Article

Characterizing News Report of the Substandard Vaccine Case of Changchun Changsheng in China: A Text Mining Approach

Ping Zhou [1], Yao He [2], Chao Lyu [1] and Xiaoguang Yang [1],*

[1] Department of Hospital Management, Key Laboratory of Health Technology Assessment of National Health Commission (Fudan University), School of Public Health, Fudan University, Shanghai 200032, China; zhouping@fudan.edu.cn (P.Z.); 14301020033@fudan.edu.cn (C.L.)

[2] Department of Global Health, School of Public Health, University of Washington, Seattle, WA 98195-7965, USA; yaohe@uw.edu

* Correspondence: yangxg@fudan.edu.cn; Tel.: +86-21-33565182

check for updates

Abstract: *Background:* The substandard vaccine case of that broke out in July 2018 in China triggered an outburst of news reports both domestically and aboard. Distilling the abundant textual information is helpful for a better understanding of the character during this public event. *Methods*: We collected the texts of 2211 news reports from 83 mainstream media outlets in China between 15 July and 25 August 2018, and used a structural topic model (STM) to identify the major topics and features that emerged. We also used dictionary-based sentiment analysis to uncover the sentiments expressed by the topics as well as their temporal variations. *Results*: The main topics of the news report fell into six major categories, including: (1) Media Investigation, (2) Response from the Top Authority, (3) Government Action, (4) Knowledge Dissemination, (5) Finance Related and (6) Commentary. The topic prevalence shifted during different stages of the events, illustrating the actions by the government. Sentiments generally spanned from negative to positive, but varied according to different topics. *Conclusion*: The characteristics of news reports on vaccines are shaped by various topics at different stages. The inner dynamics of the topic and its alterations are driven by the interaction between social sentiment and governmental intervention.

Keywords: substandard vaccine case; news report; text mining; topic model; sentiment analysis

1. Introduction

Vaccination is the most effective medical approach that can be used to eliminate suffering from the health and financial burdens caused by large-scale infectious disease transmission. While vaccination usually accompanies wide societal concern, the mass media also plays an influential role on its acceptance [1–3]. The Strategic Advisory Group of Experts on Immunization Working Group on Vaccine Hesitancy listed communication and media environment as a key influence on vaccine hesitancy [4,5]. From a research perspective, news reports are a very important vessel for information, as they convey the knowledge, attitudes, and sentiments of a society [6–9]. Such an abundance of information has been acknowledged by many researchers in the area of vaccines, and studies have tried to reveal the underlying factors that influence vaccination acceptance based on the trace implications of news texts. For example, Becker et al. measured confidence in vaccinations using a multinational media surveillance system [10], and Faasse et al. analyzed the influence of news coverage and Google searches on Gardasil adverse event reporting [11], among other approaches.

Vaccine incidents or related events are an important aspect of vaccine research, particularly when covered by the media and reflecting extensive concern from the society [12–14], as they can have a

potential impact on vaccination confidence [15,16]. These events can cause average citizens with little professional knowledge on vaccinations to worry, particularly for their children [17–19]. On 15 July 2018, a scandal broke out when the Changchun Changsheng Biotech Co. Ltd., one of the major vaccine manufacturers in China, was confirmed to have fabricated production and inspection records and arbitrarily changed process parameters and equipment during its production of freeze-dried human rabies vaccines [20]. Within a week, a huge wave of concern and discussion was raised from mass media. This event drew the attention of the Chinese President Xi Jinpin, who characterized the scandal as "veiled in nature and shocking" and demanded a thorough investigation on 24 July [21]. In the meantime, the World Health Organization also addressed the importance of vaccinations, and called for actions on regulation [22]. The case reached a conclusion on 16 August when the standing committee of the Communist Party of China, the top power authority of China, held more than 40 government officials accountable, including seven at the provincial or ministry level [23], and the Changsheng Limited was ordered to pay about 9.1 billion yuan (USD 1.3 billion) in penalties on 16 October [24].

As a shocking incident, the case of the substandard vaccine by the Changchun Changsheng Company (referred to as the "Changsheng case" for short) triggered an outburst from news reports both in China and aboard, and the incident provides an opportunity to inspect how the media reacts to a sensational public health event. To investigate the large scale and great diversity of the news reports, we employed computer-assisted text mining tools to review the news texts directly. We believe such innovative method could potentially reduce laborious human review or coding work, and could lead to a better presentation of the characteristics of the reporting on the Changsheng case from a holistic perspective.

Topic is an important entry point to inspecting the characteristics of news reports [25–27]. In regards to the Changsheng case, the news reports comprised a great variety of topics such as vaccine safety, weak regulating systems, as well as the financial misconduct of the company. Such topics, characterized by proportion, temporality, and sentiment, are helpful to understanding the inner structure, temporal variation, and sentiment of the reports. Specifically, we want to focus on the following research questions:

RQ1: What topics emerged in the news reports during the Changsheng case, and how were they are featured in terms of prevalence and key words?

RQ2: How did the different topics change over time?

RQ3: What sentiments were expressed through the media and how did they change over time from a topical perspective.

In this paper, we first introduce the methods used for analyzing the news report texts. Then we examine the media presentation of the event, including the major topics that emerged from the news reports and their distribution over time. Sentiments present in the news reports were identified based on keywords from temporal and topical perspectives. Finally, we interpret the research findings and draw conclusions.

2. Materials and Methods

Text mining approaches were used in this study to characterize news reports of the Changsheng case. Specifically, we tried retrieving the quantified features, like topic prevalence, temporal alteration and sentiment propensity, and demonstrated the features statistically to provide a full picture of the case. Retrieving information from news texts can be challenging, as it is highly unstructured compared with other semi-structural text sources such as legal documents [28], patent documents [29], or electronic health records [30]. In this case, we used a topic modeling approach to deconstruct the news contents, and a lexicon-based method to analyze the news sentiments. We tried to analyze a broad selection of the mainstream newswires inside China to reduce preference or bias from any individual news source.

2.1. Materials

We designated the data collection period between 15 July and 25 August 2018, which covered the break out of the event until the official conclusion. We identified 2211 news articles related to the Changsheng case from 384,254 pieces of news published by 83 media outlets in China during the specified period (see Supplementary Materials for detail). The news texts were processed by conventional natural language processing procedures including word segments, removals of numbers, punctuation, and stop words [31]. The minimum word length was kept to two. The final corpus contained the columns of title, date, source, and content for each included news article.

2.2. Topic Model: Primary Analytical Approach

2.2.1. Overview of Topic Model

Topic model analysis is an important approach by which to inspect a large quantity of textual data [32]. It provides an intuitive way of identifying what topics potentially exist in the corpus and captures quantities of interest [33]. Such identification can transfer the unstructured textual data into a low-dimension quantitative feature, and could be combined with other methods such as regression [34], time series [34], and sentiment analysis [35].

The topic model method assumes that there are a number of potential topics available during the collection of documents, and that each document or word belongs to a certain topic with a different probability [36]. The classical topic model, basing on latent dirichlet allocation (LDA), assumes that specific document is generated by first nominating a topic from the potential topic-document distribution, then selecting a word from the potential word-document distribution, as shown in Figure 1. Assuming there are N documents, K topics, and V words in the whole corpus, θ is the length-K per document-topic distribution for document d, β is the length-V per topic-word distribution for a certain k-th topic, and $Z_{d,n}$ is the selected topic from which the observed words $W_{d,n}$ are chosen [36]. α and η are the hyperparameters initially set for model fitting.
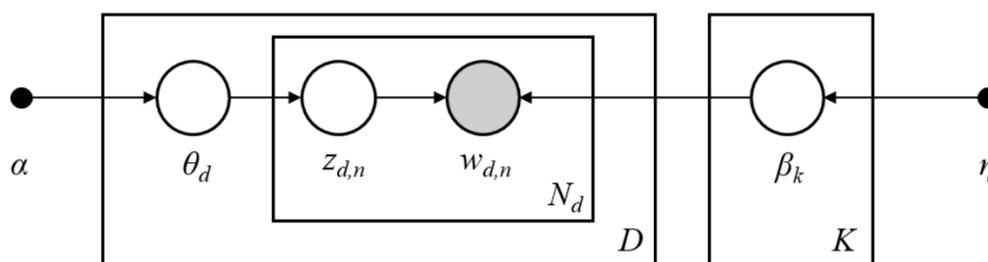

Figure 1. Textual data generative mechanism of the Topic Model [33].

The mathematic definition of the classic topic model proposed by Blei et al. is as follow:

$$\theta \sim Dir(\alpha)$$

$$\beta \sim Dir(\beta)$$

$$Z_{d,n}|\theta_d \sim Mutinomial(\theta_d)$$

$$W_{d,n}|Z_{d,n} \sim Multinomial\left(\beta_{Z_{d,n}}\right)$$

Fitting the topic model helps us identify the parameters from a given corpus. For domain-specific research, the most interesting measured quantity is the θ: the proportion of topics relevant to a certain document [33]. Using a statistical aggregation, we can calculate the prevalence of the topics in the whole corpus then probe the semantic structure of the corpus [33].

2.2.2. Structural Topic Model-Based Data Analyses

Topic models are widely applied to a variety of text formats such as newspapers [37,38], patent contents [29], social media [39,40], research articles or reports [41,42], and so on. As an extension of the topic model, the recently developed structural topic model (STM for short) could provide a way of quantifying the effects of document properties (e.g., time of creation, sources) to a specific topic's prevalence, which is useful for exploring the features of the topics [33,43]. Robert et al., the author of the STM, proposed the model as follows:

$$\theta_d \big| \big(C_{d,\gamma}, \Sigma \big) \sim LogisticNorm\big(C_{d,\gamma}, \Sigma \big)$$

$$\beta_{d,k} \propto exp\big(m + k_k + k_{g,d} + k_{k_{g,d}} \big)$$

$$Z_{d,n} \big| \theta_d \sim Multinomial(\theta_d)$$

$$W_{d,n} \big| Z_{d,n} \sim Multinomial\big(\beta_{Z_{d,n}} \big)$$

The format and data generation mechanisms are similar between the LDA and STM. The major difference is the change in topic prevalence from Dirichlet distribution into logistic normal distribution, which can incorporate covariates. The parameters k_k, $k_{g,d}$, $k_{k_{g,d}}$ represent the specific deviations of the topics, covariates, and interaction topic-covariates, respectively.

In this paper, we used the STM to investigate: (1) what topics emerged from the news reports that were related to the Changsheng case, (2) the quantitative prevalence of the topics from all the reports, and (3) how topic prevalence changed over time. Selecting the suitable number of topics (K) is a challenge in topic model analysis. Despite this, there are quantitative criteria to support selection [32,37], and most studies eventually rely on human judgement [34,37,41]. In this study, we followed the method of Roberts et al. [33], using the build Semantic Coherence and Exclusivity to get an overview of the coverage of topics under different K, then determined the final topic number by manually reviewing the results.

2.2.3. Sentiment Analysis

Sentiment polarity and strength [44] on the news reports of the Changsheng case were calculated from both time and topical perspectives to describe the overall sentiment expression and its variation across the whole event. We also identified the top sentiment terms that contributed to the news text [45]. In this paper, we employed a lexicon-based analysis [46] to calculate the quantitative sentiment propensity of the news report. The lexical dictionary created by the Dalian University of Science and Technology (DUST) [47] was used in this study with minor adaptations. For example, we removed the word "Changsheng" (which means "long life" in Chinese), a very positive word in the DUST dictionary, while incorporating new terms that appeared in the reporting on Changsheng case such as "violated the moral bottom line" (mentioned by Premier Li Keqiang) as a passive word [47].

The above works were implemented using the R (3.4.4) programming language and packages. Specifically, the fitting and visualization of the topic model were conducted using the *stm* [48] and *stminsights* [49] packages, while the sentiment analysis was implemented using the *tidytext* package [50].

3. Results

3.1. News Occurrence of the Case

The first chart in Figure 2 shows the change in the amount of news reporting over the time period that the study sample covered (15 July to 25 August 2018) in the Chinese context. Compared to Google Trends, which shows the popularity of topics on the internet (second chart), and the Weixin Index (the most popular social media mobile application similar to WhatsApp in China), which shows the

popularity of items on the mobile internet (third chart), the time distribution of the news reports included in this study matched that of the internet and mobile internet closely.
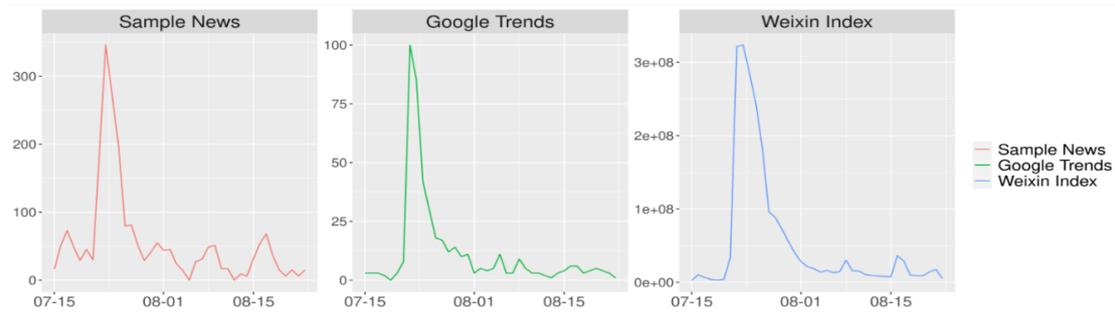

Figure 2. Temporal trends of the news reports included in the study with the corresponding Google trends and Weixin index, indicating the social attention to this incident.

3.2. Topics that Emerged from the Corpus

3.2.1. Overall Topical Presentation

Table 1 shows the 17 topics that automatically emerged from fitting the STM, as well as their proportion in the corpus. We listed the top 10 words (ranked by β_k in Figure 1) belonging to each topic, added a representative label to each topic, and categorized them into six groups.

To discover how the topics of news reporting changed over time, we divided the observation period into three-day intervals to calculate the proportional distribution of the 17 reporting topics within each interval. Based on the topic distribution, we categorized reports into four periods, namely, the initial period, outbreak period, continuation period, and ending period (Figure 3).
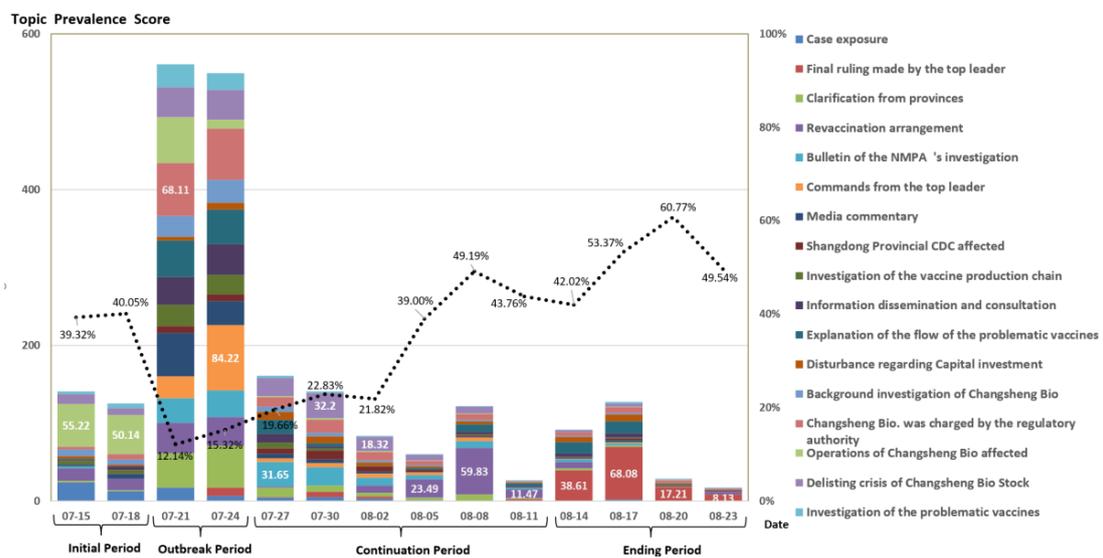

Figure 3. Prevalence and temporal variation of the topics. Note: The dotted line is the proportion of the highest prevalence topics at a certain time point.

Table 1. Topics and keywords in the news reports on the Changsheng Bio case.

Category	Label of Topics	Key Words	Proportion
Media Investigation	1. Case exposure	counterfeit, manufacturing, pharmaceutical, corporation, sales, product, Changsheng Bio.	0.034
	2. Background investigation of Changsheng Bio	Changsheng Bio, company, BioKangtai, shareholding, life sciences, 100 million-yuan, transfer, Changchun High & New Tech Industry Inc	0.036
	3. Investigation of the vaccine production chain	Changsheng Bio, sales, 100 million yuan, e10,000-yuan, sales expenditure, Changchun Changsheng, company	0.03
	4. Investigation of the problematic vaccines	DPT, issuance, off grade, potency, manufacturing, Wuhan Institute of Biological Products Co. Ltd., inspection	0.036
Response from the Top Authority	5. Commands from the top leader	work, adamant, conduct, case, investigation team, investigation, State Council	0.063
	6. Final ruling made by the top leader	regulation, meeting, pharmaceutical, problematic, work, case, safety, company	0.073
Government Action	7. Bulletin of the NMPA's investigation	corporation, manufacturing, underway, NDA, batch, DPT, company, Changchun Bio	0.083
	8. Shangdong Provincial CDC affected	journalist, Shandong, procure, Changsheng Bio, Changchun Changsheng, DHPPi, injection	0.023
	9. Charged by the regulatory authority	company, Changsheng Bio, bulletin, Changchun Changsheng, disclose, provision, Changsheng Bio	0.081
	10. Clarification from provinces	Changsheng Bio, DPT, case, problematic, vaccination, response, off grade	0.048
	11. Explanation of the flow of the problematic vaccines	revaccination, vaccination, DPT, children, off grade, work, dose, immunization	0.063
	12. Revaccination arrangement	vaccination, rabies vaccine, Changchun Changsheng, company, revaccination, observation, CDC, national	0.108
Q&A	13. Information dissemination and consultation	vaccination, DPT, children, tetanus, kids, pertussis, off grade, batch number	0.053
Finance Related	14. Operations of Changsheng Bio affected	Changsheng Bio, company, Changchun Changsheng, manufacturing, product, freeze-dried human rabies vaccine, pharmaceutical	0.107
	15. Disturbance regarding Capital investment	project, account, Changsheng Bio, Changsheng, company, capital, journalist, investment,	0.022
	16. Delisting crisis of Changsheng Bio Stock	Changsheng Bio, fund, delisting, company, valuation, Changsheng, limit down	0.085
Commentary	17. Media commentary	problematic, case, regulation, corporation, general public, China, counterfeit, manufacturing, media	0.056

Note: The Key Words and Proportion columns are automatically created by the STM algorithm, and indicate the significance of the keywords within specific topics, and the semantic proportion of specific topics among the whole corpus, respectively. The Category and Label of Topics columns are annotated by the authors. Q&A: questions and answers. DPT: diphtheria, pertussis, tetanus. NMPA: National Medical Products Administration. CDC: Center for Disease Control and Prevention. Changsheng Bio (长生生物) and Changchun Changsheng (长生生物): Changchun Changsheng Biotech Co. Ltd. In Chinese.

During the initial period (15–20 July), most news reports were case investigations, and the topics of more than half the reports regarded Operations of Changsheng Bio affected (Shenzhen Stock Exchange) and Case Exposure (53.51%).

During the outbreak period (21–25 July), the amount of reporting increased rapidly once the president addressed the case, and the distribution of each topic was fairly similar. Some of the major topics were Changsheng Bio Charged in the Investigation (12.07%), Commands from the Top Leader (10.14%), Clarification from Provinces (8.78%), Explanation of the Flow of the Problematic Vaccines (8.24%), and Media Commentary (7.79%). Government entities at all levels and various sectors in society were highly concerned with the Changsheng case during this period. The peak of media attention lasted for five days, then the number of news reports started to decrease on 26 July.

During the continuation period (26 July–12 August) the dominant topics of reporting shifted towards Results of the Investigation of the National Medical Products Administration (NMPA) and Revaccination Arrangement of the National Health Commission (NHC). The number of reports on these two topics grew rapidly and became mainstream. During ending period (after 13 August), the amount of reporting on the Changsheng case experienced little fluctuation, and the major topic remained Final Ruling (42–61%). Once the State Council published the progress of their investigation into the Changsheng case on 6 August, and the amount of reporting decreased rapidly.

3.2.2. Time Trends of the Topic Categories

Figure 4 shows the trends of topic categories over time to further illustrate the progress of the case. During the observed period, the prevalence of topics in the Media Investigation category decreased steadily over time. The prevalence of topics in the Commentary category increased rapidly during the initial period, reaching its peak on 22 July (eight days after the case exposure), then decreased rapidly afterwards and remained at a low level towards the end. The prevalence of the Q&A of Vaccination topics fluctuated over time, peaking when the authorities officially announced the investigation results, then decreasing steadily after the second peak. The category of Finance Related received a lot of media attention at the beginning of the initial period, but the prevalence of topics in this category decreased and fluctuated afterwards.
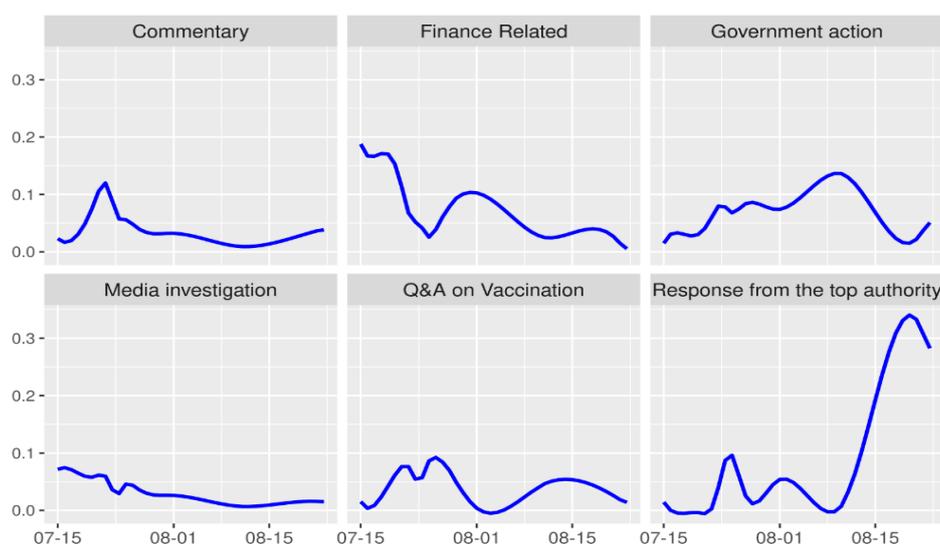

Figure 4. Time trends of the topic categories.

The prevalence of topics in the Response from the Top Authority and Government Action category increased following the initial case exposure. Three peaks in the prevalence of Response from the Top Authority appeared around the 10th, 20th, and 35th days following exposure, which corresponded to the critical time points when the president issued commands, the State Council published the progress of the investigation, and the president announced the final ruling, respectively. The third peak indicates that the final ruling from the president almost dominated news reporting at that time, after which the media attention to this case would start to dissipate. The prevalence of topics in Government Action increased steadily and peaked on the 25th day following exposure.

3.3. Sentiment Analysis of the Case

3.3.1. Daily Sentiment Score

We found that, despite some fluctuation, sentiment was mostly negative during the first 23 days following the initial case exposure, and that the absolute values of the negative sentiment scores were exceptionally high on the 1st, 6th, 17th, 19th, and 20th days, when there was extensive discussion in the media. Following the 24th day, sentiment became increasingly positive, and there was a sharp increase in positive sentiment towards the end of the observation period (Figure 5).
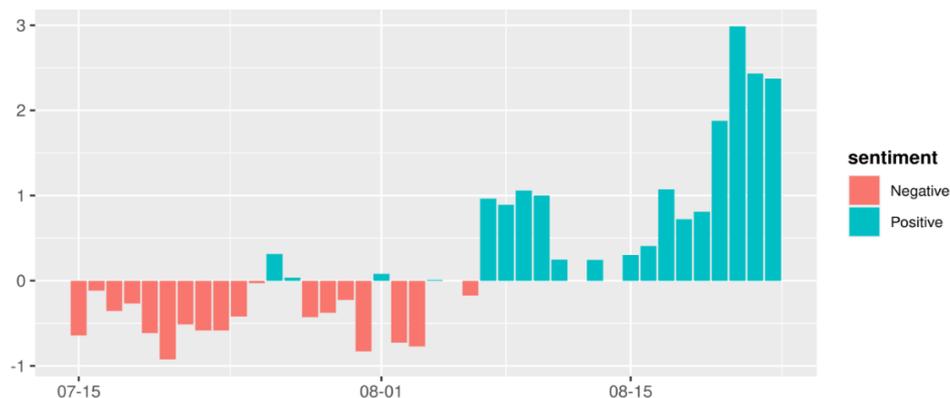

Figure 5. Daily sentiment score of the news.

3.3.2. Sentiment by Topic

According to our sentiment analysis by topics (Figure 6), the news reports on Revaccination Arrangement, Explanation of the Flow of the Problematic Vaccines, and Final Ruling Made by the Top Leader expressed positive sentiment, particularly those on Final Ruling. Such topics included the confirmative attitudes of the government, such as "guarantee", "adamant", and "crystal clearly", among others. While news reports on the rest of the topics mostly expressed negative sentiment, particularly on Operations of Changsheng Bio Affected (Shenzhen Stock Exchange), Changsheng Bio was Charged by the Regulatory Authority, Clarification from Provinces, and Media Commentary. These results reflected the strong negative attitudes from the government, society, and media towards the Changsheng case.
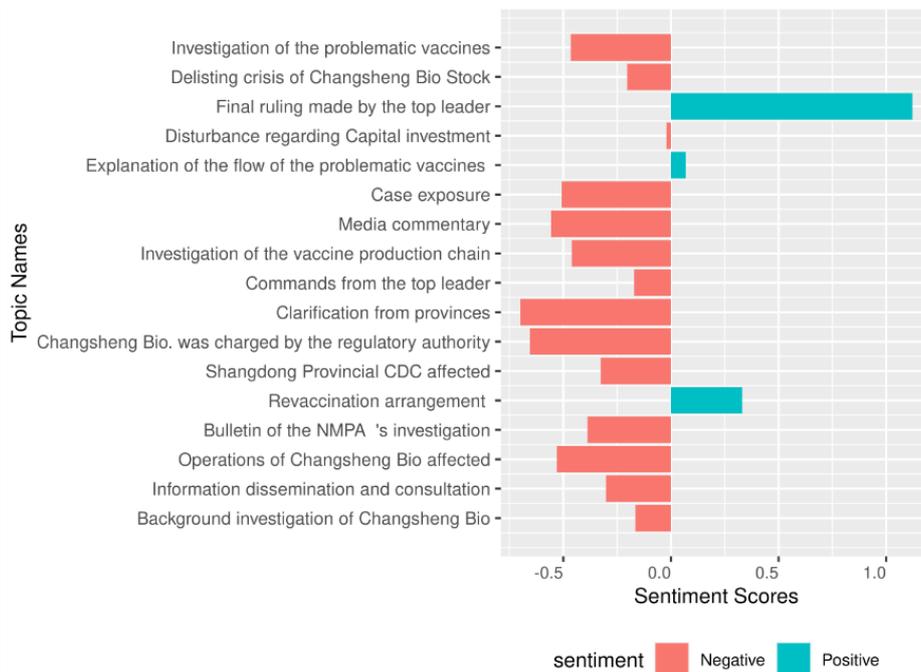
Figure 6. Sentiment scores of the 17 topics.

3.3.3. Most Frequently Used Sentiment Words

Figure 7 shows the most frequently used sentiment words from the news texts. In terms of negative sentiment, the most frequently used words were "off grade", "suspected", "bribery", and "moral bottom line", with contributions of −0.086, −0.078, −0.052, and −0.039, respectively. In terms of positive sentiment, the most frequently used words were "guarantee", "adamant", "health", and "ensure", with contributions of 0.075, 0.059, 0.57, and 0.056, respectively (Figure 7).
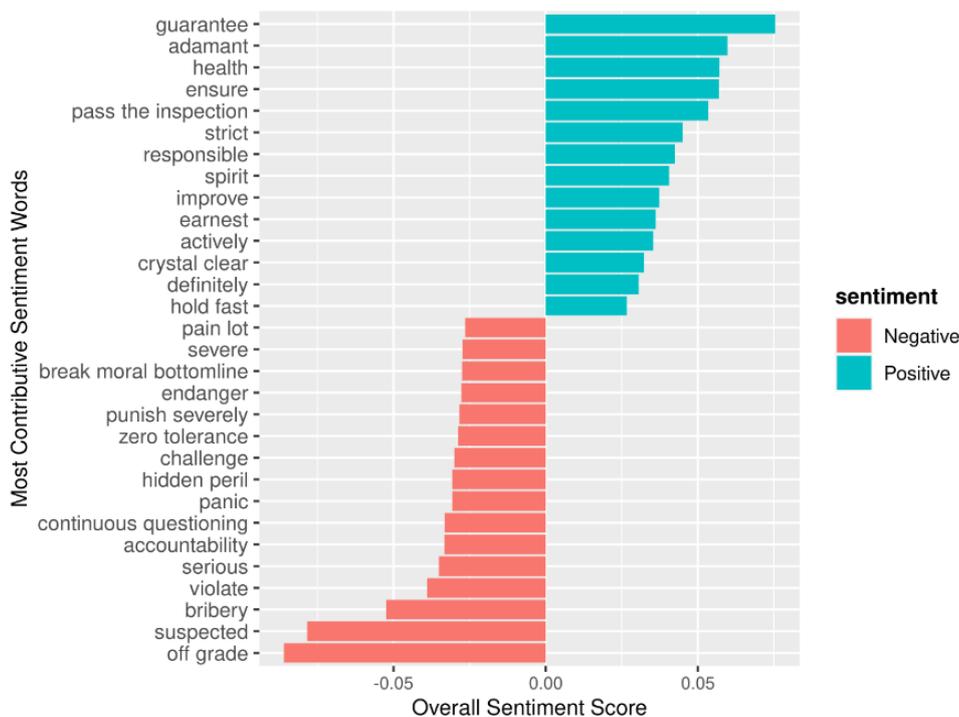

Figure 7. Sentiment words used most frequently in the corpus.

3.3.4. Temporal Alteration in the Contribution of the Sentiment Words

To show at what point during the observation period each word was most relevant, we analyzed how the three-day average score of each sentiment word changed over time. Figure 8 shows that negative sentiment was dominant during the initial period, with words such as "hidden peril", "panic", "suspected", "zero tolerance", and "violate" being used. During the outbreak period, both positive and negative sentiments coexisted. Once the continuation period began, the overall sentiment started to turn positive, with words such as "guarantee", "ensure", "health", and "pass the inspection" emerging more frequently. The appearance of negative sentiment words such as "challenge", "panic", and "violate" fell rapidly to almost zero towards the end of the continuation period, particularly after 6 August. Since the dominating news topic was Final Ruling Made by the Top Leader during the ending period, the sentiment of the media was mostly positive, and the words "spirit", "adamant", "guarantee", "earnest", and "strict" appeared with high frequency.
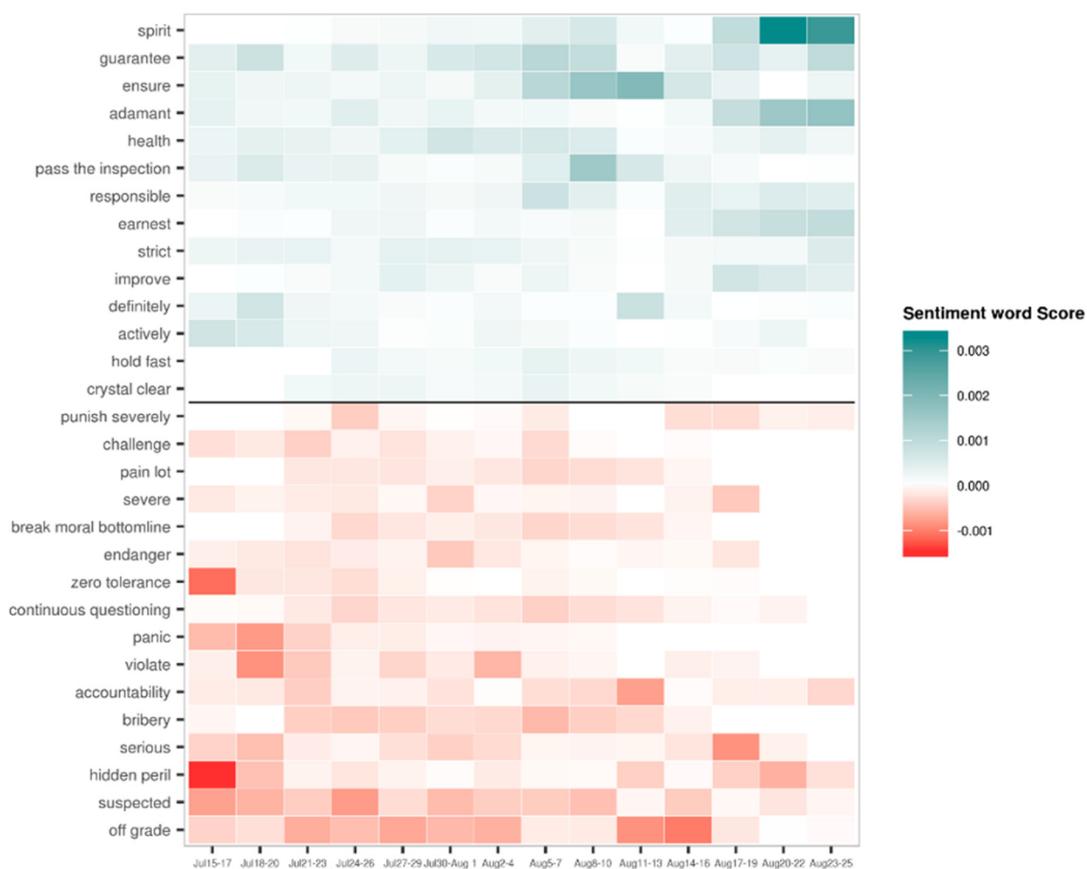

Figure 8. Temporal changes in the contribution of various sentiment words.

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretations, and the experimental conclusions that can be drawn from them.

4. Discussion

In this study we used quantitative textual analysis to show how the media reported on the Changsheng case. While many existing studies have analyzed the news coverage of vaccination issues on a long-term scale [14,51], we focused on the response to a bursting vaccine incident over a short time span only. We began by examining the temporal trends of the incident based on news volume, and distinguished the four phases, namely, the initial period, the outbreak period, the continuation

period, and the ending period (Figures 2 and 3). This is in line with the life cycle of news reporting on an emergency incident, according to journalism and communication studies [52–54].

We further deconstructed the topics and references to the Changsheng case in a mainstream media context, and identified 17 topics falling into six categories covering Media Investigation, Response from the Top Authority, Government Action, Q&A on Vaccination, Finance Related, and Commentary. The results show that the Changsheng case, amplified by news reports, went far beyond the health sector and attracted the wide attention of the whole society [55]. This implies that the vaccine issue is not only a health issue, but a public affair that relates to politics, the economy, public security, social mentality, and so on. Further research would help determine the wider implications of the results.

Interesting findings emerged by combining the temporal and topic modeling, and illustrated a shift in focus by the media over time. For instance, news reporting focused on company investigations and business affairs during the initial period, then shifted to multiple topics during in the outbreak period, and finally narrowed their focus to topics of politics and government actions in the continuation and ending periods (Figures 3 and 4). Such a shifting partly illustrates the attention of society towards the incident, and is reflective of the governmental intervention towards the case.

It is not surprising that most of the news topics during the Changsheng case had a negative sentiment propensity (Figure 5), but sentiment towards different topics also varied. Temporal sentiment shifted over the observation period, with increased negative sentiment appearing in the earlier stage of the incident (Figures 5 and 8). After President Xi expressed strong resolution to stop the counterfeit and problematic production of vaccines, overall sentiment gradually turned positive, especially towards the end of the event.

In terms of policy implications, the Changsheng case was an influential public incident for which decisive governmental action and intervention were crucial, and the news reports reflected the governmental actions throughout the event. As we can see, government-related topics had the largest proportion, including confirmation of an investigation, information dissemination, immunization consultation, revaccination arrangement, and others. Such findings confirm the research of Guofeng Wang, who showed how a top-down perspective can be adopted to legitimize the ruling party and sustain social stability during a crisis [56]. The outbreak period was the window of opportunity for government intervention, and there was a quite sharp increase in news report volume during the outbreak (see Figure 2) that then dissipated quickly. This contributed to timely and adamant government action.

From a research perspective, this study shows the potential of text mining methods for vaccine related research. Compared to existing studies on the Changsheng case that have used text mining methods such as word embedding [57], conventional topic models [58], qualitative discourse analysis [56], or human-coded key phrases [59], our research employed an enhanced topic model approach with minimal subjective judgement, and combined quantified topic proportion values with temporal and sentiment features. By analyzing qualitative textual data in a quantitative way, such methods are particularly applicable to large-scale corpora when human inspection is impossible. This sheds new light on vaccine communication studies as more and more textual data emerges on the internet.

5. Conclusions

Using a text mining approach, this study explored the characters of news reports on a sensational vaccine case in the context of China. It showed that there were four stages in the media life cycle of the vaccine case, in which the earlier period was the opportunity window for governmental intervention. With the decisive governmental action, news reporting sentiments became increasingly positive. We presented the potential for text mining to analyze vaccine-related news text, which is also applicable to other public health issues. In terms of limitations, we only focused on official news reports and did not include other data sources such as social media, which contain more information from the user side. Furthermore, the topic model was not precise enough to incorporate more specific information at

the individual level. In this case, deep learning-based natural language processing approaches would be useful for the further studies.

Supplementary Materials: The following are available online at http://www.mdpi.com/2076-393X/8/4/0691/s1, 1. Table S1: Media sources of the news reports that are included in this paper. Figure S1: Semantic coherence and exclusivity of topic number 8–30.

Author Contributions: Conceptualization: P.Z. and X.Y.; Methodology, P.Z., X.Y.; formal analysis and software, X.Y.; resource, X.Y., data curation, C.L. and Y.H., writing—original draft preparation, P.Z. and X.Y.; writing—review and editing, X.Y. and Y.H.; visualization, C.L., funding acquisition, X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the Humanities and Social Sciences Fund from the Ministry of Education of China (19YJCZH217).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Guillaume, L.R.; Bath, P.A. A content analysis of mass media sources in relation to the MMR vaccine scare. *Health Inform. J.* **2008**, *14*, 323–334. [CrossRef] [PubMed]
2. Honkanen, P.O.; Keistinen, T.; Kivelä, S.-L. The impact of vaccination strategy and methods of information on influenza and pneumococcal vaccination coverage in the elderly population. *Vaccine* **1997**, *15*, 317–320. [CrossRef]
3. Shropshire, A.M.; Brent-Hotchkiss, R.; Andrews, U.K. Mass Media Campaign Impacts Influenza Vaccine Obtainment of University Students. *J. Am. Coll. Health* **2013**, *61*, 435–443. [CrossRef] [PubMed]
4. Goldstein, S.; Macdonald, N.E.; Guirguis, S. Health communication and vaccine hesitancy. *Vaccine* **2015**, *33*, 4212–4214. [CrossRef]
5. World Health Organization. Report of the SAGE Working Group on Vaccine Hesitancy. 2014. Available online: https://www.who.int/immunization/sage/meetings/2014/october/1_Report_WORKING_GROUP_vaccine_hesitancy_final.pdf (accessed on 27 September 2020).
6. Beaudoin, C.E.; Hong, T. Health information seeking, diet and physical activity: An empirical assessment by medium and critical demographics. *Int. J. Med. Inform.* **2011**, *80*, 586–595. [CrossRef]
7. Berry, T.R.; Wharf-Higgins, J.; Naylor, P. SARS Wars: An Examination of the Quantity and Construction of Health Information in the News Media. *Health Commun.* **2007**, *21*, 35–44. [CrossRef]
8. Coleman, R.; Thorson, E.; Wilkins, L. Testing the Effect of Framing and Sourcing in Health News Stories. *J. Health Commun.* **2011**, *16*, 941–954. [CrossRef]
9. Seale, C. Health and media: An overview. *Sociol. Health Illn.* **2003**, *25*, 513–531. [CrossRef]
10. Becker, B.F.H.; Larson, H.J.; Bonhoeffer, J.; Van Mulligen, E.M.; Kors, J.A.; Sturkenboom, M.C. Evaluation of a multinational, multilingual vaccine debate on Twitter. *Vaccine* **2016**, *34*, 6166–6171. [CrossRef]
11. Faasse, K.; Porsius, J.T.; Faasse, J.; Martin, L.R. Bad news: The influence of news coverage and Google searches on Gardasil adverse event reporting. *Vaccine* **2017**, *35*, 6872–6878. [CrossRef]
12. Chai, K.-C.; Tao, R.; Chang, K.-C.; Yang, Y. Impact of China's Vaccine Incidents on the Operational Efficiency of Biopharmaceutical Companies. *Front. Public Health* **2020**, *8*, 93. [CrossRef] [PubMed]
13. Okita, T.; Enzo, A.; Kadooka, Y.; Tanaka, M.; Asai, A. The controversy on HPV vaccination in Japan: Criticism of the ethical validity of the arguments for the suspension of the proactive recommendation. *Health Policy* **2019**, *124*, 199–204. [CrossRef] [PubMed]
14. Okuhara, T.; Ishikawa, H.; Okada, M.; Kato, M.; Kiuchi, T. Newspaper coverage before and after the HPV vaccination crisis began in Japan: A text mining analysis. *BMC Public Health* **2019**, *19*, 770. [CrossRef]
15. Liu, B.; Chen, R.; Zhao, M.; Zhang, X.; Wang, J.; Gao, L.; Xu, J.; Wu, Q.; Ning, N. Vaccine confidence in China after the Changsheng vaccine incident: A cross-sectional study. *BMC Public Health* **2019**, *19*, 1564. [CrossRef] [PubMed]
16. Yang, R.; Penders, B.; Horstman, K. Vaccine Hesitancy in China: A Qualitative Study of Stakeholders' Perspectives. *Vaccines* **2020**, *8*, 650. [CrossRef]

17. Forster, A.; McBride, K.; Davies, C.; Stoney, T.; Marshall, H.; McGeechan, K.; Cooper, S.; Skinner, S. Development and validation of measures to evaluate adolescents' knowledge about human papillomavirus (HPV), involvement in HPV vaccine decision-making, self-efficacy to receive the vaccine and fear and anxiety. *Public Health* **2017**, *147*, 77–83. [CrossRef]

18. Yigit, E.; Boz, G.; Gokce, A.; Aslan, M.; Ozer, A. Knowledge, Attitudes and Behaviors of Faculty Members on Childhood Vaccine Refusal A University. *Eur. J. Public Health* **2020**, *30*, 5. [CrossRef]

19. Chen, L.; Zhang, Y.; Young, R.; Wu, X.; Zhu, G. Effects of Vaccine-related Conspiracy Theories on Chinese Young Adults' Perceptions of the HPV Vaccine: An Experimental Study. *Health Commun.* **2020**, 1–11. [CrossRef]

20. China Daily. Vaccine Producer under Investigation. Available online: http://www.chinadaily.com.cn/a/201807/15/WS5b4b31f6a310796df4df6843.html (accessed on 27 September 2020).

21. China Daily. Xi Urges thorough Probe in Vaccine Scandal. Available online: http://www.chinadaily.com.cn/cndy/2018-07/24/content_36632041.htm (accessed on 27 September 2020).

22. World Health Organization. WHO Statement on Rabies Vaccine Incident in China. Available online: https://www.who.int/china/news/detail/25-07-2018-who-statement-on-rabies-vaccine-incident-in-china (accessed on 6 October 2019).

23. China Daily. Govt Sacks 6 Officials over Vaccine Scandal. Available online: http://www.chinadaily.com.cn/a/201808/18/WS5b776d62a310add14f38674d.html (accessed on 27 September 2020).

24. China Daily. Vaccine Maker Changsheng Fined 9.1 Billion Yuan in Safety Scandal. Available online: http://www.chinadaily.com.cn/a/201810/17/WS5bc639eea310eff303282ba3.html (accessed on 27 September 2020).

25. Liu, Q.; Zheng, Z.; Zheng, J.; Chen, Q.; Liu, G.; Chen, S.; Chu, B.; Zhu, H.; Akinwunmi, B.O.; Huang, J.; et al. Health Communication Through News Media During the Early Stage of the COVID-19 Outbreak in China: Digital Topic Modeling Approach. *J. Med. Internet Res.* **2020**, *22*, e19118. [CrossRef]

26. Huang, M.; ElTayeby, O.; Zolnoori, M.; Yao, L.; Zhang, Y.; Zhang, K.; Torii, M. Public Opinions toward Diseases: Infodemiological Study on News Media Data. *J. Med Internet Res.* **2018**, *20*, e10047. [CrossRef]

27. Ghosh, S.; Chakraborty, P.; Nsoesie, E.O.; Cohn, E.; Mekaru, S.R.; Brownstein, J.S.; Ramakrishnan, N. Temporal Topic Modeling to Assess Associations between News Trends and Infectious Disease Outbreaks. *Sci. Rep.* **2017**, *7*, 40841. [CrossRef] [PubMed]

28. Petrovic, D.; Stankovic, M. Use of linguistic forms mining in the link analysis of legal documents. *Comput. Sci. Inf. Syst.* **2018**, *15*, 369–392. [CrossRef]

29. Gwak, J.H.; Sohn, S.Y. Identifying the trends in wound-healing patents for successful investment strategies. *PLoS ONE* **2017**, *12*, e0174203.

30. Leroy, G.; Gu, Y.; Pettygrove, S.; Galindo, M.K.; Arora, A.; Kurzius-Spencer, M. Automated Extraction of Diagnostic Criteria from Electronic Health Records for Autism Spectrum Disorders: Development, Evaluation, and Application. *J. Med. Internet Res.* **2018**, *20*, e10497. [CrossRef] [PubMed]

31. Wong, K.-F.; Li, W.; Xu, R.; Zhang, Z.-S. Introduction to Chinese Natural Language Processing. *Synth. Lect. Hum. Lang. Technol.* **2009**, *2*, 1–148. [CrossRef]

32. Blei, D.M.; Lafferty, J.D. A correlated topic model of Science. *Ann. Appl. Stat.* **2007**, *1*, 17–35. [CrossRef]

33. Roberts, M.E.; Stewart, B.M.; Airoldi, E.M. A Model of Text for Experimentation in the Social Sciences. *J. Am. Stat. Assoc.* **2016**, *111*, 988–1003. [CrossRef]

34. Dybowski, T.; Adämmer, P. The economic effects of U.S. presidential tax communication: Evidence from a correlated topic model. *Eur. J. Polit. Econ.* **2018**, *55*, 511–525. [CrossRef]

35. Li,, F.; Huang, M.; Zhu, X. Sentiment Analysis with Global Topics and Local Dependency. In Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10), Atlanta, GA, USA, 11–15 July 2010; pp. 1371–1376.

36. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

37. Cerchiello, P.; Nicola, G. Assessing News Contagion in Finance. *Econometrics* **2018**, *6*, 5. [CrossRef]

38. Chandelier, M.; Steuckardt, A.; Mathevet, R.; Diwersy, S.; Gimenez, O. Content analysis of newspaper coverage of wolf recolonization in France using structural topic modeling. *Biol. Conserv.* **2018**, *220*, 254–261. [CrossRef]

39. Bail, C.A. Cultural carrying capacity: Organ donation advocacy, discursive framing, and social media engagement. *Soc. Sci. Med.* **2016**, *165*, 280–288. [CrossRef] [PubMed]

40. Mishler, A.; Crabb, E.S.; Paletz, S.; Hefright, B.; Golonka, E. Using Structural Topic Modeling to Detect Events and Cluster Twitter Users in the Ukrainian Crisis. In *HCI International 2015-Posters' Extended Abstracts. HCI 2015. Communications in Computer and Information Science*; Stephanidis, C., Ed.; Springer: Berlin/Heidelberg, Germany, 2015.

41. Kuhn, K.D. Using structural topic modeling to identify latent topics and trends in aviation incident reports. *Transp. Res. Part C Emerg. Technol.* **2018**, *87*, 105–122. [CrossRef]

42. Moro, S.; Cortez, P.; Rita, P. Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Syst. Appl.* **2015**, *42*, 1314–1324. [CrossRef]

43. Roberts, M.E.; Stewart, B.M.; Tingley, D.; Lucas, C.; Leder-Luis, J.; Gadarian, S.K.; Albertson, B.; Rand, D.G. Structural Topic Models for Open-Ended Survey Responses. *Am. J. Polit. Sci.* **2014**, *58*, 1064–1082. [CrossRef]

44. Thelwall, M.; Buckley, K.; Paltoglou, G. Sentiment strength detection for the social web. *J. Am. Soc. Inf. Sci. Technol.* **2012**, *63*, 163–173. [CrossRef]

45. Silge, J.; Robinson, D. *Text Mining with R: A Tidy Approach*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2017.

46. Taboada, M.; Brooke, J.; Tofiloski, M.; Voll, K.; Stede, M. Lexicon-based methods for sentiment analysis. *Comput. Linguist.* **2011**, *37*, 267–307. [CrossRef]

47. Xu, L.; Pan, R. The Construction of Sentiment Ontology. *J. Chin. Soc. Sci. Tech. Inform.* **2008**, *27*, 6. (In Chinese)

48. Roberts, M.E.; Stewart, B.M.; Tingley, D. STM: An R Package for Structural Topic Models. R package version 1.3.3. *J. Stat. Softw.* **2019**, *91*, 1–40. [CrossRef]

49. Schwemmer, C. Stminsights: A 'Shiny' Application for Inspecting Structural Topic Models. R package version 0.4.0. 2015. Available online: https://github.com/cschwem2er/stminsights (accessed on 17 November 2020).

50. Silge, J.; Robinson, D. tidytext: Text Mining and Analysis Using Tidy Data Principles in R. *J. Open Source Softw.* **2016**, *1*, 3. [CrossRef]

51. Xu, Z. Personal stories matter: Topic evolution and popularity among pro- and anti-vaccine online articles. *J. Comput. Soc. Sci.* **2019**, *2*, 207–220. [CrossRef]

52. Kuang, W. Countermeasures Against New-Media Public Opinion. Available online: https://link.springer.com/chapter/10.1007/978-981-13-0914-4_11 (accessed on 15 November 2020).

53. Zhang, L.; Wei, J.; Boncella, R.J. Emotional communication analysis of emergency microblog based on the evolution life cycle of public opinion. *Inf. Discov. Deliv.* **2020**, *48*, 151–163. [CrossRef]

54. Monahan, B.; Ettinger, M. News Media and Disasters: Navigating Old Challenges and New Opportunities in the Digital Age. In *Handbook of Disaster Research. Handbooks of Sociology and Social Research*; Rodríguez, H., Donner, W., Trainor, J., Eds.; Springer: Cham, Switzerland, 2018.

55. Gesualdo, F.; Marino, F.; Mantero, J.; Spadoni, A.; Sambucini, L.; Quaglia, G.; Rizzo, G.; Sahinovic, I.; Zuber, R.F.L.; Tozzi, A.E. The use of web analytics combined with other data streams for tailoring online vaccine safety information at global level: The Vaccine Safety Net's web analytics project. *Vaccine* **2020**, *38*, 6418–6426. [CrossRef] [PubMed]

56. Wang, G. Legitimization strategies in China's official media: The 2018 vaccine scandal in China. *Soc. Semiot.* **2020**, *2020*, 1–14.

57. Zhou, M.; Qu, S.; Zhao, L.; Kong, N.; Campy, K.S.; Wang, S. Trust collapse caused by the Changsheng vaccine crisis in China. *Vaccine* **2019**, *37*, 3419–3425. [CrossRef] [PubMed]

58. Hu, D.; Martin, C.; Dredze, M.; Broniatowski, D.A. Chinese social media suggest decreased vaccine acceptance in China: An observational study on Weibo following the 2018 Changchun Changsheng vaccine incident. *Vaccine* **2020**, *38*, 2764–2770. [CrossRef]

59. Okuhara, T.; Ishikawa, H.; Okada, M.; Kato, M.; Kiuchi, T. Contents of Japanese pro- and anti-HPV vaccination websites: A text mining analysis. *Patient Educ. Couns.* **2018**, *101*, 406–413. [CrossRef]