote the usage of tobacco products and build positive image of them. Moreover, it aims to minimize subjective biases inherent in human coding procedures. It is important to note that models using unsupervised learning techniques have a limitation, as the results can vary depending on the random seed number used for model estimation and the selection of the analysis model. Considering this point, this study adopts a comparative approach by examining the outcomes generated by both LDA and STM, thereby avoiding reliance on a single topic model.

According to the results of this study, a notable conflict exists between a marketing-oriented position and a policy-oriented position, as depicted in Figure 12. The Marketing-oriented position advocates for electronic cigarettes as ‘less harmful’ alternatives for smokers, promoting activities related to Harm Reduction and Smoke-free environments. Multinational tobacco companies, such as British American Tobacco (BAT) and Philip Morris International (PMI), have adopted harm reduction in their public relations

initiatives and marketing communication, which provides them an opportunity to showcase their engagement in new product development of self-styled ‘next-generation products’ (Figure 13) [78]. For PMI, a ‘smoke-free world’ is largely presented as getting existing users of cigarettes and combustible tobacco to switch to innovative non-combustible products that are positioned as harm reduced [79–83]. The implementation of smoke-free policies that do not allow cigarette smoking indoors has reduced the consumption of cigarettes, nevertheless, marketing communication for next-generation products commonly encourages the use of e-cigarettes in settings where conventional cigarette smoking is not allowable. Furthermore, this stance also emphasizes an eco-friendly and forward-looking attitude, aiming to encourage users to choose better alternatives that ‘benefit’ both themselves and the environment. Conversely, policy-oriented position adopts a more cautious approach towards the use of electronic cigarettes, particularly concerning adolescents and women, with a focus on prevention.

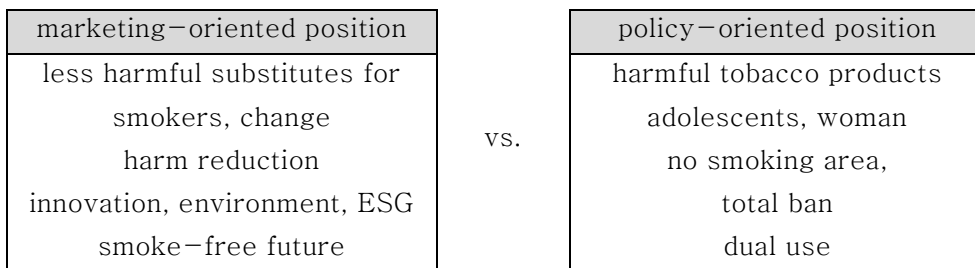


Figure 12. Conflict between Positions on Electronic Cigarettes



Figure 13. Harm Reduction Strategy by Tobacco Industry (BAT)

Considering the previous studies on the framing of public health issues, health problems are defined on the terrain, which is that once a health problem is “discovered,” assigning responsibility for causes and solutions forms the crux of public discourse on the issue [84]. Previously, there was a discourse of 'Smoking vs. Non-smoking' on conventional cigarettes. However, as a result of this study, it was found that the latest discourse was 'Conventional Cigarettes vs. Other Tobacco(or Nicotine Delivery) Products'. It seems that since Electronic Cigarettes spread to the Korean market in earnest, tobacco industry has been trying to erase the option of "complete smoking cessation" that does not use any tobacco products, by diversifying the tobacco product itself, which is the subject of the existing discourse competition.

On the other hand, as a result of this study, it was found that the proportion of advertising topics or of topics building a positive image of companies was greater in economic newspapers than in general newspapers. In addition, when comparing the frequently used words by newspaper companies, it was confirmed that some economic and general newspapers frequently mentioned specific tobacco products' or tobacco companies' names. This shows that even in the case of general newspapers, tobacco companies and tobacco product names were frequently mentioned, because of the

results Korean newspapers have increasingly relied on corporate advertising, which can act as invisible pressure for news organizations to portray advertisers positively [85].

The findings of this study reveal the widespread presence of electronic cigarette advertisements in the online news media. In addition to direct advertising through features like ‘New Product Launching’, the articles on the topic of ‘Industry Business Performance’ provided detailed descriptions on how the tobacco industry achieved economic success. Consequently, there is a risk of indirect advertising through news articles that cannot be overlooked.

Particularly noteworthy is the industry’s utilization of terms in product names that highlight functional advantages while downplaying health risks. Current regulations prohibit the use of words that may cause misunderstanding or minimize health hazards when describing tobacco products¹⁷. Furthermore, terms, symbols, pictures, and photographs that might lead to the overlooking of health hazards are also restricted by law. However, the names of electronic cigarette products include forward-thinking and positive words, such as ‘change’, ‘hybrid’, ‘available’, and ‘sense’, as well as words that evoke the anticipation of sticks or liquid’s flavors such as ‘tropical’, ‘ice’, ‘shine’, ‘cool’, and ‘tundra’. The manner in which the media, including newspapers, addresses health issues can significantly influence individual health perceptions, beliefs, behaviors, and even public

¹⁷ Article 10-2 in Enforcement Decree of The Tobacco Business Act
(1) Misleading words, etc., referred to in Article 25-5 (2) of the Act, means any of the following words, phrases, trademarks, shapes, or other indication: 1. Light; 2. Mild; 3. Low tar; 4. Genuine;
5. Indications that are likely to cause any misconceptions about tobacco by treating lightly the effects of tobacco on human health or the risks of tobacco thereto, with the indication of a symbol, figure, picture, or three-dimensional shape or a combination thereof.

opinion. These examples suggest that online media advertisements may serve as environmental factors of tobacco products use such as the prevalent advertisements in convenience stores in Korea.

Therefore, future research specifically on the impact of media marketing is essential. It is imperative to analyze how the media and reporting guidelines on marketing influence individuals' perception of health and their behavioral choices. A previous study investigating the effect of exposure to electronic cigarette advertisements demonstrated that exposure to electronic cigarette advertisements was significantly associated with the increased use of electronic cigarettes[86].¹⁸ The study further revealed that exposure to electronic cigarette advertisements was associated with the increased use of them, specifically in retail stores and on Internet or social media among adolescents and young adults. It is crucial to determine whether electronic cigarettes marketing through the news media channels positively influences the perception of tobacco company brands or products, especially among adolescents and young adults, leading to product use driven by curiosity about flavors. If extensive exposure to advertisements has a detrimental effect on smoking cessation efforts, regulatory measures aimed at controlling electronic cigarette marketing in the news media become necessary.

The topics identified in the additional STM analysis by year reveal changes in the primary anchor products of the tobacco industry between the years 2019 and 2022. In response to negative issues surrounding certain products, sales strategies have been adapted by both improving existing products and switching the main sales products from e-cigarettes to HTPs. For instance, following the suspension of e-cigarettes in 2019, there was a decline in e-cigarette sales, resulting in a scarcity of related articles by 2022.

¹⁸ aOR = 1.53, 95% CI 1.22 to 1.92; I2 = 60.1%

On the other hand, articles related to ‘New HTPs launching’ displayed consistent frequency over time despite fluctuations, and they constituted a significant portion of industry–related articles in 2022. Therefore, the findings of this study indicate that the industry not only focuses on enhancing product features and functions but also employs promotional methods to facilitate product sales in compliance with government regulations.

Article 5(3) of the FCTC mandates that public health policies should be protected from the commercial interests of the tobacco industry when establishing and implementing public health policies on tobacco regulation. This provision is based on the understanding that various activities within the tobacco industry serve as conduits for the global tobacco epidemic. Given that the tobacco industry's activities can shape the perception and trajectory of tobacco use, it is important to recognize that seemingly "positive", such as business growth, market expansion, corporate social responsibility initiatives, and media interviews beyond product launches, contribute to the overall image of the tobacco industry's image. Therefore, in order to advance and protect public health, it is crucial to closely monitor all forms of media that deliver health–related information to individuals.

Meanwhile, this study has some limitations. First, the methods used in this study assumes that each word has only one meaning. A much more reasonable assumption is that words assume different meanings based upon their appearance alongside other words. In the topic related to specific theme, for example the increase in cigarette tax, the nuance of positive–negative sentiment conveyed in the text cannot be grasped only by the composition of the word. In the future, it is necessary to grasp the opinions and arguments for and against related situation or environments through emotional analysis and network analysis. Also, since this study analyzed only

daily articles, text from other media such as SNS was not included in the analysis, so it has a limitation of exploring only part of the discourse. In particular, since comments or SNS posts on newspaper articles are the result of readers or users' direct opinions, it is expected that the terms and topics that appeared in newspapers will differ in the social perception of new tobacco products by the general public. Therefore, it is believed that more meaningful implications will be obtained if follow-up studies such as comparative analysis of differences in perceptions between experts and the general public are carried out through methods such as FGI for texts from more diverse media in the future.

By exploring the discourse on 'new and emerging tobacco products' in Korean online news articles, this study reveals the frequent presence of TAPS activity on the news coverage was frequent and consistent with using specific and changing rhetoric in Korea. To achieve tobacco endgame¹⁹, real-time and systematic monitoring on tobacco-related media contents is needed for preventing people to using tobacco products[87].

¹⁹ Tobacco endgame strategies represent a paradigm shift in tobacco control. The following elements attempt to define 'real' endgame strategies; 1. Having an explicit government intention and plan to achieve close to zero prevalence of tobacco use. 2. A clearly stated government 'end' target date within a maximum of two decades.

Chapter 5. Conclusion

This study utilized topic modeling methods to examine the contents of Korean online news articles pertaining to E-cigarettes and Heated Tobacco Products, aiming to analyze the discourse surrounding these products. The findings of this study revealed that the frequently used words varied depending on the type of newspaper and publication company. Additionally, the topics covered in the news articles on E-cigarettes and Heated Tobacco Products highlighted the conflicting positions between a marketing-oriented perspective and a policy-oriented perspective on electronic cigarettes. The topic of 'New Product Launching' consistently appeared throughout the years, providing purchase-related information. To effectively address the evolving marketing strategies employed by the tobacco industry and mitigate the gateway effect, wherein the use of e-cigarettes leads to subsequent use of conventional cigarettes, it is crucial to monitor the content of online media and implement stricter control policies that specifically target cross-border Tobacco Advertising, Promotion, and Sales.

Bibliography

- [1] KDCA. “Korea National Health and Nutrition Examination Survey FACT SHEET. Changes of Health Behaviors and Chronic Disease Prevalence for 20 years(1998–2018)”. (2020).
- [2] Current Smoking rate [Internet]. K-indicator. (2023). [cited 7 February 2023]. Available from: <https://www.index.go.kr/unify/idx-info.do?idxCd=4237>
- [3] Ministry of Economy and Finance. “Market Trends of Tobacco in 2022” Ministry of Economy and Finance press release, January 31, 2023. On the Ministry of Economy and Finance website.
https://www.moef.go.kr/nw/nes/detailNesDtaView.do?searchBbsId=MOSFBBS_000000000028&menuNo=4010100&searchNttId=MOSF_000000000062809, accessed February 10, 2023.
- [4] Lee, Sungkyu, and Jinyoung Kim. “Evolution of Tobacco Products.” *Journal of the Korean Medical Association* 63, no. 2 (2020): 88. <https://doi.org/10.5124/jkma.2020.63.2.88>.
- [5] KDCA·TCIKC. “2022 KDCA Report on Harmful Effects of Tobacco: An Overview of Tobacco Use and its Effects on Health” (2022).
- [6] World Health Organization. (2021). WHO report on the global tobacco epidemic, 2021: addressing new and emerging products.
- [7] Lee AS, Hart JL, Walker KL, Keith RJ, Ridner SL. (2018). Dual users and Electronic cigarette Only Users: Consumption and Characteristics. *Int J Healthc Med Sci.* 4(6): 111–6.
- [8] Lim, Young Hwa, Keum Ji Jung, Jakyoun Lee, Jiwoo Baek, and Heejin Kimm. “E-Cigarette and Combustible Cigarette Single Use, Ever Dual Use Pattern, and Amount of Nicotine Consumed among Korean College Students.” *Journal of the Korean Society*

- for Research on Nicotine and Tobacco 13, no. 3 (September 30, 2022): 93–104. <https://doi.org/10.25055/jksrnt.2022.13.3.93>.
- [9] Tobacco. “Tobacco,” May 24, 2022. <https://www.who.int/news-room/fact-sheets/detail/tobacco>.
- [10] CHO, HONG–JUN. "Comparison of the risks of combustible cigarettes, e–cigarettes, and heated tobacco products" Journal of the Korean Medical Association 63, no.2 (2020) : 96–104.
- [11] WHO Framework Convention on Tobacco Control. (2013). Guidelines for implementation article 13.
- [12] Enforcement Decree Of The Tobacco Business Act., Article 9. (2020).
- [13] National Health Promotion Act, Article 17472. (2022).
- [14] Hwang, Ji–eun. (2020). A Qualitative Study on the Level of Interest in Heated Tobacco Products in South Korea: Using Data of Naver Search Trend and Online News Media Coverage. Health and Social Welfare Review, 40(4), 611–635. <https://doi.org/10.15709/HSWR.2020.40.4.611>
- [15] RAND. “Media Influences on Health,” n.d. <https://www.rand.org/topics/media-influences-on-health.html>.
- [16] Noar, Seth M. “A 10–Year Retrospective of Research in Health Mass Media Campaigns: Where Do We Go From Here?” Journal of Health Communication 11, no. 1 (January 2006): 21–42. <https://doi.org/10.1080/10810730500461059>.
- [17] Coleman, Renita, and Esther Thorson. “The Effects of News Stories That Put Crime and Violence Into Context: Testing the Public Health Model of Reporting.” Journal of Health Communication 7, no. 5 (October 1, 2002): 401–25. <https://doi.org/10.1080/10810730290001783>.
- [18] YANOVITZKY, ITZHAK, and JO STRYKER. “Mass Media, Social Norms, and Health Promotion Efforts.” Communication Research 28, no. 2 (April 2001): 208–39.

<https://doi.org/10.1177/009365001028002004>.

- [19] Brodie, Mollyann, Elizabeth C. Hamel, Drew E. Altman, Robert J. Blendon, and John M. Benson. "Health News and the American Public, 1996–2002." *Journal of Health Politics, Policy and Law* 28, no. 5 (October 1, 2003): 927–50. <https://doi.org/10.1215/03616878-28-5-927>.
- [20] In Sook Cho, Doh Yeon Kim, "Coverage of Korean Newspapers on Tobacco Lawsuits : Denormalization, Favorability, and Foreign Examples", *Journal of Speech, Media & Communication Research*, 제28호, 2015, 114–153.
- [21] Nelson, D. E., Pederson, L. L., Mowery, P., Bailey, S., Sevilimedu, V., London, J., ... & Pechacek, T. (2015). Trends in US newspaper and television coverage of tobacco. *Tobacco Control*, 24(1), 94–99.
- [22] Joseph, N., Sivanantham, P., Sulgante, S., Singh, R. J., Goel, S., & Kar, S. S. (2021). Depiction of tobacco in Indian mass media: A content analysis. *Indian Journal of Tuberculosis*, 68, S23–S28.
- [23] Abraham, E. A., Egbe, C. O., & Ayo–Yusuf, O. A. (2019). News media coverage of shisha in Nigeria from 2014 to 2018. *Tobacco induced diseases*, 17.
- [24] Løchen, M. L., Torhild Gram, I., Skattebo, S., & Kolstrup, N. (2007). Tobacco images and texts in Norwegian magazines and newspapers. *Scandinavian journal of public health*, 35(1), 31–38.
- [25] Durrant, R., Wakefield, M., McLeod, K., Clegg–Smith, K., & Chapman, S. (2003). Tobacco in the news: an analysis of newspaper coverage of tobacco issues in Australia, 2001. *Tobacco control*, 12(suppl 2), ii75–ii81.
- [26] Braun, S., Mejia, R., Barnoya, J., Gregorich, S. E., & Pérez–Stable, E. J. (2011). Tobacco advertising and press coverage of

- smoking and health in 10 years of Argentinean newspapers. *CVD prevention and control*, 6(3), 71–80.
- [27] He, S., Shen, Q., Yin, X., Xu, L., & Lan, X. (2014). Newspaper coverage of tobacco issues: An analysis of print news in Chinese cities, 2008–2011. *Tobacco Control*, 23(4), 345–352.
- [28] Wackowski, O. A., Lewis, M. J., Delnevo, C. D., & Ling, P. M. (2013). A content analysis of smokeless tobacco coverage in US newspapers and news wires. *nicotine & tobacco research*, 15(7), 1289–1296.
- [29] Sun, S., Veltri, G. A., & Wang, F. (2018). Representations of electronic cigarettes in Chinese media. *BMC Public Health*, 18(1), 1–11.
- [30] Gunnar, S., Rikke Iren, T., & Ingeborg, L. (2022). Newspaper coverage of snus in an emerging norwegian snus market 2002–2011: A content analysis. *Nicotine and Tobacco Research*, 24(2), 212–219.
- [31] Haddock, C. K., Hoffman, K., Taylor, J. E., Schwab, L., Poston, W. S., & Lando, H. A. (2008). An analysis of messages about tobacco in the Military Times magazines. *Nicotine & Tobacco Research*, 10(7), 1191–1197.
- [32] Rooke, C., & Amos, A. (2014). News media representations of electronic cigarettes: an analysis of newspaper coverage in the UK and Scotland. *Tobacco Control*, 23(6), 507–512.
- [33] Smith, K. C., Terry–McElrath, Y., Wakefield, M., & Durrant, R. (2005). Media advocacy and newspaper coverage of tobacco issues: a comparative analysis of 1 year's print news in the United States and Australia. *Nicotine & Tobacco Research*, 7(2), 289–299.
- [34] Harris, J. K., Shelton, S. C., Moreland–Russell, S., & Luke, D. A. (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.

- [35] YungWook Kim, SeungKyung Ham. (2014). Discursive Struggles between Nonsmoking and Smoking Sectors : An Application of Critical Discourse Analysis. *Korean Journal of Journalism & Communication Studies*, 58(5), 333–361.
- [36] Youjin Choi. (2017). A Content Analysis of Tobacco Control Policy Coverage in 2011~2016: Story Type, Theme and Tone Comparison by Policy. *Korean Journal of Journalism & Communication Studies*, 61(2), 64–92.
- [37] Cho Insook, KIM DOH YEON. (2015). Coverage of Korean Newspapers on Tobacco Lawsuits: Denormalization, Favorability, and Foreign Examples. *한국소통학보*, (28), 114–153.
- [38] Jun, Jungmi, Sei-Hill Kim, James Thrasher, Yoo Jin Cho, and Yu-Jin Heo. “Heated Debates on Regulations of Heated Tobacco Products in South Korea: The News Valence, Source and Framing of Relative Risk/Benefit.” *Tobacco Control* 31, no. e1 (February 4, 2021): e57–63.
<https://doi.org/10.1136/tobaccocontrol-2020-056131>.
- [39] Kim, Sei-Hill, Jungmi Jun, James F. Thrasher, Yu-Jin Heo, and Yoo Jin Cho. “News Media Presentations of Heated Tobacco Products (HTPs): A Content Analysis of Newspaper and Television News Coverage in South Korea.” *Journal of Health Communication* 26, no. 5 (May 4, 2021): 299–311.
<https://doi.org/10.1080/10810730.2021.1931988>.
- [40] Mi-Young Oh. (2021). A Study on the Tobacco Industry’s Publicity Strategies: Based on the Analysis of Tobacco Company’s Documents and the Media Coverage of Tobacco and Tobacco Companies. *JOURNAL OF THE KOREA CONTENTS*

ASSOCIATION, 21 (6), 165–179.

- [41] Lee, Cheol Min. “The Impact of Heated Tobacco Products on Smoking Cessation, Tobacco Use, and Tobacco Sales in South Korea.” *Korean Journal of Family Medicine* 41, no. 5 (September 20, 2020): 273–81. <https://doi.org/10.4082/kjfm.20.0140>.
- [42] Korea Disease Control and Prevention Agency. “Electronic Cigarettes.” Korea Disease Control and Prevention Agency, August 25, 2022. <https://www.kdca.go.kr/contents.es?mid=a20205010602>.
- [43] DiMaggio, Paul, Manish Nag, and David Blei. “Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding.” *Poetics* 41, no. 6 (December 2013): 570–606. <https://doi.org/10.1016/j.poetic.2013.08.004>.
- [44] 김사승. (2013). 신문뉴스의 상품가치에 관한 미디어 경제학적 관점의 분석. *Korean Journal of Journalism & Communication Studies*, 57(3), 5–33.
- [45] Kostadinova, Petia, and Daniela V Dimitrova. “Communicating Policy Change: Media Framing of Economic News in Post-Communist Bulgaria.” *European Journal of Communication* 27, no. 2 (June 2012): 171–86. <https://doi.org/10.1177/0267323112449097>.
- [46] Tobacco: E-cigarettes. “Tobacco: E-Cigarettes,” May 3, 2023. <https://www.who.int/news-room/questions-and-answers/item/tobacco-e-cigarettes>.
- [47] Enforcement Decree of the National Health Promotion Act., Article 27–2. (2018).
- [48] Entman, Robert M. “Framing: Toward Clarification of a Fractured Paradigm.” *Journal of Communication* 43, no. 4 (December 1, 1993): 51–58. <https://doi.org/10.1111/j.1460->

[2466.1993.tb01304.x](https://doi.org/10.1136/bmj.320.7247.1468).

- [49] Bland, J M. “Statistics Notes: The Odds Ratio.” *BMJ* 320, no. 7247 (May 27, 2000): 1468–1468. <https://doi.org/10.1136/bmj.320.7247.1468>.
- [50] Bland, J M., and D. G Altman. “Statistics Notes: Transforming Data.” *BMJ* 312, no. 7033 (March 23, 1996): 770–770. <https://doi.org/10.1136/bmj.312.7033.770>.
- [51] Kaiser, S., & Ali, R. (2018). Text mining: use of TF–IDF to examine the relevance of words to documents. *International Journal of Computer Applications*, 181(1), 25–29.
- [52] Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1), 11–21.
- [53] Spärck Jones, K. (2004). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 60(5), 493–502.
- [54] Zhang, Wen, Taketoshi Yoshida, and Xijin Tang. “A Comparative Study of TF*IDF, LSI and Multi–Words for Text Classification.” *Expert Systems with Applications* 38, no. 3 (March 2011): 2758–65. <https://doi.org/10.1016/j.eswa.2010.08.066>.
- [55] Desai, A. (2015). A review on knowledge discovery using text classification techniques in text mining. *International Journal of Computer Applications*, 111(6).
- [56] M. Sukanya and S. Biruntha, "Techniques on text mining," 2012 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), Ramanathapuram, India, 2012, pp. 269–271, doi: 10.1109/ICACCCT.2012.6320784.
- [57] KDnuggets. “Text Mining 101: Topic Modeling – KDnuggets,” n.d. <https://www.kdnuggets.com/text-mining->

- 101-topic-modelling.html.
- [58] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993–1022.
 - [59] Grün, B., & Hornik, K. (2011). topicmodels: An R package for fitting topic models. *Journal of statistical software*, 40, 1–30.
 - [60] Ioana. “Latent Dirichlet Allocation: Intuition, Math, Implementation and Visualisation.” Medium, September 26, 2020. <https://towardsdatascience.com/latent-dirichlet-allocation-intuition-math-implementation-and-visualisation-63ccb616e094>.
 - [61] Roberts, M. E., Stewart, B. M., & Airoidi, E. M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, 111(515), 988–1003.
 - [62] Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software*, 91, 1–40.
 - [63] Bischof, J., & Airoidi, E. M. (2012). Summarizing topical content with word frequency and exclusivity. In *Proceedings of the 29th international conference on machine learning (icml-12)* (pp. 201–208).
 - [64] Airoidi, E. M., & Bischof, J. M. (2016). Improving and evaluating topic models and other models of text. *Journal of the American Statistical Association*, 111(516), 1381–1403.
 - [65] KIM, Hyunguk, and Jinwoong SONG*. “Analysis of Articles Related to Quantum Mechanics by Using Topic Modeling: Focus on Comparing LDA, CTM, STM Models.” *New Physics: Sae Mulli* 70, no. 5 (May 29, 2020): 493–502. <https://doi.org/10.3938/npsm.70.493>.

- [66] Baek, Young Min, and Misa Park. “Introduction to Statistical Analytic Techniques for Text Data and Suggestions for Social Sciences in the Age of Data Science.” *Social Science Review* 49, no. 1 (May 31, 2018): 188–210. <https://doi.org/10.31502/ssri.49.1.9>.
- [67] Blei, D. M., & Lafferty, J. D. (2006, June). Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning* (pp. 113–120).
- [68] Kherwa, P., & Bansal, P. (2019). Topic modeling: a comprehensive review. *EAI Endorsed transactions on scalable information systems*, 7(24).
- [69] Roberts, M., Stewart, B., Tingley, D., Benoit, K., Stewart, M. B., & Rcpp, L. (2022). Package ‘stm’. Imports matrixStats, R. & KernSmooth, (2017).
- [70] Béchara, Hannah, Alexander Herzog, Slava Jankin, and Peter John. “Transfer Learning for Topic Labeling: Analysis of the UK House of Commons Speeches 1935–2014.” *Research & Politics* 8, no. 2 (April 2021): 205316802110222. <https://doi.org/10.1177/20531680211022206>.
- [71] Arun, R., Suresh, V., Veni Madhavan, C.E., Narasimha Murthy, M.N.: On finding the natural number of topics with latent Dirichlet allocation some observations. In: Zaki, M.J., Yu, J.X., Ravindran, B., Pudi, V. (eds.) *Advances in Knowledge Discovery and Data Mining. Lecture Notes in Computer Science*, pp. 391–402. Springer, Berlin (2010)
- [72] Cao, J., Xia, T., Li, J., Zhang, Y., Tang, S.: A density-based method for adaptive LDA model selection. *Neurocomputing* 72(7–9), 1775–1781 (2009)
- [73] Deveaud, R., SanJuan, É., Bellot, P.: Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique* 17(1), 61–84 (2014)

- [74] TobaccoTactics. “PMI Promotion of IQOS Using FDA MRTP Order – TobaccoTactics,” n.d.
<https://tobaccotactics.org/wiki/pmi-iqos-fda-mrtp-order/>.
- [75] 광영섭. “S. Korea Counters Philip Morris’ Info Disclosure Lawsuit | Yonhap News Agency.” Yonhap News Agency, November 26, 2018.
<https://en.yna.co.kr/view/AEN20181126002500320>.
- [76] Implement MPOWER measures to reduce tobacco use. “Implement MPOWER Measures to Reduce Tobacco Use,” December 20, 2021.
<https://www.who.int/westernpacific/news-room/multimedia/overview/item/implement-mpower-measures-to-reduce-tobacco-use>.
- [77] Bae Jung-geun. "A Study on Advertising’s Influence on Newspaper Reports" Korean Journal of Journalism & Communication Studies (KJJCS) 54, no.6 (2010) : 103–128.
- [78] British American Tobacco – Tobacco harm reduction in focus. “British American Tobacco – Tobacco Harm Reduction in Focus,” June 7, 2018.
https://www.bat.com/group/sites/UK__9D9KCY.nsf/vwPagesWebLive/DOAZHFT9.
- [79] Dewhirst, Timothy. “Co-Optation of Harm Reduction by Big Tobacco.” Tobacco Control 30, no. e1 (August 12, 2020): e1–3.
<https://doi.org/10.1136/tobaccocontrol-2020-056059>.
- [80] Chapman, S, R Borland, M Scollo, R C Brownson, A Dominello, and S Woodward. “The Impact of Smoke-Free Workplaces on Declining Cigarette Consumption in Australia and the United States.” American Journal of Public Health 89, no. 7 (July 1999): 1018–23.
<https://doi.org/10.2105/ajph.89.7.1018>.
- [81] Heloma, Antero, Maritta S. Jaakkola, Erkki Kähkönen, and

- Kari Reijula. “The Short–Term Impact of National Smoke–Free Workplace Legislation on Passive Smoking and Tobacco Use.” *American Journal of Public Health* 91, no. 9 (September 2001): 1416–18. <https://doi.org/10.2105/ajph.91.9.1416>.
- [82] Fichtenberg, C. M. “Effect of Smoke–Free Workplaces on Smoking Behaviour: Systematic Review.” *BMJ* 325, no. 7357 (July 27, 2002): 188–188. <https://doi.org/10.1136/bmj.325.7357.188>.
- [83] US Department of Health and Human Services. (2000). *Reducing tobacco use: a report of the Surgeon General*.
- [84] Lawrence, Regina G. “Framing Obesity.” *Harvard International Journal of Press/Politics* 9, no. 3 (July 2004): 56–75. <https://doi.org/10.1177/1081180x04266581>.
- [85] 이봉수, & 제정임. (2007). 경제저널리즘의 종속성: 한국 신문의 재벌 보도와 광고와의 관계.
- [86] Luu, Ngoc Minh, Thanh Hai Phan, Jin–Kyoung Oh, and Seung–Kwon Myung. “Exposure to Electronic Cigarette Advertisements and Use of Electronic Cigarettes: A Meta–Analysis of Prospective Studies.” *Nicotine & Tobacco Research* 25, no. 5 (November 25, 2022): 983–90. <https://doi.org/10.1093/ntr/ntac266>.
- [87] Thomson, G., Edwards, R., Wilson, N., & Blakely, T. (2012). What are the elements of the tobacco endgame?. *Tobacco control*, 21 (2), 293–295.

Abstract in Korean

신종담배 관련 국내 온라인 신문 기사 내용 분석: 토픽 모델링(Topic Modeling)의 적용

강나은

보건학과 보건학전공

서울대학교 보건대학원

연구 목적: 담배업계가 다양한 미디어 채널을 통해 신종담배 홍보 활동을 적극적으로 펼쳐왔으나, 그 잠재적인 광고 위험성에 비해 국내 언론 속 신종담배에 대한 관심은 부족한 실정이다. 본 연구는 정량적인 텍스트 분석을 통해 국내 온라인 일간지 기사에서 등장한 신종담배 관련 담론에 접근하고자 하였다.

연구 방법: 분석 대상으로 2019년부터 2022년까지 전국 종합 일간지 11곳과 경제 일간지 8곳에서 보도된 기사 중, 검색어인 '신종담배', '전자담배', '가열담배', '액상담배', '썬담배' 중 하나 이상을 포함한 모든 기사를 빅카인즈(www.bigkinds.or.kr)를 통해 수집하였다. 기술 통계 분석으로 전반적인 보도 동향을 파악하고, 출현 빈도가 높은 단어들을 파악하였다. 경제지와 일반지로 구분되는 신문의 종류와 각 신문사에 따라 상대적으로 더 많이 쓰인 단어를 비교하기 위하여 단어 로그 오즈비와 TF-IDF 값을 도출하고 이들을 비교하였다. 분석 대상 기사 내용의 주요 주제들을 파악하기 위해, 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA)과 구조적 토픽 모델(Structural Topic Model, STM)을 통한 토픽 모델링 분석을 실시하였다.

연구 결과: 해당 기간 중 총 5,162건의 기사가 보도되었으며, 기사건수는 2019년에 가장 많았으나 이후 급격히 감소하는 모습을 보였다. 신문의 종류와 신문사에 따라 상대적으로 더 자주 사용된 단어의 종류가 상이하였다. 경제 일간지일수록 담배 업계의 활동과 관련된 단어를 더 많이 사용하였으며, 일부 신문사에서 특정 담배회사명 및 제품명이 두드러지게 사용되었다. 전체 기사에서 5회 이상 등장한 총 14,001개의 단어를 바탕으로 LDA에서 11개, STM에서 15개의 토픽을 확인하였다. 두 토픽 모델 모두에서 뚜렷하게 등장한 토픽은 '신제품 출시', '업계 사업 실적', '신종담배의 건강 위험과 사용 행태', '액상형 전자담배의 건강 위험과 국내 사용 중단', '전자담배 불법 사용 관련 이슈', '비가격 담배 규제 정책(P, W)', '액상형 전자담배 증세', '정부의 포괄적 과세 정책'이었다. 이중 '신제품 출시'에 대한 토픽이 두 모델에서 가장 많은 비중을 차지하였다.

결론: 국내 일간지 온라인 기사에서 이루어지는 신종담배 관련 담론을 탐색함으로써, 언론에서 담배 업계의 광고 및 홍보 활동이 빈번하고 일관적으로 전개되고 있음을 확인하였다. 또한 시간의 흐름에 따라 홍보에 새롭게 사용된 문구를 확인함으로써 담배 업계의 전략 변화를 추측할 수 있었다. MPOWER 중 이행 수준이 미비한 '광고, 판촉 및 후원(TAPS)'의 규제 수준을 강화하기 위해서는, 광고성 내용이 일간지에 등장하는 현 문제 상황을 개선할 필요가 있다. 향후 연구를 통해 신종담배 마케팅 기사에의 노출 여부가 담배 사용 및 금연에 끼치는 영향을 분명히 규명하고, 노출군의 건강 행동에 부정적 영향을 끼칠 시 언론에서의 담배 제품의 광고, 홍보 및 판촉 모니터링과 규제 정책이 강화되어야 할 것이다. 본 연구는 최근 수년간 이용이 증가하고 있는 쉐련형 및 액상형 전자담배에 초점을 두고 토픽 모델링 방법을 통해 신문기사 내용 분석을

시도한 국내 첫 연구로, 국민건강 증진을 위한 언론의 생태계를 점검하고 헬스 리터러시 증진 방안을 마련하는 데에 도움이 되기를 기대한다.

주요어: 전자담배, 신종담배, 신문기사, 토픽 모델링, LDA, STM

Appendix

Appendix 1. The Example Table Form for Cross-checking topic labelling

LDA		
Topics	Words	
	담배, 전자, 제품, 출시, 기기, 소비자, 아이코스, 판매, 사용, 쉼련, 편의점, 전용, 한국필립모리스, 프로, 제공, 전국, 고객, 적용, 가능, 브랜드	
	담배, 전자, 액상, 사용, 판매, 제품, 중단, 미국, 정부, 성분, 전자담배, 권고, 유해, 가향, 물질, 비타민, 결과, 조치, 중증, 검출	
	담배, 금연, 전자, 흡연자, 흡연, 전자담배, 쉼련, 사용, 일반, 건강, 니코틴, 결과, 그림, 경고, 사람, 일반담배, 교수, 위험, 정책, 발생	
	담배, 제품, KT&G, 전자, 시장, 쉼련, PMI, 출시, 판매, 소비자, 필립모리스, 아이코스, 2.0, BAT, 글로벌, 한국, 하이브리드, 브랜드, 일본, 세계	
STM		
Topics	Words by Prob	Words by FREX
	출시, 담배, 전자, KT&G, 제품, 쉼련, 2.0, 판매, 하이브리드, 아이코스, 소비자, 기기, 프로, 시장, 전국, 확대, 스틱, 일루마, 사용, 적용	에이블, 부스트, 에디션, Fiit, HYBRID, MIIX, 2.0, 일루마, 믹스, 죠즈코리아, 인텍션, 솔리드, 스마트코어, 미니, 프라임, 전용스틱, 에임, 프로, 1.0, 블레이드
	KT&G, 담배, 시장, 전자, 매출, 증가, 실적, 해외, 수출, 대비, 성장, 전년, 쉼련, 전망, 글로벌, 기록, PMI, 확대, 사업, 점유	영업이익, 이애텍, 이랜텍, 백복, FCP, 목표주가, 고객사, 인삼공사, 한국인삼공사, 아이티엠반도체, 실적, 배당주, 주가, 호실적, KGC, 중동, 호조, 해외법인, 컨센서스, 매수

